A 1ST STEP TOWARDS AN ABSTRACT VIEW OF COMPUTATION IN SPIKING NEURAL-NETWORKS

Pierre Kornprobst, Sandrine Chemla, Olivier Rochel and Thierry Viéville
Projet Odyssée, INRIA/ENPC/ENS, INRIA, BP 93, 06902 Sophia Email: Thierry.Vieville@sophia.inria.fr

ABSTRACT
Neural network information is mainly conveyed through (i) event-based quanta, spikes, whereas high-level representation of the related processing is almost always modeled in (ii) some continuous framework. Here, we propose a link between (i) and (ii) which allows to derive the spiking network parameters given a continuous processing and also obtain an abstract interpretation of the related processing. .

KEY WORDS
Cortical maps; Diffusion operator; Spiking neural networks.

1 Introduction

High-level specification [1, 2, 3] of how the brain represents and categorizes the causes of its sensory input allows to link "what is to be done" (perceptual task) with "how to do it" (neural network calculation). More precisely, a general class of cortical map computations can be specified representing what is to be done as an optimization problem, in order to derive the related neural network parameters considering regularization mechanisms (implemented using so-called partial-differential-equations).

A recent contribution of our team [4] revisits this framework with three add-ons.

[1] It is generalized to a larger class of (non-linear) map computations which specifies a n dimensional vectorial computation map, taking unbiased partial observation of a noisy input into account and using non-linear anisotropic diffusion in order to reduce the noise while preserving the data variation.

Using a scalar or vector valued map is an important feature when addressing the modelization of cortical processing units such as cortical columns [5]. It may also help defining improved models of neurons or small neuronal assemblies, where the state is not only defined by a scalar membrane potential [3].

Introducing non-linear constraints between the map components has several advantages, one is to take noisy measures into account avoiding statistical bias (see e.g. [6] for a development), another is to define physical parameters (e.g. 3D orientation) with complex structure.

The local solution of the previous criterion can be implemented -in the general case- using a network dynamics of the form of Cohen-Grossberg analog network, using linearized integral approximation of a diffusion operator introduced by Cottet, Degond and Mas-Gallic.

In practice, several non trivial mechanisms (segmentation using the Mumford-Shah method family, transparent motion analysis, winner-take all see above) can be derived within this framework, as illustrated during the talk.

[2] On the reverse, standard "analog" neural network, providing their weights are local and "unbiased", can be represented in this framework providing a guaranty of convergence and an abstract view of the underlying processing.

[3] Finally, we also can verify within this framework not only *one* but several cortical maps can interact, with feed-backs, in a stable way using two biological assumptions:
(i) feedback values are smoothed in space, before influencing other maps,
(ii) forward connections define a lattice (thus without loop).

Here we would like to go a step further and discuss how not only analog models of biological neural-networks, but also spiking-neural networks could be implemented using this formalism.

2 Specification of neural-networks from a variational approach.

Perceptual processes architecture, in computer or biological vision [1, 5], is based the computation of "maps" of quantitative values. Such maps encode retinotopic quantities such as contrast magnitude, contrast orientation related to edge orientation, shape curvature, binocular disparity related to the visual depth, color cues, temporal disparity between two consecutive images in relation with visual motion detection, etc. Other maps are not only parametrized by retinotopic locations, but also other parameters (e.g. orientation, retinal velocity, etc.) or more abstract quantities [7].

According to generative approaches [1, 2, 3], the cortical map computation can be modeled as an optimization problem. Let us state this in the very general form proposed in [8].

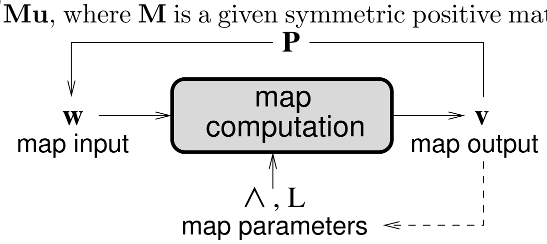
Given an input map w, one look for an output map $\bar{\mathbf{v}}$ verifying

$$\bar{\mathbf{v}} = \underset{\mathbf{v} \in H / \mathbf{c}(\mathbf{v})=0}{argmin} \; \mathcal{L}(\mathbf{v}), \quad \text{with} \tag{1}$$

$$\mathcal{L}(\mathbf{v}) = \int |\hat{\mathbf{w}} - \mathbf{w}|^2_{\mathbf{\Lambda}} + \int \phi(|\nabla \mathbf{v}|_{\mathbf{L}}) + \int \psi(\mathbf{v}), \tag{2}$$

$$\text{and } \hat{\mathbf{w}} = \mathbf{P}\, \mathbf{v}, \tag{3}$$

where ∇ stands for the gradient operator, $\phi(\cdot)$, $\psi(\cdot)$, $\mathbf{P}$, $\mathbf{c}(\cdot)$, $\mathbf{\Lambda}$ and $\mathbf{L}$ are commented hereafter. The norms defined in (2) are weighted norms defined by $|\mathbf{u}|_{\mathbf{M}} = \mathbf{u}^T \mathbf{M} \mathbf{u}$, where $\mathbf{M}$ is a given symmetric positive matrix.



Here is a representation of the model (1)–(3).

The first term in (2) is a fidelity attached term specifying how the output is related to the input, the second term is a smoothing term which defines the regularity of the output and the third term allows to constraint the form of the solution. The equation (3) shows the chosen relation between the estimation of the input given an output. So the formulation (1)–(3) specifies the cortical map computation in the sense that it explains the "goal", what is to be done, but without any reference to how it is done.

The functions $\mathbf{\Lambda}$ define a so-called *measurement information metric* which represents the *precision of the input* (the higher this precision in a given direction, the higher the value of $\mathbf{\Lambda}$ in this direction) and allows to take into account (in a statistical framework, $\mathbf{\Lambda}$ corresponds to the inverse of a covariance matrix) *partial observations and missing data* (i.e. null precision)

The functions $\mathbf{L}$, define a *diffusion tensor* modulated by a function ϕ which controls the amount of smoothness required. Low variations, assumed to be "noise", are smoothed (using e.g. quadratic and isotropic smoothing for additive white noise), while high contrasts, assumed to be the "signal" are preserved (e.g. with diffusion only in the direction tangential to the edges). Furthermore, when a problem is ill-posed, adding some a priory on the smoothness of the solution regularizes the problem (e.g. with diffusion from well-defined values to undefined or ill-defined values).

Three kinds of constraints are introduced: *structural constraints* (via $c()$) , to define a nonlinear solution, i.e. to force the solution to belong belong to a manifold defined by implicit equations; *optimization constraints* (via $\psi()$) to control the form of the solution; *measurement constraints* between both the input and the quantity to estimate (via $\mathbf{P}$) to obtain an unbiased estimation [6] in this non-linear case.

In the cortex, such "neuronal unit" is a cortical hyper-column. Our model can be be mapped onto usual computational model of cortical columns processes (see also [5] for a treatise on the subject). Regarding such a "processing a cortical map being a discrete implementation of such neural units network. unit" (see e.g. [9]), we propose here a possible interpretation of such an abstract analog network.

$\mathbf{w}$	Extra cortical input or intra-cortical input from previous layers
$\mathbf{v}$	Extra cortical or backward intra-cortical output
$\sum_j \sigma_{\cdot j}\, \mathbf{v}_j$	Local connections
$\mathbf{\Lambda}, \mathbf{L}$	Remote backward connections
Iterative operations	Internal connections

This mapping is to be understood as a working assumption. It also make explicit the scale at which such analog networks should be situated. This mapping has the chance to be compatible with the laminar architecture of the cortex or neocortex [10, 11] and with the related inter-layer circuitry. Excitatory but also inhibitory connectivity is included in the diffusion term as detailed in [12].

3 Analog network implementation

The local solution of the previous criterion can be implemented -in the general case- using a network dynamics of the form of Cohen-Grossberg analog network, using linearized integral approximation of a diffusion operator introduced by Cottet, Degond and Mas-Gallic [13, 14, 15, 12]. More precisely, for a neuron of index i, it writes:

$$\dot{v}_i = -\bar{\epsilon}_i(v_i) + \sum_j \bar{\sigma}_{ij}(v_i)\, v_j + \bar{\kappa}_i\, w_i$$

The derivation [4] is however far from being trivial with two key points:

- the term $\bar{\kappa}_i$ is a simple gain, while the corrective term $\bar{\epsilon}_i(v_i)$ is a straight-forward but very complex non-linear function of the criterion parameters. It is thus a real issue to obtain an automatic symbolic derivation. The obtained derivation turns out to be a short-term adaptive rule.

- the term $\bar{\sigma}_{ij}(v_i)$ corresponds to the synaptic weights and a linear family of solutions for $\bar{\sigma}$ is derived for a given diffusion tensor $\mathbf{L}$. Among those solutions:

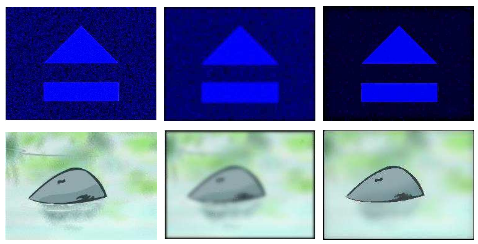
  - an optimal solution (here the closest discrete approximation with respect to the continuous one, in the least-square sense, given a well-formed distance) is chosen [12],

  - the synaptic weights are related to the related diffusion operator using a Hebbian learning scheme [8]

  obtaining, a biologically plausible mechanism of short-term adaptation.

We have been able to not only make the mathematical derivation explicit [8, 4], but, using the `maple` symbolic calculator, make also the computational derivation automatic, including the temporal discretization of the equations for computer implementation [16].

This allows, as illustrated e.g. in Fig. 3 for a simple example of isotropic/anisotropic diffusion, to simulate visual functions considering neural-networks.



Initial image Isotropic Anisotropic

Figure 1. Two examples of results using anisotropic diffusion (right image), the 1st example being the same as in [14] to validate the present method. The original image is on the left.

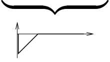
4 Spiking neuron implementation

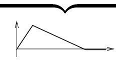
In event based neural network models, the output of a neuron of index i is entirely characterized by the sequence of spike firing times: $\mathcal{F}_i = \{\cdots t_i^n \cdots\}$

At a computational level, it has been shown (see e.g. [17] for a review) that any feed-forward or recurrent (multi-layer) analog neuronal network (à-la Hopfield, e.g. McCulloch-Pitts) can be simulated arbitrarily closely by a insignificantly larger network of spiking neurons with analog inputs and outputs encoded by temporal delays of spikes (even in the presence of noise) [18, 19].

The Gerstner and Kistler Spike Response Model [20] of a biological neuron defines the state of a neuron via a single variable:

$$
\begin{aligned}
u_i(t) \;=\; & \underbrace{\nu_i(t - t_i^*)} \\
+ \; & \textstyle\sum_j \sum_{t_j^n \in \mathcal{F}_j} w_{ij} \, \underbrace{\epsilon_{ij}(t - t_i^*, (t - t_j^n) - \delta_i)} \\
+ \; & R_i \, I(t)
\end{aligned}
$$





where
$u_i()$ is the neuron state, related to the membrane potential, while
$\nu_i()$ describes the neuronal response to its own spike (neuronal refractoriness),
t_i^* being the last spiking time of the ith neuron, and
$\epsilon_{ij}()$ describes the neuronal response to pre-synaptic spikes at time t_j^n post-synaptic potential,
w_{ij} being the *connection strength* (excitatory if $w_{ij} > 0$ or inhibitory if $w_{ij} < 0$),
δ_i being the *adaptive delay* (including axonal delay),
$I()$ being the continuous input currents, for an input resistance R_i.
The neuron fires when $u_i(t) \geq \theta_i$, for a given threshold.

At the computational level, using piece-wise linear profiles yields a closed-form calculation of the spiking events, thus allows to obtain an efficient and exact implementation of (1) in event-based massive neuronal simulators such as MVASPIKE. A complete description of such a mechanism is detailed in e.g. [21]. This is to be compared with other simulations (e.g. [22, 23]) where stronger simplifications of the S.R.M. models have been introduced to obtain a similar efficiency, whereas other authors (e.g. [21]) propose heavy numerical resolutions at each step.

Using this model and following Maas and Natchslager [18, 19], we[1]:
- represent the signal $v_i = t_i - T_\bullet$ as the last spike delay t_j with respect to a given temporal reference $T_\bullet$,
- consider piece-wise linear response profiles (as approximations of Hodhkins-Huxley related profiles),
- introduce a temporal discretization of the input current,

and obtain a direct link with continuous representation of neural map computation:

$$\dot{v}_i = -\bar{\nu}(v_i)\, v_i + \textstyle\sum_j \bar{\sigma}_{ij}(v_i)\, v_j - \bar{\epsilon}_i(v_i) + \bar{\kappa}_i \, I_i$$

while $v_i = g(v_i)$ $\quad g()$:

- the resistive coefficient $\bar{\nu}_i$ being proportional to the spiking threshold θ_i;
- the weights $\bar{\sigma}_{ij}$ being in direct relation with the synaptic weights w_{ij};
- the corrective term $\bar{\epsilon}_i$ being controlled by the axonal delay δ_i;
- the input gain κ_i being controlled by the input resistance.

with closed-form correspondence allowing to explicitly calculate the neural network parameters given an abstract continuous representation.

This relationship is valid only in a given temporal window, with saturation outside, as for analog networks. Here it appears that fast adaptive delays (as observed in recent intra-cellular experiments of e.g. [24]) is a crucial element in this model. In the derivation the constraint $\bar{\nu} = \sum_j \bar{\sigma}_{ij}$ appears. It is coherent with S.T.D.P. adaptation rules (yielding the same constraint) as derived by, e.g., [25] It also corresponds to what is obtained from a variational framework relating the neuronal weights to a continuous diffusion operator, as introduced by Cottet and also[1]. This last formulation is in direct relation with a sub-class of Cohen-Grossberg dynamical systems.

We, for instance, illustrate the previous derivation with an event-based implementation of an early-vision processing layer, (in relation to what has been shown in Fig. 3 in 2D). In Fig 2 where, a 1D spiking neural network (each abscissa corresponds to a neuron and each ordinate to its related temporal value),

[1] See http://www.inria.fr/rrrt/rr-5657.html for details

receives a noisy input (top view) its output after a few iterations (bottom view) correspond to an edge-preserving smoothing of the input, using a non-linear diffusion operator.

Figure 2. An example of 1D implementation of anisotropic smoothing using a spiking network. The abscissa represent the spatial 1D dimension of the linear neural-network with local connections. The ordinates represent, at a given instant, the last spike time with respect to a reference. This is thus in relation with the "instantaneous phase" of the spiking-neurons, thus with short-term coherence between firings.o p view: noisy input. Bottom view: edge-preserving smoothed output.

5 Conclusion

The previous derivation allows to relate variational formulation of perceptual tasks, as defined in computer vision, to neuronal spiking networks. Although, formal derivations have been worked out [8] and implemented, this is still to be considered as a working assumption and a starting point for further investigations.

At a 1st glance, this derivation seems at least valid in situations where spike delays are short, i.e. in the "fast-brain" mode (see e.g. [26]) or when large activity is observed as it is often the case in the related cortical areas (see e.g. [27]). Please refer to [18, 19, 23] for a precise discussion about the validity of considering the spiking network for computing such a weighted sum.

With respect to the state of the art, we not only consider the 1st rising part of post-synaptic potential but also the 2nd decreasing slope, with the potential effects of having excitatory post-synaptic potentials yielding a relative inhibitory effect (with a symmetric effect for inhibitory potentials); the fact we do not exclude such "reverse" effect for delayed responses is in direct relation with recent experimental finding of this kind [28].

A step further, if the previous derivation makes sense, this means that we must observed fast adaptive synaptic mechanisms, with respect to the weights, but also fast adaptive mechanism of delay. This is indeed a recent assumption (see e.g. [29]) which has been observed experimentally and which importance for computational models is to be better understood.

Both elements are key aspects of the model: they must be confronted with biological facts. This is the next step of this work.

This work is realized within the scope of the European Facets project in direct relation with scientific contributions of partners of this consortium.

References

[1] T. Lee and D. Mumford. Hierarchical bayesian inference in the visual cortex. *J. Opt. Soc. Am. A*, 20(7), 2003.

[2] Karl Friston. Functional integration and inference in the brain. *Prog Neurobiol*, 68:113–143, 2002.

[3] P. Dayan and L. F. Abbott. *Theoretical Neuroscience : Computational and Mathematical Modeling of Neural Systems.* MIT Press, 2001.

[4] T. Viéville and P. Kornprobst. Modeling cortical maps with feed-backs. In *Int. J. Conf. Neural Networks*, 2006.

[5] Y. Burnod. *An adaptive neural network: the cerebral cortex.* Masson, Paris, 1993. 2nd edition.

[6] T. Viéville, D. Lingrand, and F. Gaspard. Implementing a multi-model estimation method. *The International Journal of Computer Vision*, 44(1), 2001.

[7] T. Gisiger, S. Dehaene, and J. P. Changeux. Computational models of association cortex. *Curr. Opin. Neurobiol.*, 10:250–259, 2000.

[8] Thierry Viéville. An abstract view of biological neural networks. Technical Report RR-5657, INRIA, August 2005.

[9] J. Petitot and Y. Tondut. Vers une neuro-géométrie. fibrations corticales, structures de contact et contours subjectifs modaux. *Mathématiques, Informatique et Sciences Humaines*, 145:5–101, 1999.

[10] D. Mumford. On the computational architecture of the neocortex. i. the role of the thalamo-cortical loop. *Biological Cybernetics*, 65:135–145, 1991.

[11] R.J. Douglas and K. A. C. Martin. Neuronal circuit of the neocortex. *Ann. Rev. Neuroscience*, 27:419, 2004.

[12] T. Viéville. An unbiased implementation of regularization mechanisms. *Image and Vision Computing*, 23(11):981–998, 2005.

[13] P. Degond and S. Mas-Gallic. The weighted particle method for convection-diffusion equations. *Mathematics of Computation*, 53(188):485–525, 1989.

[14] G.-H. Cottet and M. El Ayyadi. A Volterra type model for image processing. *IEEE Transactions on Image Processing*, 7(3), March 1998.

[15] R. Edwards. Approximation of neural network dynamics by reaction-difusion equations. *Mathematical Methods in the Applied Sciences*, 19:651–677, 1996.

[16] Sandine Chemla. Biologically plausible computation mechanisms in cortical areas. Master's thesis, Master STIC, Univ. Nice Sophia-Antipolis, 2006. in press.

[17] Wolfgang Maass and Christopher M. Bishop, editors. *Pulsed Neural Networks.* MIT Press, 2003.

[18] W. Maass. Fast sigmoidal networks via spiking neurons. *Neural Computation*, 9:279–304, 1997.

[19] W. Maass and T. Natschler. Networks of spiking neurons can emulate arbitrary hopfield nets in temporal coding. *Neural Systems*, 8(4):355–372, 1997.

[20] W. Gerstner and W. Kistler. *Spiking Neuron Models.* Cambridge University Press, 2002.

[21] O. Rochel. *Une approche événementielle pour la modélisation et la simulation de neurones impulsionnels.* PhD thesis, Université Henri Poincaré - Nancy 1, 2004.

[22] Maurizio Mattia and Paolo Del Giudice. Efficient event-driven simulation of large networks of spiking neurons and dynamical synapses. *Neural Computation*, 12:2305–2329, 2000.

[23] Berthold Ruf. *Computing and Learning with Spiking Neurons - Theory and Simulations.* PhD thesis, Institute for Theoretical Computer Science, Univ. Graz, Austria, 1998.

[24] N. Gazeres, L. Borg-Graham, and Y. Fregnac. A model of non-lagged X responses to flashed stimuli in the cat lateral geniculate nucleus. *Visual Neuroscience*, 1998.

[25] R. Guyonneau, R. vanRullen, and S.J. Thorpe. Neurons tune to the earliest spikes through stdp. *Neural Computation*, 2004. In review.

[26] S. Thorpe, A. Delorme, and R. VanRullen. Spike based strategies for rapid processing. *Neural Networks*, 14:715–726, 2001.

[27] D.J. Amit and N. Brunel. Model of global spontaneous activity and local structured delay activity during delay periods in the cerebral cortex. *Cerebral Cortex*, 7:237–252, 1997.

[28] L. Borg-Graham, C. Monier, and Y. Fregnac. Visual input evokes transient and strong shunting inhibition in visual cortical neurons. *Nature*, 393:369–373, 1998.

[29] Robert Legenstein, Christian Naeger, and Wolfgang Maass. What can a neuron learn with spike-timing-dependent plasticity? *Neural Computation*, 2005. in press.