A hypothesis for parallel fiber coding in a cerebellar model of smooth pursuit eye movement

Olivier J.-M. D. Coenen †, Michael P. Arnold ‡, Terrence J. Sejnowski § and Marwan A. Jabri ‡ ⋆
olivier@salk.edu mikea@sedal.usyd.edu.au terry@salk.edu marwan@ece.ogi.edu

†Res. on the Neuroscience of Autism, Children's Hosp. Res. Ctr., La Jolla, California 92037, USA
‡Computer Engineering Lab., School of Electrical and Information Eng., The Univ. of Sydney, Australia
§Computational Neurobiology Laboratory, HHMI, The Salk Institute & UCSD, La Jolla, California, 92037, USA
⋆Electrical and Computer Engineering, Oregon Graduate Institute Beaverton, OR 97006, USA

Abstract

A neural network model based on the anatomy and physiology of the cerebellum is presented. The model learns to generate predictive smooth pursuit eye movements to follow target trajectories, and respond to large tracking error by producing corrective saccades [19]. A biologically motivated learning rule based on Bayesian analysis controls the plasticity at the input layer of the cerebellum [8]. The possibility that this unsupervised learning rule at the granule cell synapses of the cerebellum uncovers latent structures in their mossy fiber inputs is investigated. Using the unique convergence of the granule cells at the mossy fibers glomeruli, the learning rule approximates the emergence of a sparsely-distributed and statistically-independent code at the parallel fibers, in contrast with previous learning rules that only produce a decorrelated representation [13] [18]. Such a code is beneficial for learning downstream at the Purkinje cells: It simplifies the credit assignment problem between climbing and parallel fiber activities, while retaining the ability for generalization that binary codes or fixed synaptic weights used in many cerebellum models lack [19] [30]. Simulations with a spiking model of the cerebellum are used to study the resulting representation at the parallel fibers and Purkinje cells.

1 Introduction

The cerebellum (Cb) has been suggested to act as a short-term predictive engine in the brain [7] [1] [23] whose processing may require three distinct functional stages: 1) the transformation and combination of cerebellar inputs in a pre-processed form appropriate for predictions; 2) the identification and selection of the pre-processed inputs that define the context of the prediction and that can anticipate the neural activity to predict and 3) the construction of the predictions themselves. The first stage has been suggested to occur in the granular layer (fig. 1), the second stage at the Purkinje cells and inhibitory interneurons and the third at the Purkinje cells and deep cerebellar nuclei neurons [7].

The computational strength of the cerebellum may be to produce predictions or predictive neural commands dependent on precise contextual information. Hence, as context changes are recognized, the cerebellar predictions will change accordingly. We suggest that the information processing of mossy fiber inputs by the granular layer of the cerebellum gives the cerebellum the required sensitivity to contextual changes to produce precise context-dependent predictions. By definition, a context is a situation that can occur independently of another and that is composed of multiple elements happening together. Hence, contextual information at the cerebellum could be provided by a set of granule cells firing independently of other sets, which in turn specify other contexts. A sparse and distributed representation that minimizes statistical dependencies in the parallel fibers would therefore give the cerebellum the ability to recognize different contexts and construct context-specific predictions effectively. The theoretical development of learning rules in the granular layer of the cerebellum in this paper suggests that one potential advantage of cerebellar computation for machines is to acquire the ability to learn different tasks in different contexts something that is essential for continuous and long-term learning.

Recently, probabilistic and information theoretic approaches to learning in neural networks have made new interesting connections with neural functions. For example, the representation in networks (i.e. what is encoded in their units) after being presented with a series of natural images (forests, fields, mountainous terrains, ...) at its inputs is found to be similar to the responses of cells in the visual cortex [6] [26] etc. These networks learn in an unsupervised manner (i.e. without a teacher to give the correct answer) to extract latent structures in their inputs [5] [15] [20] [21] [25]. These results hint at the possibility to understand the functional processing of neurons and their diverse responses by using general principles of Bayesian probability and information theory. However, all of these re-

sults have been obtained in systems with early sensory signals and it has not been clear how to apply these principles in higher level of cortical processes (although see [33]).

In this paper, we investigate whether these techniques could be applied to a well known system in the brain that is a corner stone of automatic learning and adaptation in the nervous system, the cerebellum [8]. We present preliminary results demonstrating that such techniques may explain the first layer of processing in the cerebellum.

2 Methods and Results

The role of the granular layer which consists of mossy fiber (Mf) glomeruli (Gl), granule (Gc) and Golgi (Go) cells is two fold. One is to transmit to the Purkinje cells through the parallel fibers a complete contextual account of mossy fibers (Mfs) activity, and the other is to provide it in a form which facilitates adaptation in the Purkinje cells and inhibitory interneurons of the cerebellar cortex. We maintain that a sparse and distributed representation in the parallel fibers that maximizes the mutual information between the mossy fibers (Mfs) and the parallel fibers and minimizes the statistical dependencies among parallel fibers fulfills these two roles.

A difficult credit assignment problem The Purkinje cells, which receive on the order of 100 000 parallel fibers, axons of the granule cells (Gcs), face a difficult credit assignment problem in identifying which parallel fiber synapse must be modified in connection with climbing fibers activity, which direct learning at the Purkinje cells [2] [7] [22] [28]. Cerebellum models often solve this problem by using thresholding together with a binary code to limit the number of active parallel fibers [19] [30]. These models will often have poorer generalization abilities than an analog code would. One way to keep the generalization benefits of an analog code and solve the credit assignment problem may be to use a sparsely-distributed [12] and (nearly) statistically-independent representation [5] in the parallel fibers. A sparse and distributed code tends to minimize the time during which cells are active, and a statistically-independent representation minimizes the redundancy across active cells. Both properties reduce the complexity of the credit assignment between active parallel fibers and climbing fiber at a Purkinje cell more effectively than previous learning rules that only produced a decorrelated representation [13] at the parallel fibers [18].

The granular layer: anatomy Mossy fiber inputs to the cerebellum terminate in glomeruli where granule cell dendrites and Golgi cell (Go) axons converge to make synaptic contacts (fig. 1). A glomerulus contacts about 20 to 50 granule cells through excitatory NMDA and AMPA receptors and granule cells receive a combination of 4 to 7 mossy fibre inputs [16] [27]. The large number of Gcs in the Cb may be related to the large number of possible Mf input combinations [22]. A Golgi cell integrates the output of approximately 1000-6000 granule cells and receive inputs from a number of mossy fibres.

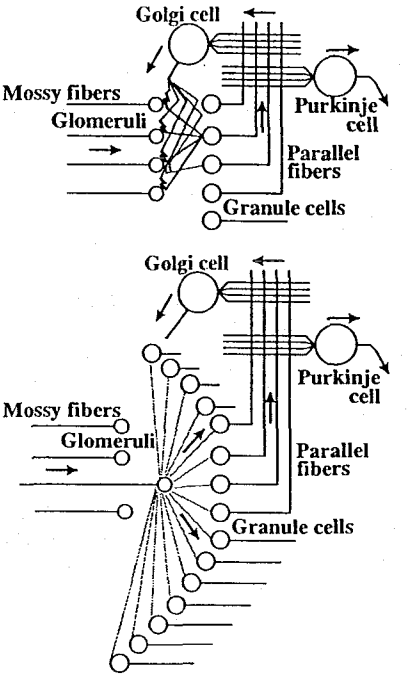

Figure 1: Granular layer of the cerebellum considered in the model. The mossy fibers (Mfs) end in glomeruli (Gl) which make contact with granule cells (Gcs) dendrites. The granule cells send their parallel fiber axons to the Purkinje cells and also make contact with a Golgi cell (Go) which inhibits every granule cell dendrites at the glomeruli. Golgi cells also receive a number of inputs from mossy fiber glomeruli directly (not shown). A granule cell (Gc) receives 4-7 mossy fiber (Mf) inputs (top) whereas a mossy fiber glomerulus contacts 20-50 granule cells (bottom) [16] [27]. The Gl-Gc and Go-Gc synapses are represented respectively by matrices $\mathbf{W}$ and $\mathbf{V}$ of identical dimensions, where v_{ji} is the weight of the inhibitory Go-Gc synapse above the excitatory Gl_i-Gc_j synapse with weight w_{ji}. Note that the Purkinje cells, output of the cerebellar cortex, are not part of the granular layer, but are shown to complete the cerebellar network. The arrows indicate the propagation of activity in the network.

The Gl is the site of complex receptor interactions. A Gc receives GABAergic inhibition from a Golgi cell axon on its dendrites directly above the mossy fibre glomerulus synapses. Glycine is also released by the Go in addition to GABA, and has been shown to potentiate the NMDA response in cultured mouse brains neurons [17]. Moreover, impairment of $GABA_A$ receptor activity by NMDA

receptor activation in rat cerebellum granule cells has been observed [29] whereas 'spillover' glutamate from mossy fibers has been shown to inhibit GABA release from Golgi cell terminals by activating presynaptic metabotropic glutamate receptors (mGluRs) [24]. These complex interactions are not modeled here but are the subject of current investigations.

The granular layer: computation The processing of a Gc in the granular layer may be visualized as two nested projections of the multidimensional Mf space [8]. The first is a projection of the multidimensional Mf space on to the 4-7 dimensional subspace defined by the 4-7 Mf inputs to the Gc. The ensemble of Gcs is therefore attempting to represent the high multidimensional Mf space through a large ensemble of projections onto 4-7 dimensional subspaces. The second projection is the projection of the 4-7 dimensional Mf subspace defined by the Gc inputs onto the one-dimensional firing of each Gc. Whereas the first projection is fixed during development, this second projection is plastic, as recent evidence of LTP at the Mf-Gc synapses show [9], and is defined by the synaptic weight values at the Gl, namely the Gl-Gc and Go-Gc synapses. The question is what should the weight values be? Ignoring the influence of the Go cell for the moment, the question is equivalent to asking in what direction in the 4-7 dimensional Mf input subspace should the Gc orient its response? One answer is in the direction which gives the Gc the sparsest probability density function (i.e. one with high kurtosis) [12] or the most non-Gaussian density function [14] [15]. (Note that a sparse density is a particular example of a non-Gaussian density function.) The reason is that Mf inputs are not random, and therefore that some degree of redundancy will be present. We hypothesized [8] that this redundancy produces sparse distributions in the Mf subspaces [12].

The Go inhibitory inputs may have different effects on the Mf to Gc projection. If the inhibitory weights onto all the dendrites of one Gc are equal, then the net effect could simply be to reduce the overall Gc activity, or set a threshold that vary with the Go activity. On the contrary, if the inhibitory weights onto each dendrite of a particular Gc can take on different values, then the Go inhibition effect is more complex. As the Go activity changes, the inhibition can change on-line, by different amount for each Gc dendrites, the *effective weights* associated with Mf inputs. The net result is a rotation of the Gc sensitivity in Mf input space that varies with the activity of the Go. This modulation may allow the parallel fibers to flexibly encode specific spatio-temporal patterns of mossy fibers activity. As the context changes and that mossy fibers activity changes, the Gc activities will change, which in turn modulate the activity of the Go that receives inputs from these Gcs. The change in Go activity in turn may change the orientation of the sensitivity of the Gcs in Mf input space, causing further changes in Gc activities. The Gcs may therefore track changes in Mf inputs in particular directions dictated by the overall context that sets the Go activity level.

One of the best description of the fundamental nature of data is given by maximizing the independence index of projection as offered by independent component analysis (ICA); this procedure has been successfully used for unsupervised exploratory projection pursuit [15]. This means that the sparsest direction (or most non-Gaussian) in the Mf input subspace, and therefore the Mf-Gc synaptic weights, can be found, in principle, by using algorithms that minimize the statistical dependencies of the outputs [5] [15].

The granular layer: model A linear relationship between the activity of the mossy fiber inputs x and the granule cells activity s is assumed. Furthermore, we assumed that a Go contacts every Gc dendrites at every Gl in its arbor so that the Go-Gc synapses can be represented by a weight matrix V. The number of Gl-Gc and Go-Gc synapses at a Gl are therefore equal, with respective matrices W and V of identical dimensions (top panel in fig. 1). The Go is modeled as having a substracting effect on the Mf input x so that the Gc activity is modeled as $s = Wx - V\underline{1}z \geqslant 0$, where z is the Go activity, and $\underline{1}$ is a column vector of 1's. We also write $\mathbf{z} = \underline{1}z$, to represent a column vector of identical z values. The granule cell activities are therefore the responses of the projection of x into s performed by the weights W and activity $V\mathbf{z}$. In the following, we assume that the number of granule cells receiving the same mossy fiber inputs is equal or smaller than the number of mossy fiber inputs to a granule cell, i.e. that if a granule cell receives 4 Mf inputs there are at most 4 granule cells receiving these same 4 Mf inputs. This corresponds to a *locally* undercomplete (< 4) or complete ($= 4$) representation of each set of 4 Mf inputs by the Gc, although, since the number of Gcs exceeds the number of Mfs by a factor of more than 10 000, the global representation at the granular layer may itself be overcomplete or not.

Role of Golgi cells If the Mf inputs have the form $x = \hat{x} + x_o$ where x_o is the mean of x and $\hat{x}$ has zero mean, x_o will be large and positive since the firing rates for the mossy fibres may be large and positive. As a result, the Gc activity $s = Wx + w_0$ will remain centered at $\mathbf{0}$ only if the bias weight vector is $w_0 = -Wx_o$ and negative.

The bias w_o is approximated by $-Vz$, giving to the Go the role of setting the threshold of Gcs so that their activities s remain sparsely distributed at 0.

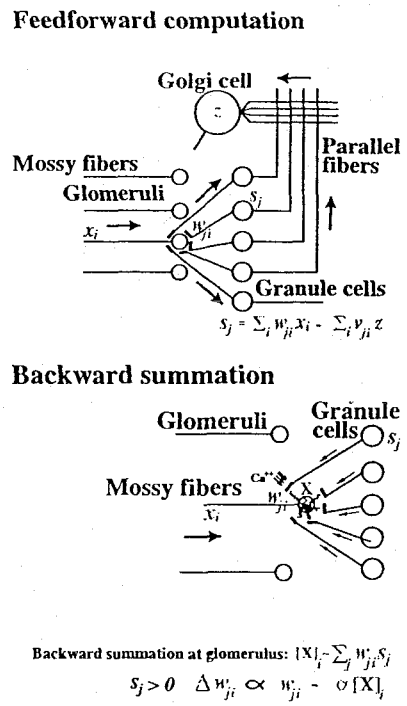

Figure 2: Feedforward computation in the model for computing cell responses and backward summation used for directing plasticity of the synaptic weights [8]. The variables x, s and $z \geq 0$ represent the firing rate activity of the mossy fibers, granule cells and Golgi cell respectively (top). The quantity X_i directs learning at the glomerulus i (bottom) and is computed using the activity of granule cells propagated back along their dendrites (granule cells are electrotonically compact, see also [31]).

Learning rules The Gc activities are modeled as $s = Wx - Vz$ with $s, x, z \geq 0$ with a sparse (high kurtosis) and statistically independent prior probability density $f_s(s) = \prod_i f_s(s_i)$ where $f_s(s_i)$ was chosen to be the same exponential density for all Gcs, $f_s(s_i) = \alpha \exp(-\alpha s_i) = \gamma_1(s_i)$, and where $\gamma_1(\cdot)$ is the gamma density of order 1 [1]. x, s and z represent the firing rate of the Mf, Gc and Go respectively. Due to the positive constraint, the prior is more precisely $f_s(s_i) = \alpha e^{-\alpha s_i} U(s_i)$ where $U(\cdot)$ is the step function. To simplify the derivation of the learning rules, this prior was approximated by two exponential priors, $f_s(s_i) = \alpha \exp(-\alpha s_i)$ for $s > 0$ and $f_s(s_i) = \alpha \exp(\beta s_i)$ for $s \leq 0$, where $\beta >> \alpha$. Taking the limit as $\beta \to \infty$, the original prior with the step function is recovered.

The Go activity is given by the sum of Gcs activity: $z = \sum_{i=1}^{N} s_i$, where the number of Gcs N is about 1000-6000 [27]. The Go density turns out to be a gamma density of order N, $f_z(z) = \gamma_N(z)$ and for large N, $N >\sim 200$, the gamma density approaches a Gaussian density with a mean $\mu(z) = N/\alpha$ and a variance $\sigma^2 = N/\alpha^2$. In the following, the simplifying assumption that the Go cell activity is independent of the mossy fiber inputs x and constant at its mean value is made. The more realistic dependent case is currently being investigated.

Bayesian derivation with maximum likelihood [5] [21] [25] The objective is to maximize the probability density of the input data (X) given the model. The likelihood function in terms of M observations x_k of x is $f_X(X|W,V) = \prod_{k=1}^{M} f_x(x_k|W,V)$. Assuming a complete representation where 4 Gcs receive the same 4 Mf inputs, the 4-dimensional Mf input can be written as $x = W^{-1}(s + Vz)$ in the linear regime of $s = Wx - Vz$ by inverting the network. Droping the index k, the density of a single data point is obtained by marginalizing over the states of the network, $f_x(x|W,V) = \int f_x(x|s,z,W,V)\, f_{sz}(s,z)\, ds\, dz$ where $f_x(x|s,z,W,V) = \delta(x - W^{-1}s - W^{-1}Vz)$ and where $\delta(\cdot)$ is the n-dimensional delta function. The joint density distribution $f_{sz}(s,z) = f_s(s) f_z(z)$ since s and z are assumed independent here, and $f_z(z)$ is a delta function, since z is assumed fixed at its mean value.

The learning rules for $\{w_{ji}\}$,$\{v_{ji}\}$ are derived by taking the gradient of the log likelihood and multiplying the results by WW^T [3] [21]:

$$\Delta w_{ji} \propto w_{ji} - \alpha \sum_j s_j\, w_{ji} \qquad (1)$$

$$\Delta v_{ji} \propto +\alpha \qquad (2)$$

for active granule cells $s_j > 0$ [8]. The complete learning rules are shown in fig. 2 (bottom); the synaptic weight update rules change sharply depending on whether the Gc is active or not. Notice that a backward summation $\sum_j s_j\, w_{ji}$ from Gc activity is required at the ith Gl and that the particular connectivity at the Gl makes its computation possible (see below). This summation

[1] The gamma density of order N is $\gamma_N(s_i) = \frac{\alpha^N}{\Gamma(N)} s_i^{N-1} e^{-\alpha s_i}$ where $\Gamma(N)$ is the gamma function. Gamma densities have the property that the density $f_z(z)$ of the sum of two independent random variables $z = s_1 + s_2$ with respective gamma densities of order p and q, $\gamma_p(s_1)$ and $\gamma_q(s_2)$, is a gamma density of order $p + q$, $f_z(z) = \gamma_{p+q}(z)$.

is unique to the ith Gl, and is the same for all weight changes Δw_{ji} at that Gl, but the difference $w_{ji} - \alpha \sum_j s_j w_{ji}$ is unique to each synapse at that Gl.

Although these learning rules were derived for the complete case, Girolami *et al.* [15] showed that the same learning rules hold whether s forms a complete or undercomplete representation of x, i.e., in our case, whether the number of Gc s with the same Mf inputs is equal or smaller than the number of Mf inputs x to s.

The backward summation $\sum_j s_j w_{ji}$ is biological plausible in the granular layer of the cerebellum due to the unique convergence of information at the glomeruli. Because the Gcs are electrotonically compact, the spiking activity at the soma is assumed to be reflected at the dendrites. In addition, patch-clamp recordings made from the dendrites of neocortical pyramidal cells in brain slices have showed that action potentials initiated first in the axon can actively propagate back into the dendritic tree [31]. Na+-dependent action potentials have also been shown to backpropagate over the dendrites in an activity-dependent manner in CA1 pyramidal neurons of the rat hippocampus [32]. The biophysical mechanisms for computing the backward summation and its distribution at every synapse at the Gl are currently being worked out [10].

Smooth pursuit model A spiking neural network model based on the anatomy and physiology of the cerebellum is presented. The model learns to generate predictive smooth pursuit eye movements to follow target trajectories, and respond to large tracking error by producing corrective saccades [4] [8] [19].

The granular layer contained 1955 Gcs that received 200 Mf inputs encoding for different aspects of eye movements and target: 10 Mfs encoded the target position and velocity for each, 90 Mfs encoded the position and velocity of the eye relative to the head for each [4]. Each Gc received inputs from a single Go and a unique set of 4 Mfs. The Gcs projected to 10 Go and a single Purkinje cell. The Purkinje cell projected to a cerebellar nucleus neuron and received a climbing fiber from the inferior olive that encoded retinal velocity slip of the target during smooth pursuit. The nucleus neuron encoded the output eye velocity to follow a target that oscillated in one dimension with angular position given by $\theta = 3\sin(2\pi t)/4$ where t is the time. The parameters of the Gc prior were $\alpha = 0.2$ and $\beta = 10$. To adapt the *analog* learning rules to a spiking network, firing rates were computed by taking a moving average of activity over a period of 50 ms and making synaptic weight updates every 100 ms. A saccade to the target was generated whenever the eye was more than 0.25°away from the target; the number of saccades in a given period of time was therefore an indication of the performance of the system in smooth pursuit.

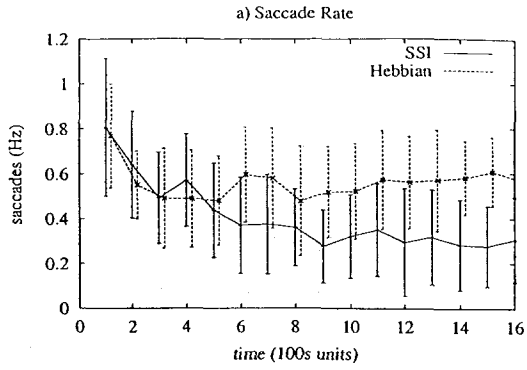

Figure 3: Performance comparison of the smooth pursuit cerebellar model for two different granular layer representations. The mean and standard deviation of the saccade rate over a sliding 100 seconds window are shown for the sparse and statistically-independent (SSI) representation obtained from the present learning rules and a decorrelated representation (Hebbian) that used Hebbian and anti-Hebbian learning rules for the Mf-Gc and Go-Gc synapses, respectively [18]. The number of saccades per second (Hz) is representative of the performance of the system since a saccade was generated whenever the eye deviated from the target by more than 0.25°away from the target [4] [8] [19].

The performance of the system with the present learning rules that approximate a sparse and statistically-independent (SSI) representation in the Gcs was compared to the performance with learning rules leading to a decorrelated representation (Hebbian) at the Gcs [18]. The learning rules for the decorrelated representation were: Hebbian for the Mf-Gc synapses and anti-Hebbian for the Go-Gc synapses. The performances for the two representations are compared in fig. 3. The SSI representation has a mean saccade rate that continues to decrease whereas the Hebbian representation remains at the same level after an initial drop.

Examination of the Gc activities shows a sparser pattern of activities developing in the SSI model than in the Hebbian model (fig. 4). With the SSI representation, the Gc distributions closely match the exponential priors used. Note that here the peak of the priors was chosen to be at 10 Hz instead of zero for easier comparison with the other representation. In the Hebbian representation the distributions have significantly higher means and variances.

This paper demonstrates the plausibility that the granule cells uncover latent structures in their mossy fiber inputs using unsupervised learning rules in the granular layer. The current results

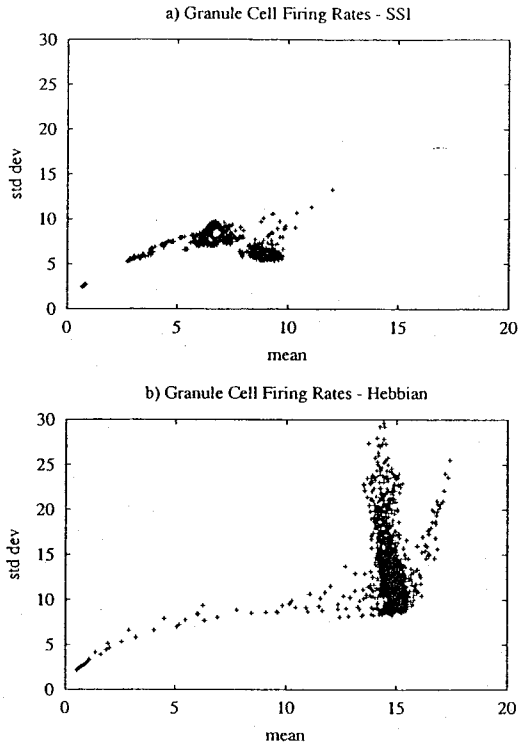

Figure 4: Comparison of the distribution of activities in the granular layer for a) the sparse and statistically-independent (SSI) and b) decorrelated (Hebbian) representations. Each point in the scatter plots corresponds to the mean and standard deviation of the firing rate for one of the 1995 granule cells over the final 100 seconds of the simulation. The firing rates were calculated from the spike trains over a 100 ms window.

support the hypothesis that these powerful unsupervised computations facilitate learning at later stages in the cerebellum. Analysis, simulations, as well as robotic experimentations are currently under way to further expand the results presented here, and biophysical mechanisms supporting our hypothesis will be presented shortly [11].

3 Acknowledgements

OJMDC thanks Dr. Tewon Lee for initial discussion early on the project. This work was supported by a McDonnell Pew Foundation grant and a Girling Watson Fellowship to OJMDC, by the Howard Hughes Medical Institute to TJS and by the Australian National Research Council to MAJ.

References

[1] N. A. Akshoomoff, E. Courchesne, and J. Townsend. Attention coordination and anticipatory control. *Int. Rev. Neurobiol.*, 41:575–598, 1997.

[2] J. S. Albus. A theory of cerebellar function. *Math. Biosci.*, 10:25–61, 1971.

[3] S. Amari. Natural gradient works efficiently in learning. *Neural Computation*, 10:251–276, 1998.

[4] M. P. Arnold. Olivo-cerebellar model for the smooth pursuit eye movement task using spike response networks. Technical report, Computer Engineering Lab., University of Sydney, Sydney, Australia, 2000.

[5] A. J. Bell and T. J. Sejnowski. An information-maximization approach to blind separation and blind deconvolution. *Neural Comput*, 7(6):1129–59, Nov. 1995.

[6] A. J. Bell and T. J. Sejnowski. The "independent components" of natural scenes are edge filters. *Vision Res*, 37(23):3327–38, Dec. 1997.

[7] O. J.-M. D. Coenen. *Modeling the Vestibulo-Ocular Reflex and the Cerebellum: Analytical & Computational Approaches.* PhD thesis, University of California, San Diego, 1998. Physics Department. Available at www.cnl.salk.edu/~olivier.

[8] O. J.-M. D. Coenen, M. Arnold, M. A. Jabri, E. Courchesne, and T. J. Sejnowski. A hypothesis for parallel fiber coding in the cerebellum. In *Society for Neuroscience Abstracts*, volume 25. Society for Neuroscience, 1999.

[9] E. D'Angelo, P. Rossi, S. Armano, and V. Taglietti. Evidence for NMDA and mGlu receptor-dependent long-term potentiation of mossy fiber-granule cell transmission in rat cerebellum. *J Neurophysiol*, 81(1):277–87, January 1999.

[10] D. Eagleman, O. J.-M. D. Coenen, P. R. Montague, and T. J. Sejnowski. Cerebellar glomeruli: Sparse encoding by ensheating a limited resource? 2000. In preparation.

[11] D. Eagleman, O. J.-M. D. Coenen, P. R. Montague, and T. J. Sejnowski. Cerebellar glomeruli: Sparse encoding by ensheating a limited resource? 2000. *Journal of Neuroscience.* In preparation.

[12] D. J. Field. What is the goal of sensory coding? *Neural Computation*, 6:559–601, 1994.

[13] P. Foldiak. Forming sparse representations by local anti-Hebbian learning. *Biol.Cybern.*, 64(2):165–70, 1990.

[14] J. H. Friedman. Exploratory projection pursuit. *Journal of the American Statistical Association*, 82(397):249–266, 1987.

[15] M. Girolami, A. Cichocki, and S.-I. Amari. A common neural network model for unsupervised exploratory data analysis and independent component analysis. *I.E.E.E. Transactions on Neural Networks*, 1998.

[16] R. L. Jakab and J. Hamori. Quantitative morphology and synaptology of cerebellar glomeruli in the rat. *Anat Embryol (Berl)*, 179(1):81–8, 1988.

[17] J. W. Johnson and P. Ascher. Glycine potentiates the NMDA response in cultured mouse brains neurons. *Nature*, 325(5):529–531, 1987.

[18] H. J. Jonker, A. C. Coolen, and J. J. Denier van der Gon. Autonomous development of decorrelation filters in neural networks with recurrent inhibition. *Network*, 9(3):345–62, Aug 1998.

[19] R. E. Kettner, S. Mahamud, H. C. Leung, N. Sitkoff, J. C. Houk, B. W. Peterson, and B. A. G. Prediction of complex two-dimensional trajectories by a cerebellar model of smooth pursuit eye movement. *Journal of Neurophysiology*, 77(4):2115–2130, 1997.

[20] M. S. Lewicki and T. J. Sejnowski. Learning overcomplete representations. *Neural Comput*, 12(2):337–65, Feb. 2000.

[21] D. J. C. MacKay. Maximum likelihood and covariant algorithms for independent component analysis. Unpublished manuscript, 1996.

[22] D. Marr. A theory of cerebellar cortex. *J. Physiol.*, 202:437–470, 1969.

[23] R. C. Miall, D. J. Weir, D. M. Wolpert, and J. F. Stein. Is the cerebellum a Smith predictor? *Journal of Motor Behavior*, 25(3):203–216, 1993.

[24] S. J. Mitchell and R. A. Silver. Glutamate spillover suppresses inhibition by activating presynaptic mGluRs. *Nature*, 404(6777):498–502, Mar 30 2000.

[25] B. A. Olshausen. Learning linear, sparse, factorial codes. A.I. Memo 1580. Technical report, Massachusetts Institute of Technology, Cambridge, Mass., 1996.

[26] B. A. Olshausen and D. J. Field. Emergence of simple-cell receptive field properties by learning a sparse code for natural images [see comments]. *Nature*, 381(6583):607–9, Jun 13 1996.

[27] S. L. Palay and V. Chan-Palay. *Cerebellar Cortex, Cytology and Organization*. Springer-Verlag, 1974.

[28] J. L. Raymond and S. G. Lisberger. Neural learning rules for the vestibulo-ocular reflex. *Journal of Neuroscience*, 18(21):9112–29, Nov 1 1998.

[29] M. Robello, C. Amico, and A. Cupello. A dual mechanism for impairment of GABAA receptor activity by NMDA receptor activation in rat cerebellum granule cells. *European Biophysics Journal*, 25(3):181–7, 1997.

[30] J. Spoelstra, N. Schweighofer, and M. Arbib. Cerebellar learning of accurate predictive control for fast-reaching movements. *Biol. Cybern.*, 82(4):321–33, 2000.

[31] G. Stuart and B. Sakmann. Active propagation of somatic action potentials into neocortical pyramidal cell dendrites. *Nature*, 367(6458):69–72, 1994.

[32] H. Tsubokawa and W. Ross. Muscarinic modulation of spike backpropagation in the apical dendrites of hippocampal CA1 pyramidal neurons. *Journal of Neuroscience*, 17(15):5782–91, 1997.

[33] P. K. Young, M. A. Jabri, S. Y. Lee, and T. J. Sejnowski. Independent components of optical flows have MSTd-like receptive fields. In *Proceedings of ICA '2000*, pages 597–601, Helsinki, Finland, June 2000.