A Model of Cerebellar Metaplasticity

Nicolas Schweighofer¹ and Michael A. Arbib

¹USC Brain Project

University of Southern California

Los Angeles, California 90089-2520

Abstract

The term “learning rule” in neural network theory usually refers to a rule for the plasticity of a given synapse, whereas *metaplasticity* involves a “metalearning algorithm” describing higher level control mechanisms for apportioning plasticity across a population of synapses. We propose here that the cerebellar cortex may use metaplasticity, and we demonstrate this by introducing the Cerebellar Adaptive Rate Learning (CARL) algorithm that concentrates learning on those Purkinje cell synapses whose adaptation is most relevant to learning an overall pattern. Our results show that this biologically plausible metalearning algorithm not only improves significantly the learning capability of the cerebellum but is very robust. Finally, we identify several putative neurochemicals that could be involved in a cascade of events leading to adaptive learning rates in Purkinje cell synapses.

Introduction

Metaplasticity is a form of “controlled plasticity” that augments the local synaptic learning rules by a “metalearning algorithm” that describes higher level control mechanisms for apportioning plasticity across a population of synapses¹. It has been described recently in neural network research (Jacobs 1988; Sutton 1992) and in hippocampal learning by Abraham and Bear (1996) who assert that “metaplasticity has occurred if prior synaptic plasticity or cellular activity (or inactivity) leads to a persistent change in the direction or degree of synaptic plasticity elicited by a given pattern of synaptic activation.” In the present paper,

we propose a model of cerebellar Purkinje cell metaplasticity and argue that the cerebellum has much to gain from “learning how to learn.”

Marr (1969) and Albus (1971) hypothesized that the cerebellar cortex is an array of perceptrons, each being a Purkinje cell, with parallel fibers providing the context in which the movements are made and the climbing fiber giving the error signal necessary for modifying each parallel fiber → Purkinje cell synapse. Over the last three decades, evidence has been mounting that corroborates a form of the synaptic plasticity they postulated (for review, see for instance Crepel et al. 1996). Moreover, it has been suggested (Kawato et al. 1987) that the cerebellum acquires internal neural models of the motor system. Given the large number of degrees of freedom and the pervasive nonlinearities of the physical systems, these internal models are extremely complex neural representations of the systems they control (Kawato et al. 1987). Thus, if assemblies of cerebellar Purkinje cells are to learn inverse models, the cells need very efficient learning capabilities (learning at the cellular level).

Theoretical considerations emphasize three important factors that could limit the learning performance (assessed at the system level of effective motor control): First, the *learning rates* of the parallel → Purkinje synapses (i.e., the rate of change in synaptic efficacy) should be adequate to allow fast learning at the system level to respond rapidly to changes in the controlled system or in the environment. However, overly large learning rates are not desirable in adaptive neural networks as they can induce oscillations in the patterning of synaptic weights and even divergence from the de-

¹We may use the term *learning* to refer to adaptive improvements at the level of an overall organism or neural network and the term *synaptic plasticity* to refer to mechanisms for changing efficacy at the synaptic level. However, it has become accepted in neural network theory to use the term “learning rule” to refer to a rule for changing the weight of a given synapse, at the risk of some confusion as to the level of “learning” involved.

¹Corresponding author. Present address: ERATO, Kawato Dynamic Brain Project, Japan Science and Technology Corporation, Seika-cho, Soraku-gun, Kyoto 619-02 Japan.

sired results at the system level (Hertz et al. 1991). Conversely, excessively small learning rates slow down the system's learning. Because it is doubtful that the optimal learning rates are genetically coded, there should exist some self-tuning properties so that "good" learning rates can be found. Second, the very large number of synapses per Purkinje cell (on the order of 200,000 synapses in humans) would induce what is called "overfitting" in the artificial neural network literature: The number of free parameters (the modifiable synapses) exceeds by far the desired number of parameters required for the learning of a specific problem. A curve fit by too many parameters follows all the small details or noise but is very poor for interpolation and extrapolation. Hence, to possess good generalization properties, the cerebellum would greatly benefit from a process that would reduce the number of potentially modifiable synapses. Third, the brain in general—and the cerebellum in particular—is "wired" overabundantly to allow for the learning of many possible combinations. Thus, it is probable that a majority of parallel fibers does not have information relevant to a particular Purkinje cell. Learning about irrelevant inputs acts as noise, interfering with learning about relevant inputs (Sutton 1992). Inputs that are likely to be irrelevant should be given small or null learning rates, whereas inputs that are likely to be relevant should be given large learning rates.

Jacobs (1988) proposed a neural model in which each synapse has its own adaptive learning rate. The number of free parameters is reduced by "freezing" the synaptic weights when their variation is not needed: Inputs to a cell that are likely to be irrelevant in a given task are given small learning rates, whereas inputs that are likely to be relevant are given large learning rates. Sutton (1992) proposed a newer version of Jacob's algorithm for incremental learning, the Incremental Delta-Bar-Delta algorithm (IDBD). The key feature of the metaplasticity rules considered in this paper is that the learning rate for a given synapse is proportional to the temporal correlation between current weight change and recent weight changes. We first review the way this is obtained in the IDBD rule. We then show how CARL (Cerebellar Adaptive Rates Learning), a modified version of IDBD, can be implemented in a biologically plausible way, that is, by a cascade of putative neurochemicals in the Purkinje cell synapse. Our simulation results show that CARL performs considerably better than a simple perceptron with a fixed learning rate and

is very robust to the choice of the parameters. Finally, we propose neurochemical mechanisms that could indeed implement adaptive learning rates in the living cerebellum.

THE IDBD ALGORITHM

The basic idea of the IDBD algorithm is that if the current weight change is positively correlated with past weight changes, this indicates that the past weight changes should have been larger, and thus the learning rate can be increased; it should be decreased otherwise. In Sutton's IDBD, the learning system is a simple one cell linear perceptron, with base learning rule the delta rule [or "least mean square (LMS)" rule]. At each instant, the synaptic inputs are weighted by the synaptic efficacies to give the cell's "firing rate" y :

$$y(t) = \sum_{i=1}^n w_i(t) x_i(t) \quad (1)$$

The weight change rule is given by

$$\Delta w_i(t) = \alpha_i(t+1) x_i(t) \delta(t) \quad (2)$$

where α_i is an adjustable "learning rate" and the error between the desired output y^* and the real output y is

$$\delta(t) = y^*(t) - y(t) \quad (3)$$

Note that if the value of the learning rate were the same for all synapses and constant over time, equation 2 reduces to the simple delta rule: The weights are modified whenever the synaptic input and the error channel are activated simultaneously. Sutton (1992) introduces metaplasticity as follows: The learning rates are of the form

$$\alpha_i(t) = e^{\beta_i(t)} \quad (4)$$

The exponential is not needed in theory but speeds up learning and forces the α_i to always be positive. The β_i vary with time and are updated according to

$$\beta_i(t+1) = \beta_i(t) + \theta x_i(t) \delta(t) h_i(t) \quad (5)$$

where θ is the *metalearning rate* and h_i is an additional per-input memory variable:

MODEL OF CEREBELLAR METAPLASTICITY

$$b_i(t+1) = b_i(t)[1 - \alpha_i(t+1)x_i^2(t)]^+ + \alpha_i(t+1)x_i(t)\delta(t) \quad (6)$$

where $[x]^+$ is x if $x > 0$, else 0. Because the first term on the right-hand side of equation 6 is a decay term [the term $\alpha_i(t+1)x_i^2(t)$ is normally zero or a positive fraction] and the second term is the last weight change (see equation 2), $b_i(t)$ is a decaying trace of the cumulative sum of recent changes of w_i .² Thus, the overall changes in β_i are proportional to the temporal correlation between the current weight change $x_i(t)\delta(t)$ and recent weight changes $b_i(t)$. When this correlation is consistently positive for the i th synapse, the learning rate α_i converges toward a close to optimal, positive value and toward zero otherwise (Sutton 1992). Simulations show that IDBD gives very good learning results compared with the delta rule with a fixed learning rate, at least in a simple tracking task (Sutton 1992).

CARL

We now relate this to the plasticity of parallel fiber $\rightarrow$ Purkinje cell synapses and propose CARL. There is evidence that the climbing fibers convey signals encoding error in the performance of the system in which the cerebellar subsystem is installed (Ito 1984). A climbing fiber signal depresses the synaptic efficacies of those parallel fiber $\rightarrow$ Purkinje cell synapses that were activated in conjunction with the climbing fiber (Ito et al. 1982). It has been suggested that this long-term depression (LTD) is expressed by a phosphorylation of the ionotropic glutamate AMPA receptors (Daniel et al. 1992). Requirements for LTD are a rise of postsynaptic Ca^{2+} induced by climbing fiber action with concurrent activation of metabotropic glutamate receptors by parallel fiber action (Linden et al. 1991; Daniel et al. 1992).

Assuming that the Purkinje cells act like one-cell perceptrons learning incrementally, we propose that a modified version of IDBD could be implemented at the parallel fibers $\rightarrow$ Purkinje cell synapses. With the following correspondences, equation 1 gives the firing rate potential of a linearized Purkinje cell, and equation 2 models cer-

ebellar learning (LTD could be accounted for by a minus sign):

$$x_i \Leftrightarrow GC_i; \quad y \Leftrightarrow PC; \quad \delta \Leftrightarrow IO$$

where GC_i , PC , and IO represent the activities of, respectively, the i th granule cell, the Purkinje cell, and the inferior olive. Thus, the Purkinje cell activity is simply the weighted sum of the granule cell (parallel fiber) inputs, and the synapses are modified by concurrent climbing fiber and parallel fiber activities.³ Note that if the learning rate was the same for all the synapses, the learning rule would simply reduce to the LTD portion of the learning rule proposed by Albus; the original contribution of the present work is the proposal that each Purkinje cell synapse can be modified at its own and (as we will see below) near optimal rate. Moreover, we show that this “learning” of the learning rate can be achieved in a biologically plausible manner.

As described above, the key feature of IDBD is that the change in learning rate is proportional to the temporal correlation between current weight change and recent weight changes. Because b_i (in equation 5) is a memory of the weight changes, its decay should be significantly slower than a cell membrane time constant, which is of the order of 10 msec. In IDBD, β_i is described as an integrator but could also be represented by a leaky integrator with a time constant much longer than the time constant of b_i . We thus propose that β_i and b_i model second-messenger concentrations, which have significantly larger time constants than electrical activities. As we shall see below, these time constants are on the order of hours for β_i and on the order of a second for b_i .

Equations 4, 5, and 6 were derived from optimization theory, not according to biological plausibility. There is no reason to believe that chemical processes in the synapse could implement the exponential function of equation 4 or the nonlinear, activity-dependent decay factor $[1 - \alpha_i(t+1)x_i^2(t)]^+$ in equation 6. Instead, provided that the diffusional delays of the neurochemicals in the synapses [in the order of 10–20 msec (Fiala et al. 1996)] are negligible compared with the time constants of the

²These forms of β_i and b_i were derived from a gradient descent analysis [see Sutton (1992) for more details].

³For the purpose of comparing the responses of CARL and IDBD, we assume here that the PC response is instantaneous, i.e., there are no dynamics in the equation giving the PC firing as a function of its inputs. A more detailed model of the Purkinje cell—including compartmental modeling—would be fully compatible with CARL.

chemical reactions, second-messenger concentrations can be modeled with first order kinetics equations (Fiala et al. 1996). Thus, we rewrite equation 5 and 6 as follow:

$$\frac{d\beta_i}{dt} = -\frac{\beta_i}{\tau_1} + k_1 (\beta_{\max} - \beta_i) GC_i IO b_i \quad (7)$$

$$\frac{db_i}{dt} = -\frac{b_i}{\tau_2} + k_2 (b_{\max} - b_i) GC_i IO \quad (8)$$

where k_1 and k_2 are input gains, $\beta_{\max}$ and $b_{\max}$ maximal concentrations, and τ_1 and τ_2 the time constants of the two second messengers. In CARL the adaptive learning rate is taken simply as β_i (and not as the exponential of β_i). Weight changes occur in those parallel fiber $\rightarrow$ Purkinje cell synapses activated in conjunction with the climbing fiber:

$$\frac{dw_i}{dt} = \beta_i GC_i IO \quad (9)$$

Equations 9 and 8 show that b_i is a decaying trace (with time constant τ_2) of the cumulative sum of recent changes of w_i . Moreover, as shown by the last term of equation 7, the overall change in learning rate β_i is proportional to the temporal correlation between current weight change $GC_i IO$ and recent weight changes b_i .

SIMULATIONS

For our purpose (namely showing that the advantages of adaptive learning rates can be obtained with a neurochemically plausible learning rule), we require only that CARL has performance comparable with that of IDBD and significantly better than the delta learning rule (i.e., learning without metaplasticity). Sutton (1992) assessed the capabilities of the IDBD by using a series of tracking tasks, supervised learning tasks in which the target drifts over time and must be tracked.⁴ Here we use a very similar task, which involves 20 real-valued inputs and one output. The inputs are chosen independently and randomly according to a normal distribution with zero mean and unit variance. The

⁴A convenient, stationary task could also be used, but nonstationary tasks are more appropriate because they show the ability of the IDBD algorithm to generalize, i.e., to use previous learning in future, related tasks.

Figure 1: Comparison of the best average asymptotic performances of LMS, IDBD, and CARL. (Solid line bars) The neuron with 20 synapses; (broken line) the neuron with 500 synapses. In the latter case, the LMS learning rule cannot decrease the error, but IDBD and CARL do almost as well as in the former case.

target is the sum of the first five inputs, each multiplied either by +1 or -1, that is,

$$y^* = s_1 x_1 + s_2 x_2 + s_3 x_3 + s_4 x_4 + s_5 x_5 + 0x_6 + \dots + 0x_{20} \quad (10)$$

where all the s_i are either +1 or -1.⁵ To make it a tracking problem, after every 100 examples one of the five s_i is selected and switched in sign. We chose a time step of 10 msec; thus, the target changes every second. If the metalearning algorithm can identify which inputs are relevant, then it should be able to track the drifting target function more accurately than ordinary delta rule. One long run of 100,000 sec is performed, and on another 1000 sec the average asymptotic error is computed. The integration method is the Euler method, the program is written in C, and simulations are run on a DEC alpha 500.

The solid line bars of Figure 1 show the best performance obtained with the three algorithms in the task described above. With its optimal learning rate (learning rate = 0.05), the LMS (delta) rule attains a mean square error (MSE) of 0.89. IDBD and CARL perform much better: IDBD has an MSE of 0.27 with the metalearning rate $\theta = 0.01$, and CARL

⁵In this simple model there are no sign constraints for the weights and neural activities; in the real cerebellum these quantities can vary around positive means.

MODEL OF CEREBELLAR METAPLASTICITY

has an MSE of 0.29 with the following parameter values: $k_1 = 5$, $\beta_{\max} = 10$, $\tau_1 = 10,000$ sec, $k_2 = 0.1$, $b_{\max} = 0.01$, and $\tau_2 = 0.32$ sec.

Biological neurons have many synapses and receive many irrelevant, noisy inputs. Consequently, we now assess the performance of CARL with a large number of modifiable synapses, that is, 500 synapses instead of 20 as formerly. The tracking task is of the same nature (only five inputs are relevant). Because the number of relevant inputs has decreased from 25% to 1%, the learning task is very arduous (such a low percentage of relevant inputs is likely to be unrealistic in the real cerebellum, but it allows us here to show the capacity of CARL). As the broken line bars of Figure 1 show, the simple LMS is unable to decrease the error significantly for any learning rate (MSE = 5.06). Both IDBD and CARL, however, reduces the error significantly: Their respective MSE results are 0.28 and 0.31 (same parameter set as above, except for LMS learning rate = 0.001).

To exhibit biological plausibility, CARL must be robust in the choice of the parameters. In Figure 2, we show CARL's performance as a function of the $\log_{10}$ of the two time constants (the cell has now 20 synapses again). Good performances are obtained for a large range of values, that is, $\tau_1 > 1000$ sec and $\tau_2 < 1$ sec. Similarly, the performance of CARL should not be too affected by the choice of the gains k_1 and k_2 . In equation 7, k_1 is the rate of adjusting of the learning rate and corresponds to the metalearning rate in IDBD. If the range of values of k_1 yielding good performances is

Figure 2: Average asymptotic error for CARL over wide ranges of the two time constants τ_1 and τ_2 (logarithmic plot). Note the robustness of the algorithm for $\tau_1 > 1000$ sec and $\tau_2 < 1$ sec.

Figure 3: Average asymptotic error for the LMS learning rule over a range of learning rates (a), and CARL over wide ranges of the two input gains k_1 (b) and k_2 (c) (logarithmic plot). Note the robustness of CARL compared with the LMS learning rule.

narrow, CARL would not solve the problem of finding the “good” learning rate, but instead the problem would be similar to the LMS, merely displaced an order higher. On the contrary if the learning performances do not depend much on k_1 , then a very rough value of k_1 could be inherited. Figure 3, b and c, shows that the values of k_1 and k_2 for which good learning occurs belong to very large ranges (note that we plotted the MSE as a function of the $\log_{10}$ of the gains), all the more when compared with the narrow range of optimal learning rates for the LMS shown in Figure 3a. Thus, there is no need for a metalearning process of higher order (i.e., the learning of k_1 and k_2).

In Figure 4, we show how cellular activity leads to a persistent change in the degree of synaptic plasticity. We plot the time course of the adaptive learning rates β_i for CARL for one relevant and one irrelevant input (an input is relevant if it has a nonzero weight in the target, equation 10) as a function of time. The learning rates for irrelevant synapses converge to values close to zero, as desired. The learning rate of the relevant inputs all converge toward ~ 0.14 . To see if this is an optimal value, we use an LMS rule where the irrelevant inputs are given zero learning rates and the relevant inputs are given fixed learning rates between 0.05 and 0.25. Simulations show that best performance is obtained for learning rates between 0.13 and 0.16. Thus CARL, as IDBD (see Sutton 1992), finds near optimal learning rates.

Discussion

In this paper we proposed a new model of metaplasticity in the cerebellar Purkinje cell. This model dramatically improves learning performance

Figure 4: Time course of two learning rates under CARL for one relevant and one irrelevant input. For the relevant input, the corresponding β_i quickly climbs to a large value to ensure responsiveness to changes; for the irrelevant input, the corresponding β_i stays near zero.

by automatically setting near optimal synaptic learning rates (which can be zero), in order to deal with the problem of input relevance or nonrelevance in regard to a certain goal.

As a concrete example, we can relate metaplasticity to the role of the Purkinje cells in multi-joint movements. The cerebellum is known to provide the motor commands necessary to compensate for the interaction forces occurring during fast reaching movements (Bastian et al. 1996). For a Purkinje cell to compute a precise motor command for a joint, convergence of kinematic information from other joints is necessary (Schweighofer et al. 1998a,b). For a shoulder-related Purkinje cell for instance, kinematic information about the elbow joint is crucial, but information about the wrist is less relevant, and proprioceptive information about the thumb is probably irrelevant because its mass is negligible. However, information about many joints reaches the Purkinje cell, because each parallel fiber spans the cerebellar cortex over a long distance (about 6 mm long in the monkey; Mugnaini 1983) and links many cerebellar functional units (or “cerebellar microzones,” each having a width of ~200 μm ; Oscarsson 1980). The degree of input relevance, however, cannot be available a priori for the shoulder Purkinje cell. We propose that the temporal correlation between (1) the error signal (carried by the climbing fiber) resulting from a noncorrect shoulder movement and (2) the kinematic information from a specific joint (brought about by the parallel

fiber synaptic input) will not only determine the value of the synaptic efficacy (cerebellar plasticity, i.e., LTD) but also determine the optimal learning rate of the synapse, allowing efficient learning of the shoulder motor command. Thus, metaplasticity would allow the cerebellum to be a neural network with powerful learning capabilities, as required for the acquisition of complex internal inverse neural models.

Two other learning algorithms are relevant to the present work. First, the basic LMS rule can be modified by adding a “momentum” term (Rumelhart et al. 1986) that can greatly improve the speed of learning. The idea is to give each weight some inertia, or momentum, so that the weight changes direction on average and not with each little kick. The momentum effectively adjusts the weight change as a function of past experience, as IDBD or CARL does. Moreover, the LMS rule including a momentum term could be implemented by cellular neurochemistry as readily as CARL. However, the augmented momentum rule does not directly deal with irrelevant inputs, and the problem of the appropriate choice of the parameters (i.e., the learning rate and the momentum parameter) is even more difficult than in the simple LMS rule. It has been shown that, for the LMS, the use of the momentum reduces the stable range of the learning rate parameter and thus could lead to instability if the learning rate is not adjusted appropriately. Moreover, the misadjustment increases with increasing learning rate (Roy and Shynk 1990). The second method that is related to CARL is the exponentiated gradient (EG) method of Littlestone (1988). This algorithm addresses the issue of irrelevant inputs, albeit with a different mechanism than CARL, and simulation results show that it outperforms the LMS when only a few inputs are relevant. The original EG rule might be difficult to implement in a biologically plausible manner because of the use of the exponential function at each synapse and the need for equal distribution of a global quantity that depends on the sum of the weight changes over the cell. The approximated EG (Kivinen and Warmuth 1994) appears more biologically plausible because it has the benefit to avoid the use of the exponential functions and has a much simpler form. However, the approximated EG cannot be implemented in the Purkinje cell because it requires computation of the correlation between the difference of the target signal y^* with the local synaptic input x_i and the error signal δ . However, as discussed above, the Purkinje cell syn-

MODEL OF CEREBELLAR METAPLASTICITY

apses are only thought to be able to compute the correlation between the error signal δ (carried by the climbing fiber) and the local synaptic input x_i (from the parallel fibers).

Computer models represent one possible solution to a given problem. To provide generalized significance, computer models have to be experimentally validated. Although the metalearning algorithm we developed is biologically plausible, the simulation protocol we used is not directly based on neuroscience experimental data; but because our simulations were computationally tractable and because we did not need to make unnecessary assumptions, we could concentrate on the learning algorithm per se and study it in detail. We can now make several testable predictions that may lead to future progress in the empirical studies of cerebellar plasticity.

Several experiments show that synaptic plasticity in the hippocampus can be dramatically modulated by prior synaptic activity (see Abraham and Bear 1996). Similarly, we predict that the rate at which cerebellar LTD occurs can be greatly modulated by prior conjoint stimulation of the parallel fiber and climbing fiber pathways. We further predict that in the cerebellum, metaplasticity and synaptic modification are induced simultaneously by the same synaptic activity, as shown by equations 8 and 9. But what would be the cascade of events in the Purkinje cell responsible for a putative mechanism underlying cerebellar metaplasticity? and what would be a minimal model describing it?

Our simulation results suggest that cerebellar LTD could be modulated by different neurochemical concentrations at the synapse. In CARL, b_i is a second messenger whose concentration depends on the concurrent activation of both the inferior olive and the parallel fiber and whose time constant is 1 sec or less. Possible candidates include high levels of Ca^{2+} arising from both external sources after climbing fiber action and internal stores after PF activity (Llano et al. 1991) and protein kinase C (Crepel and Krupa 1988). Because β_i must have a very long half-life for the system to have good performance (on the order of an hour or more), proteins would be good candidates. In the hippocampus, induction of long-term potentiation is followed by a complex pattern of changes in protein synthesis. Facilitation of activation of the protein calpain is associated with a greater degree of synaptic potentiation (Muller et al. 1995). Calpain is involved in the regulation of glutamatergic

synapses (Bi et al. 1994), and calpain activation has a slow onset (1–4 hr) that lasts for several days after stimulation (Bi et al. 1996). Because the isoform calpain II has recently been shown to exist in relatively high quantity in the Purkinje cells (Li et al. 1996), we suggest that this form of calpain is one possible candidate for a molecule that influences the Purkinje cell learning rates.

Acknowledgments

This research was supported by grant N00014-92-J-4026 from the Office of Naval Research for research on "Cerebellum and the Adaptive Coordination of Movement" and by J.S.T. We are grateful to S. Schaal for raising the issues of overfitting and metalearning and to G. Tocco, M. Kawato, and K. Doya for useful comments. Constructive criticism on an earlier draft was kindly provided by M. Baudry, R. Simantov, F. Pollick, and M. Tiede.

The publication costs of this article were defrayed in part by payment of page charges. This article must therefore be hereby marked "advertisement" in accordance with 18 USC section 1734 solely to indicate this fact.

References

- Abraham, W.C. and M.F. Bear. 1996. Metaplasticity—the plasticity of synaptic plasticity. *Trends Neurosci.* **19**: 126–130.
- Albus, J.S. 1971. The theory of cerebellar function. *Math. Biosci.* **10**: 25–61.
- Bastian, A.J., T.A. Martin, J.G. Keating, and W.T. Thach. 1996. Cerebellar ataxia: Abnormal control of interaction torques across joints. *J. Neurophysiol.* **76**: 492–509.
- Bi, X.N., G. Tocco, and M. Baudry. 1994. Calpain-mediated regulation of AMPA receptors in adult rats brain. *NeuroReport* 661–664.
- Bi, X., V. Chang, R. Siman, G. Tocco, and M. Baudry. 1996. Regional distribution and time-course of calpain activation following kainate-induced seizure activity in adult rat brain. *Brain Res.* **726**: 98–108.
- Crepel, F.C. and M. Krupa. 1988. Activation of protein kinase C induces a long-term depression of glutamate sensitivity of cerebellar Purkinje cells. An in vitro study. *Brain Res.* **458**: 397–401.
- Crepel, F., N. Hemart, D. Jaillard, and H. Daniel. 1996. Cellular mechanisms of long term depression in the cerebellum. *Behav. Brain Sci.* **19**: 347–353.
- Daniel, H., N. Hemart, D. Jaillard, and F. Crepel. 1992. Coactivation of metabotropic glutamate receptors and of voltage gated calcium channels induces long-term depression in cerebellar Purkinje cells. *Exp. Brain Res.* **90**: 327–331.
- Fiala J.C., S. Grossberg, and D. Bullock. 1996. Metabotropic

Schweighofer and Arbib

glutamate activation in cerebellar Purkinje cell as substrate for adaptive timing of the classically conditioned eye-blink response. *J. Neurosci.* **16**: 3760–3774.

Hertz J., A. Krogh, and R. Palmer. 1991. *Introduction to the theory of neural computation*. Addison-Wesley, Redwood City, CA.

Ito, M. 1984. *The cerebellum and neuronal control*. Raven Press, New York, NY.

Ito M., M. Sakurai, and P. Tongroach. 1982. Climbing fiber induced long term depression of both mossy fiber responsiveness and glutamate sensitivity of cerebellar Purkinje cells. *J. Physiol.* **324**: 113–134.

Jacobs, R. 1988. Increased rates of convergence through learning rate adaptation. *Neural Networks* **1**: 295–307.

Kawato, M., K. Furukawa, and R. Suzuki. 1987. A hierarchical neural network model for control and learning of voluntary movement. *Biol. Cybern.* **57**: 169–185.

Kivinen J. and M.K. Warmuth. 1994. *Exponentiated gradient versus gradient descent for linear predictors*. Tech. Rep. UCSC-CRL-94-16, University of California, Santa Cruz, CA.

Li, J., F. Grynspan, S. Berman, R. Nixon, and S. Bursztajn. 1996. Regional differences in gene expression activated neutral proteases (calpains) and their endogenous inhibitor calpastatin in mouse-brain and spinal cord. *J. Neurobiol.* **30**: 177–191.

Linden D.J., M. Dickinson, M. Smeyne, and J. Connor. 1991. A long term depression of AMPA currents in cultured Purkinje neurons. *Neuron* **11**: 1093–1100.

Littlestone, N. 1988. Learning quickly when irrelevant inputs abound: A new linear threshold algorithm. *Machine Learn.* **2**: 285–318.

Llano, I., J. Dressen, M. Kano, and A. Konnerth. 1991. Intradendritic release of calcium induced by glutamate in cerebellar Purkinje cells. *Neuron* **7**: 577–583.

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

Mugnaini, E. 1983. The length of cerebellar parallel fibers in chicken and rhesus monkey. *J. Comp. Neurol.* **220**: 7–15.

Muller D., I. Molinari, L. Soldati, and G. Bianchi. 1995. A genetic deficiency in calpastatin and isovaleryl carnitine treatment is associated with enhanced hippocampal long-term potentiation. *Synapse* **19**: 37–45.

Oscarsson, O. 1980. Functional organization of olivary projection to the cerebellar anterior lobe. In *The inferior olivary nucleus: Anatomy and physiology* (ed. J. Courville and Y. Lamarre), pp. 279–289. Raven Press, New York, NY.

Roy, S. and J.J. Shynk. 1990. Analysis of the momentum LMS

algorithm. *IEEE Trans. Acoustics, Speech, Signal Processing* **38**: 2088–2098.

Rumelhart, D.E., G.E. Hinton, and R.J. Williams. 1986. Learning representations by back-propagating errors. *Nature* **323**: 533–536.

Schweighofer N., M.A. Arbib, and M. Kawato. 1998a. Role of the cerebellum in reaching movements. I. Distributed inverse dynamics control. *Eur. J. Neurosci.* **10**: 86–94.

Schweighofer N., J. Spoelstra, M.A. Arbib, and M. Kawato. 1998b. Role of the cerebellum in reaching movements. II. A neural model of the intermediate cerebellum. *Eur. J. Neurosci.* **10**: 95–105.

Sutton, R. 1992. Adapting bias by gradient descent: An incremental version of the Delta-Bar-Delta. *Proceedings of the Tenth National Conference on Artificial Intelligence*, pp. 171–176. MIT Press, Cambridge, MA.

Received August 25, 1997; accepted in revised form December 1, 1997.

A model of cerebellar metaplasticity.

N Schweighofer and M A Arbib

Learn. Mem. 1998, **4**:

Access the most recent version at doi:[10.1101/lm.4.5.421](https://doi.org/10.1101/lm.4.5.421)

References

This article cites 22 articles, 1 of which can be accessed free at:
<http://learnmem.cshlp.org/content/4/5/421.full.html#ref-list-1>

License

Email Alerting Service

Receive free email alerts when new articles cite this article - sign up in the box at the top right corner of the article or [click here](#).