

TOPICAL REVIEW

Expansion coding and computation in the cerebellum: 50 years after the Marr–Albus codon theory

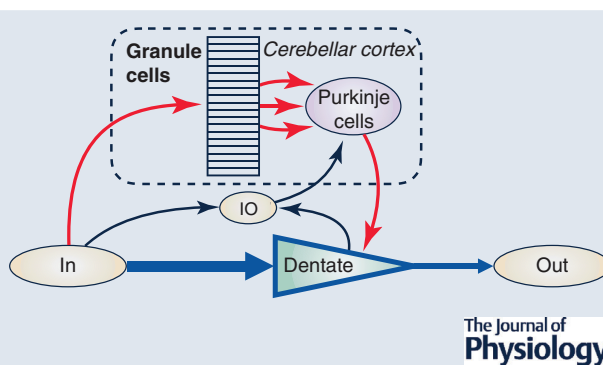
Terence D. Sanger¹ , Okito Yamashita^{2,3} and Mitsuo Kawato^{2,3}

¹Departments of Biomedical Engineering, Neurology, and Biokinesiology, University of Southern California, 1042 Downey Way, DRB 140, Los Angeles, CA, 90089, USA

²Brain Information Communication Research Laboratory Group, Advanced Telecommunications Research Institutes International (ATR), Hikaridai 2-2-2, 'Keihanna Science City', Kyoto, 619-0288, Japan

³Center for Advanced Intelligence Project (AIP), RIKEN, Nihonbashi 1-chome Mitsui Building, 15th floor, 1-4-1 Nihonbashi, Chuo-ku, Tokyo, 103-0027, Japan

Edited by: Ian Forsythe & Ole Paulsen



Abstract Fifty years ago, David Marr and James Albus proposed a computational model of cerebellar cortical function based on the pioneering circuit models described by John Eccles, Masao Ito and Janos Szentagothai. The Marr–Albus model remains one of the most enduring and influential models in computational neuroscience, despite apparent falsification of some of the original predictions. We re-examine the Marr–Albus model in the context of the modern theory of computational neural networks and in the context of expanded interpretations of the connectivity and function of cerebellar cortex within the full motor system. By doing so, we show that the original elements of the codon theory continue to make important predictions for cerebellar mechanism, and we show that evidence appearing to contradict the original model is based on an artificially narrow interpretation of cerebellar structure and motor function.

Terry Sanger holds an SM in Applied Mathematics (Harvard), PhD in Electrical Engineering and Computer Science (MIT), and MD (Harvard), with medical specialization in Child Neurology and Movement Disorders. He is currently Provost Professor of Biomedical Engineering, Neurology, and Biokinesiology at the University of Southern California (USC), Director of the Pediatric Movement Disorders Clinic and Deep Brain Stimulation Program at Childrens Hospital of Los Angeles (CHLA), and the founding Academic Director of the Health Technology and Engineering program at USC (HTE@USC). His research on disorders of developmental motor control is driven by his interest in finding new treatments for children with movement disorders including dystonia, chorea, spasticity and dyspraxia. He has a particular interest in computational motor learning, and the role of motor learning in recovery from childhood brain injury. Major focus areas of laboratory research include wearable devices to promote motor learning, EMG-driven communication devices and assistive prosthetics, and modelling of the electrophysiology of deep-brain stimulation.



(Received 1 August 2019; accepted after revision 5 December 2019; first published online 17 December 2019)

Corresponding author T. Sanger: University of Southern California, 1042 Downey Way, DRB 140, Los Angeles, CA 90089, USA. Email: terry@sangerlab.net

Abstract figure legend Overall structure of the cerebro-cerebellum. We propose that the role of the cerebellar cortex and Purkinje cells is to modulate and select signals from incoming mossy fibres for transmission through the dentate nucleus to a highly compressed representation on the outgoing fibres projecting to frontal cortex. Inferior olive (IO) provides the training signals for plasticity.

Introduction

Fifty years after it was initially proposed, the theory of cerebellar computational function of Marr (1969) and Albus (1971) remains one of the most influential and debated computational models, driving the field of cerebellar study and providing an important example of the role of Computational Neuroscience for the investigation of brain function. The theory was developed based on work by Eccles, Ito and Szentagothai (Eccles *et al.* 1967), and although there are notable differences between the details of the Marr and Albus models, the underlying structural similarities have endured. Despite the fact that empirical investigations have provided evidence that is sometimes at odds with predictions of the theory, an accepted alternative theory of cerebellar cortical function has yet to emerge.

Shared components of the Marr and Albus models include (1) plasticity at the granule cell (parallel fibre) to Purkinje cell synapses, (2) control of plasticity by climbing fibre activity, and (3) expansion recoding of mossy fibre inputs by granule cells. The existence of plasticity and its modulation by climbing fibres, as well as other sites of plasticity within cerebellum, has been studied (Ito *et al.* 1982; Hansel *et al.* 2001; Jörntell & Hansel, 2006) and reviewed comprehensively (Hansel *et al.* 2001; D'Angelo, 2014; Badura & De Zeeuw, 2017; Raymond & Medina, 2018). Here, we examine the implications of expansion recoding in terms of more recent theoretical and experimental results. We focus primarily on motor function of the cerebro-cerebellum, which controls arm and dextrous hand movements through its connections to frontal cortex via the dentate nucleus and motor thalamus.

Codon theory (Blomfield *et al.* 1970; Marr, 1970) was named after the nucleotide triplets that comprise the genetic code, and its mechanism was based on earlier theoretical work of Brindley (1969). Re-coding of the input into discrete codon patterns has been proposed as a mechanism to permit flexible representation and rapid learning of non-linear functions and pattern discrimination through plasticity at the parallel fibre to Purkinje cell synapses (Blomfield *et al.* 1970; Cayco-Gajic *et al.* 2017). Theoretical analyses have examined the computational importance of the degree of connectivity (Billings *et al.* 2014; Litwin-Kumar *et al.* 2017).

Predictions of the original codon theory included:

- (1) only a few granule cells are active at a time;
- (2) minimal overlap occurs between the granule cell representations of different states;
- (3) granule cells encode unchanging functions of the mossy fibre inputs.

There is experimental evidence that calls each of these predictions into question. There appears to be plasticity throughout the cerebellum, including at the mossy fibre to granule cell synapses, and within the deep cerebellar nuclei (Gao *et al.* 2012; D'Angelo, 2014; Sgritta *et al.* 2017; Raymond & Medina, 2018). Multi-cell recording and imaging methods have shown that very high numbers of granule cells are simultaneously active, sometimes approaching 50% at any given time (Badura & De Zeeuw, 2017; Giovannucci *et al.* 2017; Knogler *et al.* 2017; Gilmer & Person, 2018). As a result, there may be significant overlap between the pattern of activity in different states. Much of these data have been obtained using calcium-sensitive dyes with two-photon imaging methods. Because the time constants of the dyes (hundreds of milliseconds) and the relatively low rate of image sampling (30 ms or more per frame) are slow compared to the synaptic integration time scale for granule cells (10–30 ms), simultaneous activation of granule cells will be overestimated (Cayco-Gajic & Silver, 2019). We further claim that these predictions do not constitute required elements of the codon theory, and thus their falsification does not negate the theory. When placed within the context of the function and input–output connectivity of the cerebellum, we believe that the codon theory continues to describe the central computational elements that result in cerebellar function for motor control.

Early learning models were based on Rosenblatt's perceptron theory (Rosenblatt, 1958), which postulates gradual changes in network output as a result of pairing of input and a training signal. Albus assumed the climbing fibre training signal is an error signal so that Purkinje cell output should be reduced when errors occur (Albus, 1971). Marr assumed the climbing fibre training signal is a teaching signal so that Purkinje cell output should adapt to emulate its teacher (Marr, 1969). Marr predicted that climbing fibre activity would facilitate Hebbian long-term potentiation (LTP) at the parallel fibre to Purkinje cell

synapses, whereas Albus predicted that climbing fibre activity would cause long-term depression (LTD). LTD has been experimentally verified, although LTP occurs in the absence of climbing fibre activity (Ito *et al.* 1982; Hansel *et al.* 2001; Gao *et al.* 2012; Raymond & Medina, 2018).

An important omission of the Marr and Albus models was a failure to place the cerebellum in the context of the complete sensory-motor apparatus necessary for biological movement. Since the Purkinje cell output is inhibitory, it must inhibit *something*, so understanding the incoming excitatory mossy fibre drive to the deep cerebellar nuclei is essential to any model of function. Furthermore, the role of any training signal on the climbing fibres needs to be interpreted in the context of the inputs and outputs of the deep cerebellar nuclei and the eventual effect of those outputs. The elegant computational hypothesis of supervised learning failed to specify the origin or type of training signal, the need for a non-linear mapping of the sensory data, the purpose or encoding at the output, or the relationship between the cerebellum and other components of the feedback and feedforward control systems. Figure 1 shows a more complete schematic illustration of the connectivity of the cerebellum and its reciprocal connections to motor cortices.

The striking expansion of representation in humans from approximately 250 million mossy fibres to more

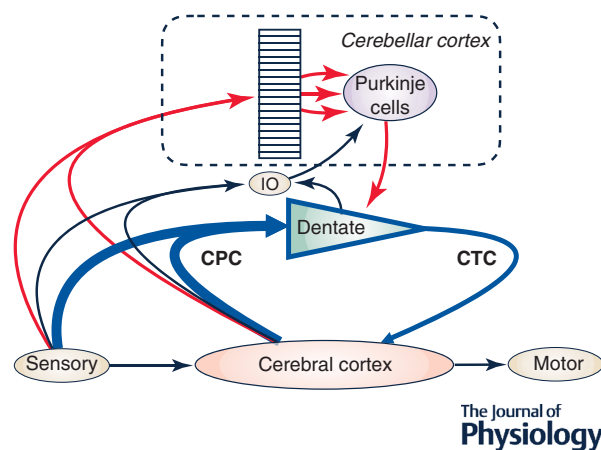


Figure 1. Input and output connectivity of the dentate nucleus of the cerebellum

Inputs arise on mossy fibres from spinocerebellar and other sensory systems, as well as from cerebral cortex via the cortico-ponto-cerebellar (CPC) pathway. Outputs from the dentate nucleus project to cerebral cortex via the cerebello-thalamo-cortical pathway (CTC) as well as to the red nucleus (not shown). The effect is a loop between cerebral cortex and cerebellar dentate nucleus as shown with the blue arrows. Dentate transmission is regulated by the inhibitory output from the Purkinje cells, which is determined by the granule cell input and the results of plasticity controlled by climbing fibres from the inferior olive. Sensory and cortical inputs from mossy fibres branch to supply the granule cells.

than 50 billion granule cells (Herculano-Houzel, 2009), followed by compression through 15 million Purkinje cells (Eccles *et al.* 1967) was the basis for the original codon theory. Equally striking, however, is the dramatic compression from 250 million mossy fibres at the input to fewer than 1 million cells in the dentate nuclear output (300,000 per side in the human; Andersen *et al.* 2004), a compression ratio of approximately 400:1. Since the dentate output may represent activity in ‘microzones’ with 100 or more Purkinje cells each (Apps *et al.* 2018), the mapping from mossy fibre input to dentate nuclear output may effectively exceed 1000:1 compression if correlated Purkinje cell outputs lead to correlated activity in the dentate nucleus. The significance and function of this dramatic compression remain unexplained and were not considered in the original theory. Figure 2 gives an approximate quantitative illustration of the relative numbers of cells and synapses in each of the cerebellar components. There are approximately 200 billion synapses between the mossy fibres and granule cells, 10 trillion synapses between granule cells and Purkinje cells, and 5 billion synapses between Purkinje cells and deep nuclear cells (Apps *et al.* 2018). Therefore, while plasticity occurs throughout the cerebellum, there is 50 times more potential for stored information at the granule cell to Purkinje cell synapses than in all other regions combined, perhaps justifying Marr’s and Albus’s emphasis on the importance of plasticity at these synapses. We will revisit the original proposed structure of the computations that serve essential motor functions of the cerebellum. Just as different muscles, despite similar structure, have very different functions depending on their origin, insertion and innervation, so we believe cerebellum may have a very different function for eye movement, balance, reaching and cognition despite its homogeneity of structure. The challenge is to propose a single computational structure that can subserve many different functions. We will focus primarily on the function of the cerebellar hemispheres, the cerebrocerebellum, projecting to dentate nucleus and believed to be most closely related to voluntary control of the upper extremity, speech and some aspects of cognitive function.

Behavioural functions of the cerebellum. Clues to the behavioural function of cerebellum come from observations of neurological disorders, including focal lesions (Holmes, 1917), Purkinje cell degeneration (Boder & Sedgwick, 1958), metabolic disorders (Steinlin *et al.* 1998) and disorders of cerebellar development (Fogel & Perlman, 2007; Machado *et al.* 2015). Common clinical signs included under the heading of ‘cerebellar ataxia’ include poor coordination of multijoint movement (dyssynergia), intention tremor, inappropriate magnitude of movement (dysmetria), poor timing or rhythmicity

(dysrhythmia), and inability to perform rapid alternating movements (dysdiadochokinesia) (Schmitz-Hübsch *et al.* 2006; Machado *et al.* 2015). Recent clinical studies document cognitive findings that form the ‘cerebellar cognitive affective syndrome’, which includes deficits of executive function (including abstract reasoning, context-shifting, sequencing and planning), language (including deficits of grammar as well as fluency), memory (particularly spatial memory) and affect (Schmahmann & Sherman, 1997, 1998; Kraan, 2016).

From deficits present in ataxia and similar disorders, we can infer some possible roles of the cerebellum in motor control. These include adaptation to biomechanical context (Bastian, 2006), context-dependent modulation of motor activity (Wolpert *et al.* 1998), coordination of movement across multiple joints (Bhanpuri *et al.* 2014), control of the dynamics of movement (Kawato *et al.* 1987; Middleton & Strick, 1998; Doya, 1999; Morton & Bastian, 2007) and cognitive tasks (Ito, 2008).

An important distinction between different functions is the need for different representations of information (Giovannucci *et al.* 2017). Control of muscles requires graded continuous activity that depends on accurate representations of position, velocity and force. Context-dependent movement, context-shifting, abstract reasoning and executive function are more likely to depend on discrete categorized representations of information. Therefore motor control will tend to have continuous representations that require calculations of smooth functions, while abstract reasoning will tend to have binary representations that require calculations of Boolean logical operators. It is possible that some functions

such as context-dependent motor control will require a combination of smooth and Boolean approximation. The original codon representation suggested by Marr (1969) and the cerebellar model articulatory controller (CMAC) architecture suggested by Albus (1975) both assume that information is encoded by binary patterns over the set of granule cells. As with many subsequent extensions of these theories (Houk *et al.* 1996), we emphasize that rate-coded or bursting activity in granule cells could represent the smooth continuous functions more relevant for motor control. Throughout the discussion below, we will address both continuous and discrete calculations that may be performed by the same cerebellar structure.

Motor functions of the cerebellum. The output of the cerebellum is generated exclusively by the deep cerebellar nuclei. Output to frontal cortex arises from the dentate and anterior interpositus nuclei, which in turn receive their input from cerebrocerebellar cortex (Lu *et al.* 2007) (see Fig. 1). Since the incoming mossy fibres synapse within the deep nuclei as well as within the granule cell layer (Zhang & Linden, 2006) and the Purkinje input to the deep nuclei is purely inhibitory, there is increasing recognition that for at least some functions, the role of cerebellar cortex may be as a modulator of transmission or learning in the deep nuclei, rather than as the desired output itself (Ito, 2006). In the context of oculomotor control, it has been pointed out that control and learning can occur in the absence of cerebellar cortical input (Ito, 1984). This is presumably mediated by the direct connection between mossy fibre inputs and the deep cerebellar nuclei that provide the outputs (Zhang &

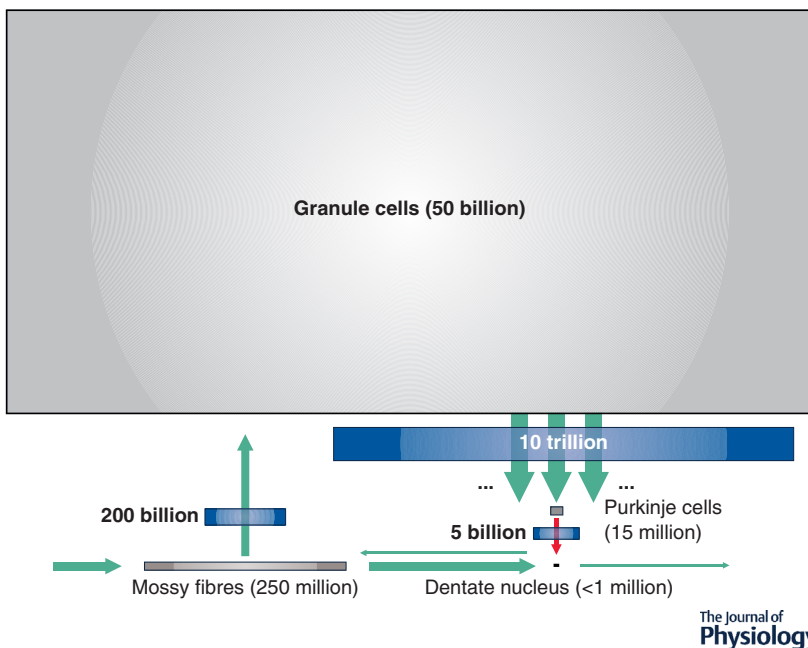


Figure 2. Illustration of the relative numbers of cells and synapses in each region of cerebellum

The area of each grey rectangle is proportional to the approximate cell count in that region. The area of each blue inset illustrates (but not to scale) the relative number of output synapses from that region. Note that the dentate nuclei are almost invisible due their relatively small counts and size.

Linden, 2006; Requarth & Sawtell, 2011; Raymond & Medina, 2018).

The cerebellum is not directly responsible for motor output to muscles. We propose that the cerebellum provides input to cortical and brainstem motor areas that is appropriate for correction of error during movement. The dentate nucleus projects to the red nucleus in the brainstem, and via thalamus to frontal cortex and striatum (Rand, 1954; Kelly & Strick, 2003; Gallay *et al.* 2008). In return, cerebral cortex (and possibly basal ganglia) project back to the cerebellum via pontine relay centres (Kelly & Strick, 2003; Palesi *et al.* 2017). The cerebellum is thus a relay station from mossy fibre input to deep nuclear output, with the relay modulated in a highly context-specific way by the cerebellar cortex. This modulation can provide gating (allow or prevent transmission), gain control (modulation of pass-through signals), multiplexing (selection of which of many possible signals to transmit) or compression (mixing input signals to produce relevant outputs). The multiplexing and compression functions may be particularly important because the number of mossy fibre inputs (250 million) far exceeds the number of dentate nuclear outputs (300,000 per side in human; Andersen *et al.* 2004). For perspective, it is worth realizing that the number of output neurons is fewer than the number of neurons in the substantia nigra compacta that project to and modulate striatal function. Therefore it seems likely that the output from the cerebellum is unable to provide detailed patterns for control of frontal cortex, but may rather provide widespread modulation, perhaps with precise timing.

Removal of some or all of the cerebellum in paediatric tumour patients is compatible with complete neurological recovery in as many as one-third of cases (Cochrane *et al.* 1994; Sonderkaer *et al.* 2003). The motor system is capable of learning and control, including feedback control, in the absence of a fully functioning cerebellum (Crisicimagna-Hemmingner *et al.* 2010). Recent evidence supports the hypothesis that at least some climbing fibres encode motor error (Giovannucci *et al.* 2017; Herzfeld *et al.* 2018). If climbing fibres indicate error, then this is error due to the current inability of the remainder of the motor system to reduce that error. Therefore we might imagine a system in which the cerebellum is called upon to correct errors only when the rest of the system is unable to do so. This could occur while the rest of the system is learning, or it could occur following learning if the rest of the system does not possess the computational structures or data connectivity sufficient to reduce certain types of error. In either case, an important role of the cerebellum could be as a rapid computational 'patch' to either temporarily or permanently permit correction or compensation for specific motor outputs in specific situations. It can do this by providing access to context-specific sensory or state data in the precise

situations in which these data are needed. This hypothesis suggests an interpretation of climbing fibre input as a request for help; it could encode a specific request for more or less activity, or a more general request for information or computational support. Depending on the way in which error information is calculated, other areas of the brain that learn more slowly than the cerebellar cortex might eventually catch up, reduce the errors to zero, and the cerebellar cortex would no longer need to perform those calculations (see Fig. 3). This could be one explanation for the transient nature of representations in cerebellum following initial exposure to new tasks (Raymond & Medina, 2018).

Stability requirement. The brain must be able to protect the body from harm (Sanger, 2014). A significant risk to the body is the presence of multiple positive and negative feedback loops that link sensory information to motor output. It is striking that despite these feedback loops, healthy vertebrates are almost never destabilized by the environment to exhibit involuntary oscillation or uncontrolled increases in output. Nevertheless, instability is a constant risk for any motor control system that includes feedback. This is particularly true in biological systems for which the very large dimensionality of sensory input could drive, either through context or by direct reflexes, any of the muscles contributing to motor output. An essential component of control and stability is to limit the sensory influence on motor output to only the components of the sensory input that are relevant to the task being performed. Failure to regulate the gain of sensory input or to limit the influence of irrelevant input will lead to instability and uncontrolled variability (Sanger *et al.* 2005).

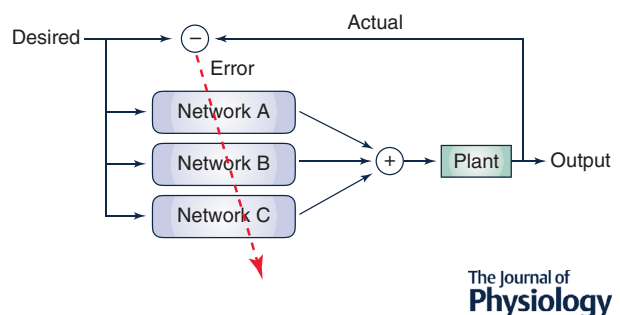


Figure 3. Illustration of the effect of multiple learning networks in parallel

The output is generated by the sum of the network outputs, and comparison with the desired output provides an error that can be used to train all the networks. As soon as any of the networks succeeds in reducing the error to zero, the remaining networks do not need to train further; the network that learns fastest and most accurately will become responsible for reducing error.

There is a continual competition between LTP and LTD at the parallel fibre to Purkinje cell synapses, with non-specific LTP competing with highly context-specific LTD (Hansel *et al.* 2001; Jörntell & Hansel, 2006). Since the Purkinje cell output inhibits the deep nuclei, this mechanism can be interpreted as inhibiting cerebellar output except in very specific contexts indicated by climbing fibre activity (Tyrrell & Willshaw, 1992). This is consistent with the observation of high spontaneous firing rates in Purkinje cells, particularly in the absence of climbing fibre input (Montarolo *et al.* 1982; Cerminara & Rawson, 2004; Zhou *et al.* 2014). High Purkinje cell firing rates provide tonic inhibition to the deep cerebellar nuclei, potentially preventing transmission of data from mossy fibres to the thalamus. By inducing LTD, the climbing fibre inputs permit very selective information transmission. Such a structure may be particularly important for reducing the risk of instability, because the overall gain in the sensori-motor loop passing through cerebellar deep nuclei is kept low except in very specific circumstances.

Computational implications of the codon theory

A striking feature of the human cerebellar cortex is not only the 200:1 expansion from 250 million mossy fibre inputs to more than 50 billion granule cells, but the fact that each granule cell contacts only a small number of different mossy fibres (typically 4, with a range from 1 to about 7) (Eccles *et al.* 1967; Marr, 1969; Herculano-Houzel, 2009; Badura & De Zeeuw, 2017; Knogler *et al.* 2017; Litwin-Kumar *et al.* 2017; Raymond & Medina, 2018). From an anatomical viewpoint, this does not have to be the case; there are certainly many areas in brain with much higher connectivity. This suggests that there is some computational reason why such a small number is needed, and why larger numbers do not occur (Billings *et al.* 2014; Cayco-Gajic *et al.* 2017; Litwin-Kumar *et al.* 2017; Sawtell, 2017). In the following sections we explore the computational consequences of this structure.

Similar to Rosenblatt's perceptron theory, the Marr–Albus theory is based on Purkinje cells forming a synaptically weighted linear combination of the granule cell inputs. This is therefore a regression model, and the optimal weights will be formed from the regression of the desired outputs on the input, as we discuss below in the section on learning. We emphasize that although the underlying computational structure performs regression, this same structure can be used to perform either smooth function approximation or Boolean function approximation, depending on whether the inputs and outputs are interpreted as continuous or binary. In the following, we discuss how the cerebellar cortical structure can be used both for smooth function approximation and for Boolean function approximation.

Smooth function approximation. The original Marr–Albus theory posited that granule cell connections were randomly selected codons that could be used to represent higher-order features of the input. We do not disagree. However, we extend the Marr–Albus model to include non-Boolean functions of the input, since dynamic control is based on calculation of continuous variables such as torque or muscle force. The original Marr codons were Boolean combinations of the instantaneous mossy fibre firing pattern (Marr, 1969; Tyrrell & Willshaw, 1992), extended by Albus to include Boolean combinations of the bits of the binary representation of continuous inputs (Albus, 1975). There are a number of different ways to represent non-binary data in neural firing patterns. The simplest is through the use of rate coding, in which the firing rate (or probability of firing) is proportional to the encoded signal. There are examples of rate coding in cerebellar granule cells (Arenz *et al.* 2008), and we will use rate coding here for illustration. Other non-binary codes include temporal codes, coincidence codes, interspike interval codes and burst codes (Rancz *et al.* 2007).

Suppose the granule cell output rate can be modelled as a non-linear-output weighted sum

$$g_i = \sigma \left(\sum_{j=1}^4 b_{ij} m_j \right) \quad (1)$$

where g_i is the i th granule cell output, m_j is the mossy fibre input firing rate, b_{ij} is the weight, and σ is a non-linearity. If we consider the Taylor expansion $\sigma(x) = a_0 + a_1x + a_2x^2 + \dots$, then we see that depending on the nature of σ the output may have linear or polynomial terms in its output. These terms will include $m_1^{n_1} m_2^{n_2} m_3^{n_3} m_4^{n_4}$ that indicate the non-linear interactions between mossy fibre inputs for integer values of n_i . If the granule cells are summed by the Purkinje cell, then the Purkinje cell will fire when

$$\theta < \left(\sum_k w_k m_{1,k}^{n_{1,k}} m_{2,k}^{n_{2,k}} m_{3,k}^{n_{3,k}} m_{4,k}^{n_{4,k}} \right) \quad (2)$$

where θ is the firing threshold, and w_k is the weight for the k th monomial, which may be carried on multiple parallel fibres. Therefore we see that this Purkinje cell output represents a fourth order polynomial in the mossy fibre inputs. The number of granule cells is insufficient to represent all possible fourth order monomials of the mossy fibres, so it is possible that plasticity at the mossy fibre to granule cell synapse selects monomials that occur with higher likelihood.

Calculation of force output requires access to position, velocity, acceleration and opposing force, so fourth order non-linear interactions may be required to solve dynamics problems. Mossy fibre inputs may indicate different derivatives of elements of state, or time-delayed

representations of state (van Beugen *et al.* 2013). For instance, calculation of the required torque to drive a serial-link manipulator such as an arm requires calculation of inertial forces due to joint angular acceleration, as well as centrifugal and Coriolis forces. Acceleration forces depend on the second derivative of joint angle. Coriolis forces depend on the products of the first derivatives of interacting joints. Therefore successful control of reaching requires knowledge of the first and second derivatives of the angle of all joints, as well as cross products between all interacting joints. These interactions must be learned in each behavioural context, because they change, for example, in a viscous environment (such as in water), an inertial environment (such as when holding a heavy object), or a mechanically constrained environment (such as when wearing a long-arm cast).

Boolean function approximation. If the mossy fibre inputs represent discrete states rather than continuous variables, then granule cells will represent a Boolean combination of up to four of the inputs. For Boolean functions, the inputs and outputs are interpreted as a 1 if there is a spike during a particular time interval, and 0 otherwise. Expansion coding enables linear combinations of binary inputs to represent a more complete set of Boolean functions. Typical uses of Boolean functions include pattern separation, classification, or any task for which the desired output is a discrete or categorical variable.

Depending on the threshold for Purkinje cell firing, a linear combination of binary inputs could reflect the conjunction (OR) of the inputs, the disjunction (AND) of the inputs, or a combination of OR and AND. For illustration, suppose the granule cell represents disjunction (AND), meaning that all four mossy fibres would have to fire simultaneously to trigger the granule cell. Then if the Purkinje cell performs a conjunction (OR) of the parallel fibres, the result is disjunctive normal form:

$$\bigvee_k m_{1,k} \wedge m_{2,k} \wedge m_{3,k} \wedge m_{4,k} \quad (3)$$

which is the Boolean equivalent of a fourth order polynomial. It is not necessary that each granule cell performs a disjunction (AND) of the mossy fibre inputs, so it is probably more accurate to write:

$$\bigvee_k b_k(m_{1,k}, m_{2,k}, m_{3,k}, m_{4,k}) \quad (4)$$

where b_k is the Boolean function computed at the granule cells. The effect of the Boolean calculation is the same: the Purkinje cell will inhibit the deep cerebellar nuclei whenever some or all of the input conditions are satisfied. Conversely, the Purkinje cell will not fire *only if most of*

the connected inputs are not firing, so the deep cerebellar nuclei are disinhibited whenever:

$$\bigwedge_k \bar{b}_k(m_{1,k}, m_{2,k}, m_{3,k}, m_{4,k}) = 1 \quad (5)$$

Synchrony between firing of different Purkinje cells (Person & Raman, 2012a,b) would represent simultaneous Boolean computations, leading to the intriguing hypothesis that the population output is a synchronized digital representation.

There is evidence from recordings that multi-modal interactions of multiple sensory modalities occur in granule cells (Huang *et al.* 2013; Ishikawa *et al.* 2015). Therefore granule cells need access to several mossy fibre inputs in order to be able to perform multi-modal processing. They can pass multi-modal information between multiple parts of the body and multiple representations in order to fine-tune the selection of specific behavioural contexts. Parallel fibre inputs whose effect on the Purkinje cell have been suppressed through climbing fibre-induced LTD can be active without necessarily causing Purkinje cell firing, and thus they reflect multi-modal components of input patterns whose transmission is permitted.

Expansion recoding. Pattern separation has been identified as an important justification for expansion coding (Billings *et al.* 2014; D'Angelo, 2014; Cayco-Gajic *et al.* 2017; Gilmer & Person, 2018; Cayco-Gajic & Silver, 2019). Discrete patterns are more likely to be linearly separable in a high-dimensional space, and therefore expansion coding enhances the ability of a linear Purkinje cell model to have flexible learned responses to differing inputs. For example, Minsky & Papert (1969) showed that no linear function of two binary inputs a and b could compute exclusive-OR, whereas expansion to include the additional input a and b resolves this problem.

The calculation of Boolean or polynomial functions from the mossy fibre inputs leads to a potentially much larger number of functions than the original set of inputs. This phenomenon is well-known in mathematics, and was the original concept behind the 'radial basis function' neural network algorithm (e.g. Broomhead & Lowe, 1988). The underlying idea is that any desired function $y(x)$ can be approximated as the weighted sum over a set of basis functions $g_i(x)$:

$$y(x) \approx \hat{y}(x) = \sum_{i=0}^N w_i g_i(x) \quad (6)$$

If the basis set $\{g_i\}$ (represented by the granule cell outputs from eqn (1)) is sufficiently rich, then the expected error $E[y - \hat{y}]$ can be made arbitrarily small by choosing a sufficient number of basis functions in eqn (6). The number of possible basis functions N is typically much

larger than the dimensionality M of the input x . In extreme cases, there can be an infinite number of basis functions even for x with only one dimension, as in the Taylor expansion of a univariate function $g_i(x) = x^i$. The generalization to multivariate x leads to the polynomial basis. Other examples of multivariate bases include the Fourier basis $g_i(x) = \sin(f_i x + \phi_i)$, radial basis functions $g_i(x) = h(\|x - x_i\|)$, wavelet expansions, and generalized splines. The basis proposed by Albus for the CMAC architecture is a particularly interesting one, consisting of all possible Boolean combinations of the bits of the digital representation of x (Albus, 1975). Unlike the preceding examples, this allows discontinuous functions and perhaps for this reason it has shown utility for the control of robotic manipulators.

Non-linear basis function approximation methods can be considered examples of 'expansion recoding'. They non-linearly transform an input x of dimension M into an expanded code g of dimension $N \gg M$. While certain codes may have a sparseness property (for which only a subset of the outputs g_i are non-zero), this is not a requirement of expansion recoding, and many common codes (including the Fourier and Taylor bases) do not have this property. The purpose of expansion recoding for function approximation is not sparseness, but rather the ability to approximate a large class of functions. For example, the Taylor expansion can approximate arbitrary analytic functions to arbitrary accuracy, and the Fourier expansion can approximate arbitrary cyclic functions to arbitrary accuracy. In the absence of (non-linear) expansion recoding, only linear functions $\sum w_i x_i$ can be approximated, and the set of linear functions is not sufficient for motor control. In the absence of expansion recoding, certain Boolean functions cannot be approximated (this was the basis for Minsky and Papert's famous criticism of the Perceptron theory; Minsky & Papert, 1969).

For 18 years until the publication and widespread investigation of multilayer perceptrons (Rumelhart *et al.* 1987), the codon theory, CMAC architecture and other basis-function representations were the only biologically inspired examples of universal non-linear function approximators. With our current appreciation of the larger class of such approximators and their importance and utility for biological and robotic control, we now recognize that the codon theory implements expansion recoding and fourth order polynomial approximation. This permits the cerebellum to calculate non-linear and non-somatotopic functions that are inserted as needed to modulate and correct errors in the otherwise (approximately) linear and somatotopic dynamics implemented by other motor areas. When the codon theory is interpreted in the broader context of expansion recoding, it is synonymous with basis-function representations and contains all of their power and flexibility.

Modulation of dynamics. When non-linear basis-function approximators are inserted into a control system, they permit control and stabilization of dynamics. Consider a linear system

$$\dot{x} = Ax + Bz \quad (7)$$

where x is cortical state and z is sensory input. The matrix A represents local recurrent connections that implement dynamics within cerebral cortex, and the matrix B represents the transmission and transformation of sensory data z to the cerebral cortex. If there are only local somatotopic interactions, then the matrices A and B will be approximately diagonal, which permits only very limited control possibilities. Such a system would not be able to compensate for interactions between the joints of a multi-joint manipulator, for example, and it would be expected to show poor coordination, possibly similar to what is seen in cerebellar ataxia.

This suggests that one role of the cerebellum is to augment dynamics with non-linear and non-local interactions (Kawato *et al.* 1987; Kawato & Gomi, 1992; Ishikawa *et al.* 2015). In order to maintain consistency with anatomical structures for cortically controlled movement, we include separate terms for the cerebellar and extracerebellar dynamics. Therefore suppose that Ax and Bz are the state update and sensory input to cortex that are independent of cerebellum (eqn (7)), and let Cx and Dz be the state update and sensory input to cortex that are routed from mossy fibre connections to the deep cerebellar nuclei and back to cortex. A , B , C and D are all assumed to have somatotopy, in the sense that localized inputs project to localized outputs. This means that these matrices are all near-diagonal.

Now let $\phi_C(x, z)$ and $\phi_D(x, z)$ be the non-linear functions calculated by the outputs of Purkinje cells that modulate Cx and Dz , respectively. Then we can write the combined dynamics as:

$$\dot{x} = Ax + Bz + (1 - \phi_C)Cx + (1 - \phi_D)Dz \quad (8)$$

where $\phi_C(x, z)$ and $\phi_D(x, z)$ are diagonal matrices multiplying Cx and Dz corresponding to the assumption that the Purkinje cell outputs are modulators of signals that are transmitted through the deep cerebellar nuclei. The terms are written as $(1 - \phi)$ to correspond to inhibitory outputs since greater firing rates of the Purkinje cells ϕ correspond to greater inhibition of the deep cerebellar nuclei (with the assumption that $\phi \leq 1$). ϕ_C and ϕ_D can depend on both cortical state x and sensory state z , and they can calculate smooth modulation functions or Boolean on/off functions. Time-dependent, frequency-dependent, or delayed inputs $x(t - \delta)$ are possible, and such inputs provide additional non-linearities capable of implementing time-dependent dynamics (Requarth & Sawtell, 2011; van Beugen *et al.* 2013; Sawtell, 2017).

In the absence of climbing fibre synaptic modulation, the Purkinje cells will fire for almost all mossy fibre inputs (see discussion below). Thus ϕ_C and ϕ_D will by default be close to maximal firing leading to inhibition of transmission through the deep cerebellar nuclei Cx and Dz . Therefore the system defaults to stabilization of dynamics, preventing excessive feedback gain $(A + C)x$ or $(B + D)z$.

ϕ_C and ϕ_D not only calculate non-linear interactions, but also allow non-somatotopic or multi-modal elements of x and z to have influences that affect distant parts of the body. This is particularly important for balance and multi-joint interactions at higher speed. Equation (8) shows that $1 - \phi_C$ and $1 - \phi_D$ could be zero (corresponding to complete Purkinje cell inhibition of deep cerebellar nuclei) whenever the linear dynamics in eqn (7) are sufficient for behaviour. Equation (7) will be expected to perform well for slow, single-joint, or smooth cyclic movements since it implements a linear filter for the sensory input z . For movements that are rapid, multi-joint, or require sudden transitions, non-linear functions such as those in eqn (8) are required, and this will necessitate cerebellar intervention and reduction in Purkinje cell inhibitory outputs.

This structure permits great flexibility of control. In principle, eqn (8) could calculate

$$\dot{x} = f(x, z) \quad (9)$$

for any function f , which is the most general form of a non-linear dynamic system. Equation (8) predicts that complete cerebellar loss that includes injury or loss of the deep nuclei will lead to loss of movement precision and coordination of non-local interactions, but would not lead to total loss of control since the basic cortical dynamics in eqn (7) would be preserved. However, it also predicts that loss or injury of Purkinje cells alone could be much worse, since disinhibition of the deep nuclei could cause excessive gain in eqn (8) and potentially lead to destabilization, tremor, or involuntary sustained muscle activation.

Learning algorithm

Neural-inspired learning algorithms have been studied extensively. Mathematical models of un-normalized LTP are unstable and can grow without bound, but models of un-normalized LTD can never be unstable because synaptic strength cannot decrease below zero. The presence of Golgi cells and other interneurons can be used to stabilize the firing rates of both the Purkinje cells and the granule cells (Albus, 1971; Hansel *et al.* 2001; Ito, 2006). The result of climbing fibre activation is to reduce via LTD the effect of a particular set of basis functions (granule cells) on the Purkinje cell, selectively disinhibiting only in very specific contexts.

Effect of expansion coding on learning. Most of the literature on granule cell encoding assumes that the decoding is performed by a linear mixing operation at the Purkinje cells. Equation (6) is a linear mixture of non-linear functions $g_i(x)$, and all the non-linearity arises from the granule cell input layer g_i , with linear mixing performed by weighted summation at the Purkinje cell. Equation (6) shows that once the granule cell outputs g_i are calculated, the output \hat{y} is formed by a weighted linear combination that can be written $w^T g$. The optimal weights w^* in the linear approximation are given by the regression of the desired output vector y on the vector of granule cell inputs g :

$$w^* = E[gg^T]^{-1}E[gy^T] \quad (10)$$

where $E[gy^T]$ is the cross correlation between y and g , and $E[gg^T]^{-1}$ is the inverse of the autocorrelation matrix of g .

Under the assumption that the encoding $g(x)$ is not changing, the weights w can be learned through several different iterative algorithms. The least-mean-square (LMS) or Widrow–Hoff rule (Widrow & Hoff, 1960) is a recursive learning algorithm that performs gradient descent on the mean-squared error $E[||y - w^T g||^2]$ and is given by:

$$\frac{dw}{dt} = \gamma g(x)(y(x) - \hat{y})^T \quad (11)$$

where γ is a learning rate, $y(x)$ is the desired output for state x , $\hat{y} = w^T g(x)$ is the current approximation, and dw/dt is the instantaneous change in the weights. LMS has the advantage of being a local learning algorithm, meaning that only information present at the granule cell to Purkinje cell synapse is needed for learning. Interestingly, the rate of convergence of this algorithm is not affected by size, sparseness, or orthogonality of the basis $g(x)$, but rather it depends only on γ and the eigenvalues of the cross-correlation matrix $E[gg^T]$ (Haykin & Widrow, 2003).

If the inputs are uncorrelated so that $E[gg^T]$ is diagonal, then $w^* = E[gy^T]$ and the desired output can be learned using the simpler algorithm:

$$\frac{dw}{dt} = \gamma \bar{g} y^T \quad (12)$$

where the total weight $||w||$ is normalized, and \bar{g} is the normalized granule cell input (a process that has been attributed to the Golgi cells; Albus, 1971; Hansel *et al.* 2001; Ito, 2006; D'Angelo, 2014). (To obtain the exact correlation requires $||w|| = E[||y||]$ and $E[||\bar{g}||] = 1$.) The advantage of this algorithm is that it does not depend on calculation of the network output \hat{y} but requires only the teaching signal y . Therefore the climbing fibres can represent the desired output y and not the error signal

$y - \hat{y}$. Although error signals may be present (Herzfeld *et al.* 2018), y could be a performance error that needs to be corrected, rather than the explicit Purkinje cell output error. Lack of requirement for explicit representation of \hat{y} or $y - \hat{y}$ in the climbing fibres would solve the theoretical problem that the inferior olive (the source of the climbing fibres) has only indirect access to the Purkinje cell output \hat{y} via transmission through deep cerebellar nuclei which compress and mix the Purkinje outputs with mossy fibre inputs. Since expansion coding tends to decorrelate inputs, one important purpose of the codon representation may be to permit use of eqn (12) for learning.

Generalization. Generalization describes the behaviour of the network when presented with previously unseen data. While smooth basis functions (including radial basis functions, and the Fourier and Taylor bases) provide some local generalization, we expect that for stability, generalization in the cerebellum will be designed to inhibit the output for any data points that are far from previously seen data. Given the large number of granule cells and the constant pressure of LTP to increase their ability to drive Purkinje cell inhibition, this is likely to happen automatically. The system starts out as inhibiting everything, with full generalization of inhibition over all possible mossy fibre inputs. It then selectively permits disinhibition (Tyrrell & Willshaw, 1992). This is a sculpting operation: the default behaviour is globally generalized inhibition, which is the safest solution. The Appendix shows an example of a rapid sculpting algorithm of this type. Generalization is determined not only by the granule cell representation, but also by the representation that is incoming on the mossy fibres. Mossy fibres of cortical origin entering via the pons will reflect the representation in the cortex (Kennedy *et al.* 2014; Giovannucci *et al.* 2017).

There may not be a single unique signal transmitted on the climbing fibres (Simpson *et al.* 1996). In some cases, this signal is postulated to be the sensory prediction error: the difference between the predicted sensory result of an action and the actual sensory result. If this were used for training in eqn (12), it would lead to the cerebellar output reflecting computations required for correction of forward dynamics. Inferior olive has a rich sensory representation (Gellman *et al.* 1983), although it is not clear whether sensory prediction error is directly encoded. On the other hand, if the climbing fibres represent an error computed in motor output coordinates, as might be the case for oculomotor control, then the output of the Purkinje cells will resemble computations required for inverse dynamics (Herzfeld *et al.* 2018). The output of a (potentially low-quality) feedback controller is sufficient to produce motor error signals for this purpose, and this is the basis for the feedback-error learning algorithm

(Kawato *et al.* 1987; Miyamoto *et al.* 1988). If the errors reflect higher-level behaviours such as social behaviour, then the Purkinje cell outputs will represent higher-order functions (Wolpert *et al.* 2003). If the climbing fibres indicate the presence of error but not its sign, then the output will predict the presence of errors in each context, and the weightings and sign would have to be learned by a downstream network. If the climbing fibres indicate a forward-model prediction of future state, then the Purkinje cells will learn to predict (and cancel) the sensory effects of self-generated actions (Requarth & Sawtell, 2011; Kennedy *et al.* 2014). If the climbing fibres indicate the expected reward in a given state, then the output will learn a value function of state that can be useful for optimal control, as in temporal-difference learning. In all cases, the outputs will be predicted to show generalized inhibition with very context-specific disinhibition that is related to everything from low-level motor state to high level social and behavioural context.

An assumption of the original Marr and Albus models was that the granule cell representation was random and unchanging. While this can be effective for computation, it is likely that many random combinations of mossy fibre inputs would never occur or would have little use for computation. Therefore it is not surprising to find plasticity at the mossy fibre to granule cell synapses that could either be unsupervised (reflecting the statistics of the mossy fibre inputs) or supervised (tailored to the needs of the specific functions to be approximated) (Sgritta *et al.* 2017). We note that learning a granule cell representation would be expected to increase the representational and computational power of the network (Schweighofer *et al.* 2001; D'Angelo, 2014).

Role in cognition

The role of the cerebellum for cognitive function has been increasingly recognized (Schmahmann & Sherman, 1997). The effects of cerebellar injury on multiple higher-order functions including behaviour, language and affect suggests a strong modulatory or computational role (Leiner *et al.* 1993; Schmahmann & Sherman, 1997, 1998; Kraan, 2016). The computational basis of higher-order cognitive and language processing is a matter of speculation. Presumably cerebellar influence is mediated through connections from the ventral dentate nucleus to prefrontal cortex (Middleton & Strick, 2001). Higher cognitive function is less dependent on sensory input compared with lower-level motor control, and therefore we expect that the mossy fibre inputs to prefrontal-projection regions of dentate nucleus and cerebellar cortex will contain data from cerebral cortex, most likely prefrontal and higher order association areas. The cerebellum is connected in a way that permits

modulation of cortical dynamics which can be modelled similar to eqn (8) as:

$$\dot{x} = Ax + (1 - \phi(x))Cx \quad (13)$$

where x represents the cortical state, Ax is the internal cortical dynamics, and $\phi(x)Cx$ is the component of the dynamics controlled by cerebellum. Just as for motor control, the cerebellum in this case allows non-linear and non-local information to contribute to the ongoing cortical dynamics. To prevent oscillation or saturation of the cortical activity, the positive cerebellar feedback loop is inhibited as much as possible, with disinhibition permitted only in very specific situations.

The nature of the climbing fibre training signal for cognitive function is not known. One can conjecture that it could be used to train forward or inverse models of human behaviour, social interactions, or planned sequences of action. Since language involves the manipulation of discrete symbols, much of language processing would be likely to use Boolean basis functions, whereas emotional states or social interactions might be more likely to use continuous basis functions.

Conclusion

We propose that the cerebro-cerebellar cortex is an inhibitory modulator of incoming sensory signals and state feedback signals that drive cerebro-cortical dynamics. We propose that it does so by using expansion coding and a non-linear basis function representation to identify specific states. When these states occur, the Purkinje cell output modulates, selects or multiplexes task-relevant signals in the deep cerebellar nuclei. The structure permits great flexibility, allowing smooth inputs or binary inputs, computation of continuous functions or Boolean functions, and learning many different types of functions depending on the nature of the teaching signal provided by the climbing fibres.

Our hypotheses remain consonant with the initial predictions of Marr and Albus. The expansion recoding from mossy fibres to granule cells continues to serve the purpose of allowing non-linear (or Boolean) interactions to be modelled, and for learning to occur using a Hebbian algorithm equation 12 rather than the more complex supervised learning rule equation 11. The main difference we propose is that we see the overall purpose as being highly selective disinhibition, and we interpret the role of LTP as creating a default of widespread inhibition and stabilization. We propose that this is beneficial because it creates maximum safety for survival by reducing the chance of dynamic instability. The learning algorithm can be reinterpreted as sculpting small regions of disinhibition from a large region of inhibition.

In light of the above discussion, we suggest that Marr's and Albus' original predictions based on the codon theory including sparse and unchanging granule cell representations are not necessary consequences of expansion recoding, and are not essential to learning or representation. Therefore their falsification does not falsify the codon theory. We suggest that the true significance of the codon theory lies in the computational power of expansion recoding. When cerebellar function is interpreted in the context of both smooth and Boolean computation, and when the inhibitory effect of the Purkinje cells on the deep cerebellar nuclei is recognized, the persisting implications of the Marr–Albus model become evident.

We hypothesize that the overall purpose of the cerebellum is to provide otherwise unavailable signals to drive the temporal dynamics of multiple cerebral cortical and brainstem regions. Cerebellum is called to action when other brain regions do not have access to the data or computational ability required to perform their functions. In this interpretation, the climbing fibre input can be considered a call for help; it indicates computational assistance is needed, and the cerebellum responds by supplying the most selective and specific information in order to solve the problem without creating instability. The cerebellum works in tandem with the other components of motor function to ensure stability and to guarantee selectivity of context and appropriateness of sensory and state data for computation and behaviour. The Marr–Albus model continues to provide a strong theoretical basis for how this can be accomplished.

Appendix: Sandcastle Algorithm

Consider a non-orthogonal and possibly non-sparse code that takes random Boolean functions of N -dimensional input x , but which is allowed to depend on at most four elements of x . There are N^4 possible basis functions, so a complete code is not possible, but instead we can construct a random code $\phi(x)$. Each $\phi_i(x)$ is quite simple. If it is only AND gates then it will be selective for a particular pattern. If it is only OR gates then it will respond to any patterns with an on bit. All possibilities in between are allowed. This is intended to model the granule cell responses to the incoming mossy fibres.

Now the Purkinje cells can implement an AND function of a subset of $\phi_i(x)$, or an OR function of a subset of $\phi_i(x)$, or both. Suppose that initially the Purkinje cell thresholds are low so that they implement an OR gate. Then every time that the climbing fibre signals an error, LTD will reduce the effect of all the currently firing inputs, essentially causing a 'sculpting' of its response. If the climbing fibre is an error signal, then sculpting will occur whenever the Purkinje cell fires in an incorrect context,

so it is deleting an inappropriate generalization from the current context. There may also be changes in the overall threshold, so that a larger number of inputs are required to fire, which will lead to selectivity, although there is no guarantee that the coincidence of all of the granule inputs can actually be achieved by any real input.

LTP will gradually cause an increase in the synaptic efficacy of all synapses, proportional to the environment statistics. So there is an ongoing balance between LTP causing a statistical response to the environment, and LTD deleting inappropriate generalization. This process could stabilize at a cell that generalizes appropriately or that forms a suitable basis for generalization. This is consistent with the hypothesis that the purpose of the cerebellum may include finding basis functions that have generalization properties that are tailored to the task at hand.

Sandcastle algorithm

Suppose we have a desired Boolean function $y_\alpha(x)$ and let $\chi_1 = \{x | y_\alpha(x) = 1\}$ be its positive support. Instead of constructing y_α as a combination of the output of other Boolean functions, think of constructing χ_1 as a combination of the positive support of other Boolean functions. Consider the algorithm:

- (1) Initialize $\phi(x) = \phi_1(x) \vee \phi_2(x) \vee \dots$ for a large set of functions $\{\phi_i\}$, and assume that the positive support of y_α is fully contained in the positive support of the initial $\phi(x)$.
- (2) Repeat: for each x_k , if $y_\alpha(x_k) < \phi(x_k)$ remove all functions ϕ_i for which $\phi_i(x_k) = 1$

This process will ‘sculpt’ the positive support of ϕ . It is possible that too much will be removed, so that at some point the support of y_α will no longer be contained in the support of ϕ . In that case, we could add back some functions, perhaps the disjunction (AND) of pairs or groups of functions, in order to achieve greater specificity. Rather than do so, we use an iterative algorithm to change thresholds:

- (1) Initialize

$$\phi(x) = \begin{cases} 1 & \text{if } \sum w_i \phi_i(x) > \theta \\ 0 & \text{else} \end{cases}$$

for a large set of functions $\{\phi_i\}$, weights $w_i \geq 1$, and firing threshold θ . Assume that weights are chosen such that the positive support of y_α is fully contained in the positive support of the initial $\phi(x)$.

- (1) Repeat: for each x_k , if $y_\alpha(x_k) < \phi(x_k)$ set $w_i \leftarrow w_i - \gamma \phi_i(x_k)$ for all i .

This reduces the weights for components that over-generalize. Components with $w_i < 1$ still contribute, because if several trigger together then the weights may add

to a number greater than θ and thereby trigger the output. So this means that the output calculates the AND of the over-generalizing functions.

In combination with the above algorithms, we can have a continual increase

$$w_i \leftarrow w_i + \gamma \phi_i(x_k) \quad (\text{A1})$$

which will cause weights to reflect the statistics of the input. Weight normalization is required to maintain boundedness. The overall behaviour is a competition between the continual LTP increase based on the statistics of the input, and the LTD decrease based on sculpting from the output.

This is different from radial basis functions, which assemble the output through synthesis. Sandcastle creates the output by removal. It will automatically generalize, and then reduce generalization whenever an error occurs. This is similar to the idea in the basis-function tree algorithm for which in the absence of information to the contrary, the network generalizes as much as possible (Sanger *et al.* 1992).

Simulation

To illustrate the algorithm in the case of continuous approximation, we simulate 400 mossy fibres, 20,000 parallel fibres, and a single Purkinje cell. The input $x = (x_1 \dots x_4)$ has four dimensions. There are 100 mossy fibres m_{ij} for each dimension j , and each mossy fibre is tuned to one of the scalar inputs with either a linear or non-linear smooth tuning function $m_{ij}(x_i)$. These functions are chosen using a randomly weighted Fourier basis (a random combination of sin and cos functions). The tuning curves for $m_{1j}; m_{2j}; m_{3j}; m_{4j}$ are the same, so the representation is the same for each of the four dimensions, although on different groups of mossy fibres. There are 20,000 granule cells. Each takes input from four mossy fibres, one from each of the four input dimensions. The choice of mossy fibre inputs is random but unchanging. In the non-linear case, the four mossy fibre inputs are multiplied to obtain the granule cell activity. In the linear case, they are added. A randomly selected target function is obtained by multiplying six randomly chosen smooth Fourier basis functions $f(x) = f_1(x) \times f_2(x) \times \dots \times f_6(x)$ to obtain a smooth non-linear random function that is different from any of the mossy fibre or granule cell functions. The single Purkinje cell output is calculated as a weighted sum of the granule cell outputs, with weighting vector w .

For training, 2000 samples of x are presented in sequence. Samples are constrained to a linear subspace of the four-dimensional space. For each sample, the granule cell outputs $g(x)$ are calculated, and the output of the Purkinje cell is given by $y = 1 - w^T g(x)$,

since the output is inhibitory. An error is calculated as $e = f(x) - y$, and this is assumed to be carried on the climbing fibre. The weights are updated according to a Hebbian LTD rule, $\Delta w = -\gamma(ew^Tg(x))$, where γ is a learning rate set to 0.0003 for the simulations. Figure 4 shows a comparison of the output for the non-linear and the linear granule cell model. The test

set of inputs is the full set of training inputs (we do not test generalization). The non-linear model learns quickly, converging within fewer than 2000 iterations, even though there are 20,000 granule cell inputs that need to be trained. The linear model does not converge, and it shows poor approximation. Even when the linear model is applied to approximate a linear function (not

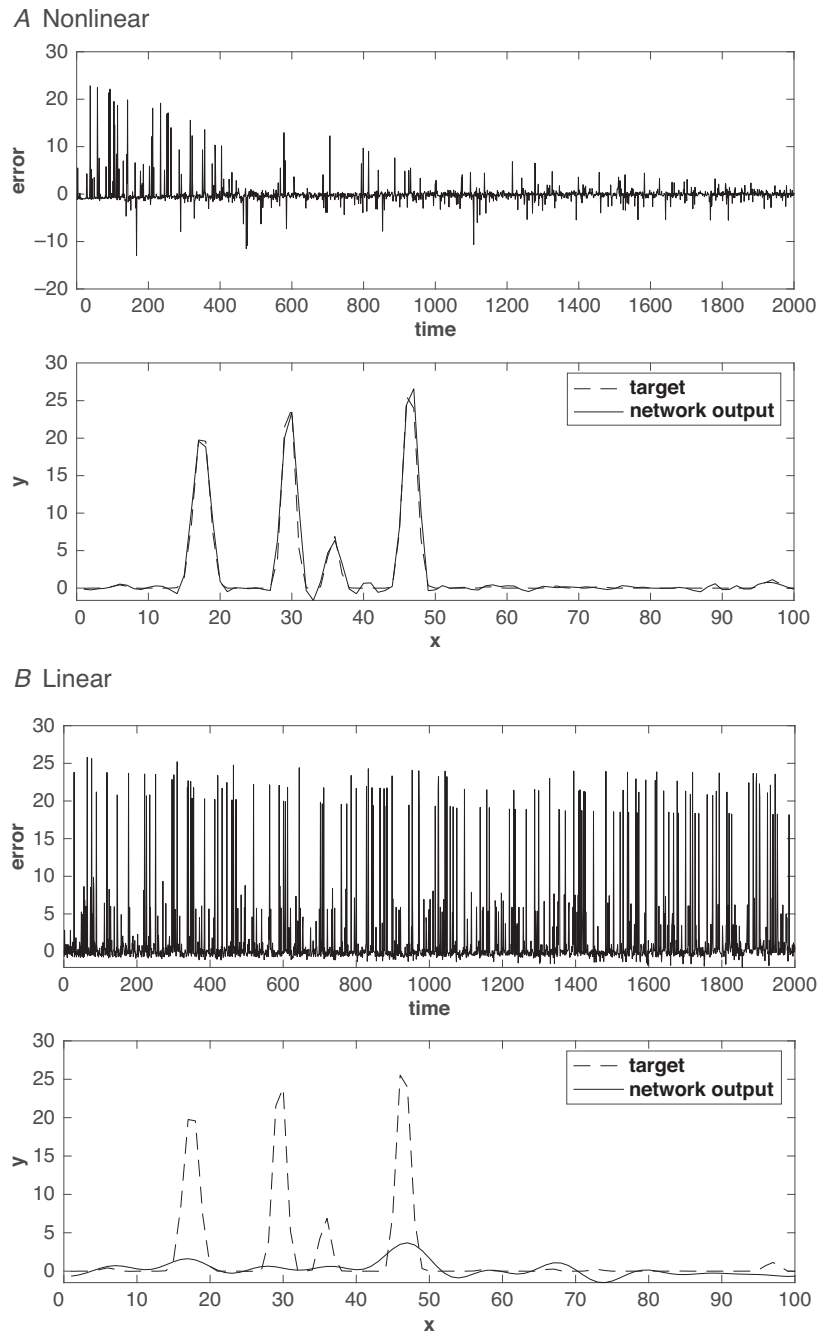


Figure 4. Comparison of the sculpting algorithm using granule cells representing fourth-order non-linear interactions between mossy fibres (A), and granule cells representing linear interactions between groups of four mossy fibres (B)
 Top traces show error as a function of time, with one example input/output pair per unit time. Bottom traces compare the learned network output (continuous line) to the target function (dashed line).

shown) the approximation is poor, most likely because even with 20,000 granule cells, linear combinations only have 100 linearly independent representations. (Matlab code for this simulation can be downloaded from <http://www.sangerlab.net/SandcastleAlgorithm.m>.)

References

- Albus JS (1971). A theory of cerebellar function. *Math Biosci* **10**, 25–61.
- Albus JS (1975). A new approach to manipulator control: The cerebellar model articulation controller (cmac). *J Dyn Syst Meas Contr* **97**, 220–227.
- Andersen B, Fabricius K, Gundersen H, Jelsing J & Stark A (2004). No change in neuron numbers in the dentate nucleus of patients with schizophrenia estimated with a new stereological method – the smooth fractionator. *J Anat* **205**, 313–321.
- Apps R, Hawkes R, Aoki S, Bengtsson F, Brown AM, Chen G, Ebner TJ, Isope P, Jörntell H, Lackey EP, Lawrenson C, Lumb B, Schonewille M, Sillitoe RV, Spaeth L, Sugihara I, Valera A, Voogd J, Wylie DR & Ruigrok TJH (2018). Cerebellar modules and their role as operational cerebellar processing units. *Cerebellum* **17**, 654–682.
- Arenz A, Silver RA, Schaefer AT & Margrie TW (2008). The contribution of single synapses to sensory representation in vivo. *Science* **321**, 977–980.
- Badura A & De Zeeuw CI (2017). Cerebellar granule cells: dense, rich and evolving representations. *Curr Biol* **27**, R415–R418.
- Bastian AJ (2006). Learning to predict the future: the cerebellum adapts feedforward movement control. *Curr Opin Neurobiol* **16**, 645–649.
- Bhanpuri NH, Okamura AM & Bastian AJ (2014). Predicting and correcting ataxia using a model of cerebellar function. *Brain* **137**, 1931–1944.
- Billings G, Piasini E, Lőrincz A, Nusser Z & Silver RA (2014). Network structure within the cerebellar input layer enables lossless sparse encoding. *Neuron* **83**, 960–974.
- Blomfield S, Marr D & Cowan JD (1970). How the cerebellum may be used. In *From the Retina to the Neocortex*, ed. Vaina L, pp. 51–57. Birkhäuser, Boston.
- Boder E & Sedgwick RP (1958). Ataxia-telangiectasia: a familial syndrome of progressive cerebellar ataxia, oculocutaneous telangiectasia and frequent pulmonary infection. *Pediatrics* **21**, 526–554.
- Brindley GS (1969). Nerve net models of plausible size that perform many simple learning tasks. *Proc R Soc Lond B Biol Sci* **174**, 173–191.
- Broomhead DS & Lowe D (1988). Radial basis functions, multi-variable functional interpolation and adaptive networks. Memorandum 4148. Royal Signals and Radar Establishment, Malvern, UK.
- Cayco-Gajic NA, Clopath C & Silver RA (2017). Sparse synaptic connectivity is required for decorrelation and pattern separation in feedforward networks. *Nat Commun* **8**, 1116.
- Cayco-Gajic NA & Silver RA (2019). Re-evaluating circuit mechanisms underlying pattern separation. *Neuron* **101**, 584–602.
- Cominari NL & Rawson JA (2004). Evidence that climbing fibres control an intrinsic spike generator in cerebellar purkinje cells. *J Neurosci* **24**, 4510–4517.
- Cochrane DD, Gustavsson B, Poskitt KP, Steinbok P & Kestle JR (1994). The surgical and natural morbidity of aggressive resection for posterior fossa tumors in childhood. *Pediatr Neurosurg* **20**, 19–29.
- Crisicimagna-Hemming SE, Bastian AJ & Shadmehr R (2010). Size of error affects cerebellar contributions to motor learning. *J Neurophysiol* **103**, 2275–2284.
- D'Angelo E (2014). The organization of plasticity in the cerebellar cortex: from synapses to control. *Prog Brain Res* **210**, 31–58.
- Doya K (1999). What are the computations of the cerebellum, the basal ganglia and the cerebral cortex? *Neural Netw* **12**, 961–974.
- Eccles J, Ito M & Szentagothai J (1967). *The Cerebellum as a Neuronal Machine*. Springer-Verlag, Berlin, Heidelberg.
- Fogel BL & Perlman S (2007). Clinical features and molecular genetics of autosomal recessive cerebellar ataxias. *Lancet Neurol* **6**, 245–257.
- Gallay MN, Jeanmonod D, Liu J & Morel A (2008). Human pallidothalamic and cerebellothalamic tracts: anatomical basis for functional stereotactic neurosurgery. *Brain Struct Funct* **212**, 443–463.
- Gao Z, Van Beugen BJ & De Zeeuw CI (2012). Distributed synergistic plasticity and cerebellar learning. *Nat Rev Neurosci* **13**, 619.
- Gellman R, Houk JC & Gibson AR (1983). Somatosensory properties of the inferior olive of the cat. *J Comp Neurol* **215**, 228–243.
- Gilmer JI & Person AL (2018). Theoretically sparse, empirically dense: New views on cerebellar granule cells. *Trends Neurosci* **41**, 874–877.
- Giovannucci A, Badura A, Deverett B, Najafi F, Pereira TD, Gao Z, Ozden I, Kloth AD, Pnevmatikakis E, Paninski L, De Zeeuw CI, Medina JF & Wang SS (2017). Cerebellar granule cells acquire a widespread predictive feedback signal during motor learning. *Nat Neurosci* **20**, 727.
- Hansel C, Linden DJ & D'Angelo E (2001). Beyond parallel fibre LTD: the diversity of synaptic and non-synaptic plasticity in the cerebellum. *Nat Neurosci* **4**, 467.
- Haykin S & Widrow B (2003). *Least-Mean-Square Adaptive Filters*. John Wiley & Sons, Hoboken, NJ, USA.
- Herculano-Houzel S (2009). The human brain in numbers: a linearly scaled-up primate brain. *Front Hum Neurosci* **3**, 31.
- Herzfeld DJ, Kojima Y, Soetedjo R & Shadmehr R (2018). Encoding of error and learning to correct that error by the purkinje cells of the cerebellum. *Nat Neurosci* **21**, 736.
- Holmes G (1917). The symptoms of acute cerebellar injuries due to gunshot injuries. *Brain* **40**, 461–535.
- Houk JC, Buckingham JT & Barto AG (1996). Models of the cerebellum and motor learning. *Behav Brain Sci* **19**, 368–383.
- Huang C-C, Sugino K, Shima Y, Guo C, Bai S, Mensh BD, Nelson SB & Hantman AW (2013). Convergence of pontine and proprioceptive streams onto multimodal cerebellar granule cells. *Elife* **2**, e00400.

- Ishikawa T, Shimuta M & Häusser M (2015). Multimodal sensory integration in single cerebellar granule cells in vivo. *Elife* **4**, e12916.
- Ito M (1984). *The Cerebellum and Neural Control*. Raven Press, New York.
- Ito M (2006). Cerebellar circuitry as a neuronal machine. *Prog Neurobiol* **78**, 272–303.
- Ito M (2008). Control of mental activities by internal models in the cerebellum. *Nat Rev Neurosci* **9**, 304.
- Ito M, Sakurai M & Tongroach P (1982). Climbing fibre induced depression of both mossy fibre responsiveness and glutamate sensitivity of cerebellar purkinje cells. *J Physiol* **324**, 113–134.
- Jörntell H & Hansel C (2006). Synaptic memories upside down: bidirectional plasticity at cerebellar parallel fibre-purkinje cell synapses. *Neuron* **52**, 227–238.
- Kawato M, Furukawa K & Suzuki R (1987). A hierarchical neural-network model for control and learning of voluntary movement. *Biol Cybern* **57**, 169–185.
- Kawato M & Gomi H (1992). A computational model of four regions of the cerebellum based on feedback-error learning. *Biol Cybern* **68**, 95–103.
- Kelly RM & Strick PL (2003). Cerebellar loops with motor cortex and prefrontal cortex of a nonhuman primate. *J Neurosci* **23**, 8432–8444.
- Kennedy A, Wayne G, Kaifosh P, Alviña K, Abbott L & Sawtell NB (2014). A temporal basis for predicting the sensory consequences of motor commands in an electric fish. *Nat Neurosci* **17**, 416.
- Knogler LD, Markov DA, Dragomir EI, Štih V & Portugues R (2017). Sensorimotor representations in cerebellar granule cells in larval zebrafish are dense, spatially organized, and non-temporally patterned. *Curr Biol* **27**, 1288–1302.
- Kraan C (2016). Cerebellar cognitive affective syndrome. In *Developmental Disorders of the Brain*, ed. Rinehart NJ, Bradshaw JL & Enticott PG, pp. 41–56. Psychology Press.
- Leiner HC, Leiner AL & Dow RS (1993). Cognitive and language functions of the human cerebellum. *Trends Neurosci* **16**, 444–447.
- Litwin-Kumar A, Harris KD, Axel R, Sompolinsky H & Abbott L (2017). Optimal degrees of synaptic connectivity. *Neuron* **93**, 1153–1164.
- Lu X, Miyachi S, Ito Y, Nambu A & Takada M (2007). Topographic distribution of output neurons in cerebellar nuclei and cortex to somatotopic map of primary motor cortex. *Eur J Neurosci* **25**, 2374–2382.
- Machado AS, Darmohray DM, Fayad J, Marques HG & Carey MR (2015). A quantitative framework for whole-body coordination reveals specific deficits in freely walking ataxic mice. *Elife* **4**, e07892.
- Marr D (1969). A theory of cerebellar cortex. *J Physiol* **202**, 437–470.
- Marr D (1970). A theory for cerebral neocortex. *Proc R Soc Lond B Biol Sci* **176**, 161–234.
- Middleton FA & Strick PL (1998). The cerebellum: an overview. *Trends Neurosci* **21**, 367–369.
- Middleton FA & Strick PL (2001). Cerebellar projections to the prefrontal cortex of the primate. *J Neurosci* **21**, 700–712.
- Minsky M & Papert SA (1969). *Perceptrons: An Introduction to Computational Geometry*. MIT Press, Cambridge, MA, USA.
- Miyamoto H, Kawato M, Setoyama T & Suzuki R (1988). Feedback-error-learning neural network for trajectory control of a robotic manipulator. *Neural Netw* **1**, 251–265.
- Montarolo PG, Palestini M & Strata P (1982). The inhibitory effect of the olivocerebellar input on the cerebellar Purkinje cells in the rat. *J Physiol* **332**, 187–202.
- Morton SM & Bastian AJ (2007). Mechanisms of cerebellar gait ataxia. *Cerebellum* **6**, 79.
- Palesi F, De Rinaldis A, Castellazzi G, Calamante F, Muhlert N, Chard D, Tournier JD, Magenes G, D'Angelo E & Wheeler-Kingshott CAG (2017). Contralateral cortico-ponto-cerebellar pathways reconstruction in humans in vivo: implications for reciprocal cerebro-cerebellar structural connectivity in motor and non-motor areas. *Sci Rep* **7**, 12841.
- Person AL & Raman IM (2012a). Purkinje neuron synchrony elicits time-locked spiking in the cerebellar nuclei. *Nature* **481**, 502.
- Person AL & Raman IM (2012b). Synchrony and neural coding in cerebellar circuits. *Front Neural Circuits* **6**, 97.
- Rancz EA, Ishikawa T, Duguid I, Chadderton P, Mahon S & Häusser M (2007). High-fidelity transmission of sensory information by single cerebellar mossy fibre boutons. *Nature* **450**, 1245.
- Rand RW (1954). An anatomical and experimental study of the cerebellar nuclei and their efferent pathways in the monkey. *J Comp Neurol* **101**, 167–223.
- Raymond JL & Medina JF (2018). Computational principles of supervised learning in the cerebellum. *Annu Rev Neurosci* **41**, 233–253.
- Requarth T & Sawtell NB (2011). Neural mechanisms for filtering self-generated sensory signals in cerebellum-like circuits. *Curr Opin Neurobiol* **21**, 602–608.
- Rosenblatt F (1958). The perceptron: a probabilistic model for information storage and organization in the brain. *Psychol Rev* **65**, 386.
- Rumelhart DE, McClelland JL & PDP Research Group (1987). *Parallel Distributed Processing, volume 1, Explorations in the Microstructure of Cognition: Foundations*. MIT Press, Cambridge, MA, USA.
- Sanger TD (2014). Risk-aware control. *Neural Comput* **26**, 2669–2691.
- Sanger TD, Kaiser J & Placek B (2005). Reaching movements in childhood dystonia contain signal-dependent noise. *J Child Neurol* **20**, 489–496.
- Sanger TD, Sutton RS & Matheus CJ (1992). Iterative construction of sparse polynomial approximations. In *Advances in Neural Information Processing Systems 4*, ed. Moody JE, Hanson SJ, Lippmann RP, pp. 1064–1071, Morgan Kaufmann, San Mateo, CA, USA.
- Sawtell NB (2017). Neural mechanisms for predicting the sensory consequences of behaviour: insights from electrosensory systems. *Annu Rev Physiol* **79**, 381–399.
- Schmahmann JD & Sherman JC (1997). Cerebellar cognitive affective syndrome. *Int Rev Neurobiol* **41**, 433–440.
- Schmahmann JD & Sherman JC (1998). The cerebellar cognitive affective syndrome. *Brain* **121**, 561–579.

- Schmitz-Hübsch T, Tezenas du Montcel S, Baliko L, Boesch S, Bonato S, Fancellu R, Giunti P, Globas C, Kang J-S, Kremer B, Mariotti C, Meleghe B, Rakowicz M, Rola R, Romano S, Schöls L, Szymanski S, van de Warrenburg BP, Zdzienicka E, Dsürr A, Klockgether T (2006). Reliability and validity of the international cooperative ataxia rating scale: a study in 156 spinocerebellar ataxia patients. *Mov Disord* **21**, 699–704.
- Schweighofer N, Doya K & Lay F (2001). Unsupervised learning of granule cell sparse codes enhances cerebellar adaptive control. *Neuroscience* **103**, 35–50.
- Sgritta M, Locatelli F, Soda T, Prestori F & D'Angelo EU (2017). Hebbian spike-timing dependent plasticity at the cerebellar input stage. *J Neurosci* **37**, 2809–2823.
- Simpson J, Wylie D & De Zeeuw C (1996). On climbing fibre signals and their consequence(s). *Behav Brain Sci* **19**, 384–398.
- Sonderkaer S, Schmiegelow M, Carstensen H, Nielsen LB, Muller J & Schmiegelow K (2003). Long-term neurological outcome of childhood brain tumors treated by surgery only. *J Clin Oncol* **21**, 1347–1351.
- Steinlin M, Blaser S & Boltshauser E (1998). Cerebellar involvement in metabolic disorders: a pattern-recognition approach. *Neuroradiology* **40**, 347–354.
- Tyrrell T & Willshaw D (1992). Cerebellar cortex: its simulation and the relevance of Marr's theory. *Phil Trans R Soc Lond B Biol Sci* **336**, 239–257.
- van Beugen BJ, Gao Z, Boele H-J, Hoebeek F & De Zeeuw CI (2013). High frequency burst firing of granule cells ensures transmission at the parallel fibre to Purkinje cell synapse at the cost of temporal coding. *Front Neural Circuits* **7**, 95.
- Widrow B & Hoff ME (1960). Adaptive switching circuits. Technical report. Stanford Electronics Laboratories, Stanford University, CA, USA.
- Wolpert DM, Doya K & Kawato M (2003). A unifying computational framework for motor control and social interaction. *Philos Trans R Soc Lond B Biol Sci* **358**, 593–602.
- Wolpert DM, Miall RC & Kawato M (1998). Internal models in the cerebellum. *Trends Cogn Sci* **2**, 338–347.
- Zhang W & Linden DJ (2006). Long-term depression at the mossy fibre–deep cerebellar nucleus synapse. *J Neurosci* **26**, 6935–6944.
- Zhou H, Lin Z, Voges K, Ju C, Gao Z, Bosman LW, Ruigrok TJ, Hoebeek FE, De Zeeuw CI & Schonewille M (2014). Cerebellar modules operate at different frequencies. *Elife* **3**, e02536.

Additional information

Competing interests

None of the authors has a conflict of interest with the work presented here.

Author contributions

The study was performed at ATR Telecommunications Laboratory. T.D.S. contributed to conception, design, simulation, analysis, interpretation of data and drafting of the manuscript. O.Y. contributed to interpretation of data and drafting the work and revising it critically for important intellectual content. M.K. contributed to conception, design, interpretation of data, and drafting the work and revising it critically for important intellectual content. All authors have approved the final version of the manuscript and agree to be accountable for all aspects of the work ensuring that questions related to the accuracy or integrity of any part of the work are appropriately investigated and resolved. All persons designated as authors qualify for authorship, and all those who qualify for authorship are listed.

Funding

Support for this project was provided by Riken AIP, ATR and the USC Department of Biomedical Engineering. T.D.S. was supported by a travel grant from RIKEN AIP, Japan. O.Y. is supported by RIKEN AIP. M.K. is supported by ATR Telecommunications and RIKEN AIP. This research was partially supported by AMED under Grant Number JP19dm0307009.

Keywords

cerebellum, control theory, dynamics, feedback control, motor control, motor learning, neural control