Model of the granular layer of the cerebellum

David Philipona **Edouard Dognin** **Olivier J.-M. D. Coenen**
{davidp, dognin, coenen}@sony.csl.fr
Neuroscience Team
Sony Computer Science Laboratory
6, rue Amyot, 75005 Paris, France

Abstract

Involved in sensorimotor and even cognitive coordination, the cerebellum receives inputs from diverse brain areas. These inputs are recoded by the granule cells, the most numerous cells in the brain, and together with the inhibitory Golgi cells form the granular layer. A granular layer model is presented. It suggests that the granular layer performs a recoding of the mossy fibers into a sparse representation that permits optimal noise reduction by the Golgi cell and facilitates learning in the molecular and Purkinje layer of the cerebellum. Hence, although simple, we suggest that the granular layer contains different types of signal processing that lead to a robust and efficient representation for coordination.

1 Introduction

The architecture of the cerebellum suggests that it's functionality can be divided in two main processing regions: the granular layer and the Purkinje layer. The anatomical cohesion of the granular layer with the inhibitory feedback of Golgi cell (Go) is considered relatively independent functionally. This idea is reinforced considering that a parallel fiber with a length of about 2.6 mm on each side in human will contact about 225 Purkinje cells (PCs), each possibly receiving very different climbing fiber inputs. Therefore, each Purkinje cell may be directed to learn very different functions, yet sharing the same parallel fibers (PFs) inputs. Hence, this diversity in Purkinje cell specialization suggests that, to remain beneficial to all Purkinje cells, the processing in the granular layer should remain essentially unsupervised. Its purpose is a general recoding of the mossy fibers (Mfs) inputs ([13]; [1]): 1) to transmit to Purkinje cells complete contextual account of mossy fiber activity, 2) in a form that facilitate learning at the Purkinje cells and 3) that minimizes destructive interference between learning tasks. We have suggested earlier that a sparse and distributed representation in the parallel fibers that is constructed to maximizes the mutual information between mossy fibers and parallel fibers and to minimize the statistical dependencies among parallel fibers would fulfills the three roles listed above ([5]; [6]). For such a representation to occur, the codon of Marr must be adaptive. The adaptation should rely on the different statistical dependencies in the mossy fibers signals that are created when an animal perform different activities. We consider the case when mossy fibers inputs contain noise, and suggest new roles for the Golgi cell inhibition in the granular layer network.

Anatomy The granular layer receives mossy fibers that form the glomeruli; these make contact with the granule cells (Gcs) dendrites. Mossy fiber inputs to the cerebellum terminate in glomeruli where granule cell dendrites and Golgi cell axons converge to make synaptic contacts (fig. 1). A glomerulus contacts about 20 granule cells through excitatory receptors and granule cells receive a combination of 4 to 7 mossy fiber inputs [14]. The granule cells send their parallel fiber axons to the Purkinje cells and make contact with Golgi cell(s), which inhibit(s) every granule cell dendrites at the glomeruli. A Golgi cell integrates the output of approximately 1000-6000 granule cells and receive inputs from a number of mossy fibers. A granule cell receives inhibition from a Golgi cell axon on its dendrites on top of the mossy fiber glomerulus synapses. The outward expansion in the diagram suggests that there are many more granule cells than mossy fibers inputs. In the granular layer, all synapses are excitatory, except for the Golgi cell axon, which is inhibitory.

2 Methods

Although anatomically the Golgi cell inhibition is located at the dendrites of granule cells, we model the inhibition as summing linearly at the soma due to the electrotonic compactness of these cells ([7]). Hence, we write the relation between the firing frequencies as (fig. 1):

$$\boldsymbol{S} = \sigma(\boldsymbol{W}\boldsymbol{X} - v\boldsymbol{z}); \quad \boldsymbol{z} = \sigma(\mu^T \boldsymbol{S} + \gamma^T \boldsymbol{X} - \theta) \tag{1}$$

where $\boldsymbol{X}$ and $\boldsymbol{S}$ are vectors that represent the firing frequency of the mossy fibers and granule cells respectively, $\boldsymbol{W}$ is the weight matrix of their excitatory synapses, $\boldsymbol{z}$ is the firing rate of the Golgi cell with $\boldsymbol{v}$ the weight vector of its inhibitory synapse onto a granule cell, $\boldsymbol{\mu}$ is the weight vector of the parallel fibers synapses onto the Golgi cell, $\boldsymbol{\gamma}$ is the weight vector of the excitatory synapses of the mossy fibers onto Golgi cell dendrites, and $\boldsymbol{\sigma}$ is a function describing the firing rates relationship between mossy fibers and granule cells. All vectors and matrix are constrained to be positive, and the function $\boldsymbol{\sigma}$ is typically a squashing function limiting the upper firing rate of the cells. Nevertheless, in a non-saturated regime, this function is taken to be linear as the granule cells respond linearly to input currents ([7]).

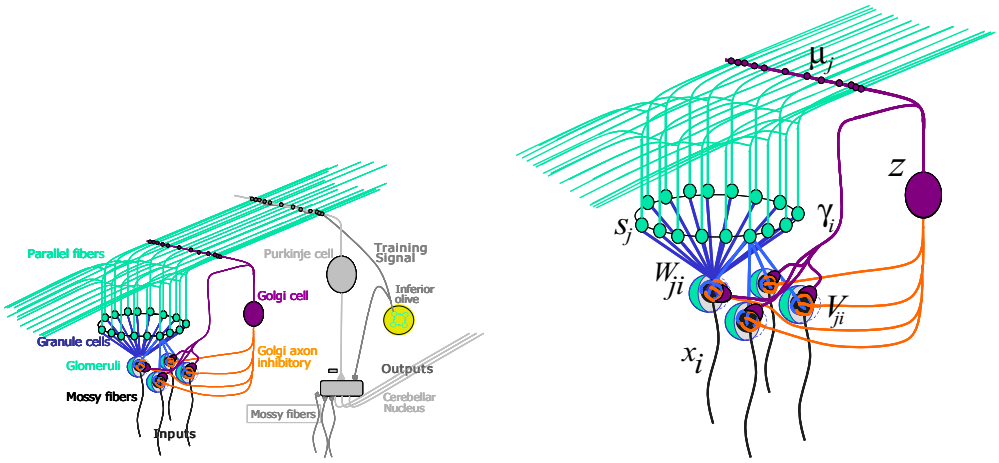

Figure 1: The left diagram illustrates the connectivity in the cerebellum. The right diagram presents the notation used. $\boldsymbol{x_i}$: $\boldsymbol{i^{th}}$ mossy fiber firing rate, $\boldsymbol{W_{ji}}$: the $\boldsymbol{i^{th}}$ mossy fiber to the $\boldsymbol{j^{th}}$ granule cell synaptic weight, $\boldsymbol{s_j}$ is the $\boldsymbol{j^{th}}$ granule cell firing rate, $\boldsymbol{\gamma_i}$: the $\boldsymbol{i^{th}}$ mossy fiber synaptic weight to the Golgi cell, $\boldsymbol{\mu_j}$: the $\boldsymbol{j^{th}}$ parallel fiber synaptic weight to the Golgi cell, $\boldsymbol{V_{ji}}$: the inhibitory synaptic weight of the Golgi cell to the granule cell dendrite superimposed over the $\boldsymbol{W_{ji}}$ synaptic connection, and finally $\boldsymbol{z}$ the firing rate of the Golgi cell.

We note that for this system, the positivity constraint for the weight matrix $\boldsymbol{W}$ can be lifted through a simple rewriting since we have: $\left\{\boldsymbol{W} - v\gamma^T | \boldsymbol{W} \in \mathcal{M}(\mathcal{R}_+, n, n), \gamma \in \mathcal{R}_+^n\right\} = \mathcal{M}(\mathcal{R}, n, n)$. If we assume that the Golgi cell is away from saturation[1], we can rewrite Eq. 1: $\boldsymbol{S} = \sigma((\boldsymbol{W} - v\gamma^T)\boldsymbol{X} - v\bar{z})$ with $\bar{z} = \mu^T \boldsymbol{S} - \theta$ and any optimization using $\boldsymbol{W} \in \mathcal{M}(\mathcal{R}_+, n, n), \gamma \in \mathcal{R}_+^n$ based on these equations is equivalent to an optimization using $\widetilde{\boldsymbol{W}} \in \mathcal{M}(\mathcal{R}, n, n)$ based on

$$\boldsymbol{S} = \sigma(\widetilde{\boldsymbol{W}}\boldsymbol{X} - v\bar{z}); \quad \bar{z} = \mu^T \boldsymbol{S} - \theta \tag{2}$$

where $\widetilde{\boldsymbol{W}}$ is not constrained to be positive (but where $\boldsymbol{v}$ is now strictly positive, i.e. non-null). In the non-saturated region of the granule cells, the complete recoding of the mossy fibers is obtained from Eq. 2:

$$\boldsymbol{S} = \left(\boldsymbol{I} - \frac{1}{1 + \mu^T v} v\mu^T \widetilde{\boldsymbol{W}}\right) \boldsymbol{X} + \frac{\theta}{1 + \mu^T v} v \tag{3}$$

where $\widetilde{\boldsymbol{W}} = \boldsymbol{W} - v\gamma^T$.

2.1 Sparse coding in the parallel fibers

The hypothesis is that different statistical dependencies in the mossy fibers signals are created when an animal perform different activities, whether sensory, motor or cognitive, or mixtures of all three. The

[1] Although Golgi cell recordings suggest a low mean firing of about 15 Hz, the cell can follow input current up to 300 Hz (Dieudonné, D'Angelo, personal unpublished observations).

statistical dependencies are assumed to create structure in the mossy fibers that will be sparse along some directions. These sparsest directions or most non-Gaussian directions in the mossy fibers input space, and therefore the mossy fiber-granule cell synaptic weights, can be found using independent component analysis (ICA) algorithms ([6]; [8]). By discovering such directions, the resulting representation in the granule cells should be sparse, distributed and nearly statistically independent, which should facilitate further processing in the molecular and Purkinje layer of the cerebellum.

To better illustrate the granular layer model, the mossy fibers inputs in the following are assumed to be encoding the pixels of natural images (fig. 2, left), although we know particular well that mossy fibers encode neural signals from many different areas of the brain going well beyond visual encoding alone. Natural images are used to provide a well-characterized statistical structure ([15]) in the mossy fibers to complete the illustration. In the simulations, the granule cells received between 4 and 64 mossy fibers inputs, organized in patches of different sizes, $\mathbf{2} \times \mathbf{2}$, $\mathbf{3} \times \mathbf{3}$, $\mathbf{4} \times \mathbf{4}$, $\mathbf{6} \times \mathbf{6}$ and $\mathbf{8} \times \mathbf{8}$ patches.

Figure 2: (Left) Images are used to illustrate the potential statistical structure present in mossy fiber inputs. The hypothesis is that the granular layer takes advantage of the structure present in mossy fiber inputs to construct a representation that is *lifetime* sparse (see text) and as statistically independent as possible. A pixel represents the firing rate activity of a mossy fiber. The white square illustrates the receptive field of a single granule cell. (Right) Learned receptive fields ($\mathbf{8} \times \mathbf{8}$ pixels) for 64 of the granule cells. Each small image shows the mossy fibers inputs that make the granule cell response maximally. These are essentially independent component analysis (ICA) basis for the natural images used as mossy fiber inputs.

In order to build the desired sparse and distributed representation at the granule cells, a conventional ICA algorithm was used for simplicity to learn the mossy fibers-granule cells synaptic weights [3] with the *natural gradient* improvement ([2]; [12]). Note that a biologically plausible implementation of the learning algorithm could also have been used ([5]; [6]). The basis obtained after learning over 16 000 patches of natural images are shown in (fig. 2, right).

The ICA algorithm was used to compute the weights $\widetilde{\boldsymbol{W}}$ of Eq. 2, which has no positivity constraint on the weights. With the equivalence established between Eq. 2 and Eq. 1, the positive restricted weights $\boldsymbol{W}$ and $\boldsymbol{\gamma}$ are computed. As expected, the results below were the same whether using Eq. 2 or Eq. 1. The results using directly a learning rule using the biological constraints of positivity of responses and synaptic weights will be presented elsewhere.

2.1.1 The role of direct mossy fiber inputs to the Golgi cell

The equivalence between Eq. 1 and Eq. 2 suggests that the role of the direct mossy fiber inputs to the Golgi cell is to permit a linear transformation between the mossy fibers and granule cells that is not restricted by the positivity constraint on these synapses. The weights $\boldsymbol{\gamma}$, which need to be learned at the same time as the $\boldsymbol{W}$, permit a richer encoding in the granule cell than would be possible otherwise.

2.2 Denoising: A new role for the Golgi cells

The noisy mossy fibers inputs are modeled as $\boldsymbol{X} + \boldsymbol{\epsilon}$, where $\boldsymbol{\epsilon}$ is gaussian with zero mean and variance $\boldsymbol{\sigma^2}$. Consistent with ICA, the granule cell activities, constrained to be positive, are modeled with a sparse and statistically independent prior density $\boldsymbol{f_s(s)} = \prod_i \boldsymbol{f_s(s_i)}$ where $\boldsymbol{f_s(s_i)}$ is the exponential density $\boldsymbol{f_s(s_i)} = \boldsymbol{\alpha} \exp(-\boldsymbol{\alpha} s_i) = \boldsymbol{\gamma_1(s_i)}$ and where $\boldsymbol{\gamma_1(\cdot)}$ is the gamma density of order 1 [2].

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

The expectation of the S given the inputs with noise, is $E[(S+\varepsilon)^+] = \frac{1}{\alpha} + \frac{\alpha}{2}\,\varsigma^2 + \vartheta(\varsigma^4)$, where ε is the noise as seen by the granule cell, $(x)^+$ indicates positive values only, and ς^2 is the variance of the noise ε. From Eq. 2, we see that $\varepsilon = \widetilde{W}\,\epsilon$, so that its variance $\varsigma^2 = E[\varepsilon\varepsilon^T] = \widetilde{W}\,\sigma I\,\widetilde{W}^T$.

The optimal denoising for sparse variables with an exponential distribution $f_s(s_i) = \alpha \exp(-\alpha s_i)$ is given by $\hat{s}_i = s_i - \alpha\,\sigma$, where σ is the input noise variance ([9]). In fact, for sparse variables, the sparser S is, the better denoising works. In ICA, we made the approximation that the granule cells probability density factorizes, this is equivalent to considering the components s_i separately for denoising ([9]).

Taking $\mu = 1$, the Golgi activity is given by the sum of Gcs activity: $\bar{z} = \sum_{i=1}^{N} s_i - \theta$, where the number of Gcs N is about 1000-6000 [14]. With no noise in Mfs, the Go density turns out to be a gamma density of order N, $f_{\bar{z}}(\bar{z}) = \gamma_N(\bar{z})$ and for large N, $N >\sim 200$, the gamma density approaches a Gaussian density with a mean $\mu(\bar{z}) = N/\alpha$ and a variance $\sigma^2 = N/\alpha^2$. With noisy inputs, $X + \epsilon$, using $E[(S+\varepsilon)^+]$, the Go expectation is $E[\bar{z}] = N\,E[(S+\varepsilon)^+] - \theta = \frac{N}{\alpha} + \frac{N\alpha}{2}\,\varsigma^2 + \vartheta(\varsigma^4) - \theta$, assuming same noise variance and same exponential density for all s_i. Therefore the inhibition required for optimal denoising $\alpha\,\sigma$ is estimated for each image inputs by $\widetilde{W}\,\widetilde{W}^T\,\frac{2}{N}\,\bar{z}$, with $\theta = \frac{N}{\alpha}$ in computing $\bar{z}$.

Coupling between excitatory and inhibitory synaptic weights As seen above, because the input $X + \epsilon$ is multiplied by $\widetilde{W}$, the noise variance is amplified as $\widetilde{W}\sigma I\widetilde{W}^T$. To reduce the activity of the Gc accordingly, one choice is to set v in Eq. 2 appropriately. Since the inhibition is summed in the Gc, we have $v_j = \sum V_{ji}$, where V_{ji} has the same dimensions as W_{ji} and opposed to them (see fig. 1). Therefore one solution is to set $V_{ji} = W_{ji}^2$, so that the total inhibition to the Gc s_j is optimal.

3 Results

Encoding by the granule cells of noisy mossy fibers inputs for different level of Golgi cell inhibition is shown in fig. 3. As inhibition level increases, the number of Gcs active decreases, whereas the image quality degrades gracefully. The optimal denoising was computed using the known noise variance injected in the mossy fibers inputs. The receptive field of the Gcs were 6×6 Mfs pixels. Similar results were obtained for more realistic Gcs receptive field sizes (2×2 or 3×3). The results are shown for non-overlapping Gcs receptive fields to encode the image. The patches observed with the highest level of inhibition are reduced or disappear when overlapping receptive fields are used. Similar results (fig. 4) were obtained as well when the Golgi cell was used to estimate the noise variance as described above. The cell summed over the responses of 6000 Gcs, and where α = 0.13. In fig. 5, the effects of the Golgi cell inhibition on different representations in the Gcs are compared.

4 Discussion and Conclusion

Denoising with sparse variables is illustrated with the following arguments. Since the density of a super-gaussian random variable, like the exponential random variables that we used, has a sharp peak at zero, it can be assumed that small values of S correspond in fact to pure noise, i.e. that the true value should be $S = 0$. Thresholding such values to zero thus reduce noise ([9]).

Many cerebellar models have used random weights for the mossy fiber-granule cell synapses with arbitrary Golgi cell inhibition to determine the number of granule cells that should remain active ([11]). Fig. 5 shows that this strategy leads systematically to information loss in the encoding of the parallel fibers. Other models have suggested to use a representation that decorrelates the mossy fibers inputs, (this is equivalent to performing principal component analysis (PCA)) ([4]; [10]; [16]). The retained information is high, but this encoding is not lifetime sparse as the number of granule cells active in the subset shown is at its maximum most of the time. Willmore and Tolhurst reached the same conclusion comparing different neural codes: PCA minimizes *population* sparseness and not *lifetime* sparseness ([17]). For lifetime learning and minimizing interference across tasks, the desired property for the Gcs representation is *lifetime* sparse and not *population* sparse.

We suggested that one role of the Golgi cell (Go) accomplished by summing over all parallel fibers is to construct an estimate of the noise variance in the mossy fibers (Mfs) inputs. To construct this estimate, we assumed that all Mfs that send their inputs to the Go via the parallel fibers had the same noise. This is a strong approximation, but one that may be realistic considering that the Gos have a limited Mfs receptive field and that they may overlap. This suggests that they are as many noise estimates as Gos. Moreover, if they are independent Mfs noise contributions to one Gc, which would be estimated by multiple Go inputs, the current model still holds, since the variance of independent variables adds, e.g. Var$[x + y]$ = Var$[x]$ +

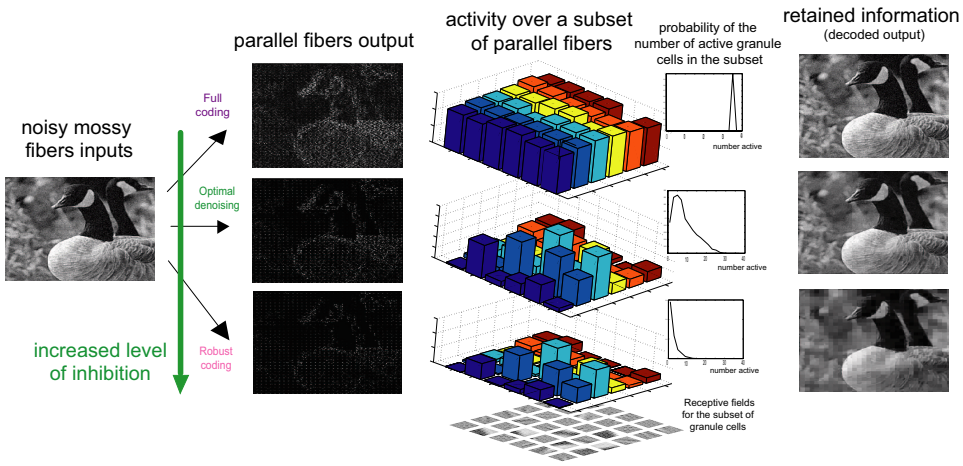

Figure 3: Encoding by the granule cells of noisy mossy fibers inputs for different level of Golgi cell inhibition. The same set of 72 granule cells is used for coding the entire image by patches of $\mathbf{6} \times \mathbf{6}$ pixels. In total, it is equivalent to using 80000 granule cells to encode the entire image, whose output is shown under *parallel fibers output*. The average activity of the granule cells over all patches is shown in the next column, where as the probability for a given number of granule cells to be active for an image patch is shown next. The rightmost column shows the decoded output using all 80000 granule cells. The optimal denoising is obtained when the level of inhibition from the Golgi cell is equal to the variance of the noise injected in the image. *Robust coding* demonstrates that the coding is robust to higher level of inhibition and that the image quality degrades gracefully using this encoding scheme.

Var$[y]$. Therefore, for optimal noise reduction, one simply needs to sum the inhibition from multiple Golgi cells at the granule cell, a likely scenario.

5 Acknowledgements

We would like to thank David Eagleman for contributing to fig. 1 and 2. This research was funded in part by the EC contract IST-2001-35271 Project SpikeFORCE (www.spikeforce.org).

References

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

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

[3] 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.

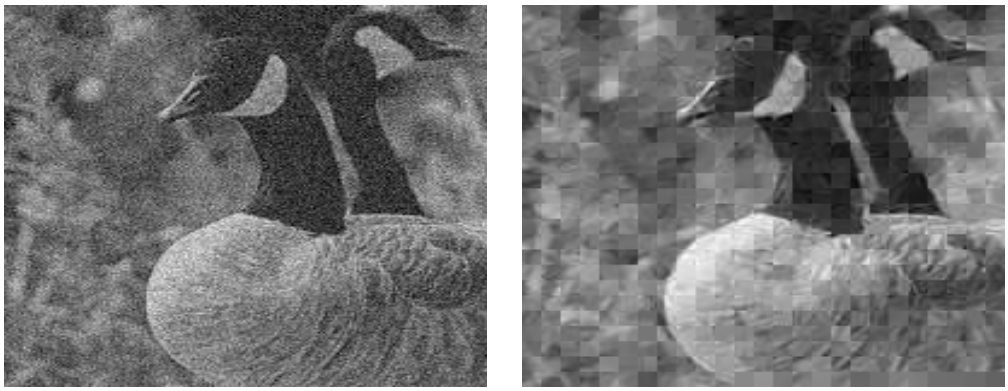

Figure 4: Denoising by Golgi cell inhibition. The Golgi cell computes an estimate of the noise variance by summing over 6000 parallel fibers (see text). (Left) The noisy mossy fibers inputs, using gaussian noise with a standard deviation of 0.5 of the original image' standard deviation. (Right) Denoised image, reconstructed from parallel fibers activity.

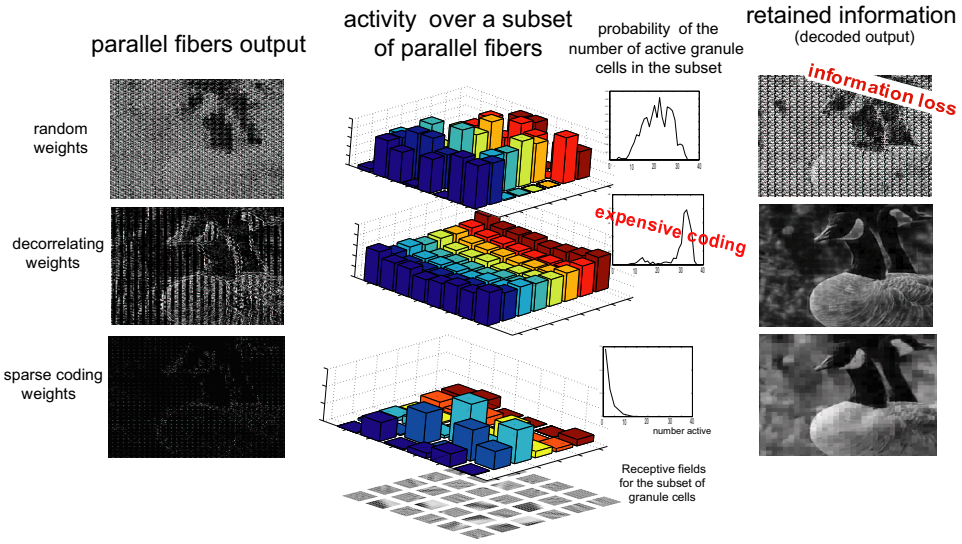

Figure 5: Comparison of different encoding strategy by the granule cells. The first row illustrates the resulting encoding and decoding when random synaptic weights are used between mossy fibers and granule cells with arbitrary level of Golgi cell inhibition. Note the dramatic loss in information in the decoded output. The second row uses weights computed to obtain the most decorrelated granule cell outputs (equivalent to principal component analysis). The retained information is high, but this encoding is not sparse as the number of granule cells active in the subset shown is at its maximum most of the time. Finally, the last row of the previous figure is repeated here for direct comparison with our sparse coding strategy. For this high level of inhibition, the decoded output is more coarse than with the decorrelated weights, but the strategy only uses 4-5 active cells compared to 32. Note that the level of inhibition remained at a similar level for all encoding strategy for comparison.

[4] G. Chauvet. Habituation rules for a theory of the cerebellar cortex. *Biological Cybernetics*, 55:201–209, 1986.

[5] 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.

[6] O. J.-M. D. Coenen, M. Arnold, T. Sejnowski, and M. Jabri. Parallel fiber coding in the cerebellum for life-long learning. *Autonomous Robots*, 11(3):291–7, 2001. Kluwer Academic Publishers, Netherlands.

[7] E. D'Angelo, G. De Filippi, P. Rossi, and V. Taglietti. Synaptic excitation of individual rat cerebellar granule cells in situ: Evidence for the role of NMDA receptors. *J Physiol (Lond)*, 484 (Pt 2):397–413, Apr 15 1995.

[8] 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*, 9(6):1495, 1998.

[9] A. Hyvärinen. Sparse code shrinkage: Denoising of nongaussian data by maximum likelihood estimation. *Neural Computation*, 11:1739–1768, 1999.

[10] 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.

[11] 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.

[12] D. J. C. MacKay. Maximum likelihood and covariant algorithms for independent component analysis. Unpublished manuscript, available at http://wol.ra.phy.cam.ac.uk/mackay/BayesICA.html, 1996.

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

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

[15] P. Reinagel and S. Laughlin, editors. *Network: Computation in Neural Systems. Special issue featuring selected papers from Natural Stimulus Statistics Workshop 2000 (October 2000), Cold Spring Harbor, USA*, volume 12. Institute *of* Physics, August 2001.

[16] N. Schweighofer, K. Doya, and L. F. Unsupervised learning of granule cell sparse codes enhances cerebellar adaptive control. *Neuroscience*, 103(1):35–50, 2001.

[17] B. Willmore and D. J. Tolhurst. Characterizing the sparsness of neural codes. *Network: Computation in neural systems*, 12(3):255–270, August 2001.