Dynamical Neural Networks: modeling low-level vision at short latencies

Laurent Perrinet
E-Mail: Laurent.Perrinet@incm.cnrs-mrs.fr,
WWW: http://www.incm.cnrs-mrs.fr/LaurentPerrinet,
DyVA team, INCM/CNRS (UMR 6193)
31, ch. Joseph Aiguier, 13402 Marseille Cedex 20, France.

April 26, 2007

This review is available as a chapter [Perrinet, 2007], pages 163–225, in

- *Topics in Dynamical Neural Networks: From Large Scale Neural Networks to Motor Control and Vision*, volume 142 of *The European Physical Journal (Special Topics)* by Bruno Cessac, Emmanuel Daucé, Laurent U. Perrinet, and Manuel Samuelides [Cessac et al., 2007]
- URL http://www.incm.cnrs-mrs.fr/LaurentPerrinet/Publications/Perrinet06.
- URL http://dx.doi.org/10.1140/epjst/e2007-00061-7.

Note: this document is an up-to-date version of the document referred above. The differences between this on-line version and the published one may be found on the author's website.

Abstract

Our goal is to understand the dynamics of neural computations in low-level vision. We study how the substrate of this system, that is local biochemical neural processes, could combine to give rise to an efficient and global perception. We will study these neural computations at different scales from the single-cell to the whole visual system to infer generic aspects of the underlying *neural code* which may help to understand this cognitive ability. In fact, the architecture of cortical areas, such as the Primary Visual Cortex (V1), is massively parallel and we will focus on cortical columns as generic adaptive micro-circuits. To stress on the dynamical aspect of the processing, we will also focus on the transient response, that is during the first milliseconds after the presentation of a stimulus.

In a generic model of a visual area, we propose to study the neural code as implementing visual pattern matching, that is as efficiently inverting a known model of image synthesis. A possible solution offered by the architecture of the visual pathways could be to represent at first on the surface of the cortical area how well the prototypical visual features are matched by a combination of inferential mechanisms as *ideal observers*. We studied the efficiency of this representation by rating the statistics of the output using natural scenes, that is scenes occurring frequently. We show that this may be finally used to provide a behavioral output such as an eye movement.

However, constraints specific to the visual system imply that the set of prototypical features is not independent and that the cortical columns should communicate to produce an efficient, sparse solution. We will present efficient algorithms and representations based on the event-based nature of neural computations. By explicitely defining this efficiency, we propose then a simple implementation of *Sparse Spike Coding* using greedy inference mechanisms but also how the system may adapt in a unsupervised fashion. These computations may be implemented in simple models of neural networks by explicitly setting the lateral connectivity between populations of columns. Using natural scenes, this algorithm provides a model of V1 which exhibit prototypical properties of neural activities in that area. We show simple applications in the field of image processing as a quantitative method to evaluate these different cortical models.

Keywords: *Neuronal representation, over-complete dictionaries, inverse linear model, distributed probabilistic representation, spike-event computation, efficient coding, Matching Pursuit, Sparse Spike Coding, Sparse Spike Learning.*

PACS: 87.19.Dd: Information processing, in vision, 07.05.Mh: Neural networks, 42.30.Sy: Pattern recognition, 07.05.Pj: Image processing, 02.50.-r: Probability theory, stochastic processes, and statistics, 02.50.Tt: Inference methods, 42.66.Si: Psychophysics of vision, visual perception.

Contents

What is the substrate of our cognitive abilities? How much do we differ from other forms of intelligence, such as animals or computers? How do we adapt in harmony with the outside world and other individuals? And how can we interact with the brain to cure pathologies from neuro-degeneration to autism? These broad questions still have no clear answer today and scientists agree that we still stand at the Middle Age of our knowledge of the brain.

A major challenge in neuroscience is to understand the content of the activity that is observed in biological neurons. This neural activity constitutes the link between the structure of the nervous system and the function of our cognitive abilities. However, these complex activity patterns that are the basis of our cognitive abilities remain a mystery and there is yet no known unifying model explaining the "language", or *neural code*, that could be eventually used by neurons at the various scales of the central nervous system.

Herein, we will approach this challenging question by focusing on the particular efficiency of low-level vision. In fact, vision is an apparently simple cognitive ability, which has always been a very active research area at the convergence of neurophysiology, psychology and computational neuroscience.

1 Neural computations in low-level vision

Along the history of cognitive and mind sciences, metaphors were used to help us understanding the mechanisms generating our thoughts. In the 18^{th} century for instance, painting dominated the cultural world and the mind was imagined as a painter depicting ideas, thus using the thoughts as brush strokes. Similarly, the brain was depicted in the century of industrialism alternatively a steam engine, a precise clock or an automaton. The major metaphor is nowadays the computer and many neuroscientists take this analogy as a source of inspiration to understand neural computations. As we will show, this metaphor has shown limits and new candidates emerge with the raise of distributed communication systems — such as the omnipresent InterNet and more generally to multi-scale networks, *e.g.* communication, political or social nets [Barabasi and Bonabeau, 2003]. We may therefore think of the emerging rules of these networks —such as a crowd in an urban area— as a novel metaphor of the dynamics of the brain.

We are interested in this paper in understanding the processes that give raise to our cognitive abilities and we will focus on low-level vision in humans. In particular, we will review some of the dynamics of the electrochemical signals occurring in the brain —especially the brief peak of activity known as action potentials or *spikes*— and their link to the efficiency of neural computations in populations of neurons (see also the paper of Bruno Cessac in the same issue for a more detailed account on spike dynamics). Following the particular example of a visual neural implant as a full-featured model of low-level vision, we will put at stake the metaphor of the computer to describe the algorithms that may take place in the brain. We will review some facts on neural computations and on vision and some new insights into the dynamical processing of neural information will help us to propose some alternatives to the computer architecture that are more adapted to mimic and understand the brain.

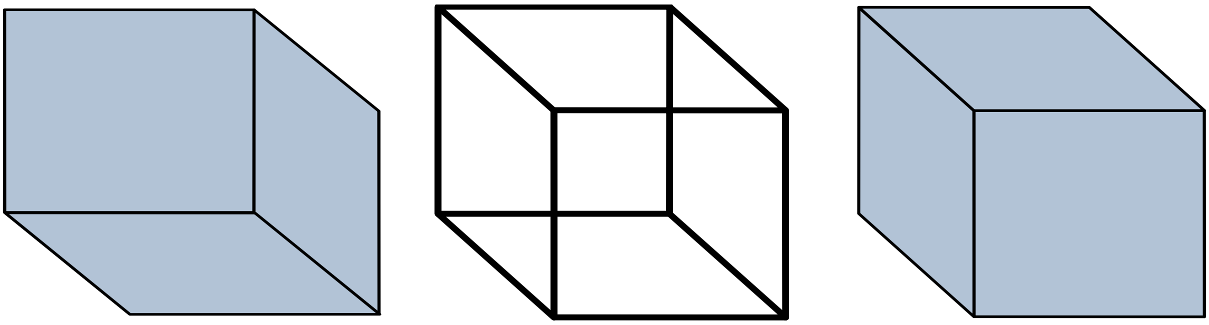

Figure 1: **Ambiguity of visual information.** To illustrate the ambiguity of visual stimuli, we show how a simple visual stimuli could generate different interpretations. When looking at the central cube, the most probable interpretations oscillate between two configurations of a plain and solid cube, as seen on the left and right of the figure. One of these configurations may be favored (e. g. by masking with the hand the other one), thus changing the context. Theoretically, an infinite set of different polyhedrons, of solids or networks of lines could have given raise to the stimulation (such as this flat figure printed on paper or elongated "cubes"). We argue that the visual system assigns probabilities to all these possible and ambiguous configurations knowing the context to finally choose the most appropriate solution.

1.1 What's special about the brain?

1.1.1 Efficiency of low-level vision

Vision is our ability to extract information from the luminous energy reflected by our environment by using the *image*[1] focused by the eye on the retina[2]. In fact, pushed by evolution's pressure, a majority of animals have developed the ability of seeing. This is particularly proeminent in primates which devote a large proportion of their brain resources to visual processing (in proportion of cortical tissue, approx. 52% in monkeys and 27% in humans [Orban et al., 2004]) and allow for that quick and reliable perception of the visual world. The information is processed in the Central Nervous System (CNS), that is the system constituted by the brain and the spinal chord, until it may reach a behavioral decision. But this task of statistical pattern recognition (or "feature detection") proves to be very difficult because a striking aspect of the visual information is the inherent ambiguity of low-level images. This is well illustrated in the rich collection of visual illusions, such as the Necker cube (see Fig. 1), proving that a given retinal stimulation may correspond to a multitude of different configurations. These *ambiguities* are omnipresent in natural images (*i.e.* images that are behaviorally relevant) and arise for instance from occlusions, transparencies and conflicting 3D features. However, our daily experience shows us that our visual system resolves these ambiguities —and in most of the time in an effortless and unconscious fashion— quickly and reliably. However complex these mechanisms may appear, they are particularly robust in biological vision. In fact, this system remains efficient during development from

[1] We will use the term image in his generic meaning of a topological map of —possibly vectorial— values. A PNG or even compressed JPEG image being particular architectures of images.

[2] Note that the retinal image is inverted optically by the eye's lens, both upside-down and left-right.

a newborn child to adulthood despite the changes in our environment such as the growth of our body, which implies drastic changes of configuration. In fact, the genetical material is not sufficient to code for all the architecture of the visual system, and adaptive strategies should exist to tune the system according to the visual function. This *plasticity* is illustrated by artificial changes of configuration, such as when using continuously during several weeks transforming glasses such as prisms. After some time, the world is perceived such as it was before this change : the system adapted so as to come back to a perception which is coherent with the outside world[3] [O'Regan and Noë, 2001]. However, this system is relatively fragile to particular pathologies such as age-related macular degeneration (ARMD) and a clearer understanding of the mechanisms underlying the robustness of neural mechanisms could provide solutions, such as efficient alternatives for these different cases of blindness.

1.1.2 Shortcomings of the sequential computer metaphor for vision

The situation is different with present models of cognitive abilities which usually follow the computer metaphor. In fact, for cognitive tasks such as vision which does not require an expert's knowledge (in opposition to playing chess, inverting a matrix, searching in a database), computer solutions to vision so far are ineffective either by their lack of functionality or by their energy consumption. For instance, autonomous retinal implants at present may only process a few pixels or else they will produce too much heat (by their own power consumption) for their practical use on the retina. Applications of automated visual systems based on sequential computers (such as desktop computers) are today still seldom and limited to simple tasks (such as segmentation or intrusion detection). More importantly, they apply only to a very limited range of situations which exclude their use in natural conditions (for instance when changing the context). More complex applications of feature detection, for instance the automated processing of satellite images, are very demanding in computer power and mobilize the most powerful computers in the world such as the most efficient solution available in 2006, IBM's BlueGene/L[4]. This system consumes 2 MW while the visual system, as a part of the central nervous system, uses only a fraction of the estimated 20 W used by the brain[5]. Let's explore what may be the key difference that makes computers still so ineffective compared to brains.

First, computers have historically privileged *sequential* algorithms for the resolution of experts' problems. Today, most computer solutions use the architecture inherited from the Van Neuman computer which may be linked to the early computers engineered by Blaise Pascal or Charles Babbage and formalized by the one-tap universal Turing machine. However, sequential processing is different from parallel processing since some computations are done without knowing the state of some

[3] More surprisingly, when removing this perturbation, the system as to adapt (though on a shorter time scale) to go back to the original configuration. It has also been reported that in case of the reversing glasses, text and characters still appeared reversed after adaptation. This is arguing for a modularity of the visual systems capability, since positioning the body in the visual world is somewhat separated from reading.

[4] Such as found on `http://www.top500.org/`.

[5] In humans, we may therefore estimate that the first levels of the visual system consume of the order of 5 W.

variables and with many different simultaneous threads[6]. As a metaphor, on one side verbal communication is sequential: it is difficult for us to hear different conversations at the same time. This type of processing therefore relies on a sequence of processes. On the other side, the visual system implements a huge memory by its structure and by the interaction of a multitude of parallel events allowing to process different qualities (shape, motion, identity) of the image in parallel and finally to give a result at any time.

Secondly, unlike computers, we saw that the central nervous system is adaptive and may *learn* new configurations according to given functions. In particular, most classical algorithms use knowledge that is clearly defined from the programmer (that is for instance the collection of all famous chess moves in a chess simulator) and rarely adapt to interpret and include new knowledge. For instance, most feature detection systems assume that the features may be found at any position, and therefore that detection is translation invariant. In the brain, though, this knowledge is learnt by the stability of a corresponding physical law in time (namely of the translation of objects) and is not implemented *a priori* on the visual areas. However, there is no unified theory for unsupervised learning (that is without any "teacher"). As the Baron Münchhausen, who was able to lift himself out of the sea by pulling himself up by his own boot straps[7], the system may autonomously emerge from the interaction of the visual system with the outside world.

1.1.3 Computational neuroscience & new computing paradigms

We saw the superiority of neural computations in handling cognitive tasks such as vision: the response is quick and reliable and at a longer time scale it is also robust and adaptive. However, in order to understand neural computations and their efficiency, we are bound to explore the properties of the brain as a "black box": we don't know *a priori* its mechanisms. As stated by Marr [1982], this may be understood by studying the different levels of the problems a neural function has to face : a definition of the function (computational level), a notion of the form it takes (algorithmic level) and eventually the generic rule that implement it (implementation level). This method may be approached by using an integrated methodology: by interacting with neuroscience (as neuro-physiology or psychology), we may propose models which can then be validated by using simulations to a wider range of neuroscientific data. This methodology for finding an eventual "neural code" defines the goal of *computational neuroscience*[8].

At present, many computational neuroscientists claim here that the set of neural computations that give raise to cognitive abilities such as vision may be understood as generic rules and are expressed in the behavior of populations of neurons. In the following, we will describe generic principles of neural computations and then in particular in the visual system. In this view, the brain is constituted by a multi-scale network of different modules (or micro-circuits) which have a common generic structure but that each adapt their interactions according to the function to perform.

[6] It should be noted that parallel architectures are thus necessarily dynamical systems.

[7] This tale seems to be at the origin for the terminology of the *bootstrap* theory.

[8] There is however some confusion in the community, leading to an understanding of computational neuroscience as using computers as a tool of analysis in neuroscience rather than a theory of computations in neural systems. The term "theoretical neuroscience" [Dayan and Abbott, 2001] stands then in this perspective as an alternative but is certainly inadequate from its etymology.

The success in modeling this "neural code" is then validated by studying how well the implementation in the neural activity (the structure of the code) corresponds to the particular cognitive ability that is being modeled (its function).
Going further in the direction of David Marr, we may also state that to understand this "black box", we have to understand the function it provides. We may then propose respectively an algorithm and an implementation of this function. One should however note that this methodology is by no way finalist: more efficient functions emerged blindly against other functions from natural selection since they are more fit for survival and therefore of being present in the future. This observation applies also for the learning mechanisms: a learning algorithm that will adapt at best to the environment and that may replicate itself at best will therefore have a better chance of survival in its progeny. In fact, models at the different levels of the network may use a high number of free parameters and constraining the function and task in a sub-system of the model allows to constrain some of these parameters. Neural computations may therefore be understood at different spatial time scales in terms of an optimization problem.

1.2 The different scales of neural computations

To define this hypothetical neural code, we will review some generic properties of the CNS, from its basic elements to the whole system. These results have been revealed using a multitude of techniques which permits the observation of neural anatomy and activity at different spatial and temporal resolutions.

1.2.1 Are neurons and spikes the elementary bricks of neural computations?

In fact, since the seminal work of Cajal [1911] which closely followed the technical advances in staining neural tissue from Golgi, we know that the brain is not a continuous medium, but that is rather a dense network of specialized cells, the *neurons*. Neurons[9] are specialized cells in the CNS which are supposed to form the substrate of neural computations. These may be distinguished in three classes : sensory neurons (such as the photoreceptors neurons), then motor neurons which are connected to muscles and the vast majority of remaining neurons, the associative neurons which are neither sensory nor motor. In fact, the brain is constituted by approx. 10^{12} neurons which are supported by approx. 10 fold more supporting cells, the glial cells, the role of which is still not completely determined. Neurons vary in size from 4 μm to 100 μm in diameter and from a fraction of a cm to a meter in length. Neurons are very diverse but are prototypically constituted of a cell body, the soma, and of an arborization of their membrane which take different shapes, thanks to an inner architecture (cytoskeleton) modulated by the synthesis of spiraling polymers, the micro-tubules. Anatomically, dendrites are arborized and closer to the soma, integrating signals from all its surface and the axon is on the other hand most frequently elongated and terminates far from the soma.
The membrane of the neuron is covered with a variety of ionic gates whose goal is to modulate the ionic concentrations inside the cell and therefore the electrical potential through the membrane. In addition to the passive propagation along

[9] For a more precise account on neurons and spikes see the paper of B. Cessac in this issue.

the membrane, these ionic gates may actively modulate the propagation of this potential along the membrane. These variations of electrochemical potential may be propagated to other cells (neurons or muscles fibers) and the morphology and direction of this propagation separates functionally the arborization between a receiving dendritic tree and an emitting termination, the axon. This propagation is denoted respectively feed-forward, in the direction from sensory neurons to motor neurons, feed-back in the opposite direction or lateral if no preferred direction may be determined. With their supporting cells, neurons may be considered as elementary bricks of the CNS where the information is processed in its arborizations by variations of its membrane potential.

The neuronal connections occur at special contact points called *synapses*. They are very dense, approximately 10^4 per neuron (and thus there is an estimated total of 10^{16} synapses in humans) and transmit the signal from neuron to neuron through direct electrical junctions (the so-called *gap junctions*) or more frequently through chemical synapses. These chemical contacts are mediated through the release of vesicles of neurotransmitters (in the milli-second temporal scale) the type of which determines specifically the quality of the connection both pre- and post-synaptically. Finally, they are highly plastic and may change in morphology and strength in short time scales, changing locally the response characteristics of the neuronal circuitry, as will be discussed in Sec. 3.2. Synapses thus constitute the locus of information transmission between neurons and their plasticity plays a major role in the adaptation of the neural computations.

In certain conditions, the raise of the membrane potential may drive the neuron to an all-or-nothing state: the Action Potential (AP, or *spike*). Depending on the local morphology of the membrane and of ion channels, the propagation mechanism may enter a positive feed-back, thus an "exploding" state, which elicits a rush of ions and produces a brief (approximately 1 ms) change of membrane polarity. This 'spike' propagates then along the membrane, until it reaches synapses where it is often necessary for the release of neuro-mediator, hence the name action potential. Even if spikes may also occur in dendrites, this signal is particularly adapted to the propagation of all or none signals along the axon. This is particularly true for *myelinated axons* for which a sheath of specialized glial cells permits the fast and robust propagation of spikes (90 m/s in sheathed axons vs. less than 0.1 m/s in unsheathed axons). In general, the raise of potential occurs when excitatory post synaptic potentials converge from the branches of the dendrite and cross a certain threshold. In general, one may observe that the spike will be released faster if the driving input is faster. Moreover, the spiking signal is a robust, timely precise [Bair and Koch, 1996] and reproducible [Mainen and Sejnowski, 1996] signal. In the framework of single-neuron computing [Koch and Segev, 2000], knowing the exact mechanisms of neuronal integration and transmission and given the current state of all neurons, one may predict that a sensory input will propagate along axons, be integrated analogically through the synapses and then integrated by the dendrites of further associative neurons [Koch, 1998]. This chain will generate a dynamical flow of spikes until it reaches motor neurons, thus closing the loop from perception to action. Spikes —associated with corresponding neurons— appear to be thus the unit of neuronal information and one could understand a neuron with a label corresponding to a set of preferred stimulations it is selective to.

1.2.2 Cortical columns

However, the situation is not so simple and in contrast with what was often assumed in theoretical models of neural networks there is a great influence from the context in the network. This is implemented by the feedback and lateral links that occur in the network at different scales and which thus form recurrence loops of signal propagation, thus defining indivisible populations of neurons, or cell assemblies [Hebb, 1949] . Actually, if we consider that the low-level system has evolved to form an efficient canal of information, one could see the population of cells as a democratic network where every cell tries to maximize its metabolism by adapting his configuration. Neurons are therefore in a constant competition to produce the highest activity (and the highest metabolism). But at the same time, as in a theoretically democratic society, a neuron that would monopolistically transmit all information would lower the overall efficiency of the whole assembly. Thus, neurons have to cooperate to maximize the global metabolism of the assembly in the long term. In this view, the smallest scale for describing the coding of the information is not on individual neurons but seems rather to lie on the scale of cell assemblies and that these constitute local circuits based on generic rules by implementing efficiently tuned cooperation (*i.e.* homeostatic) and competition rules. Economical laws are thus more appropriate than logical ones to describe neural computations.

A prototypical example of cell assemblies is given by the columnar structure of the cortex which will be of particular interest in this paper. In fact, the cortex reveals paradoxically a wide diversity of functionalities but a similar repetitive structure. It is anatomically constituted by a laminated surface (approx. 2 mm deep in total) with a vertical organization of bundles of approx. 110 neurons[10] —a *cortical column*— receiving and emitting spikes to the rest of the CNS or from other columns. These columns (of a diameter of approx. $20 - 60\,\mu$m) have vertically a similar 6-layer organization across the cortex [Mountcastle, 1998] : the middle layer (granular or layer IV) receives input from the thalamus, relays it to the upper layer (supra-granular). Cortico-cortical connections run then horizontally or through the white-matter for far-ranging interactions. This layer is connected to the deep layers (infra-granular) which connect back to the thalamus (layer VI) or directly to the basal ganglia and the spinal cord (layer V). As we saw, the propagation of information takes more time with distance and by economy we may predict that inside an area the columns in cortical areas will be arranged so as to minimize the time of local cooperation by arranging similar features on neighboring cortical locations (as will be implemented in Sec. 3.3.4). This principles is generally observed in biology and is at the base of Self-Organizing Maps [Kohonen, 1982]. The approx. 600 million cortical columns of the human cortex therefore constitute a network of micro-circuits with an apparently generic architecture.

However, it is not clear if neural computations may be reduced to the scale of the cortical column (for a review, see [Horton and Adams, 2005]). In fact, depending on the function of the cortical area, the code produced by this population of neurons may have different characteristics. In the cortical motor areas, for instance, the infra-granular layers are thicker and the average output firing rate of a population of columns correlates well with the action actually being done [Georgopoulos et al.,

[10] This number is stable in mammals except a count of around 260 in the primary visual cortex of primates.

1986]. In low-level perceptual areas on the other side, the activity of neighboring columns may have drastic and highly non-linear effects and a column activated by a preferred stimulus may be excited or inhibibited depending on its surround (see [Ringach, 2002] and section 3.3.3). In fact, cortical columns may often be horizontally organized in assemblies or *hyper-columns* (of a diameter of approx. $0.4 - 1.0\,\mathrm{mm}$) and seem to interact by cooperation and competition rules therein. Populations of neurons or of mini-columns interact therefore recursively and non-linearly, and this behavior may be understood functionally as an enhancement of the efficiency in terms of information processing.

1.2.3 The CNS as a dynamical architecture

We may extend this analysis to the scale of the CNS to reveal a complex modular architecture. Anatomically, the CNS may be subdivided in different parts which correlate with their apparition during evolution and also with the development of the embryo. Respectively, the oldest parts are the spinal cord (over 450 million years old, controls movements of limbs and trunk), the brain stem and the midbrain (approx. 200 million years old, which receive sensory information and control most vital functions), and the diecephalon which contains the thalamus (where most sensory information converges) and the hypothalamus (which seems to modulate emotions and the homeostasis of brain activity). More recently in evolution, appeared the cerebral hemispheres with in their center the hippocampus (or archeo-cortex) and in the deep layers the basal ganglia (which regulates motor performance). On its surface, lies a thin layer wrapped in the skull on top of the midbrain : the cerebral cortex (or neocortex, aprox. 0.4 million years) which is supposed to play a major role in our cognitive abilities (from perceiving to reasoning). The CNS also includes the cerebellum on the dorsal part of the midbrain and the function of which is to coordinate planning, timing and patterning of motor responses at the interface with the cerebral hemispheres [Kandel et al., 2000, p. 292]. This introduces a hierarchy in the CNS from the most primitive to earlier modules which are built successively on the top of the oldest.

As was stated before, information flows in parallel in the architecture along feed-forward, lateral but also in feed-back pathways. First, the propagation on the membrane of neurons takes time and this parallel architecture implies the existence of a dynamical hierarchy in the CNS depending on the latency of an area compared to an other [Bullier, 2001]. In fact, a sound evolutionary strategy is to propagate the most relevant information first and this dynamical architecture must therefore minimize in the architecture of the CNS the delay between the sensation and a behaviorally relevant response. This constraint on the shortest path will be naturally driven by the hierarchy of the feed-forward links between the modules of the CNS. Secondly, the lateral and feed-back loops may explain mechanisms of memory or of a progressive maximization of the efficiency of the activity. This interpretation may therefore explain the different rhythms that may be observed during different cognitive tasks [Freeman and Barrie, 1994; Rodriguez et al., 1999] but also more generally the relative synchronization of large groups of neurons [Fries et al., 2002]. It is therefore necessary to consider the role of lateral and feed-back connections to understand dynamical properties of perception such as temporal masking or multi-stability [Logothetis et al., 2001]. In particular these studies show that the dynamics of this network may correlate with cognitive abilities and thus seem to give hints on the neural code at the scale of the CNS [Jirsa, 2004].

Anatomical and functional observations reveal that the cortical surface is itself divided in modules —or *cortical areas*— corresponding to different functions which are themselves organized in association maps. In the human cortex, it may be subdivided anatomically in 52 areas (categorization from Brodmann [1909]), which are themselves assembled in 32 functional areas. These areas are interconnected by 232 reciprocal cortico-cortical connections[11]. The cortex is a highly convoluted surface in humans and when flattened, it may be considered as a surface (approx. 1.4 m^2 in humans) which consists of sensory areas, associative areas, the somato-sensory areas and the prefrontal cortex. The function of these areas may be described by exploring the general condition they are specifically selective to (for instance, to all auditory signals or speech)[12]. The cortex is therefore a complex network of maps interacting to transform sensations into a dynamical flow of spikes which may generate actions. Dynamics of neural computations should therefore be studied at all these rather different scales : the neuron, the cortical area, the CNS (see Tab. 1 for a synthesis).

spatial scale	**unit**	**temporal scale**	**relevant technique**
0.001 mm-0.01 mm	synapses	2 – 5 ms	electrode
0.1 mm-0.1 mm	neurons	10 ms	electrode
1 mm	columns	10 – 50 ms	electrodes - LFP
10 mm	hyper-columns	20-50 ms	optical imaging
10 mm 100 mm	cortical areas	ms	optical imaging
1000 mm	CNS	110 ms	EEG / MEG / fMRI

Table 1: **Scales of neural computations** This table summarizes the scale order in temporal and spatial domains for neural computations in human for the different bricks that we extracted in Sec. 1.2. It lists some particularly relevant techniques used in neuroscience to study that particular scale.

1.3 Properties of the low-level visual system

The prevalence of the study of vision correlates with its importance and relative size in the brain. It constitutes a major theme in neuroscience and the amount of literature makes it the best known cognitive function. The general properties of the CNS that we reviewed above showed some general principles that apply to the visual system and we will now focus on the particularities of the low-level visual system (by definition the part of the CNS which is specially devoted to vision) at the different scales of description.

[11] For these connections, an important and rather unexplained feature is that the time of upstream in the hierarchy is equal to the time of downstream.

[12] One should note that this popular view and caricatured in phrenology, is in mathematical terms intractable, since one may only guess from the neural response the selectivity to a limited range of the presented stimuli (which are themselves often chosen according to this guess). However, an actual response may take different forms and be selective to a wider range of stimuli and interactions of stimuli in a non trivial way, leading to a biased description of the function of neuronal structures [Olshausen, 2004]

1.3.1 Hierarchical and dynamical architecture of the low-level visual system

As seen above for the whole CNS, we may describe a hierarchy in the visual system going from sensation to action. We may review it by studying the propagation of visual information elicited by the presentation of an initial image between two saccades or blinks of the eyes[13] or more precisely by a "flashed" image to reveal the transient dynamic of the network [Bullier, 2001]. First, the retina captures the luminous information that hits the back of the eye. The retina consists of approx. 10^8 neurons, among them 4 million photo-receptors, and one million output neurons, the ganglion cells, whose axons form the optic nerve. These are of several types, in majority M and P cells : the 10^5 magno-cellular (in short *M*) cells which are quick (responding after approx. 10 ms) but at a lower spatial resolution (coarse coding of up to 100 rods) and the 8.10^5 parvo-cellular (or in short *P*) cells (after a latency of over 30 ms) which have a better spatial resolution (up to a one to one connection with photoreceptors). These cells convey information about the local contrast in the image and is relayed by the optic nerve to the Lateral Geniculate Nucleus (LGN), an olive shaped nucleus of the thalamus which keeps the M and P cells segregated into its different layers and adds a delay which is the sum of the propagation time (from 2.5 ms to 10 ms, depending on the fiber) and an integration time depending on the stimulus (see Fig. 2). This nucleus concentrates the luminous sensory input and separates rough information (such as the information used to regulate the circadian rythm) while the visual information heads toward the cortex.

This information is then conveyed by the optic radiations through the white matter and enters the thick granular layer of the *Primary Visual Area* (labelled V1 in humans; the granular layer is easily observable anatomically, hence the name striate cortex). The information from the magno and parvo cells still keeps segragated and enters after resp. approx. 40 ms and 60 ms (Fig. 2-Middle) in separate sub-layers of the granular layer. V1 is situated in the occipital lobe and is the convergent area for up-stream and down-stream visual information in the visual system's hierarchy, thus forming a "black-board" for all visual information. As suggested by Marr [1982], the function of this cortical area could be to form a synthetic representation of the visual information gathered from the converging visual input, as an elaborate sketch of the visual world, relatively independant of some crude lighting conditions such as the global luminance or contrast. This area could also serve to segregate the figure from the ground [Bullier, 2001].

After V1, the visual information branches in two pathways which originate from the M and P pathways and gradually represent more "abstract" visual features. First, the dorsal pathway is particularly sensitive to the position and motion of objects, hence the name of "Where" pathway: the heavily myelinated path from V1 to mediotemporal area (MT) leads to fast activation of this area selective to local velocities (the earliest latencies lying at 45 ms). The major part of this pathway naturally originates from the fast cells of the magno-cellular pathway. On the other hand, the ventral pathway (areas V4 to the infero-temporal areas) shows a selectivity to the identity of features or objects and is progressively independent to object position or motion, hence the name of "What" pathway. Moreover, there exist also cross-links in the low-level visual system between cortical areas which enable

[13] To keep the description simple we will abstract binocular and color cues and the eye will be considered as fixed but also micro-saccadic movements [Martinez-Conde et al., 2000].

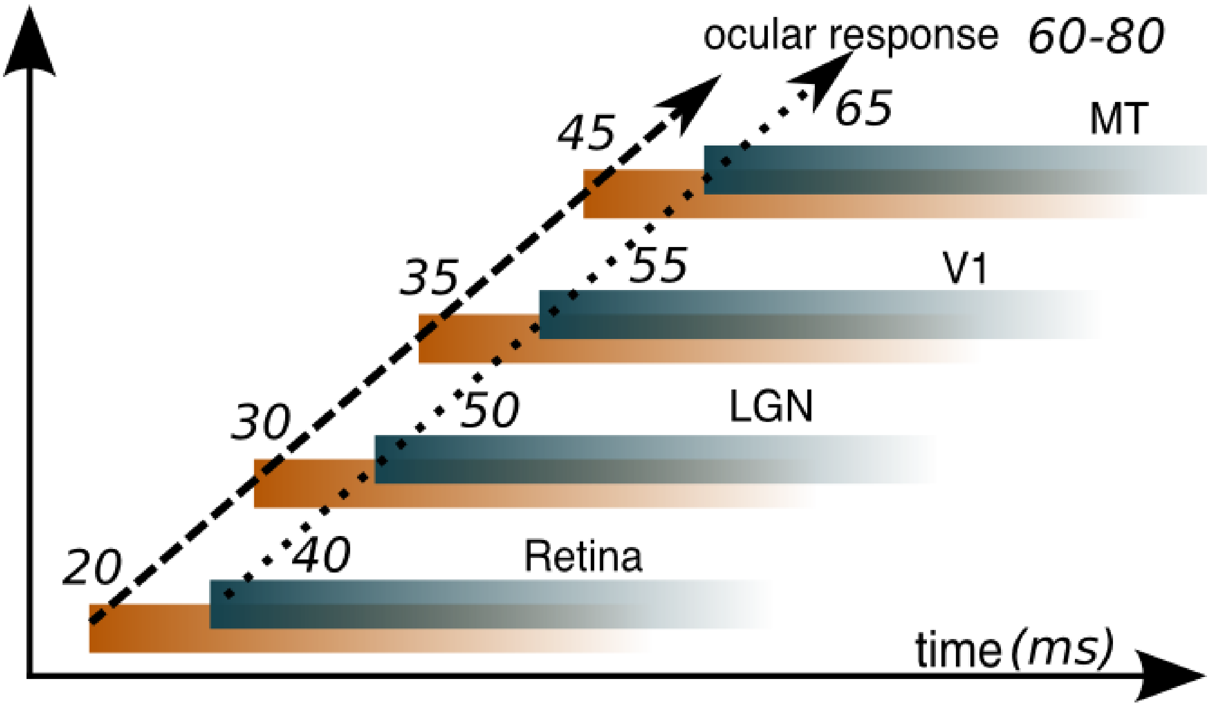

Figure 2: **Simplified dynamical view of the visual "Where" pathway.** When presenting a full contrast image, the corresponding information flows from the retina to higher visual areas to finally reach an eye movement. The time of propagation of information builds a parallel and hierarchical order of activation of the different visual areas until a —possibly progressive— decision is taken. Indicative latencies are given in milli-seconds for the macaque brain (numbers in italics). At the activation of an area, the information may be fed-back to a lower-level area with a similar propagation time. The dashed line correspond to the early "Magno pathway" —that is to the earliest possible latencies reached by the information transmitted through the M cells— while the dotted line corresponds to the later "Parvo pathway". It should be noted that until an eye movement decision is made, the oculo-motor system is still in an open-loop phase.

the sharing of inter-modal features and, as time goes, the anatomical separation between the M and M+P pathways progressively vanishes (for a review, see [Salin and Bullier, 1995]). This distinction remains functionally and we may consider the visual pathways as a "fast brain" driven by M cells and then a finer pathway which allows for a progressively more reliable output [Bullier, 2001]. The visual system thus consists in a dynamically functional hierarchy which progressively transforms the retinal information in more abstract maps progressively losing the retinotopical information.

1.3.2 Receptive fields

A major breakthrough in our knowledge of the visual system was the discovery of the specific response of neurons in these cortical areas [Hubel and Wiesel, 1959; Henry et al., 1974]. In fact, by exploring the firing rate response of neurons of the primary visual cortex (V1), they observed that most neurons responded preferently to local oriented edges, thus bridging for the first time the presentation of visual features with a selective cortical activity. By moving the measurement electrode in V1, the preferred orientation evolved as they moved tangentially to the cortical surface, forming a series of interdigidated stripes on the surface of V1. However, the corresponding preference did not change when moving perpendicularly to the surface [Mountcastle, 1957; Hubel and Wiesel, 1962] (as we mentioned in Sec. 1.2.2). For any neuron, the corresponding field of visual space that helped to change significantly the response of the neuron is called its *Receptive Field* (RF)[14] and Hubel and Wiesel [1968] proposed a model to explain these results, for which neurons integrate information on their RF and that the final response, mediated by the mean firing rate, is non-linearly transformed to match neurophysiological observations [Carandini et al., 1997]. This popularized the view of neurons as simply matching templates field [Barlow, 1972] and that, by extending this theory to higher levels, one neuron could be preferentially activated by the image of one own's grand-mother image, hence the nick-name of a "grand-mother cell".

However, this theory is not sufficient to describe all observations and tends to be incomplete or to reflect a distorted description of natural conditions. First, the experimental conditions tend to remove arbitrarilly "outliers" and to use non-natural stimuli [Olshausen, 2004], leading to an over-estimation of the overall mean firing rate and of a biased sampling of neurons' properties in V1. Secondly, as was suggested in Sec. 1.2.2, dynamics of the signal on the RF of a column may be influenced by the context and this dynamic may be a necessary condition for the emergence of a percept [Jancke et al., 2004]. In particular, there is an intracortical propagation of information tangential to the area (at approx. $0.2 - 1\,\mathrm{m.s}^{-1}$ corresponding to approx. 1.5 ms between two neighboring cortical columns) but also a back-propagating flow from higher level areas (see Sec. 1.2.3) which greatly influences the response of cortical columns. These results in selectivity differences which may influence the shape of the RF [Gilbert and Wiesel, 1979; Monier et al., 2003].

This non-linear weighting may be related to a generic interaction taking functionally the form of a divisive normalization [Schwartz and Simoncelli, 2001; Wainwright et al., 2001] which basically optimizes the independance of the output of the local circuit. This may be related to the extraction of the maximum likelihood of the feature [Denève et al., 1999] and there is a trend in explaining these behaviors as the emergence of a generic local microcircuit that would implement an efficient coding of the neural information. Other models such as those proposed by Grossberg [2003] implement in recurrent circuits of cortical columns a similar optimization argument and give raise to functional algorithms which are applied to image processing [Grossberg and Yazdanbakhsh, 2005]. This introduces an indirect dynamical influence, the response of columns in the area being influenced by the direct feed-forward visual input but also by the converging dynamical context of lateral interactions.

We may therefore rather describe the function of the RFs of columns by dynamically separating the function of the fast statistical pattern matching from the latter slower interactions within the population using lateral connections. First, columns quickly integrate afferent information to match the feed-forward visual input with some range of features which are specific to the column (this will developed in Sec. 2).

[14] Functional definition of the RF were first defined by Sherrington [1906] for the tactile sense as "the whole set of points of skin surface from which the scratch-reflex can be elicited", and then in other areas, for instance by [Hartline, 1940].

Secondly, the neighboring columns —which were also excited by some correlated activity if their RFs overlap— interact at the same time to form a robust and efficient representation of the visual map according to the range of selective features which are specific to the area. We will develop this hypothese in section 3 and propose a model of the behavior of the population of columns. In parallel, we will study how slowly varying rules may adapt the weights of the cortical micro-circuitry to achieve an efficient processing of natural images (see Sec. 3.2). In this view, the overall system will achieve at best to alleviate visual ambiguities by locally detecting features on their RF while dynamically optimizing the representation.

1.3.3 Feature maps: distributed visual representations

A key concept in this dynamical model of the visual system is the topological organization of the visual information on the surface of the cortex, forming dynamical *feature maps* [Miikkulainen et al., 2005]. We saw that cortical areas are in general organized in maps of features and that these are organized by arranging neighboring features in close cortical positions (see Sec. 1.3.1). We will in our models isolate a function for every visual area and thus consider that the different visual images will correspond to a special filtering of the retinal image to extract a set of features relevant to that function and independently to another set of features. These particular sets of features are most often defined by their intuitive level of complexity and are tested experimentally on the different cortical areas. The retina seems to represent the possible levels of contrast at different scales in the image rather independently to smooth illumination variations. In V1, the representation on the cortical map is more complex and maps are selective to several variables (for instance, edge orientation and phase in V1) on the flat surface of the cortex [Basole et al., 2003]. It seems to be rather independent of the global complexity of local contrast values, a visual scene eliciting the same map if it was presented as a drawn sketch. It is argued that this is in general achieved by bundling in macro-columns sets of micro-columns centered on the same spatial position but consisting of the set of different variables to code (see Fig. 3). In a higher visual level area, such as the cortical area MT, macro-columns regroup mini-columns sensitive to different possible local speeds relatively independent of other features of the visual image (such as the local texture) [Albright, 1984], mapping therefore an image of the motion flow of the visual scene. The dynamical propagation of information therefore creates an architecture of feature images which represent different aspects of visual features in parallel and which correspond to progressively more abstract functions in the hierarchy of the visual architecture.

Thus, these feature images are vectorial images of features organized by the spatial topology. This organization reflects at a first order the regularity due to the translation of objects in the visual space and therefore that features in the similar portion of visual space should be located on neighboring sites in the cortex. It is in particular retinotopic in low-level visual areas (that is that they represent the space as it is on the image formed on the retina). In particular, the arrangement of cells shows a structure centered on the axis of the eye, at the fovea, with a sampling which is approx. uniform on the macula (the region around the fovea) and than logarithmically proportional to eccentricity. This structure may be interpreted as a way for the visual system to focus information located around the axis of the eye. This structure must therefore be related to the mechanisms of attention and the

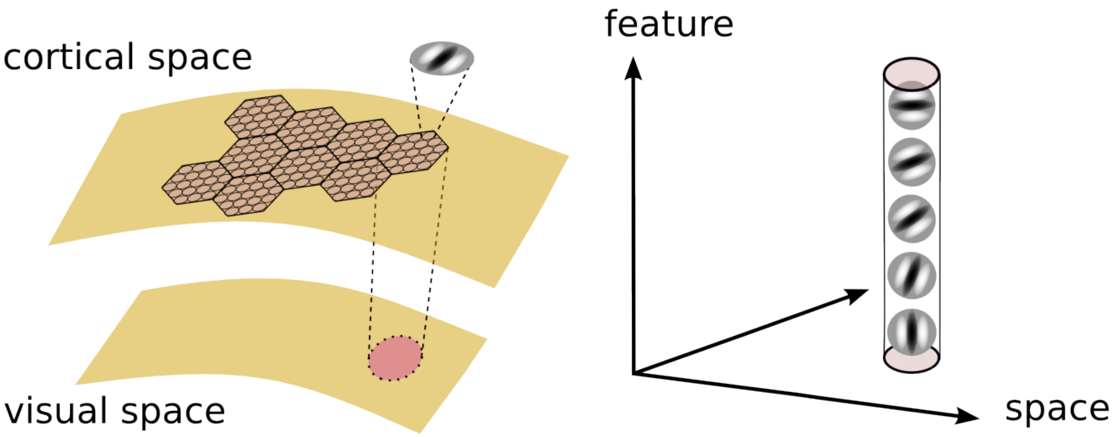

Figure 3: **Feature images.** *(Left)* Anatomical studies show that neurons on cortical surfaces share the same selectivity to visual features (in this example an elongated edge as in V1) perpendicularly to the surface but gradually change tangentially to the cortical area according to the retinotopy (visual space). These populations of neurons may be grouped by cortical columns in an "hyper-column" sharing a whole set of features at approx. the same visual space position. *(Right)* A useful view is to understand this topological arrangement as a *feature image*, that is as the representation of the selectivity to the different features (a vector) as a function of space. We may understand the previous arrangement as an optimal projection of this feature space on the two-dimensional map of the cortex [Petitot, 2003].

dynamics of eye movements. On cortical areas, the density of neurons is uniform (2.10^5 neurons/mm^2 in V1 [Hubel and Wiesel, 1974]) but the log-polar topological representation is preserved. However, it is progressively less precise in higher visual areas (In humans, at visual eccentricities of resp. 5 and 20 degrees, the diameter of V1 neurons' RFs is resp. 1 and 5 deg. while it is 5 and 20 degrees in MT) and in particular in the ventral pathways where the cortical areas are progressively independent on the position of the objects. It is also known to be less anchored on the retinal space (which in particular moves with every eye movement) and may get instead anchored on an ego-centered spatial reference. Low-level visual areas may thus be more generally described as these feature maps, the "closer" in feature space, the closer on the cortex.

From the columnar architecture of the cortical surface, features are projected on this two-dimensional surface and activities may be interpreted as a distributed representation of the quality of the match with the features. In fact, the structure of V1 revealed for instance by optical imaging [Grinvald et al., 2001] shows that the information of orientation is distributed in macro-column, forming "pin-wheels" of orientation selectivity[15]. This representation may be viewed as an optimal projection of the multidimensional feature image on the cortical surface [Petitot, 2003] (see Fig. 3), thus forming a distributed representation of the different features[16].

[15] This is not true for all species, such as the cat, and may reveal a hierarchy of organizations during evolution

[16] This principle may be true also at the scale of the CNS, and for instance neighboring areas (e.g. V1 and V2) represent relatively reflection symmetric visual maps along their common border.

However, it is not clear what and how information of the feature image is encoded in the neural activity, should it be explicitely (for instance in the neuronal firing frequency [Adrian, 1928]) or non-explicitely (through a non-linear transform of the feature image). Cortical maps therefore represent in a distributed manner the feature images at different positions and we will study this hypothesis in the following section.

2 Matching of low-level visual features

As we saw in the previous section, the goal of low-level vision is first to detect different known visual patterns rapidly and at best by using the actual visual input. We will in this section propose a generic inferential model thanks to an explicit representation of the quality of a match. In this framework, columns in a representation map will be regarded as ideal observers representing locally the probability of having matched the features they are selective to. We will illustrate this method on models of progressively higher visual areas: from the photoreceptors layer, to the output of the retina and finally to a model of the perception of motion for the control of the motion of the eyes.

2.1 Neural computations and the ideal observer approach

We will first present the problem of feature detection in the classical *statistical pattern matching* framework. It will be based on a generative model of the signal that will allow to represent explicitly the probability of feature detection.

2.1.1 Distributed probabilistic representation: example of photoreceptors

As an illustration, let us first understand the map of photoreceptors in the retina as transforming the image of luminosities (that is of the energy of absorbed light, the actual physical image measured in Trolands which corresponds to a physical energy of the incoming photonic flow) to a relevant variable, the *luminance* which, as we will see, is directly related to a probability. In fact, the light emitted or reflected by the various objects in our environment is concentrated by the eye on an image of photoreceptors of which the electrical activity is modified proportionally to the energy of the light influx. It was observed that this transform may be well described by a generic point-wise operation. It is implemented by the roughly logarithmic response of the aligned photoreceptors which transforms the probability distribution function of luminosities into a flatter distribution of luminances. This is similar to the histogram equalization operation which maximizes the entropy (for a bounded output $O(\vec{x})$ at location $\vec{x}$) and therefore the coding efficiency of the representation in terms of compression [van Hateren, 1993] or dually of the minimization of intrinsic noise [Srinivasan et al., 1982]. This process has been observed in a variety of species and for instance perfectly illustrated in the salamander [Laughlin, 1981][17]. It may evolve dynamically to slowly adapt to varying changes in luminances, such as when the light diminishes at dawn but also to some more elaborated scheme within

[17] This non-linear operation is by coincidence similar to the response of the cathode tubes used in traditional monitors and incidentally to the gamma correction applied to images.

a map [Hosoya et al., 2005]. The photoreceptor array thus non-linearly processes the image of luminous energies into a more tractable and robust image.
Let's understand this transformation mathematically thanks to the simple hypothesis that the output of the array should remain bounded. Let's note $P(I(\vec{x}))$ the probability of the luminosity $I(\vec{x})$ at position $\vec{x}$. This probability is highly skewed toward low energies in natural images (and generally independent on $\vec{x}$ if the lighting conditions are uniform). Let's now note

$$f_{\vec{x}} = \int_{\infty}^{I(\vec{x})} dP(I(\vec{x})) \tag{1}$$

the cumulative probability distribution[18] computed over a certain time period. Under the hypothesis that the output is bounded, the output maximizes entropy if it has uniform probability distribution over this range and a characteristic time period[19] and the solution may be written

$$\mathbb{L}(\vec{x}) = f_{\vec{x}}^{-1}(I(\vec{x})) \tag{2}$$

We therefore have an uniform distribution of the output probability distribution and in particular $\mathbb{L}(\vec{x}) = \int dP(\mathbb{L}(\vec{x}))$ models well the normalized luminance at position $\vec{x}$ (assuming that the output is in the range between 0 and 1). This process therefore allows a transformation of the image into a more efficient representation which directly derives from a probability.
Let's draw from this example general principles for understanding and modeling neural computations in low-level vision. The new variable, the luminance, is transformed into a variable independent of a global choice of a measurement unit: it is therefore more adapted to statistical pattern matching where the choice of a match should be for instance independent of the lighting conditions. This important transformation, also called histogram equalization, transforms arbitray values in an efficient representation and we will see how it may be used in practice below for the construction of look-up-tables (see for instance Sec. 2.2.2, Eqs. 1 or 58). It may be linked to the rank of the ordered values [Perrinet, 1999][20] and therefore may be a pre-processing stage which may have a role in transforming generic analogical values in a value which may be of relevance for detecting features. In fact, this variable is optimally sampled on its dynamic range as it is adaptive to slightly varying change of luminosity's statistics. Secondly, this statistical measurement is particularly adapted to inference mechanisms but also to the convergence of features from possibly different modalities as seems to be omnipresent in the CNS. This representation, from the direct relation to a probability measure, seems therefore to provide creative insights into the mechanisms underlying neural computations.

2.1.2 The ideal observer as a generic function

Let's define here a functional methodology to understand the mechanisms of neural computations. In fact, a recurrent problem in computational neuroscience is the under-constrainted number of free parameters which prompts a fastidious

[18] the symbol $dP(X)$ will here denote in general the probability distribution measure over variable X.

[19] This assumes the stationarity of the signal over this time period or over the image under the assumption of ergodicity.

[20] This was further applied to the retina, see [van Rullen and Thorpe, 2001] and Sec. 2.2

exploration of parameters. Testing the validity of fits between models and biological data may then reveal ambiguous results since they may not reveal general principle of computing but only a mere *ad hoc* description. As in the example above, a solution is rather to understand the functionality of the process (there from luminosity into luminance) in terms of its efficiency and thereby lowering the number of the parameters. Similarly, to resolve this problem in higher visual maps, we will thus assume that visual computations are tuned so as to match at best known visual features (knowing neural physiological constraints) and that they will do so in an optimal way, that is as an *ideal observer* [Geisler and Albrecht, 1997; Mamassian, 2002]. This view will allow us to explicit an interpretation for neural activity as directly linked to a probability measure and then model neural mechanisms as elementary inferences.

As coined by von Helmholtz [1925] under the term of unconscious inference, the goal of the visual perceptual system is to optimally solve an estimation problem of matching visual features for which we will see that probabilistic representation may give a generic computational scheme. This *ecological* view has multiple ramifications in the fields studying vision and there is growing evidence that neuro-physiological signals [Zemel and Sejnowski, 1998] and psychological responses [Kersten et al., 2003; Rao et al., 2002] may be interpreted as the response of an ideal observer according to an hypothesized model of the function of the system under consideration. This evolution is not in rupture with previous work based on linear models but rather extends these models to larger set of hypotheses. Moreover, these models and their emergent properties may now be explicitly be confronted with the hypotheses and physical causes they explicitly represent and thus allow a "dialectical" validation process. They may also explain the need of complex non-linear responses as the response of a network of inferences as an ideal observer is related to a particular information (should it be "the pixel $\vec{x}$ is at luminance 0.9", "there is a vertical edge at $\vec{x}$" or "the most likely motion is leftward"). The notion of ideal observer may thus provide a common language to understand some cognitive abilities.

In that view, the system should have a knowledge of the statistics of the patterns forming natural images. In fact, the underlying concept of an ideal observer is interdependent with the concept of patterns, or more generally of *causes*, and if we need to estimate a probability of a match we should first establish a model of the synthesis of the physical signal. According to the mathematician and economist Antoine-Augustin Cournot (1801-1877), the collection of physical signals, and therefore of images, may be understood as the fortuitous interference of independent causes, that is as the interplay of independent unknown sources. However, these patterns interfere according to known physical laws (such as transparency or occlusion in optics) which allow to build a model for image synthesis[21]. This generative model for the features that constitute the image is therefore necessary to describe inference mechanisms [MacKay, 2003, p.55]. In that sense, we may understand a goal of the feature images (see Sec. 1.3.3) as representing the probability of a match knowing a model of the synthesis and of the mixing of the causes they correspond to.

[21] In that sense, the different visual areas may correspond to relatively independent *models* or concepts of the visual features in an image: motion, detection of a pattern, ...

2.1.3 The Linear Generative Model (LGM) and natural scenes

Let's first try to find a general model of physical image formation which should encompass diverse levels of complexity. In fact, let's try to model the combination of simple luminous static features which interact according to the laws of transparency. It will be used in our model of the retina (see Sec. 2.2) for static images and is only valid on a local visual distance where occlusions are not predominant. If we consider that thanks to the adaptation of the photoreceptors, the incident light energy is roughly uniform, the luminosity from a position $\vec{x}$ in the visual space $\mathcal{X}$ is by the multiplicative law of light absorption the product of the reflectance of the different primitive causes (in a dictionary $\mathcal{S}$) that transformed the luminous influx:

$$I(\vec{x}) \propto \prod_{j \in \mathcal{S}} R_{j\vec{x}}^{s_j} \tag{3}$$

In this equation, the R_j correspond to spatial shapes (such as edges) which correspond to stable physical topological objects while s_j is its relative strength or contrast. This regularities may then be transformed in hypotheses which would be transcribed in ideal answers thanks to the bayesian formulation [Purves and Lotto, 2003] . This "laws" also correpond also to psychological experiments on the perception of transparency [Metelli, 1974] when manipulating the overlap of different simple transparent shapes. By the log-transform of photoreceptors (see Sec. 2.1.1), we thus have

$$L(\vec{x}) \sim \log(I(\vec{x})) \propto \sum_{j \in \mathcal{S}} s_j.(\log R_j)(\vec{x}) \tag{4}$$

A simple physical description of the image as the superposition of transparent patterns may therefore be given in terms of this Linear Generative Model (or LGM). The LGM may be also understood in a statistical framework similarly as we understood the transformation of luminosities into luminances [Atick, 1992]. In fact, the luminance as a probability measure may be seen as the factorial combination of independent features (or sources). These maps for the luminance of source probabilities (that we note respectively $\mathbb{L}(\vec{x})$ and s_j) thus by the causal link between the hypothetical sources $\mathbf{A}_j \propto \log R_j$ and the input image (see Fig. 12):

$$\mathbb{L} = \sum_{j \in \mathcal{S}} s_j.\mathbb{A}_j \tag{5}$$

where $\mathcal{S}$ is the set of sources. This defines the receptive field of a neuron as mapping explicitly the causal link between sources and luminance. This model is thus more general (and corresponds more to statistical pattern matching) since it is not strictly corresponding to physical assumptions as before. This will allow us to derive simple inference mechanisms using bayesian inference since we explicitly use probabilities [MacKay, 2003, p. 27].

2.1.4 Computing the probability of a match

Having set the forward model we are now interested in computing the match of a particular instance of the signal (here an image) with the model.

Theorem 1 (Best Match of a Single Source) *In the low-noise limit, for a given signal* $\mathbb{L} \in \mathcal{I}$*, the log-probability corresponding to a* single *source* $s_j.\mathbb{A}_j \in \mathcal{I}$ *knowing it is a realization of the LGM as it is defined in Eq. 5 (and for which we assume no prior knowledge) is maximal for the scalar projection coefficient*

$$s_j =< \mathbb{L}, \frac{\mathbb{A}_j}{||\mathbb{A}_j||^2} >= \frac{\sum_{\vec{x}\in\mathcal{X}} \mathbb{L}(\vec{x}).\mathbb{A}_j(\vec{x})}{\sum_{\vec{x}\in\mathcal{X}} \mathbb{A}_j(\vec{x})^2} \qquad (6)$$

and is then up to a constant proportional to

$$< \mathbb{L}, \frac{\mathbb{A}_j}{||\mathbb{A}_j||} >= \frac{\sum_{\vec{x}\in\mathcal{X}} \mathbb{L}(\vec{x}).\mathbb{A}_j(\vec{x})}{\sqrt{\sum_{\vec{x}\in\mathcal{X}} \mathbb{A}_j(\vec{x})^2}} \qquad (7) \quad \square$$

Proof We will note in general a single source by its index and strength by $\{j, s\}$ so that the corresponding vector in $\mathcal{S}$ corresponds to a vector of zero values except for the value s at index j. First, given the signal $\mathbb{L} \in \mathcal{I}$, we are searching for the probability corresponding to a *single* source $s.\mathbb{A}_j \in \mathcal{I}$ knowing it is a realization of the LGM. It is defined thanks to the conditional probability (Laplace, 1774) as the *a posteriori* probability noted $P(\{j, s\}|\mathbb{L})$. To evaluate this probability, we derive from this definition the theorem of Bayes [Bayes, 1764]:

$$P(\{j, s\}|\mathbb{L}) = 1/Z.P(\mathbb{L}|\{j, s\}).P(\{j, s\})] \qquad (8)$$

where Z is a normalization constant independent of the source[22], $P(\mathbb{L}|\{j, s\})$ is the likelihood probability of a signal knowing the single source and $P(\{j, s\})$ is the *a priori* probability of the sources. The prior thus corresponds to the available information when nothing is known from the input signal.

Following the interpretation of Cournot, we will first assume that we are in a low-noise limit environment (the global contrast is optimal and the eye/camera is adapted to the scene) so that we have no or little measurement noise. Knowing one component $\{j, s\}$, the only "noise" from the viewpoint of our ideal observer, that is the column j, is the combination of the unknown sources $\{\alpha_k\}_{1\leq j\leq N}$:

$$\mathbb{L} = s.\mathbb{A}_j + \nu \text{ with } \nu = \sum_k \alpha_k.\mathbb{A}_k \qquad (9)$$

The residual of the signal (an image) is thus considered as an undetermined perturbation[23]. Assuming that the α_k are independent random variables (since we know only $\{j, s\}$), from the central limit theorem it comes that for a sufficiently high number of sources, the distribution of the random variable ν converges to a normal distribution with known mean and covariance matrix. From the work of Field [1987], we know that for natural images this normal distribution is fairly homogeneous so that we may assume some prior knowledge on these second order statistics. In fact, over natural images, the correlation between the luminance of neighboring pixels is known to decrease inversely proportional to the distance, so that the covariance matrix has a regular shape [Atick, 1992]. We may therefore either use another metric (based on the Mahalanobis distance, as exposed in [Perrinet et al., 2004])

[22] We will keep this notation even if the constant may be different in the following.

[23] It should be stressed that the image model is still deterministic.

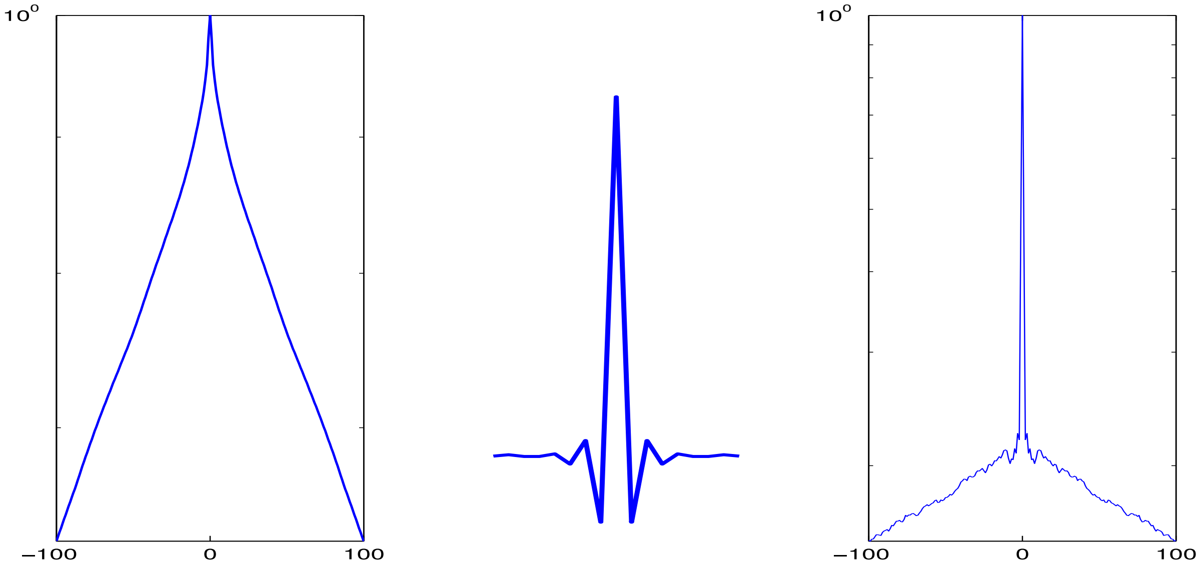

Figure 4: **Spatial decorrelation** *(Left)* Profile of pairwise spatial correlation in a set of natural images. It shows the typical decrease in $\frac{1}{f^2}$ of the power spectrum. *(Middle)* decorrelation filter computed from the methods of [Olshausen and Field, 1997] (see text). This profile is similar to the interaction profile of bipolar and horizontal cells in the retina. *(Right)* Profile of pairwise spatial correlations of the filtered images . As observed in the LGN [Dan et al., 1996], the power spectrum is relatively whitened.

or use a decorrelating kernel such as is defined in [Olshausen and Field, 1997] to transform this "noise" into a spherical probability distribution centered around the origin ($E(v) \ = 0$ with a covariance matrix equal to the identity times a variance σ^2, corresponding to the mean energy of images in $\mathcal{I}$). Note that this re-normalization according to the scale (or temporal frequency when using moving images [Dong and Atick, 1995]) leads to a different distribution of the Fourier components in the spatial frequency space: the image's power spectrum distribution is "spherized"[24]. This decorrelation process corresponds simply to a pre-process of filtering by the decorrelating kernel. It roughly corresponds to the physiological response of the layer of horizontal and bipolar cells in the retina between the photo-receptors and the ganglion cells. It results in modifications in the spatial frequency tuning of cells [Enroth-Cugell and Robson, 1966] which may be observed by the contrast sensitivity of retinal columns as a function of spatial contrast (see Fig. 4) and matches with the response measured in visual neurons [Dan et al., 1996]. This step will provide a preprocessing stage which spherize the non-gaussian statistics of natural images on the receptive fields of neurons.

The residual signal is thus considered as a decorrelated noise and from $P(\mathbb{L}|\{j, s\}) =$

[24] This is similar with the maximum entropy principle studied in Sec. 2.1.1. It extends to a fair competition between all different contrasts shapes. It should be noted that this is similar to diagonalizing the covariance matrix, such as in Principal Component Analysis and that this normalization could be easily learned by a linear hebbian rule [Oja, 1982].

$P(\mathbb{L} - s.\mathbb{A}_j) = P(\nu)$, it follows

$$\begin{aligned} P(\{j,s\}|\mathbb{L}) &= \frac{1}{Z}.P(\mathbb{L}|\{j,s\}).P(\{j,s\}) \\ &= \frac{1}{Z}.\exp(-\frac{||\mathbb{L} - s.\mathbb{A}_j||^2}{2.\sigma^2}).P(\{j,s\}) \end{aligned} \tag{10}$$

We will further consider that the dictionary $\mathbb{A}$ was learned so that over a long period the different sources have similar statistics: the prior is uniform across sources and values (we thus have no prior knowledge or preference for any source and was proposed in Sec. 1.3.3 and will be implemented in Eq. 55).

$$\begin{aligned} \log P(\{j,s\}|\mathbb{L}) &= -\log Z + ||\mathbb{L} - s.\mathbb{A}_j||^2/\sigma^2/2 \\ &= -\log Z + [s^2.||\mathbb{A}_j||^2 - 2.s. < \mathbb{L}, \mathbb{A}_j >]/\sigma^2/2 \end{aligned}$$

It should be noted that to minimize this bi-variate function in s and j, we may first minimize for every element j the coefficient s_j to get the corresponding $s^*_j = \text{ArgMax}_s P(\{j,s\}|\mathbb{L})$. From the above equations, this is equivalent to minimizing in the last equation the quadratic function of s which is minimal for the scalar coefficient

$$s^*_j = \frac{< \mathbb{L}, \mathbb{A}_j >}{||\mathbb{A}_j||^2} \tag{11}$$

that is for the scalar projection of the input on $\mathbb{A}_j$. Then, since for every element j, $s^*_j.\mathbb{A}_j$ is the projection of $\mathbb{L}$ on $\mathbb{A}_j$, so that $s^*_j.\mathbb{A}_j$ and $\mathbb{L} - s^*_j.\mathbb{A}_j$ are orthogonal, it follows from Pythagoras's theorem:

$$-\log P(\{j,s\}|\mathbb{L}) = \log Z + \frac{< \mathbb{L}, \mathbb{A}_j >^2}{2.\sigma^2.||\mathbb{A}_j||^2} \tag{12}$$

and that the filter with maximum *a posteriori* probability is given by:

$$\begin{aligned} j^* &= \text{ArgMin}_j[||\mathbb{L} - s^*_j.\mathbb{A}_j||^2] \\ &= \text{ArgMax}_j| < \mathbb{L}, \frac{\mathbb{A}_j}{||\mathbb{A}_j||} > | \end{aligned} \tag{13}$$

Finally, as defined in Eq. 13, we found that the source component that maximizes the probability is the *projection of the signal* on the normalized elements of the dictionary, that is up to $||\mathbb{L}||$, to the cosinus of the angle between the data and the feature vectors. ■

This justifies the computation of the correlation in the perceptron model [Rosenblatt, 1960] as it provides a measure of the log-probability under the assumptions that we used. Moreover, it is popular to introduce a prior favorizing small coefficients as it is similar to a regularization strategy. One should note that we could easily constrain for the scalar values to be positive by setting an appropriate prior or simply by only looking for positive correlations in Eq. 13. Finally, using a generative model for the image, we could easily compute —under defined hypotheses — the argument of the maximum of this probability , that is the *Maximum A Posteriori* ((or MAP)) of the different sources. This log-probability correlates with the linear component of a good proportion of visual neurons in the low-level visual system and may be used to model the first level in the visual hierarchy, the retina.

2.2 Application to modeling the retina: transforming light into a wave of spikes

Based on the previous results, we will design a simple model of the transform of luminances into a contrast map. We will in particular explore how this map can be translated into a spiking pattern and build an efficient code for the transmission of information to the LGN and the visual cortex.

2.2.1 Architecture: detecting multi-scale contrasts

The retina is a thin layer at the back of the eye (see Sec. 1.3.1) which transforms the signal of the photoreceptors into an image of multi-scale contrasts. This feature map (see Sec. 1.3.3) is itself converted into a spiking signal along the optic nerve. We will further assume that the transform should be invariant to some translations and scaling and thus that the features to detect are generated by one similar *mother function* which will be replicated at different positions and scalings. This assumption is the natural counterpart of the changing position of objects in tridimensional space. A continuous sampling would constitute a continuous wavelet transform. However, the limited number of neurons constrains the transform to be discrete. The geometric sampling of the scale of the discrete wavelet optimizes a compromise of precision versus localization in the case of a uniform repartition of the spectral information. A popular solution (chosen for instance in van Rullen and Thorpe [2001]) is to choose a *dyadic* progression, that is where filter radius and grid spacing both grow as powers of two[25]. From these assumptions, we can define a single *mother function* ψ from which every filter can be derived using translation and scaling.
As the architecture is defined, an important task is to choose an appropriate mother wavelet to detect contrasts in the image. As in [van Rullen and Thorpe, 2001] and from [Field, 1994], neurons j are defined here according to their position $\vec{x}^*$ and scale σ as dilated, translated and sampled *Mexican Hat* (or Difference Of Gaussian — DOG) filters (see [Mallat, 1998, pp. 77], and Fig. 5) as

$$\mathbb{A}_j = \frac{1}{\sigma}\texttt{DOG}(\frac{\vec{x}-\vec{x}^*}{\sigma}) \text{ with } \texttt{DOG} \quad = \quad G_1 - 1/K^2.G_K \tag{14}$$

$$\text{and } G_\sigma(\vec{x}) \quad = \quad \frac{1}{2\pi.\sigma^2}.\exp(-\frac{\|\vec{x}\|^2}{2.\sigma^2}) \tag{15}$$

where we denote G_σ as the 2D Gaussian function of variance σ which equals .5 for the mother function. These filters fit to the receptive fields that can be observed in the biological retina for a constant K approx. equal to 5 [Enroth-Cugell and Robson, 1966]. It also fits the Laplacian-of-Gaussian function defined by [Marr, 1982] with $K \sim 1.6$. The choice of the mother function defines the prototypical contrast to detect. In accordance with the previous results using the LGM model (see Sec. 2.1.4), we will set the linear representation at the output of the retina to be the *a posteriori* log-probability of a match with a set of features. This is in accordance with neurophysiological data on the linear response of retinal neurons [Rodieck, 1965], and will drive the activity of the retinal "columns" according to :

$$C_j :=< \mathbb{L}, p_j.\mathbb{A}_j >= p_j. \sum_{\vec{x}\in\mathcal{R}_j} \mathbb{L}(\vec{x}).\mathbb{A}_{j\vec{x}} \tag{16}$$

[25] The total number of columns is thus proportional to the number of pixels by a factor of approx. 4/3.

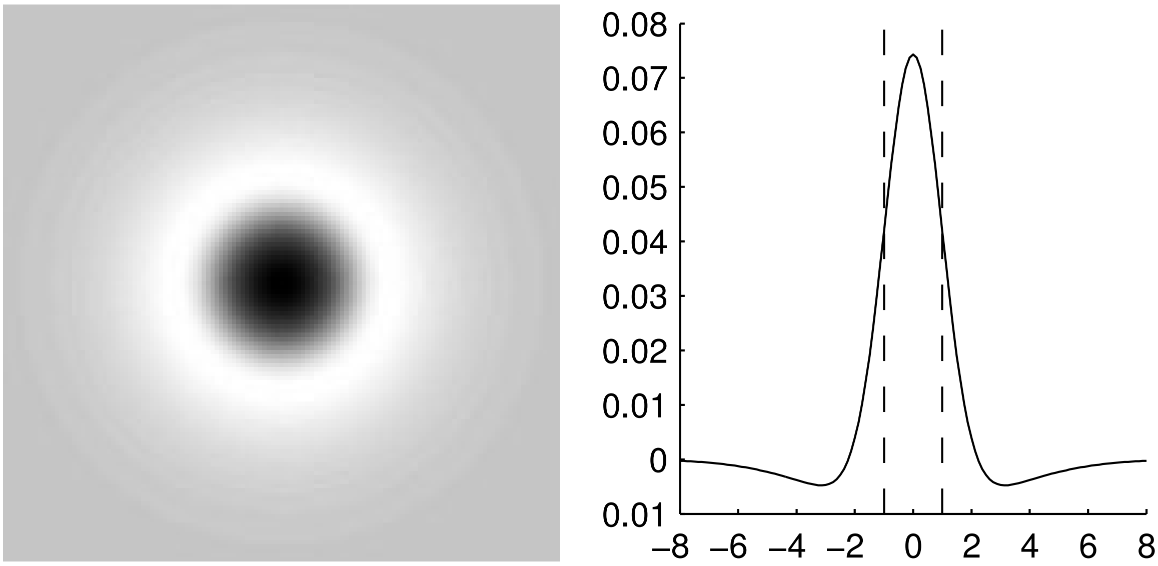

Figure 5: **Mother function for the filters in the retina**. *(Left)* The DOG filter is a center-OFF (in black) surround-ON (in white) contrast detector. *(Right)* Central profile of the DOG filter. As a unit measure, the vertical striped lines represent the variance of the narrower Gaussian used to generate the DOG filter and thus corresponds to the center of neighboring filters.

where $L(\vec{x})$ is the luminance at pixel $\vec{x}$ and $\mathcal{R}_j$ is here the receptive field of the column j. Instead of differentiating ON or OFF cells (so that the number of neurons is doubled), we will consider for simplicity and because it is exactly equivalent that each neuron $\vec{x}$ is assigned a polarity $p_{\vec{x}}$ which is either +1 or −1, so that the coefficients are rectified (*i.e.* $|C_{\vec{x}}| = p_{\vec{x}}.C_{\vec{x}}$). To compare this algorithm with standard computerized images, the photo-receptors and neurons are here placed uniformly over rectangular grids[26]. The bayesian inference is here synthesized by a linear filtering with filters corresponding to luminous contrasts.

2.2.2 Method: transmission of the spiking image

We will consider the optic channel as an information channel and rate the quality of a reconstruction using the coefficients computed above. In fact, there may certainly not be an explicit reconstruction of the image in the visual system, but as proved by Sec. 2.1.4, the squared error between the original and reconstructed image gives an explicit measurement of how well the parameters extracted conform to the model of natural images that we use. Here, this transform forms an approximate *orthogonal wavelet transform* [Mallat, 1998] of the image, *i.e.* the responses of different fibers are approx. uncorrelated (that is $< \mathbb{A}_j, \mathbb{A}_{j'} >\sim 0$ for $j \neq j'$) and may be inverted simply by the Calderón formula :

$$\mathbb{L}_{\text{rec}} := 1/Z.\sum_{j\in\mathcal{D}} C_j.\mathbb{A}_j = \mathbb{L} * \texttt{PSF} \tag{17}$$

[26] A more biological mapping would be a log-polar such as described in Sec. 1.3.3

$$\text{where } \mathtt{PSF} = \sum_{1\leq s\leq s_{\mathtt{max}}} \frac{1}{\sigma_s^2} \mathtt{DOG}(\frac{\vec{x}}{\sigma_s}) * \mathtt{DOG}(\frac{\vec{x}}{\sigma_s}) \tag{18}$$

where $*$ represents the correlation. The PSF function is the "point spread function"[27] of the coding system and acts as a blur on the image. This linear layer therefore exhibits two problems : first, the reconstruction is approximate and second, its implementation may be computationally slow because the size of the filters can become very large. A common alternative is to use a Laplacian pyramid as defined by [Burt and Adelson, 1983] which is computationally more tractable and since it is perfectly orthogonal, the reconstruction of the image is then perfect.

When presenting an image at an initial time, the output ganglion cell neuron of each column of the model integrates the analog contrast information C_j at its soma until the activity eventually reaches a threshold: it then emits an action potential or spike (see Sec. 1.2.1). As in most models of neuronal integration, we will simply assume that the stronger the activation, the earlier the cell will reach the threshold. Such behavior happens in most biological neurons and can be be implemented [Perrinet, 2005] by both detailed and more simple models such as the Integrate-and-Fire neuron [Lapicque, 1907]. Finally, the spike then propagates along the axon and the neuron's activity is reset. Classically, this generates a pattern of spikes whose *firing frequency* may constitute the image's code. But the code may also be equivalently carried by the exact spiking time (or *latency*) of the first spike. We may thus consider that the code consists of this latency for each of the different fibers j and which is inversely proportional to the neuron's excitation current, that is to the corrected activity. This algorithm defines a coding scheme that transforms an analog matrix pyramid into a spike 'wave front' that travels along the optic nerve.

Using this framework, the coefficients are emitted and transmitted in order, starting with the highest rectified contrast. If we know exactly the corresponding contrast values when trying to decode the spike wave, we may reconstruct progressively the image as

$$\mathbb{L}_{\text{rec}}(t) = \sum_{r=1\ldots t} C_{o(r)}.\mathbb{A}_{o(r)} \tag{19}$$

where t is the corresponding discrete time corresponding to the count of fired spikes (*i.e.* their rank) that we use for the reconstruction and $o(r)$ is the address of the neuron of rank r. In fact, if we assume that the filters are orthonormal and from Pythagoras' theorem, since o is a permutation of the addresses of neurons, the squared error SE(t) at time t is simply:

$$\begin{aligned} ||\mathbb{L}_{\text{rec}}(t) - \mathbb{L}||^2 &= ||\sum_{r=1\ldots t} C_{o(r)}.\mathbb{A}_{o(r)} - \sum_j C_j.\mathbb{A}_j||^2 \\ &= ||\sum_{r=t+1\ldots r_{max}} C_{o(r)}.\mathbb{A}_{o(r)}||^2 \\ \text{SE}(t) &= \sum_{r=t+1\ldots r_{max}} |C_{o(r)}|^2 \end{aligned} \tag{20}$$

where r_{max} is the final time (and therefore corresponds to the total number of rectified coefficients). From Eq. 20, this strategy of coefficient propagation corresponds thus to a "greedy" minimization of the MSE at each step of the algorithm. This also leads to the convergence of $\mathbb{L}_{\text{rec}}(t)$ toward $\mathbb{L}_{\mathtt{rec}}$ (and therefore to $\mathbb{L}$ for the Laplacian

[27] Similarly as in optics, this is the response of the whole system (coding and decoding) to an impulse, here to the image of a single pixel of luminance 1. From the linearity of the transform, it proves the assertion.

Pyramid), leading to a progressive compact coding of the image (see Fig. 6).
But, how can this information be encoded and decoded using only one spike per axon? In fact, these contrast values observe regularities across natural images as they were ordered from the largest to the lowest. A solution is therefore to use the mean analog value to form a *Look-Up Table* (LUT) to decode the analog values back from their rank. Let's thus define

$$\texttt{LUT}(r) = E[|C_{o(r)}|] \tag{21}$$

where E denotes the average over a set of randomly chosen images from the database[28]. In practice, the average was computed using a stochastic algorithm. For instance in that case, after the n^{th} image using $\texttt{LUT}(n)$ as a modulation function,

$$\texttt{LUT}^{(n+1)}(r) = (1 - \mu^{(n)}).\texttt{LUT}^{(n)}(r) + \mu^{(n)}.|C^{o(r)}| \tag{22}$$

where t is as before the discrete time corresponding to the decomposition and $\mu^{(n)}$ the stochastic learning gain (typically, $\mu^{(n)} = 1/\tau$ where τ is the characteristic time scale of the learning). Then, we can reconstruct the image from the spike list using

$$\tilde{\mathbb{L}}_{\text{rec}}(t) = \sum\nolimits_{r=1\ldots t} \texttt{LUT}(r).p_{o(r)}.\mathbb{A}_{o(r)} \tag{23}$$

where $\tilde{\mathbb{L}}_{\text{rec}}(t)$ is the image reconstructed using the spikes rank at step t and $p_{o(r)}$ the polarity of neuron corresponding to the r^{th} spike. Using the orthogonality of the filters, the error $\text{SE}_{\text{Lut}}(t)$ is therefore using a same method as above (see Eq. 20)

$$\begin{aligned} \text{SE}_{\text{Lut}}(t) &:= ||\tilde{\mathbb{L}}_{\text{rec}}(t) - \mathbb{L}||^2 && (24) \\ &= ||(\tilde{\mathbb{L}}_{\text{rec}}(t) - \mathbb{L}_{\text{rec}}(t)) + (\mathbb{L}_{\text{rec}}(t) - \mathbb{L})||^2 \\ &= \sum_{r=1\ldots t} (\texttt{LUT}(r) - |C_{o(r)}|)^2 + \text{SE}(t) && (25) \end{aligned}$$

The reconstruction error is therefore the sum the quantization error added to the energy that has not yet been transmitted. Eq. 25 also justifies the choice of the LUT as the mean (see Eq. 21) since it is the optimal estimator for the rectified coefficient as a function of its rank in the MSE metric. Neuro-physiological mechanisms for producing this decrease of the coefficients may involve a set of separate neurons (namely fast spiking inter-neurons) using shunting inhibition [Delorme and Thorpe, 2003] or directly the collaterals of afferent fibers to a pool of inhibitory neurons. We propose in Eq. 58 that these "rank counter interneurons" could be tuned used an incremental adaptive rule with an on-line hebbian learning scheme. In a more general framework, instead of using explicitly a rank, which mathematical definition is hard to relate with a sound biological interpretation[29], this modulation may be based on a divisive normalization by using the contextual information (see Sec. 3.1.4) instead of the rank the probability of the match.

2.2.3 Results: Spike Coding of natural images

This algorithm was experimentally validated using a database of natural images. These images were chosen in the publicly available database of linearly calibrated

[28] Further averaging or learning schemes used here 200 randomly chosen images.

[29] How for instance is it possible to define an initialization time in the brain? How to handle ex-aequos or analog precision?

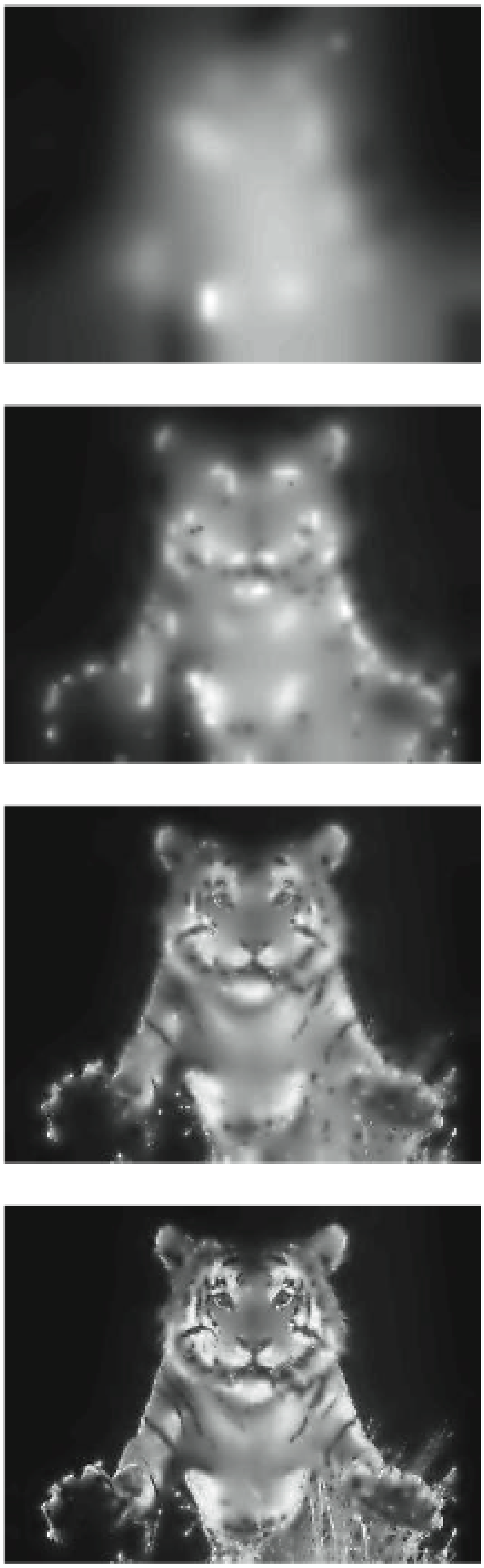

Figure 6: **Progressive reconstruction of the spiking image in the retina.** We compared for a natural image the model retinian code as defined in Sec. 2.2. We show the reconstructed images from the spike code after resp. 100, 750, 3000 and 9000 spikes from up to down. We recognize the original image in both cases after a few hundreds spikes as the coefficients rapidly decrease to zero.

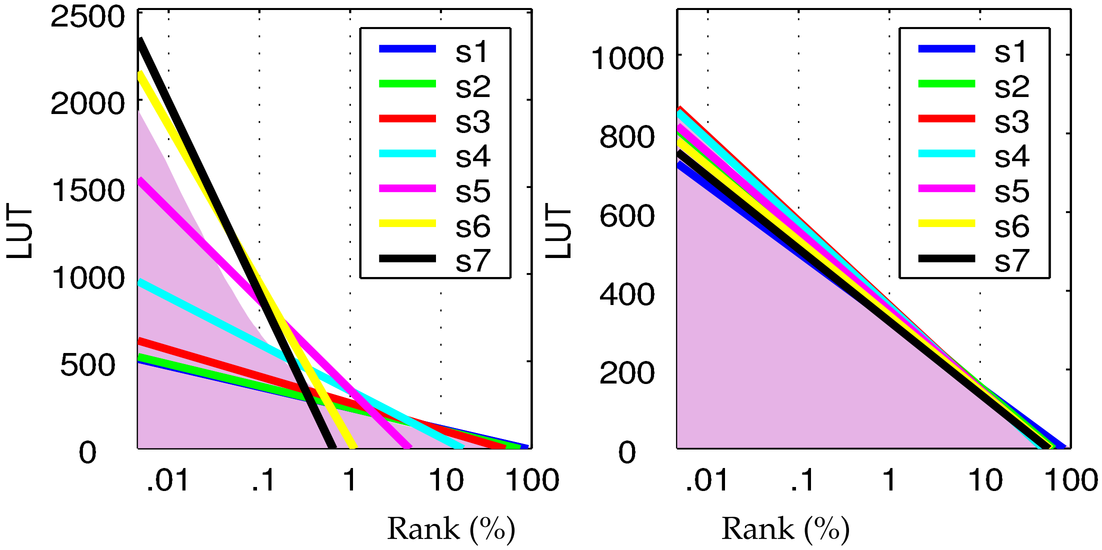

Figure 7: **Optimization of the regularity of the wavelet coefficients with harmonized scales.** *(Left)* The LUT (Eq. 21) is shown in the background in plain color as a function of the relative rank (in %) using a logarithmic scale on the abscissa. When separating the LUT for the different scales (from lowest to highest : *s*1 to *s*7 in the legend), one may observe that they correspond to similar regularities —linked to the regular distribution of singularities at different scales— but are mistuned (lower frequencies, as the 7^{th} scale 's7' are stronger and thus decrease more rapidly as a function of the overall rank). These regularities are therefore lost and mixed when ranking all scales together. *(Right)* By normalizing the different scales according to the statistics of natural images, the "vote" by the ranking process becomes "fair" and the LUTs for the different scales better match. The resulting LUT for harmonized scales preserves the underlying regularity and information transmission is therefore more robust (see Eq. 25): it represents a more effective way to encode the analog value by the rank of a spike.

natural images from van Hateren and van der Schaaf [1998]. These images were corrected using a γ correction [Poynton, 1999] to assure the balance of luminance and mimic the analogical response of photo-receptors to luminosity, *i.e.* to light intensity (see Sec. 2.1.1).

The coding efficiency of this analog-to-spike scheme is dependent on the regularity of the contrast coefficients. In fact, when analyzing this regularity separately for the different scales, the coefficients of particular scales are not well tuned to the overall LUT (see Fig. 7-Left). The coefficients corresponding to the lowest frequencies have *a priori* a higher probability to be transmitted first whatever the image to be coded [Perrinet et al., 2002]. From the "donut" shaped Fourier transform of the DOG filters, it is easy to see that there is a direct correspondence between the activities of the neurons at a given scale and the Fourier components of the image at a certain frequency. From what we saw in Sec. 2.1.4, there is also a direct relation between the probability distribution of contrast coefficients and the power spectrum which was observed on natural images. The mistuning of neurons at different scales thus corresponds in Fourier space to the shape of the mean power spectrum function in natural images. From Eq. 25, and since this leads to less variance, we are thus assured that this regularity results in a more effective information transmission, a result that was verified experimentally (see Fig. 8 and [Perrinet et al., 2004]). The regularity of contrast coefficients is therefore better when considering the second order statistics of natural images.

When the scales are tuned, rectified coefficients follow a very regular linear decrease (see Fig. 7-Right) in the log-linear plot, starting at rank 1 to a value proportional to the mean energy in the image and ending at the final rank at zero. It suggests the existence of a relation of the rectified contrast value as a function of the logarithm of the relative rank, that is a power-law probability distribution of contrast coefficients and one may wonder of the origin of this regularity. In our framework, since this multi-scale contrast representation gives a local measurement of the *Lipschitz exponents* in the image [Mallat and Hwang, 1991]. In fact, this regularity may be linked to the distribution of these exponents in natural images since they correspond to a measure of the order of the singularities which are present in the image, and that can be qualitatively ranked from the highest to the lowest Lipschitz exponents as : isolated dots, lines, edges, slopes, gradients until uniform surfaces [Mallat, 1998, p.513]. We may interpret the relative regularity of the distribution of Lipschitz exponents physically as (1) the whitening process removes the correlations between spatial frequencies due to size and depth of objects [Alvarez et al., 1999], (2) then, the distribution of complexity of shapes and textures of objects in nature is regular. This last point is linked to the inherent properties of auto-similarity in images [Turiel et al., 1998]. This generative model approach justifies the use of the LUT along with the decorrelation in the algorithm since it corresponds to a more robust physical interpretation of the visual input.

We estimated the efficiency of this coding scheme on the set of natural images. The Mean-Squared Error (MSE) and Mutual Information (MI) are popular criteria to rate the efficiency of the coding and we measured these values for different numbers of propagated coefficients and compared this result to the case where we used or not the LUT. In our framework, we defined the MSE based on the new metric, which leads to a new distance between images directly proportional to the log-likelihood defined in Sec. 2.1.4. The MSE appears then to be more correlated to a subjective measure of distances between images, and since there is a non-uniform

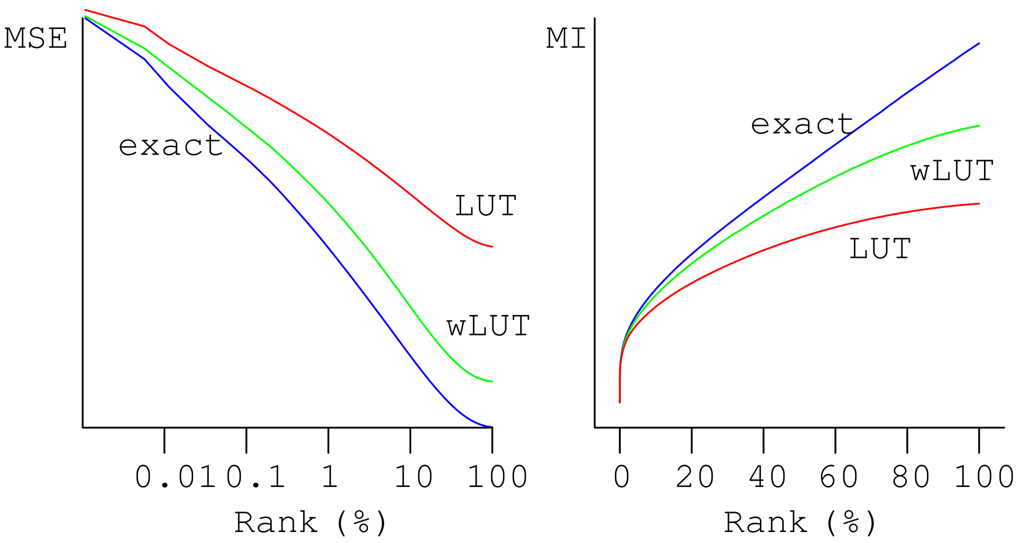

Figure 8: **Progressive reconstruction of the image from the spike list using rank-order coding.** We plotted *(Left)* the Mean-Squared Error (MSE, logarithmic scale on the abscissa) and *(Right)* the Mutual Information (MI) using the different temporal spike codes described in the text. We compared the results of the propagation when knowing the coefficients (exact) with the method described in Eq. 21 (LUT) which uses an optimized Look-Up-Table to "guess" the value of the coefficients from their rank. Finally we compared these strategies to the optimized method that uses the regularity found in natural images through the statistics of natural images (wLUT). The reconstruction from this latter method is close to the method with exact values and proves that the analog values may be transmitted using rank order coding. It therefore constitutes a compact spike code which provides a simple implementation of rank-order coding for static images.

prior in the energy of coefficients as a function of spatiall frequency, it corresponds in fact to the Mahalanobis distance [Mahalanobis, 1936] applied to our set of natural images. It removes some of the disadvantages of the standard MSE measurement, such as its dependence to a constant component and provides thus a more robust criteria for image reconstruction (see Fig. 8). As a conclusion for this model, we have provided a general scheme for spike coding in the retina using the relative rank and using the statistics of natural images. It also provides an efficient code for the preservation of edges [Sen and Furber, 2006]. This proves that this strategy can build a complete and efficient code from the retina (analog to spike coding) which can be decoded (spike to analog coding) using solely a temporal cooperation and provides a compact *temporal spike code* in the retina.

However, this dynamical algorithm transforming an analogic image in a train of spikes may be adapted to different goals such as is inspired by the architecture of the visual system (see Sec. 1.3.1). In fact, it may be behaviorally more relevant to still propagate the lowest frequencies, that is potentially closer contrast features first. As is implemented in the retina by the differentiation between the Magno- and Parvo-cellular pathways, low and high spatial frequency bands show different initiation latencies, the neurons from the Magno-cellular pathway being significantly faster (by approx. 20 ms). At the same time, we can still evaluate the weighted coefficients to produce a highly regular LUT (as in Fig. 7-Right), hence a better transmission of the coefficients but now rank the propagation of the coefficients according to their energy so as to choose the order in which the spikes are emitted (therefore using a similar algorithm as the first scheme)[30]. In a statistical inferential model, this will correspond to the inclusion of a gain for low frequencies. Practically, the scheme uses two parallel sorting mechanisms, one based on the regularity of the distribution of Lipschitz exponents and the other based on the progressive transmission of the parts of the image starting with the most informative. Together, they provide an algorithm that can efficiently decode the analog values corresponding to each spike using only the relative rank information and we will show now how it may be applied to a model of *short latency ocular following*.

2.3 Application to modeling short-latency ocular following

To further validate the previous hypotheses, we will describe an application of this method to describe a particular behavioral output through psycho-physical results on the visual perception of motion. In particular, *ocular following* is a reflexive and continuous pursuit by the eyes of an object which unexpectedly moves in the field of view and we will show here how the quantitative theory of matching visual features that we derived above may contribute to the understanding of the psychological results. This dynamical probabilistic model will be built in interaction with psycho-physical results and will be validated in comparison with other models.

2.3.1 Experimental psycho-physical setup

Motion processing is an essential piece of the complex visual machinery involved in controlling our actions. For instance, a brief and unexpected translation of a large visual scene elicits machine-like ocular following responses at ultra-short latencies in both humans (∼ 85 ms) and monkeys (∼ 55 ms, see [Masson, 2004] for a review). The initial, open-loop part of these reflexive eye movements exhibit many of the properties attributed to low level motion processing. These results suggest that the very earliest phase of reflexive tracking initiation relies on a linear motion detection followed by a rapid linear integration over a large part of the visual field. A key to understand the information flow is to study how this system reacts to different levels of information quality by measuring its output, the eye movements.

The experiment consisted in presenting different moving stimuli under strictly controlled conditions (see Fig. 9) to the subjects while measuring their eye movements with a very precise temporal and spatial precision thanks to the search coil technique [Masson et al., 2000]. The stimuli are standard in the sense that (1) they all elicit a very early horizontal motion with a latency ∼ 85 ms in the direction of the grating for humans (or *1D*) motion; (2) ocular following responses to the motion stimuli such as in the diagonal of the barber-pole stimulus [Masson et al., 2000] or of the uni-kinetic plaid [Masson and Castet, 2002] exhibit a later component with a latency ∼ 105 ms which rotates the tracking direction towards the global motion direction of the surface. This late component seems to depend on 2D local motion cues such as line-endings in barber-poles and blobs in uni-kinetic plaids: affecting one or the other specifically changes the late component but leaves

[30] These constraints of course don't occur in biology where all the processing is parallel and LUTs may be computed individually at every point.

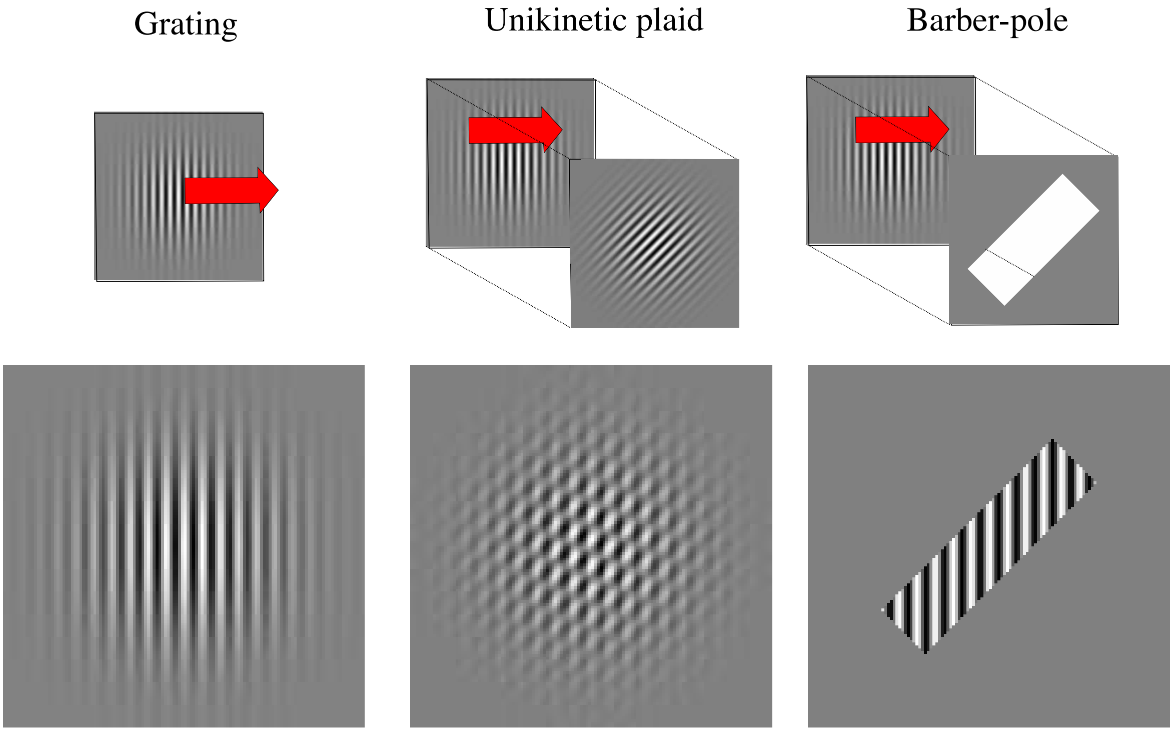

Figure 9: **Stimuli for ocular following experiments.** The presented stimuli were chosen as prototypical examples of the composition of 1D and 2D features in natural images, resp. from left to right Grating, Plaid and Barberpole. All three have the same main horizontally drifting vertical frequency with a velocity of 1 pixel/frame (grating), but merged with a static slanted frequency (plaid) or through an occluding aperture (barberpole). For the latter two, non-Fourier components (see text) are almost the only to contribute to vertical speed.

the early component intact [Masson et al., 2000; Masson and Castet, 2002] (see Fig. 11). Several recent studies have found similar dynamics directly at the level of MT neurons suggesting a direct link at short latencies between neural activity and the ocular response. With barber-pole motions, the direction selectivity of MT neurons evolves over time from grating (1D) to global (2D) motion-driven responses [Pack et al., 2004]. The latency between these two components is fairly constant across conditions at $\sim$ 20 ms (with an added delay inversely proportional to the contrast) and suggest that the 1D and 2D components are driven by two different independent visual pathways.

To quantify the characteristics of the two components in the different stimuli we measured their contrast response functions (CRFs). To achieve these experiments, the different stimuli were presented at different contrasts and we measured the dynamics of the gaze in both the 1D (horizontal) and 2D (vertical) directions (see Fig. 11). These curves show that from the onset of the response, the acceleration gain is fairly constant on the early part and we therefore measured this slope as the increase of speed in a 20 ms time window. Results show that the CRFs are very similar across stimuli for the 1D component but that, as perceived intuitively, the response to the unikinetic plaid is rather sluggish as the contrast increases, while the response to the barber-pole saturates quicker, in a more "binary" fashion (see Fig. 11).

2.3.2 A 2-pathway Bayesian model

We will create a probabilistic map (as described in Sec. 1.3.3 which quantitatively represents the probability of different possible translation speed $\vec{v}$ according to an internal model of the translation [Weiss et al., 2002]. This area will roughly correspond to area MT in primates (see Sec. 2) which is known to be organized retinotopically and representing specifically velocities [Albright, 1984]. As described before (see 2.1.3), we make the assumption of a generative model for translation in the input flow :

$$\mathbb{L}(\vec{x}, t) = \mathbb{L}(\vec{x} - \vec{v}.dt, t - dt) + \nu \tag{26}$$

This means that knowing the translation speed vector $\vec{v}$, the image intensity is approximately conserved along this direction. This approximation is noted as ν which is a noise image flow that is in general a colored Gaussian noise (with an emphasis on low frequencies in natural scenes [Dong and Atick, 1995]). If we observe an image flow $\mathbb{L}$ defined in space and time, and we wish to determine the probability *a posteriori* $P(\vec{v}|\mathbb{L})$ of the speed of translation $\vec{v}$ knowing $\mathbb{L}$, we have similarly as in Sec. 2.1.4:

$$\log P(\vec{v}|\mathbb{L}) = Z + \frac{-||\mathbb{L}(\vec{x}, t) - \mathbb{L}(\vec{x} - \vec{v}.dt, t - dt)||_K^2}{2.\sigma^2} + \log P(\vec{v}) \tag{27}$$

Since the background noise is constant across experiments, we will assume that the Michelson contrast C (the ratio of the extremal amplitudes of luminance over the extremal possible amplitude of luminance) of $\mathbb{L}$ controls the overall amount of noise to signal ratio[31]. Moreover, we will assume that the prior is normal with

[31] We may thus compute theoretically the probability for all contrasts. The visual system of course computes it on every instance of the signal.

variance σ_p so as to favor slow speeds [Weiss et al., 2002]. Finally,

$$\log P(\vec{v}|\mathbb{L}) = Z - [\frac{C^2}{\sigma^2}\Delta\mathbb{L}_{100}(\vec{v}) - \frac{1}{\sigma_p^2}.||\vec{v}||^2]/2 \tag{28}$$

where $\mathbb{L}_{100} = 1/C.\mathbb{L}$ is the normalized 100% contrasted image and $\Delta\mathbb{L}_C = C^{-2}.\Delta\mathbb{L}_{100}$ is the contrast-normalized gradient constraint for the normalized image[32]. It is therefore only necessary to know the constraint for the full contrast image flow to deduce analytically the probability in any case.
To overcome the static structure of the model from Weiss et al. [2002] and render the dynamical properties of short-latency ocular following suggested by the psychophysical results, we designed explicitly a duplexing in the input flow. This description is separated in the information conveyed quickly about raw 1D features and then by a more precise 2D information.

1. a first early component which computes linearly the transient temporal aspect of the image flow. It has a low spatial resolution, hence the name of 1D input (or Fourier since it is obtained by a simple convolution). It may yield ambiguous cues (such as for the grating), and static features are removed (for instance, the unikinetic plaid's 1D component becomes equivalent to the grating).
2. a later response which is selective to the outline of objects, hence the name of 2D input (or non-Fourier since it is obtained by non-linear operations). It is computed by selecting the most salient features in the moving flow.

Since these images carry the relevant information for each channel they each become an input flow for the probabilistic representation (Eq. 28).
Finally, the probabilistic information is then pooled to provide an optimal decision by minimizing the risk of error. If we consider the motor response as a simple first-order linear system acting as an *ideal observer* (see Sec. 2.1.2), we may predict that the eye's velocity gain is proportional to the mean velocity $\gamma = E(\vec{v})/\tau$ and that its latency is inversely proportional to the gain added to a fixed latency. To model the short-latency gain, we thus simply need to compute the mean translation speed command $\vec{v}$ knowing the visual input. As in [Weiss et al., 2002], if the constraint map is quadratic, then the *a posteriori* probability is Gaussian (and its mean is equivalent to its maximum). This allows us therefore to bridge an input image flow with a behavioral response and thus to validate the model by comparing it to the biological data.

2.3.3 Results : predicting CRFs using the model

A common procedure in psycho-physics that was also applied here is to study the results of the experiment with varying levels of noise [Albrecht and Hamilton, 1982]. However, it is in general analytically intractable from Eqs. 28 and $\gamma = E(\vec{v})/\tau$. As in Weiss et al. [2002], we may assume that the log-likelihood is quadratic. This assumption is very accurate for the grating and still a very good approximation for the plaid, in both 1D and 2D pathways. From Eq. 28, we deduce that the log-posterior is also quadratic and that the mean coincides with the MAP. Moreover,

[32] This formulation is very similar to the luminance conservation equation[Aubert et al., 2000] but gives a more rational explanation for this choice.

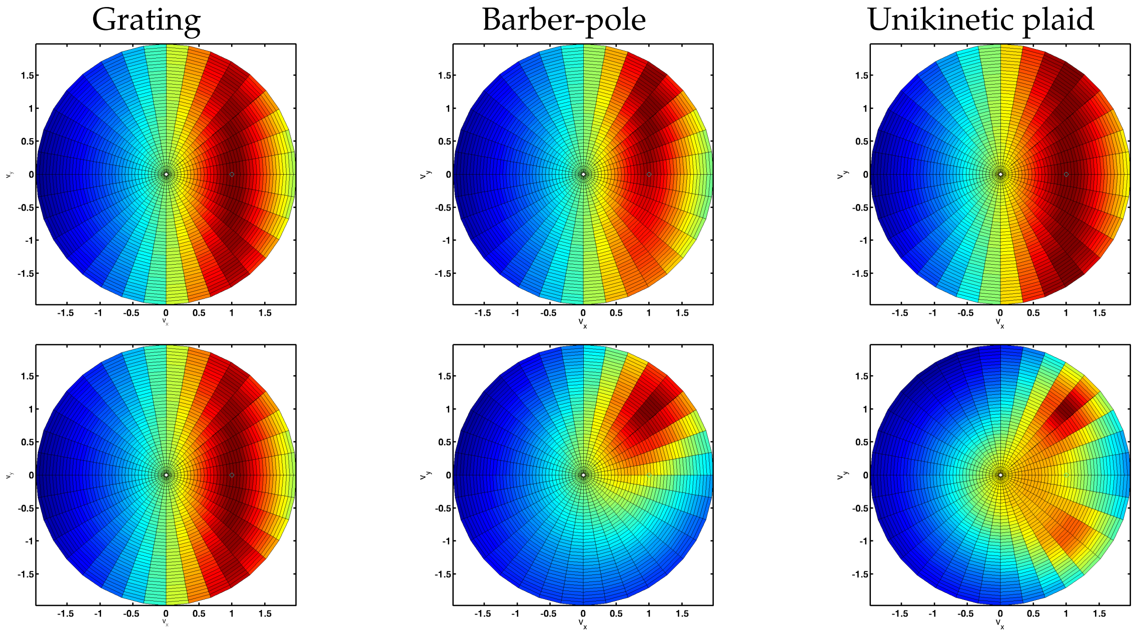

Figure 10: **Probabilistic motion constraints maps.** According to the model of short-latency ocular following, *(Left)* grating, *(Middle)* unikinetic plaid and *(Right)* barberpole. The *(Top row)* shows the result from the 1D early response and the *(Bottom row)* the late 2D response. The maps from the 1D information and for both the grating and plaid are quadratic and the CRFs are slope 2 Naka-Rushton curves. The map for the barberpole is however multi-modal and non-quadratic and generates a different CRF.

if we write that $\Delta\mathbb{L}_{100} = K.||\vec{v} - \vec{v}_0||^2$ (where K and $\vec{v}_0$ are constants), then the solution $\vec{v}_m$ satisfies :

$$\frac{d}{d\vec{v}} \log P(\vec{v}_m|\mathbb{L}) = \frac{K^2.C^2}{\sigma^2}(\vec{v}_m - \vec{v}_0) + \frac{\vec{v}_m}{\sigma_p^2} = 0$$

that is

$$\vec{v}_m = \frac{C^2}{C^2 + C_{50}^2}.\vec{v}_0 \text{ with } C_{50} \propto \frac{\sigma_p}{\sigma}$$

We recognize here the prototypical Naka-Rushton curve of slope 2 [Naka and Rushton, 1966], which is characteristic of the CRF at multiple levels of the visual system. Contrast dynamics of 1D and 2D motion for ocular tracking with the quadratic approximation involves thus necessarily that the Contrast Response Functions all follow a Naka-Rushton function with a slope of 2.

However, before making any such assumptions, we will at first keep all generality and try then to validate them. The algorithm will compute $\Delta\mathbb{L}_{100}(\vec{v})$ for each stimulus as a measure of the ambiguity of the motion and then compute the mean velocity $E(\vec{v}) = \int \vec{v}P(\vec{v}|\mathbb{L})$ to compute the contrast response functions of the 1D and 2D motion processing for short-latency ocular following. As expected, the constraint maps for the grating and the 1D component of the unikinetic plaid are quadratic along the line of constraint (a result that may be analytically proved). However, this distribution keeps quadratic for the 2d component of the unikinetic plaid but with a lower standard deviation, while it is multi-modal for the 2D component of the barberpole[33] (see Fig. 10). We used the Nelder-Mead simplex method [Lagarias et al., 1998] to fit the models with the data. Knowing the variability of the data, we minimized the χ^2 value to evaluate how close the model described the data [Cavanaugh et al., 2002] :

	Naka-Rushton	Bayesian model
Horizontal gain (1D)	0.1943	0.1324
Vertical gain (2D)	0.1567	0.0929

We have thus shown that the fit of the data obtained with our full model and the quadratic approximation (i.e. the model presented by Weiss et al. [2002] on the 2-pathway architecture) were comparable [Perrinet et al., 2005]. Moreover, when integrating in a time window of increasing size Δt and assuming that noise is independent over time, we found that the C_{50} was inversely proportional to Δt.

This approach provides a general framework for describing our model of translation in natural scenes. The explicit detail of the derivatives used to extract 1D and 2D features allows to draw analytically the similarity with techniques used in image processing and which use the luminance conservation constraint while providing a Bayesian explanation. We have also shown that the time course of this mechanism is separated in 2 distinct pathways with distinct CRFs. In particular, constraint map may not be quadratic and correspond to a different distribution of the visual information across different possible visual features. Moreover, this model does not use an hypothese concerning the quadratic shape of the constraint function.

[33] the second peak corresponds to the probability of a motion parallel to the narrow width of the aperture and grows as the ratio of the aperture gets closer to 1.

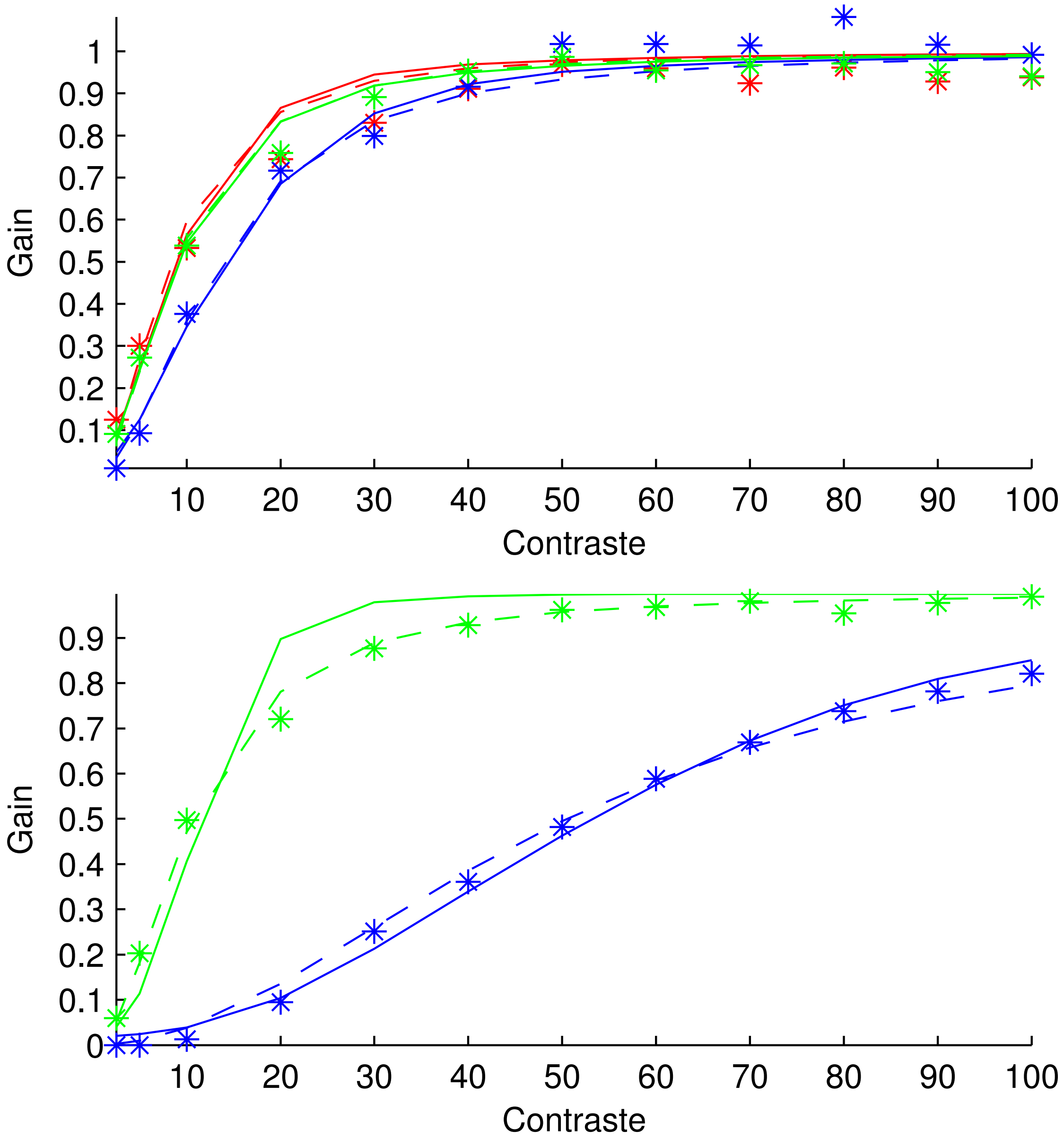

Figure 11: **Eye velocity in the ocular following task.** We show here the averaged velocity response of the eye on *(Up)* the horizontal axis and *(Down)* on the vertical axis in response to the barberpole stimulus (see Fig. 9). In the open-loop condition, responses follow a ramp pattern. The latency of the horizontal response is earlier of 20 ms. As the contrast decreases, both the gain decreases and the latency increases. The observation (stars) are well fitted by the Naka-Rushton (continuous line) and the bayesian model (dotted line).

In our model, the resulting contrast response function is in general analytically intractable, resulting in a richer family of possible functions. In particular, a higher kurtosis as in the 2D component of the barber-pole resulted in a CRF which fitted a Naka-Rushton with a slope higher than 2. The quantitative results suggests that the full model is already justified in this simple example, but this should be even more salient in more complex (and natural) experiments. This framework allows us thus to study how different features may be compared to be finally pooled together to give a single sensory-motor decision. In particular this should be extended to a model of a functional receptive field for ocular following which accounts for the spatial integration of local motion features as is for instance done by Bayerl and Neumann [2004]. A further extension is to study the rules of spatio-temporal integration of local features [Perrinet et al., 2006] and the adaptation of the visual system to repeated stimulations [Montagnini et al., 2006]. However, we will see in the next section that to efficiently represent multi-modular features, we need to extend our model to take into account "neighboring" information.

3 Sparse Spike Coding: building efficient representations

In the previous section, we were confronted to a main limit of classical feed-forward representations which is crucial for low-level visual processing. In fact, whether filters have to be orthogonal or the representation will necessarily be redundant (see Sec. 2.2.2). However, we will see that this constraint is incompatible with efficient visual algorithms and we will show here how to build a dynamical and adaptive algorithm resulting in an efficient representation that we will apply to a model of V1. We will then propose an implementation using Integrate-and-Fire neurons and test the efficiency of this artificial neural code and finally try to define a generic event-based computational approach.

3.1 Sparse Spike Coding for low-level visual processing

3.1.1 Low-level vision as a (hard) inverse problem

We saw that we may describe the goal of vision as representing matches according to an internal model (see Sec. 2). To further go in this direction, we will describe the goal of each low-level visual map as implementing an *inverse problem*. This goal is that for any of these images, this map must *efficiently* and *rapidly* (in the order of a fraction of a second) represent a set of relevant features characteristic of this map (see Sec. 1.3.3). This representation, including for instance in V1 the location and orientation of the edges that outline the shape of an object, is then relayed to higher level areas to allow, for instance, a recognition of more complex patterns (shape or motion of an object). The hypothesized function over the long term (in the order of hours to years) will thus be to adapt to natural scenes (that is images that occur most frequently) so as to progressively build this model. Actually, this is similar to numerous tasks in engineering and applied mathematics, where a reverse-engineering process allows to find a representation of the data (such as an estimation of the internal state of a system in control theory) by identifying the so-called hidden parameters of the system (see Fig. 12). The success of this

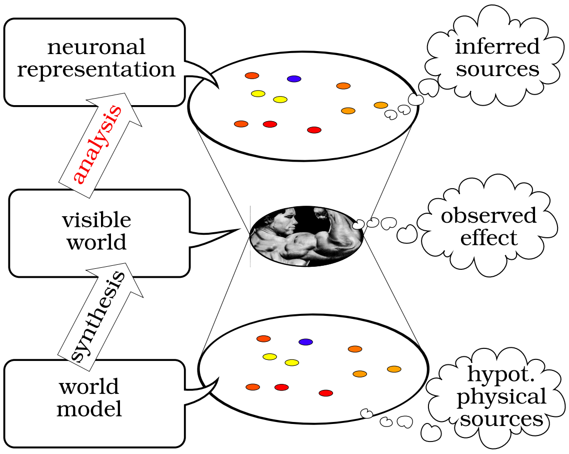

Figure 12: **Inverse-mapping as a a goal for sensory neural coding**. The visible world is modeled as the interaction of a large set of hypothetical physical sources (world model) according to a known model of their interactions ("synthesis"). We will consider that for sensory cortical areas, the goal of the neural representation (and its implementation by the *neural code*) is to analyze the signal so as to recover at best and as quickly as possible the sources that generated the signal ("analysis") . The analysis may thus be considered as an inverse mapping of the synthesis. A proposed solution for this problem is to *infer* at best the most probable hidden state.

algorithm over the long term (in the order of days to generations) allows then to validate through the pressure of evolution the model that was learned. In this framework, we will thus describe neuronal activity as the result of the efficient inversion (or analysis) of an internal model of the world.

To simplify the problem of the inverse problem, a first solution is to constrain the LGM to use orthogonal filters that form a complete basis as in Sec. 2.2.2. However, in this case, slight continuous transforms of the input may yield non-continuous transforms of the representation although it is highly desirable for the representations of natural images to be robust to these natural transforms (see Sec. 1.3.3). This view is similar to the 'rules' that were described by psychologists of the Gestalt school as Metzger [1936] since the system will exploit the regularities of the world to efficiently extract relevant information. As is the case for natural images in low-level vision, we will consider that the observed signals are generated by sources that share certain features which differ by continuous transformations such as edges at different time, position, orientation or scale. Since these regularly occurring changes in the physical world (translations, rotations and scaling) are very common, if there exists a corresponding transformation in the source space (that is if this transformation of all sources are in the dictionary), the resulting representation of the transformed image should simply be derived by a transformation (in the source space) of the original representation. This idea is supported as well by neurophysiological data (see Sec. 1.3.3) as by psychological experiments [Shepard and Metzler, 1970] and may also be a feature of some mid-level visual area, for instance with dynamic remapping [Pouget, 2002]. Typically, this robustness constraint implies in a wavelet architecture that the tiling of the filters is smoother than an orthogonal representation [Perrinet et al., 2004][34]. As a consequence, the dictionary will be *over-complete*, *i.e.* the number of dictionary elements will be of several orders of magnitude larger than the dimension of the image space. We deduce that the efficiency of the inverse mapping in low-level visual areas, conditioned by the relevance of the LGM , requires for the dictionary to be over-complete.

Given this constraint on the forward model, the representation should also optimize its information capacity [Atick, 1992; van Hateren and van der Schaaf, 1998], a task which may therefore relate to remove redundancies, but more generally to separate features into independent features. Using the linear forward model, for any signal $\mathbb{L}$, there exist at least one set of parameters $\mathbf{s}$ which recovers the observed signal if the dictionary $\mathbb{A}$ is complete. However, in this case where the dictionary is over-complete, the inversion of the LGM will not yield an unique solution in $\mathcal{S}$ to any given signal in $\mathcal{I}$: the problem is ill-posed. To any observation may correspond a wide range of causal configurations: the overlap between atoms generates a potential ambiguity as was illustrated in Fig. 1. The coding strategies corresponding to possible 'analysis' algorithms (see Fig. 12) have different efficiencies and, in particular, the solution given by the wavelet coefficients (as in [van Rullen and Thorpe, 2001]) with an over-complete dictionary yields an highly redundant —hence inefficient— representation. More importantly, they do not correspond to the actual physical causes and thus do not invert the forward problem (which is as we will see especially inappropriate for the learning). According to Barlow [2001], the goal of sensory processing would be rather to choose the most efficient representation: for instance the representation should reduce its redundancy. In a more general view and following the same argument as the Occam razor, whenever there is the choice between two equivalent representations, the most probable is the one that is the most sparse [Perrinet, 2006], a property supported by physiological experiments [Hahnloser et al., 2002; DeWeese et al., 2003]. A possible goal would thus be to achieve the coding strategy that describes at best the images with the least coefficients [Olshausen and Field, 1997]. This *sparseness* constraint thus allows to restrict the different solutions of the inversion of the forward model so as to find an appropriate candidate for the neural code.

However, the combinatorial complexity of the inverse problem for the LGM grows very quickly as the dimension of the dictionary increases (it's NP-complete, see [Mallat, 1998]). There exists therefore no simple algorithm that optimizes exactly the problem in reasonable time as we handle more complex signals such as natural images, but acceptable sub-optimal strategies to approach this problem do exist (see a review in [Pece, 2002]). Most popular solutions optimize the reconstruction error and the sparsity by using a Lagrangian multiplier to tune the compromise between both constraints [Olshausen and Field, 1997]. Their solution is based on a gradient-based optimization approach (for a review, see [Simoncelli and Olshausen, 2001]) which are heuristics particularly well adapted to computations on sequential computers. Focusing on the nature of neural computations, we will rather present an alternate *parallel* and *event-driven* heuristic.

[34] Another argument says that this could solve the binding problem by distributing the information of multiple copies of the neurons.

3.1.2 One solution: Greedy inference pursuit

Let's first define a sparse spiking algorithm in mathematical terms. It is a dynamical algorithm which should provide from an input (should it be continuous or spiking) an output consisting of a list of spikes that we may for instance define at time t as a list $\Lambda(t) = \{\lambda^k : (j^k, t^k, s^k)\}_{t^0 \leq t^k \leq t}$ of events to which different values may be attached (here respectively address, latency and amplitude). We will restrict first the input to this model to flashed static images[35]. Assuming that the LGM is known, we will define the goal of our sparse spiking algorithm as recovering the correct sources (corresponding to some hidden state variables, see Fig. 12) from an observed static image as quickly as possible. Spikes are fairly similar across time and over the CNS (they have metabolically the same cost and their precision make them carry *a priori* the same information). In an over-complete parallel scheme all information should therefore be in the address and timing of the spiking neurons neuron (rather than than in the activity it represents), so that it is proportional to $I_{spike} = \log_2(N)$, where N is the number of filters In the dictionary. We may define the cost for any algorithm to be a cost over the transmitted information as a function of the number n of emitted spikes. The information knowing $\Lambda(t)$ may be evaluated for the list of generated spikes as $I(t) = -\log(P(\mathbb{L}|\Lambda(t))$ so that the dynamical cost (in bits) may be evaluated for instance as

$$T(t) = -\log(P(\mathbb{L}|\Lambda(t)) - \log_2(N) * ||\Lambda(t)||_0 \tag{29}$$

where $||\Lambda(t)||_0$ is the number of spikes in the list (its l_0-norm)[36]. This formalization could be used in a cost function to reflect how much we do care about the transmission of information as a function to the number of spikes and thus give an explicit definition of the Occam's razor for our system. This expression allows thus one to compare different dynamical spiking algorithms knowing solely the LGM (which is necessary to derive the probability) and the expression of the cost.

We will here propose a solution for inverting the forward model that we defined for natural images based on a Bayesian inference framework using feature-matching neurons and spikes as events representing primitive "decisions". In fact, as in numerous optimization problems, a solution is to begin the algorithm with a subset of the problem which is easy to solve, take it into account and then to resume the algorithm in a recursive manner on the transformed observation signal : it's the greedy approach [Perrinet, 2004b]. Following this process and focusing on every single spike, a greedy solution could use recursively two steps: Matching (M) and Pursuit (P).

(M) To each neuron is assigned a vector (or weight pattern) corresponding to its preferred stimulus. Neurons compete in parallel to find the most probable *single source* component by integrating evidence according to their weight patterns (see Sec. 2.1.4). The first source to be detected should be the one corresponding to the highest activity, that is to the most information about the image knowing that source.

(P) A decision is assigned to this best match which, once it has been taken, is taken into account *before* performing any further computations (and in particular

[35] In particular, we will study the transient response of the network and neglect the information fed back by higher areas. This latter information will be necessary in more complex algorithms which take into account the context of a local feature.

[36] This definition is similar to measure of model complexity such as AIC [Akaike, 1974].

finding a new match). It thus yields a new observation signal (and therefore a modified internal representation) where we 'removed' the detected source.

We will see that this method is similar to the approach developed in the method of Matching Pursuit [Mallat and Zhang, 1993] but also to other techniques such as Projection Pursuit in statistics [Friedman and Stuetzle, 1980]. However, instead of a heuristic scheme, and thanks to the description of the successive steps that may lead to the greedy pursuit, it may be considered as an optimization strategy of the goal that we defined above (namely maximizing the transfer of information).

Greedy inference pursuit

Theorem 2 (Gredy Inference Pursuit) *Under the same hypotheses as Eq. 10, the greedy algorithm optimizing the goal set in Eq.29 for the succession of spikes is set for a known architecture $\mathbb{A}$ by transforming any image $\mathbb{L}$ in a event list Λ by*

- *initializing activities as*

$$C_j^{(0)} = < \mathbb{L}, \frac{\mathbb{A}_j}{||\mathbb{A}_j||} > \tag{30}$$

- *and recursively generating a new event $\{(j^{(n)}, t^{(n)}, s^{(n)})\}_n$ by:*

$$j^{(n)} = ArgMax_j |C_j^{(n-1)}| \quad \textit{(Matching)} \tag{31}$$

$$C_j^{(n)} = C_j^{(n-1)} - s^{(n)}.R_{\{j,j^{(n)}\}} \quad \textit{(Pursuit)} \tag{32}$$

where $R_{\{j,j^{(n)}\}} = < \frac{\mathbb{A}_j}{||\mathbb{A}_j||}, \frac{\mathbb{A}_{j^{(n)}}}{||\mathbb{A}_{j^{(n)}}||} >$ and $s^{(n)} = C_{j^{(n)}}^{(n-1)}$.

- *the input signal may be reconstructed from the spike list as*

$$\mathbb{L} = \sum\nolimits_{k=1...n} s^{(k)}.\mathbb{A}_{j^{(k)}} + \mathbb{L}^{(n)} \tag{33}$$

where $\mathbb{L}^{(n)}$ is the residual image at rank n. □

One can note that in this formalization, the timing is carried by the relative succession of the spikes and not in the absolute timing. This is coherent with the homeostatic processes occurring in the system which make the average firing frequency stationary. Moreover, we keep for the moment in the spike list the amplitude of the coefficients, but we will see that in natural images they obey to regularities and therefore carry little information knowing their rank in the list.

Proof Since the property has to be sound for all spike lists until a certain time, it has to be right for the first spike. From section Sec. 2, we computed the probability of any possible source as :

$$P(\{j, s\}|\mathbb{L}) \propto \exp(-\frac{||\mathbb{L} - s.\mathbb{A}_j||^2}{2.\sigma^2}) \tag{34}$$

The cost defined in Eq. 29 is equivalent to minimizing

$$C = \frac{||\mathbb{L} - s.\mathbb{A}_j||^2}{2.\sigma^2} + \log_2(N) * ||s||_0 \tag{35}$$

which is a hard problem [Chen, 1995]. Initializing the activity as $C_j^{(0)} =< \mathbb{L}, \frac{\mathbb{A}_j}{||\mathbb{A}_j||} >$ and selecting $j^{(1)} = \text{ArgMax}_j |C_j^{(0)}$ will give the best first single source match of the greedy algorithm.
As we found the MAP source knowing the signal $\mathbb{L}$, we may pursue the algorithm by accounting for this inference on the signal knowing the element that we found before detecting another single source component. Sources are supposed to have conditionally independent activities[37] and the pursuit algorithm assumes that —knowing the previous detection— we may resume the detection on this residual signal:

$$P(\{j, s\}|\mathbb{L}, \{j^*, s^*\}) = P(\{j, s\}|\mathbb{L} - s^*.\mathbb{A}_{j^*}) \tag{36}$$

with s^* given by Eq. 11: $s^* = C_{j^*}/||\mathbb{A}_{j^*}||$. We will thus use this new residual signal in which we will then find a new component corresponding to the most probable single source.
In this recursive approach, we will note as n the rank of the step in the pursuit (which begins at $n = 0$ for the initialization). Let's therefore set initially $\mathbb{L}^{(n)}$ the successive residuals and $\mathbb{L}^{(0)} = \mathbb{L}$. Let's also note the address of the successive winning neuron from the first step $n = 1$ as $j^{(n)} = \text{ArgMax}_j |C_j^{(n-1)}|$. Knowing $j^{(n)}$, in order to resume the pursuit at the next step, we saw that we need to compute the projection of the signal on the elements of the dictionary. In this greedy approach, we may thus update the residual and the corresponding activities $C_j^{(n-1)} =< \mathbb{L}^{(n-1)}, \frac{\mathbb{A}_j}{||\mathbb{A}_j||} >$ by subtracting to $\mathbb{L}^{(n-1)}$ its projection on the winning element of index $j^{(n)}$ (see Eq. 11):

$$\mathbb{L}^{(n)} = \mathbb{L}^{(n-1)} - C_{j^{(n)}}^{(n-1)}.\frac{\mathbb{A}_{j^{(n)}}}{||\mathbb{A}_{j^{(n)}}||} \tag{37}$$

Furthermore, we don't need to feed this information back to the signal (which would be very inefficient in a neural implementation) and we may directly compute the activity again for all vectors thanks to the linearity of the scalar product operator:

$$\begin{aligned} C_j^{(n)} &= < \mathbb{L}^{(n)}, \frac{\mathbb{A}_j}{||\mathbb{A}_j||} > \\ &= < \mathbb{L}^{(n-1)} - C_{j^{(n)}}^{(n-1)}.\frac{\mathbb{A}_{j^{(n)}}}{||\mathbb{A}_{j^{(n)}}||}, \frac{\mathbb{A}_j}{||\mathbb{A}_j||} > \end{aligned}$$

and finally

$$C_j^{(n)} = C_j^{(n-1)} - C_{j^{(n)}}^{(n-1)}. < \frac{\mathbb{A}_j}{||\mathbb{A}_j||}, \frac{\mathbb{A}_{j^{(n)}}}{||\mathbb{A}_{j^{(n)}}||} > \tag{38}$$

This activities' update (Eq. 38) corresponds in neuro-physiological terminology to a lateral interaction. It will be proportional to $R_{j,j^{(n)}}$ where $R_{j,j^{(n)}} =< \frac{\mathbb{A}_j}{||\mathbb{A}_j||}, \frac{\mathbb{A}_{j^{(n)}}}{||\mathbb{A}_{j^{(n)}}||} >$ is the normalized correlation of any element j with the winning element $j^{(n)}$. This is in accordance with neuro-physiological data suggesting that most lateral interactions

[37] For any realization of the images, individual sources have independent activities since in our framework they would correspond to different independent causes. Thus, by removing one source, one gets a new image (conform with the LGM model) and one does not change the probability distribution of the other sources.

are symmetric and often proportional to the similarity of neurons receptive fields []. The greedy pursuit therefore transforms an incoming signal $\mathbb{L}$ in a spike list with decreasing coefficient values $\{j^{(n)}, s^{(n)}\}$ From Eq. 37), the signal may be reconstructed as in Eq. 33 which should converge to $\mathbb{L}$ if the norm of the residual signal $\mathbb{L}^{(n)}$ converges to zero. ■

Properties of the greedy pursuit In this algorithm, the choice of the best match and the update rule are independent of the choice of the norm $||\mathbb{A}_j||$ of the filters (see Eq. 30 and 31), so that we may indifferently use in the following normalized filters (that is $||\mathbb{A}_j|| = 1$ for all neurons) so as to simplify the equations. This algorithm is exactly equivalent to Matching Pursuit [Mallat and Zhang, 1993]. This algorithm is familiar in signal processing and is increasingly used for image and video processing [Neff and Zakhor, 1997; Durka et al., 2001; Capobianco, 2003; Fischer et al., 2006, 2007a] and signal processing [Blinowska and Durka, 1994]. Moreover, we have shown that the use of the statistics of natural images statistically optimizes the coding efficiency by modifying the image space metric [Perrinet et al., 2004] compared to an heuristic optimization of Matching Pursuit [Pati et al., 1993]. The Bayesian inference framework allows to precisely tune the heuristic approach of the Matching Pursuit and is somewhat similar in the approach developed by Denève et al. [1999] for "reading" neural population codes by matching activity responses with the envelope of the correlation of neighboring neurons. It allows for instance to set a different prior or to include knowledge of the measurement noise that is adapted to the goal of the system (and hence a different matching criteria that may depend on the norm $||\mathbb{A}_j||$). This algorithm presents similar computational complexity and properties [Mallat, 1998, pp.412–9] which is suboptimal but more rapid compared to Basis Pursuit [Chen, 1995]. In particular

$$C^{(n)}_{j^{(n)}} = C^{(n-1)}_{j^{(n)}} - C^{(n-1)}_{j^{(n)}} = 0 \tag{39}$$

and as a consequence the activity of a winning neuron is totally canceled.

Theorem 3 (Convergence of Greedy Inference Pursuit) *The residual squared error is strictly decreasing and may be computed recursively as*

$$\begin{aligned} \mathbf{SE}^{(n)} &= ||\mathbb{L}^{(n)}||^2 \\ &= \mathbf{SE}^{(n-1)} - |s^{(n)}|^2.||\mathbb{A}_{j^{(n)}}||^2 \end{aligned} \tag{40}$$

In addition, the algorithm converges (that is $\lim_{n\to\infty} ||\mathbb{L}^{(n)}|| = 0$*) exponentially in the space generated by the dictionary (that is in the image space if it is complete).* □

In practice, it implies that the stopping criteria may be computed using this computation without computing $||\mathbb{L}^{(n)}||$.
This theorem gives a practical way of controlling the convergence of the algorithm and gives the proof that the algorithm will stop, that is that the residual error energy will be in finite time below the threshold ε.

Proof Although filters in the dictionary are here generally not orthogonal, the residual image is orthogonal to the winning filter and from Pythagora's theorem

$$||\mathbb{L}^{(n-1)}||^2 = ||\mathbb{L}^{(n)}||^2 + |s^{(n)}|^2.||\mathbb{A}_{j^{(n)}}||^2 \tag{41}$$

so that we may easily compute the Squared Error (SE) of the residual signal at every step of the coding :

$$\begin{aligned}
\mathbf{SE}^{(n)} &= -\log P(\Lambda|\mathbb{L}^{(n)}) \\
&= ||\mathbb{L} - \textstyle\sum_{k=1\ldots n} s^{(k)}.\mathbb{A}_{j^{(k)}}||^2 = ||\mathbb{L}^{(n)}||^2 \\
&= \mathbf{SE}^{(n-1)} - |s^{(n)}|^2.||\mathbb{A}_{j^{(n)}}||^2 \qquad (42) \\
\mathbf{SE}^{(n)} &= ||\mathbb{L}||^2 - \textstyle\sum_{k=1\ldots n} |s^{(k)}|^2.||\mathbb{A}_{j^{(k)}}||^2 \\
&= ||\mathbb{L}||^2 - \textstyle\sum_{k=1\ldots n} |C_{j^{(k)}}^{(k-1)}|^2 \qquad (43)
\end{aligned}$$

The exponential convergence is *e.g.* provided in [Gribonval and Vandergheynst, 2006]. ■

The residual squared error is a highly significant criteria here since it is proportional to the probability of the image knowing the spike list and therefore of how well the image is described by the coefficients already propagated (see Eq. 34). We see also from Eq. 40 that the SE is strictly decreasing and since it is bounded, it therefore converges. A further consequence of the monotonous decrease of the SE from Eq. 40 is that under the condition that the dictionary is at least complete, it convergences to the exact reconstruction [Mallat, 1998, p.414]. Moreover, Frossard and Vandergheynst [2001] has shown that it is still true up to an upper bound in a case where the coefficients are quantized.

Theorem 4 (Transformation invariance of the representation) *If the architecture $\mathbb{A}$ is invariant to a transform T (and therefore, there exist a dual transform F between the elements of the dictionary), then the spike list obtained for the transformed image $T(\mathbb{L})$ is the transformed spike list of the original image.*

$$MP(T(\mathbb{L}) = F(MP(\mathbb{L})) \qquad (44) \quad \square$$

This feature complies with our constraint that we set at the beginning of this section. In practice, this will be applied in low-level areas for usual spatial transform such as translations and scalings but also to shifts in time. However, since both the images and the dictionaries are finite, the invariance will always be approximate for these transforms.

Though simple, the greedy pursuit is a complex non-linear algorithm. In fact, the study of its behavior is non trivial and may involve chaotic dynamics [Davis, 1994]. In particular, it is obvious that the choice that is made at a giving step may influence all future steps. This implies that a failed match may propagate wrong information to following steps and therefore that the probability of a failure grows higher as the rank increases. These properties are discussed in [Perrinet et al., 2004] and in particular we illustrated that the speed of convergence increases as the dictionary becomes more over-complete so that it provides an efficient representation for natural scenes in image processing tasks.

3.1.3 Results for a multi-resolution model of the retina

This algorithm was first tested by extending the model of the retina presented in Sec. 2.2 with an over-complete set of DOG filters corresponding to more accurate neuro-physiological data. Considering the same spike coding scheme, we may ask

whether an increase in the number of filters used to describe the image can enhance the representation, *i.e.* if there would be an advantage in using an over-complete representation in the retina. The filters are thus defined as the standard dyadic scale, but the image pyramid now includes respectively $\{1, 2, 4, 8\}$ scales *per octave*, *i.e.* the scale level characteristic variances now grow as $\sigma(s) = \sigma(1).\rho^s$ where s is the scale index and $\rho = \{2, \sqrt{2}, \sqrt[4]{2}, \sqrt[8]{2}\}$.
These experiments proved that as the number of neurons increased, the coefficients decreased more rapidly as a function of the relative rank and also the MSE. This behavior is understandable, because choosing a higher number of filters allows the construction of a more fine grained multi-scale representation of the image. In fact, the number of neurons is multiplied by a factor of approximately $\chi = (1 - \rho^{-2})^{-1}$. This results in our different cases to an over-completeness of respectively $\{4/3, 2, 2 + \sqrt{2} \sim 3.41, 1/(1 - 1/\sqrt[4]{2}) \sim 6.28\}$. The information (in bits) needed to code the address of each spike (position and scale) is thus $\log_2(n_{pixel}) + \log_2(1 - (1/\rho)^2) + 1$ (n_{pixel} being the number of pixels and one bit being allocated for the polarity). We may therefore compute the performance of the coding scheme in terms of the mean decrease in MSE as a function of the number of bits necessary to code the spike list (see Fig. 13-Left). However, the situation is different if we compare the trade-off between efficiency (MSE decrease) and the architecture's complexity (we assumed here that it is proportional to the number of neurons). We obtained different results as a function of the degree of over-completeness (see Fig. 13-Right) and thus conclude that under this constraint, the greedy dyadic algorithm seems to be optimal in the retina [Perrinet et al., 2004].

This appears to be mainly due to the nature of DOG filters (and to circularly symmetric wavelet filters in general) which to a certain extent overlap too much and do not capture any new information since their complexity is low. In fact, the evolution of the retina is certainly constrained by its function, so that the argument may be reversed. First, the retina plays a key role in the visual pathways since it is the first processing layer : it is therefore very demanding in terms of robustness and the neurons are highly active. Moreover, the eyes are in a wide range of living species are mobile elements which permit the active exploration of the visual environment. Thus, the number of neurons in the retina is presumably limited not only by the total energy it can devote but also by physical restrictions such as the size of the optic nerve. Since this number is limited (its over-completeness is limited), the representation may only use more general filters.

We further compared the method we describe here with similar techniques used to yield sparse and efficient codes such as the conjugate gradient method used by Olshausen and Field [1997]. We used a similar context and architecture as these experiments and used in particular the database from the Sparsenet algorithm. Namely, we used a set of 10^5 10×10 patches (so that $M = 100$) from whitened images drawn from a database of natural images. From the relation between the likelihood of having recovered the signal and the squared error in the new metric, the mean squared reconstruction error (L2-norm) is an appropriate measure of the coding efficiency for these whitened images. This measure represents the mean accuracy (in terms of the logarithm of a probability) between the data and the representation. We compared here this measure for different definitions and values for the "sparseness". First, by changing an internal parameter tuning the compromise between reconstruction error and sparsity (namely the estimated variance of the noise for the conjugate gradient method and the stopping criteria in the pursuit),

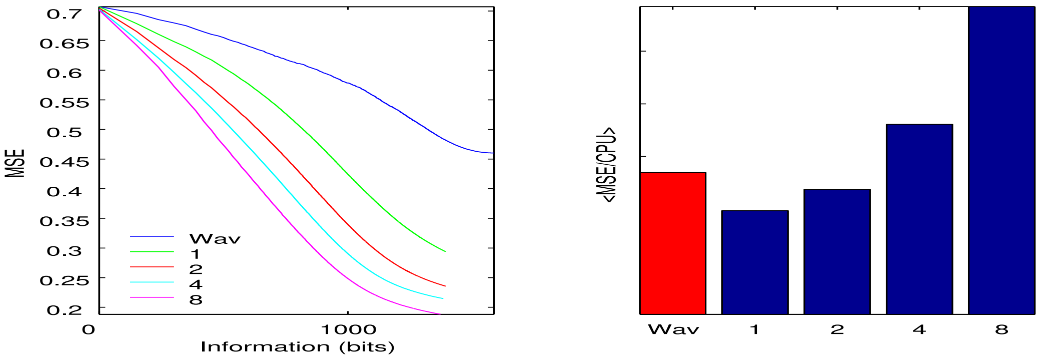

Figure 13: **Is the spike representation over-complete in the retina?** *(Left)* We compared the progressive transmission of information for different degrees of over-completeness in the retina by plotting the average MSE of the residual as a function of the information to code the spike list (in logarithmic scale, propagation up to 12.5% of the relative rank for clarity). The set of neurons used rotation symmetric Mexican hat filters, with scales from layer to layer growing as $\rho = \{2, \sqrt{2}, \sqrt[4]{2}, \sqrt[8]{2}\}$ (and denoted on the legend respectively as 1, 2, 4 and 8). As a comparison we plotted the method used in [van Rullen and Thorpe, 2001] (line 'Wav'). As a function of rank, the MSE decreases more rapidly for increasing degrees of over-completeness. *(Right)* But if we plot the trade-off of MSE with CPU usage as a function of the over-completeness, we find that for the same amount of information the adaptive dyadic strategy is optimal. One should note that the results of the method described in the text is better than the wavelet method of [van Rullen and Thorpe, 2001] since it is adaptative.

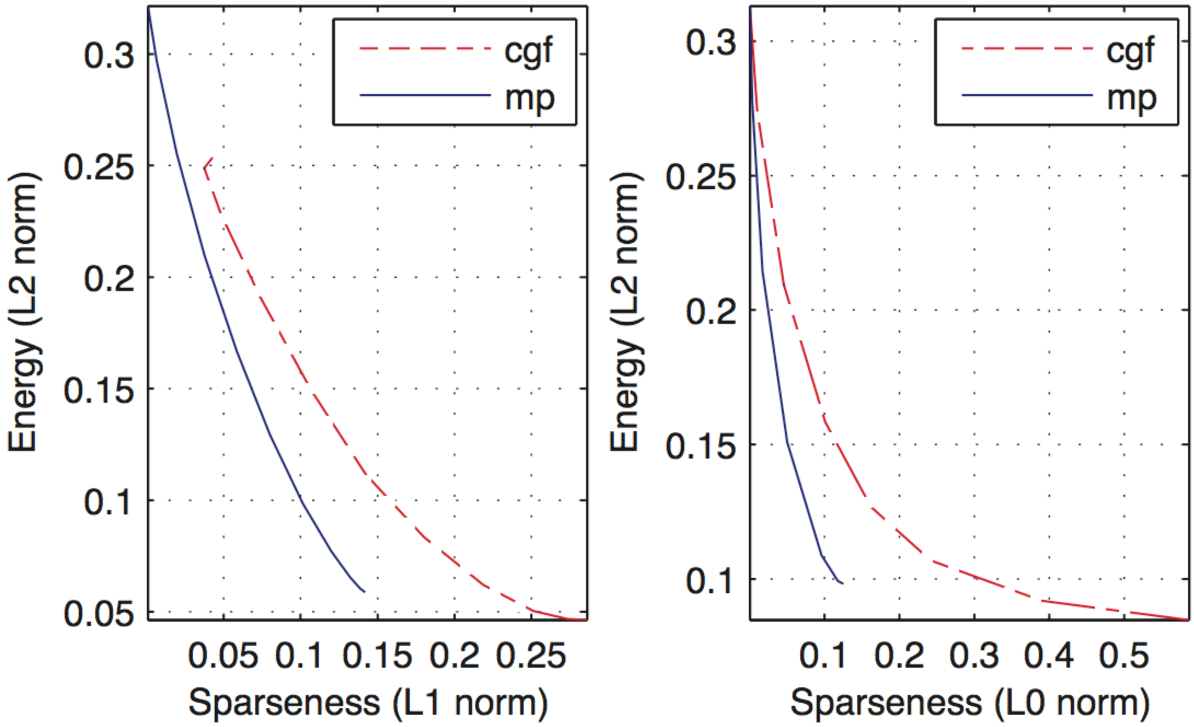

Figure 14: **Efficiency of the matching pursuit compared to conjugate gradient**. We compared here the matching pursuit ('mp') method with the classical conjugate gradient function ('cgf') method as is used in [Olshausen and Field, 1997]. We present the results for the coding of a set of image patches drawn from a database of natural images. These results were obtained with the same fixed dictionary of edges for both methods. We plot the mean final residual error for two definitions of sparseness: **(Left)** the mean absolute sum of the coefficients and **(Right)** the number of active (or non-zero) coefficients (the coding step for MP). For this architecture, the sparse spike coding scheme appears to be more efficient to code natural image patches.

one could yield different mean residual error with different mean absolute value of the coefficients (see Fig. 14, left) or L1-norm. In a second experiment, we compared the efficiency of the greedy pursuit while varying the number of active coefficients (the L0-norm), that is the rank of the pursuit (as defined in Eq. 29). To compare this method with the conjugate gradient, a first pass of the latter method was assigning for a fixed number of active coefficients the best neurons while a second pass optimized the coefficients for this set of "active" vectors (see Fig. 14, right). Computationally, the complexity of the algorithms and the time required by both methods was similar, though an advantage came for MP on larger dimensions (for $M > 200$). However, the pursuit is by construction more adapted to provide a progressive and dynamical result while the conjugate gradient method had to be recomputed for every set of parameter. Best results are those giving a lower error for a given sparsity or a lower sparseness (better compression) for the same error. In both cases, the Sparse Spike Coding provides a coding paradigm which is of better efficiency as the conjugate gradient.

3.1.4 Neural implementation of Sparse Spike Coding

We will derive now an implementation of this algorithm using a network of spiking neurons [Perrinet, 2005]. It is based on the same feed-forward architecture as the perceptron (see Fig. 15) since this architecture provides a simple *ArgMax* operator (see Sec. 1.1.3) and we will implement the greedy pursuit using *lateral interactions*. To implement the computation of the match of an input with stored patterns, we first define a dictionary which will be implemented by normalized weight vectors $\mathbb{A}_j$ and —assuming that the raw input was pre-processed as described in Sec. 2.1.4— the input activity $\mathbb{L}$ is decorrelated. The linear feed-forward perceptron integrates synaptically the input into an initial activity C_j such that

$$C_j = p_j . < \mathbb{L}, \mathbb{A}_j > \tag{45}$$

The neurons are duplicated with opposite polarity $p_j = \pm 1$ so that $C_j = p_j.|C_j|$ to model the ON / OFF symmetry of simple cells [Ringach, 2002]. The scalar projection will therefore drive the potential of the neuron.

The activity is represented by a driving current $C_j(t)$ that drives the potential V_j of leaky Integrate-and-Fire neurons [Lapicque, 1907] from the initialization time. This model gives a good fit of the dynamical behavior of neurons [Carandini et al., 1997]. For illustration purposes, the dynamics of the neurons will here be modeled by a simple linear integration of the driving current C_j (other monotonic integration schemes lead to similar formulations):

$$\tau.\frac{d}{dt}V_j = -V_j + \frac{1}{g_j}.C_j \text{ if } V_j \leq \theta \tag{46}$$

where τ is the time constant of membrane integration and g_j an inverse gain (in Siemens if C_j is considered as a current) corresponding to the conductance of the membrane (this may act as a modulator in time, which is important in integrating the influence of context).

Neurons generate a spike when their potential reach an arbitrary threshold θ that we set here to 1. We may predict from the monotonous integration that the first neuron to generate a spike will be the one that corresponds to the maximal rectified

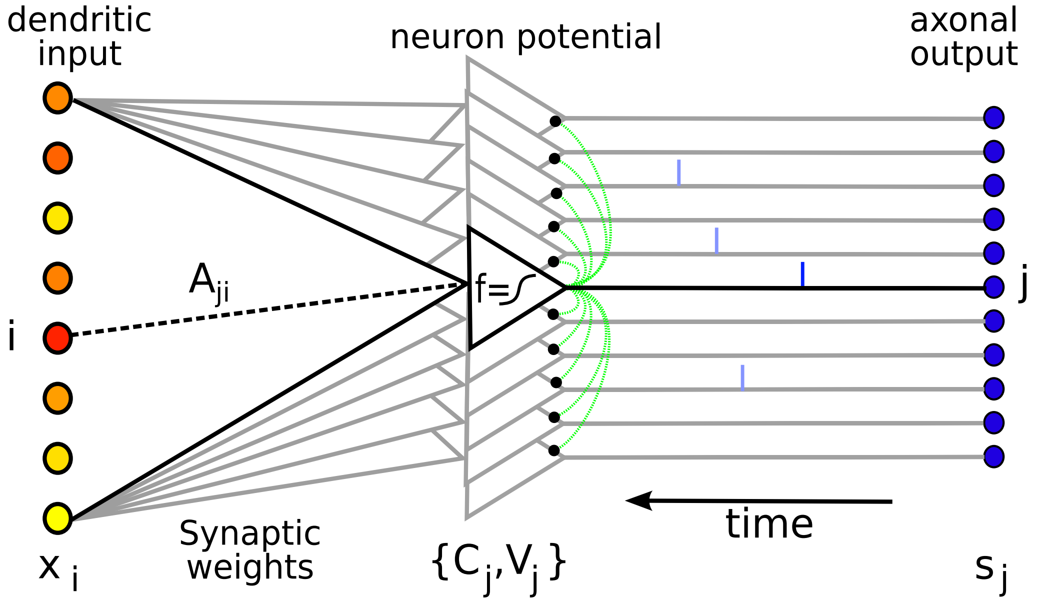

Figure 15: **Model of a neuronal layer as a communication channel**. To understand the content of neural activity, we consider here that the neuronal layer implements the inverse of a forward model (that is the analysis in Fig. 12). The architecture is similar to the perceptron: the input (noted x_i) is matched with normalized weight patterns A_{ji} (which are fixed in this section) so as to provide an integrative activation value (the membrane potential) which in turn is non-linearly transformed to achieve a membrane potential which grows proportionally to the probability of matching a feature. Spikes represent decisions that are fed back on the correlated neighboring neurons using lateral interactions (that we represented for the first spiking neuron) but also on the axonal output which yield a spiking output s_j.

scalar projection of the input signal with the weight vectors of the network, that is

$$j^* = \mathrm{ArgMax}_j \frac{C_j}{g_j} \tag{47}$$

at the latency

$$t^* = -\tau . \log(1 - \frac{\theta . g_j^*}{C_{j^*}}) \tag{48}$$

This is therefore a simple and biologically plausible implementation of a MAP estimate (Matching step of Eq. 31) using the parallel architecture of the network which is in contrast with the complexity of this implementation on a single-processor computer.
To further implement the greedy algorithm (the Pursuit step given in Eq. 31), we then need to implement a lateral interaction on the neighboring neuron. In our scheme the interaction should yield the same configuration in the network (activity and potential) as if the source that was detected was originally absent from the signal. In this model, if j^* is the winning neuron, the activity should have been subtracted by $|C_{j^*}|.R_{\{j,j^*\}}$ (see Eq. 31) and the potential by this value integrated over t^*. The lateral interaction is thus achieved by updating after each spike the activity of the neighboring neurons proportionally to their cross-correlation $R_{\{j,j^*\}}$ with the corresponding winning neuron:

$$C_j \leftarrow C_j - |C_{j^*}|.R_{\{j,j^*\}} \tag{49}$$

and therefore of the potential of every neuron by the potential $C_{j^*}/R_{\{j,j^*\}}.(1 - e^{-t^*/\tau})$, that is simply by:

$$V_j \leftarrow V_j - R_{\{j,j^*\}} \tag{50}$$

and then resume the algorithm. This lateral interaction is here immediate and behaves as a refractory period on the winning neuron (in fact $C_{j^*} \leftarrow 0$ and $V_{j^*} \leftarrow 0$) but also on correlated neurons. It involves a subtractive hyper-polarizing term on the potential and on the activity. Biologically, it is improbable that the lateral interaction could be instantaneous, but this lateral interaction could be implemented in a fast manner using a lateral interaction mediated by fast-spike inter-neurons[38]. Finally, this simple implementation therefore implements the Matching Pursuit algorithm that we defined in Eq. 31 and we will apply it to simple visual tasks.
Moreover, we may as in Sec. 2.2.2 use the regularity of the decrease of the coefficients as a function of their rank to be able to reconstruct the signal[39]. A similar equation as Eq. 25 may be written so that the energy of the quantization error adds to the error of the coefficients given by the matching Pursuit [Frossard and Vandergheynst, 2001]. This could be implemented in cortical areas by shunting inhibition [Borg-Graham, 1999] or synaptic depression. The best estimator for the squared error is therefore the one minimizing the variance, that is the mean of the absolute coefficient:

$$\mathtt{LUT}(r) = E[|C_j^r|] \tag{51}$$

[38] It should be stressed that we don't explicitly integrate a refractory period in these equations.

[39] It should be again noted that this is not actually used in the visual system (there is no reconstruction of the image) but that this scheme provides a tool to quantify the information transfer contained in the spike list thanks to Eqs. 33 and 40.

This representation proposes an alternative to classical paradigms of neural coding such as the spike-rate coding approach of the *perceptron* (see Fig. 15). Instead of coding information in the mean firing frequency of neurons, it proposes an original approach solving the problem that we defined above. It uses a distributed probabilistic representation constituted by two signals. On one side, we assumed here that the continuous activity of neurons (such as the membrane potential) in the layer represented the evidence of a correct match such as defined in Sec. 2. On the other side, the discrete spiking signal signifies a set of elementary decisions made by the neurons.

To illustrate the properties of the algorithm, we may model a network of linear Integrate-and-Fire neurons forming a simple model of an hyper-column in the granular layer of the primary visual area (V1). This model consist of an isolated network of 16 neurons selective to different orientations of contours and which are modeled as Gabor filters (which are here symmetric with circular envelopes, see Sec. 3.3 for a more detailed account on V1). We compared a pure feed-forward model to a network implementing the lateral interactions that we described above (see Eq. 49 and 50). We show here the resulting spiking activity when one of the preferred stimuli (the horizontal edge) was continuously presented from time $t = 0$ (see Fig. 16).

We observe that the neuron corresponding to that preferred stimulus fires with the shortest latency but also produces the highest spike rate. Moreover, the activity of the neurons corresponding to non-preferred directions shows a lower spiking activity when implementing the greedy pursuit. This dynamic reflects the lateral interaction (here an inhibition to the positively correlated neurons) generated at every spike which is observed in V1 [Celebrini et al., 1993]. In fact, compared to the linear model, the latency and the frequency of the neighboring neurons show a sharper response for neighboring edge orientations (see Fig. 16) which corresponds to the high selectivity observed in simple cells from V1 [Ringach et al., 2002]. The selectivity of this model was compared with the model of *divisive normalization* [Schwartz and Simoncelli, 2001], suggesting that this simple implementation of Integrate-and-Fire neurons —linked by lateral interactions and removing dynamically the redundancy in the signal— could provide a functional model for the complex processing occurring in cortical areas.

3.2 Building adaptive low-level vision systems

As was emphasized in the first section, the connectionist view of cognitive functions emphasizes on the construction of a network of similar micro-circuits (see Sec. 1.1.3). To allow an efficient processing of the input, these should therefore *adapt* according to variable internal parameters (such as the weights of the neural connections) thanks to a generic "learning program" . However we miss a general theory for learning in the CNS and in particular for low-level visual system. We will here try to define and study a learning rule based on an efficiency criteria at the level of a neural assembly and then apply it to a model of adaptation in the primary visual cortex.

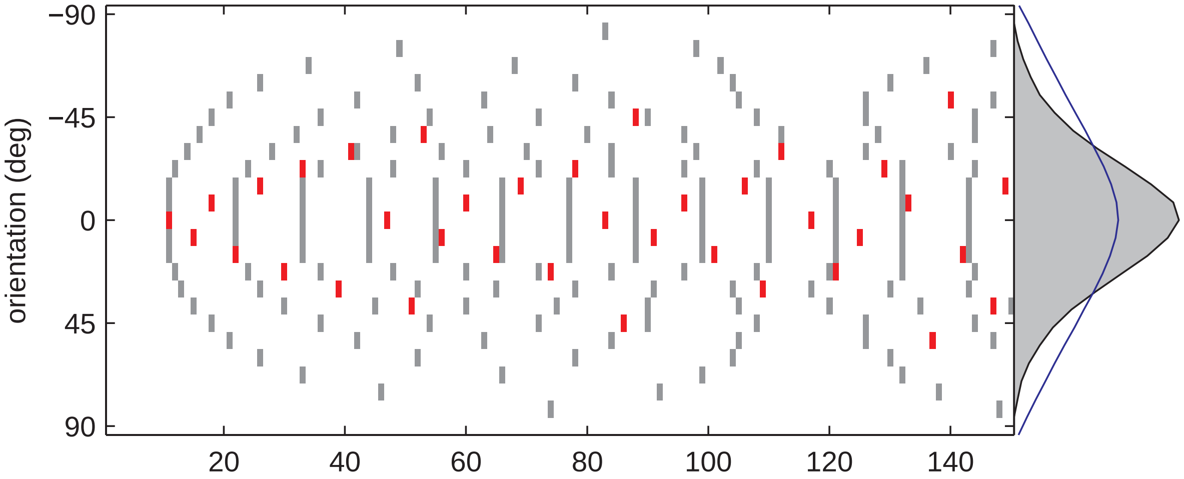

Figure 16: **Implementation of the greedy pursuit using Integrate-and-Fire Neurons**. We simulated here the activity of a network of Integrate-and-Fire neurons tuned to form a simple model of an hyper-column in the primary visual area (V1) to the presentation of a horizontal edge at $t = 0$. We show in this image the output spiking activity of 16 neurons tuned for different orientations for the feed-forward (black bars) and the sparse spike coding (white bars) models during the first 150 ms. In this latter model, the correlation linked to the information already detected is propagated as a hyper-polarizing and shunting lateral interaction to the neighboring neurons: the response in both latency and spiking frequency to the oriented edge is clearly more selective. *(Right Inset)* Output spike firing rate to the presentation of a horizontal edge at time $t = 0$. For the linear feed-forward model (plain line), the sparse spike coding scheme (filled curve) for different orientations of the input stimulus. The narrower tuning curve for the latter method represents a more selective response to the features learned in synaptic weights and mimics the behavior of the neural response in the primary visual area [Ringach, 2002].

3.2.1 Sparse Spike Learning: adapting towards efficient representations

Let's first formalize the long-term goal of a neural assembly —such as cortical mini-columns defined in Sec. 1.2.2— as the optimization of the efficiency defined in the previous section. A major keystone in modeling learning in neural assemblies was established by Hebb [1949]. It simply stated that :

> "When an axon of cell A is near enough to excite a cell B and repeatedly or persistently takes part in firing it, some growth process or metabolic change takes place in one or both cells such that A's efficiency, as one of the cells firing B, is increased."

It should be noted that this "Hebb rule" was only a small fraction of Hebb's theory which insisted on the notion of *cell assemblies* and of associative memories as was defined above in Sec. 1.3.3. In this aspect, the rule takes another formulation than what is usually applied to the single neuron and rather focuses on populations of neurons. In fact, applying simply the "Hebb rule" on a linear representation using the correlation generates preferred directions as the axis of a Principal Components Analysis [Oja, 1982], which is in particular known to badly represent natural visual scenes (but which may serve as a processing as explained in Sec. 2.1.4). In our framework, we would rather interpret this statement as adjusting the relation between different points in the inferential network, for instance : "if a set of events A is more probable than expected in successfully *causing* an event B, then increase their link to adjust the prediction of B knowing the set A".

The dictionary should adapt in an unsupervised manner as a function of the input 's statistics so that the coding is the most efficient, that is that the data is at best explained by the LGM model. Inspired by the measure of the log-likelihood of the data knowing the model as an upper-bound for the description length [Shannon and Weaver, 1964], we may thus try to maximize it over a subset $\mathcal{I}$ of images corresponding to natural behavioral conditions:

$$\mathcal{L} = E[-\log(P(\mathbb{L}|\mathbb{A})]_{\mathbb{L}\in\mathcal{I}} \tag{52}$$

where $E[.]$ represents the mean. However, maximizing this cost requires the evaluation over a huge set of images and conditions, a task which is not compatible with biological constraints.

Based on the work of Olshausen and Field [1997], we will derive a learning as an algorithm gradually optimizing this efficiency criteria. A solution is to use the Sparse Spike Coding representation that we defined in the previous section since we have an evaluation of the log-likelihood by the distance of the residual image to the selected filter (see Eq. 34), the rest being regarded as a perturbation which should cancel out by integrating it in time. Translating the Hebb rule to our model of sensory coding (see Fig. 12) at every single coding step n, it could be translated in "if an image $\mathbb{L}$ is causing the efficient response of j, then $\mathbb{A}_j$ should be adjusted toward $\mathbb{L}$". At this step of the pursuit and using the gradient descent approach as in [Olshausen and Field, 1997], we infer that we may slowly modify the weight vector corresponding to the winning filter $\mathbb{A}_{j^{(n)}}$. This learning consists in taking the winning weight vector closer to $\frac{\mathbb{L}^{(n)}}{s^{(n)}}$ by applying:

$$\mathbb{A}_{j^{(n)}} \leftarrow \mathbb{A}_{j^{(n)}} + \eta.s^{(n)}.(\mathbb{L}^{(n-1)} - s^{(n)}.\mathbb{A}_{j^{(n)}}) \tag{53}$$

or equivalently

$$\mathbb{A}_{j^{(n)}} \leftarrow \mathbb{A}_{j^{(n)}} + \eta.s^{(n)}.\mathbb{L}^{(n)} \tag{54}$$

where η is the learning rate (this is exactly Eq. 17 in [Olshausen and Field, 1997]). By slowly adjusting $\mathbb{A}$, this will therefore in the long term enhance the coding efficiency of the coding algorithm. Moreover, this method provides when it converges a set of optimal filters in the sense that their activity is optimally independent [Lewicki and Sejnowski, 2000]. It is thus a hebbian-like learning, but instead of being applied to the linear representation it is based on the sparse representation implemented by the Sparse Spike Coding.

Moreover, to keep the assumption of an uniform prior (see Sec. 2.1.4), it is important to include a regulatory mechanism in the learning algorithm. This will correspond to an implementation of the homeostatic aspect of the "fair competition" among neurons as stated in Sec. 1.2.2. In fact, the first neurons to learn will be more prone to be selected again and the following representations will have a non-uniform prior on the probability of being activated. Inspired by the SparseNet algorithm, we introduced a gain control mechanism which influenced the choice of the winning neurons. By taking advantage of the probabilistic representation by explicitly changing the probability of being selected, we may for instance modify Eq. 31 as

$$j^{(n)} = \mathrm{ArgMax}_j[P(\mathbb{L}^{(n)}|\mathbb{A}_j)/P^{(n)}(j)] \tag{55}$$

where $P^{(n)}(j)$ is the probability of choosing a neuron j at rank n computed over the previous coding steps [40]. This provides an homeostatic parameter which ensures that the competition in the neural assembly keeps fair in a certain time window.

This sparse hebbian rule is therefore a simple approach that has many similarities with other algorithms. It may be proved that in a stationary environment, this rule would yield progressively sparser representations [Perrinet, 2004a] and would converge to the optimal solution defined by Independent Component Analysis [Bell and Sejnowski, 1995, 1997]. In fact, the learning rate gives a temporal scale and the learning focuses on slowly varying features as is used in Slow Feature Analysis [Wiskott and Sejnowski, 2002]. We may also compare the algorithm as Vector Quantization (VQ) which would be applied on a multidimensional unit sphere corresponding to the different features. Our approach is therefore relatively conservative, and we will see its implications on modeling learning in a population of neurons.

3.2.2 Neural implementation of Sparse Spike Learning

Similarly as the Sparse Spike Coding algorithm and since they are based on similar mechanisms, this set of linear event-based computations are easily implemented in a network of spiking neurons. We may extend the neuronal implementation of the Sparse Spike Coding by integrating the Hebbian rule at every event. Using similar assumptions and notations as in Sec. 3.1.4, only the winning neuron will be modified and as in 54 :

$$\mathbb{A}_{j^*} \leftarrow \mathbb{A}_{j^*} + \frac{1}{\tau}.s^*.\mathbb{L}^* \tag{56}$$

where $\tau = 1/\eta$ is the learning time constant (in number of spikes) as in Eq. 22 and $\mathbb{L}^*$ the residual image. We may at the same time modify the correlations (that is the

[40] We will see that it is easy to compute $P(\mathbb{L}^{(n)}|\mathbb{A}_j)$ using a similar LUT as in Sec. 2.1.1. This homeostatic constraint could be included in the generic cost defined above since a variability in the behavior of the neurons would result in a quantization cost as in Eq. 25.

lateral connections) with

$$\Delta R_{\{j,j^*\}} = \frac{1}{\tau}(< \frac{\mathbb{L}^*}{C_{j^*}}, \mathbb{A}_j > - R_{\{j,j^*\}}) = \frac{1}{\tau}(\frac{C_j}{C_{j^*}} - R_{\{j,j^*\}}) \tag{57}$$

and implement a rule to learn the LUT :

$$\mathtt{LUT}(r) \leftarrow (1 - 1/\tau_{\mathtt{LUT}}).\mathtt{LUT}(r) + 1/\tau_{\mathtt{LUT}}.C_{j^*} \tag{58}$$

where $\tau_{\mathtt{LUT}}$ is the learning time constant of this rule. For a given value C_j this function gives a way to compute $P(\mathbb{L}^{(n)}|\mathbb{A}_j)$. Finally, we may implement the homeostatic rule in this distributed approach by changing the gain of the activity. More simply, using Eq. 48, we may vary the threshold by setting a different threshold θ_j for every neuron and at step n:

$$\theta_j = -\theta.P^{(n)} \text{ with} \tag{59}$$

$$P^{(n)}(j) \leftarrow (1 - 1/\tau_{homeo}).P^{(n)}(j) + 1/\tau_{homeo}.\delta(j^{(n)} = j^*) \tag{60}$$

where τ_{homeo} is the homeostatic time constant and δ is the Kronecker function.
As a causal temporal learning rule, the sparse hebbian rule is related to Spike-Time Dependent Plasticity (STDP) [Debanne et al., 1995; Bell et al., 1997; Bi and Poo, 1998; Abbott and Nelson, 2000]. In fact, this algorithm enhances temporally causal relationships, and is by construction similar to STDP, but the exact learning modification will depend in this functionalist view to the activity of the whole population of neurons. The net weight change between neurons may thus be alternatively hebbian and non-hebbian, and the learning time window in particular may have varying shapes during the learning. In particular this shows that the hebbian learning on the linear representation defined in the method of van Rullen and Thorpe [2001], that is the learning proposed in [Guyonneau et al., 2005] could not yield independent components when applied to multiple inputs. Our approach may therefore explain the variety of learning time windows that may be observed [Abbott and Nelson, 2000] and that may be a consequence of this more general rule. The importance of the learning lies instead in constantly maximizing the efficiency of the network which should —in a stationary environment— increase monotonously.

3.2.3 Performance of the Sparse Spike Learning and Coding

We compared this "sparse-hebbian" learning scheme with the SparseNet algorithm. In fact, our algorithm (equations 30, 31 with 55 and 54) mainly differs by the method used to obtain the sparse representation. The latter uses the conjugate gradient method [Olshausen and Field, 1997] to optimize a trade-off between sparseness and reconstruction quality. We used a similar context and architecture as these experiments and used in particular the database of inputs of the SparseNet algorithm. Here, we show the results for 12 × 12 patches (so that $M = 144$) from the whitened images and we chose to learn 169 filters. We optimized some parameters for instance by evaluating the variance of noise in the database. However, varying every learning parameter showed to rarely change the results qualitatively and illustrated the stability of both algorithms[41]. The convergence was mostly quick

[41] In particular, we studied modifications —such as the use of natural gradient [Lewicki and Sejnowski, 2000] or using rectified coefficients— with little qualitative differences.

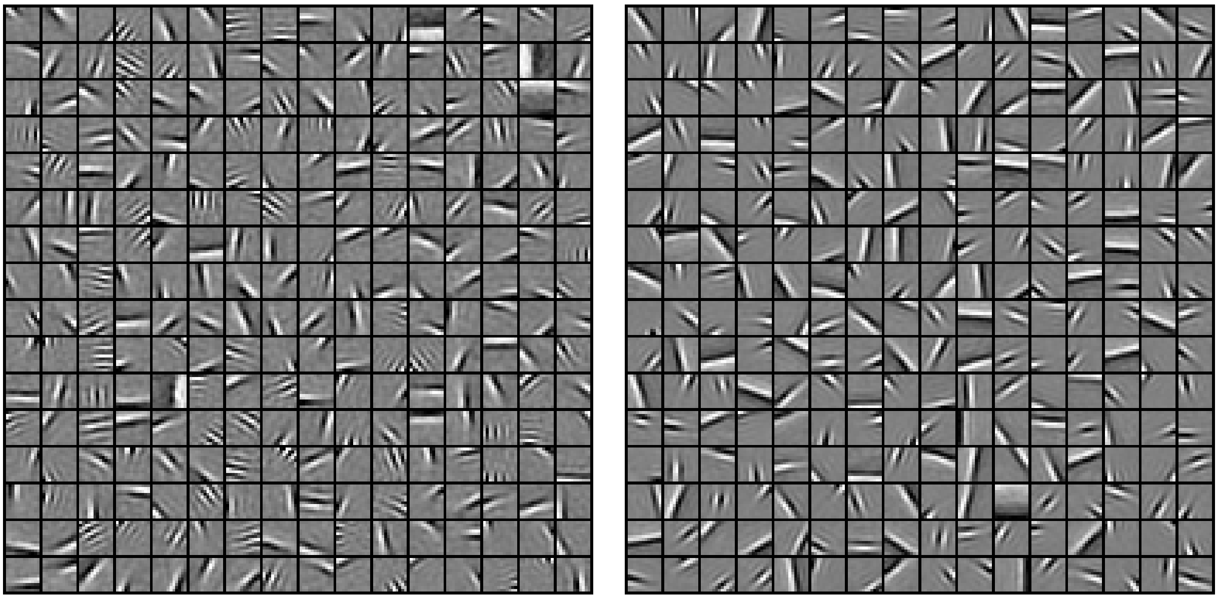

Figure 17: **Results using the sparse-hebbian learning scheme with** 196 **filters.** Starting with random filters, we compared here the results of the learning schemes using *(Left)* the classical conjugate gradient function ('cgf') method as is used in [Olshausen and Field, 1997] with *(Right)* the Sparse Spike Coding method. Results replicate the original results of [Olshausen and Field, 1997] and are similar for both methods: both dictionary consist at convergence of Gabor-like filters which are similar to the receptive fields of simple cells in the primary visual cortex [Ringach, 2002]. Edges appear in these conditions to be the independent components of natural images. However, it should be noted that the Sparse Spike Coding method introduces less localized filters and a higher proportion of high-frequency Gabors when no efficient homeostatic rule was defined [Perrinet, 2004a, 2006]. All scripts generating figures and control experiments are available, see Sec. 3.4.3)

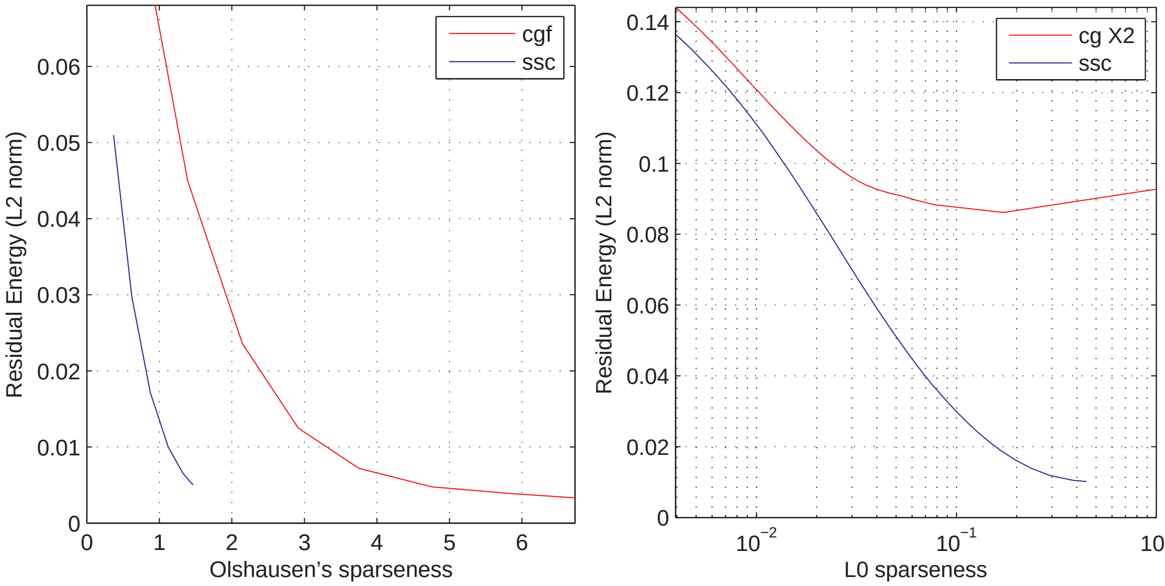

Figure 18: **Efficiency of the matching pursuit compared to conjugate gradient**. We compared here the matching pursuit ('ssc') method with the classical conjugate gradient function ('cgf') method as is used in [Olshausen and Field, 1997] for the coding of a set of 10000 image patches drawn from a database of natural images. We plot the mean final residual error and the signal to noise ratio (in dB) as a function of two definitions of sparseness : *(Left)* the mean absolute sum of the coefficients (the sparseness defined in [Olshausen and Field, 1997]) and *(Right)* the number of active (or non-zero) coefficients (the coding step for MP) which provides an estimate of the coding efficiency (in bits) for the image patch. The conjugate gradient function approach was applied twice ('cg X2', see text). For this architecture, the sparse spike coding scheme proves to be more efficient to learn and code natural image patches. The approximate solution for the hard constraint (Eq. 29) is therefore more efficient on both constraints costs.

(after ∼ 500 learning steps for $\eta = 1/20$) and the homeostatic rule was efficient to keep the prior flat. As in SparseNet, we observe the emergence of similar structure as the receptive fields of simple cells in the primary visual cortex [Ringach, 2002] (see Fig. 17). It thus experimentally provides a simple neural implementation for Independent Component Analysis [Lewicki and Sejnowski, 2000].

We also compared the efficiency of both coding algorithms on the learned representation basis (see Fig. 18) with a similar method as in Fig. 14. Computationally, the complexity of the algorithms and the time required by both methods was similar on the different simulations on a computer. However, the Sparse Spike Coding is by construction more adapted to a parallel architecture. It also provides a progressive result while the conjugate gradient method had to be recomputed for different number of coefficients. Best results are those giving a lower error (or higher SNR) for a given sparsity or a lower sparseness (better compression) for the same error. In both cases, the Sparse Spike Coding provides a coding paradigm which is of better efficiency as the conjugate gradient and of lower complexity.

3.3 A model of computations in the Primary Visual Cortex (V1)

To illustrate the efficiency of the Sparse Spike Coding on a larger and more realistic scale model of a low-level visual area, we applied the algorithm to a model of the primary visual cortex (V1). As was stated in Sec. 1.3.1, we will assume that the function of V1 is to represent a "sketch" of the visual scene. We will restrict here the model to the response to a flashed static grayscale image (as in Sec. 2.1.3) and in particular study the dynamics of the response to the oriented edges that delimit visual objects. We will present some particular extensions permitted by this algorithm such as adding long range prediction interactions which may enhance the overall efficiency of the algorithm and that we will put in a neuro-physiological perspective.

In fact, a model of the edge content of natural images is of great importance for understanding the visual scene. It is widely used in image processing for multiple tasks the human visual system is supposed to perform such as segmenting objects by their shape or denoising using prior knowledge on the statistics of edges. This has led to numerous formalizations [Marr, 1980; Canny, 1986; Deriche, 1987; Castan et al., 1990] which we will try to understand in the framework of our formalization. This extension of the previous framework will serve to forge a general framework of an efficient dynamical distributed event-based computing scheme.

3.3.1 Multi-scale edge representation: Sparse Edge Coding

In natural scenes, edges exist at multiple scales and we may extend the previous algorithm (see Sec. 3.1.4) to a more appropriate architecture such as multi-scale edge detection [Mallat, 1998, p. 452]. As in section 2.2, we will generate all filters by replicating mother functions according to translations and scalings so as to transform the image in a map as in Sec. 1.3.3. As pointed by Jones and Palmer [1987], simple cells in V1 approximately fit Gabor filters which in turn are well approximated by partial derivatives of 2D-Gaussians. In particular, from the sensibility of vision to the relative frequency (an effect known as Weber's rule), the mother wavelet may be defined in the polar coordinate (r, θ) of the Fourier domain as log-Gabor filters [Field, 1987] :

$$G_k(r, \theta) = \exp(-\frac{\log^2(r/r_k)}{2.\sigma_r{}^2}).\exp(-\frac{(\theta - \theta_k)^2}{2.\sigma_\theta{}^2}) \tag{61}$$

where (r_k, θ_k) is the center of the filter and $(\sigma_r, \sigma_\theta)$ are respectively the frequency and angular bandwidth of the filter (see Fig. 19). Optimizing this representation, we may set up the tiling of the discrete wavelet architecture using 5 scales and 6 to 8 different orientations [Fischer et al., 2005a,b; Redondo et al., 2005; Fischer et al., 2007b]. In particular, the activity images are complex numbers and the argument (the angle of the complex number) of a particular neuron will in fact correspond to the phase of the Gabor filter (see Fig. 19). This analytical formalization permits thus to capture the diversity of receptive fields types in a simple abstract fashion.

As a linear over-complete transform, it fits particularly well to the Matching Pursuit algorithm. As in Sec. 2.2, the modulus operator will be used to detect the best match and we then extract at every step the particular location, phase, orientation and scale of the winning neuron. It thus implements in a simple manner a population

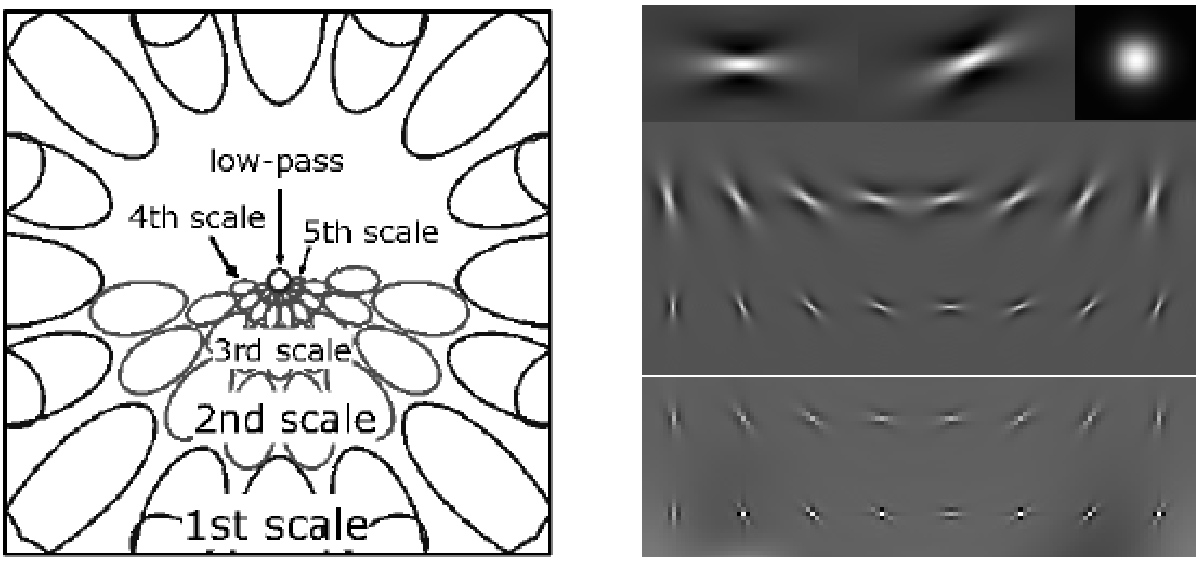

Figure 19: **Multiresolution scheme with 6 orientations and 5 scales.** *(Left)* Schematic contours of the filters in the Fourier domain. The Fourier domain origin (DC component) is located at the center of the inset and the highest frequencies lie on the border. *(Right)* Real (symmetric) part of the filters in the space domain. Scales are arranged in rows and orientations in columns. The two first scales are drawn at the bottom magnified by a factor of 4 for a better visualization. The low-pass filter is drawn in the upper-right part.

of simple cells with different phase preferences and gives a phase-independent evaluation of the edge content on the receptive field as may be observed in complex cells of V1 [Liu et al., 1992]. In this particular case, the correlation between filters could be analytically computed in the Fourier domain and the algorithm is therefore of relatively low-complexity (see [Fischer et al., 2005a] for details). Moreover, this transformation was designed to be self-invertible [Fischer et al., 2007b] to ease the reconstruction of the image from the coefficients and applications to image processing (as a consequence the point spread function from Eq. 18 is a discrete dirac function). As in Sec. 2.2.2, we observed that over a database of natural images, the decrease of coefficients as a function of their rank was smooth (see Fig. 20). This observation extends the experiments of Ruderman and Bialek [1994] which suggest that the local contrasts of different orders follow regular statistics to this non-linear transformation. The Sparse Spike Coding approach would thus yield here a small quantization error.

To evaluate the efficiency of this algorithm, we compared it with existing solutions on a set of natural images. We first computed on a first set of natural images the LUT (see Fig. 20). The LUT was then used to code the amplitude of coefficients according to their rank and was used to code a second set of 200 natural images. We therefore could compare the efficiency of this algorithm in terms of data compression with the standard JPEG method and the SEC algorithm described in [Fischer et al., 2005a]. We varied the entropy of the output signal by varying the quality parameter in JPEG and the number of chosen coefficients (that is the rank) in both other methods. Assuming that the goal of low-level sensory system is to transform the

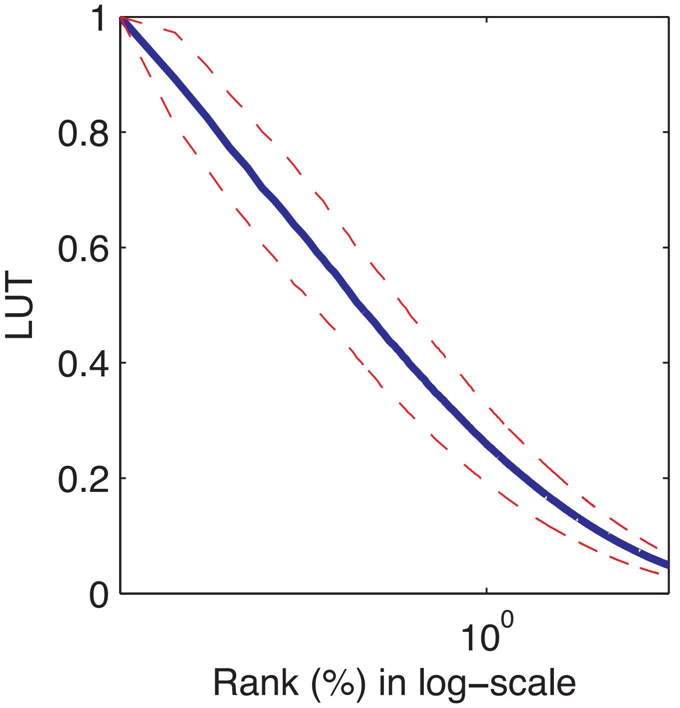

Figure 20: **Regularity of edge coefficients distribution in natural images**. We plotted for 200 natural images the mean and variance (depicted by the dotted lines representing the mean ± one standard deviation) of the decrease of coefficients values in the probabilistic Matching Pursuit scheme described above as a function of the rank. Assuming that these values represent the edge content of the images, it shows that the probability distribution of edge coefficients is regular in natural images and may be used in an efficient compression scheme. It permits to evaluate the coefficients quickly as a function of the rank, and since the transmission error being proportional to the variance to transform efficiently an analog image in a wave of spikes. It should be noted that this decrease is much more rapid than the one observed in the model retina the coefficients are below .15 after 1% of relative rank (to compare with Fig. 7).

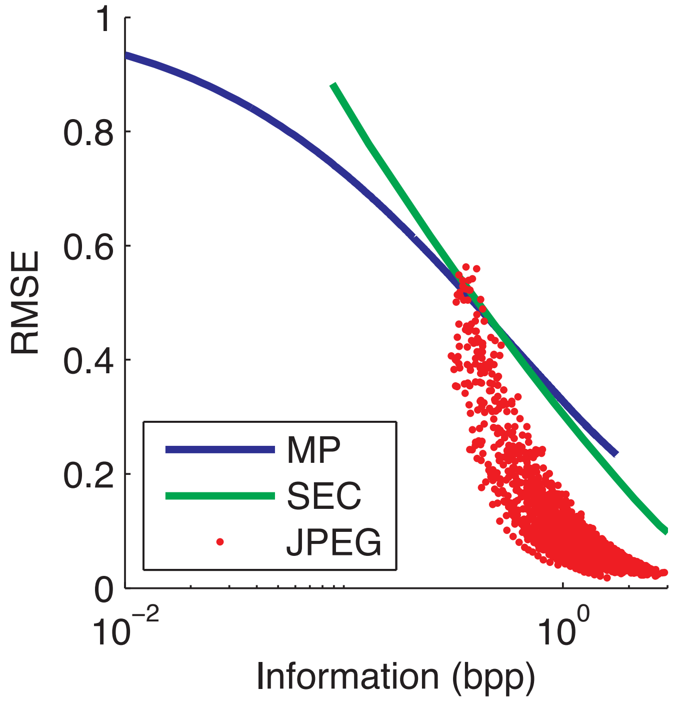

Figure 21: **Compression efficiency of MP compared to SEC and JPEG**. We compared the coding efficiency measured as the normalized root mean squared error (RMSE) as a function of the coding entropy (in bytes per pixel). A RMSE of 0 means a perfect reconstruction, while a RMSE of 1 means that the error is maximum (null image). Three phases corresponding to the three method appear in the graph. The MP method is capable of coding with very low entropy starting at relatively low efficiency, but getting progressively better. The JPEG method is not adapted to low-entropy coding but gets very good results after a medium quality parameter (around 60 and 1 bpp). The SEC method gives intermediate results of good efficiency in rather low entropy ranges.

representation while loosing the least information, we measured the RMSE (see Fig. 21) as a measure of the efficiency of the methods. The RMSE was computed after decorrelation which accounted for the spatial frequency sensibility of the human visual system (as seen in Sec. 2.1.4). As a conclusion, the MP method appears to be particularly efficient for a crude and quick "sketch" of the image. Adaptive methods as the SEC algorithm appears therefore to be a promising method for an efficient coding scheme on a wider range of behavioral situations in particular in noisy environments.

3.3.2 Handling uncertain input and filters

In Eq. 31, we made at every step the "greedy" assumption that the detection of an edge and its amplitude was correct and we explore here if we may enhance the algorithm by accounting that our knowledge is non-uniform on the image and receptive field. In fact, it is often argued that the Matching Pursuit scheme provides an approximate solution since the decisions made at every step influence the rest of the algorithm. An error at a given step may therefore be propagated and

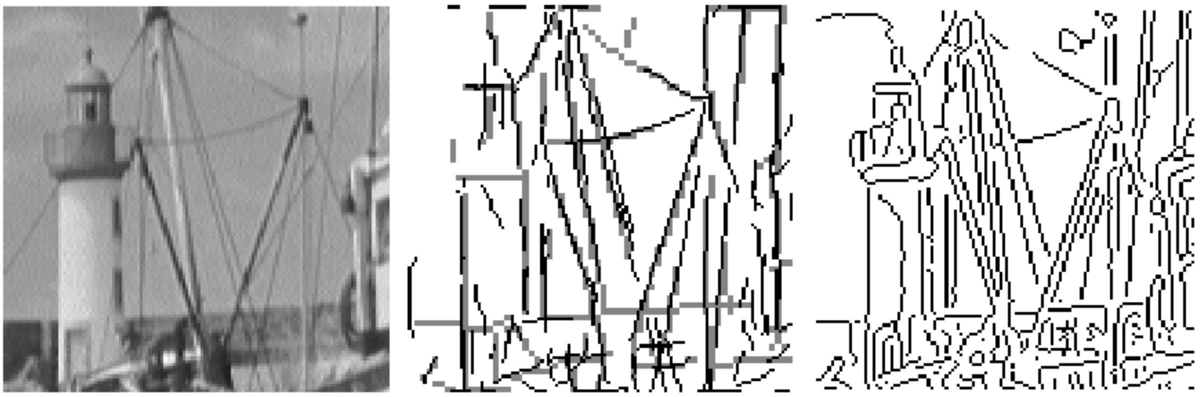

Figure 22: **Edge extraction using the multi-scale representation**. **(Left)** 128 × 128 pixel image "Boats". **(Middle)** Edges extracted by sparse Gabor wavelets. 1st scale and 2nd scale edges appear in black. The 3rd scale edges appear in gray, they are plotted larger since the 3rd scale is downsampled and offers coarser spatial localization. **(Right)** Edges extracted by Canny method [Canny, 1986; Deriche, 1987]. The SEC method extracts edges which are subjectivelly more accurate. Reconstruction of the image using these edges revealed qualitatively better results (see [Fischer et al., 2007a]).

amplified in the following steps. Though the squared error is bound to decrease, the activity may at certain steps increase. A first solution is to introduce a prior in Eq. 13 that favors small amplitudes and which implies that only a fraction of the correlation is removed to neighboring neurons is removed at every step (this "smoothness" regularization factor is determined by the variance of the prior). This comes however on a price for the sparsity of the signal (correlated neurons are more likely to be selected when this factor increases). Another solution is to take into account our knowledge of the evidence of the inferential decision: the "greediness" of the algorithm being now regulated by our actual evidence of the action potential. However, a problem with Matching Pursuit [Mallat and Zhang, 1993] is that any decision in this recursive scheme is propagated to the following steps and that the algorithm is progressively more prone to detection errors. Thanks to the interpretation of MP in a probabilistic framework [Perrinet, 2004b], we may lower the risk of false detection by explicitly taking into account the variability of image formation in space but also specifically to the knowledge of one source:

$$\mathbb{L} - s.\mathbb{A}_j = \mathbf{n}_j \tag{62}$$

with $\mathbf{n}_j$ being a gaussian centered noise image totally characterized by the image of variances $\hat{\mathbb{A}}_j$.
As in Sec. 2.1.4 and [Perrinet et al., 2005], we may derive the optimal choice as the maximum *a posteriori* now as:

$$j^* = \text{ArgMax}_j |C_j| \text{ with } C_j = \frac{\sum_i \alpha_{ji}.\mathbb{A}_{ji}.\mathbb{L}_i}{\sqrt{\sum_i \alpha_{ji}.\mathbb{A}_{ji}^2}} \tag{63}$$

with i spanning the indexes of image pixels and $\alpha_{ji} = 1/\hat{\mathbb{A}}_{ji}$ acts as a "transparency" channel on the image or the receptive field j: all pixels will not have the same weight on the decision, typically, the surround accounting for less information than

the center. This new metric measure is equivalent to a dot product and similar to the χ^2 measure, accounting for the non-uniform variability of measures in the receptive field of the neuron. It is also somewhat similar to the Mahalanobis [1936] distance if we measure statistically the variances $\hat{\mathbb{A}}_{ji}$. We may also implement in this framework a image based alpha-channel corresponding to outside the image (and for which $\alpha_i = 0$) or for retinotopic areas, as the scotoma, where the information is permanently uncertain. This is for instance present in the *blind spot*, an area of the retina where the convergence of the axons from ganglion cells doesn't allow for the reception of luminous information.
We may then resume the algorithm as in MP by removing the pattern corresponding to the winning neuron as:

$$\mathbb{L} \leftarrow \mathbb{L} - s^*.\mathbb{A}_{j^*} \text{ with } s^* = \frac{\sum_i \alpha_{j^*i}\mathbb{A}_{j^*i}.\mathbb{L}_i}{\sum_i \alpha_{j^*i}.\mathbb{A}_{j^*i}^2} \tag{64}$$

It corresponds to the mean estimated best source, and it is in this framework minimizing the risk of a false detection. This may be implemented directly by a lateral interaction using for all j

$$C_j \leftarrow C_j - s^*.R_{\{j,j^*\}} \tag{65}$$

where

$$R_{\{j,j^*\}} = \frac{\sum_i \alpha_{ji}.\mathbb{A}_{ji}.\mathbb{A}_{j^*i}}{\sqrt{\sum_i \alpha_{ji}.\mathbb{A}_{ji}^2}} \tag{66}$$

is the correlation of the neurons with the winning neuron. Note that the activity of the winning neuron is canceled.
We applied this framework to a simple reconstruction task where

1. the alpha-channel of each receptive fields was circular
2. the scotoma was defined as the border of the image and a known image of alpha values (by modeling the extent of the blind spot)

The results show that an original image could be reconstructed despite the incomplete input information. This method provides therefore an explicit method for handling uncertainties by taking advantage of the probabilistic representation. It also provides a generic method for handling edges which outperforms heuristic methods which are typically used (e.g. mirroring) and which could be of use in image processing when having an *a priori* knowledge of the non-uniformity of uncertainties.

3.3.3 Predicting cocircular edges using the local context

It is a *priori* more probable in natural images that edges are aligned according to smooth contours and we may introduce this knowledge in our algorithm. This has been exploited in numerous studies and models and has been formalized using neuro-physiological evidence [Kovacs et al., 1999; Geisler et al., 2001; Sigman et al., 2001] by defining an *associative field* [Field et al., 1993; Seriès et al., 2002]. It also correlates with the observed lateral propagation of information in V1 [Grinvald et al., 1994; Georges et al., 2002; Bair et al., 2003; Jancke et al., 2004] and could serve to understand the functionality of the flow of information at the different scales of

V1. This field modifies the excitability of neighboring neurons knowing one edge and is implemented by long-range lateral interactions modifying the threshold or equivalently the global conductance. It may be understood physically by the larger co-occurrence of oriented edges in natural scenes and is related to the Gestalt law of good continuity. It may be formally modeled as a prior on co-circular edges with a prior for low curvatures (that is for straight lines). Introducing this information will enhance the fair competition between neurons of V1 and thus increase efficiency.
This information is particularly adapted to the dynamical approach of our distributed probabilistic method. In fact, the probabilistic representation enables to combine information from different modalities, as here the information from the input image (that is the residual image at rank n) and the information from the edges already extracted at this time. This knowledge modifies the assumption of independence of the activity of neighboring cells at a larger scale than a cell's RF and we may assume that the a posteriori probability from Eq.13 is modified by the knowledge on the edge that we already extracted (noted $\mathcal{J} = \{j^{(k)}\}_{1 \leq k < n}$) by:

$$P(\{j^{(n)}, s^{(n)}\}|\mathbb{L}, \mathcal{J}) = P(\{j^{(n)}, s^{(n)}\}|\mathbb{L}).P(\{j^{(n)}, s^{(n)}\}|\mathcal{J}) \tag{67}$$

The prior from the known edges combines (by conditional independence) as

$$\log P(j^{(n)}|\{j^{(k)}\}_{1 \leq k < n}) = \sum\nolimits_{1 \leq k < n} \log P(j^{(n)}|j^{(k)}) \tag{68}$$

The probability of occurrence of an edge k knowing an edge l is here parametrized by

$$-\log P(k|l) = \log Z + \frac{d(j,k)^2}{2.\sigma_d^2} + \frac{c(j,k)^2}{2.\sigma_c^2} + \frac{\delta(j,k)^2}{2.\sigma_\delta^2} \tag{69}$$

where $d(j,k)$, $c(j,k)$ and $\delta(j,k)$, are respectively the distance, the curvature and the angle between the two edges (see Fig. 22-Left) and σ_d, σ_c and σ_δ the respective standard deviations for the measures. This parameters are tuned to match the effective probability in natural images, but from the circular definition of the edges obtained by the method, special care has to be put on the stability of this tuning. In fact, even if it is intractable on a sequential machine, the strategy used in cortical columns is certainly simpler and based on a slowly varying rule which statistically "computes" the *a priori* co-occurrence between neurons. From the extension presented in Sec. 3.3.3, we may implement a simple rule to learn associative fields. This rule is again based on an evaluation of the probability of the activation of one neuron k knowing that another neuron l was activated before. This may be written :

$$P(k|l) \leftarrow (1 - 1/\tau_{pred}).P(k|l) + 1/\tau_{pred}.\delta(l) \tag{70}$$

where τ_{pred} is the time constant of this learning rule. Finally, this approach similar to advanced algorithms in image processing such as contourlets [Le Pennec and Mallat, 2005] but it introduces a more general adaptive framework which reflects aspects of the processing occuring in V1. These extensions suggest that local information is integrated (up to the cost in metabolic "wet-ware") using a generic predictive mechanism which could be described in terms of the local micro-circuitry of cortical columns.
Let's draw an application from these formalizations. In particular, if we note in a matrix form W the linear transform and h^* the set of selected coefficients

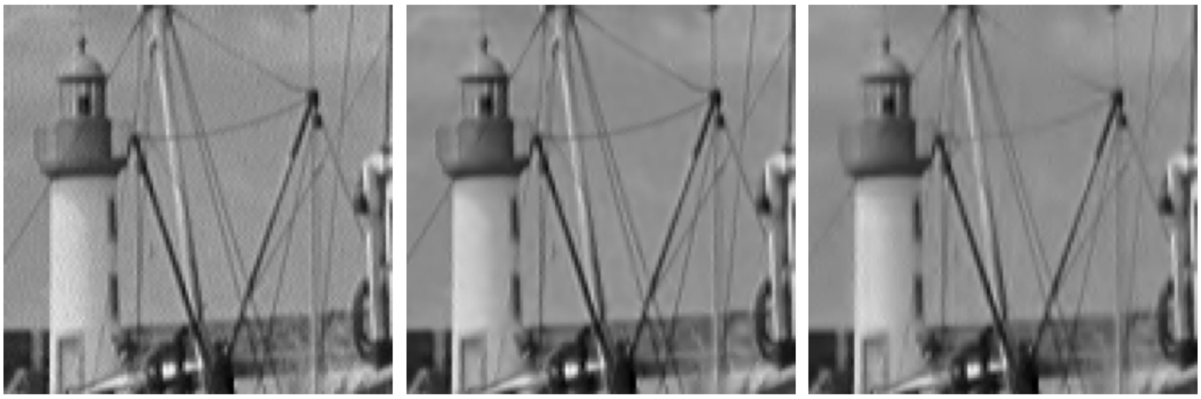

Figure 23: **Denoising results.** *(Left)* This 120 × 120 detail of the image "Boats" is corrupted by Gaussian noise for a Peak-Signal to Noise Ratio (PSNR) of 20.22 dB and contains an important level of noise. *(Middle)* After denoising by bi-orthogonal wavelets 'db4' and soft-thresholding a high level of artifacts appears, PSNR=24.65 dB. In "Boats" many details, i.e. the wires, appear smoothed and almost disappear while some noise points still remain. *(Right)* The image denoised by sparse log-Gabor wavelets shows a dramatic decrease of the level of artifacts, moreover some important details now appear preserved and the blur is lower. The PSNR shows an important improvement up to 25.30 dB.

for a perturbed image $\mathbb{L}$, it may be proved that the image minimizing the error cost for a denoising task will be given by $\overline{W}^T(\overline{W}^T\overline{W})^{-1}\mathbb{L}$ where $\overline{W}$ is the matrix corresponding to the linear transformation but limited on the selected coefficients h^*. The pseudo-inverse may be approximated by the Landweber algorithm. This was implemented with the log-gabor architecture and exhibits promising results (see Fig. 23).

3.3.4 Learning transform-invariant representations: Sparselet Analysis

As stated in the introduction of this section (see Sec. 3.1.1), a major feature of biological vision is its robustness to frequently occurring changes in the visual scene such as those produced by our self-motion: translations, zooms, rotations. We will here describe a general method to achieve a transform invariant representation by setting that for a transform in the image space, there exist a transform in the representation space. To implement this feature on a computer model for translation of the image, we will as a first approximation *impose* the weight patterns to be exactly the same at different locations and scale in a Laplacian Pyramid [Simoncelli and Freeman, 1995][42] as was done by Sallee and Olshausen [2003]. This approximation was verified on the small patches of images where we obtained filters that were similar up to a translation (see Fig. 17). The Sparse Spike Coding algorithm may be easily extended to the translation constraint [Perrinet et al., 2002] and was applied to the image pyramid. The activity of the neurons was computed at every scale

[42] In particular, we used the same representation algorithm as in [Burt and Adelson, 1983]. The invariance is therefore exact over discrete translations and dyadic scales. This procedure is based on the hypothesis that stimuli appear similarly according to these transformations and is a major constraint of the limited amount of memory (see Sec. 1.1.2).

of the pyramid using the correlation operator (that is the convolution with the central symmetric transformation of the filters) which produced the initial activities. We then used the same operator to compute the correlation between filters and to define lateral interactions. By this way, we implemented the Sparse Spike Coding algorithm in this architecture by virtually replicating a "column" of filters at *every position and scale* of the pyramid.

We used the same parameters as in the previous section, but on larger images of 100×100 pixels and with an increasing number of filters (this number thus corresponds to the over-completeness of the dictionary). As in the previous section, we see the emergence of edge selective filters. However, since these are replicated at all positions, similar filters but with different positions would compete and only the strongest —generally a centered filter— is selected (see Fig. 24). This allow for a greater competition of the selection of features in the images, and the center of winning filters will be implicitly more probable to be chosen first on irregular parts of the image (which corresponds to higher activities)[43]. This was enhanced by the spatial kurtosis of natural images since they typically show large "uninteresting" regions and more localized regions where the algorithm will by construction be more likely to fire spikes (see Fig. 20). As we increase the over-completeness, the number of orientations increases as well as the length of the edges. It is of special interest to see that under this efficiency criteria, the different qualitative features of edges, orientation, relative scale, aspect ratio, phase will appear in different orders, the information on orientation being for instance prioritary on the phase information. We also see filters with a similar shape but scaled with a factor inferior to 2, as was used in Sec. 3.1.3 and [Perrinet et al., 2004]. In fact, a lower over-completeness forces the filters to be more general and therefore to be selective to a wider range of orientations. As is observed in progressively higher level of the hierarchy in the visual pathways, there is a progressive sharpening of the selectivity of neurons (typically, edge filters get "longer"). Finally, we obtain a set of mother wavelets, replicated at the different positions and scales of the pyramid and which allow for a translation and scale invariant representation of images, hence the name *Sparselet Analysis*.

We studied the coding efficiency of this architecture for different over-completeness values and compared it to the linear laplacian and steerable pyramids. At the end of the learning scheme, we coded a set of 1000 images in every over-completeness condition. As stated above, the MSE provided an efficient measure of the information carried by the code and we measured it during the algorithm for different number of coefficients. As in Fig. 25, we rated the MSE in function of the number of selected coefficients (see Fig. 24-Left). However, as the over-completeness increases, so would be the number of bits necessary to code every address. We therefore plotted the same results but as a function of the bytes necessary to code the address (see Fig. 24-Right)[44]. The results show that the Sparselet analysis is better adapted to code natural images. In fact, the information transfer gets better as the over-completeness increases. However, the information transfer rate is already optimal for 16 filters per pixel and having more filters per pixels may give an over-fitted representation of visual features. A limit of this approach is that we don't know *a priori* the number of neurons necessary to represent a special feature, that is the

[43] This is desirable since in natural images, large portions, such as the sky or smooth surfaces, have relatively low edge content

[44] It is a similar in the method to Fig. 21.

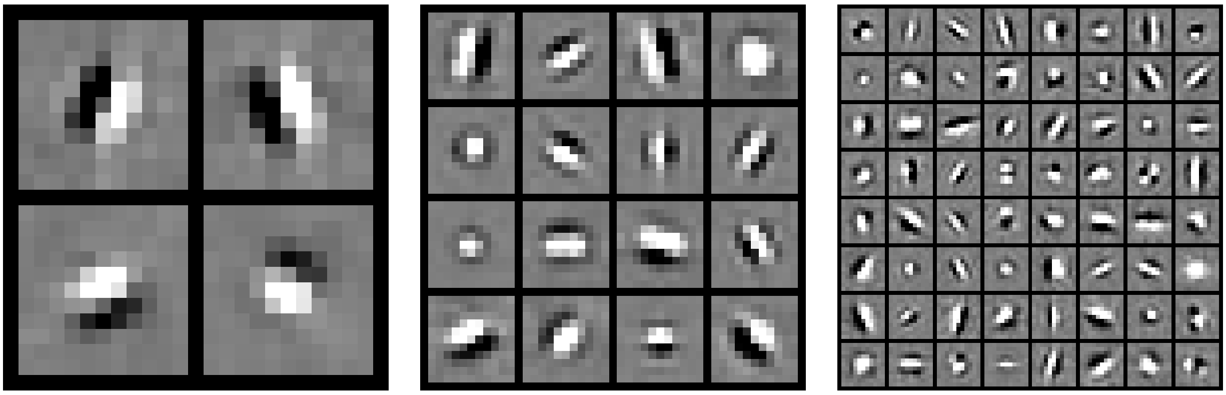

Figure 24: **Filters obtained using the Sparselet Analysis.** We show the resulting mother wavelets from the Sparselet Analysis scheme at different levels of over-completeness, respectively **(Left)** 4, **(Middle)** 16 and **(Right)** 64 filters per pixel. As in Fig. 17, we see the emergence of filters selective to different orientation but also to different sub-scales. These are centered since similar filters in the pyramid (translation and 2-fold scaling) compete in the algorithm. The results exhibit a sharpening of the selectivity of the filters as the over-completeness increases. Note that this sets of filters correspond to the feature vectors defined in Sec. 1.3.3.

complexity of the edge content that is coded in V1. However, we may see that for Fig. 24-Right, filters relative to higher order information (such as corners) appear which are no longer characteristic to V1. This over-completeness (approx. 16) also roughly corresponds to estimation made on cells in V1.

3.3.5 Topological Sparselet Analysis

The filters that were learned by this method where randomly permutable[45] and we may use the information of the position of the address of each learned filter in the feature vector to enhance the representation's efficiency. In fact, a natural extension of this model was to introduce lateral connectivity between these neurons in the Sparselet Analysis learning scheme. This introduces topological relations between neurons as implementing the association of neighboring spatial responses to neighboring neurons in the hyper-column. By taking advantage of the probabilistic framework, this facilitation may take the form of the a priori knowledge of the selection of a neighboring (similarly as in 3.3.3). We parametrized it by an exponentially decreasing gain on the chance of selection of the winning neurons (that is by modulating the threshold) whose amplitude and fall constant experimentally did not change the qualitative results but only the convergence rate of the solution. This strategy leaded in general to a better convergence since learned features in filters —and which therefore appeared more frequently in comparison to random filters— *cooperated* to neighboring filters. This scheme led to the emergence of a "pin-wheel" by associating the learning for neighboring neurons (see Fig. 26) and could be a the origin of the emergence of pinwheel hyper-columns as a basic module

[45] In fact, their particular order on the grid as presented in Fig. 17 and 24 is totally random.

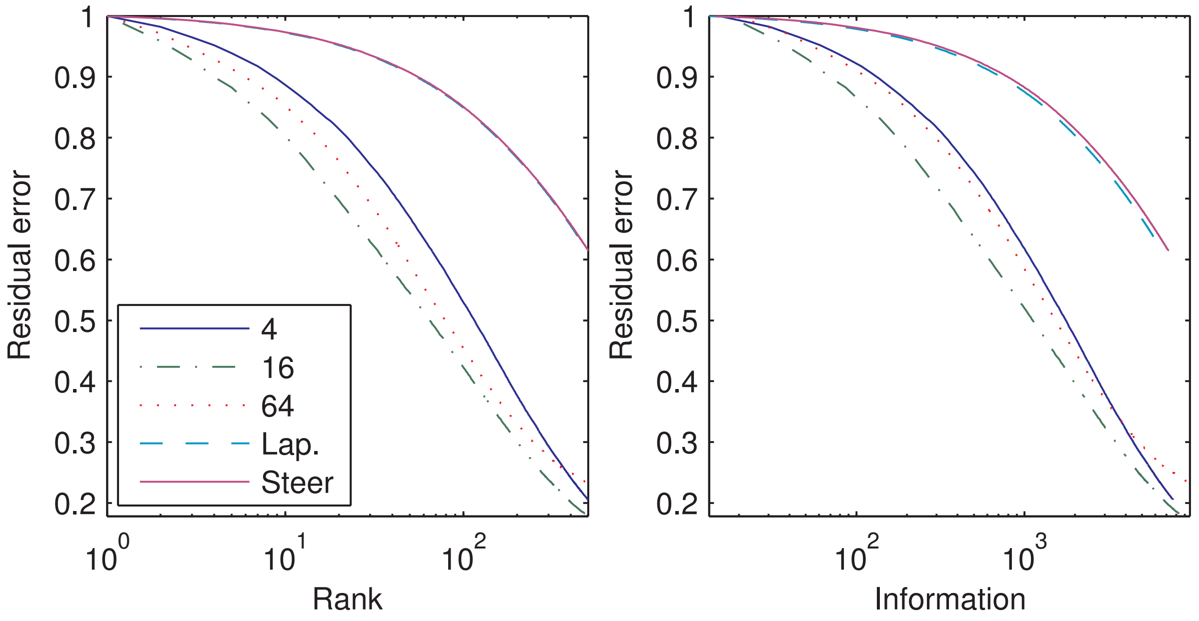

Figure 25: **Efficiency of the Sparselet Analysis**. We compared here the sparselet analysis scheme at different levels of over-completeness with classical methods. We plot the residual Mean Squared Error (MSE) over a set of 1000 natural images for different numbers of spikes. *(Left)* The MSE is plotted in function of this number : the sparse spike coding algorithm outperforms the laplacian (Lap.) and steerable pyramid (Steer.) linear representations and gets better as the over-completeness increases. *(Right)* When plotting the MSE as a function of an approximate of the information needed to code the spike list, the increase of efficiency is less marked: these strategies have similar efficiency.

in V1 for detecting the orientation of edges[46].

3.4 Causal Sparse Spike Coding: an event-based computational approach

A particular challenge in this class of model is to extend the input to a continuous flow of information. In fact, we so far used flashed images to study the transient processing of the information which is of particular importance in the CNS, but a realistic model should be able to handle natural scenes, that is varying images stimuli. Moreover, this is necessary when considering a higher level area after the retina, for which the visual input is dynamically transformed into a spatio-temporal signal thanks to the dynamics of visual pathways.

[46] The particular projection of this feature map on the 2D surface of the cortex is not addressed here but we refer to [Petitot, 2003].

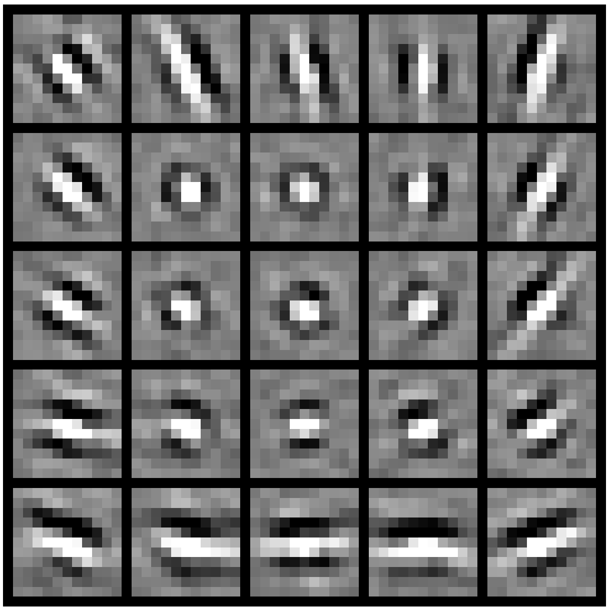

Figure 26: **Emergence of a topological set of orientation selective filters.** As presented above, we simulated the emergence of filters in the primary visual area but introduced lateral connections between filters so that neighboring spiking filters (and thus learning filters) was enhanced. This lead to the emergence of topological relations between neighboring filters similar to the neuro-physiologically observed *pinwheels*: all different orientations are present while preserving a smooth transition between neighboring filters. Note that the center gets less selective (since the constraint is higher than in the periphery), an observation which is contested in physiology.

3.4.1 Causal Sparse Spike Coding

If extending the Matching Pursuit to the time domain is straight-forward, accounting for the direction of time flow (that is to causality) impose to build a modified algorithm. In fact, Matching Pursuit was first used in the time domain [Mallat and Zhang, 1993] and is used for instance in processing time-varying signals or video flows [Blinowska and Durka, 1994; Neff and Zakhor, 1997]. However, these methods don't provide an adequate model of what we know from the nature of time: we should restrain the available information to the past (for the future is not known) and constraint the possible decision times (events) to the present. In all generality, we may write a dictionary of filters as sliding windows of spatiotemporal signals $\mathbb{A}_j$ which we want to detect in the input signal. As in Sec. 2.1.4, we may use the same hypotheses (conditional independence of noise in time) and state that for any given spatiotemporal signal $\mathbb{L}$ at the present time t the information is only known from the initial time $t = 0$ to the present. Therefore, thanks to Eq. 11, we may compute at t for each element j the log-likelihood of the match of $\mathbb{L}$ with $\mathbb{A}_j$ as a new spatiotemporal signal :

$$C_j(t) = 1/||\mathbb{A}_j||. \int_{u \leq t} < \mathbb{L}(u), \mathbb{A}_j(u-t) > du \qquad (71)$$

where $||\mathbb{A}_j||$ is the norm of filters, that is $||\mathbb{A}_j|| = (\int_{u \leq 0} < \mathbb{A}(u), \mathbb{A}_j(u) > du)^{1/2}$. Most often, this is computed in signal processing from the note that $C_j(0) = 0$ for all j and using a recursive formulation based on the shape of $\mathbb{A}_j$.
Using an event-based approach and since $C_j(t)$ represents a log-likelihood, we may state that if it crosses a threshold θ, then we detected a feature and —similarly as in matching pursuit— we will account for that decision on the activity. It means that at this moment[47] t^*, for a certain j^*, $C_{j^*}(t^*) = \theta$ and that we may subtract $\theta.\mathbb{A}_{j^k}(t - t^*)$ from $\mathbb{L}$. This is fed back on C_j as

$$C_j(t) \leftarrow C_j(t) - \theta.R_{\{j^*,j\}}(t - t^*) \qquad (72)$$

where $R_{\{j^*,j\}}(t) = 1/||\mathbb{A}_j||. \int_{u \leq t} < \mathbb{A}_{j^n}(t - t^*), \mathbb{A}_j(u - t) > du$. It should be noted first that as in matching pursuit $C_{j^*}(t^*) = 0$ but also that this feed back is not limited to the moment of the decision but also to the (near) future, since in the general case where $R_{j^*,j}(t)$ is not null for $t \geq 0$. This will be also true if we define predictive fields as in Sec. 3.3.3. This formalization should be put in parallel with the range of methods using Partial Derivative Equations (PDEs) in signal processing. By combining a linear diffusion implemented by the linear kernel corresponding to the RF of the column with the spiking activity, it implements a generic approach for the anisotropic diffusion of the information. In particular, fruitful collaborations should be foreseen by the interaction of this approach (based on a Bayesian and neural formalism) and the mathematical knowledge on PDEs. These links already exist in the literature [Degond and Mas-Gallic, 1989; Aubert et al., 2000; Grimbert and Faugeras, 2005; Viéville and Kornprobst, 2006].

3.4.2 Application: generalized LIF

This general formalism may be first applied to dynamically detect static features. In fact, this formalization generalizes the algorithm presented in Sec. 3.1.4 if we restrict

[47] This subtends a certain continuity of C_j.

the input to a constant image $\mathbb{L}$ after $t \geq 0$ and define filters as separable sliding windows with a constant image profile $\mathbb{A}_j$ and a first-order linear temporal filtering. It may be then easily proved that filters with an exponential decay in time will implement the linear integration of LIF neurons, the time constant τ corresponding to a trade-off between present activity and the short-term trace of activity :

$$C_j(t) = \frac{< \mathbb{L}, \mathbb{A}_j >}{||\mathbb{A}_j||} . \int_{-t \leq 0} e^{-t/\tau} du = \frac{< \mathbb{L}, \mathbb{A}_j >}{||\mathbb{A}_j||} .(1 - e^{-t/\tau}) \tag{73}$$

with $C_j(0) =$ for all j. The choice of filters and times is then exactly similar to Sec. 3.1.4 and leads to the same observations thanks to this "backward compatibility".
In practice, we observed that the threshold was a parameter to control the trading off between speed vs. accuracy. In fact, a lower threshold would give more rapid results but at the same, the information needed to reach that threshold may not be enough to discriminate between different features. A higher threshold will on the other side mean a higher accuracy by a longer latency. These observations were applied by changing the overall threshold but may also change the sensitivity of every single neuron. This is a desirable feature when adjusting the relative importance of neurons in a population and in particular for learning.

3.4.3 Application: Multi-layer architecture

The previous example shows that our formalism may be generalized to a multi-layer dynamical architecture. In fact, we only considered until now the case where all information was flashed at an initial time and how cortical layers would react of this input. In particular it was not possible to handle a multi-layered network since the output of a children layer to a flashed image would be transformed in a spatiotemporal pattern of spikes (which themselves would correspond in a theoretical "read-out" reconstruction to a continuous flow). Thanks to the extension of Sparse Spike Coding to the causal model, we may give a simple expression for its extension to multi-layer architectures, as was sketched in [Perrinet et al., 2002]. In fact, similarly as in Eq. 33, the image can be *virtually* reconstructed in a layer (here denoted by index (2)) from the spike list generated at the previous layer (1):

$$\mathbb{L}^{(2)}(t) = \sum_k a_k^{(1)} . \mathbb{A}_{j_k^{(1)}}^{(1)}(t - t_k^{(1)}) \tag{74}$$

and therefore the activity at this layer may be directly computed as

$$C_j^{(2)}(t) = \sum_k a_k . \int_{u \leq t} < \mathbb{A}_{j_k^{(1)}}^{(1)}(t - t_k)(u), \mathbb{A}_j^{(2)}(u - t) > du \tag{75}$$

the kernel $< \mathbb{A}_j^{(1)}(v)(u), \mathbb{A}_k^{(2)}(u) >$ corresponding to the construction of a higher level receptive field with lower level RFs (to simplify notations, all filters are normalized). This would therefore reduce, according to the model of [Hubel and Wiesel, 1974] to recursively updating $C_j^{(2)}$ with a linear factor of known functions (as one edge was modeled as an alignment of LGN filters). This is a desirable feature since —as in the hierarchy of visual pathways— it reduces the computational complexity by minimizing the crosstalk to the minimum required. However it increases the structural complexity of the network.

Conclusion: distributed, event-based and adaptive nature of neural computation

We proposed in the previous section how we could build an adaptive, multi-layered, fully parallel and dynamic model of cortical areas using discrete events. We saw that to better understand the dynamics of neural computations occurring in low-level vision, we may use a functionalist and integrative approach. This conducted the elaboration of a model using:

1. a distributed probabilistic representations using an explicit model of the function of the area and of the representation that is used. In particular, we saw that cortical areas may explicitly represent maps of features and that the neural activity may correspond to the detection of these features.
2. computations are driven by binary events, spikes which are generated so as to build efficient representations on the cortical area. This efficiency is regulated by homeostatic rules which tune the competition between neurons so that it maximizes its fairness, finally generating a sparse code.
3. learning algorithms may be set so as to learn in an unsupervised manner the features that may at best represent the image while capturing regularities.

Originality of this approach Though limited to the transient response of the visual system, this approach introduced several originalities:

1. The computation of a match was linked to statistical inference, enumerating thus all the hypothesis underlying the chosen formulation (see Sec. 2.1.4). This leaded to a better understanding of spatio-temporal integration in the visual perception of motion (see Sec. 2.3).
2. We have linked this linear representation with an efficient algorithm, Sparse Spike Coding (see Sec. 3.1) based on a neural implementation with lateral interactions (see Sec. 3.1.4). Thanks to the regularity of the coefficients of this non-linear representation on natural images (see Fig. 20), this could lead to efficient applications in image processing (Sparse Edge Coding, see Sec. 3.3).
3. We have shown a method of learning in this Sparse Spike Coding scheme which gave an efficient neural implementation of Independent Components Analysis (see Sec. 3.2). This was extended to wavelet representation (Sparselet Analysis, see Sec. 3.3.4).
4. We extended the formalization of Sparse Spike Coding to the time domain, allowing the extension of these formalisms to more complex stimuli but also to feed-backs (see Sec. 3.4)

Cortical columns as computational bricks These different features are inherited by a fundamental hypothesis on neural computations, that is the uniformity of computational principles in cortical micro-circuits. In fact, the previous principles are the direct consequence of the construction of the cortex as a tiling of similar micro-circuits. From the limited length of the genomic information, this strategy provides a robust solution which may adapt to the development on the CNS. This hypothesis thus provides a conceptual framework which may enable us to further understand the principles behind our neural computations. It should however not be taken strictly, and —as in all biological systems— exceptions abound. In particular, the present definition of the cortical column as an independent system seems more due to the history of neuroscientific discoveries [Horton and Adams, 2005].

Future challenges for modeling neural computations However, it provides a fruitful tool to confront models of neural computations and to validate them with biological experiments. It also allows to segregate different aspects of neural computations between different scales, the study at the scale of cortical areas gaining from knowledge on detailed models of the neuron but also at a higher scale (integration at the level of the CNS, social nets).
On an epistemological level, we see that this method opens a number of future challenges to understand the nature of neural computations. A main goal of integrative neuroscience is the dialectal methodology between abstract models and observations and sees at present a great development thanks to the number of inter-disciplinary cross-talk between corresponding fields (computational neuroscience with neuro-physiologists but also cellular neuroscience with integrative neuroscience).
This challenges the metaphor of the brain that we defined in the beginning of this paper. Neural computations are definitely not of the same nature as the electronic changes occurring in present day sequential computers but relate more to scale-free interactions of interconnected agents each adapting to the rest of the system in a dialectical recurrence, in analogy to the social nets building now the culture of the world.

Reproducible research

Scripts reproducing all figures may be obtained from the author upon request. In particular, scripts reproducing the learning experiments (see Sec. 3.2) are available on the author's web-site at `http://www.incm.cnrs-mrs.fr/LaurentPerrinet`.

Acknowledgments

The author thanks the team at the Redwood Neuroscience Institute for stimulating discussions and particularly Jeff Hawkins, Bruno Olshausen, Fritz Sommer, Tony Bell, Dileep George, Kilian Koepsell and Matthias Bethge. Special thanks to Artemis Kosta for essential comments on this work. This work was supported by a grant form the French Research Council (Action Concertée Incitative "Temps & Cerveau") and by EC IP project FP6-015879, "FACETS".

References

Larry F. Abbott and Sacha B. Nelson. Synaptic plasticity: taming the beast. *Nature Neuroscience*, 3:1178–83, 2000. URL `http://www.nature.com/cgi-taf/DynaPage.taf?file=/neuro/journal/v3/n11s/full/nn1100_1178.html`.

Edgar Adrian. *The basis of sensation: the action of sense organs*. London: ChristoPhers., 1928.

Hirotugu Akaike. A new look at the statistical model identification. *IEEE Transactions on Automatic Control*, 19:716–23, 1974.

Duane G. Albrecht and David B. Hamilton. Striate Cortex of Monkev and Cat: Contrast Response Function. *Journal of Neurophysiology*, 48(217–238), 1982.

Thomas D. Albright. Direction and orientation selectivity of neurons in visual area MT of the macaque. *Journal of Neurophysiology*, 52:1106–30, 1984.

Luis Alvarez, Yann Gousseau, and Jean-Michel Morel. The Size of Objects in Natural Images. Technical Report 9921, Centre de Mathématique et de Leurs Applications, 1999.

Joseph J. Atick. Could information theory provide an ecological theory of sensory processing? *Network: Computation in Neural Systems*, 3(2):213–52, 1992. URL `http://ib.cnea.gov.ar/~redneu/atick92.pdf`.

G. Aubert, Rachid Deriche, and P. Kornprobst. Computing Optical Flow via Variational Techniques. *SIAM Journal on Applied Mathematics*, 60(1):156–82, 2000.

Wyeth Bair and Christof Koch. Temporal Precision of Spike Trains in Extrastriate Cortex of the Behaving Macaque Monkey. *Neural Computation*, 8(6):1185–1202, 1996.

Wyeth Bair, James R. Cavanaugh, and Anthony Movshon. Time course and time-distance relationships for surround suppression in macaque V1 neurons. *Journal of Neuroscience*, 23(20):7690 —701, August 2003.

A.-L. Barabasi and E. Bonabeau. Scale-free networks. *Scientific American*, pages 50–9, 2003.

Horace B. Barlow. Single units and sensation: A neuron doctrine for perceptual psychology? *Perception*, 1(4):371–94, 1972.

Horace B. Barlow. Redundancy reduction revisited. *Network: Computation in Neural Systems*, 12:241—25, 2001.

Amit Basole, Leonard E. White, and David Fitzpatrick. Mapping multiple features in the population response of visual cortex. *Nature*, 423:986–90, jun 2003.

Pierre Bayerl and Heiko Neumann. Disambiguating Visual Motion Through Contextual Feedback Modulation. *Neural Computation*, 16:2041–66, 2004.

Thomas Bayes. An Essay Toward Solving a Problem in the Doctrine of Chances. *Philosophical Transactions of the Royal Society of London*, 53:370–418, 1764.

Anthony J. Bell and Terrence J. Sejnowski. An Information-Maximization Approach to Blind Separation and Blind Deconvolution. *Neural Computation*, 7:1129–59, 1995.

Anthony J. Bell and Terrence J. Sejnowski. The 'independent components' of natural scenes are edge filters. *Vision Research*, 37(23):3327–38, 1997.

CC. Bell, VZ. Han, Y. Sugawara, and K. Grant. Synaptic plasticity in a cerebellum-like structure depends on temporal order. *Nature*, 387:278–81, 1997.

G-Q Bi and M-M Poo. Synaptic Modifications in Cultured Hippocampal Neurons: Dependence on Spike Timing, Synaptic Strength, and Postsynaptic Cell Type. *Journal of Neuroscience*, 18:10464–72, 1998. URL http://www.jneurosci.org/cgi/content/full/18/24/10464?axtoshow=&HITS=10&hits=10&RESULTFORMAT=&author1=bi&searchid=QID_NOT_SET&stored_search=&FIRSTINDEX=.

K. J. Blinowska and P. J. Durka. The Application of Wavelet Transform and Matching Pursuit to the Time-Varying EEG signals. In C. H. Dagli and B. R. Fernandez, editors, *Intelligent Engineering Systems through Artificial Neural Networks*, volume 4, pages 535–540. ASME Press, New York, 1994. ISBN 0-7918-045-8.

Lyle J. Borg-Graham. Interpretations of Data and Mechanisms for Hippocampal Pyramidal Cell Models. In *Cerebral Cortex*, volume 13. P. S. Ulinski, E. G. Jones and A. Peters, New York: Plenum Press, 1999.

Korbinian Brodmann. *Vergleichende Lokalisationslehre der Grosshirnrinde in ihren Principien, dargestellt auf grund des Zellenbaues.* Johann Ambrosius Barth Verlag, Leipzig, 1909. URL http://en.wikipedia.org/wiki/Brodmann_area.

Jean Bullier. Integrated model of visual processing. *Brain Research Reviews*, 36: 96–107, 2001. URL http://dx.doi.org/10.1016/S0165-0173(01)00085-6.

Peter J. Burt and Edward H. Adelson. The Laplacian Pyramid as a compact image code. *IEEE Transactions on Communications*, COM-31,4:532–40, 1983.

Santiago Ramòn Y Cajal. *Histologie du système nerveux de l'Homme et des vertébrés.* Maloine, Paris, 1911.

J. Canny. A computational approach to edge detection. *IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI)*, 8:679–98, 1986.

Enrico Capobianco. Independent multiresolution component analysis and matching pursuit. *Comput. Stat. Data Anal.*, 42(3):385–402, 2003. ISSN 0167-9473. doi: http://dx.doi.org/10.1016/S0167-9473(02)00217-7.

Matteo Carandini, J. Heeger, and Anthony Movshon. Linearity and normalization in simple cells of the macaque primary visual cortex. *Journal of Neuroscience*, 17 (21):8621—44, November 1997.

S. Castan, J. Zhao, and J. Shen. Optimal Filter for Edge Detection Methods and Results. In *In Proceedings of the First European Conference on Computer Vision (Eccv)*, pages 13–7, 1990.

James R. Cavanaugh, Wyeth Bair, and Anthony Movshon. Nature and interaction of signals from the receptive field center and surround in macaque V1 neurons. *Journal of Neurophysiology*, 88(5):2530–46, November 2002. doi: 10.1152/jn.00692.2001. URL http://dx.doi.org/10.1152/jn.00692.2001.

S. Celebrini, Simon J. Thorpe, Y. Trotter, and M. Imbert. Dynamics of orientation coding in area V1 of the awake primate. *Visual Neuroscience*, 5(10):811–25, 1993.

Bruno Cessac, Emmanuel Daucé, Laurent Perrinet, and Manuel Samuelides. *Topics in Dynamical Neural Networks: From Large Scale Neural Networks to Motor Control and Vision*, volume 142 of *The European Physical Journal (Special Topics)*. Springer Verlag, Berlin / Heidelberg, mar 2007. doi: 10.1140/epjst/e2007-00061-7. URL `http://www.springerlink.com/content/q00921n9886h/?p=03c19c7c204d4fa78b850f88b97da2f7&pi=0`.

Shaobing Chen. *Basis pursuit*. PhD thesis, Stanford, 1995.

Y Dan, Joseph J. Atick, and RC Reid. Efficient coding of natural scenes in the lateral geniculate nucleus: experimental test of a computational theory. *Journal of Neuroscience*, 16(10):3351–62, May 1996.

Geoffrey Davis. *Adaptive Nonlinear Approximations.* PhD thesis, New York University, 1994.

Peter Dayan and Larry F. Abbott. *Theoretical neuroscience*. The MIT Press, Cambridge, MA, 2001.

D. Debanne, D. Shulz, and Yves Frégnac. Temporal constraints in associative synaptic plasticity in hippocampus and neocortex. *Can. J. Physiol. and Pharmacol.*, 73:1295–311, 1995.

P. Degond and S. Mas-Gallic. The Weighted Particle Method for Convection-Diffusion Equations. *Mathematics of Computation*, 53(188):485–525, 1989.

Arnaud Delorme and Simon J. Thorpe. Early cortical orientation selectivity: How fast shunting inhibition decodes the order of spike latencies. *Journal of Computational Neuroscience*, 15:357–65, 2003.

Sophie Denève, Peter E Latham, and Alexandre Pouget. Reading population codes: a neural implementation of ideal observers. *Nature Neuroscience*, 2(8):740–5, 1999.

Rachid Deriche. Using Canny's criteria to derive a recusively implemented optimal edge detector. *Int. Journal of Computional Vision*, pages 167–87, 1987.

Michael R. DeWeese, Michael Wehr, and Anthony M. Zador. Binary coding in auditory cortex. *Journal of Neuroscience*, 23(21), August 2003.

W. D. Dong and Joseph J. Atick. Statistics of natural time-varying images. *Network: Computation in Neural Systems*, 6(3):345–58, 1995.

P. J. Durka, D. Ircha, and K. J. Blinowska. Stochastic time-frequency dictionaries for matching pursuit. *IEEE Transactions on Signal Processing*, 49(3):507–510, March 2001.

C. Enroth-Cugell and J. G. Robson. The Contrast Sensitivity of Retinal Ganglion Cells of the Cat. *Journal of Physiology*, (187):517–23, 1966.

David J. Field. Relations between the statistics of natural images and the response properties of cortical cells. *Optical Society of America A*, 4(12):2379–94, 1987.

David J. Field. What is the goal of sensory coding? *Neural Computation*, 6(4):559–601, 1994.

David J. Field, Anthony Hayes, and Robert F. Hess. Contour Integration by the Human Visual System: Evidence for a Local 'Association Field'. *Vision Research*, 33(2):173–93, 1993.

Sylvain Fischer, Rafael Redondo, Laurent Perrinet, and Gabriel Cristóbal. Efficient representation of natural images using local cooperation. In Ricardo A. Carmona and Gustavo Linan-Cembrano, editors, *Perception*, volume 34 of *ECVP*, page 241, 2005a.

Sylvain Fischer, Rafael Redondo, Laurent Perrinet, and Gabriel Cristóbal. Sparse Gabor wavelets by local operations. In Gustavo Linan-Cembrano Ricardo A. Carmona, editor, *Proceedings SPIE*, volume 5839 of *Bioengineered and Bioinspired Systems II*, pages 75–86, Jun 2005b. doi: doi:10.1117/12.608403.

Sylvain Fischer, Gabriel Cristóbal, and Rafael Redondo. Sparse Overcomplete Gabor Wavelet Representation Based on Local Competitions. *IEEE Transactions in Image Processing*, 15(2):265, February 2006.

Sylvain Fischer, Rafael Redondo, Laurent Perrinet, and Gabriel Cristóbal. Sparse approximation of images inspired from the functional architecture of the primary visual areas. *EURASIP Journal on Advances in Signal Processing*, pages Article ID 90727, 16 pages, 2007a. doi: doi:10.1155/2007/90727. URL `http://www.hindawi.com/GetArticle.aspx?doi=10.1155/2007/90727&e=cta`.

Sylvain Fischer, Filip Sroubek, Laurent Perrinet, Rafael Redondo, and Gabriel Cristóbal. Self-invertible 2D log-Gabor wavelets. *Int. Journal of Computional Vision*, 2007b.

W. J. Freeman and J. M. Barrie. Chaotic Oscillations and the Genesis of Meaning in Cerebral Cortex. In G. Buzsáki, editor, *Temporal Coding in the Brain*, pages 13–37, Berlin Heidelberg, 1994. Springer-Verlag.

Jerome H. Friedman and Werner Stuetzle. Projection pursuit regression. *Journal of the American Statistical Association)*, 1980.

P. Fries, J. H. Schroder, P. R. Roelfsema, Wolf Singer, and A. K. Engel. Oscillatory neuronal synchronization in primary visual cortex as a correlate of stimulus selection. *Journal of Neuroscience*, 22(9):3739–54, 2002.

Pierre Frossard and Pierre Vandergheynst. A Posteriori Quantized Matching Pursuit. *IEEE Data Compression Conference*, 2001.

Wilson S. Geisler and Duane G. Albrecht. Visual cortex neurons in monkeys and cats: Detection, discrimination, and identification. *Visual Neuroscience*, 1(14): 897–919, 1997.

Wilson S. Geisler, J. S. Perry, B. J. Super, and D. P. Gallogly. Edge co-occurance in natural images predicts contour grouping performance. *Vision Research*, 41(6): 711–24, 2001.

Sébastien Georges, Peggy Seriès, Yves Frégnac, and Jean Lorenceau. Orientation dependent modulation of apparent speed: psychophysical evidence. *Vision Research*, 42(25):2757–72, Nov 2002.

A. P. Georgopoulos, A. B. Schwartz, and R. E. Kettner. Neuronal population coding of movement direction. *Science*, 233(4771):1416–9, Sep 1986.

Charles D. Gilbert and T. N. Wiesel. Morphology and intracortical projections of functionally characterised neurones in the cat visual cortex. *Nature*, 280(5718): 120–5, Jul 1979.

R. Gribonval and P. Vandergheynst. On the exponential convergence of matching pursuits in quasi-incoherent dictionaries. *IEEE Transactions in Information Theory*, pages 255– –61, jan 2006. doi: 10.1109/TIT.2005.860474.

François Grimbert and Olivier Faugeras. Analysis of Jansen's model of a single cortical column. Technical Report 5597, Projet Odyssée, 2005.

A. Grinvald, D. Shoham, A. Shmuel, D. Glaser, I. Vanzetta, E. Shtoyerman, H. Slovin, and A. Sterkin. In-vivo optical imaging of cortical architecture and dynamics. In *Modern Techniques in Neuroscience Research*. U. Windhorst and H. Johansson (Editors) Springer Verlag, 2001.

Amiram Grinvald, Edmund E. Lieke, Ron D. Frostig, and Rina Hildesheim. Cortical Point-Spread Function and Long-Range Lateral Interactions Revealed by Real-Time Optical Imaging of Macaque Monkey Primary Visual Cortex. *Journal of Neuroscience*, 14(5):2545–68, May 1994.

Stephen Grossberg. How does the cerebral cortex work? development, learning, attention, and 3-d vision by laminar circuits of visual cortex. *Behavioral and Cognitive Neuroscience Reviews*, 2(1):47–76, March 2003.

Stephen Grossberg and Arash Yazdanbakhsh. Laminar cortical dynamics of 3D surface perception: stratification, transparency, and neon color spreading. *Vision Research*, 45(13):1725–43, Jun 2005. URL `http://dx.doi.org/10.1016/j.visres.2005.01.006`.

Rudy Guyonneau, Rufin van Rullen, and Simon J. Thorpe. Neurons tune to the earliest spikes through stdp. *Neural Computation*, 17(4):859–79, 2005. URL `http://neco.mitpress.org/cgi/content/abstract/17/4/859`.

Richard H. R. Hahnloser, Alexay A. Kozhevnikov, and Michale S. Fee. An ultra-sparse code underlies the generation of neural sequences in a songbird. *Nature*, 419:65–70, 2002.

H. K. Hartline. The receptive fields of optic nerve fibers. *American Journal of Physiology*, 130:690–9, 1940.

Donald O. Hebb. *The organization of behavior: A neuropsychological theory*. Wiley, New York, 1949.

Geoffrey H. Henry, B. Dreher, and P. O. Bishop. Orientation specificity of cells in cat striate cortex. *Journal of Neurophysiology*, 37:1394–409, 1974.

Jonathan C. Horton and Daniel L. Adams. The cortical column: a structure without a function. *Philosophical Transactions of the Royal Society of London*, 360(1456): 837–62, 2005. doi: doi:10.1098/rstb.2005.1623.

Toshihiko Hosoya, Stephen A Baccus, and Markus Meister. Dynamic predictive coding by the retina. *Nature*, 436(7047):71–7, Jul 2005. doi: 10.1038/nature03689. URL http://dx.doi.org/10.1038/nature03689.

David Hubel and Torsten Wiesel. Receptive Fields of Single Neurones in the Cat's Striate Cortex. *Journal of Physiology*, 148:574–91, 1959.

David Hubel and Torsten Wiesel. Receptive fields, binocular interaction and functional architecture in the cat's visual cortex. *Journal of Physiology*, 160:106–54, 1962.

David Hubel and Torsten Wiesel. Receptive Fields and Functional Architecture of Monkey's Striate Cortex. *Journal of Physiology*, 195:215–44, 1968.

David Hubel and Torsten Wiesel. Uniformity of Monkey's Striate Cortex: a parallel relationship between field size, scatter and magnification factor. *Journal of Physiology*, 158:295–306, 1974.

Dirk Jancke, Frédéric Chavane, Shmuel Naaman, and Arvind Grinvald. Imaging cortical correlates of illusion in early visual cortex. *Nature*, 428:423–6, 2004.

Viktor K. Jirsa. Connectivity and dynamics of neural information processing. *Neuroinformatics*, 2004.

Judson P. Jones and Larry A Palmer. An evaluation of the two-dimensional gabor filter model of simple receptive fields in cat striate cortex. *Journal of Neurophysiology*, 58(6):1233–58, 1987.

E.R. Kandel, J.H. Schwartz, and T.M. Jessel. *Principles of Neural Science*. McGraw Hill, New York, 4th edition, 2000.

D. Kersten, Pascal Mamassian, and A. Yuille. Object perception as Bayesian inference. *Annual Review of Psychology*, 55, 2003.

Christof Koch, editor. *Biophysics of computation: information processing in single neurons*. Oxford University Press, New York, 1998.

Christof Koch and Idan Segev. The Role of Single Neurons in Information Processing. *Nature Neuroscience*, 3:1171–7, 2000.

T. Kohonen. Self-organized formation of topologically correct feature maps. *Biological Cybernetics*, 43:59–69, 1982.

Ilona Kovacs, Petra Kozma, Akos Feher, and Gyorgy Benedek. Late maturation of visual spatial integration in humans. *Proceedings of the National Academy of Sciences USA*, 96(21):12204–12209, 1999. URL http://www.pnas.org/cgi/content/abstract/96/21/12204.

J.C. Lagarias, J. A. Reeds, M. H. Wright, and P. E. Wright. Convergence Properties of the Nelder-Mead Simplex Method in Low Dimensions. *SIAM Journal of Optimization*, 9(1):112–47, 1998.

L. Lapicque. Recherches quantitatives sur l'excitation électrique des nerfs traitée comme une polarisation. *Journal of Physiology (Paris)*, 9:620–35, 1907.

Simon B. Laughlin. A simple coding procedure enhances a neuron's information capacity. *Zeitung für Naturforschung*, 9–10(36):910–2, 1981.

Erwan Le Pennec and Stéphane Mallat. Sparse Geometric Image Representations With Bandelets. *IEEE Transactions in Image Processing*, 14(4):423, april 2005.

Michael S. Lewicki and Terrence J. Sejnowski. Learning overcomplete representations. *Neural Computation*, 12(2):337–65, 2000. URL `citeseer.nj.nec.com/lewicki98learning.html`.

Zheng Liu, James P. Gaska, Lowell D. Jacobson, and Daniel A. Pollen. Interneuronal interaction between members of quadrature phase and anti-phase pairs in the cat's visual cortex. *Vision Research*, 32(7):1193–8, 1992.

Nikos K. Logothetis, Jon Pauls, Mark Augath, Torsten Trinath, and Axel Oeltermann. Neurophysiological investigation of the basis of the fMRI signal. *Nature*, 412: 150–7, 2001.

David J. C. MacKay. *Information theory, inference, and learning algorithms*. Cambridge University Press, 2003. URL `http://www.inference.phy.cam.ac.uk/mackay/itila/`.

Prasanta Chandra Mahalanobis. On the generalized distance in statistics. *Proceedings of the National Academy of Sciences USA*, 12:49–55, 1936.

Zachary F. Mainen and Terrence J. Sejnowski. Influence of Dendritic Structure on Firing Pattern in Model Neocortical Neurons. *Nature*, 382:363–366., 1996.

Stéphane Mallat. *A wavelet tour of signal processing*. Academic Press, 1998.

Stéphane Mallat and Wen Liang Hwang. Singularity Detection And Processing with Wavelets. Technical report, Courant Institute of Mathematical Sciences, New York University, New York, 1991.

Stéphane Mallat and Zhifeng Zhang. Matching Pursuit with time-frequency dictionaries. *IEEE Transactions on Signal Processing*, 41(12):3397–3414, 1993.

Pascal Mamassian. Bayesian modelling of visual perception. In Rao et al. [2002], pages 13–36.

David Marr. Visual Information Processing: The Structure and Creation of Visual Representations. *Philosophical Transactions of the Royal Society of London*, 290: 199–218, 1980.

David Marr. *Vision*. W. H. Freeman and Company, NY, 1982.

S. Martinez-Conde, SL Macknik, and David Hubel. Microsaccadic eye movements and firing of single cells in the striate cortex of macaque monkeys. *Nature Neuroscience*, 3, 2000.

Guillaume S. Masson. From 1D to 2D via 3D: surface motion integration for gaze stabilization in primates. *Journal of Physiology (Paris)*, 98:35–52, 2004.

Guillaume S. Masson and Eric Castet. Parallel motion processing for the initiation of short-latency ocular following in humans. *Journal of Neuroscience*, 22(12):5149–63, 2002. URL http://eutils.ncbi.nlm.nih.gov/entrez/eutils/elink.fcgi?cmd=prlinks&dbfrom=pubmed&retmode=ref&id=12077210.

Guillaume S. Masson, Daniel R. Mestre, F. Martineau, C. Soubrouillard, C. Brefel, O. Rascol, and O. Blin. Lorazepam-induced modifications of saccadic and smooth-pursuit eye movements in humans: attentional and motor factors. *Behavioral Brain Research*, 108(2):169–80, 2000. URL http://eutils.ncbi.nlm.nih.gov/entrez/eutils/elink.fcgi?cmd=prlinks&dbfrom=pubmed&retmode=ref&id=10701660.

F. Metelli. The perception of transparency. *Scientific American*, 230(4):90–8, 1974.

Wolfgang Metzger. *Gesetze des Sehens*. Verlag Waldemar Kramer (Frankfurt), 1st edition, 1936.

Risto Miikkulainen, James A. Bednar, Yoonsuck Choe, and Joseph Sirosh. *Computational Maps in the Visual Cortex*. Springer Verlag, February 2005.

Cyril Monier, Frédéric Chavane, Pierre Baudot, Lyle J. Graham, and Yves Frégnac. Orientation and direction selectivity of synaptic inputs in visual cortical neurons: A diversity of combinations produces spike tuning. *Neuron*, 37(4):663–80, 2003. URL http://www.sciencedirect.com/science/article/B6WSS-4CC2YG4-4J/2/32e4310b7c254407325bdd8430b58d97.

Anna Montagnini, Pascal Mamassian, Laurent Perrinet, Eric Castet, and Guillaume S. Masson. Bayesian modeling of dynamic motion integration. In *1ère conférence francophone NEUROsciences COMPutationnelles (NeuroComp)*, 2006.

Vernon B. Mountcastle. Modality and topographic properties of single neurons of cat's somatic sensory cortex. *Journal of Neurophysiology*, 20(4):408–434, Jul 1957.

Vernon B. Mountcastle. *Perceptual neuroscience: the cerebral cortex*. 1998.

K. I. Naka and W. A. Rushton. S-potentials from luminosity units in the retina of fish (Cyprinidae). *Journal of Physiology*, 185(3):587–99, August 1966.

R. Neff and A. Zakhor. Very Low Bit-Rate Video Coding based on Matching Pursuits. *IEEE Transactions on CSVT*, 7(5):158–71, Oct. 1997.

Erkki Oja. A Simplified Neuron Model as a Principal Component Analyzer. *Journal of Mathematical biology*, 15:267–273, 1982.

Bruno A. Olshausen. What is the other 85% of V1 doing? In Terrence J. Sejnowski and J. Leo van Hemmen, editors, *Problems in Systems Neuroscience*. Oxford University Press, 2004.

Bruno A. Olshausen and David J. Field. Sparse coding with an overcomplete basis set: a strategy employed by V1? *Vision Research*, 37:3311–25, 1997.

Guy A Orban, David Van Essen, and Wim Vanduffel. Comparative mapping of higher visual areas in monkeys and humans. *Trends in Cognitive Science*, 8(7): 315–24, Jul 2004. URL http://dx.doi.org/10.1016/j.tics.2004.05.009.

Kevin J. O'Regan and Alva Noë. A sensorimotor account of vision and visual consciousness. *Behavioral Brain Research*, 24(5), 2001.

Christopher C Pack, Andrew J Gartland, and Richard T Born. Integration of Contour and Terminator Signals in Visual Area MT of Alert Macaque. *Journal of Neuroscience*, 24(13):3268–80, Mar 2004. doi: 10.1523/JNEUROSCI.4387-03.2004. URL `http://dx.doi.org/10.1523/JNEUROSCI.4387-03.2004`.

Y. Pati, R. Rezaiifar, and P. Krishnaprasad. Orthogonal Matching Pursuit: Recursive Function Approximation with Applications to Wavelet Decomposition. In *Proceedings of the 27 th Annual Asilomar Conference on Signals, Systems, and Computers*, 1993.

Arthur E. C. Pece. The problem of sparse image coding. *Journal of Mathematical Imaging and Vision*, 17:89–108, 2002.

Laurent Perrinet. Apprentissage hebbien d'un reseau de neurones asynchrone a codage par rang. Technical report, Rapport de stage du DEA de Sciences Cognitives, CERT, Toulouse, France, 1999. URL `http://www.risc.cnrs.fr/detail_memt.php?ID=280`.

Laurent Perrinet. Finding Independent Components using spikes : a natural result of hebbian learning in a sparse spike coding scheme. *Natural Computing*, 3(2): 159–75, January 2004a. doi: 10.1023/B:NACO.0000027753.27593.a7. URL `http://www.incm.cnrs-mrs.fr/LaurentPerrinet/Publications/Perrinet04nc`.

Laurent Perrinet. Feature detection using spikes : the greedy approach. *Journal of Physiology (Paris)*, 98(4-6):530–9, July-November 2004b. doi: 10.1016/j.jphysparis.2005.09.012. URL `http://hal.archives-ouvertes.fr/hal-00110801/en/`.

Laurent Perrinet. Efficient Source Detection Using Integrate-and-Fire Neurons. In W. Duch et al., editor, *ICANN 2005, LNCS 3696*, volume 3696 of *Lecture Notes in Computer Science*, pages 167–72, Berlin Heidelberg, 2005. Springer. doi: 10.1007/11550822_27. URL `http://www.incm.cnrs-mrs.fr/LaurentPerrinet/Publications/Perrinet05icann`.

Laurent Perrinet. An efficiency razor for model selection and adaptation in the primary visual cortex. In *Fifteenth Annual Computational Neuroscience Meeting*, 2006. URL `http://www.incm.cnrs-mrs.fr/LaurentPerrinet/Publications/Perrinet06cns`.

Laurent Perrinet. Dynamical neural networks: modeling low-level vision at short latencies. In *Topics in Dynamical Neural Networks: From Large Scale Neural Networks to Motor Control and Vision* Cessac et al. [2007], pages 163–225. doi: 10.1140/epjst/e2007-00061-7. URL `http://www.incm.cnrs-mrs.fr/LaurentPerrinet/Publications/Perrinet06`.

Laurent Perrinet, Manuel Samuelides, and Simon J. Thorpe. Sparse spike coding in an asynchronous feed-forward multi-layer neural network using Matching Pursuit. *Neurocomputing*, 57C:125–34, 2002. URL `http://www.incm.cnrs-mrs.fr/LaurentPerrinet/Publications/Perrinet02sparse`. Special issue: New Aspects in Neurocomputing: 10th European Symposium on Artificial Neural Networks 2002 - Edited by T. Villmann.

Laurent Perrinet, Manuel Samuelides, and Simon J. Thorpe. Coding static natural images using spiking event times: do neurons cooperate? *IEEE Transactions on Neural Networks*, 15(5):1164–75, September 2004. ISSN 1045-9227. doi: 10.1109/TNN.2004.833303. URL `http://hal.archives-ouvertes.fr/hal-00110803/en/`. Special issue on 'Temporal Coding for Neural Information Processing'.

Laurent Perrinet, Frédéric Barthélemy, Eric Castet, and Guillaume S. Masson. Dynamics of motion representation in short-latency ocular following: A two-pathways bayesian model. In Ricardo A. Carmona and Gustavo Linan-Cembrano, editors, *Perception*, volume 34 of *ECVP*, page 38, 2005.

Laurent Perrinet, Frédéric V. Barthélemy, and Guillaume S. Masson. Input-output transformation in the visuo-oculomotor loop: modeling the ocular following response to center-surround stimulation in a probabilistic framework. In *1ère conférence francophone NEUROsciences COMPutationnelles - NeuroComp*, 2006.

Jean Petitot. The neurogeometry of pinwheels as a sub-Riemannian contact structure. *Journal of Physiology (Paris)*, 97(2-3):265–309, 2003. URL `http://dx.doi.org/10.1016/j.jphysparis.2003.10.010`.

Alexandre Pouget. Dynamic Remapping. In Michael A. Arbib, editor, *The handbook of brain theory and neural networks.* The MIT Press,, Cambridge, MA, second edition, 2002.

Charles Poynton. Frequently Asked Questions about Gamma. Technical report, 1999.

Dale Purves and R. Beau Lotto. *Why We See What We Do: An Empirical Theory of Vision*. Sinauer Associates, Sunderland, Massachusetts, 2003. doi: ISBN:0-878-93752-8.

Rajesh P. N. Rao, Bruno A. Olshausen, and Michael S. Lewicki, editors. MIT Press, 2002.

Rafael Redondo, Sylvain Fischer, Laurent Perrinet, and Gabriel Cristóbal. Modeling of simple cells through a sparse overcomplete gabor wavelet representation based on local inhibition and facilitation. In Ricardo A. Carmona and Gustavo Linan-Cembrano, editors, *Perception*, volume 34 of *ECVP*, page 238, August 2005.

Dario L. Ringach. Spatial structure and symmetry of simple-cell receptive fields in macaque primary visual cortex. *Journal of Neurophysiology*, 88:455–63, 2002.

Dario L. Ringach, Robert M Shapley, and M.J. Hawken. Orientation selectivity in macaque V1: diversity and laminar dependence. *Journal of Neuroscience*, 22 (13):5639–51, 2002. URL `http://eutils.ncbi.nlm.nih.gov/entrez/eutils/elink.fcgi?cmd=prlinks&dbfrom=pubmed&retmode=ref&id=12097515`.

R. W. Rodieck. Quantitative analysis of cat retinal ganglion cell response to visual stimuli. *Vision Research*, 5:583–601, 1965.

E. Rodriguez, N. George, J.-P. Lachaux, J. Martinerie, B. Renault, and F. Varela. Perception's shadow: long-distance gamma band synchronization of human brain activity. *Nature*, 397:430–3, 1999.

Franck Rosenblatt. Perceptron simulation experiments. *Proceedings of the I. R. E.*, 20: 167–192, 1960.

Daniel L. Ruderman and William Bialek. Statistics of Natural Images: Scaling in the Woods. *Physical Review Letters*, 73(6):551–8, 1994.

P. A. Salin and J. Bullier. Corticocortical connections in the visual system: structure and function. *Physiological Review*, 75(1):107–54, January 1995.

Phil Sallee and Bruno A. Olshausen. Learning Sparse Multiscale Image Representations. In Michael I. Jordan, Michael J. Kearns, and Sara A. Solla, editors, *Advances in neural information processing systems*, volume 15, pages 1327–34. The MIT Press, Cambridge, MA, 2003.

Odelia Schwartz and Eero Simoncelli. Natural Signal Statistics and Sensory Gain Control. *Nature Neuroscience*, 4(8):819–25, 2001.

Basabdatta Sen and Steve Furber. Information recovery from rank-order encoded images. In *Workshop on Biologically Inspired Information Fusion University of Surrey*, Aug 2006.

Peggy Seriès, Sébastien Georges, Jean Lorenceau, and Yves Frégnac. Orientation dependent modulation of apparent speed: a model based on the dynamics of feed-forward and horizontal connectivity in V1 cortex. *Vision Research*, 42(25): 2781–97, Nov 2002.

Claude Elwood Shannon and Warren Weaver. *The mathematical theory of communication*. The University of Illinois Press, Urbana, 1964.

Roger N. Shepard and Jacqueline Metzler. Mental Rotation of three-dimensionnal Objects. *Science*, 171:701–4, 1970.

C. S. Sherrington. Observations on the scratch-reflex in the spinal dog. *Journal of Physiology*, 34(1–2):1—50, March 1906. URL `http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=1465804`.

Mariano Sigman, Guillermo A. Cecchi, Charles D. Gilbert, and Marcelo O. Magnasco. On a common circle: Natural scenes and Gestalt rules. *Proceedings of the National Academy of Sciences USA*, 98(4):1935–40, 2001.

E. P. Simoncelli and W. T. Freeman. The Steerable Pyramid: A Flexible Architecture for Multi-Scale Derivative Computation. In *Second International Conf. on Image Processing*, Washington, DC, October 1995.

Eero P. Simoncelli and Bruno A. Olshausen. Natural Image Statistics and Neural Representation. *Annual Review of Neuroscience*, 24:1193–216, 2001. doi: 10.1146/annurev.neuro.24.1.1193. URL `http://dx.doi.org/10.1146/annurev.neuro.24.1.1193`.

Mandyam V. Srinivasan, Simon B. Laughlin, and A Dubs. Predictive coding: A fresh view of inhibition in the retina. *Proceedings of the Royal Society of London. Series B, Biological Sciences*, 216(1205):427–59, Nov 1982.

Antonio Turiel, Germán Mato, Néstor Parga, and Jean-Pierre Nadal. Self-similarity Properties of Natural Images. In Michael I. Jordan, Michael J. Kearns, and Sara A. Solla, editors, *Advances in neural information processing systems*, volume 10. The MIT Press, Cambridge, MA, 1998.

J. Hans van Hateren. Spatiotemporal contrast sensitivity of early vision. *Vision Research*, 33:257–67, 1993.

J. Hans van Hateren and A. van der Schaaf. Independent component filters of natural images compared with simple cells in primary visual cortex. *Philosophical Transactions of the Royal Society of London B*, 265:359–66, 1998.

Rufin van Rullen and Simon J. Thorpe. Rate coding versus temporal order coding: what the retina ganglion cells tell the visual cortex. *Neural Computation*, 13(6): 1255–83, 2001.

Thierry Viéville and Pierre Kornprobst. Modeling Cortical Maps with Feed-Backs. In *International Joint Conference on Neural Networks*, 2006.

Hermann von Helmholtz. *Treatise on Physiological Optics*, volume 3. New York: Optical Society of America, 1925.

Martin J. Wainwright, Odelia Schwartz, and Eero P. Simoncelli. Natural Image Statistics and Divisive Normalization: Modeling Nonlinearities and Adaptation in Cortical Neurons. In *Statistical Theories of the Brain*. The MIT Press, Rajesh Rao, Bruno Olshausen and M Lewicki, 2001.

Yair Weiss, Eero P Simoncelli, and Edward H Adelson. Motion illusions as optimal percepts. *Nature Neuroscience*, 5(6):598–604, Jun 2002. doi: 10.1038/nn858. URL `http://dx.doi.org/10.1038/nn858`.

Laurenz Wiskott and Terrence J. Sejnowski. Slow Feature Analysis: Unsupervised Learning of Invariances. *Neural Computation*, 14(4):715–770, 2002.

Richard S. Zemel and Terrence J. Sejnowski. A model for encoding multiple object motions and self-motion in area MST of primate visual cortex. *Journal of Neuroscience*, 18(1):531–47, Jan 1998.