A HYPOTHESIS OF THE OPERATION OF CEREBELLAR CIRCUITS WITH A FOCUS ON CIRCUITS INVOLVED IN AXIAL AND LIMB MOVEMENTS

by

MIKE GILBERT

A thesis submitted to the University of Birmingham

for the degree of

DOCTOR OF PHILOSOPHY

School of Psychology

College of Life and Environmental Sciences

University of Birmingham, UK

Date submitted: 17 January 2020

ABSTRACT

For most brain regions the only realistic approach to theory is a top down, systems view because the evidence is too sparse to 'join the dots'. For a number of reasons the cerebellum is probably the best-suited large brain structure to take a different, evidence-based approach. That is the approach here, combining evidence of anatomy, electrophysiology, molecular biology and behavioural conditioning studies with mathematical and computational modelling, to build a model of the way the cerebellum derives output from input. Modelling is used to test predictions and to generate hypothetical data not available with current experimental techniques, which feed back into the model. This allows modelled behaviour of different parts of the circuit to be tested against evidence of other parts. The focus is on circuits involved in control of axial and limb movements, although parts of an explanation are likely to be portable to other circuits (because cerebellar circuit wiring is modular). An important part of the proposals is that the functions of pattern recognition and output coding are separate. It is a function of recoding in the granular layer to turn input variables into independent (and fewer) internal variables. Independence means they can be used in different functions without mutual interference of the execution of those functions with each other. Pattern recognition determines which circuits have output and when, but does not code output. Instead, the response to a known pattern following training is permissive, creating a time window in which output cells are controlled ad hoc by internally generated information about movement. Recoding in

the granular layer, as well as having a long-suspected role in pattern detection, also has a role in control of output rates, by turning (what is from a cerebellar view) an arbitrary range and frequency distribution of input rates into a narrow range of granule cell rates with a fixed bandwidth and a frequency distribution with a fixed shape, so that the only functional variable of internal signals traffic at the scale of input to a Purkinje cell is the adjustable range.

ACKNOWLEDGEMENTS

I would like to thank my supervisor, Prof Chris Miall, without whom…. In my experience good help and advice is hard to find. I was fortunate.

I would also like to thank Alex Mathy for discussing some of the ideas at an early stage, Pete Holland and Olivier Codol for their comments on a draft of one of the early chapters, Orna Rosenthal for her comments on another one, and Joe Galea, Ned Jenkinson and Peter Hanson for their insight into the mysteries of journal publication and funding.

Finally I would like to thank Ali who helped me to speak and write more clearly.

On Ernest Rutherford and Niels Bohr. ‘Both regarded mathematics as an important tool…but never as an end in itself. Rutherford was fond of making disparaging remarks about theoreticians who were too attached to formal mathematics.’

‘Rutherford and Bohr’ in *Atomic Histories*, Sir Rudolf E. Peierls

TABLE OF CONTENTS

CHAPTER 2

CHAPTER 6

LIST OF FIGURES

CHAPTER 5

CHAPTER 6

LIST OF TABLES

CHAPTER 1

INTRODUCTION

1.1 Scope

This thesis is a theory of the cerebellum with the scope set out in this section and section 1.3. The work is confined to the internal operation of the cerebellum – how its repeating circuits, and therefore the cerebellum itself, derive output from input. The scope accordingly does not include, for the most part, where input is from or where output is sent. The aim is to explain the functional design of the basic wiring that circuits are thought to share – the repeating anatomy which, together with the functional division of the cerebellar cortex into microzones (Oscarsson 1979), provides the organisation which makes the cerebellar wiring scheme modular. It is likely to be important to understand the functional significance of circuit wiring at this fundamental level of organisation (Apps and Garwicz 2005, Ramnani 2006, Ruigrok 2011).

The explanation of circuit function is not the expansion or development of a single underpinning principle. Circuit design (it is proposed) is not based around a single idea but has a number of parts, none of which has precedence. Circuits (in this view) are not designed to exploit a central principal but are a composite of solutions to

different problems faced by different parts of the circuit, which may have been added piecemeal over evolutionary time.

A brief synopsis appears in the abstract and a summary appears in Chapter 6. The model is in irreconcilable conflict with the popular idea that the cerebellum implements a machine learning algorithm. It is not adapted from or aligned with existing models.

'Circuit' refers to repeating microcircuits and not to the wider (1-2 mm) longitudinal zones (described in Buisseret-Delmas and Angaut 1993) which have been shown in some cases to contain them. It remains unconfirmed whether the whole of the cerebellum is divided into microzones. Evidence is still limited to only certain regions (such as the C3 and C1 zones: Garwicz, Jörntell et al. 1998, and Cerminara, Aoki et al. 2013, respectively). While some aspects of circuit wiring are probably common to all circuits, cerebellar circuits are modified in different regions for different functions (Cerminara, Lang et al. 2015). For example, vestibular circuits lack deep nuclei, so that their output is carried by Purkinje cells. Vermal circuits responsible for muscle tone and state maintenance – holding still against resistance – may exert control by a sustained balancing act between antagonistic outputs (Shadmehr 2017). The focus here, in this thesis, is on circuits which are dedicated to control of axial and limb movement. This decision is partly because there are data available from the C3 region of the cerebellar cortex that are not available for other types of circuits (much of it from Lund University, referenced in the main text). So, while an explanation of

that evidence, argued across this thesis, may in parts apply to circuits generally, it is unknown how much functional organisation circuits share. (See generally Bengtsson and Jorntell in Apps, Hawkes et al. 2018 for a discussion.)

The scope does not include an explanation of other ways that the organisation of the cerebellum as a whole can be divided (Apps and Hawkes 2009). These include longitudinal banding defined by the expression of Zebrin II (Brochu, Maler et al. 1990), a metabolic enzyme, and other molecular markers which are co-expressed with Zebrin II, including the glutamate transporter EAAT4, which can affect the direction of plasticity at the parallel fibre-Purkinje cell synapse (Wadiche and Jahr 2005). Another example is the subdivision of the cerebellum into five or six transverse zones based on patterns of gene expression (Ozol, Hayden et al. 1999).

1.2 A revisionist approach

In a sentence, the present approach has been a non-selective collation of the evidence of anatomy, electrophysiology, molecular biology and conditioning studies, at cellular, circuit and behavioural level, re-ordered into a functional picture, using mathematical and computer modelling to test ideas and make predictions, and also to generate estimates of data that are not available with current experimental methods, which are used in the model.

All models of the cerebellum have problems. Briefly – using a broad top-down/bottom-up distinction (Medina and Mauk 2000) – top-down models simplify or do not include detailed physiological data, instead attempting to explain presumed functions of the cerebellum at a more abstract level, while bottom-up models, where physiology is more to the fore, are narrowly focussed at the expense of the wider picture (and other evidence). (See Llinás and Negrello 2015 which this paragraph roughly paraphrases on this point.)

Part of the motivation for this thesis is a belief that work on the cerebellum has reached a critical mass, so that it is possible to assemble the evidence into a functional mosaic which explains the principles of the operation of cerebellar circuits without sacrificing detail – so, without a scale preference. As a result it is unnecessary to take a selective view of the evidence, and possible to model the behaviour of networks without making unphysiological assumptions.

1.3 A note on structure

A research degree reporting experimental work would usually contain a literature review. The nature of the present work means it is not front loaded in this way. The structure of this document takes the form of five chapters (excluding this introduction), each in the form of a draft paper. The reason is that they were written in this form with the intention of publishing. While they have been adapted to a thesis

format, (few but) occasional references remain to the author as ‘we’ (because the papers will be submitted jointly with Professor Miall who supervised the work).

The chapters each discuss a different part of circuit design and function. While this covers a significant part of the operation of the basic cerebellar circuit, other parts are missing because of limits on space and time. There is little on basket cells and nothing on Lugaro cells, for example, or on the substantial dopaminergic or serotonergic input to the cerebellum, or the two inhibitory deep nuclear projections that terminate on Golgi cells ipsilaterally and in the inferior olive contralaterally, which form part of the largely closed circuit architecture discussed in section 1.5.

1.4 Anatomy in brief

Because they are written for publication each of the chapters includes in its introduction background relevant to that paper. Since they are also written to be part of this thesis every effort has been made to minimise duplication. A precis of anatomy and parallel fibre synaptic learning also appears here.

The majority of (glutamatergic) input to the cerebellum is from mossy fibres which terminate on granule cells in the inner layer of the cerebellar cortex, the granular layer. Granule cell axons rise into the outer layer of the cortex, the molecular layer,

where they divide to form parallel fibres, so called because they run parallel to the surface of the cerebellum and to each other, which make contact in passing on Purkinje cells and inhibitory molecular layer interneurons (MLIs).

Purkinje cells are organised functionally into long thin groups of several hundred cells – microzones (Oscarsson 1979). A microzone occupies a thin sagittal slice of the cerebellar cortex perpendicular to the surface, penetrated at right angles by parallel fibres, and measuring some 15 mm long x 50-100 μm wide (an estimate for the C3 region in cats: Dean, Porrill et al. 2010), although dimensions vary.

Purkinje cells each receive powerful contact from a single climbing fibre. Contact on all Purkinje cells in a microzone is from a small group of climbing fibres, originating on the contralateral inferior olive. These are connected by dendritic gap junctions that help to cause them to fire together (Long, Deans et al. 2002), so that climbing fibre input to a microzone is a broadside received by the whole population of Purkinje cells at the same time. In fact, microzones are defined by their climbing fibre input (Oscarsson 1979, Garwicz, Apps et al. 1996, Garwicz, Ekerot et al. 1998).

A microzone receives granule cell signals from a roomy area of the cerebellar cortex equal to the product of its length and the span of parallel fibres. A parallel fibre span of 6 mm is reported for cats, chickens and monkeys (Brand, Dahl et al. 1976,

Mugnaini 1983); 5 mm is reported in rats (Harvey and Napper 1988, Harvey and Napper 1991).

A pattern of parallel fibre input to a Purkinje cell which is repeatedly twinned with climbing fibre input causes long-term weakening of transmission at the corresponding set of active parallel fibre-Purkinje cell synapses (Hansel, Linden et al. 2001, Ito 2001, Qiu and Knopfel 2009). The same conditions train strengthening of synapses onto inhibitory interneurons that inhibit Purkinje cells (Jörntell and Ekerot 2003, Rancillac and Crépel 2004, Smith and Otis 2005, Jörntell and Ekerot 2011). Following training circuits acquire a response to a repeat of parallel fibre input in the same pattern at least partly as a result of synaptic plasticity. The response is a transient reduction of the Purkinje cell firing rate, which can be a full pause (Jirenhed, Bengtsson et al. 2007, Rasmussen, Jirenhed et al. 2008).

Feed-forward inhibition of Purkinje cells by MLIs contributes to the pause. MLI morphology depends on the depth of the cell body in the molecular layer. There is continuous variation of phenotype from the outer to inner molecular layer. Within that there are distinct characteristics at different levels, which are expressed most at that level and less at others. There is not agreement whether these express different genotypes or are just steps in graded variation. It has become typical for MLIs to be divided into stellate cells, which occupy roughly the outer two thirds of the molecular layer, and (in species that have them) basket cells, which occupy the inner third.

Stellate cells at outer level have a short main axon which does not extend beyond the range of the dendritic field. At mid level the main axon is longer, follows a horizontal sagittal course and gives off more and longer collaterals. There is a more modest increase in the size of the mid-level dendritic field. Stellate cell dendritic and axonal fields are both severely flattened in the sagittal plane. Basket cells give rise to descending axon collaterals which make intimate contact with the Purkinje cell body and first axonal segment. Basket cell axons have a still greater sagittal range (of what may be around 10 Purkinje cells, though estimates vary) and give off mediolateral side branches, so that the basket cell axonal field is not planar. (This description is derived from: Eccles, Ito et al. 1967, Palay and Chan-Palay 1974, Paula-Barbosa, Tavares et al. 1983, Sultan and Bower 1998. A detailed description appears in Chapter 5.)

1.5 Closed circuit architecture

Olivary cells contact Purkinje cells with which they co-terminate collaterally in deep nuclei (Llinás, Walton et al. 2004) (see also Ruigrok 1997, Bengtsson and Hesslow 2006). Collaterals are reciprocated by an inhibitory nuclear projection to the contralateral olivary complex (Hesslow and Ivarsson 1996, Bengtsson, Svensson et al. 2004), so that they complete what are thought to be closed circuits (Shinoda, Sugihara et al. 2000, Pijpers, Voogd et al. 2005, Bengtsson and Hesslow 2006, Fujita and Sugihara 2013). Functional closed-loop organisation has been reported (Chaumont et al 2013). Discrete regions of the inferior olive swap reciprocal

connections with (also discrete) functional groups of nuclear cells (Llinás et al 2004; see also Ruigrok 1997; Bengtsson and Hesslow 2006).

The same discrete cluster of olivary cells may project to two or more microzones which in turn project to the same deep nuclear cells (termed a multizonal microcomplex: Apps and Garwicz 2005). So that climbing fibres co-terminate in deep nuclei with the output of the region of the cortex where they send the same signal – that is, circuits are still closed – even where they target more than one microzone. ‘CF [climbing fibre], olivonuclear, and corticonuclear axons project in a map-like fashion: neighbouring neurons in one region (olive, cortex, or deep nucleus) innervate neighbours in the other two regions’ (Ozden, Dombeck et al. 2012 p.14).

‘Mossy fibers projecting to a certain group of Purkinje cells through the granular layer also project to the deep cerebellar nucleus neurons receiving input from those Purkinje cells’ (D’Angelo et al 2013 p.9 citing Voogd et al 2003; see also Ruigrok 2010). Mossy fibres and climbing fibres which co-terminate in deep nuclei also co-terminate in the same sagittal zone of the cerebellar cortex (Ruigrok 2010).

The evidence for the strict topography and ubiquity of the olivo-cortico-nucleo-olivary loop is incomplete, and it may be a simplification. For example, nucleo-olivary projections in some nuclear sub-regions project to the ipsilateral as well as the contralateral inferior olive (Ruigrok and Voogd 1990, Teune, van der Burg et al.

2000), and a cortical region – and even a single Purkinje cell (Wylie, De Zeeuw et al. 1994) – may project to more than one nuclear group. Nonetheless, cortical areas that project to the same nuclear targets are innervated from single olivary regions (Pantò, Zappalà et al. 2001), and the general principle seems to hold (for a review see Uusisaari and De Schutter 2011).

FIGURE 1.1

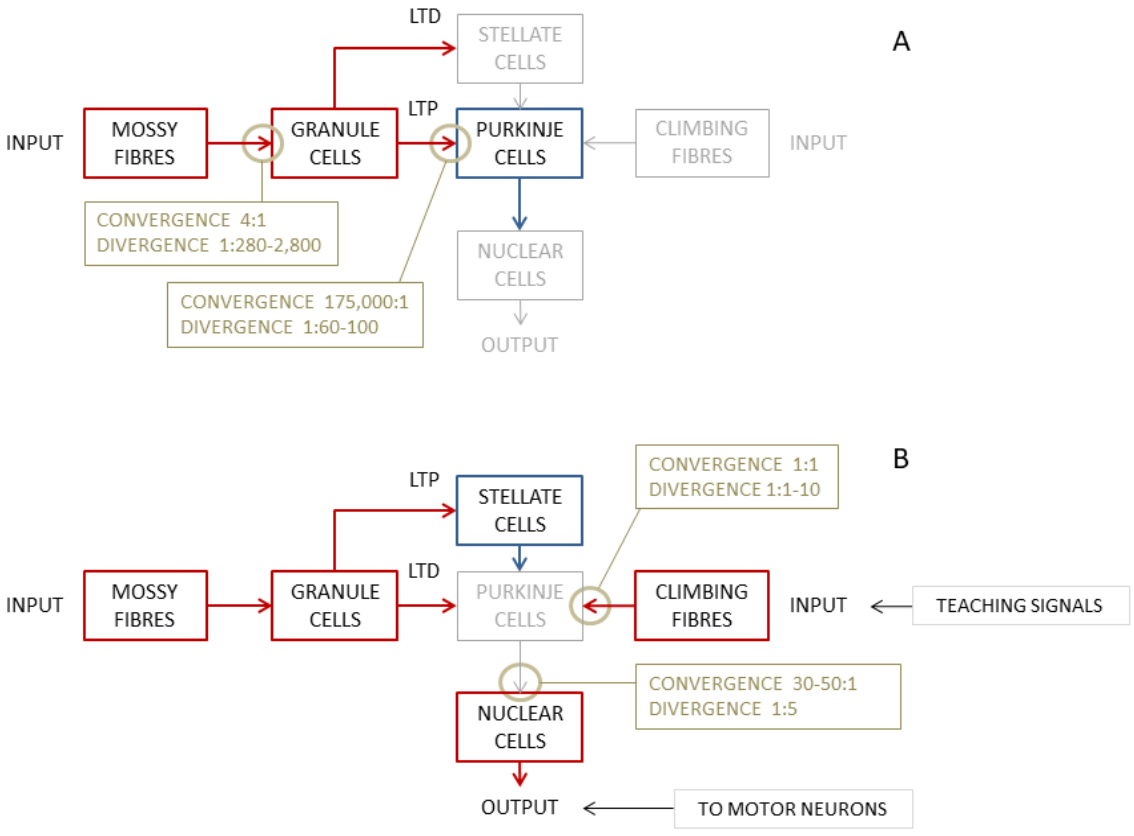


FIGURE 1.1 Schematic of cerebellar circuitry. A: Prior to training, strong Purkinje cell firing inhibits nuclear cells, the output cells of the circuit. Red boxes and arrows: active glutamate cells; blue: active GABA cells; grey: silent neurons. **B:** Following

training, firing of Purkinje cells is weakened or suspended in the conditioned response, because of LTD of parallel fibre inputs and LTP of stellate cells, releasing nuclear cells from inhibitory restraint.

1.6 Theoretical context

This section is a background review of theoretical models of the cerebellum. Attempts to explain cerebellar function have been dominated by a motor learning role for climbing fibre signals since its introduction (Marr 1969), and learning models are dominated by the perceptron (Albus 1971, Brunel, Hakim et al. 2004) and adaptive filter models (Fujita 1982, Dean, Porrill et al. 2010). Chapter 2 is a critique of the adaptive filter model. It includes a brief synopsis of the Marr and Albus models and a more detailed description of those models appears in Appendix 1. An account of the perceptron and adaptive filter models is therefore not duplicated here. Experimental support on some key points is mentioned briefly.

1.6.1 Evidence that helped to establish learning models

Experimental evidence appeared to verify predictions made by learning models, making them popular. They gained in credibility from confirmation of the prediction that pairing of parallel fibre signals with climbing fibre signals teaches long-term

depression of the parallel fibre Purkinje cell synapse (Ito 1989, Hansel and Linden 2000), and from tracing of the conditioned eyeblink reflex pathway through the cerebellum (McCormick and Thompson 1984, Christian and Thompson 2003). The idea that climbing fibre signals are driven by errors, first proposed by Albus (Albus 1971), became embedded when it was reported that incorrect reaching movements by monkeys were associated with an increase in complex spikes, known to be a reliable indicator for climbing fibre discharge (Gilbert and Thach 1977). (See Streng, Popa et al. 2018 for a review of the evidence for the idea that complex spikes signal errors, which concludes they probably don't.)

1.6.2 Oscillation/timing models

Braitenberg proposed that slow transmission by parallel fibres would allow temporal associations within a time window – experimental evidence contradicted his original proposal (Braitenberg and Atwood 1958) but he later proposed a revised form (Braitenberg, Heck et al. 1997), which received subsequent refinement (Sultan and Heck 2003). The original idea was that parallel fibres provided ‘delay lines’ with Purkinje cells only firing when the received a coincident parallel fibre volley and climbing fibre signal, thus providing output timing. An example was the time-delayed command sent to antagonistic muscle to terminate a movement. In its later form (Braitenberg, Heck et al. 1997) slow transmission of parallel fibres allows the cerebellum to detect sequences of inputs and respond with a sequence of outputs. An input sequence of mossy fibre signals that passes along a folium causes a robust

volley, or ‘tidal wave’, of parallel fibre signals travelling in the same direction. The amplitude of the wave is strongest when input travels at the same speed as the conduction velocity of parallel fibres and weaker when it is faster or slower, triggering a corresponding sequence of Purkinje cell activations.

Among problems with the delay lines idea is that the cerebellum is seen as ‘a large collection of individual lines’ (beams), requiring a narrow focus of granule cell activations travelling in the mediolateral direction. This has not been reported and is inconsistent with the sagittal alignment of mossy fibre terminal branches. Also, despite the slow conduction velocity of parallel fibres they are only capable of delays of a maximum of a few milliseconds, so could only code sequences that are very rapid and very short, too short to code most movements on a typical behavioural timescale.

Llinás rejected learning theories which he thought lacked evidence of long-term modification of Purkinje cell activity, and proposed instead that climbing fibre signals acted as a kind of pacemaker which provided coordination in the control of movements (Llinás and Welsh 1993). This was based on electrophysiological investigation of the inferior olive (Llinás and Yarom 1981), electrical coupling between olivary cells (Llinás, Baker et al. 1974), and the effect on Purkinje cells (Sugihara, Lang et al. 1993). According to the hypothesis, subthreshold oscillations by olivary cells and electrotonic coupling between them is seen as suggesting that olivary cells provide a timing device, or pacemaker, used to control onset time of

movements. The uniform length of olivary cells is cited as evidence that the rhythm of olivary timing is conveyed with precise synchrony to Purkinje cells, and through the cerebellum to the rest of the motor system. Selection of cerebellar targets depends on how Purkinje cells are functionally grouped by their climbing fibre input, which in turn depends on how olivary cells are gap-junction grouped. Olivary grouping is controlled by inhibitory nucleo-olivary feedback which, by selectively decoupling cells, create discrete clusters of coupled cells, each oscillating to a slightly different rhythm. Different patterns of muscle activation are produced by changing group boundaries.

The pacemaker hypothesis most notably fails to take account of the fact that parallel fibres are also ‘true afferents’ (as Llinas describes the relationship of climbing fibres to Purkinje cells), so that excitatory input to Purkinje cells – and therefore also the effect on proposed olivary functional clustering of nucleo-olivary feedback – is from more than one source, with an integrated outcome. An effect of parallel fibres must explain by extension an effect of other cells they act through or which act on them, such as MLIs. Later work – for example on olivary coupling (De Gruijl, Bazzigaluppi et al. 2012) gap junctions generally (Vervaeke, Lorincz et al. 2012) and complex spike synchrony (De Gruijl, Hoogland et al. 2014) – has not improved experimental support.

1.6.3 Gain change and Purkinje cell bistability

Lesions or inactivation of the inferior olive result in a large increase in simple spike rates (Cerminara and Rawson 2004, Zucca, Rasmussen et al. 2016), suggesting an ongoing role of climbing fibre signals in cerebellar function. An early idea was that climbing fibres adjust gain control such that the effect of peripheral input to a Purkinje cell depends on timing relative to a climbing fibre signal, either increasing or decreasing responsiveness, as shown in cats following removal of the cerebrum (Ebner, Yu et al. 1983, Ebner and Bloedel 1984). The same timing dependent sensitivity reported during locomotion in decerebrate ferrets led to the dynamic selection hypothesis, the idea that climbing fibres selected Purkinje cells for sensitisation in a performance-related manner (Lou and Bloedel 1992). However, sensitisation appears to be absent in awake rabbits (Simpson, Wylie et al. 1996), suggesting evidence is inadequate in intact, behaving animals.

Purkinje cells were subsequently reported to switch between up and down states, where they were more or less readily activated, leading to the 'bistability' hypothesis. 'However, CS [complex spike]-coupled changes in SS [simple spike] firing rates are only prominent in reduced or anaesthetised preparations …[There are some discrepencies between studies, but,] in the awake animal, climbing fiber input appears to have a very restricted effect on SS firing rate' (Streng, Popa et al. 2018 p.739).

1.6.4 Internal models

A note on internal models. The models mentioned above are an attempt to explain how the cerebellum converts input into output. That is, they are concerned with what happens inside the cerebellum. The forward and inverse models (together ‘internal’ models) are concerned with a function they propose the cerebellum may have in a wider scheme of motor control. In the forward model the cerebellum learns to convert a motor command into a prediction of its sensory consequences, that is, the sensory feedback the movement is expected to generate (Wolpert and Miall 1996). The inverse model inverts this idea to compute the motor command needed to generate a desired state (Wolpert, Miall et al. 1998). Both are examples of a systems analysis – the cerebellum is represented as a functional stage in a chain (or loop) of connected parts of the motor system as a whole – rather than an attempt to explain internal cerebellar function in physiological detail, so that internal models have in this way different scope to model proposed here.

GLOSSARY

All the chapters use defined terms. A term used in more than one chapter, as most of them are, has the same definition in all of them.

Trained and untrained states

Training is with a conditioning protocol, or direct stimulation of paired input that simulates a conditioning protocol, so that a ‘trained’ and ‘conditioned’ microzone (or circuit) are synonyms. A circuit may be trained to many different patterns. So a circuit can be trained one moment and untrained the next, depending on the input it receives. Single cells can be trained (in this sense). A ‘match’ is a learned pattern of parallel fibre activity (therefore in a ‘known’ pattern).

Circuit

‘Circuit’ refers to the wiring described in section 1.5 above. A circuit may contain a single microzone or more than one (a multizonal microcomplex: Apps and Garwicz 2005). Where a circuit contains more than one microzone, all of them must receive a match for the circuit to be trained.

Pattern

A ‘pattern’ refers to a binary pattern of active cells, some on some off. It does not mean, for example, fluctuations in the firing rate of a cell, or a changing pattern of active cells. References to the ‘amount’ or ‘level’ of activity are to the number of

active cells, and more precisely the proportion that are active, roughly equating to density.

File

A file of Purkinje cells is a mediolateral row that spans a microzone. The estimated dimensions of a microzone in the C3 region (Dean, Porrill et al. 2010) would mean that a file contains around 5 Purkinje cells.

Input

An ‘input’ is synaptic contact from an active cell, and not mere innervation. In chapter 3 (only) the meaning is narrower: ‘input’ and ‘output’ are input and output of the granular layer unless otherwise stated.

Nuclear cells

References to ‘nuclear’ cells are to excitatory projection neurons that carry the main output of the cerebellum unless otherwise stated (also therefore ‘output cells’).

Beam

A ‘beam’ is a region defined in section 3.3.7 of chapter 3 or a permutation of active parallel fibres in a region of that size.

Horizontal and vertical

Horizontal means parallel to the surface of the cerebellum, usually in the direction of the long axis of a microzone. Vertical means in the direction orthogonal to the surface of the cerebellum.

CHAPTER 2

REVIEW OF THE PHYSIOLOGICAL FEASIBILITY OF THE ADAPTIVE FILTER MODEL

2.1 Introduction

The dominant class of theories of cerebellar function are learning models and the dominant models in that class are the perceptron (Albus 1971, Brunel, Hakim et al. 2004) and adaptive filter (Fujita 1982, Dean, Porrill et al. 2010) models, both developed for applications in physical sciences and later adapted to the cerebellum. Both propose that the cerebellum implements a supervised learning algorithm which uses iterative adjustment of parallel fibre synaptic weights either to generate binary output (Albus 1971, Brunel, Hakim et al. 2004) or to reduce deviation from desired output (Fujita 1982, Dean, Porrill et al. 2010). Output is driven by input in remembered patterns of parallel fibre activity – 'remembered' in the sense that transmission of input signals is modulated by training-adjusted synaptic weights. Weights are determined by an algorithm (the algorithm depends on the model). Supervision is provided by climbing fibres which typically signal errors. Purkinje cell firing is treated as output of the system. Output is learned, that is, determined by training and not input rates – training displaces the naïve response to input rates. The unit of learning is a Purkinje cell – storage and expression of pattern memory is independent of learning by other Purkinje cells.

Support for the adaptive filter model is appraised in this chapter. The focus is on physiological implementation – what and how good is the evidence that the cerebellum provides the hardware to put into operation the proposals made by the model, and that this is what it is designed to do, and does? The conclusion is that the model conflicts with the evidence on fatal points.

Learning theories of the cerebellum share the idea that climbing fibres provide an instruction signal that teaches modification of parallel fibre synaptic transmission strength. In Marr's original proposal (Marr 1969), learning was the result of coincidence - patterns of active parallel fibres that are repeatedly paired with climbing fibre input to the same Purkinje cell 'at about the same time' induce strengthening of the corresponding pattern of active parallel fibre synapses. This trained a spatial pattern of operational synapses. A Purkinje cell fired if a sufficient proportion of a learned pattern was subsequently received at trained synapses. The proportion needed to be high enough to overcome an inhibitory threshold set by feed-forward inhibition of Purkinje cells by molecular layer interneurons which sampled the same parallel fibre activity. A Purkinje cell would fire if excitation was greater than inhibition. Thus Purkinje cells learn to respond selectively to granule cell representations of mossy fibre activity.

This mechanism could identify a spatial pattern of activity but didn't explain how output rates were derived. Rather, output is driven by input in a known pattern, so implicitly at rates that vary with input rates, but Marr does not explore the effect on

parallel fibre-Purkinje cell synaptic transmission of a graded result of learning, assuming instead (rather than proposing) that transmission strength is binary.

Marr is associated with the perceptron algorithm previously developed by Rosenblatt (Rosenblatt 1961) but in fact does not mention it (or reference Rosenblatt). Albus, however, proposes the possible suitability of Purkinje cells to classify patterns on a perceptron model if state space can be sufficiently expanded (Albus 1971, from p.33 onwards). Albus predicts the effect of learning is a pause in Purkinje cell firing because training weakens transmission – like Marr he treats the Purkinje cell rate as synonymous with output. Also like Marr, output is cued by a learned spatial pattern of parallel fibre activity, and supervision of learning is provided by a climbing fibre signal. But the learning rule – coincidence for Marr – is more complex because the perceptron algorithm detects patterns of input rates – not just the pattern of active cells but the rates they each fire at (at times selected by a climbing fibre signal) – and generates binary output: all learned input patterns generate output in one of two (unspecified) forms.

Albus does not include a proposal of a physiological mechanism that implements the algorithm (such as a mechanism that would inform learned synaptic changes that control or modify transmission). Indeed the discussion is confined to an argument that the cerebellar cortex could in theory implement the algorithm (as opposed to a reason or other evidence that it does). Albus does not propose, for example, what function classification would have in any specific or real case, or how binary output

codes continuously variable motor commands, or how the number of input permutations (known or not) is capped (necessary because unknown permutations cause arbitrary output, and on the other hand if too many are learned – so they are all known but the system exceeds capacity – it compromises the fidelity of the response to all of them). (See Appendix 1 for a more in-depth description of the Marr and Albus models.)

The idea that the cerebellum may be an analogue of an adaptive filter was first proposed by Fujita (Fujita 1982). Adaptive filters are used in physical sciences to modify signals. Fujita incorporates climbing fibre supervision of parallel fibre synaptic learning into a model that handled analog signals. Fujita considered that in theory Purkinje cells could provide the filter in an adaptive filter model of cerebellar function, which converted analog input into a 'desired' (also analog) output signal. In Marr's model output was episodic, in the perceptron model it was binary; the adaptive filter model instead converts a time-varying input signal into an also continuous (but different) output signal. The reason for the conversion is to turn raw mossy fibre signals into error-free motor output.

Fujita hypothesised that learning at the parallel fibre-Purkinje cell synapse is under tuition of a variable rate of climbing fibre discharge. Unlike Marr and Albus, climbing fibre signals are not discrete events, but also time-varying. The purpose of synaptic plasticity is to iteratively adjust synaptic weights so that the postsynaptic effect taken together generates the 'correct' (thus learned) output. The adjustment is calculated at

each synapse with a mathematical function – such as the covariance learning rule (Sejnowski 1977). Like Albus, the proposal does not include a physiological mechanism that implements the function. (Fujita himself knew he made unphysiological assumptions.) Thus ‘a PF [parallel fibre] signal that is positively correlated with an error signal has its weight reduced (through LTD), whereas a signal that is negatively correlated with an error signal has its weight increased (through LTP)’ (Dean, Porrill et al. 2010 p.33).

Since Fujita there has been an expanding theoretical literature. A problem of evaluating computational models which have had multiple modifications and patches is that they have a hydra-like immunity – none of them is individually necessary to claim the class is viable. But because ultimately all forms must be implemented at physiological level, shared practical features, the chassis that defines the class (because the model claims it has them), can be checked against the evidence. A list includes: a preprocessing stage that converts raw input into a form received by the filter; analog input; individually adjustable synaptic transmission of parallel fibre signals; sufficient memory; appropriate supervision; and conversion of post-filtered signals into output of the system.

2.2 Arguments against

2.2.1 Unevidenced assumptions

Fujita's model 'is based on a number of assumptions and simplifications' (Fujita 1982 p.202). Some are: an unphysiologically compartmentalised (i.e. closed) system – he models one such compartment; the whole population of granule cells of a compartment (1,000s) all receive the same mossy fibre signal, and provide input to a single Purkinje cell (with interference from no other input), and receive inhibition from a single Golgi cell;[1] Golgi cells act as 'leaky' integrators in order to provide a way that recoding causes phase shifts – phase shifts are one of the ways that non-biological applications of the adaptive filter pre-process raw input signals; stellate cells and basket cells are functionally equivalent; all parallel fibres that contact a Purkinje cell also contact basket and stellate cells that contact the same cell – thus, each parallel fibre exerts an effect via a direct and indirect path (and there is no interference from other parallel fibres that contact the same basket and stellate cells); parallel fibre-Purkinje cell synaptic weight adjustment implements an algorithm; the result following training is a correct motor command; Purkinje cell firing is synonymous with output of the cerebellum. Importantly, these were not assumptions in the sense of filling a gap in the evidence or making a prediction, they are for convenience.

[1] This compartmental arrangement, where a single Golgi cell inhibits all granule cells without significant overlap or gaps between compartments, was incorrectly reported by: Eccles, J. C., M. Ito and J. Szentágothai (1967). The cerebellum as a neuronal machine. Berlin, New York etc., Springer-Verlag.

2.2.2 Non-analog granule cell signals

Studies that include granule cell recordings from awake, behaving animals are not numerous. However, available recordings suggest that granule cells often fire in short bursts at high instantaneous frequencies. For example, vestibular granule cells fire in bursts of around 8-40 ms in rabbits on a rotating turntable (5-18 spikes per burst at around 530 Hz: van Beugen, Gao et al. 2013). In adult head-fixed mice on a cylindrical treadmill, recorded in the apex of lobule V, 'movement-evoked spiking in individual GCs [granule cells] was highly variable, with some GCs displaying continuous high frequency firing, while others displayed intermittent spike bursts' (Jelitai, Puggioni et al. 2016 Fig.3d; Supplementary Table 4; modal no. of spikes/burst 3, range 3-168, mean intraburst frequency 94 Hz, mean burst duration 76 ms). The significance is that most recorded granule cells do not provide *analog* forward transmission of mossy fibre signals.

Moreover, reported sensory-evoked mossy fibre signals are not themselves analog signals (Jörntell and Ekerot 2006, Rancz, Ishikawa et al. 2007), although recordings may not be representative of activity in freely-behaving animals. For example, some skin receptors, which fire in bursts under experimental stimulation, may provide 'cutaneous proprioception' during movement (Jörntell and Ekerot 2006), meaning: distortion of the skin during movement, especially cyclic movement, may cause some mechanoreceptors to generate a time-varying signal.

Burst firing of only some inputs to a granule cell may preclude analog firing. To fire, a granule cell is thought to need mossy fibre input to at least 3 of its (average) four dendrites (Jörntell and Ekerot 2006, Billings, Piasini et al. 2014, Cayco-Gajic, Clopath et al. 2017), which each receive contact from a single mossy fibre. So, a granule cell that does not receive at least 3 analog inputs and receives a mixture of analog and burst inputs only meets threshold for the duration of the shortest input burst, and therefore itself (burst) fires only for that duration.

2.2.3 The restricted bandwidth of climbing fibre rates

Both background and maximum firing rates of olivary cells are unusually low. In ‘both anaesthetized and awake animals [olivary neurons discharge] about once or twice per second’ (Simpson et al 1996 p.369). The maximum seems to be around 10 Hz (reported, for example, when nociceptive stimulation is applied: Ekerot et al 1987). 'Olivary neurons are refractory for approximately 100 ms following discharge, which limits their ability to fire at high rates' (Gibson et al 2004 p.214). To teach a graded or even varied range of Purkinje cell rates climbing fibre rates would themselves need to be graded or varied. The restricted range of variation argues against instruction coded in the climbing fibre rate.

2.2.4 Climbing fibre rates may be undetectable

The low climbing fibre rate, as distinct from the narrow range of rates, comes with other problems. First, it is not capable (even hypothetically) of transmitting rate coded information at more than slow speeds, because the elapse of sufficient time is necessary to 'count' enough discharges to infer a rate. Even then it would depend on the assumption that the rate in that period was functionally constant, or put another way that nothing coded in the rate changed faster than the time needed for an inference. This would make it far too slow to code information at behavioural speeds. Second, in any event the physiological equivalent of an inference – integration by the postsynaptic cell – may be unavailable, because the postsynaptic effect of climbing fibre input to Purkinje cells is temporally segregated. The rest of this section is about that.

Climbing fibres discharge in short bursts at rates > 250 Hz (Maruta, Hensbroek et al. 2007, Mathy, Ho et al. 2009) in groups of 1-6 spikes (average 2-3) (Mathy, Ho et al. 2009, Bazzigaluppi, De Gruijl et al. 2012), so that an average cluster has a maximum duration of 4-8 ms (i.e. the time between the first and last spikes at 250 Hz), and the maximum duration is 20 ms, with a group of 6 spikes. So bursts are invariably well separated even at the fastest olivary discharge rate (of ~10 Hz, which is atypical and possibly toxic).

Climbing fibre discharge is signalled to the site of synaptic learning (the distal branches of the Purkinje cell arbour) by elevation of intradendritic calcium. Transient postsynaptic calcium elevation is very rapid (a few ms) and all but simultaneous in all dendrites (Kitamura and Häusser 2011). A Purkinje cell receives all its climbing fibre input from the same cell, so that input is invariably a synchronised volley. Calcium transients in spines on distal branchlets, to which parallel fibre synaptic contact is confined, decay faster (179 ± 24 ms) than in dendrites (in anaesthetised rats: Kitamura and Häusser 2011). Thus postsynaptic integration of internal signals that inform synaptic learning are separated up to climbing fibre rates of around 5 Hz, and never sum across more than two transients. So in practice Purkinje cells cannot 'decode' the climbing fibre rate, i.e. they receive an invariant effect up to around 5 Hz and at rates between 5 and 10 Hz would need to infer rate coded information from calcium dynamics as a function of the interval between a single pair of discharges. Even at high rates the interval between two discharges is the duration of more than half a mouse step cycle (around 200-300 ms: Sarnaik and Raman 2018).

Finally, while the timing of a climbing fibre transient is exactly synchronised in all dendrites, peak amplitude is variable in different branches. Variability is reduced, but by no means eliminated, by blocking inhibitory synaptic transmission (Kitamura and Häusser 2011). The authors propose that climbing fibre signalling may be regulated branch-by-branch by GABAergic synaptic inhibition by interneurons. An alternative (and our) view is that there is no evidence that differences are regulated, and differences in peak amplitude in different branches are random. If so, the amplitude of integrated calcium in a branch – should consecutive transients overlap – varies

partly at random, and not in a climbing fibre rate-dependent manner, because the sum of calcium does not vary only or reliably with the time between peaks.

2.2.5 Climbing fibre instruction trains LTD, not bi-directional plasticity

The direction of parallel fibre-Purkinje cell plasticity is thought to depend on the level of postsynaptic calcium elevation. Levels generated by parallel fibre synaptic activation alone induce long-term potentiation (LTP). Higher levels caused by near coincidence of parallel fibre activation and a climbing fibre calcium transient induce long-term depression (LTD) (for example Coesmans, Weber et al. 2004, for a review: Jörntell and Hansel 2006). Paired activation of parallel fibres and climbing fibres is consistently reported to teach LTD, and not a bidirectional, graded outcome. This well-established finding does not support a bi-directional outcome of paired input. The next two paragraphs consider what might seem on the face of it to be exceptions, where training with paired input does not teach LTD.

There is a parallel fibre frequency threshold for LTD induction of around 40 Hz (Bidoret, Ayon et al. 2009, because this is necessary to recruit presynaptic NMDARs), so that at lower frequencies paired climbing fibre activation does not induce LTD. This is thought to act as a high pass filter, so that learning is limited to paired parallel fibre bursts. It functions, that is, as a threshold – set where, for

induction of LTD, a parallel fibre must spike at least twice within a short time window – and is not otherwise rate sensitive (either parallel fibre or climbing fibre rates).

It has been reported that a minimum of 3 spikes in a climbing fibre burst is necessary to teach parallel fibre-Purkinje cell LTD (Rasmussen, Jirenhed et al. 2013). There is no reported dependence of the outcome on the parallel fibre rate (or other parallel fibre signal coding), or other dependence of the outcome on the climbing fibre rate. This would suggest again a threshold, set at the minimum spike number for induction of LTD, rather than providing functional flexibility, in that sense doing the same job as the high pass filter.

2.2.6 The widespread view of instruction signals – that they are graded – is contradicted by the evidence

It has been suggested that teaching signals vary in their discharge signature (as opposed to discharge rate) so that they have an effect that depends on the signature, and that the all-or-nothing model of olivary firing has been consigned to ‘history’ (Hansel 2009 p.309, Najafi and Medina 2013, Rasmussen 2019). There is an alternative view. Olivary discharge is typically in short, high frequency (> 250 Hz) bursts (Maruta, Hensbroek et al. 2007, Mathy, Ho et al. 2009), as noted. Recorded at the soma the signature is a strong spike followed by a variable number of smaller spikes or wavelets (Armstrong and Harvey 1966, Crill 1970, Armstrong and Rawson

1979). All are reflected as full spikes in the axon (Mathy, Ho et al. 2009). The average number of spikes in a cluster is 2-3 and the range is 1-6 (Mathy, Ho et al. 2009, Bazzigaluppi, De Gruijl et al. 2012). Spikes are generated in the initial axonal segment, causing the effect measured at the soma. Some axonal spikes fail to propagate far (>125 μm), so that what is initially a group of 4 (say) can become a group of 3, or two. The first spike always propagates, and the others propagate with variable probability (range p = 0.66-0.89) (Mathy, Ho et al. 2009) depending on their position in a burst. The number of spikes is further reduced by synchronisation of sub-threshold oscillations between dendritically gap-junction-connected inferior olive cells, thought to synchronise climbing fibre input to a microzone (Blenkinsop and Lang 2006). The amplitude of oscillations is larger between gap-junction-connected than non-gap-junction-connected cells, and the number of spikes is in inverse proportion to amplitude (Bazzigaluppi, De Gruijl et al. 2012). As a result we might expect the number of spikes in a synchronised burst to be mainly confined to a reduced range, that is, the reduction of the number by oscillation and non-propagation condense the functional range of variation (down from a maximum of only 6 in the first place).

The low number of spikes in a burst would seem (to us) too low for a functionally graded effect. The narrow range, and olivary refractory period, would seem rather adapted to restrict rather than to exploit variation. Moreover, the failure of spikes to propagate sometimes, seemingly at random, would indicate that in the (hypothetical) event that information was coded in the number of spikes initially generated, it is unreliably transmitted. The mean probability of transmission of the second, third,

fourth and fifth spikes in a burst is 0.805 (Mathy, Ho et al. 2009). That gives a probability in any given burst of five spikes of around 0.66 that at least one is not transmitted, and a probability of around 0.25 that at least two are not transmitted. With a burst of 4 the equivalent probabilities are 0.58 and 0.17, and so on.

The number of studies of olivary burst firing is relatively modest. Najafi and Medina (2013) (Najafi and Medina 2013) advise caution in the interpretation of the results. The ‘number of spikes per CF [climbing fibre] burst was quite variable from one burst to the next and always fell within the same limited range (1–6 spikes), regardless of condition or behavioral state. Therefore, the changes in burst size for any given situation were small (<1 spike per burst) and could only be detected in the average as a slight probability bias toward generating more bursts with many (>4) or few (1) spikes’ (Najafi and Medina 2013 p.4).

In fact, climbing fibre signals are resistant to variation. The discharge signature is stable – more precisely: equally variable in a stable range – whether depolarisation is just over threshold or stronger. ‘Successive responses evoked by identical stimuli had a variable number of secondary spikes. … A just-threshold stimulus was as likely to evoke responses of multiple spikes as a stronger stimulus. … [Both] the mean number of spikes and the interspike interval of the secondary spikes in the unitary response were independent of the stimulus intensity’ (Crill 1970 p.201). This would argue that information represented by the strength of excitatory input to the inferior olive is not represented in the signature of climbing fibre signals, even before spikes

are lost to transmission failure. (Recordings were made from the somata of olivary neurons, which very reliably reflect the initially generated number of axonal spikes.)

Variation of the instantaneous rate of spikes in a burst has not been reported and is not thought to be a vector for graded climbing fibre tuition. Interspike intervals are highly reliable, indicating ‘that the timing of spikes within a burst in the olivary axon is highly stereotyped…, with only the number of spikes varying’ (Mathy, Ho et al. 2009 p.392).

So, evidence is lacking that the climbing fibre discharge signature varies functionally. Functionally invariant bursts (and an information neutral discharge rate, discussed above) would mean the effect of an instruction signal is not time variant except in depending on whether or not there is one at all – the coincidence learning rule. If so, the rule is not covariance.

2.2.7 The effect on circuit output of pair-trained parallel fibre synaptic depression

The adaptive filter function of synaptic depression at the parallel fibre-Purkinje cell synapse is to weaken transmission of signals that cause errors. By the covariance learning rule, positive correlation of a parallel fibre signal with a climbing fibre signal teaches LTD. This ‘makes sense when correlations represent causal relations,

because in that case reducing the impact of PF signals that are correlated with an error signal will reduce the error itself' (Dean, Porrill et al. 2010 p.33). This is not the reported effect of learning. Depression of the parallel fibre-Purkinje cell synapse contributes to the learned reduction of the simple spike rate in the conditioned response (Jirenhed, Bengtsson et al. 2007, Rasmussen, Jirenhed et al. 2008). Depression of the rate causes nuclear cells to fire at higher rates (Telgkamp and Raman 2002, Pedroarena and Schwarz 2003, Telgkamp, Padgett et al. 2004) and may also gate direct mossy fibre input to nuclear cells, in circuits where nuclear cells receive mossy fibre collaterals. That is, parallel fibre-Purkinje cell depression causes or contributes to the learned response of the circuit, it does not weaken or block it.

2.2.8 Problems with storage: pair-trained parallel fibre-Purkinje cell synaptic weights are not graded

Separation of learned patterns is important to the adaptive filter model in order that a set of learned weights is pattern-specific – otherwise (if stored patterns overlap significantly) the response would receive interference from unrelated memories. (In the original Fujita model this is not an issue because he assumes that all granule cells receive the same mossy fibre signal – the facts are put aside to allow the model to work.) Transmission at operational synapses is iteratively adjusted so that input signals in a known pattern are passed through a corresponding set of graded synaptic weights. The idea is that the combined effect homes in on the desired modulation of firing by the postsynaptic cell: learned output. In fact, the large majority

of parallel fibre-Purkinje cell synapses – estimated to be as many as 85% – are silent (Isope and Barbour 2002). The adaptive filter explanation is that this is the predictable result of the proposition that most parallel fibre signals received by a Purkinje cell are noise in respect of the function that the circuit has, and therefore drive ‘erroneous output’, so triggering error signals that drive down synaptic weight to nil (Dean, Porrill et al. 2010 p.37). If so, the result would be that a small minority of synapses are ever functional, proportionately limiting pattern storage capacity. Even assuming sparse parallel fibre activity, such that a functional pattern contains 1% of parallel fibres that contact a Purkinje cell, any two stored patterns would overlap by an expected 6.7% (assuming the pattern of active cells in a recoded pattern is random, and that 15% of synapses are functional). It would only need 5 patterns to be stored for approximately 24.1% of each pattern to be made up of synapses which also belonged to at least one of the others, 10 for around 46.3% overlap, 20 for 73%, and so on. So, i) the adaptive filter prediction of graded synaptic weights is not true for the large majority of synapses; ii) the model is obliged (in order to explain this result) to propose (without evidence) that the large majority of synapses only receive noise; iii) storage is confined to a low fraction of synapses such that storage capacity is severely restricted.

2.2.9 The covariance rule is an assumption

The covariance learning rule (and for that matter the least mean squares rule) is a formula which adjusts (a parameter representing) synaptic weight to achieve a

prescribed result. The idea that learning has this outcome is not derived from evidence but simply a proposal – a stipulation of the model. There is no evidence that individual adjustments of parallel fibre synaptic weights are calibrated for a collective effect on output, or that this contributes to motor control. It is simply a mathematical way of deriving from many parallel fibre input rates a learned response (of a Purkinje cell) that depends on (hypothesised) variable characteristics of the climbing fibre signal. Also, there is no clear physiological proposal how such a rule would be implemented – for example, how it controls induction pathways – or basis in evidence for the assumptions it makes.

2.2.10 The motor error problem

Climbing fibre signals in the adaptive filter model are generated by behavioural errors. The model does not explain how error signals are appropriately triggered and coded (sometimes termed the motor error problem – how does an autonomous system know what correct motor output is?). Retinal slip may feasibly generate error signals in the vestibulo-ocular reflex (for example Porrill, Dean et al. 2004). However, floccular circuits lack deep nuclei, which contain the output cells in the large majority of cerebellar circuits. So that floccular error signals, if that's what they are, are weak evidence that climbing fibre signals code errors in other circuits (so other circuits may still have a motor error problem).

2.2.11 ‘Diversification’ by recoding conflicts with the data

The ‘data suggests that the granular layer transmits MF [mossy fibre] inputs with relatively modest alterations – certainly far more modest than those required by typical adaptive filter models’ (Dean, Porrill et al. 2010 p.38). So here also the evidence is in conflict with adaptive filter models.

2.2.12 The minimum unit of learning

Purkinje cells are organised functionally into groups – microzones – whose output is channelled down onto a much smaller, discrete group of nuclear cells which include the projection neurons that carry the output of the circuit. To date, there is no evidence that the output of a microzone (carried exclusively by Purkinje cells) to a nuclear group is internally organised (Bengtsson and Jorntell in Apps, Hawkes et al. 2018 p. 663). Each Purkinje cell makes contact on an average of around 4-5 nuclear cells (rats) and each nuclear cell receives contact from 30-50 Purkinje cells (mice) (Person and Raman 2012a) (both at random). It follows that if Purkinje cell rates are not coordinated nuclear cells each receive a random mixture of rates, and each receives a different mixture.

Fujita deals with microzone organisation (briefly, in his closing remarks) by saying ‘the “one Purkinje cell model” may be expanded into a primitive model of a microzone…although this model does not explain any more functioning of the cerebellum than a single cell model’ (Fujita 1982 p.203). That is, he simply proposes that Purkinje cells each work on the one-cell model but under tuition of the same climbing fibre input, with output signals from Purkinje cells ‘summed up possibly in the nucleus’. The output of the system is the Purkinje cell firing rate and this is derived independently for each cell, so that there is no clear explanation what is to be gained by being in groups. This is a problem generally for the adaptive filter explanation of function. For example, it does not explain how the covariance rule deals with a different effect at random on different nuclear cells of a particular permutation of Purkinje cell rates, or how randomly different firing rates of different nuclear cells in a functional group have a coordinated effect on their targets, or how the many variables of input to Purkinje cells – the number and pattern of active cells, the permutation of rates they each fire at, the relative timing that they are received, the range and frequency distribution of rates, and so on – are controlled or selected so as to generate a functionally integrated Purkinje cell firing across the population of a microzone.

2.2.13 An arbitrary effect of unknown patterns

It is not clear how operation on the adaptive filter model would prevent an arbitrary effect of an unknown pattern of parallel fibre input to a Purkinje cell. Assuming

recoded patterns are randomly decorrelated, so that transmission-capable synapses are randomly distributed, a random (because it is also decorrelated) pattern would fall on a ratio of operational to inoperational synapses that's close to their relative proportions (because the numbers are large; even sparse parallel fibre activity – 1% active – is sufficient). So a predictable fraction of an unknown pattern should be received at operational synapses (unless you make the physiologically baseless assumption that all input patterns are known), with an unlearned postsynaptic effect. This would be dysfunctional both because it is triggered at random and, in the adaptive filter model, because the naive response to input rates causes errors.

2.2.14 No explanation of pattern-blind MLIs

Paired input teaches LTP at the parallel fibre-MLI synapse which is reversed by subsequent training with unpaired parallel fibre-only stimulation (Jörntell and Ekerot 2003, Rancillac and Crépel 2004, Smith and Otis 2005, Jörntell and Ekerot 2011). The MLI dendritic field, like the Purkinje cell arbour, is severely flattened in the sagittal plane (Eccles, Ito et al. 1967, Palay and Chan-Palay 1974). Assuming that parallel fibre activity is randomly distributed, and contact by a parallel fibre on a stellate cell is at the same average number of synapses as on a Purkinje cell, the proportion of pair-trained parallel fibre synapses on an MLI should be around the same as on a Purkinje cell (that is trained to the same number of patterns).

As a result, the large majority are operational (by this reasoning). Indeed the probability distribution for input to operational synapses with a known and unknown pattern are so nearly the same that MLIs cannot discriminate between patterns to which a microzone is trained and patterns to which it is not. That is, the effect on firing of the postsynaptic cell is indistinguishable except for a slight probability bias. In the adaptive filter model learned input patterns directly drive output. The model does not include or explain the function and effect of MLIs that respond indiscriminately whether input is in a well known pattern or random (much less tell known patterns apart).

2.2.15 No role or explanation of erratic Purkinje cell spiking

Purkinje cells spike erratically, even during a movement cycle where the spike count in ~10 ms step-locked bins totalled across many cycles gives a smooth curve (Sauerbrei, Lubenov et al. 2015). Variability of individual spike timing of the same cell from one cycle to the next is wide. The interspike interval is unpredictable both from one interval to the next and in the same phase-locked bin in different cycles, though it reliably fits a smoothly changing probability of spiking. Accordingly the firing rate of a single Purkinje cell varies erratically from moment-to-moment even at the fastest theoretically detectable speed (extrapolated from the interval between two consecutive spikes). Output of the adaptive filter model is learned firing of a Purkinje cell. The model does not explain how trained synaptic weights code erratic spike

timing or how erratic spiking of Purkinje cells codes motor commands (or even codes rates at all on a behavioural timescale).

2.3 Discussion

Adaptive filters in physical sciences adjust the amplitude (voltage, for example) of selected parts of the input signal, such as selected frequencies. Adaptive filters are used, for example, to remove mains hum from a heart trace and feedback from an amplified acoustic signal. Fujita uses synaptic weight for the equivalent of an amplitude adjustment, to selectively modulate transmission, to change the effect of input on output frequencies, which is to say the effect of parallel fibre on Purkinje cell rates. It is a widespread (and well-funded) idea that AI may provide insight into brain function. The idea is that physical systems may provide an analogue for biological function. However, in physical sciences frequency and amplitude are different; in the cerebellum spike amplitude is constant: signal strength refers to firing rate. Without an explicit physiological explanation how parameters that describe non-biological signals correspond to parameters that describe nerve signals – and why they are functionally equivalent – it is not clear that the adaptive filter was ever a promising analogue for cerebellar function.

There is another difference between what adaptive filters do and the hypothesised function of the cerebellum-as-an-adaptive-filter. In the given examples of applications

of adaptive filters, the result is a clean output signal. It does not stop a patient flatlining or improve a talentless performance. In the cerebellar model the desired output *is* improved performance. There is no contaminant of the input feed that the instruction signal is calibrated to weaken. Instead, the desired output is performance enhancement. Control by parallel fibre rates is overridden by synaptic adjustments to give the correct output. So: there is a disconnect (unexplained difference) between what adaptive filters are designed for, and do, and the biological function it is claimed they have in the cerebellum.

Conclusion: the cerebellum is not an adaptive filter, for the reasons given in the main text that the evidence is weak, and that the justification is poor or missing for the idea that an adaptive filter, and machine learning generally, provide a useful analogue of cerebellar function. It is worth repeating that there may be more than one model of circuit function, because circuits with different functions may work in different ways. Possibly, the adaptive filter model is a better fit for floccular circuits – which lack deep nuclei – than for other circuits. Pooling the evidence from different functional regions of the cerebellum would amount to an assumption that they work the same way. It may be that in some ways they do, but to date it is unknown how, or how much. A discussion of circuit function may need to proceed region by region.

CHAPTER 3

A PHYSIOLOGICAL MODEL OF THE GRANULAR LAYER OF THE CEREBELLAR CORTEX WITH COMPUTATIONAL SUPPORT

3.1 Introduction

A model of a brain structure or network must explain how it works (how it derives output from input), what form output takes, and what the translation is for (what function output is fit for that input isn't). The cerebellum is widely thought to implement a supervised learning algorithm which uses iterative adjustment of parallel fibre synaptic weights either to generate binary output (Albus 1971, Brunel, Hakim et al. 2004) or to reduce deviation from a desired outcome (Fujita 1982, Dean, Porrill et al. 2010). Supervision is typically provided by error signals provided by climbing fibres. Circuit function on these principles assumes that output is driven by remembered patterns of active parallel fibres.

The granular layer is packed with granule cells which receive input to the cerebellum from mossy fibres. The role of the granular layer depends on the model. It has long been hypothesised that recoding of mossy fibre signals into granule cell signals sparsens and decorrelates the binary pattern of parallel fibre activity (which parallel fibres are active and which silent) (Marr 1969, Albus 1971). In service to this function, Golgi cells adjust the granule cell input threshold – the number of mossy fibre inputs needed to make a granule cell fire (Marr 1969, Albus 1971, Billings, Piasini et al.

2014, Cayco-Gajic, Clopath et al. 2017, Cayco-Gajic and Silver 2019). In doing so they provide homeostatic control of parallel fibre activity (with the result of keeping levels low). So control is by Golgi cell adjustments of the granule cell input threshold, output is sparse, and the function is to recode mossy fibre signals into a larger state space (to facilitate pattern recognition by Purkinje cells).

It may be, however, that it is unnecessary to assume that output is driven by learned patterns in order to find a function for learning. This chapter argues a separation of the cerebellar functions of pattern detection and circuit output rate coding – so: *output is not learned* – each concurrently using differently coded data which are independently variable, so that neither interferes with the other function. It is part of the function of recoding to make them independent. It is also part of the function of recoding to block an effect of other input variables, so that they do not interfere with either function.

The output of recoding is sparse and decorrelated, in agreement with existing models, increasing storage capacity. But the new proposal disagrees on the mechanism that self-adjusts the level of parallel fibre activity, and also that these are the only functions, and that the function is confined to pattern separation. Instead, self-regulation is by adjustments that maintain a *fixed* input threshold (and implements a loop of mutually-regulating probabilities for which a detailed physiological explanation is provided). The additional functions are: the elimination from internal traffic of redundant input variables and independence of operative

internal variables. The reason is to confine output of the circuit to the effect of operative variables, to permit a separation of internal functions without mutual interference, and to enable learning at microzone level.

Separation of functions is a key point. Learning models assume learning controls output rates, and that synaptic weights, used to control output rates, are controlled by an algorithm. The present proposal argues instead that there is no need to assume that control of output rates is by signals in a known pattern and no need for an algorithm. ‘Proposal’ (and ‘model’) both refer to the physiological hypothesis as well as the computational simulation of the proposed operation of regulation.

3.2 The mossy fibre-granule cell relay

This section models the numerical relationship of input to output of the granular layer but omitting the effect of Golgi cells (added in section 3.3). Input is measured in active mossy fibres, without considering their firing rate, and output in active granule cells directly driven by that input.

3.2.1 Anatomy

Mossy fibres end in a cluster of glomeruli (each a ‘terminal’) (Wu, Sugihara et al. 1999, Shinoda and Sugihara 2013). A mossy fibre axon may give rise to several terminal clusters, aligned in the direction of the long axis of a microzone (Wu, Sugihara et al. 1999, Sultan 2001, Sultan and Heck 2003, Shinoda and Sugihara 2013). Clusters are estimated to average around 7 terminals and to measure around 200 x 150 μm (sagittal x mediolateral) (Sultan and Heck 2003). Cluster fields vary in number (per cell and terminals per cluster) and size, and can be larger (Wu, Sugihara et al. 1999). It is estimated that 100 mossy fibres contribute input to a cluster field (Sultan and Heck 2003) so that a cluster field contains 700 terminals.

Granule cells are very numerous. Estimates vary of the number of granule cells that receive synaptic contact from a single mossy fibre terminal; the range of estimates is 10-100 (Billings, Piasini et al. 2014, Ritzau-Jost, Delvendahl et al. 2014). For example, 12 has been used for modelling on the basis that it is ‘consistent with estimates of 15-20 in monkey and cat (Eccles et al 1967)’ (Billings, Piasini et al. 2014 p.962). But Gao and colleagues (Gao, Proietti-Onori et al. 2016) report an average of 44.6 +/- 10.5. We use 50 (estimated for rats: Jakab and Hámori 1988).

Granule cells have 3-5 (average 4) dendrites. A granule cell is thought to receive contact from a different mossy fibre to each of its dendrites. (See Appendix 2 for a

calculation of the probability that a mossy fibre makes contact on the same granule cell twice.) It is of note (but incidental to recoding) that a pattern of mossy fibre input to the area of the granular layer that supplies parallel fibre input to a microzone is condensed into a parallel fibre pattern that occupies a smaller area (in the sagittal plane) by more than an order of magnitude (see Appendix 3).

3.2.2 Input:output ratios without Golgi cells

The numerical relationship of input to output of the granular layer is the result of the anatomical ‘rules’ that govern contact, together with the assumptions that mossy fibre contact on granule cells is at random in a cluster field, and that the probability that a mossy fibre makes contact on a particular granule cell is the same for all granule cells.

If 100 mossy fibres each contribute 7 terminals to a cluster field, a field contains 8,750 granule cells (see Appendix 4 for convergent estimates of the number of granule cells in a cluster field). Because input to each dendrite of a granule cell is from a single mossy fibre, the fraction of mossy fibres which are active (of the total number which contact the field) gives the probability that a particular dendrite of a particular granule cell receives input. For x active mossy fibres out of y which terminate in a field, the probability that a granule cell receives input to k out of n dendrites is

$$p = \frac{n!}{k!\,(n-k)!} * \left(\frac{x}{y}\right)^{k} * \left(1 - \frac{x}{y}\right)^{n-k}$$

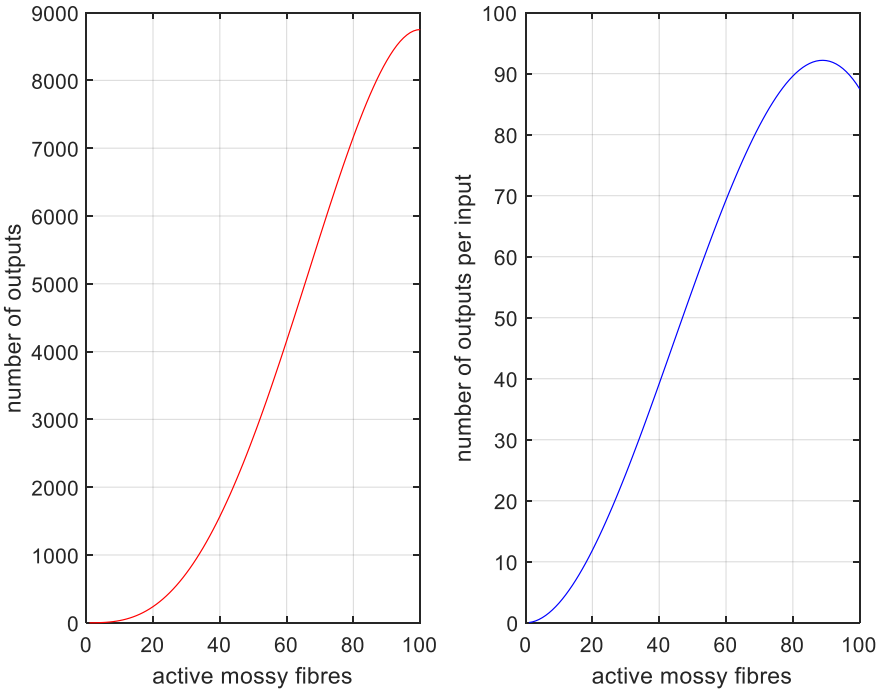
(1)

to a near approximation. (It is not exact because contact by a first active mossy fibre very slightly alters the probability that a second active mossy fibre makes contact on the same cell, because the probability for each of the other three dendrites falls from $\frac{7x}{7y}$ to $\frac{7x-1}{7y}$. A second contact alters the probability of further contact, and so on.)

The effect of changing the number of inputs to a cluster field, or of swapping some for others, is shown in Figure 3.1. There is electrophysiological evidence, discussed later (Section 3.3.4), that a minimum of 2 – and more likely 3 – inputs are necessary to make a granule cell fire, subject to modulation by Golgi cells. The simulations graphed in Figure 3.1 use a threshold of 3.

The relationship of the number of mossy fibre inputs to granule cell outputs is highly predictable. This is the result of the anatomy of contact by mossy fibres on granule cells, the large number of granule cells, and the law of large numbers, which states that over a large number of independent trials (a few hundred is enough) with outcomes that have a fixed probability, the ratio of outcomes will converge towards the expected proportions.

FIGURE 3.1 A B



C D E

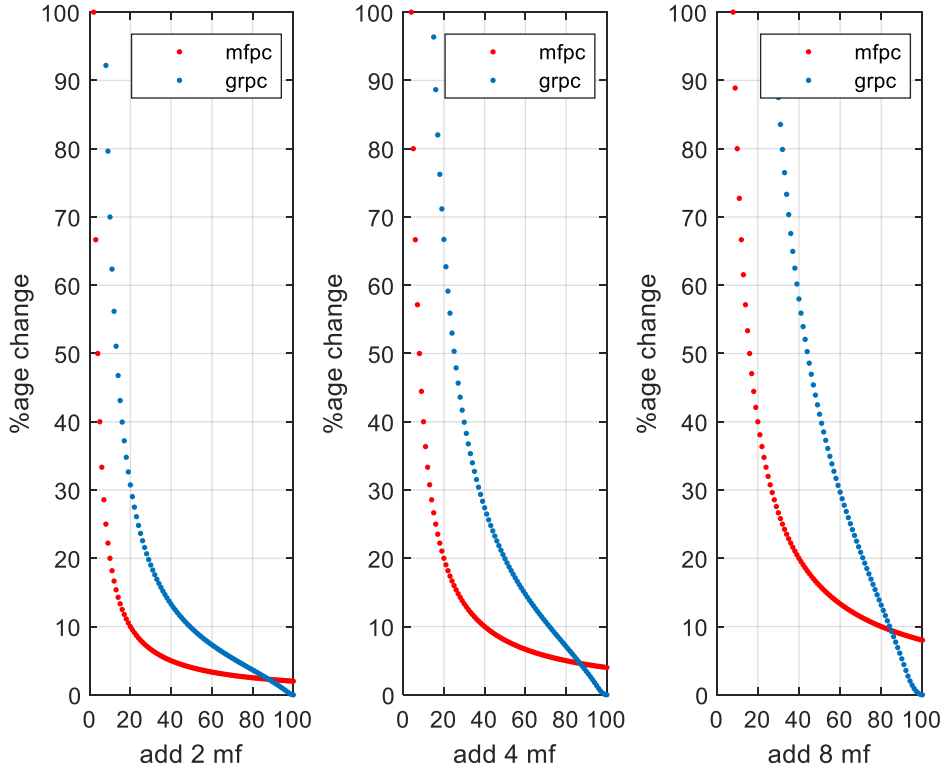


FIGURE 3.1 contd

F

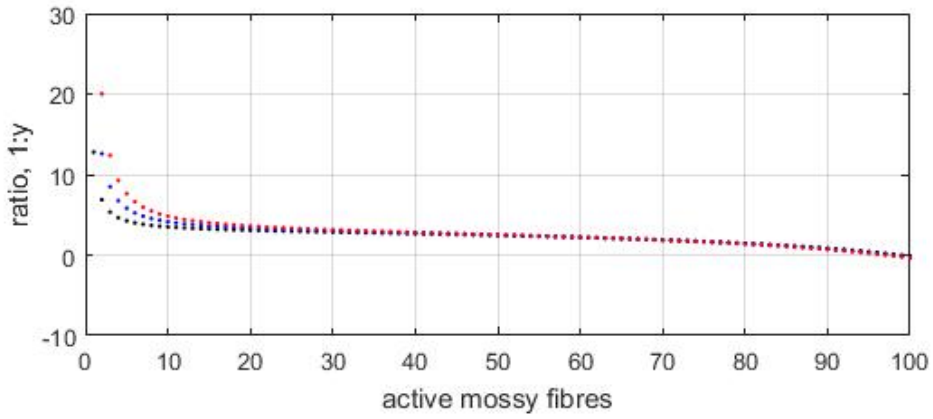


FIGURE 3.1 A: The number of active mossy fibres with input to a cluster field (x axis) plotted against the number of outputs (granule cells excited by input). To fire, a granule cell needs mossy fibre input to either 3 or all 4 dendrites. Modulation by Golgi cells is disregarded. **B:** The number of outputs per input increases as the number of inputs rises. That is, a single mossy fibre (either more or less) has a larger effect if it is one of a larger number of inputs. This relationship breaks down – the curve peaks – at the high end of the range because an extra input is more likely to be to the fourth dendrite of a granule cell which already fires. **C-E:** The percentage change in the number of active granule cells (blue data) caused by the addition of an extra 2, 4 and 8 active mossy fibres (C, D and E respectively) to a number that are already active, x. The red data show the percentage change in the number of mossy fibres. For most of the range a change to the mossy fibre number causes a larger percentage change to the granule cell number. Blue values are lower than red values at very high values of x because at those high levels, many granule cells receive 4 inputs, and so still receive enough input

to fire if one is removed. **F:** The percentage change to output caused by the addition of 2 (black), 4 (blue) and 6 (red) mossy fibres to the number on the x axis, expressed as the ratio of the %age change in the number of mossy fibres to the %age change in the number of granule cells. The ratio is between 1:2 and 1:4 across the range ~20-80 mossy fibres, that is, the percentage change to output is between 2 and 4 times greater than the percentage change to input. For most of the range x > 10 the data converge.

Because mossy fibre input is to a region with a minimum size of a terminal cluster field, which contains thousands of granule cells even at the lowest estimates, the law of large numbers always applies. As a result, the probability that a single granule cell will fire predicts the proportion of the population that do so. In this way, the circuit deploys probability without risking an uncertain outcome, notwithstanding – indeed requiring – random contact by mossy fibres on granule cells.

3.2.3 Random separation of input and output patterns: decorrelation

Decorrelation (like sparse coding) is a long-standing idea (Marr 1969, Albus 1971, Cayco-Gajic, Clopath et al. 2017). It refers to the quasi-randomisation of activity in the granular layer, so that the pattern of output is unrelated to the pattern of input except that a particular pattern of input (at a particular set of firing frequencies) uniquely and always generates the same output pattern. The result is that even

substantially-overlapping input patterns recode as randomly-separated output patterns. Separation by decorrelation follows from the assumption that contact by a mossy fibre on granule cells is at random, provided the fraction of active granule cells is small (Marr 1969, Cayco-Gajic, Clopath et al. 2017). This does not guarantee that output patterns do not overlap, but it means that the proportion of overlap between two patterns is a function of pattern density and predicted by probability, and is the same for any two patterns (assuming uniform density).

Decorrelation by only this mechanism is imperfect,[2] so that where input patterns overlap very heavily, overlap of output patterns may be by more than an amount that is only random, because generation of output patterns is not fully random. Golgi cells add a second layer of randomisation. Golgi cells are discussed in the next section.

3.3 Adding Golgi cells

3.3.1 Outline

This section models the effect of Golgi cells on granular layer recoding. The output of the model is the number of active granule cells in a mediolateral row of 20 cluster fields and therefore the number of parallel fibres that are active overhead (in a closed

[2] Because it is a biological approximation.

system). The following discussion provides the basis in evidence for the parameters used in the model, and for their relationships.

3.3.2 Background

Transmission from mossy fibres to granule cells is inhibited by Golgi cells. Golgi cell axons ramify profusely and strongly inhibit granule cells via a broad axonal arborisation (Eccles, Llinás et al. 1964, D'Angelo, Solinas et al. 2013) that covers the entire depth of the granular layer (Eccles, Ito et al. 1967). Within that field, contact is at random, so that a granule cell may be inhibited with equal probability by any of the Golgi cells which inhibit a region, and indeed each of its dendrites may be inhibited by different Golgi cells (see Appendix 5).

Golgi cell apical dendrites penetrate and traverse the molecular layer, branching modestly. Golgi cell basal dendrites are shorter and do not leave the granular layer. Both initially spread out in all directions but apical dendrites then turn towards the outer surface of the cortex following a little-deviating, broadly vertical course (Eccles, Ito et al. 1967). Golgi cells receive excitatory contact from mossy fibres and granule cells, and light contact from climbing fibres (Palay and Chan-Palay 1974, Shinoda, Sugihara et al. 2000) at the soma (Eccles, Ito et al. 1967). Mossy fibre contact on Golgi cells is onto basal dendrites within glomeruli. There is also contact on Golgi cell basal dendrites, as well as apical dendrites, by local granule cells (Eccles, Llinás et

al. 1966). A Golgi cell may receive 400 synapses from the ascending axons of granule cells and another 400 from parallel fibres arising from local granule cells (Cesana, Pietrajtis et al. 2013). There is parallel fibre contact from non-local granule cells at a further 1,200 synapses, onto apical dendrites.

The most numerous contact is therefore on apical dendrites (Palay and Chan-Palay 1974) from granule cells which originate in other circuits, although local granule cells are individually more likely to make contact. Transmission at individual parallel fibre synapses is weak (Dieudonné 1998).

In vitro, Golgi cells fire spontaneously at ~12 Hz (Kanichay and Silver 2008 p.8955) (but see Dieudonné 1998 who reports 3 Hz plus or minus 1.7, and 1.9 to 11.9 Hz in Ruigrok, Hensbroek et al. 2011). Evoked firing in brief bursts of 40-50 Hz has been reported *in vivo* (van Kan, Gibson et al. 1993).

3.3.3 A glomerular switch

Mossy fibre terminals are each ensheathed in a semi-permeable membrane, an arrangement termed a glomerulus. The membrane restricts diffusion of synaptically-released neurotransmitters that escapes into the surrounding space (spillover). The intraglomerular balance of the concentrations of GABA (from Golgi cells) and

glutamate (from mossy fibres) is adjustable and functional. GABA spillover inhibits glutamate release from mossy fibre terminals acting through presynaptic $GABA_B$ receptors (Mitchell and Silver 2000a), while glutamate spillover inhibits GABA release from Golgi cell terminals acting through presynaptic group 2 metabotropic glutamate receptors (Mitchell and Silver 2000b). Thus mossy fibre and Golgi cell terminals compete to suppress each other. An imbalance, where one gains ascendency, may create a positive feedback loop. The outcome is therefore not proportionate to the balance of input but may be a winner and loser (not confirmed but a proposal).

As a result, a glomerulus acts like a switch, so that a dendrite either counts towards the granule cell firing threshold or is prevented from counting.

However, there may not always be an outright winner, even at a dendrite where a granule cell receives strong excitatory input. An alternative to the winner-take-all proposal is that there is an independent balance of influence at each dendrite and the effect at the soma is the sum of the net effects. A granule cell still needs, to fire, a minimum number of excitatory inputs which are individually strong enough to have the combined effect of depolarising the soma to firing threshold, and inhibition still has the effect that some cells fire and some don't, depending on the rate of inhibition.

In this scenario, separately-sourced inhibition of each dendrite creates a two-tier firing threshold: First, each input must exceed an inhibitory threshold to count

towards somatic depolarisation, and second, there must be enough of them (and the integrated effect must be strong enough) to meet the somatic depolarisation threshold. The result is an input threshold that counts inputs (discussed in the next section) incorporating graded sensitivity of granule cell rates to input rates. Weakening of glomerular GABA release by stronger excitatory signals may improve fidelity of transmission (but it is unnecessary to assume all-or-nothing dendritic signalling). A numerical threshold does not imply a binary effect on granule cell rates, simply that control of whether they fire or not (as opposed to the rate they fire at) can be modelled as a switch controlled by the rate of inhibition (and this is not an unphysiological simplification).

This does not (alone) preclude a near balance of excitatory and inhibitory input to a glomerulus which generates a weak dendritic signal. Weak granule cell signals are mopped up downstream. The parallel fibre-Purkinje cell synapse acts as a high pass filter – both transmission (van Beugen, Gao et al. 2013) and synaptic learning (Bidoret, Ayon et al. 2009) are contingent on a minimum parallel fibre signal frequency, so that there is a de facto rate threshold for a functional effect of a granule cell signal.

In the model inhibition is represented as a probability that a mossy fibre signal counts towards the granule cell input threshold – the number of mossy fibre wins needed to make a granule cell fire – or else doesn't: a veto.

3.3.4 Setting the granule cell input threshold

How many mossy fibres does it take to make a granule cell fire? Seemingly, input to a single dendrite is not enough. Tonic inhibition effectively blocks activation of a granule cell by a single mossy fibre input (Chadderton, Margrie et al. 2004, Duguid, Branco et al. 2012). These studies used juvenile animals – in mature animals tonic inhibitory activity is stronger, so transmission would be more strongly inhibited (Brickley, Cull-Candy et al. 1996, Wall and Usowicz 1997). Thus, ‘a single mossy fibre will not [elicit firing]...even when activated at 1000 Hz’ (Bengtsson and Jörntell 2009 p.2393).

A threshold requiring input to all four dendrites is also improbable. Granule cells activated by cutaneous stimulation were found to receive input to 3 or 4 dendrites (Jörntell and Ekerot 2006), prompting the inference that this might be because they must receive input to all their dendrites to fire. But this would mean that only one combination of mossy fibre inputs could trigger firing. Expression of Golgi cell regulation would be very severely limited (because, again, only one permutation of inhibition – all losers – would activate the cell). Also, this would generate an output pattern too sparse to provide even the modest input to stellate cells estimated from experiments (Jörntell and Ekerot 2003) – in our simulation only around 80 granule cells are active, or 0.00024 of the population (Figure 3.2D). In contrast, a threshold of 2 (Figure 3.2B) would generate more than the reported input to stellate cells by an order of magnitude (both predicted by the distributed probability for contact by active

parallel fibres on a stellate cell – see Figure 3.4). The working assumption is therefore made here that the threshold is 3. A minimum of 3 is consistent with the view (based on Jörntell and Ekerot 2006) that granule cell 'activation depends on the synchronous activation of 3 or 4 mossy fibres' (Bengtsson and Jörntell 2009 p.2393). Three is also the minimum threshold estimated to be necessary for sparsening (Billings, Piasini et al. 2014, and used in Cayco-Gajic, Clopath et al. 2017), contributing to pattern separation.

The threshold is fixed by adjustable Golgi cell rates, that is, always fixed regardless of mossy fibre rates. This is a departure from the idea that Golgi cell regulation adjusts the number of mossy fibre inputs needed to make a granule cell fire (in order to keep parallel fibre activity low) (Marr 1969, Albus 1971).

3.3.5 A probability loop

The model does not receive firing rates as input because it uses the probability of an inhibitory veto to determine the outcome of the glomerular competition, and because the output of the competition is a cell count and not firing rates. (Exclusion of input rates from the model is discussed in section 3.4.6.) The numbers are large enough to express probability because the minimum scale of input is a cluster field. Incorporation of probability into the operation of the system is not an expedient to simplify modelling but the proposed mechanism.

Parallel fibre input to Golgi cells drives a probability loop. The number of parallel fibre inputs to a Golgi cell determines its probable firing rate (although it is not the only determinant: see next section). Dense parallel fibre activity causes a higher probability of more input to Golgi cells and vice versa. A higher Golgi cell firing rate increases the probability that its inhibition drives a glomerular veto (and a lower rate reduces the probability). Thus the density of parallel fibre activity controls the probability that a mossy fibre signal counts towards the input threshold of a granule cell that receives it.

Golgi cell firing is irregular (Ruigrok, Hensbroek et al. 2011) within adjustable rates so that on a short time scale spiking is not precisely coordinated even between cells firing at the same rate. The number of active parallel fibres that make contact on a Golgi cell is in a range predicted by a probability distribution, so that even Golgi cells that sample the same parallel fibre activity may receive contact from a different number of active cells. So individual Golgi cell spike timing and rates are not exactly coordinated even in a (defined) beam, where they all sample the same activity. But the expected number of their excitatory inputs is a function of the density of parallel fibre activity, so that the probability of an inhibitory veto can be derived from (and is controlled by) density of parallel fibre activity.

The effect of density on the probability (p) of a Golgi cell veto is treated as a sigmoidal relationship. The probability flatlines at low and high values of input density

– below a minimum number of granule cells there are not enough for a veto ($p \rightarrow 0$), while at high densities increasing the number has no further effect ($p \rightarrow 1$). The slope represents the part of the range where the probability of a veto varies with the density of parallel fibre activity. As the probability of a veto in turn regulates density, it creates a regulatory feedback loop, acting through mutually regulating probabilities.

The number of active parallel fibres in a beam is calculated using a randomly generated number of mossy fibre inputs to each of a row of mediolaterally-aligned cluster fields. The calculation uses the convention that, to fire, a granule cell must receive mossy fibre input to a minimum number of dendrites which do not also receive a Golgi cell veto. Excitatory input to dendrites which receive a veto has no effect. For example, there is no graded contribution of mossy fibre input in those cases towards depolarisation of a granule cell.

The probability that a dendrite counts towards the total is thus the product of the probabilities that it receives no veto (1 – *P(v)*) and also receives excitatory mossy fibre input (*e*). The probability that the cell fires is therefore given by

$$\sum_{k=m}^{n} \frac{n!}{k!\,(n-k)!} * \Big(\big(1 - P(v)\big) * e\Big)^{k} * \big[1 - \big((1 - P(v)) * e\big)\big]^{n-k}$$

(2)

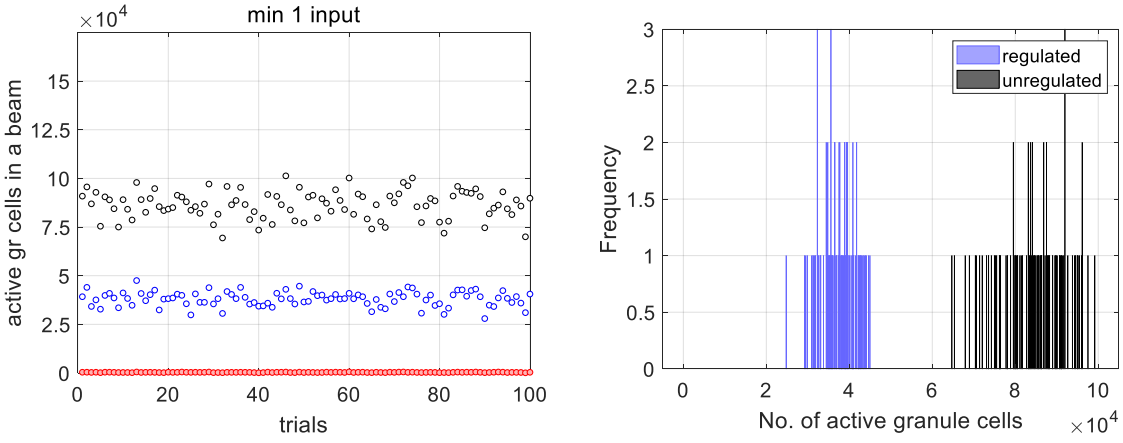
to a near approximation (see equation 1), where *n* is the number of dendrites, and *m* is the minimum number which must receive the correct combination of input to drive firing. So that the expected (denoted by *E*) number of granule cells in a cluster field (say f_1), out of a total *t*, that fire is

$$E(f1) = t\left[\sum_{k=m}^{n} \frac{n!}{k!\,(n-k)!} * \left((1 - P(v)) * \frac{x}{y}\right)^{k} * \left[1 - \left((1 - P(v)) * \frac{x}{y}\right)\right]^{n-k}\right] \tag{3}$$

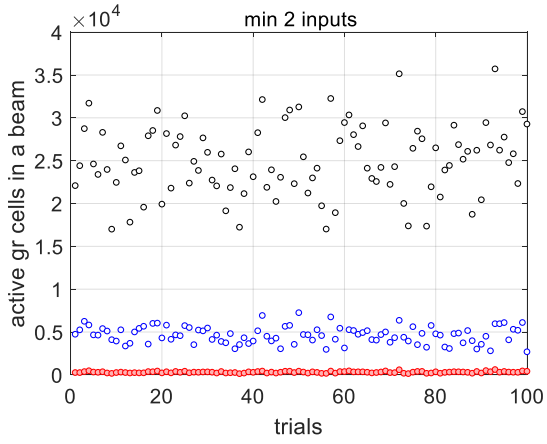
where *y* is a physiological constant (100) and *x* is in a range 3-30 generated randomly for each field in each trial. *t* (also a constant) is 8,750 and it is assumed all granule cells have *n* = 4 dendrites.

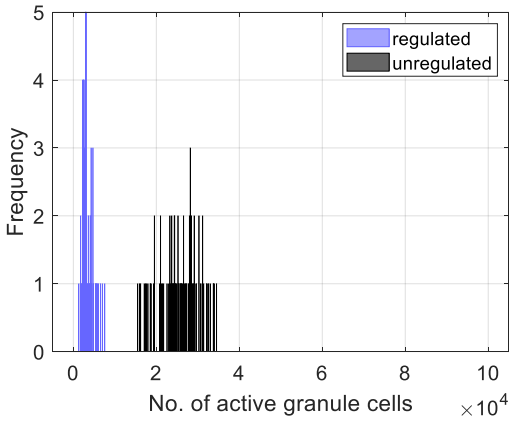
FIGURE 3.2

A

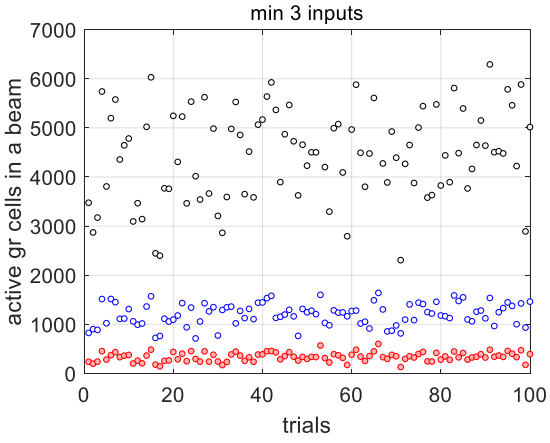


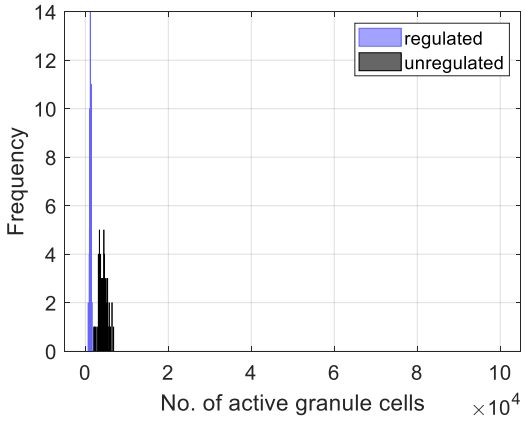
B


min 2 inputs
×10^4
active gr cells in a beam
trials

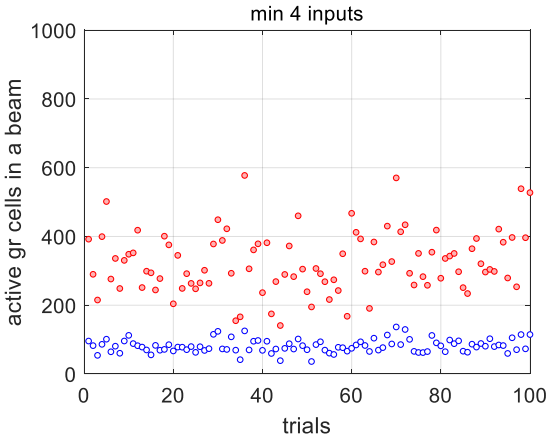

regulated
unregulated
Frequency
No. of active granule cells
×10^4

C


min 3 inputs
active gr cells in a beam
trials


regulated
unregulated
Frequency
No. of active granule cells
×10^4

D


min 4 inputs
active gr cells in a beam
trials

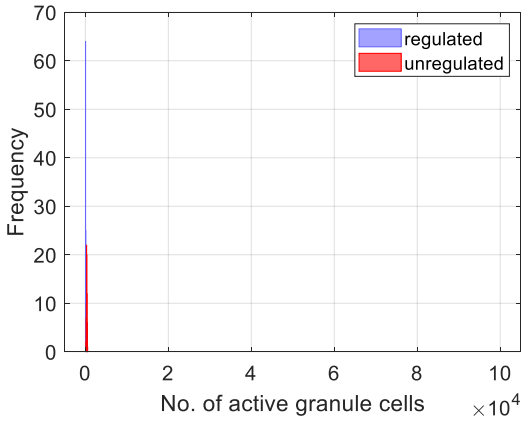

regulated
unregulated
Frequency
No. of active granule cells
×10^4

FIGURE 3.2 The total number of excited granule cells in a beam of 20 cluster fields, trial by trial, in each of 100 trials, generated by a computer model of recoding in the granular layer. The minimum number of mossy fibres needed to make a granule cell fire is varied from 1 (panel A) to 4 (panel D). The strongest (but not conclusive) experimental evidence is for a threshold of 3. For each trial, estimates are derived of the number of granule cells without Golgi cell regulation (black dots), the number with Golgi cell regulation (blue), and the number which receive 4 mossy fibre inputs (red, identical in all graphs, for comparison). In panel D there are no black data because red is the unregulated number. Note: *y* axes are scaled to the data. The number of active mossy fibres with input to each field on each trial is random, in the range 1-30. The blue data in panel C are the most probable physiological range. Blue activity is sparsened by regulation by Golgi cells and also maintained in a narrow and stable range. Panels on the right show the equivalent to the black and blue data in A-C, and blue and red in D, but as histograms. In this format the narrowness of the range of the blue data is more evident. Even so, the range of the x-axis (100,000) is considerably lower than the parallel fibre population (300,000-400,000) of a beam.

This is used to calculate the number of granule cells that fire in each field. Excitatory and inhibitory input to a glomerulus vary independently. This is important physiologically because the outcome at each glomerulus must be independent of the outcome at other glomeruli in order for granule cell population activity to express the probability of an inhibitory veto (much as, for a predictable outcome of rolling a dice

many times, the outcome of each roll must be independent of the others). The evidence that they are independent is discussed in Appendix 5.

The sum of active granule cells in all fields is the number of active parallel fibres overhead. These are randomly and therefore evenly dispersed in the sagittal plane, so that the sum has a linear relationship with the uniform density sampled by Golgi cells. In order to take account of the mutual influence of fields on each other, the total is used to recalculate the output of each field, giving a new total. The calculation is looped using the new total each time, so that it homes in on a value for each field where the number of excited granule cells and the effect of inhibition by Golgi cells are in equilibrium (after a few iterations).

To reflect encroachment of Golgi cell axons from neighbouring beams the same calculations are run for five beams, two on either side of the test beam, and in each iteration, for each field, using an average of the probability of an inhibitory veto in that field and the two on either side, in the two neighbouring beams. Otherwise the system is closed. The choice of the dimensions of a beam is discussed next.

3.3.6 Dimensions of a beam

The model calculates the number of active granule cells in a ‘beam’. The choice of the dimensions of a beam inevitably means defining a region which is in reality neither anatomically nor functionally discrete.

A beam represents a 3,000 x 200 µm folia-aligned slice of the cerebellar cortex, the horizontal dimensions of a row of 20 mossy fibre terminal cluster fields lined up in the mediolateral direction. The reason for basing it on a row of cluster fields is that estimates are available for the number of mossy fibres with input to a field and terminals per cluster, and therefore terminals per field. The length is half the span of parallel fibres, assuming a 6 mm parallel fibre span, so that all granule cells in the underlying granular layer give rise to parallel fibres that traverse the whole distance. A parallel fibre span of 6 mm is reported for cats, chickens and monkeys (Brand, Dahl et al. 1976, Mugnaini 1983); 5 mm is reported in rats (Harvey and Napper 1988, Harvey and Napper 1991).

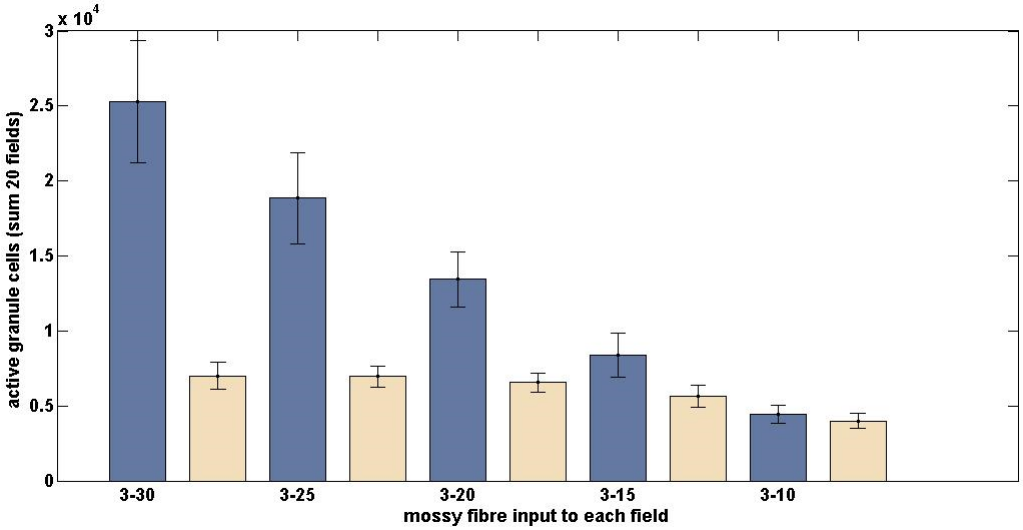
The granule cells in a 3 mm beam in reality provide only half the parallel fibres that pass through it. Yet a longer beam would still need boundaries – no matter how long you make it, there is always encroachment from granule cells which lie outside. Also, a larger population of parallel fibres in a beam would not mean that proportionately more are active, because the number is independent of the size of the population, as

regulation is modulated by the number of parallel fibres that are active, not the proportion that are active.

Granule cells have a distance-dependent strength of effect. The frequency of synapses declines along a parallel fibre with increasing distance from its bifurcation, such that there are half as many at its distal ends than at the centre (and synaptic densities are around half the size) (Pichitpornchai, Rawson et al. 1994). While an effect of distance is not absent with a shorter beam, it is smaller. All granule cells are treated as having the same strength of effect. By the same token the excluded effect of more distant granule cells is relatively weak.

FIGURE 3.3

A



B

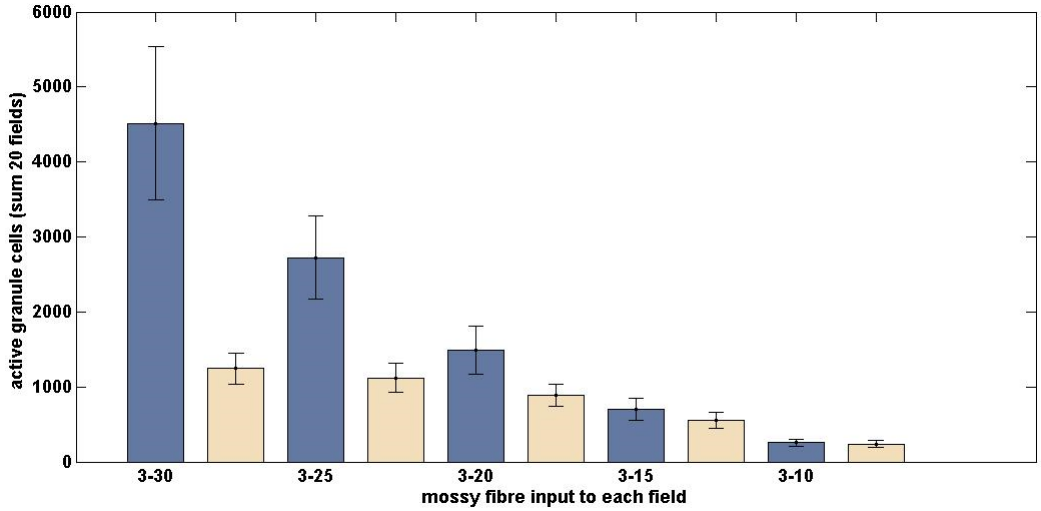


FIGURE 3.3 All bars show the total number of active granule cells in a row of 20 mediolaterally-aligned cluster fields averaged over 100 trials, to show the effect of low mossy fibre input to a beam. Input to each field is a random number in the range shown under the blue bars, also the input for the white bar to the right. Blue bars show the level of granule cell activity without Golgi cell regulation. White bars show the level of granule cell activity with Golgi cell regulation in what are otherwise the same conditions. The difference between A and B is the input threshold of granule cells. In A there must be input from at least 2 mossy fibres, in B input must be from at least 3. On-beam Golgi cells act together to effectively cap the volume of parallel fibre traffic (best seen in A). However, they have no effect in the opposite direction at weak levels of input – they cannot compensate for low mossy fibre input by making up the difference (best seen in B). Maintenance of a stable level of parallel fibre traffic therefore depends on a minimum total of mossy fibre input to the beam (while the effect

of more is capped). Provided that threshold is met, however, the density of parallel fibre activity overhead is insensitive to the number and distribution of inputs.

3.3.7 Functional anatomy of Golgi cells

Golgi cell apical dendrites fully traverse the molecular layer vertically, branching little, so that they cover the same distance at all depths, and the Golgi cell axonal plexus fills the depth of the granular layer (Eccles, Ito et al. 1967), so that the probability and strength of the influence that a Golgi cell receives from parallel fibres, and the probability and strength of the influence it has on granule cells, receive no effect from depth-dependent variation of Golgi cell morphology (which therefore the model does not need to reflect).[3]

3.3.8 What the simulation does and does not do

The program simulates self-regulation of granule cell activity in a 200 x 3,000 micron corridor – a row of 20 cluster fields – using a probability loop to calculate the expected number of excited granule cells in each of 100 trials. The purpose of the simulation is to test (by quantifying) the proposal that a physiological loop of

[3] The majority of Golgi cells lie close beneath the Purkinje cell layer. A minority are smaller and deeper and may in some cases lack apical dendrites. Deeper Golgi cells are not represented separately in the model.

probabilities returns a functionally fixed level of parallel fibre activity regardless of the number and distribution of mossy fibre inputs to the system.

To calculate, for each field, *P(v)* in the next and later iterations, the active fraction of parallel fibre contact on Golgi cells, calculated separately for each field, is received by a function which uses an iterative formula to generate a series of data pairs that describe a sigmoidal curve (see Figure 3.4B). As a Golgi cell receives contact from around 1 in 30 local granule cells and 1 in around 300 granule cells in other fields, and from around 2,000 in total (Appendix 6), the proportion that are active is

$$\left[\left(E_{(f1)}/30 + \left(\left(\sum E_{(other\ fields)}\ /300\right) - E_{(f1)}\right)\right)\right]/2000 \quad (4)$$

This is the *x* coordinate of a data pair, returning a value *j* that it is paired with, which is used to adjust *P(v)*. The adjustment is by half of the difference (with *P(v)* in the previous iteration) to represent dampening of oscillations. Dampening is to reflect, within the constraints of the model, that oscillations are not discrete – that is, not isolated moments in time but graduated changes.

The curve is a means to an end. The shape is chosen so that very low levels of parallel fibre input to a Golgi cell have little effect and high levels have little further effect, with a functional range in between where an effect is proportional to input. The

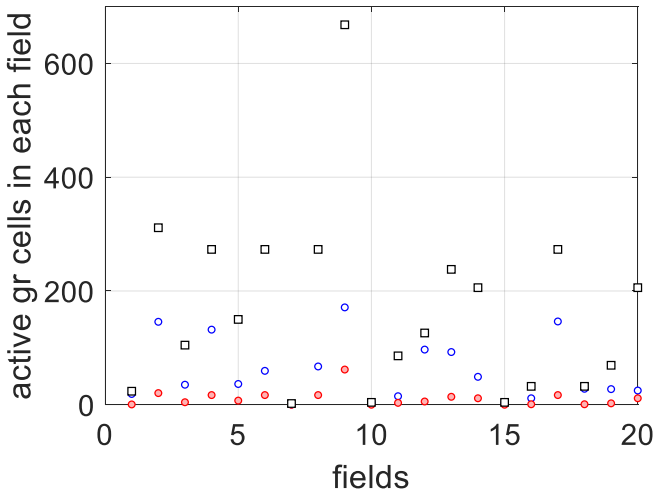
limits and size of the functional range are unreported. In the model the active percentage of parallel fibres that contact a Golgi cell is around 0.0035%, so 7 if Golgi cells receive contact from 2,000 in total (Cesana, Pietrajtis et al. 2013), in the approximate range 3-11. (This is at the low end of x values under the curve because there is also a contribution to *P(v)* from direct mossy fibre contact, discussed below.)

The curve is intuitive but a surrogate, necessary to stand in the place of missing data. If the gradient was in a different range – further to the right, say – predicted parallel fibre activity would be in a higher range. (This would also be the effect if the mossy fibre contribution to *P(v)* was smaller.) Even knowing the proportion of parallel fibres that are active it is unknown in any particular case how many make contact because the exact number is chance, with a distributed probability, (nor would we know the spatial distribution of contact, or relative timing of inputs, or how that might vary the postsynaptic effect).

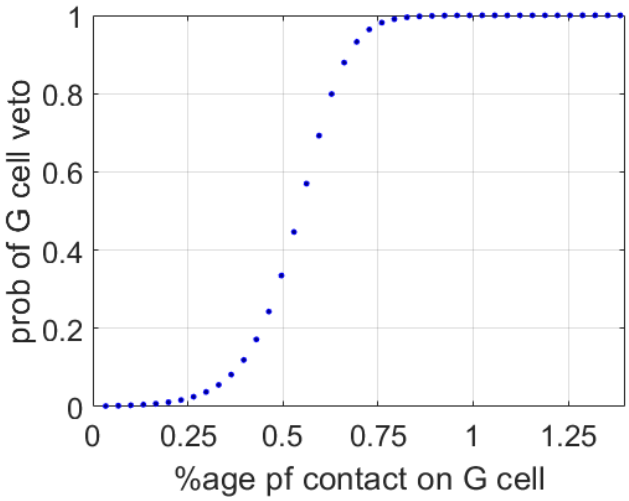
That is not a problem for the cerebellum because (it is proposed) the functional effect does not depend on individual signals or the particular pattern, it exploits probability. Nor does it matter that two Golgi cells that sample exactly the same parallel fibre activity are likely to receive contact from a different number of active cells, and very likely to receive contact from a different subset. In fact (as noted) this contributes to the necessary independence of the outcome of the glomerular competition.

FIGURE 3.4

A



B



C

Active:	0	1	2	3	4	5	6	7	8	9	10
P:	.006	.033	.083	.140	.176	.177	.147	.105	.065	.036	.018

FIGURE 3.4 A: The same simulation as Figure 3.2 but showing a single trial. A column of red, black and blue data are, respectively, the number of granule cells that receive contact from 4 active mossy fibres, the number that receive contact from at least 3, and the number that receive at least 3 to dendrites that also receive no inhibitory veto. The (randomly generated) number of mossy fibre inputs to each field, from left to right, is 9,22,15,21,17,21,4,21,29,5,14,16,20,19,5,10,21,10,13,19. Each column represents granule cell activity in each of a row of 20 mossy fibre cluster fields making up a 3,000 x 200 μm region of the granular layer whose long axis lies mediolaterally. The blue data represent physiological activity. The purpose is to show that the tightly regulated density of parallel fibre activity (in Figure 3.2) belies an uneven distribution of granule cell activity in the granular layer beneath. **B:** The curve used to make adjustments to *P(v)*. The curve was moved to the right for Figure 3.2A – otherwise, the number of granule cells meeting the input threshold was too high to fall under the functional part of the curve. **C:** The binomial distribution for active contact by parallel fibres on a superficial stellate cell derived from the density of parallel fibre traffic predicted by the simulation, the size of the stellate cell dendritic field in the sagittal plane (Palay and Chan-Palay 1974), and an estimate of the number of parallel fibres that make contact (1,000). There is a probability p = 0.9074 that the number lies between 2 and 8, in good agreement with the finding that input to an outer level stellate cell is made up of 'two to eight substantial EPSPs [excitatory postsynaptic potentials]' (Jörntell and Ekerot 2003 p.9628).

To be clear, the simulation does not predict the exact range of the density of parallel fibre activity but that the range is narrow and stable (and low), and standard, and that this is feasibly the result of recoding with the proposed mode of operation, and that this can account for the evidence. That said, the range of the blue data in Figure 3.2C, obtained with what is thought to be the physiological input threshold for granule cells, generates a level of activity which is a very good match for the reported number of parallel fibre inputs to stellate cells (Jörntell and Ekerot 2003) (Figure 3.4C).

3.3.9 Direct mossy fibre contact on Golgi cells

Direct mossy fibre contact on Golgi cell basal dendrites is reflected in a field-by-field adjustment to the probability of an inhibitory veto derived by calculating the probability in each trial that a Golgi cell receives contact from at least 4 active mossy fibres, reported to be the minimum necessary for an effect (Kanichay and Silver 2008). Around 1 in 10 glomeruli receive a Golgi cell dendrite (Hámori 1992), and maybe fewer. Given a known number of active mossy fibres that innervate a field, the probability of contact from a minimum of 4 is given by 1 minus the sum of the probabilities of contact from 0, 1, 2 and 3, derived by a binomial. This is necessarily a coarse measure (again it is a stand in for missing data) but proportionate to the number of mossy fibre inputs to a field, and in an appropriate range, so that it provides a commensurate adjustment to $P(v)$.

This fits functionally with the idea that the local probability of a veto is weighted to reflect the amount of local mossy fibre input. Possibly, a function of weighting is to mitigate clustering of granule cell activity. It would be suited to this function because the downward pressure on granule cell activity would be proportionate to the need for it. Severe clustering – or the reverse: severely depleted local activity – would impair the uniform density of parallel fibre activity overhead. It is unnecessary for underlying granule cell activity to be equal across all fields – and it isn't (Figure 3.4A).

There is also an adjustment for the higher probability of contact on Golgi cells by local than distant granule cells (see Appendix 6), re-calculated for each field in each iteration of the program.

3.4 Discussion

3.4.1 Simulation

Figure 3.2 shows the sum of granule cell activity in a beam in each of 100 trials. This can be thought of as either representing the same beam on 100 occasions, or 100 beams on the same occasion. Golgi-cell-mediated regulation has the result that parallel fibre activity is confined to a narrow and stable range. Regulation is robust across all conditions. The blue data in Figure 3.2C are a best estimate of the

physiological range, because they are derived using the most probable granule cell input threshold.

Regulation is homeostatic notwithstanding a fixed granule cell input threshold. Any number and pattern of mossy fibre inputs to a beam recodes as a functionally invariant number and pattern of parallel fibres. This relationship is subject to there being a minimum total number of mossy fibre inputs to the system (Figure 3.3). The large number of granules cells turns random contact of mossy fibres on granule cells, and inhibition by Golgi cells of granule cells, into a functional and numerically-predictable outcome.

3.4.2 Circuit-level memory

Decorrelation by layers of randomisation (provided by random contact of mossy fibres on granule cells and independently-controlled inhibition of transmission by Golgi cells) means that the pattern of parallel fibre activity is randomly and therefore uniformly distributed in the sagittal plane (at the scale of activity sampled by a Purkinje cell). Assuming the effect of regulation by Golgi cells is ubiquitous, input to a microzone is evenly distributed and received all along its full length. A microzone is defined by its climbing fibre input. It follows that plastic modulation of parallel fibre synaptic weights trained with a conditioning protocol is stored across a whole

microzone – memory is at circuit level rather than at the level of input to each Purkinje cell.

3.4.3 Functional anatomy

The proposed recoding mechanism does not use Golgi cells to adjust the number of mossy fibres needed to make a granule cell fire in order to keep parallel fibre activity low. If it did, granule cells would receive an effect on their firing rate from the number of mossy fibre inputs they receive – that is, if the threshold was set at one, say, the rate would depend in any particular case on whether firing was triggered by one mossy fibre, two, three or four: a random variable.

It is proposed instead that Golgi cells control the local probability that granule cells fire. This permits adjustments of the number of granule cells that fire in a continuous range of variation (because the number that fire reliably reflects graded changes to probability). It also permits a gradient between regions with different probabilities, rather than a step change. A stochastic chain of effects explains the morphology of the links and the anatomical rules that govern contact between them, including convergence and divergence ratios.

The mossy fibre-granule cell relay is adapted to transmit short duration mossy fibre action potentials received at high frequency (Ritzau-Jost, Delvendahl et al. 2014 p.152). The granule cell firing rate is reported to vary faithfully with input signal frequency (Rancz, Ishikawa et al. 2007). Granule cells 'have a relatively linear and uncomplicated conversion of depolarisation level to spike rate' (Bengtsson and Jörntell 2009 p.2393, citing Jorntell and Ekerot 2006 and D'Angelo et al 1998). With the winner-take-all glomerular model, the effect of excitatory input rates on output rates is also unaffected by rates of convergent inhibition from Golgi cells, because expression of inhibitory rates is confined to glomeruli which do not contribute to somatic depolarisation. So, a fixed input threshold and the glomerular competition clears the way for linear transmission.

3.4.4 Redundant variables

Parallel processing – working with multiple inputs – means that input to the cerebellum contains multiple variables – the number and pattern of active cells, for example and the permutation of firing rates (which cells fire at what rates), as well as other variables contained in a spread of rates, such as the frequency distribution and standard deviation. It is equally important to block an effect of redundant variables as it is to receive an effect of 'good' ones, because an effect of redundant variables would be arbitrary. The cerebellum is therefore designed to do both. This is an important feature.

3.4.5 Conversion of input variables into two operational internal variables

The area of the cerebellar cortex that provides parallel fibre input to a microzone is substantial: around 6 mm (the range of parallel fibres) by 15 mm (the length of a microzone), so 90 mm^2. Parallel fibre signals driven by mossy fibre input to this area are condensed into an area in the orthogonal plane that is smaller by an order of magnitude (~4.5 mm^2,the area in the sagittal plane of the molecular layer of a microzone). Scaled up, the region supplying parallel fibre input to a microzone would be around the size of four-and-a-half tennis courts side by side supplying input to a net (the microzone) 50 m long and 0.5 m thick. Granule cell bodies would be ¼ the diameter of a tennis ball and 12-14 million of them would be packed under the court to an average (but variable) depth of 70 cm. At any moment mossy fibre input to that region can be made up of any number of signals and in any configuration of active cells, firing in any permutation of frequencies.

External variables are converted to a reduced number of operational internal variables in order that redundant external variables are excluded from an effect on learning and on control of output rates. This proposal began with the idea that pattern detection and output rate coding are separable functions, so that there is no need for output rates to be derived by a complex function from learned patterns. An effect of the number of mossy fibre signals is excluded (by recoding) by converting any number to an invariant density of parallel fibre activity. An effect of the binary distribution of mossy fibre inputs is excluded by decorrelation, that is, by randomising

the recoded pattern of active parallel fibres, so that the pattern of internal signals is functionally invariant, except that it is highly specific to the mossy fibre input pattern (of active cells) and the permutation of rates they fire at. Indeed in theory mossy fibres may *all* be active and still give rise to the regulated density of active parallel fibres.

The second internal variable, really a class, is firing rates. It is part of the function of circuit design to isolate rates from other variables, by converting external variables into internal variables that are (functionally) fixed by the homeostatic regulation of parallel fibre density, so the number and binary distribution of inputs to parallel fibre targets are functionally constants. The independence of variation of patterns and rates is considered next.

3.4.6 Independence of internal variables

Input rates are not included in the simulation because the available data fall well short of the minimum necessary to simulate the effect of each stage on the next without a raft of unphysiological assumptions. Fortunately input rates are unnecessary to model regulation because it forms part of our proposal that the number of active parallel fibres and the rates they fire at are independent variables.

This independence of patterns and rates[4] can feel counterintuitive. On the face of it, generally stronger input to granule cells should mean that more fire, because more meet firing threshold. But higher parallel fibre activity predicts a higher distributed probable number of inputs to Golgi cells, and therefore stronger inhibition of granule cells. Accordingly, the inhibitory threshold is proportionately adjusted, with a self-rectifying effect on the density of active parallel fibres mediated by the probability loop. As a result, the proportion of active parallel fibres is self-regulating regardless of input rates.

Conversely, the model postulates that the glomerular competition prevents an effect of regulation on translation of mossy fibre to granule cell rates. For example, interference would result if input at variable (say stronger) rates drove a higher density of parallel fibre activity, causing a rise in the Golgi cell rate that weakened granule cell rates. Instead (in the present proposal) stronger input signals limit winners of the glomerular competition to a stable number of stronger signals. Thus there is a mutual disconnect between an effect on regulation of rates and an effect on rates of regulation.

The isolation of rates – by removing an effect of other variables on transmission – is consistent with the reported linear rate coding in the cerebellum (for example Rancz, Ishikawa et al. 2007, Jelitai, Puggioni et al. 2016), and the consistently-reported

[4] More fully, the independence of the regulated number of active parallel fibres and rates they fire at.

linear relationship of firing rates and task-related parameters. ‘The firing rate of many cerebellar neurons is a linear function of task related parameters…[and this] has been found at all levels of the cerebellar circuit’ (Raymond and Medina 2018 p.239, who provide references). Of interest in this connection is that temporal coding is reported to be absent at the parallel fibre-Purkinje cell synapse (van Beugen, Gao et al. 2013), seemingly leaving the field clear for rates to control the simple spike rate, and through the simple spike rate the firing rate of nuclear projection neurons that carry the output of the circuit.

If it is correct that internal patterns and rates contain separable, independently variable data, it allows them to be used in separate functions, because neither variable would interfere with the execution of the other function. This suggests a challenge to the idea that synaptic weights are a filter that controls output rates, because it obviates the assumption that rates are controlled by learned patterns.

3.4.7 Post script: meaning of stable density

‘Density’ of active parallel fibres is shorthand for the proportion that are active, rather than number per unit area. This distinction is important because a stable number per unit area would mean the variable size of the Purkinje cell arbour (which varies substantially between peaks and furrows of the folded cerebellar cortex) (Eccles, Ito et al. 1967) would affect the number of active cells that make contact – so that the

number would a variable. A stable proportion does not have this effect because the size and shape of the Purkinje cell arbour and the thickness of the underlying granular layer covary (Eccles, Ito et al. 1967) – the granular layer is thinner in furrows and thicker in peaks, commensurate with the larger Purkinje cell arbour size in peaks.

The idea of a stable density of parallel fibre activity needs clarifying in another way also. The small bore and slow transmission times of parallel fibres makes them 'probably the slowest…in the whole brain' (Sultan 2000 p.41). Transmission time along a 2-3 mm parallel fibre branch is ~10 ms or less (Sultan 2000). Granule cells typically fire in bursts, although firing is variable between individual cells (some fire continuously, for example, in self-paced locomotion in mice: Jelitai, Puggioni et al. 2016). The length of bursts varies: 10-20 ms has been reported in adult cats (Jörntell and Ekerot 2006) and 8-40 ms in rabbits (van Beugen, Gao et al. 2013). Bursts, accordingly, are usually longer than transmission time. It follows that for at least part of the duration of a burst the whole cell is active. For example in a 25 ms burst, in the first 10 ms the active portion of the axon spreads outwards from the soma until the whole axon is active. That lasts for 5 ms. Then in the last 10 ms the cell returns to its resting state, which also spreads out from the soma, when the cell stops generating more spikes.

In practice, a stable density of parallel fibre activity means that a microzone receives a steady turnover of parallel fibre signals, and that the rate of turnover is the same all

along a microzone (and that active cells are distributed at regulated density). A ‘volley’ of input to a microzone, accordingly, does not refer to an exactly synchronised fusillade – firing in a scintilla of time – but to turnover in a functional window, such as a synaptic learning window.

CHAPTER 4

BINARY PATTERN CLASSIFICATION BY THE CEREBELLUM

4.1 Introduction

4.1.1 Background: reported

It was proposed almost 50 years ago that output of the cerebellum is an acquired response mediated by learning under instruction of climbing fibres (Marr 1969). Purkinje cells, which exclusively carry the output of the cerebellar cortex, are organised functionally into long thin groups, or microzones (Oscarsson 1979). Purkinje cells are inhibitory and fire intrinsically (Cerminara and Rawson 2004) at robust rates (Zheng and Raman 2010). The rate correlates to the presence of the marker protein zebrin: it is approximately 60 Hz in zebrin +ve and 90 Hz in zebrin –ve Purkinje cells (in awake mice: Zhou, Lin et al. 2014). The output of a microzone is channelled down onto a smaller, discrete group of nuclear cells containing projection neurons which carry the output of the circuit. In vitro with synaptic inputs removed or blocked, nuclear neurons fire spontaneously at an average of around 90 Hz (interpositus in mice: 70 Hz in males, 110 Hz in females, hence the average: Person and Raman 2012b, Mercer, Palarz et al. 2016). In vivo, under inhibition, they continue to fire but at substantially lower rates: 10-20 Hz in resting animals (Telgkamp and Raman 2002).

A coordinated fall in the Purkinje cell firing rate causes nuclear cells[5] to fire at a transiently elevated frequency (Telgkamp and Raman 2002, Pedroarena and Schwarz 2003) and may in some cases provide the main form of control of the nuclear firing rate (Ishikawa, Tomatsu et al. 2014, Jirenhed and Hesslow 2016). In blink reflex conditioning studies, following training by repeated pairing of convergent parallel fibre and climbing fibre input, parallel fibre input alone evokes a temporary drop in the Purkinje cell firing rate and in some cases a full pause (Jirenhed, Bengtsson et al. 2007, Rasmussen, Jirenhed et al. 2008). The pause is thought to be the result of learning at parallel fibre synapses on Purkinje cells and on inhibitory interneurons, which in turn make contact on Purkinje cells. It may also involve an unknown internal mechanism that causes Purkinje cells to suppress their own firing (Johansson, Jirenhed et al. 2014, Johansson, Carlsson et al. 2015).

4.1.2 Background: theory

Established theories of the cerebellum posit that graded output is derived from input by passing it through a learned pattern of synaptic weights (Marr 1969, Albus 1971, Fujita 1982, Ito 1989), trained under climbing fibre tuition. Weighting is at the parallel fibre-Purkinje cell synapse. Graded plastic changes cause (in this model) Purkinje cells to modify their response, contributing to a learned effect on motor output. Later

[5] References to 'nuclear cells' in this paper are to excitatory projection neurons which carry the main output of the circuit unless something different is clear from the context.

models have extended earlier work in an attempt to include a role for other sites of plasticity. But the premise remains that learned synaptic weights determine the response to a particular set of inputs, which in that sense the system remembers. In this arrangement, pattern matching and control of Purkinje cell firing, and therefore nuclear rates, are performed by the same mechanism, that is, a remembered pattern[6] provides the drive that modulates the Purkinje cell rate.

4.1.3 Proposal

It will be argued instead that pattern matching is at circuit level, rather than the level of individual Purkinje cells, and there is no memory of individual patterns. Discrimination is instead at the level of, and between, the whole class of known patterns, and the residual class of all other patterns. The memory linking parallel fibre input and Purkinje cell output is not stored as a pattern of graded synaptic weights. The response does not discriminate between patterns within a class. Pattern matching is, rather, binary in the sense both that classification of input has only two classes, and that the response is one of two things: to allow or (the default state) veto output of the circuit. There is no graded or intermediate response to a partial match. Just a fraction of the Purkinje cells which innervate a nuclear group is sufficient to veto output of the circuit, so that if any part of a known pattern (received all along a microzone) is a mismatch, it blocks the response of the whole circuit. This is not to

[6] 'Pattern' is used here and throughout to mean the configuration of co-active cells, and not (for example) the permutation of firing rates or fluctuation of firing rates over time. 'Input' is used to mean synaptic contact by an active cell, rather than only innervation.

suggest that the output of the circuit is binary, but that the mechanisms of pattern detection and control of nuclear rates are separate, and the output of the mechanism of pattern detection is binary. The function of a determination – match or not – is to select which circuits have output and when, but does not control the nuclear rate. The response to a match is permissive – permitting but not coding output. Contrary to the separation of stored patterns predicted by traditional theory, they overlap very substantially, and that overlap is well tolerated and useful.

4.2 Functional anatomy

This section argues that it takes only a handful of intrinsically active Purkinje cells – a small fraction of the population of a microcomplex (Garwicz and Ekerot 1994, Garwicz, Apps et al. 1996, Apps and Garwicz 2005), for which ‘circuit’ is used as shorthand – for the majority of a nuclear group to receive inhibitory input at strong rates at hundreds of synapses.

4.2.1 The Purkinje cell-nuclear projection neuron contact rules

Substantially all of the output of the cerebellar cortex which converges on a functional group of nuclear cells is from the same microzone or the same functional but dispersed group of microzones which form part of a multizonal circuit (Pantò, Zappalà

et al. 2001, Apps and Garwicz 2005). In part of the C1 zone, for example, whose output is to the anterior interpositus nucleus and whose circuits control hind limb movements in rats, ‘a fine grain topography exists’ (Cerminara, Aoki et al. 2013 p.16440). This may not hold for all circuits. Purkinje cell termination patterns are less focussed in the fastigial nucleus, for example (Sugihara, Fujita et al. 2009). However, it is assumed in this paper that circuits are segregated, so that the output of a microzone is exclusively to the nuclear cells that carry the output of the same circuit. A nuclear group may receive the output of more than one microzone if they are part of the same circuit – the same group of olivary cells may project to two or more microzones whose output in turn converges on the same nuclear group (Apps and Garwicz 2005) – but the ‘split loops’ still form a closed circuit.

The ratio of Purkinje cells to deep nuclear neurons has been estimated (for mice) at approximately 11:1, with convergence of approximately 30-50:1, and divergence (in rats) of approximately 4-5, with each Purkinje cell contributing on average around 30 boutons per nuclear cell (range 24-36) (Person and Raman 2012a). There is variation between estimates and between species; for example, in the cat Purkinje cells have been estimated to outnumber nuclear neurons by 26:1 (Palkovits, Mezey et al. 1977).

In more detail for the bouton estimate: ‘We…calculated that each Purkinje cell has a quantal content of ~12-18 (Telgkamp and Raman 2002). With a release probability of 0.5 (Telgkamp, Padgett et al. 2004), we estimate an average of 24-36 functional

contacts per Purkinje cell on each nuclear neuron (Person and Raman 2012b)' (Indira Raman, personal correspondence dated 4 December 2018). 70 out of 86 boutons examined with electron micrographs of mouse medial and lateral cerebellar nuclei had multiple synaptic densities (Telgkamp, Padgett et al. 2004). Out of 10 boutons reconstructed from a single slice all 10 had multiple synapses with an average of 9.2 $\pm$ 1.3 densities per bouton.

4.2.2 Power of veto by a modest fraction of a Purkinje cell population

The strong firing of Purkinje cells and their individually strong contact on nuclear cells mean that a single Purkinje cell may significantly impact on firing of its targets (Pedroarena and Schwarz 2003). Modulation of nuclear cell firing requires the 'substantial co-modulation of a large proportion of the PCs [Purkinje cells] that innervate the cell' (Bengtsson, Ekerot et al. 2011, abstract).

It will be assumed that output of a microzone is not topographically organised – so that a single Purkinje cell makes contact at random on any 4 or 5 nuclear cells (for convenience we use 5) in the nuclear target group. 'To date, there is no evidence to support [the idea] that different PCs [Purkinje cells] of the microzone control specific CN [cerebellar nuclei] cells within the micro-group [associated group of nuclear cells]' (Bengtsson and Jorntell in Apps, Hawkes et al. 2018 p. 663). Assuming microzone dimensions of 40 x 5 Purkinje cells (an estimate for the C3 region in cats: Dean,

Porrill et al. 2010), a single, uninhibited file (mediolateral row) of 5 Purkinje cells would therefore together inhibit between 5 and 25 nuclear neurons, or 10-50% of a nuclear group of 50 output cells.

Direct evidence is lacking for the exact size of a nuclear group. Estimates vary and doubtless so does the actual number. Also, output of a microzone may be received by more than one nuclear location (Pantò, Zappalà et al. 2001), probably as the result of Purkinje cell collateralisation (De Zeeuw, Wylie et al. 1994). Nonetheless, closed circuit organisation seems to be broadly preserved (Pantò, Zappalà et al. 2001). The number is derived here from convergence and divergence ratios.[7] This approach comes with a caveat. In addition to the main excitatory output cells of a nuclear group, there are also 2 types of inhibitory projection neurons which carry output to the inferior olive and back to the cerebellar cortex (Uusisaari and De Schutter 2011). Possibly estimates for divergence and convergence ratios included inhibitory cells inadvertently (because the range for soma size overlaps) or because the estimate does not discriminate.

Microzone dimensions vary. For example, they can span 'mediolateral distances of several tens to a few hundred micrometers (Oscarsson, 1979; Ozden et al., 2009)' (De Gruijl, Hoogland et al. 2014, p.8937). The fatter a microzone, the more Purkinje

[7] 'I would say your estimate [of 50] is not unreasonable, but to my knowledge direct evidence is lacking. Part of the problem is that the level of convergence is likely to vary considerably depending on the cortico-nuclear target, e.g., focused patterns of termination for interpositus versus more widespread termination patterns for fastigius.' Richard Apps, personal correspondence dated 7 January 2017.

cells in a file (other things being equal). The C3 microzone population of 40 x 5 = 200 Purkinje cells (Dean, Porrill et al. 2010) is at the low end of the range of estimates. Calculated by $\frac{Purkinje\ cells * divergence}{convergence}$, and assuming divergence and convergence ratios of 1:5 and 40:1 respectively, a C3 model microzone of 200 Purkinje cells would contact a nuclear group of 25 cells. A nuclear group of 50 cells would need 400 Purkinje cells to preserve contact ratios, which would therefore need to be supplied either by a larger microzone or a multizonal microcomplex (Apps and Garwicz 2005), that is, a circuit which contains more than one microzone. Incidentally, this would give a ratio of Purkinje cells to nuclear cells of 8:1, near the reported ratio for mice (Person and Raman 2012a).

Assuming 9 synapses on a nuclear cell per Purkinje cell bouton, and contact by a Purkinje cell on 5 nuclear cells, a single file of 5 spontaneously active Purkinje cells may inhibit at high frequency as much as half a nuclear group, making around 216-324 synapses on each nuclear cell. The activation of many synapses across several boutons would represent substantial inhibitory drive, from only 2.5% of the Purkinje cells in the microzone.

Adding a second file of active Purkinje cells would result in inhibition by 10 Purkinje cells of 5-50 nuclear neurons representing 10%-100% of the target nuclear group, at 216-324 synapses per cell if input is distributed to 100% of the group, or up to 2,160-3,240 synapses in the improbable event that input converges on just 10% of the nuclear group, and so on. This would also be the result of inhibition by a single active

file in a fatter, 40 x 10 microzone. These examples are intended to illustrate that a small fraction (2.5-5%) of the population of Purkinje cells in a microzone may powerfully inhibit nuclear firing across a substantial proportion of the target group. *How* substantial is illustrated in Figure 4.1.

4.3 Simulation

In this section convergence and divergence ratios of Purkinje cells onto nuclear cells are used to calculate and illustrate how many nuclear projection neurons receive inhibition, and at what strength, at low numbers of active Purkinje cells. The actual number may be lower still, because of the effect of nuclear interneurons (discussed later). That effect is not included in the illustration for lack of quantitative data about nuclear interneurons (specifically how much contact they make and what they make it on). The simulation is accordingly a conservative estimate but with what's intended to be confidence in the data it uses.

4.3.1 20–30 Purkinje cells are enough for most nuclear cells to receive strong inhibition

Figure 4.1A shows the probability that any particular number of nuclear cells receive contact from a minimum or 1, 2 or 3 active Purkinje cells. In the first panel 20 Purkinje

cells are active, in the second 20, and in the third 30. The total Purkinje cell population size is immaterial. The effect depends on the number rather than the proportion of active Purkinje cells. Other numbers of Purkinje cells are not shown because more are unnecessary to illustrate the point that a low number is sufficient for a high probability that all or almost all of the output cells of a circuit receive heavy inhibition.

Part of the purpose of the figure is to show that while the anatomy of contact does not predict the exact proportion of nuclear cells, there is, for each condition, a predictable range (the width of the distribution). The most likely outcome is at the centre of the distribution where the probability is highest.

If m = 5 and n = 50, the probability of contact on a particular nuclear cell by a particular Purkinje cell is $\frac{5}{50} = 0.1$. The probability that a particular nuclear cell receives contact from none out of y active Purkinje cells is therefore $P(0) = (1 - 0.1)^y$, and from exactly 1 = $P(1) = \binom{y}{1} * 0.1 * (1 - 0.1)^{y-1}$. The probability of contact from exactly 2 is $P(2) = \binom{y}{2} * 0.1^2 * (1 - 0.1)^{y-2}$. The probability of contact from 1 or more, or $P(\geq 1)$, is therefore $1 - 0.9^y$. And

$$P(\geq 2) = 1 - 0.9^y - \left(\binom{y}{1} * 0.1 * 0.9^{y-1}\right)$$

$$P(\geq 3) = 1 - 0.9^y - \left(\binom{y}{1} * 0.1 * 0.9^{y-1}\right) - \left(\binom{y}{2} * 0.1^2 * 0.9^{y-2}\right)$$

The probability that exactly 1 nuclear cell out of 50 receives contact from 1 or more Purkinje cells (and the rest receive contact from none) is

$$\binom{50}{1} * [P(\geq 1)]^1 * [1 - P(\geq 1)]^{49}$$

To take another example, the probability that, say, exactly 3 out of 50 receive contact from 2 or more Purkinje cells (and the rest receive contact from none or 1)

$$\binom{50}{3} * [P(\geq 2)]^3 * [1 - P(\geq 2)]^{47}$$

and so on. Or, expanded,

$$\binom{50}{3} * [1 - P(0) - P(1)]^3 * [P(0) + P(1)]^{47}$$

This is what the program calculates (and the graphs show) for each of 1 to 50 out of 50 nuclear cells, in each of 3 conditions: 10 (Figure 4.1Ai), 20 (Figure 4.1Aii) and 30 (Figure 4.1Aiii) active Purkinje cells. Expressed algebraically, for contact from at least x out of y active Purkinje cells on z out of n nuclear cells,

$$\binom{n}{z} * (1 - P(0) - P(1) - \cdots - P(x-1))^z * (P(0) + P(1) + \cdots + P(x-1))^{n-z}$$

$$where\ P(x) = \binom{y}{x} * p^x * (1-p)^{y-x}$$

where $p = m/n$ and m is divergence from Purkinje to nuclear cells.

For values of y greater than the number of Purkinje cells which converge on a nuclear cell, the calculation must be adjusted, because it becomes possible that input

to some nuclear cells is saturated (i.e. to all synapses), affecting the probability of contact on others. Assuming convergence of 30-50:1, y must exceed at least 30 and perhaps as much as 50 to make an adjustment necessary. The amount of an adjustment would be very small at values not much higher than the convergence ratio. Nonetheless the range is limited to 30 here.

4.3.2 How contact is distributed at low numbers of active Purkinje cells

In Figures 4.1Ai-1Aiii we use the number of active Purkinje cells that make contact as a measure of inhibition, and calculate the probability of contact by at least 1 active Purkinje cell, at least 2 and so on. Figure 4.1B shows how this contact is distributed among the cells in a nuclear group: how many receive none, how many receive 1, how many 2 and so on up to the number of Purkinje cells which are active. As before, the probability of contact on a particular nuclear cell by x out of y active Purkinje cells, with divergence of m is

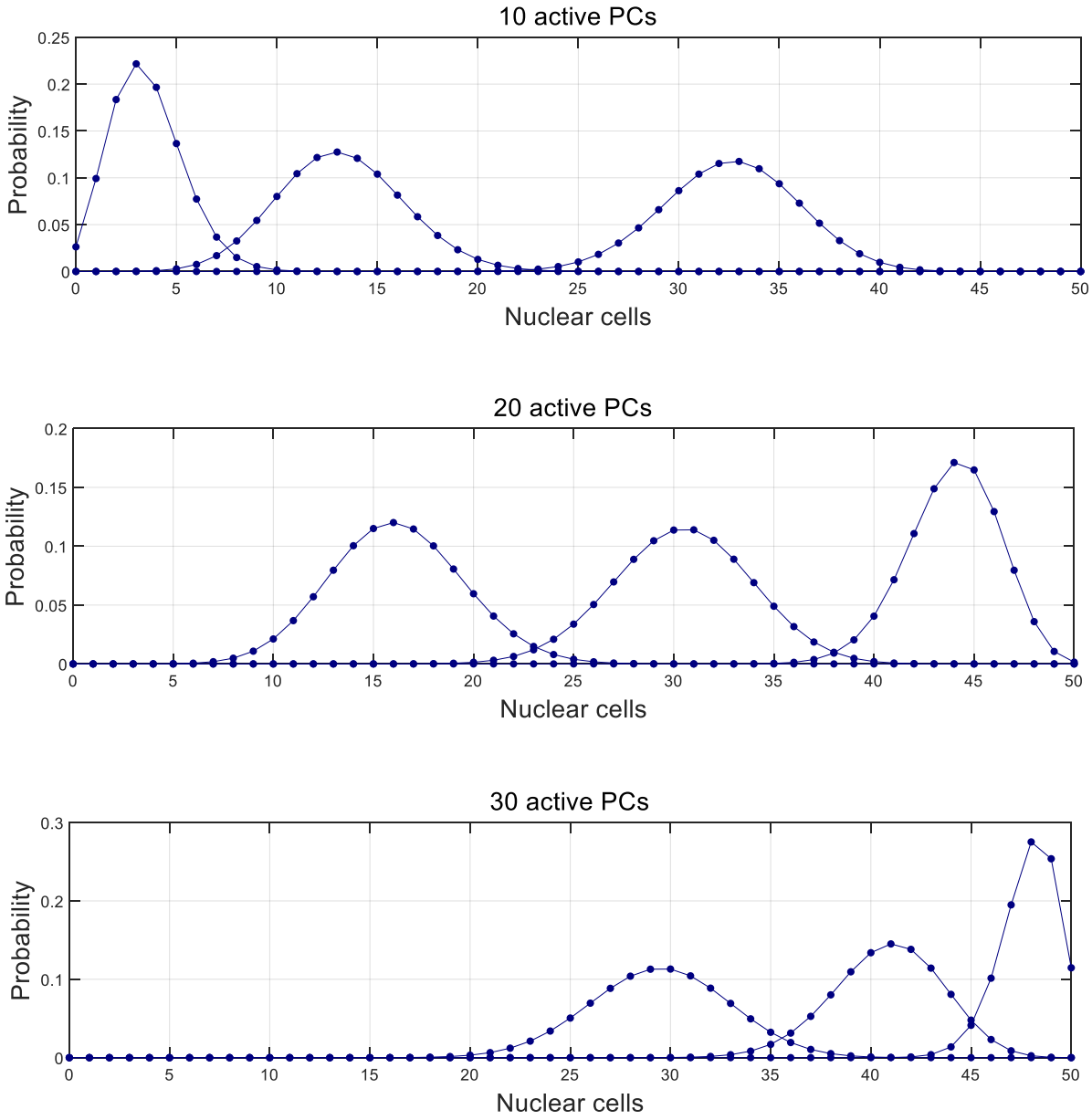
$$P(x) = \binom{y}{x} * \left(\frac{m}{n}\right)^{x} * \left(1 - \frac{m}{n}\right)^{y-x}$$

for a nuclear group size of n, so that the expected number of nuclear cells which receive contact from x active Purkinje cells is $Expected\ number = P(x)n$.

As with Figure 4.1A, at values of y greater than convergence some probabilities are no longer independent. The figure 4.1B simulations do not exceed y = 30.

FIGURE 4.1

Ai - Aiii



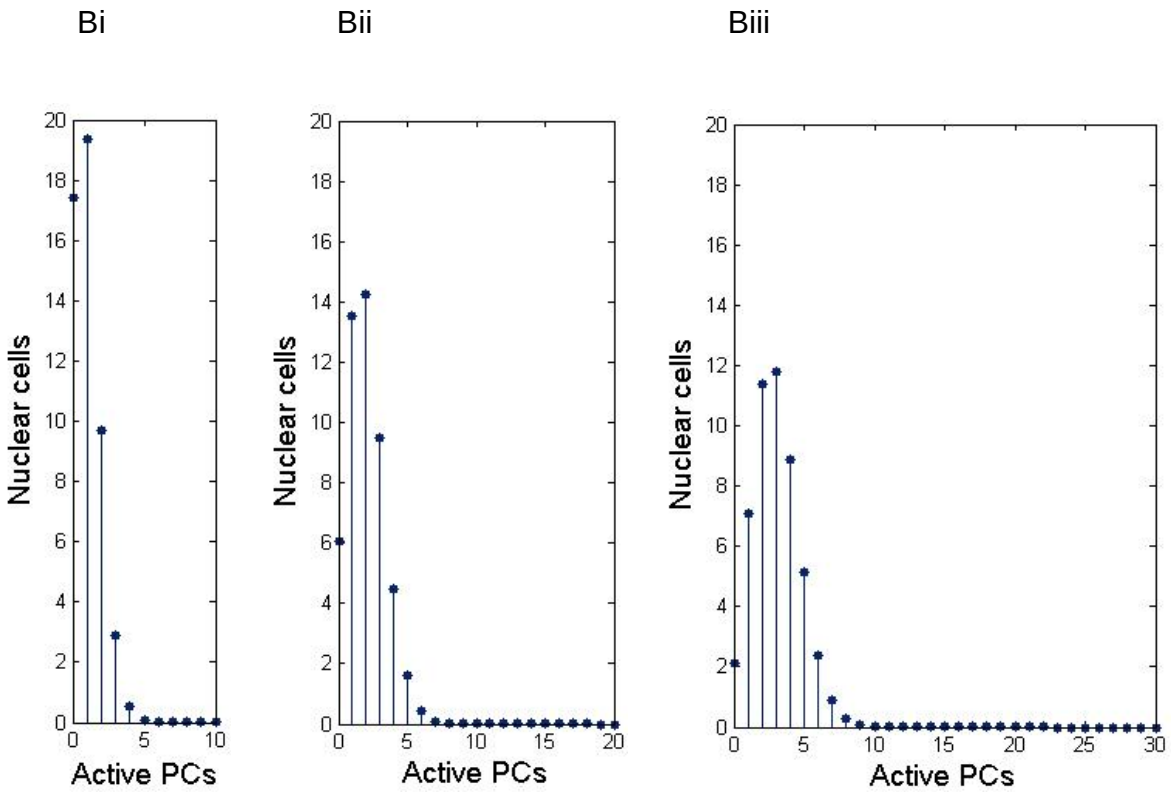


FIGURE 4.1. (Ai - iii) The probability that the number of nuclear cells on the x axis (in the range 1-50) receives contact from 1 or more (right peak), 2 or more (centre), and 3 or more (left) out of 10, 20, or 30 (Ai - iii respectively) active Purkinje cells. The important feature is not the height *per se* (y axis calibration varies to fit the data) but how near the right the peaks are, and the narrowness of the majority of the area under the curve (indicating a narrow range of the most likely number of nuclear cells). In Aii (20 active Purkinje cells), for example, it is likely that all but a small number of nuclear cells receive contact from at least 1 Purkinje cell and the majority receive input from at least 2. **(Bi - iii)** Stem diagrams showing the expected number of nuclear cells receiving contact from 0 active Purkinje cells, from 1, from 2 and so on, where the total number of active Purkinje cells is 10 (Bi), 20 (Bii) and 30 (Biii). With 10 active Purkinje cells, a

probable ~35% of nuclear cells receive no contact at all, suggesting that 10 Purkinje cells may not be enough to veto output of a nuclear group (without help from nuclear interneurons). With 30 active Purkinje cells only 2 nuclear cells receive no contact from active cells and over 80% receive contact from 2 or more.

4.4 Contact of Purkinje cells on nuclear cells is wired for a default veto

The anatomy of contact by Purkinje cells on the output cells of the cerebellar circuit allows us to draw inferences about its function, assuming random distribution of contact by a Purkinje cell within a nuclear group, and convergence and divergence of 40:1 [check simulations] and 1:5 respectively.

1) The inhibition of a nuclear group reaches functional saturation efficiently, that is, inhibition by Purkinje cells of nuclear cells is broadly equally distributed even at low numbers of active Purkinje cells. The distribution is relatively flat, such that most nuclear cells receive contact from 1 or 2 active Purkinje cells before many receive contact from 3 or 4 (Figure 4.1 all panels).

2) A modest number of spontaneously active Purkinje cells is enough to strongly inhibit the whole of a nuclear group. These do not have to be clustered together but can be any group of the necessary minimum number. Indeed a group could be distributed across more than one microzone if those

microzones are part of the same multizonal microcomplex (Apps and Garwicz 2005). However, it is likely that a file of Purkinje cells behaves as a group because they sample the same parallel fibre activity.

3) A handful of Purkinje cells is sufficient to veto output of the whole circuit assuming the effect of inhibition of a nuclear cell is not in proportion to the number of Purkinje cells but depends on exceeding a threshold (which we propose is met by a low number of Purkinje cells because a low number is enough to make strong contact). Therefore the coordinated suppression of Purkinje cell firing across the whole population of a microzone is necessary in order to disinhibit a nuclear cell group, because less than full coordination means that each nuclear cell is at high risk of receiving strong inhibition. (Note 'coordinated' and not – necessarily – strong suppression.)

4) Most of the inhibition of a nuclear cell is functionally supernumerary for most of the time (because the threshold can be met by a much lower number of cells). However, this is necessary in order that it can be any handful of Purkinje cells, in any part of a microzone, which blocks output of the circuit if that location receives a mismatch. Pattern detection is discussed in section 4.6 and physiological adaptations (to routinely excessive inhibitory bombardment of nuclear cells) in the Discussion section.

4.5 Adaptations that may assist a veto by a low number of Purkinje cells

Efficiency of the inhibition of a nuclear group by a low number of Purkinje cells may be increased by nuclear interneurons, and by the distribution of inhibitory and excitatory contact on nuclear projection neurons (so that these may be mechanisms that help a small number of Purkinje cells to punch above their weight).

4.5.1 Excitatory nuclear interneurons

Deep nuclei contain excitatory (presumed glutamatergic) interneurons that fire spontaneously and which are inhibited by Purkinje cells (Uusisaari and Knopfel 2008). Accordingly, they provide spontaneous drive to their targets, subject to Purkinje cell restraint. The convergence and divergence ratios of Purkinje cells onto interneurons and of interneurons onto nuclear cells are unknown (and the internal circuitry of deep nuclei generally is poorly understood).[8] However contact by Purkinje cells on interneurons and presumed contact of interneurons on nuclear cells may increase divergence of suppression of nuclear firing by a Purkinje cell, because firing of a Purkinje cell has a suppressing effect on more nuclear cells than only the ones it contacts directly. It may also increase the strength of suppression of the nuclear cells

[8] For example, 'The shape of the Purkinje cell axon termination field in the CN [cerebellar nuclei] varies a lot between subnuclei, and there are no reliable data on differential targeting of the CN cell types (and the distribution of cell types varies among the CN subnuclei)' (Marylka Uusisaari, personal correspondence dated 8 January 2017). Also, projection neurons have been found to have axon branches that terminate in deep nuclei, thereby giving them an interneuronal effect.

it contacts directly if these also receive an indirect effect through excitatory interneurons.

GABA/glycinergic interneurons (Husson, Rousseau et al. 2014), and possibly others (Uusisaari and De Schutter 2011), have also been reported. Inhibitory interneurons are reported to probably receive light or no inhibition from Purkinje cells (Uusisaari and De Schutter 2011), so that there is little or no simultaneous (competing) effect on nuclear cells acting through inhibitory interneurons.

To illustrate the (possible, proposed) effect of nuclear interneurons, Figure 4.2Ai-iii shows the effect of higher divergence on the expected number of nuclear cells that receive Purkinje cell contact, and Figure 4.2Bi-iv graphs the probability that x nuclear cells receive contact from at least 1, 2 or 3 Purkinje cells, where 5, 10, 15 or 20 Purkinje cells are active, with a divergence ratio of 1:10. This is the same idea as Figure 4.1Ai-iii but with higher divergence. Now, half as many active Purkinje cells are needed for around the same probable contact, and with 20 active Purkinje cells virtually full coverage of a nuclear group is effectively certain – just 5% of the estimated afferent population assuming a nuclear group of 50.

The higher divergence ratio is hypothetical – it is not known how many nuclear cells may receive an indirect influence of a Purkinje cell through interneurons. However, Figure 4.2B indicates that there would be a robust rise in the probable number of

nuclear cells that receive an effect for only a relatively modest rise in divergence, or what amounts to divergence.

In a sentence, it takes fewer Purkinje cells for a high chance that all or very nearly all of the cells in a nuclear group receive strong contact, so that fewer Purkinje cells need to be free firing (because they receive a mismatch) for a veto.

4.5.2 Derivation

Derivation of Figure 4.2Ai - iii. For divergence of 1 to m and a nuclear group of n, the probability of contact on a particular nuclear cell by a single Purkinje cell is m/n. The probability that a particular nuclear cell receives contact from x out of y active Purkinje cells is given by

$$P(x) = \binom{y}{x} * \left(\frac{m}{n}\right)^{x} * \left(1 - \frac{m}{n}\right)^{y-x}$$

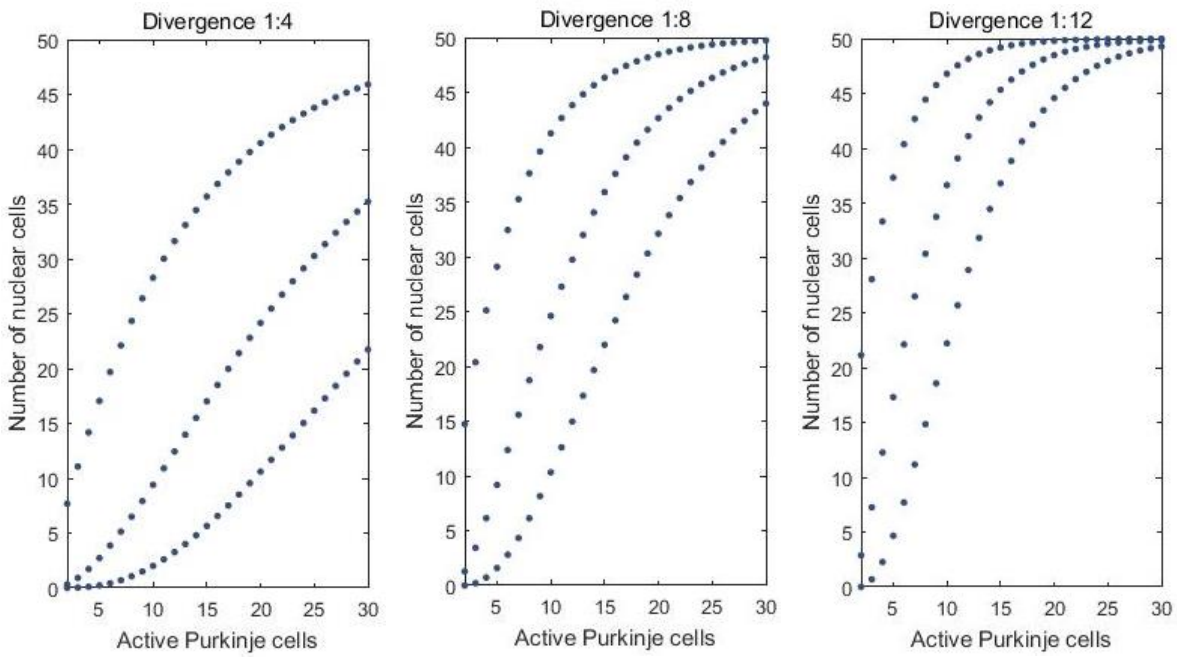
The probability that a particular nuclear cell receives contact from x = 1 or more Purkinje cells is accordingly $1 - P(0)$, multiplied by n for the expected number of nuclear cells receiving it. The probability for x = 2 or more is $1 - P(0) - P(1)$, again

multiplied by n for the expected number, and for x = 3 or more is $1 - P(0) - P(1) - P(2)$, and so on.

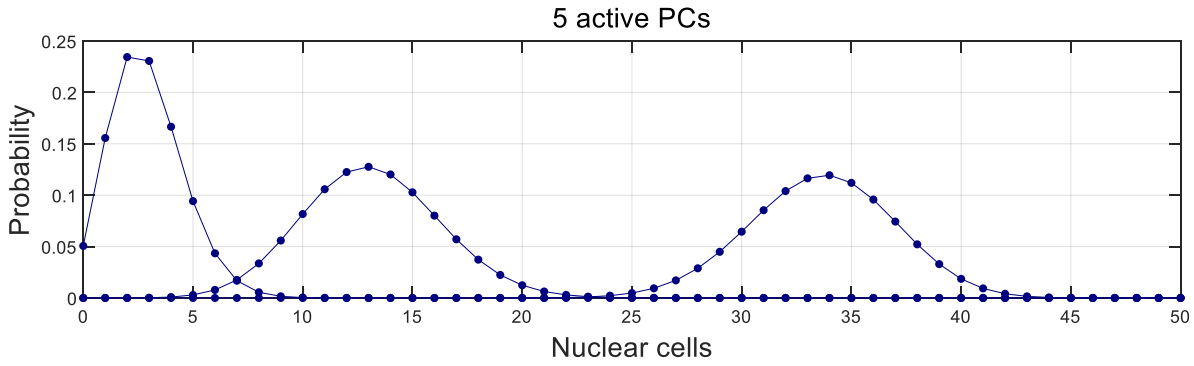
For derivation of Figure 4.2Bi - Biv see section 4.3.1.

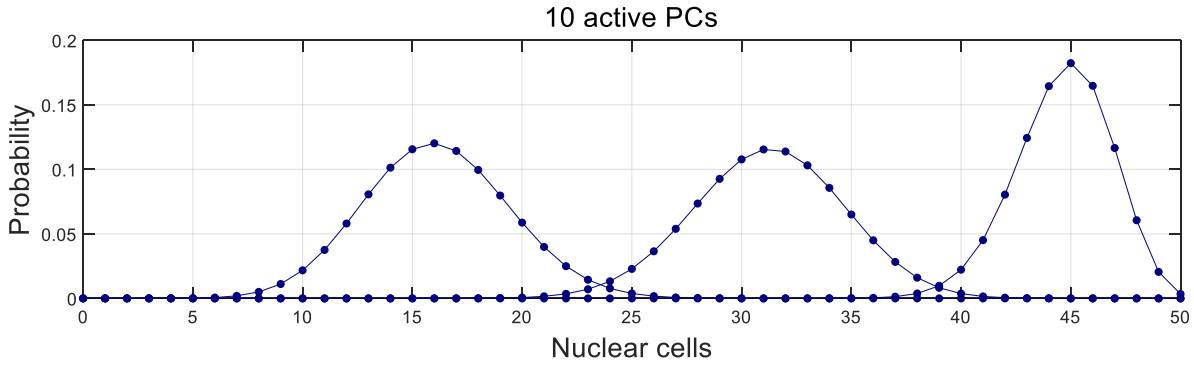
FIGURE 4.2

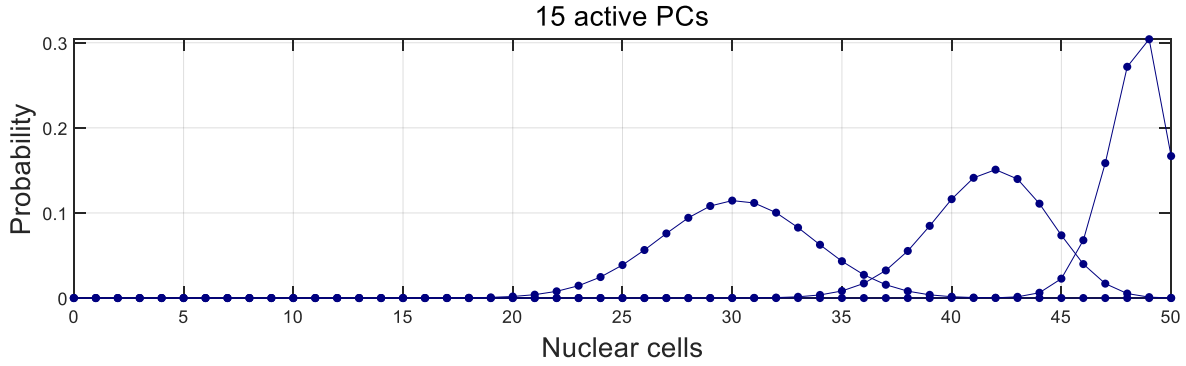
Ai Aii Aiii



Bi – Biv


5 active PCs
Probability
Nuclear cells


10 active PCs
Probability
Nuclear cells


15 active PCs
Probability
Nuclear cells

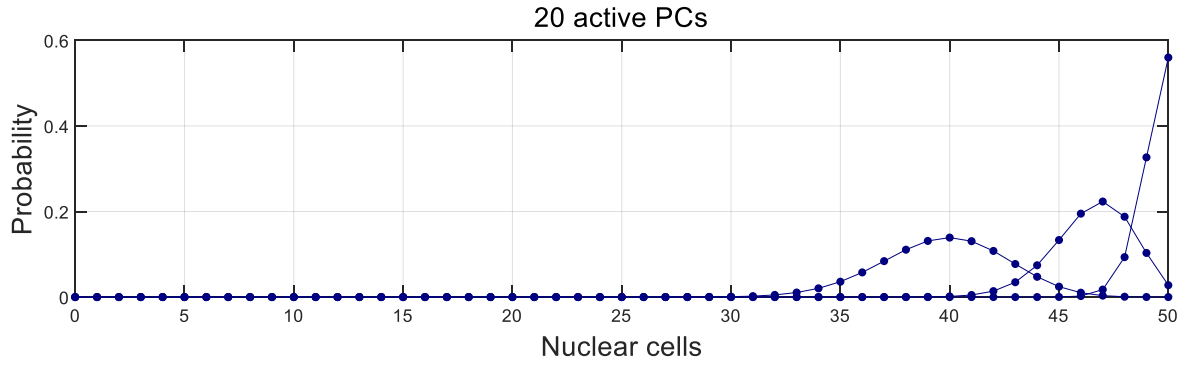

20 active PCs
Probability
Nuclear cells

FIGURE 4.2 (Ai - iii) The expected number of nuclear cells, out of a group of 50, which receive contact from at least 1 Purkinje cell (top curve), at least 2 (middle curve) and at least 3 (bottom curve), out of a number x which are active. A steeper curve indicates that fewer active Purkinje cells are needed to inhibit a nuclear group. Efficiency is in proportion to the divergence of Purkinje cell contact on nuclear cells. (Note the proportion of Purkinje cells which are active – as opposed to the number – makes no difference to their effect.) The estimated divergence ratio in rats is 1 to 4 or 5. With divergence of 1:4 (first panel) around 20 active Purkinje cells are needed for an expected 40 out of 50 nuclear cells to receive contact from at least 1, and only around half a nuclear group receives contact from at least 2. The divergence ratio is doubled in the second panel. Now, 20 active Purkinje cells are enough for almost all nuclear cells to receive contact from at least one, and for over 40 to receive contact from at least 2. In the third panel, with divergence of 1:12 and 20 active Purkinje cells, all nuclear cells receive contact from at least one active Purkinje cell and almost all from at least 2. **(Bi - iv)** The probability that a number of nuclear cells, x, receives contact from 1 or more (right peak), 2 or more (centre), and 3 or more (left) out of 5, 10, 15 or 20 (Bi - iv respectively) active Purkinje cells. A hypothetical illustration of higher functional divergence of Purkinje cells onto nuclear cells acting through excitatory nuclear interneurons, discussed in the main text (in section 4.5.1).

4.5.3 Distribution of excitatory and inhibitory contact on nuclear cells

Contact by Purkinje cells on nuclear cells ‘is characterised by preferential targeting of cell somata rather than dendrites’ (Uusisaari and De Schutter 2011, p.3443), while the majority (75%) of excitatory inputs are distal (de Zeeuw and Berrebi 1996). Purkinje cell synapses are therefore positioned to block an effect of excitatory input to nuclear cells. This powerful inhibitory veto is strengthened by inhibition of excitatory interneurons (Uusisaari and Knopfel 2008) because it weakens or silences intrinsic firing of interneurons, so that the tonic effect of intrinsic Purkinje cell activity is both to directly inhibit nuclear cells and to block tonic inhibition.

An example of a comparable inhibitory block is provided by stellate cells – interneurons in the molecular layer. Stellate cells contact and inhibit each other at the soma. Inhibition shortens the integration window for excitatory input, which is mostly dendritic, and weakens transmission of subthreshold dendritic signalling (Häusser and Clark 1997), so that an effect of excitatory input, to which stellate cells are otherwise sensitive, is only seen if tonic inhibition is removed (Carter and Regehr 2002).

The significance of a strong inhibitory bottleneck is that it increases the potency of inhibition by a Purkinje cell of its nuclear targets, so that fewer are need to be potent. This is consonant with the proposal of an inhibitory threshold set where the majority

of Purkinje cell input to a nuclear cell is normally supernumerary (because the threshold can be met by a much lower number, and maybe even by only one or two Purkinje cells).

4.6 Overlap of stored patterns is well tolerated and functional

This section and the next one bring together the conclusions of the analysis with the conclusions of Chapter 3.

Contrary to traditional learning theories there is substantial overlap of patterns stored as trained parallel fibre synapses, challenging the idea that synaptic memory is stored as bespoke weights. The estimate in Chapter 3 of the density of parallel fibre activity (number active per unit area in the sagittal plane) would mean that a Purkinje cell receives contact from several hundreds at a time (n = ~900, assuming a Purkinje cell sagittal span of 300 µm). Because activity is evenly and randomly distributed at the scale of input to a Purkinje cell (Chapter 3), training teaches an evenly and randomly distributed pattern of synapses. Pattern density is regulated, uniform and ubiquitous. This consistency is undisturbed by folding of the cerebellar cortex because the size and shape of the Purkinje cell arbour and the thickness of the granular layer are adjusted so that the number of parallel fibres that contact a Purkinje cell is unaffected (to reason from the unaffected number that intersect the Purkinje cell dendritic field) (Eccles, Ito et al. 1967).

The narrow range of the regulated density of active parallel fibres causes a predictable result of overlap of stored patterns. The amount of overlap is predicted by the number stored, and the same for all patterns. The proportion of synapses which also belong to 1 other pattern, and to 2, and 3 and so on is also predictable (Figure 4.3) and also the same for all stored patterns. As more patterns are stored the ratio of total trained to total untrained (with paired input) synapses[9] shifts towards more trained, in predictable proportions. Moreover the split is the same for all Purkinje cells trained to the same number of patterns and therefore all Purkinje cells in the same microzone (and all microzones in the same circuit), because climbing fibre signals are received in a volley across the whole population, and are always paired (an effect of the regulated density of parallel fibre activity: Chapter 3).

An estimated 80-85% of parallel fibre-Purkinje cell synapses are strongly long-term depressed, to the extent that there is 'no detectable somatic response' to granule cell stimulation (Isope and Barbour 2002 p.9676). This is consistent with a high estimate of 'electrically silent' synapses made by parallel fibres activated by cutaneous stimulation (Ekerot and Jörntell 2001). The probability that k out of a random pattern of n active parallel fibres make contact at depressed synapses is $\frac{n!}{k!(n-k)!} * 0.85^k * (1-0.85)^{n-k}$, so that the probability that the proportion k of n is between, say, 83% and 87% is given by $P\big(k \geq 83(n/100)\big) - P\big(k > 87(n/100)\big)$, or

[9] Here and elsewhere references to untrained synapses do not mean there is no plastic effect on transmission of experience but that they have not been trained with a course of paired input.

$$\sum_{k=83(n/100)}^{87(n/100)} \frac{n!}{k!\,(n-k)!} * 0.85^{k} * (1-0.85)^{n-k}$$

Using our previous estimate of n, this gives a probability of 0.9841, rising to a near certainty if the range is only slightly increased. (This paragraph is to illustrate why the split and proportions in the previous paragraph are predictable.)

Overlap is a problem for the graded-synaptic-weight model because it means synaptic weights are not bespoke, so that the response to a set of inputs receives interference from other stored patterns. Moreover, trained synapses do not, arguably, *have* graded weights (because they are silent, Isope and Barbour 2002), at least parallel fibre-Purkinje cell synapses. The Isope and Barbour detection threshold would in theory leave room for compound responses (Boris Barbour, private correspondence dated 7 December 2018). But as activation of 150 operating (that is, *non*-depresse d) synapses is needed to generate an action potential, it would suggest that input to depressed synapses at substantially larger numbers would be necessary for an effect. Also, an effect would not be bespoke but standard. Only around 15% of a known pattern is received at synapses trained only to that pattern. The rest is to synapses that participate in more than one pattern – some participating in 2, some 3, and so on – in well-predicted (and therefore the same) proportions on all Purkinje cells (Figure 4.3).

That stops being a problem, however, if trained synapses are not graded but functionally inert, and the cerebellum does not, and does not need to, remember patterns individually. Then, overlap is immaterial (or put another way well tolerated), because it means there is no effect of unrelated memories on the response to other patterns. Trained synapses are (functionally) the ‘same’ weight whether they are part of one pattern or several. The function of parallel fibre-Purkinje cell synaptic depression is not to store bespoke weights, as once assumed, but to simply render transmission functionally negligible (it is suggested). Graded weights are not necessary for this task – in fact graded transmission would impair function.

Overlap is not merely tolerated but increases storage capacity. The ability to ‘reuse’ a synapse – incorporate it into more than one pattern (indeed any number of patterns) – means that more patterns can be stored before the supply of synapses is exhausted, which is where the trained:untrained ratio reaches the limit that does not interfere with class discrimination. Class discrimination is discussed in the next section.

FIGURE 4.3 A and B

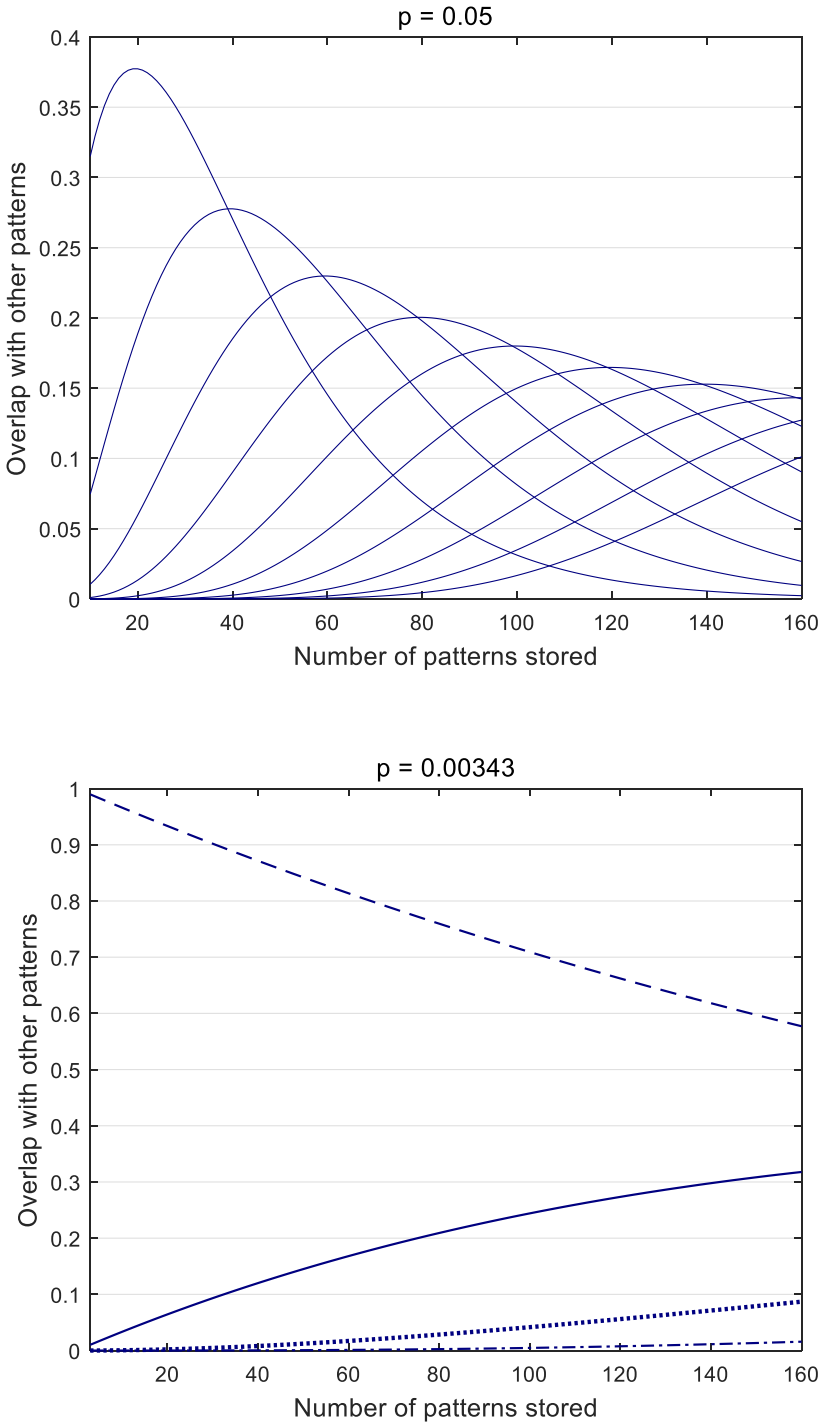


FIGURE 4.3 The changing relative proportions of a pattern which overlap with 1 other pattern, and with 2, and 3 and so on as more patterns are stored is given by a binomial distribution where the solution is the proportion of each pattern (the same for all of

them) that overlaps with k other patterns, $n+1$ is the total number of patterns stored, and p is the fraction of parallel fibres that are active (in reality a constant, because the level of parallel fibre activity is regulated (Chapter 3)). In (A) a pattern contains 5% of the total number of parallel fibres. The left peak is the proportion of synapses which also belong to one other pattern, the next one to the right is the proportion that also belong to 2 other patterns, the next is the proportion that belong to 3 other patterns, and so on, up to 10 other patterns. The x axis starts at 10 for convenience in running the calculations. (B) assumes 1,200 out of 350,000 parallel fibres are active, the estimated number at internally regulated levels of activity (Chapter 3), so $p = 0.00343$. The pattern is greatly stretched as if pulled from the right. (The purpose of A, with a higher, non-physiological number active, is to show the relationship of the relative proportions of overlap not evident only from the range in B.) The dashed line is the fraction of a pattern that does not overlap with others. The solid line is the proportion of synapses which also belong to one other pattern. The dotted line is the proportion that also belong to 2 other patterns, and the dots and dashes are the proportion that overlap with 3 other patterns. Overlap with 4 and higher numbers is either extremely low or absent in the displayed range.

Important note: There remains a functional requirement of separation and sparsening by recoding in the granular layer (and this is important), but separation now means optimal rather than complete, and is provided by decorrelation (Chapter 3) (and large targets).

4.7 Class discrimination and binary pattern matching

If the cerebellum does not discriminate between patterns (because they overlap) what does it learn? Put another way, in what way is the response different as a result of learning?

Classification of patterns is binary. A known pattern of input to a Purkinje cell (or to a microzone) is received at exclusively inert synapses, and unknown patterns are received at a well-predicted split. The effect of the latter is through a random sample of 15% of their signals which are received at synapses which are not trained. The postsynaptic response is therefore confined to a binary range: a negligible effect or the effect of the randomly sampled 15% that is received at a random sample of operational synapses. That is to say, discrimination is between the class of learned patterns and the residual class of input in any other configuration. There is no need for discrimination between individual patterns, because learned patterns do not control the output firing rate (in this contention). The role of pattern matching in the conditioned response is instead permissive – it selects which circuits have output and when, but does not code output. Control of output rates is not covered in this paper, only pattern detection.

Only a full match with a broadside of known input, received along the whole length of a microzone, is sufficient to displace a default veto (if the inferences in section 4.4 are correct). A failure to meet that threshold does not trigger a graded response but no response at all, regardless of how good or bad the match is overall, or which parts are good and which parts are bad. There is no proportionate or middling (or any other) response to a partial match. This is important because, without knowing what signals code (because the cerebellum has no information about the signals it receives except what it learns), the response to a partial match would be arbitrary. In this sense, counter to traditional expectation, pattern matching controls the response to input that a circuit does *not* recognise, and not patterns it does. That response is the ongoing and functionally unweakened inhibition of nuclear cells, that is, of its own output. This is accordingly the default state, interrupted by a full and accurate match.

4.8 Discussion

4.8.1 Note

This chapter models the contact on a nuclear group of a low number of Purkinje cells and discusses the effect if they fire at intrinsic (and therefore robust) rates. This is not to say that firing of all Purkinje cells must be silenced or strongly depressed for an effect on nuclear cells. Rather, the proposal is simply that if only a few Purkinje cells

are omitted from co-modulated firing of the rest of the population, such that those few fire at intrinsic rates, they override the effect of the others.

The effect of a mismatch, even confined to a small fraction of the area (in the sagittal plane) of a microzone, is accordingly to veto an effect of even a perfect match received everywhere else. A discussion of the effect of unsynchronised Purkinje cell rates in other scenarios is intentionally left to one side – control of nuclear rates is discussed in Chapter 6.

The rest of the Discussion section is divided between a discussion of adaptations of Purkinje cell-nuclear transmission with mitigate the problems associated with default torrential inhibition of nuclear cells, and a discussion of the proposal that the design of the cerebellum is as much to eliminate an unwanted effect of redundant variables as it is to act on ‘good’ data (input variables the cerebellum is designed and able to exploit).

Note: the present focus is on circuits involved in the control of axial and limb movements, partly because the impressive work at Lund has made a large amount of data available, much of it for the C3 region of the cerebellar cortex, which is involved in forelimb control. Part of an explanation of circuit function is likely to be portable (apply to other circuits) if an to the extent other regions share, as thought, the same basic circuit wiring. However, there are also regional differences which are likely to

be important, so that it is unsafe to pool the evidence from different regions. For this reason the literature of control of eye movements, for example – although the most studied cerebellar behaviour – does not form part of the present analysis, because the operation of vestibular circuits may be different.

4.8.2 Adaptations that mitigate high convergence and strong afferent rates

A power of veto by a modest fraction of Purkinje cells would suggest that most Purkinje cell firing is supernumerary almost all of the time. On the face of it this is wasteful and counter to the usual expectation that biological designs are efficient. Contact by Purkinje cells on nuclear cells is adapted to temper the inhibitory bludgeoning received by nuclear cells and mitigate the heavy demands on Purkinje cells, in the (proposed) ways discussed in this section.

4.8.2.1 Short-term depression and synaptic facilitation

Inhibition of a nuclear cell by 30-50 intrinsically-firing Purkinje cells each making contact at 200-300 synapses (referenced previously) might be expected to hold nuclear cells in an inactive state of hyperpolarisation. Instead, nuclear cells sustain baseline firing (at 10-20Hz in resting animals: [Telgkamp and Raman 2002]). This is partly because a steady rate of simple spike firing rapidly weakens its own synaptic

transmission (Telgkamp and Raman 2002, Pedroarena and Schwarz 2003, Telgkamp, Padgett et al. 2004). Insensitivity to the rate of steady state inhibition is actively modulated by short-term synaptic facilitation which varies with the presynaptic rate, so that transmission is frequency independent (Turecek, Jackman et al. 2017).

Nuclear cells remain sensitive, however, to the dynamics of the simple spike rate (Telgkamp and Raman 2002), such that their firing is transiently modulated by a *change* of the rate, which can be in either direction (Pedroarena and Schwarz 2003, Baumel, Jacobson et al. 2009). This arrangement thus contrives to hold nuclear cells in a state which remains sensitive to modulation of the simple spike rate but is at a stable background rate otherwise, regardless of the simple spike rate.

4.8.2.2 Bouton design

Purkinje cell bouton design is adapted to provide a reliable and fast response to dynamics, to prevent depletion of the neurotransmitter supply by sustained heavy firing, and (it is suggested) to prevent an effect leaking to other nuclear groups.

Adaptation is in the form of ‘many specializations…[which include] large boutons, glial ensheathment, GABA transporters confined to astrocytes, [and] multiple release

sites' (Telgkamp, Padgett et al. 2004 p.123). Neurotransmitter depletion is avoided by what may be a mainly presynaptic form of depression (Pedroarena and Schwarz 2003). The result is spillover-mediated transmission with 'a high response probability of postsynaptic receptors, without an unusually high release probability' (ibid, abstract). Spillover confinement to multisynaptic Purkinje cell boutons, with transporters confined to astrocytic processes at the bouton perimeter, maintains intrabouton GABA concentration at a level which permits conservation (by reduced release) by Purkinje cells, and slows down escape from synapses (into the intrabouton space, down the concentration gradient).

The same arrangement (in this contention) also importantly restricts diffusion of extrasynaptic GABA, i.e. escape from boutons, thereby limiting an effect on nuclear cells to the group that receives contact, and preventing an effect on other groups, even if group boundaries contain some intermingling. Otherwise, independent and functionally discrete control of nuclear groups would receive interference from diffusion of spillover from other groups.

4.8.3 Data can take multiple forms; the form is mutable at different steps in transmission; different functions use different forms; an effect of redundant forms is blocked

It forms part of the thinking behind this paper and the wider model it is part of that cerebellar design is equally concerned with blocking an effect of ‘unwanted’ input variables as with exploiting an effect of useful ones. Unwanted means variables that the cerebellum is not designed or able to use, and which therefore have an arbitrary effect on internal signalling, output and function, and might therefore be said to be unintelligible form a cerebellar view. In traditional cerebellar theory information is thought of as being contained in coding of individual signals, in the form either of an averaged spike rate (rate coding), or the particular timing of individual spikes (time coding).

But where input is from more than one signal there are also many other variables, such as the number of cells that are co-active, the pattern they are active in, the spatial configuration of the rates they fire at (which fire at what rates), the frequency distribution of rates in a volley of input, and so on. This is equally true on the scale of input to a single cell as of input to a functional group, such as a microzone. It is therefore important (to correct cerebellar function) to isolate an effect of wanted variable(s) by blocking an effect of others.

A functional variable is the form that information takes at that stage in transmission. A stage may contain information in more than one form. Also, the form may change from one stage to the next – information expressed as a particular variable (or variables) in one stage may be expressed as a different variable (or variables) in the next. Among other things, this permits multiple variables to be converted into a form whose effect can be isolated so that it does not interfere with other functions. Pattern matching is an example.

Input variables used in pattern matching are the spatial pattern of mossy fibre signals (which cells are on or off), the number that are active (because that affects the spatial pattern) and the relative frequency they each fire at – the pattern of input rates. These data are represented internally (for the purpose of pattern detection) as the spatial configuration of active granule cells/parallel fibres and the timing with which they are active together – so timing relative to each other, as well as to a climbing fibre signal.

Thus the physical form of data – as well as taking the shape of a stage-dependent number of variables – is mutable at different steps in transmission. The effect of the recoded, internal form is isolated by reducing the number of internal variables, to exclude an effect (on pattern matching) of anything other than the pattern of active cells. Excluded are: the number of cells per pattern, the spatial configuration, the de facto density, and the rates they fire at, because these are either invariant (the first three) or unrepresented in the response (because the response is binary, the fourth).

(Other rate-related variables are also not represented in the pattern of internal activity or therefore the effect bit has.)

The data contained in the internal pattern is that it is input specific (like fingerprinting), and timing. There is no effect of it on granule cell/parallel fibre rates, and no effect of rates on it. This is the same principle as removing or blocking an effect of other variables but instead variables are independent. Rates are used in control of output rate coding. Thus pattern matching and output rate coding, operating side by side, do not receive interference from each other, because the data they use takes different forms that vary independently.

Thus, the form of data (that is, the variables) used in pattern matching is consistent with the proposed requirement of the design of the cerebellum to isolate, at each step in transmission, the effect of ‘selected’ variables and remove or block the effect of others.

CHAPTER 5

STELLATE CELL NETWORKS

5.1 Introduction

5.1.1 Background

Microzones are 15-20 mm long and only a few cells wide (Oscarsson 1979). They are defined by their climbing fibre input so that learning has the footprint of a microzone. It was argued in Chapter 3 that the density of parallel fibre activity self-regulates homeostatically so that it is confined or attracted to a stable and relatively narrow range (Chapter 3). Regulation is mediated by recoding in the granular layer. Uniform density is partly the result of quasi-randomisation by recoding of the permutation of active parallel fibres.

Mossy fibre input to the cerebellum is vertically topographically organised (reported for the C3 region: Jörntell and Ekerot 2006, Quy, Fujita et al. 2011). Topography is preserved in the molecular layer (Palay and Chan-Palay 1974, Zhang and Linden 2012), such that superficial granule cells generally bifurcate in the outer molecular layer and deeper granule cells at deeper level. Stellate cells occupy roughly the outer

two thirds of the molecular layer in mammals and birds, whose innermost layer is occupied by basket cells, so that deeper stellate cells occupy the middle third.

It is a mainstay of learning theories that the cerebellum remembers patterns stores as synaptic weights. Training with paired parallel fibre and climbing fibre input teaches long-term depression (LTD) at the parallel fibre-Purkinje cell synapse (Ito 1989, Hansel and Linden 2000, Qiu and Knopfel 2009, Hoxha, Tempia et al. 2016) and long-term potentiation (LTP) at the parallel fibre-stellate cell synapse (Jörntell and Ekerot 2003, Rancillac and Crépel 2004, Smith and Otis 2005, Jörntell and Ekerot 2011). Following training there is a transient reduction or suspension of simple spike firing in response to a repeat of the same parallel fibre stimulation (Jirenhed, Bengtsson et al. 2007, Rasmussen, Jirenhed et al. 2008), thought to be partly mediated by feed-forward inhibition of Purkinje cells by stellate cells.

5.1.2 Proposals

The following proposals are made.

In the main text:

1. Mid-level stellate cells in fact have bad pattern memory. That is, the response is the same to parallel fibre input whether it is a pattern they are trained to or not, except for a slight probability bias.

2. The amount of contact made by stellate cells on each other is depth-dependent. Functional networking – the supply of tonic inhibition by stellate cells to each other – is confined to mid-level.

3. At normal, regulated levels of parallel fibre activity (Chapter 3) the effect of excitatory input to networked stellate cells is insignificant (because an effect is blocked by tonic inhibition and for other reasons). This is the default mode. A default block is a primary function of mutual inhibition of stellate cells (that is, it provides a reason they are networked and fire intrinsically).

In the Discussion:

4. This suggests that stronger – that is, more numerous – excitatory input is needed to overcome the veto. A good candidate is nucleocortical feedback, present in the conditioned response and terminating as a sagittal strip of mossy fibre input to the same circuit. The feedback pathway does not directly contact stellate cells but provides more input to underlying granule cells so that more fire (because more receive enough input). This creates an adaptively-timed window in which stellate cells are sensitive to excitatory input. As stellate cells have bad pattern memory, this does not need to be input in a learned pattern.

5.2 Mid-level stellate cells are pattern blind

The first step to justify proposal no.1 is to estimate the number of inputs to a stellate cell.

5.2.1 How many active parallel fibres in a random pattern make contact on a stellate cell?

It has previously been argued that parallel fibre activity is randomly and uniformly distributed and that pattern density is functionally invariant. It will be assumed that stellate cells receive contact from a random sample of passing parallel fibres, and therefore receive a random sample of parallel fibre activity. It follows that within these constraints we can use any random pattern to estimate the probable number of active parallel fibres that make contact on a stellate cell (for present purposes we can disregard whether contact is to trained or untrained synapses).

A mid-level stellate cell dendritic field (in adult rats) extends around 110-130 μm in both horizontal and vertical directions (it is ~80 μm in both directions for outer-level stellate cells) (Palay and Chan-Palay 1974, pp.217-221). Stellate cell dendritic and axonal fields are severely flattened in the same plane as Purkinje cells, with which they are interleaved (Eccles, Ito et al. 1967). A Purkinje cell receives an estimated

175,000 parallel fibre synapses (Napper and Harvey 1988), one each from around the same number of granule cells (Harvey and Napper 1991), while a roughly equal number of parallel fibres pass through the same space without making contact. Assuming a mid-level stellate cell dendritic field of 120 x 120 µm, a C3 Purkinje cell arbour is therefore around 6 times the size. So that by this coarse estimation some 350,000/6 or ~60,000 parallel fibres pass through a mid-level stellate cell dendritic field, assuming parallel fibres are evenly distributed.

It is thought a stellate cell receives 'several hundred excitatory inputs from granule cells' (Palay and Chan-Palay 1974, Carter and Regehr 2002, p.1310), perhaps as many as 1,000 (Boris Barbour, private correspondence dated 26 March 2015). 1,000 would mean around 1 in every 60 parallel fibres which pass through a field makes contact (i.e., 1,000 out of 60,000). Of 60,000 parallel fibres which pass through a stellate cell territory, around 300 are active, i.e., 1 in ~200 based on the estimate of an average of 1,800 active granule cells in a Purkinje cell arbour-sized beam (based in turn on Chapter 3). There is accordingly a probability of 1/60 = ~0.01667 that an active cell makes contact. The probability of contact by k (out of 300) active cells is therefore given by

$$\frac{300!}{k!\,(300-k)!} * 0.01667^{k} * (1-0.01667)^{300-k}$$

Eq.1

This gives the following probabilities for the range k = 0 to 10.

Table 5.1 – The probability of contact on a mid-level stellate cell from k active parallel fibres

k	0	1	2	3	4	5	6	7	8	9	10
p	.0064	.0328	.0831	.1400	.1763	.1769	.1474	.1050	.0652	.0358	.0176

where k is the number of active fibres that contact a stellate cell and p is the probability of that number given a random pattern of parallel fibre activity. The weighted average (the sum of the products) is ~5. So assuming that a stellate cell receives contact from 1,000 parallel fibres, the most likely number of inputs (the number out of 1,000 that are active) is 5, with a high probability (p = 0.9074) that the number lies between 2 and 8.

This is a good fit with the low number of parallel fibres inputs to a stellate cell based on experimental work. It has been reported that input to an outer level stellate cell is made up of 'two to eight substantial EPSPs [excitatory postsynaptic potentials]' (Jörntell and Ekerot 2003 p.9628). Convergence of our numbers and these findings has the usefulness of cross-corroborating the number of mossy fibre inputs to a granule cell which must be active for it to fire, which was used to estimate the density of parallel fibre activity (Chapter 3). The estimate was obtained using an input threshold of 3, that is, 3 or 4 inputs are necessary for a granule cell to fire. Granule

cells average 4 dendrites, and each receives contact from a different mossy fibre to each of its dendrites, so that for the average granule cell 4 inputs is the maximum.

For comparison, with an input threshold of 2 the expected number of inputs to a stellate cell is ~28 (estimated the same way): outside the reported range. (For the probability distribution recalculated using an estimate of 500 for the number of parallel fibres which contact a stellate cell see Appendix 7.)

5.2.2 What proportion of parallel fibre synapses on a stellate cell are trained?

The parallel fibre-stellate cell synapse is potentiated by a course of training with paired input, while unpaired (parallel fibre only) input trains severe depression of synaptic transmission (in vivo: Jörntell and Ekerot 2002, Ekerot and Jörntell 2003, Jörntell and Ekerot 2003, in vitro: Rancillac and Crépel 2004, Smith and Otis 2005, Soler-Llavina and Sabatini 2006). Depression of an untrained synapse is severe, blocking a postsynaptic effect (even with input at several 100 Hz, Henrik Jörntell, private correspondence dated 31 March 2017).

The total trained synapses following training with n patterns is given by

$$x_n = x_{n-1} + a((m - x_{n-1})/m)$$

Eq.2

such that the percentage left untrained is

$$y_n = 100\big(1 - (x_n/m)\big)$$

Eq.3

where m is the total number of parallel fibres which make synaptic contact on the efferent cell (175,000 on a Purkinje cell and some hundreds on a stellate cell) and a is the (weighted) average number of inputs. The value of m is derived from observation and $x_1 = ar$, where $a = df$ and $r = m/w$. Here, d is the density of parallel fibre activity, f is dendritic field size, and w is the total number of parallel fibres intersecting a field. Field 'size' means area in the sagittal plane and the density of active parallel fibres is the number per unit area. The total number of parallel fibres intersecting a field w and field size f are from observation, and density d is an estimate (derived in Chapter 3).

We might expect that an even density of parallel fibre activity would cause the same proportion of synapses to be trained on stellate cells as on Purkinje cells if we simulate training of stellate cells with the same number of patterns, and this was verified by this method. In fact we get the same result for all values of f and r. That

is, the result is independent of cell morphology, including depth-dependent morphology of stellate cells.

FIGURE 5.1

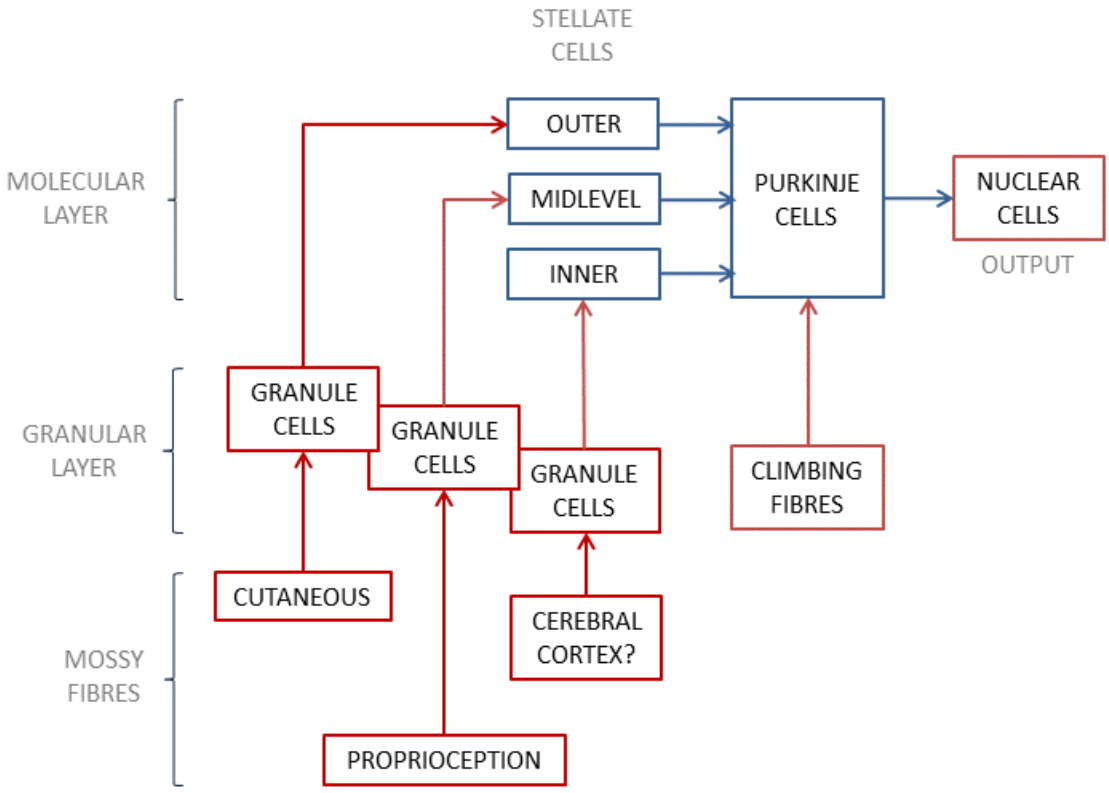


FIGURE 5.1 Vertical topography of mossy fibre input to the granular layer is preserved in the molecular layer such that the source of input to the superficial granular layer drives signals received by stellate cells in the outer molecular layer, and the source of signals received by deeper lying granule cells are received by deeper lying stellate cells.

Note that there is no variable for the amount of contact made by a parallel fibre on its targets. Contact by a parallel fibre on a Purkinje cell is reported to be at one and sometimes two synapses (average 1.24: Barbour 1993, citing then recently published work by Harvey and Napper in Dunedin). There are no equivalent data for contact on a stellate cell. The calculation assumes that contact (if any) made by a parallel fibre in passing is at a single synapse.

5.2.3 Why does that mean (mid-level) stellate cells have poor pattern memory?

A high ratio of trained to untrained synapses means that a pattern of random input is likely to excite firing, because few inputs are needed to excite firing (with tonic inhibition removed) (Häusser and Clark 1997, Carter and Regehr 2002). The number of active parallel fibres that make contact on a stellate cell is low (Table 5.1), so that (unlike parallel fibre input to Purkinje cells) random input does not fall on a predictable split of trained and untrained synapses. The chances that an unknown pattern has less input to trained synapses than a known pattern are not high. The number of inputs to trained synapses is predicted by a similar probability distribution to the distribution for a known pattern (Table 5.2). So on a trial by trial basis the postsynaptic effect is not predictably different although over many trials there is a slight probability bias (towards a lower number).

To be said to discriminate, the response must be different. In traditional theory the response (of Purkinje cells) is bespoke to the individual pattern: the response depends on the pattern. It was previously argued (Chapter 4) that discrimination is not between individual patterns but by class: the response depends on the class. Stellate cells do not discriminate in either manner – they cannot discriminate at all. The following section expands on this.

5.2.4 Probable number of inputs to operational synapses with an unknown pattern

There is no postsynaptic effect of input to untrained synapses (because they are severely depressed). So in considering the effect on the response of the cell it is unnecessary to consider an effect of signals received at untrained synapses (any more than an effect of active fibres that do not make contact, or fibres that make contact but are not active).

An unknown pattern of input to a stellate cell may be to all trained synapses, all untrained, or a mixture (with all untrained the least likely). The probability of n inputs at trained synapses is the probability of n inputs to any synapse reduced by the probability that at least one input is to an untrained synapse (so: 1 – *P*(none)) but increased by the sum of the probabilities that a higher total number of inputs, a, is reduced by a number, b, to untrained synapses, such that $a - b = n$. For example,

the odds that there are two inputs to trained synapses is increased by the product of the probabilities that there are three inputs in total, and any one is to an untrained synapse. It is also increased by the product of the probabilities that there are four inputs, and any two are to untrained synapses, and so on.

So the probability of n inputs at trained synapses with a random pattern is:

$$P(n) - P(n)(1 - 0.85^n) + \sum_{y=1}^{z} P(n+y)\left(\frac{(n+y)!}{y!\,n!} * 0.15^y * (1-0.15)^n\right)$$

Eq.4

where $P(n)$ is the probability of n inputs (to any synapse, derived in Table 5.1) and $n + z$ is the maximum number of inputs with more than insignificant odds (around 10 in Table 5.2 because higher numbers have a very low probability). The results are displayed in Figure 5.2.

Table 5.2 – The probability of contact on a stellate cell from n active parallel fibres in an unknown pattern

n	0	1	2	3	4	5	6	7	8	9	10
p	.0138	.0596	.128	.1828	.1951	.166	.1172	.0703	.0358	.0144	.0035

The probability of n is slightly higher than k in Table 5.1 in the bottom half of the range and slightly lower in the top half. The close similarity of Tables 1 and 2 means that stellate cells do not 'know' from the number of inputs they receive, or how many are to trained synapses, whether a pattern is known or unknown. That is, the response depends much more on chance than on whether a pattern is one that a cell has been trained to. Put another way, mid-level stellate cells are pattern blind. Not only are they unable to discriminate between individual patterns, but the response does not discriminate between the class of learned patterns and random activity. Even zero input to trained synapses does not decisively class a pattern as unknown because it is perfectly possible for a known (to the circuit) pattern to make no contact on an individual stellate cell.

FIGURE 5.2

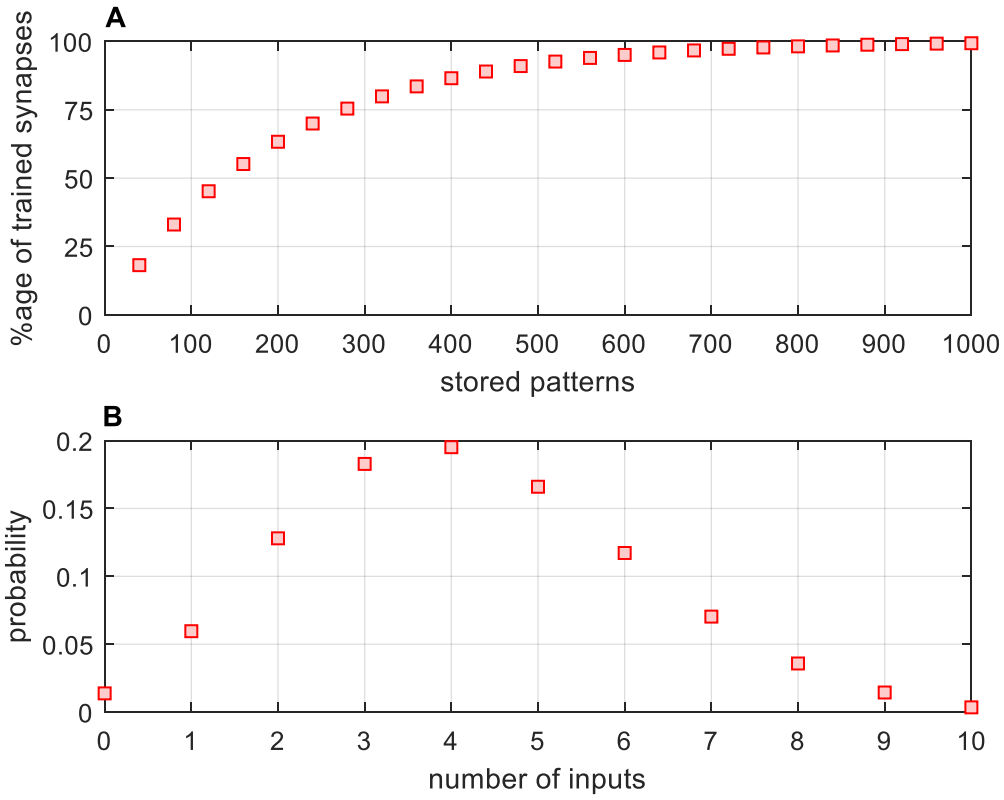


FIGURE 5.2 **A:** Estimate of the storage capacity of a stellate cell assuming that a total of 1,000 parallel fibres make contact. Data are thinned for clarity. The number that make contact per pattern is the weighted average of the Table A probability distribution. If 85% of synapses are trained – predicted by the reported proportion for the parallel fibre-Purkinje cell synapse – a stellate cells stores around 370 patterns. **B:** The probability distribution for the number of active parallel fibres that make contact on a stellate cell at trained (therefore operational) synapses, assuming 85% of synapses are trained, with an unknown pattern of input – Table B in graph form. The number is in the range 0 – 10 (a higher number is very unlikely). Each number of inputs has an associated probability.

5.3 The size and shape of a stellate cell network

This section discusses the basis in evidence for proposal 2.

5.3.1 Background

As noted, but in more detail, there is ‘a specific depth distribution of granule cells depending on the type of input they received’ (in adult cats: Jörntell and Ekerot 2006, p.11795, Quy, Fujita et al. 2011) in the C3 region, which in mammals is involved in forelimb movement. Input to superficial granule cells is triggered by stimulation of receptors at the body surface. The type and mixture of receptors differs between taxa. The middle level of the granular layer receives input from internal sources, from receptors in specialised cells – such as muscles and tendons – that execute movement. The lower third receives some of its input in the form of fast signals from distance receptors (exteroreceptors) via the pontine nuclei (Steinmetz, Logan et al. 1987, Freeman and Rabinak 2004, Freeman and Steinmetz 2011) and may also receive input from the cerebral cortex via the pons. Topography is preserved in the molecular layer, with ‘granule cells in the inner granule cell layer giving rise to PFs [parallel fibres] in the inner molecular layer and granule cells in the outer granule cell layer giving rise to PFs in the outer molecular layer' (Zhang and Linden 2012, p.122).

There are exceptions, but this is the ‘prevalent rule’ (Palay and Chan-Palay 1974, p.66).

In the C3 region mossy fibre input to the superficial granular layer is triggered by the same stimulus as climbing fibre input to the same circuit (Garwicz, Jörntell et al. 1998, Jörntell and Ekerot 2003, Apps and Garwicz 2005, Jörntell and Ekerot 2011). At mid-level mossy fibre and climbing fibre input are triggered independently and mossy fibre input is not from a single site but from mixed sources.

Equivalent data are not available for all regions of the cerebellar cortex. Organisation may be different in other regions, or absent. C3 circuits are assumed here to be generally representative of circuits which control axial and limb movements.

5.3.2 Estimate of convergence and divergence of mid-level stellate cells onto Purkinje cells

There have been several attempts to classify molecular layer interneurons based on their morphology (for example, Eccles, Ito et al. 1967, Palay and Chan-Palay 1974). But it has become the industry standard to divide them into stellate cells, which occupy the outer two thirds of the molecular layer, and basket cells, which occupy

roughly the inner third (in taxa that have them, so: mammals, birds and – debatably – some reptiles (see, for example, Midtgaard 1992)).

Stellate cells fire spontaneously and inhibit Purkinje cells, basket cells and each other (Häusser and Clark 1997, Kondo and Marty 1998, Ruigrok, Hensbroek et al. 2011). Stellate cell networks are planar, occupying the 'thin spaces…between neighbouring Purkinje cell dendritic arborizations' (Eccles, Ito et al. 1967 p.58). They extend indefinitely in both directions, with no anatomical boundaries, but may be thought of as functionally the size of the afferent network of a Purkinje cell. 'Functional' network refers to the afferent network of a Purkinje cell. Networks therefore overlap, a design with the impressive economy that a single stellate cell can participate in many of them.

It can be helpful to visualise what that looks like. Both 'the dendrites and the axons of ... interneurons extend orthogonally' to parallel fibres (Sultan and Bower 1998 p.372, confirming previous reports, including Palay and Chan-Palay 1974). Dendritic and axonal fields are flattened in the same plane. Stellate cells occupy the spaces between Purkinje cells, so that the two are interleaved (Eccles, Ito et al. 1967, Palay and Chan-Palay 1974). A mid-level stellate cell typically makes contact on 2 to 6 Purkinje cells. These are the pair of Purkinje cells whose dendritic plates are immediately adjacent to the stellate cell body, and the next neighbouring pair or two in the direction of the stellate cell main axon. As contact, if any, on a third pair of Purkinje cells is light, and few mid-level stellate cells are too short to reach a second

pair of Purkinje cells, we might generalise that mid-level divergence of stellate cells onto Purkinje cells is 1:4.

The derivation of this estimate is based on anatomy. The Purkinje cell dendritic arbour measures around 9 µm thick (mediolateral) (Hámori 1992 p.30) by 250 - 300 µm (horizontal) (Dean, Porrill et al. 2010) and fills the molecular layer vertically. It is very common for the two primary dendrites of Purkinje cells to give rise to dendritic plates that overlap, doubling the mediolateral span (Eccles, Ito et al. 1967, Hámori 1992). Stellate cell axonal fields measure ~40 µm mediolaterally. A 40 µm mediolateral 'spread is necessary if the axons are to synapse with the dendrites of Purkinje cells on either side' (Palay and Chan-Palay 1974 p.219).

Stellate cell main axon length increases with the depth of the cell body in the molecular layer. At outer level it does not leave the dendritic field, in which it wanders about (Palay and Chan-Palay 1974). The main axon of deeper-lying stellate cells maintains a 'distinct horizontal course' (Sultan and Bower 1998 p.359) parallel to the long axis of the microzone. (The difference between levels is not discrete – see section 5.3.3.) The range varies from 50-450 µm (Palay and Chan-Palay 1974). Therefore, even the maximum range of deeper-lying stellate cells is enough to reach a third pair of Purkinje cells only in few cases, where, lacking collaterals, they make sparse contact. Contact is on smooth, i.e. spineless, dendrites (Eccles, Ito et al. 1967).

The ratio of stellate cells to Purkinje cells is around 16:1 (Llinás and Negrello 2015). This can be visualised as a 4 x 4 matrix in the sagittal plane of equidistant cell bodies in the space between two Purkinje cells, subdivided into 2 x 4 mid-level cells, and 2 x 4 outer level cells. With this set up, divergence of a mid-level stellate cell onto Purkinje cells of 1:4 would mean that a Purkinje cell receives contact from an average of 16 mid-level stellate cells from each mediolateral flank. This is made up of 8 mid-level stellate cells that are immediately adjacent (regardless of the direction their axon takes), and half of the cells whose cell bodies are not immediately adjacent but which have the necessary sagittal range (so on average another 8 per flank, assuming there is a probability of 0.5 that their main axon extends in the 'right' direction). So contact is received by a Purkinje cell from a total of an average of 32 deeper-lying stellate cells (and more – another 16 – at outer level, in which we are less interested).

5.3.3 Functional stellate cell networks are confined to mid level

This subsection is an argument that graded depth-dependent variation of stellate cell morphology (Paula-Barbosa, Tavares et al. 1983) means that outer-level stellate cells are not functionally networked because (1) they make less, and perhaps in some cases no, contact on each other; (2) parallel fibres are more densely packed at outer level; and (3) weakening of transmission of excitatory dendritic signals at outer level is absent (or diminished) at outer level (because there is less inhibition).

Outer level stellate cells do not need to be networked because the need to adapt to mixed input is a mid-level problem (in this contention). In fact, outer-level networking would disrupt normal function, so that it is not merely unnecessary but undesirable.

The anatomy of stellate cells does not seamlessly morph between levels. Outer-level and mid-level stellate cells show depth-dependent expression of different features (rather than only different expression of the same features). This is consistently reported (Eccles, Ito et al. 1967, Palay and Chan-Palay 1974) even when it is used to illustrate continuous phenotypic variation of a (hypothetically) single genotype (whose expression includes basket cells: Paula-Barbosa, Tavares et al. 1983). There is intermingling of features. Features that are characteristic of outer level stellate cells are present but less common and less fully expressed at mid level, and vice versa. But even basket cell features are seen at outer level, in a few cells, and are not uncommon at mid level.

Stellate cell morphology is thus, feasibly, a transition between discrete forms, in addition to graded expression of those forms. Outer-level stellate cells usually have a fairly short, more meandering and haphazardly-branching axon, confined to the same field as their dendrites (Palay and Chan-Palay 1974, Paula-Barbosa, Tavares et al. 1983 p.758). By contrast, the mid-level stellate cell axon usually extends in one horizontal direction or the other, giving rise to ascending and descending collaterals concentrated (and longer) near the soma (Eccles, Ito et al. 1967, Chan-Palay and Palay 1972, Paula-Barbosa, Tavares et al. 1983, Sultan and Bower 1998).

Outer-level stellate cells occupy a territory of around 80 μm^2 (Palay and Chan-Palay 1974 Figures 186 - 188). Axons are confined to the same range as dendrites, with the cell body roughly central. Networks are planar. This restricts mutual contact to those cells whose cell body is both in range and in the same plane. Contact on other stellate cells is exclusively on the soma (Lemkey-Johnston and Larramendi 1968). Stellate cells make only a few synaptic contacts on other stellate cells (Trigo, Sakaba et al. 2012), an observation not limited to outer level.

At a maximum, a cell receives contact from a number of other cells which is given by the number of axonal fields that overlap at the coordinates of the cell body (it may be less because they may not all make contact). Assuming a uniform 80 x 80 µm outer-level field size, for cells to exchange contact their somata would need to be no more than 40 µm apart. In a 2 x 4 matrix, if they were equally spaced, they would be 75 µm apart horizontally and around 50 µm vertically. If horizontal rows were staggered (so that somata were not in a neat vertical row) the distance would be greater. Of course actual cell body distribution and field dimensions vary and do not have this hypothetically neat organisation, but this would suggest nonetheless that contact is sparse at outer level (and more sparse if not all cells that are in range of each other make contact).

By contrast, the larger axonal field of mid-level stellate cells – up to some hundreds of µm horizontally and extended vertically by collaterals – predicts more contact. Assuming the amount of contact exchanged by stellate cells at this level is

proportionate to their range, we might expect mid-level cells to make and receive substantially more contact than outer-level stellate cells because multiple mid-level axonal fields overlap at any point in mid-level space, such as the coordinates occupied by a stellate cell body.

The difference between the ratio of inhibitory (from other stellate cells) to excitatory (from parallel fibres) input at outer and mid level is increased by the higher concentration of parallel fibres at outer level. Parallel fibres are more densely packed at outer level (Sultan 2000). If the same proportion of parallel fibres is active at all levels, a heavier concentration of parallel fibres at outer level would tilt the balance of input to outer level stellate cells towards a higher proportion of excitatory input. In fact, in the conditioned response the proportion may be even higher because it is boosted by nucleocortical feedback (see the Discussion), which terminates preferentially on superficial granule cells (Gao, Proietti-Onori et al. 2016), whose axons divide at outer level (Zhang and Linden 2012).

At mid level the balance is reversed. There is more inhibitory contact and lower parallel fibre density. Tonic inhibitory input dramatically weakens dendritic transmission of excitatory signals in stellate cells (discussed later). So that in addition to a level-dependent ratio of inputs, there is level-dependent weakening of the effect per excitatory input, such that (we will argue) an effect of excitatory input is blocked at mid level (at regulated levels of parallel fibre activity).

So (by this reasoning) the difference between levels is larger than the direct effect only of depth on graded morphology, and variation with depth of the size of the difference is not linear – the depth-dependent increase in contact/area is non-linear and sharper demarcation of functional strata is provided by higher superficial parallel fibre density.

5.4 A model of firing by a functional network of mid-level stellate cells

Turning now to proposal no. 3, the effect of network interactions on the firing rate of individual cells is considered.

5.4.1 Background

Unequal firing rates and intermittent and unpredictable silences are reported in spontaneously firing stellate cells in slices. The spontaneous rate varies from cell to cell, average ~12 Hz (Häusser and Clark 1997, range 1-41 Hz) and firing of a single cell is characterised by irregular and occasionally lengthy silences (Häusser and Clark 1997, Ruigrok, Hensbroek et al. 2011). Several explanations have been advanced for this behaviour. One is that it is the random outcome of the effect of inhibition by stellate cells of each other. That is the view taken here, with the function that it blocks feed-forward inhibition of Purkinje cells driven by the default,

homeostatically-constrained level of excitatory input to stellate cells (Chapter 3) (that is, at low numbers, in the range given in Table 5.1). Accordingly, it is unnecessary to argue that irregular firing of networked stellate cells codes anything in order to explain its function. It also reduces the aggregate inhibitory output of the network (leaving room for an uplift in the conditioned response).

5.4.2 Physiological derivation of model parameters

Irregular firing of a networked stellate cell (in a slice) is transformed into a regular firing pattern by blocking inhibition received from other stellate cells, and transformed back by removing the block (Häusser and Clark 1997). Regular firing is at the intrinsic rate for that cell. This varies between cells but is at a higher mean rate than irregular firing because no spikes are 'missing'. Although (in some cells) intermittent and unpredictable silences of an irregularly firing cell can occasionally be for hundreds of milliseconds (Ruigrok, Hensbroek et al. 2011), the main and more common effect is on a much smaller time scale, the equivalent of the omission of roughly every other spike, and sometimes two, from the intrinsic pattern (Häusser and Clark 1997 fig. 3).

'Omissions' are because discharge of a spike is delayed by inhibitory input. A single action potential can significantly delay the occurrence of the next spike. Inhibition by multiple input spikes is thought to summate to produce longer delays. The duration of the delay depends on the number of spikes. The short integration window relative to

the stellate cell refractory period means the number of spikes depends on the number of afferent cells which spike at the same time or in a short window.

We can simulate the effect of inhibition by stellate cells of each other, and therefore irregular firing by a network, by using the number of networked cells that fire in a short time interval, and which ones they are, to calculate how many fire in the next interval (and which they are), and so on. Spikes are counted in bins of 25 ms because the maximum reported intrinsic firing rate is ~40 Hz, so a cell generates either one or no spikes per bin. It will be assumed all cells have this intrinsic rate. This is a simplification – rates vary – but it allows us to assume that all cells fire in a bin unless they receive enough input for a postponement, calculated on a cell by cell basis. The sum of inputs to a cell in a bin determines whether or not it fires in the next bin. Each round of calculations (once per bin) alters the input that cells receive in the next bin.

A cell can be silent for more than one bin if it receives enough input. This can go on indefinitely in theory – perhaps accounting for occasional extended silences (Ruigrok, Hensbroek et al. 2011) – but a long silence is improbable. Although a single input spike can delay discharge, the delay is relatively modest compared to bin size, so integration of a minimum number of inputs is necessary for a postponement (because the next spike must be delayed till *after* the next bin, so by a minimum of 25 ms). For consecutive silent bins, a cell must receive enough input in each. There is

no carry over. Bin length is effectively (in the simulation) the integration window for inhibitory input (in reality a rolling period).

The number of excitatory inputs to a stellate cell has the Table 5.1 probability distribution. The network is not closed but includes an effect from ‘outside’, that is, from more distant cells that contact the network, and cells that contact them, and so on. There is no weight in the simulation for input signal strength because discharge by a cell in a bin is exclusively in the form of a single action potential (because cells can’t discharge more than once in a bin), and the shape and amplitude of spikes is invariant (Häusser and Clark 1997). There is, though, an adjustment to reflect the random variation of the effect of a signal on different targets, thought to be probably the effects of release probability and the number of sites of synaptic contact (Häusser and Clark 1997).

The simulation assumes the 2x4 matrix of mid-level stellate cells and estimated divergence of a stellate cell onto Purkinje cells of 1:4 discussed in section 5.3.2. A Purkinje cell therefore receives contact from the 16 immediately flanking stellate cells (8 each side) and has a 50/50 chance of receiving contact from cells in a sagittally neighbouring matrix, because the main axon of a stellate cell may extend in either sagittal direction with equal probability. So a Purkinje cell may receive contact from a maximum of 48 stellate cells (24 each flank) and the average is 32. Networks are planar. The cells on each flank are networked with each other but not the other flank. A Purkinje cell also receives contact from stellate cells at outer level. These are not

included in the model because they are not (or are weakly) networked (argued previously).

The direction of the main axon of each stellate cell is assigned at random, by the toss of a coin. Stellate cells do not, in reality, make a lot of contact on each other. So in the simulation not all stellate cells in a network contact each other. To determine how much contact each cell receives from other cells, the probability of contact is weighted to allow for axon direction and the effect of depth on axon length and collateralisation. The amount of contact on each stellate cell, and which other cells it is from, is derived by dividing its neighbours (all the other cells in the same network) randomly into four groups and assigning to each group a different probability of making contact (p = 0.25, 0.333, 0.375, and 0.5, coarsely reflecting a stronger probability of contact with increasing depth). Contact (or not) on each cell from each of the other cells in the network is then decided by a dice roll, so that each cell has an afferent network selected from among the others in a way partly reflecting chance, as in reality. The sum of inhibition received by a cell in a bin is accordingly the number of those cells that fire. If that number exceeds a threshold (chosen by us) that cell does not fire in the next bin.

The threshold is simply the number needed to block firing. The same threshold is used for all cells. This means some cells never receive enough spikes to reach threshold because they don't receive enough contact, so that they fire in every bin. This is not unphysiological. Reported variation of firing of stellate cells include some

that fire strongly. A standard threshold for cells that receive a variable amount of contact that depends on depth and chance may be among the reasons for wide variation of the firing rate of stellate cells (Häusser and Clark 1997, Ruigrok, Hensbroek et al. 2011).

5.4.3 Model stellate cells

How does that translate into output of the simulation? Each cell in a network is a variable assigned the value 1 or 0 in each iteration of the program – an iteration represents a 25 ms bin. There are two ‘networks’, representing the population of stellate cells that contact a Purkinje cell, divided in two because a Purkinje cell receives contact from both sides. A network is represented by a set of variables. The number in a set is determined at the outset and fixed. A hypothetical Purkinje cell therefore receives ‘input’ from two sets.

The value assigned to a variable represents discharge (or not) by that cell (in that bin/iteration) – so, 1 means it spikes and 0 means it is silent. Each variable receives ‘contact’ from a sub-group of the other variables in the same set. Membership of a sub-group is determined at the outset and fixed across all iterations. Membership is determined, for each variable, by effectively rolling a loaded dice that reflects the higher probability of contact from deeper cells. This is done for each of the other cells

in the network, so that in theory contact can be received from any number between one and all of them.

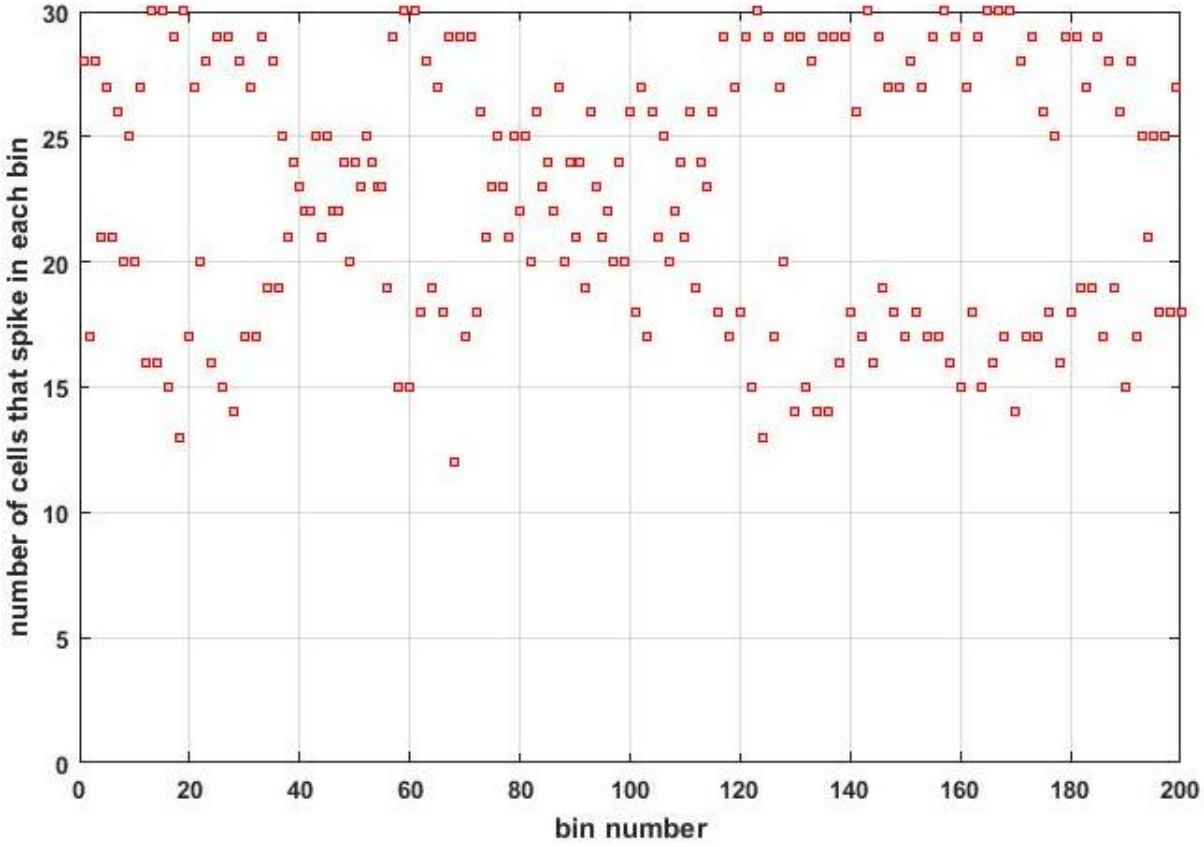
In each iteration, each cell, say cell x, receives total input equal to the number of its afferents that spike. This determines the value assigned to x in the next iteration. Assignments are 1 or 0 because a cell cannot exceed 1 spike per bin. If the count is below threshold the value assigned to x is 1. If the count is equal to or exceeds threshold the value assigned to x is 0. The threshold is the same for all cells – 8 spikes or more means a cell is silent in the next bin, less than 8 means it fires.

The total for both sets in each bin/iteration is displayed as a red square in Figure 5.3. The values assigned to a set of variables are then used to repeat the process in the next iteration. This can continue indefinitely. Figure 5.3 simulates 5 seconds.

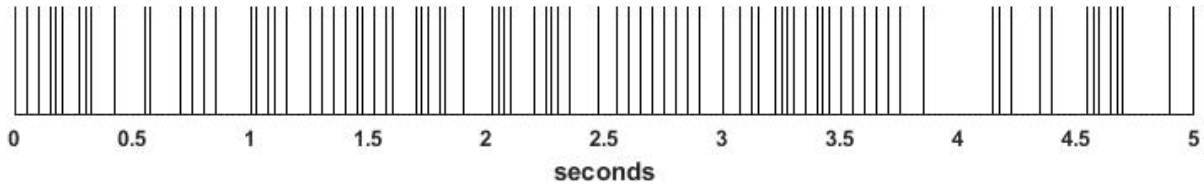
The caption contains our comments on it. One further observation. Inhibitory signals are assigned random strength to represent random variation of the effect of the same signal on different targets, and of the effect of signals from the same source at different times (Häusser and Clark 1997). The strong effect (in our simulation) of this on the firing pattern of the network is seen by removing it (not shown). The network firing signature (and the erratic firing pattern of single cells) is entirely lost even with what are otherwise the 3E settings. The dramatic effect would indicate that physiological variation is functional and necessary, rather than, say, haphazard or merely tolerated.

FIGURE 5.3

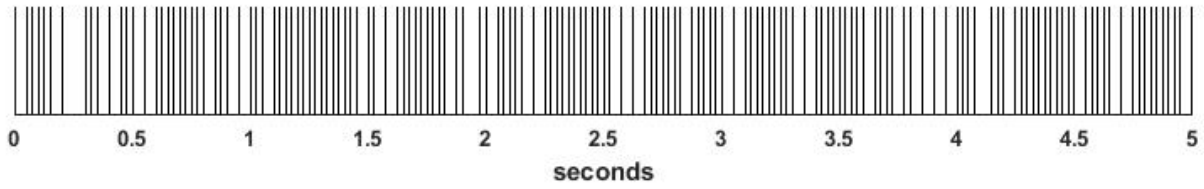
A



B (17 Hz)



C (31.4 Hz)



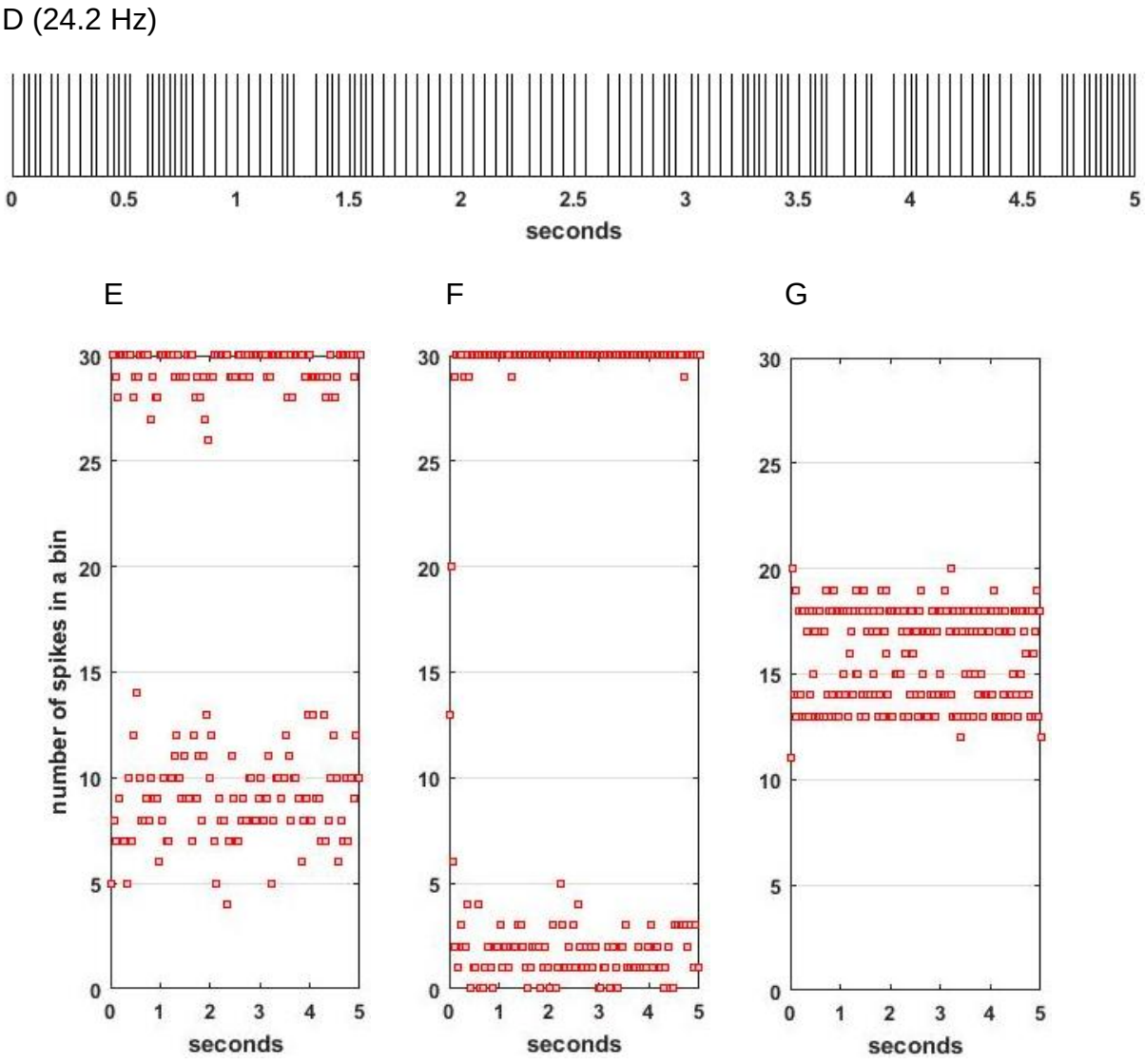


FIGURE 5.3 Simulation of a mid-level stellate cell network at homeostatically self-regulated density of parallel fibre activity. Figure 5.3A shows the number of mid-level stellate cells (out of 30 which contact a simulated Purkinje cell) that are active in each of 200 25 ms bins (so for 5 seconds). The firing pattern of individual cells generated by this model is a good match with the reported firing pattern. Figures 5.3B-D show spiking of three randomly-selected cells (compare Häusser and Clark 1997 figs 1 and 3). Their un-networked, intrinsic firing rate is 40 Hz. Figures 5.3E-G are to illustrate the

effect of reducing the number of inputs (contact by other cells that fire) needed to block firing in the next bin. In Figure 5.3E this number is reduced from 8 to 6. The result is that the total number of cells that fire in each bin oscillates, bouncing back and forth in successive bins. The reason that it does not use the whole range (down to zero) is that some cells do not receive enough contact to ever meet threshold, so that they fire in every bin. This is not unphysiological. Reported variation of firing of stellate cells include some that fire strongly. In Figure 5.3F the threshold is dropped to 4, the minimum contact received by any cell in our simulation. The result is (near to) a rapidly alternating all on/all off firing pattern. The irregular firing pattern of stellate cells in vitro (Midtgaard 1992, Häusser and Clark 1997, Carter and Regehr 2002) and (even more so) in vivo (Jörntell and Ekerot 2002, Barmack and Yakhnitsa 2008) would indicate that such a low threshold is unphysiological (because synchronised oscillation is unphysiological). In fact the same settings (as 5.3F) generate two patterns of firing. The other is shown in Figure 5.3G. In 5.3F the two halves of the network are in phase. In 5.3G they are out of phase. The settings for the simulation are exactly the same. Each time the simulation is run it can be either. (Both in and out-of-phase firing is seen in Figure 5.3A, although it is less pronounced.)

5.4.4 Nil effect of excitatory input at mid-level in default conditions

The author had expected that the reported irregular firing pattern of stellate cells would reflect a mixture of excitatory and inhibitory input. Instead, however, even very

low strength excitatory input caused cells to fire in almost every bin. This would indicate that stellate cells that fire irregularly in vivo – as they do in anaesthetised cats and mice (Jörntell and Ekerot 2002, Barmack and Yakhnitsa 2008) – do not receive excitatory input, or, rather, do not receive an *effect* of excitatory input. An effect (in the simulation) on firing even of very weak excitatory input would indicate that an effect of excitatory input is absent in natural conditions. Since input is in fact present (because parallel fibre activity is maintained at a regulated level (argued in Chapter 3)), its effect must be blocked.

A possible explanation is that in the default condition firing is exclusively under inhibitory control (as in Figure 5.3). This is an unsafe conclusion without corroboration and also would seem, on the face of it, to contradict reports that stellate cells are sensitive to excitatory input. On a close reading of the literature, however, it turns out that a default inhibitory block of an effect of excitatory input to stellate cells at low numbers (predicted by the Table 5.1 probability distribution) has robust support.

A list of reasons follows. (See also Appendix 8, ‘Examples from the literature’.)

1. The narrow calibre of stellate cell dendrites (Sultan and Bower 1998) means passive transmission of subthreshold signals is weakened by cable properties (Häusser and Clark 1997, Abrahamsson, Cathala et al. 2012). Dendrites are

thin, and even thinner distally – average diameter 0.55 µm (Sultan and Bower 1998). Because they have a large surface area to volume ratio there is more rapid leakage of charge than with a larger bore, other things being equal. The rate is increased by large dendritic depolarisations (Abrahamsson, Cathala et al. 2012), creating a strong electrical gradient across the cell membrane.

2. Tonic inhibition of mid-level stellate cells by other stellate cells, received at and spreading out from the soma, strongly increases the cable-like behaviour of subthreshold dendritic transmission (Häusser and Clark 1997). So there is strong attenuation of passive current flow, reducing charge transfer from synapse to soma.

3. The stellate cell dendritic membrane time constant is reduced by tonic inhibition (Häusser and Clark 1997). With an artificially applied 0.5 ms current pulse the voltage across the postsynaptic membrane returns to normal in a few ms. As a result there is a narrow window of integration of excitatory inputs.

4. For integration, inputs must be near in space as well as time. They must be to the same dendrite, and near each other. The normally low number of concurrent inputs to a stellate cell (Table 5.1) means that there is a relatively modest probability that two of them are to the same dendrite, and more modest that they are to the same compartment. The strong attenuation of the passive spread of subthreshold dendritic signals makes it less probable that inputs to different parts of the same dendrite summate. Summation (at Table

5.1 levels of input) is therefore likely to be absent or ‘severely affected’ (i.e., much weakened: Häusser and Clark 1997, p.675).

5. There is sublinear integration of excitatory inputs to the same dendritic compartment. Excitatory input sums sublinearly because it drives large dendritic depolarizations that reduce synaptic driving force of additional inputs (Abrahamsson, Cathala et al. 2012, working with mature stellate cells). Sublinearity is seen with just 2 quanta, assisted by large quantal conductance (Carter and Regehr 2002). The amount of the reduction is increased if input signals are strong, so that larger synaptic conductances increase sublinearity (Abrahamsson, Cathala et al. 2012). Granule cells fire in strong bursts (Ruigrok, Hensbroek et al. 2011), so summation of granule cells bursts should be – proportionately – strongly sublinear.

6. Elevation of postsynaptic intracellular calcium triggered by synaptic activation sums linearly or supralinearly (Tran-Van-Minh, Abrahamsson et al. 2016). Supralinear calcium summation causes a branch-specific, endocannabinoid-mediated, short-term suppression of neurotransmitter release (Tran-Van-Minh, Abrahamsson et al. 2016). (Calcium ion current density is too low to contribute to depolarisation, so it doesn’t interfere with sublinear synaptic voltage integration (Abrahamsson, Cathala et al. 2012).)

7. Although a stellate cell receives more numerous contact from parallel fibres than from stellate cells, the number of each type that are concurrently active is

normally in a similarly narrow and low range. This, and the distribution of inhibitory and excitatory contact – inhibitory somatic and excitatory predominantly dendritic (Lemkey-Johnston and Larramendi 1968) – suggest a tonic inhibitory override (consistent with (1) – (6).)

5.4.5 Hypothesis

The effect of excitatory input to a stellate cell is ultimately at the soma – the final site of synaptic integration (Stuart, Spruston et al. 1997). The foregoing deliberations led to the hypothesis that an effect on firing at low numbers of excitatory dendritic inputs (where the majority of excitatory contact is made: Lemkey-Johnston and Larramendi 1968) is vetoed by tonic somatic inhibition, and a default block is the function of the mutual inhibition and spontaneous firing of networked stellate cells, to prevent what would otherwise be indiscriminate (because constant) feed-forward inhibition. Excitatory input in larger numbers is needed to overcome the veto. The normal amount of regulated parallel fibre traffic (Chapter 3) is not enough. The extra input must be timely, in a narrow integration window, and received only in a trained circuit. A candidate to (indirectly) provide it is nucleocortical feedback, argued next.

5.5 Discussion

This section argues that there is a timely and targeted conditioned suspension of the default veto, and how this creates an adaptively-timed window for control of stellate cell firing rates in the conditioned response.

5.5.1 Nucleocortical feedback

Internal feedback from the cerebellar nuclei to the cerebellar cortex in classical eyeblink conditioning increases the amplitude of learned blink responses (Gao, Proietti-Onori et al. 2016). The feedback pathway is formed of collaterals of excitatory nuclear projections neurons, which turn back towards the cerebellar cortex where they terminate as mossy fibres on granule cells (Houck and Person 2015). By comparison with pre-cerebellar mossy fibre rosettes, nucleocortical rosettes have a larger diameter and nucleocortical fibre endings have more and longer filipodia-like structures. Training with a conditioning protocol causes a 70% increase in density of filipodia boutons, and the termination pattern is condensed into a narrow strip (Gao, Proietti-Onori et al. 2016).

5.5.2 Feedback primarily causes an uplift in the number and not strength of granule cell signals

The output of a microzone is carried by a modest group of nuclear projection neurons. Collaterals of these cells make up the feedback pathway. Feedback is duly precision timed and terminates back at the microzone whose output controls firing of the nuclear group. The modest number of cells in a nuclear group would suggest, on the face of it, a weak effect of feedback. However, several mechanisms combine to increase the number of granule cells contacted, and to increase the probability of an effect on granule cells that receive contact.

Mossy fibres end in several sagittally-aligned clusters (Sultan 2001, Sultan and Heck 2003) of (average) 7-8 terminals (Wu, Sugihara et al. 1999, Shinoda and Sugihara 2013). A mossy fibre makes contact via each terminal on a number of granule cells in the range of estimates 10-100 (Jakab and Hámori 1988, Ritzau-Jost, Delvendahl et al. 2014). As a result, a single mossy fibre may contact 1,000s of granule cells in a region with a sagittally-elongated footprint.

The feedback pathway terminates preferentially in the outer granular layer (Gao, Proietti-Onori et al. 2016). Cutaneous input is also received in the outer layer (Jörntell and Ekerot 2006). In natural conditions, cutaneous signals are always paired with climbing fibre input because they are evoked by the same stimulation (Garwicz,

Jörntell et al. 1998, Jörntell and Ekerot 2003). The anatomy of contact by mossy fibres on granule cells means the number of outputs per input to a location increases from low to high numbers of inputs (Chapter 3), so that the combined input to the superficial granular layer drives more than the sum of the output they would drive individually. Thus, confinement of feedback to a single level – rather than termination in a more dispersed pattern across all levels – excites more granule cells.

The author disagrees with the interpretation (Gao, Proietti-Onori et al. 2016, ten Brinke 2017) that the nucleocortical pathway is part of a positive feedback loop. In our view amplification is of the number and not strength of granule cell signals. Granule cells need a minimum of three excitatory inputs to fire (Chapter 3 and discussed earlier). It is more likely that a feedback signal is received by a cell where it makes up the number of inputs from 2 to 3 than from 3 to 4. It is also twice as likely that a cell that had previously received 2 receives another one. So feedback signals are received by more cells that did not fire before, and fire as a result, than cells that did.

5.5.3 Unconfirmed granule cell ascending axon contact on stellate cells

Termination of nucleocortical feedback mainly in the superficial granular layer means that it excites mainly granule cells whose parallel fibres are in the outer molecular layer (Palay and Chan-Palay 1974, Zhang and Linden 2012). These accordingly

mainly make parallel fibre contact on stellate cells at that level. Ascending axon contact by granule cells on stellate cells (is unconfirmed but) would in theory supply extra input to mid-level stellate cells only in the circuit receiving feedback, and not in other circuits. Moreover, ascending axons lie in the plane of the stellate cell dendritic field, which is severely flattened sagittally (Palay and Chan-Palay 1974), so that the probability of contact by a local granule cell is – presumably – higher than for non-local granule cells which can only make contact through parallel fibres, which pass through the stellate cell dendritic field at right angles, and only contact 1 in 60.

Furthermore parallel fibre signals at superficial level selectively drive feed-forward inhibition by local stellate cells and not stellate cells in other circuits, because they are received at very heavily depressed – to the extent of being inert – synapses in other circuits (Ekerot and Jörntell 2001, Jörntell and Ekerot 2002, Jörntell and Ekerot 2003, Jörntell and Ekerot 2011). Feedback is thus here also confined to a local effect.

5.5.4 An adaptively-timed window for control of stellate cells

As nucleocortical feedback is a copy of the output of the circuit (following training), this creates an adaptively-timed window for control of (networked, deeper-lying) stellate cell firing rates in the conditioned response, when they become transiently sensitive to excitatory input rates. Because stellate cells at deeper level have bad

pattern memory, this does not need to be input they are trained to. It is not the pattern used for pattern matching (although they may overlap), and does not need to be in a learned pattern at all. The control window is shorter and later than the learning window for synaptic plasticity. Timing of the window is tightly coupled to the climbing fibre signal that trains it.

The window is coordinated along the length of a microzone. Coordination is provided by adaptive timing of the intrinsic plastic response of Purkinje cells (Johansson, Jirenhed et al. 2014, Johansson, Carlsson et al. 2015) which therefore orchestrates a coordinated fall in the simple spike rate (because timing at all locations is coupled to the same climbing fibre volley) despite the range of times of signals in a learned pattern. Purkinje cells learn this response even with interneurons pharmaceutically blocked (Johansson, Jirenhed et al. 2014). This allows intrinsic plasticity of Purkinje cells to unblock feed-forward inhibition without needing feed-forward inhibition to do it.

Thus the default embargo on an effect of excitatory input to networked stellate cells is lifted in the conditioned response by nucleocortical feedback (indirectly, by exciting more local granule cells). Stellate cells become as a result transiently sensitive to excitatory input, in a window coordinated across the population of a microzone.

5.5.5 Why don’t nucleocortical signals interfere with control of granule cell rates or regulated density of parallel fibre activity?

The firing rate of the extra granule cells in the conditioned uplift should receive an influence of nucleocortical feedback rates. Also, how is the uplift reconciled with homeostatic regulation of parallel fibre traffic? This section is an argument that there is no interference with either.

Mossy fibre terminals are each ensheathed in a semi-permeable membrane, an arrangement termed a glomerulus. The membrane restricts diffusion of synaptically-released neurotransmitters that escapes into the surrounding space. The intraglomerular balance of the concentrations of GABA and glutamate is adjustable and functional. Mossy fibre and Golgi cell terminals compete to suppress each other with the result that mossy fibre-granule cell transmission is confined to competitive signals (Mitchell and Silver 2000a, Mitchell and Silver 2000b) (also Chapter 3).

Collaterals that carry feedback terminate as mossy fibres. The glomerular competition means that nucleocortical feedback signals compete on the same terms as other mossy fibre signals. The inhibitory Golgi threshold is impartial so that, for example, weak feedback signals do not dilute granule cell rates. There is no impact of feedback below competitive rates. As the main effect is on the number of granule

cells that fire (because more meet the input threshold) granule cell rates are not amplified by feedback signals.

The conditioned uplift in the local proportion of active granule cells does not compromise homeostatic regulation of parallel fibre activity (proposed in [ref PI]) because feedback terminates with the same footprint – a narrow sagittal strip – as a microzone (Gao, Proietti-Onori et al. 2016). The regulation of parallel fibres is along the mediolateral axis, defined by the range of a mutual influence of granular layer recoding at locations on beam. It regulates activity in the strip under a beam where the number of active granule cells is ‘counted’ for that purpose. The footprint of feedback is at right angles so the boosted number at the intersection only makes a very modest difference to the count. The width of the intersection measures around 0.8-1.6% of the mediolateral dimension of the region that supplies parallel fibres that pass overhead, assuming a microzone span of 50-100 μm (Oscarsson 1979, Dean, Porrill et al. 2010) and a parallel fibre range of 6 mm (Brand, Dahl et al. 1976, Mugnaini 1983). Indeed it is a hallmark feature of regulation of parallel fibre activity that it is insensitive (to the point of indifference) to variation of the number and spatial distribution of mossy fibre inputs to and within a regulated strip (Chapter 3). Therefore, neither the addition caused by feedback to the total number of active granule cells nor feedback-augmented clustering of input at the intersection interferes with regulation.

CHAPTER 6

SUMMARY INCLUDING AN OUTLINE OF OUTPUT CODING

6.1 Introduction

The cerebellum is widely thought to implement a supervised learning algorithm which uses iterative adjustment of parallel fibre synaptic weights either to generate binary output (Albus 1971, Brunel, Hakim et al. 2004) or to reduce deviation from a desired outcome (Fujita 1982, Dean, Porrill et al. 2010). Supervision is typically provided by error signals provided by climbing fibres. Circuit function on these principles assumes that output is driven by remembered patterns of active parallel fibres. The unit of learning is a Purkinje cell, storage and expression of pattern memory are independent of other Purkinje cells, and Purkinje cell rates are treated as output. Output is learned, that is, training displaces the naïve response to input rates.

We are in conflict with this view on all points. Instead, we suggest, the cerebellum is designed to separate the functions of pattern detection (which cues output) and control of firing of output cells (which codes output), so that pattern memory has a gating but otherwise permissive role and output rates are controlled separately. Separation of the functions of pattern learning/detection and control of output rates requires that the data they each use are represented as different and independent variables, so that there is not mutual interference of the execution of either function

with the other. It is partly for this reason that mossy fibre input to the cerebellum is recoded in the granular layer: to cause the number of variables coded in the binary pattern and permutations of rates received as input into a dramatically reduced number expressed by internal signals activity, in parallel fibre input to modular circuits. This is also to prevent interference (with correct function) by redundant external variables, variables for which the cerebellum has no use.

This is only half of the picture. The cerebellum also has the function of turning linear rate coded data (spikes/t, so that the speed of data transfer is limited by *t*) into a form that can be transmitted much faster, so that output rates can be under finely graduated control in a short nuclear integration window. It does this by converting input rate variables into parallel fibre signals with a fixed bandwidth in an adjustable range that follows the mean of mossy fibre input rates, and by coordinating Purkinje cell firing on a very short timescale across the population of a microzone. This is a reason for the functional organisation of Purkinje cells into microzones, and of nuclear projection neurons in groups. Coordination of Purkinje cells is a functional imperative and accounts for duplication of mossy fibre signals in sagittal strips (mossy fibres have a sagittally-aligned terminal branching pattern), and multiple random samplings of input rates by granule cells.

Memory and control of output are not at the level of single Purkinje cells but microzones. There is no requirement (or norm) that climbing fibre instruction signals are driven by errors (also unnecessary to explain the evidence), and – importantly –

output is not learned. Synaptic weights are functionally binary, and input rates control output rates ad hoc – indeed graded weights would interfere with function.

The proposed model is a hybrid, a detailed physiological model with support from computational modelling. Explained in this way, there is no need for an algorithm to account for the evidence. The ideas presented here are a summary and not a full quantitative description of the proposed mechanism of cerebellar function (which will be published separately). The idea that function of the cerebellum will have a computational explanation has displaced other approaches. In that approach, the cerebellum is the physiological implementation of a machine learning answer to the problem of turning input into whatever output the model says it has – an abstract solution to a hypothetical problem. We offer instead a synthesis of a physiological and mathematical model – a detailed mosaic of the evidence tested quantitatively by simulation. There is no overarching computational formula.

FIGURE 6.1

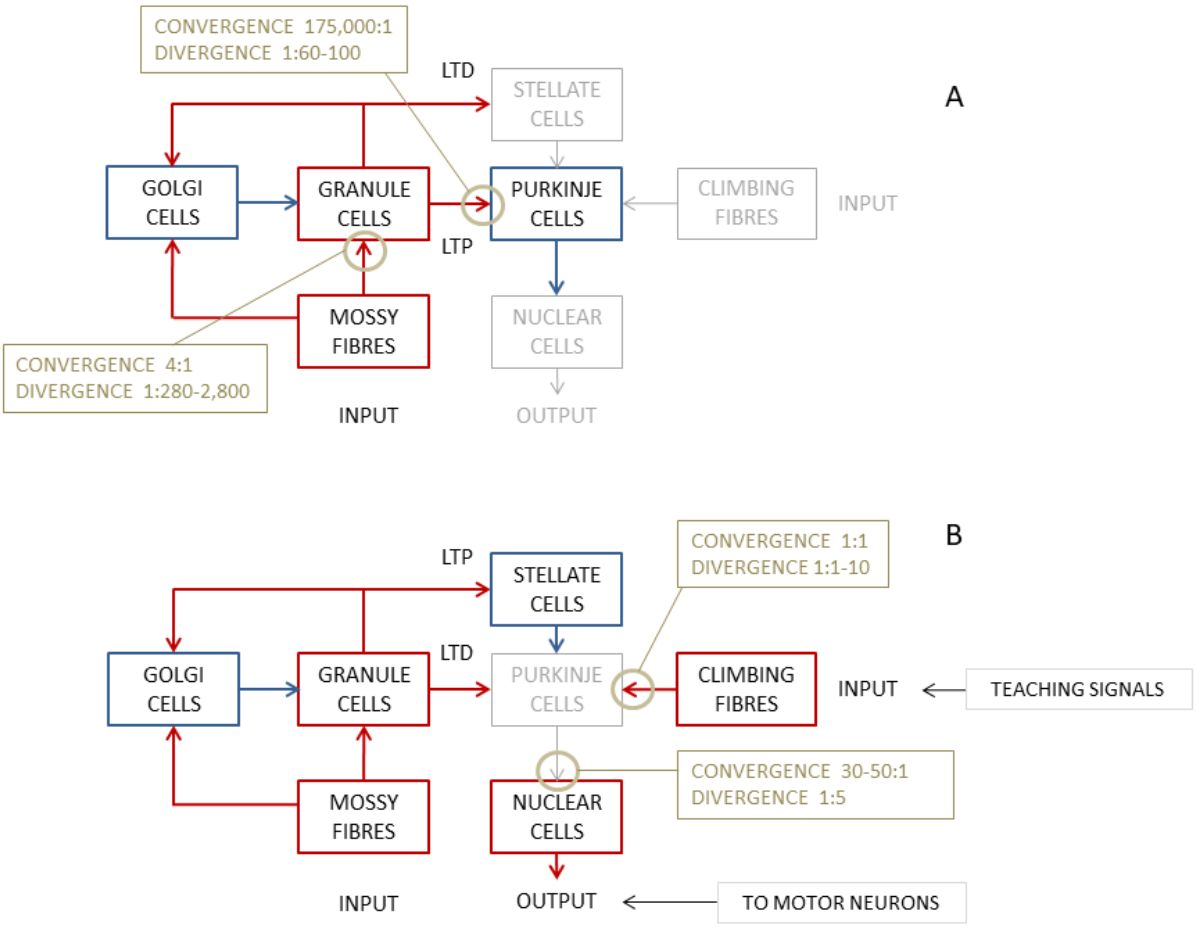


FIGURE 6.1

Schematic of cerebellar circuitry. A: Prior to training, strong Purkinje cell firing inhibits nuclear cells, the output cells of the circuit. Red boxes and arrows: active glutamate neurons; blue: active GABA neurons; grey: silent neurons. **B:** Following training with a conditioning protocol or direct stimulation that mimics a conditioning protocol, firing of Purkinje cells is weakened or suspended in the conditioned response partly as a result of long-term depression (LTD) of the parallel fibre-Purkinje cell synapse and long-term potentiation (LTP) of the parallel fibre-stellate cell synapse, causing a phasic reduction of inhibition by Purkinje cells of nuclear cells.

6.2 Recoding

6.2.1 Background

Most glutamatergic input to the cerebellum is carried by mossy fibres which terminate on granule cells in the inner layer of the cerebellar cortex, the granular layer. Granule cell axons rise into the outer layer of the cortex, the molecular layer, where they divide to form parallel fibres, so called because they run parallel to the surface of the cerebellum and to each other, which make contact in passing on Purkinje cells and inhibitory interneurons including Golgi cells. Golgi cells in turn inhibit granule cells. Purkinje cells are organised functionally into long thin groups of several hundred cells – microzones (Oscarsson 1979, Ozden, Sullivan et al. 2009, Bengtsson, Ekerot et al. 2011, Ramirez and Stell 2016). A microzone occupies a thin sagittal slice of the cerebellar cortex perpendicular to the surface.

The number of mossy fibre signals that drive input to a microzone is large. The area of the cerebellar cortex that provides parallel fibre input to a microzone is measures around 6 x 15 mm, the range of parallel fibres (Brand, Dahl et al. 1976, Mugnaini 1983)[10] and the length of a microzone (an estimate for the C3 region: Dean, Porrill et

[10] Reported for cats, chickens and monkeys; 5 mm is reported for rats Harvey, R. J. and R. M. Napper (1988). "Quantitative study of granule and Purkinje cells in the cerebellar cortex of the rat." J Comp Neurol **274**(2): 151-157, Harvey, R. J. and R. M. Napper (1991). "Quantitative studies on the mammalian cerebellum." Prog Neurobiol **36**(6): 437-463.

al. 2010) respectively. Scaled up, that's around the size of four-and-a-half tennis courts side by side supplying input to a net (the microzone) 50 m long and 0.5 m thick. Granule cell bodies would be ¼ the diameter of a tennis ball and 12-14 million of them would be packed under the court to an average (but variable) depth of 70 cm.

It has long been hypothesised that recoding of mossy fibre signals into granule cell signals sparsens and decorrelates the binary pattern of parallel fibre activity (which parallel fibres are active and which silent) (Marr 1969, Albus 1971), in order to increase storage capacity. The mechanism includes Golgi cells in the role of adjusting the granule cell 'input threshold' – the number of mossy fibre inputs needed to make a granule cell fire.[11] A higher threshold reduces the number that receive enough input to fire, and vice versa. By adjusting the threshold Golgi cells dampen the effect of variation of the number of mossy fibre inputs to the system. Golgi cells thus (in this contention) provide homeostatic control of parallel fibre activity, which is maintained at a low level (Marr 1969, Albus 1971, Billings, Piasini et al. 2014, Cayco-Gajic, Clopath et al. 2017, Cayco-Gajic and Silver 2019). So in this model regulation of parallel fibre activity is by Golgi cell adjustments of the granule cell input threshold, output is sparse, and the function is to increase storage capacity (by recoding mossy fibre signals into a larger state space).

[11] 'Input' (here and generally) means contact by an active cell as opposed to mere innervation.

Absent from the physiological support for this arrangement are (1) how the role of Golgi cells is implemented on a quantified physiological level that can be simulated; (2) how recoding affects granule cell firing rates; (3) whether and how control of parallel fibre activity (numbers active) and the rates they fire at are kept separate so that they vary independently; (4) how external variables (variables contained in mossy fibre input to four-and-a-half nominal tennis courts) are selected for an effect (on either active numbers or rates) or prevented from having one; and (5) what other functions parallel fibre regulation has.

We claim to answer these questions. The discussion is in two parts. In addition to recoding variables contained in the binary pattern of input to the granular layer, the mossy fibre-granule cell relay also has the function of recoding variables contained in the permutation of rates they each fire at. These are considered in turn.

6.2.2 Recoding variables contained in the binary pattern of input

It is part of the present proposal that the function of Golgi cells is not to adjust the input threshold of granule cells in response to input variables (and in particular in response to a changing volume of mossy fibre input) but to make it *independent* of input variables – not only the number of mossy fibre inputs to the system but mossy fibre firing rates – so that it is constant in all conditions. Part of the argument is a

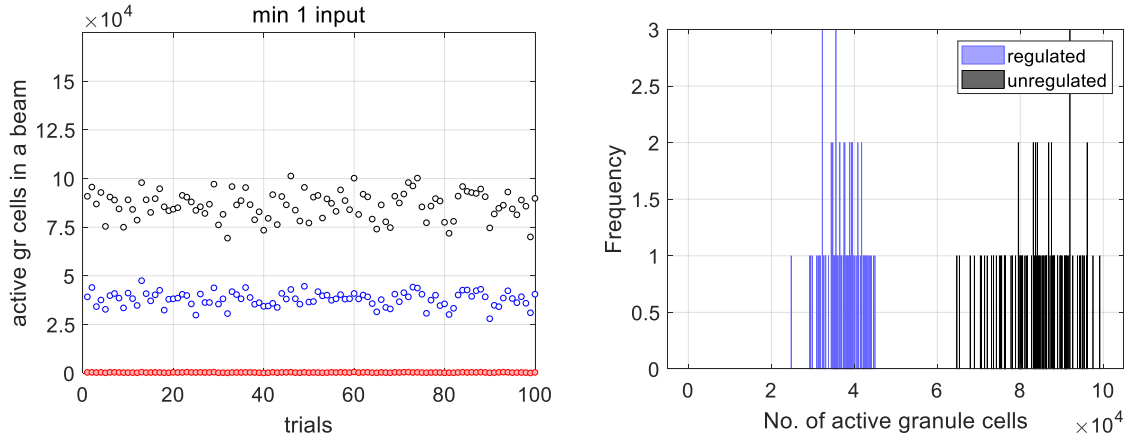
simulation of regulation of parallel fibre activity that shows it is confined to a fixed and low level with a fixed threshold.

Assuming a fixed input threshold, then, the probability that a granule cell will meet threshold can be derived with a binomial from the proportion of active mossy fibres out of the total that supply that location. This gives the expected number that fire because the law of large numbers holds that at large numbers of independent trials, the ratio of the outcomes will converge towards the proportions predicted by their probability.

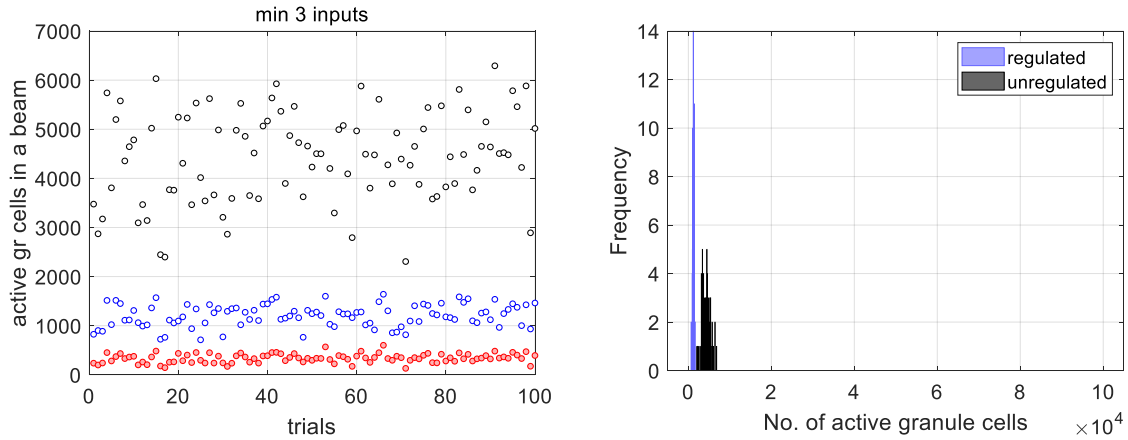
Granule cells have 3-5 dendrites (average 4) each of which receives contact from a single mossy fibre. Assuming contact is at random (i.e. mossy fibres do not select which granule cells they contact) it is very likely that a granule cell receives input to each of its 4 dendrites from 4 different mossy fibres. Mossy fibres branch terminally. Each branch ends in a cluster of terminals (average 7-8 terminals per cluster) (Wu, Sugihara et al. 1999, Sultan and Heck 2003, Shinoda and Sugihara 2013). An estimated 100 mossy fibres terminate in a region the size of cluster field so that a cluster field contains an average of around 700-800 terminals (Sultan and Heck 2003). Each terminal receives a single dendrite from each of what may be around 50 granule cells (Jakab and Hámori 1988, Gao, Proietti-Onori et al. 2016) though estimates vary (Ritzau-Jost, Delvendahl et al. 2014).

FIGURE 6.2

A



B



C

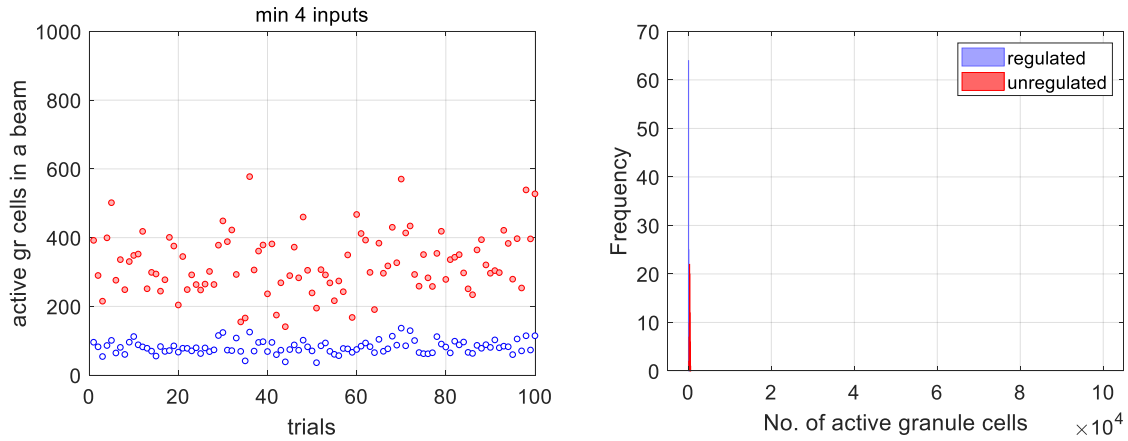


FIGURE 6.2

Homeostatic regulation of parallel fibre traffic by Golgi cell regulation of granule cell firing. The model estimates the total number of excited granule cells in a mediolaterally-aligned row of 20 mossy fibre terminal cluster fields (together measuring 3,000 x 200 μm), trial by trial, in each of 100 trials. Each field receives input from a random number of mossy fibres in the range 3-30 (out of 100) and contains 8,750 granule cells. The minimum number of mossy fibres needed to make a granule cell fire is 1 in panel A, 3 in panel B (the probable physiological input threshold) and 4 in panel C. Black dots are the estimated number of granule cells meeting the input threshold, disregarding Golgi cell regulation; the number with Golgi cell regulation are in blue, and the number which receive 4 mossy fibre inputs is red, identical in all graphs, for comparison. Note y axes are scaled to the data. The histograms are frequency distributions that show the same black and blue data, and in C the blue and red data. With an input threshold of 3 the average regulated number of active granule cells in a beam is approximately 1,200 out of perhaps 250,000. Of those only a few tens receive 4 inputs – the blue data in panel C. The very large majority all receive 3. This forms part of the control of variables that (would otherwise) affect translation of mossy fibre to granule cell rates.

Golgi cells inhibit the mossy fibre-granule cell relay in an arrangement termed a glomerulus. The glomerulus is ensheathed by a semi-permeable membrane which restricts neurotransmitter diffusion. Inhibition has a strong tonic component (Duguid, Branco et al. 2012). Activation of presynaptic $GABA_B$ receptors on mossy fibres

inhibits glutamate release (Mitchell and Silver 2000a), and activation of presynaptic mGluR2 receptors on Golgi cells inhibits GABA release (Mitchell and Silver 2000b). The result is competitive mutual suppression whose outcome depends on phasic adjustments of the neurotransmitter balance (Kanichay and Silver 2008, Cesana, Pietrajtis et al. 2013).

So, meeting the input threshold is not alone sufficient to make a granule cell fire. Inputs must each be strong enough to be competitive – or, put another way, to prevail over the inhibitory veto – with the combined effect, integrated at the soma, of depolarising the target to firing threshold.

Assuming inhibition vetoes an effect of excitatory input to a glomerulus with a probability $P(v)$, and x out of y mossy fibres are active, the probability that a particular dendrite receives excitatory input and no veto is $(1 - P(v))(x/y)$. Given n = 4 dendrites per granule cell and an input threshold m, the expected number of granule cells in a cluster field (say f_1), out of a total t, that fire is

$$E(f_1) = t\left[\sum_{k=m}^{n} \frac{n!}{k!\,(n-k)!} * \left((1 - P(v)) * \frac{x}{y}\right)^{k} * \left[1 - \left((1 - P(v)) * \frac{x}{y}\right)\right]^{n-k}\right] \quad (1)$$

where y is a physiological constant (100) and x is in a range 3-30 generated randomly for each field in each trial. t (also a constant) is 8,750 (calculated from convergence of 4:1 and divergence of 1:50).

Golgi cells receive excitatory input from local mossy fibres and from parallel fibres. There is as a result a mutual effect mediated by Golgi cells on fields aligned mediolaterally, joined by parallel fibres, and in a row of fields there is an effect of all fields on all others. $P(v)$ is derived from the density of parallel fibre activity. As a Golgi cell receives only a small sample of passing parallel fibre activity, the number of active parallel fibres that make contact is determined by chance, with a distributed probability. Even Golgi cells on the path of the same parallel fibre traffic are unlikely to receive the same number of inputs and very unlikely to receive contact from the exact subset (or in fact any) of the same active cells. The probable number that make contact[12] is accordingly a measure of, and quantifies, the effect of raising or lowering density of parallel fibre activity.

The Figure 6.2 data are derived by summing activity in a row and using it to calculate the influence of fields on each other, then rerunning the calculation for each field to get a new total, and repeating this procedure to home in on a stable value. Other parameters included in the calculation – different for each field and with an effect recalculated in each iteration – include incursion of Golgi cells from neighbouring

[12] More accurately, the probabilities associated with each possible number.

fields, direct input to Golgi cells from mossy fibres, and the higher probability that local granule cells make contact on a Golgi cell than distant granule cells (that make only parallel fibre contact).

The result – the blue data in Figure 6.2 – is that parallel fibre activity is confined to a narrow and stable range regardless of the (fixed) granule cell input threshold. There is no need to adjust the input threshold to keep output stable, or to keep it low. So: Golgi regulation works in the reverse of the expected way (in this contention). It does not have the function of adjusting the input threshold (to maintain self-rectifying parallel fibre levels), but of keeping it the same.[13]

Regulated parallel fibre activity is low regardless of the shifting underlying number and binary pattern of mossy fibre inputs, as previously predicted (Marr 1969, Albus 1971, Billings, Piasini et al. 2014, Cayco-Gajic, Clopath et al. 2017, Cayco-Gajic and Silver 2019), but the mechanism of regulation is different. This is not merely a detail of implementation. It makes an important difference to the function it contributes to recoding, and in turn to what we can infer about the function that recoding contributes to circuit operation. It was previously thought the function of recoding (in models that need the number of active parallel fibres to be a low fraction) was to increase storage capacity by sparsening and separating recoded patterns of activity. We agree it increases capacity, but not on the way that regulation works or on how pattern

[13] To achieve stable density of parallel fibre activity, it is not, at least proportionately, reflected in the active fraction of the local population, but derived at the scale of the whole strip that provides parallel fibres that pass overhead.

memory is stored and expressed. Nor is it the only function recoding has, or correct to think it has more important status than other functions.

6.2.3 Recoding variables contained in the permutation of input firing rates

The probable physiological granule cell input threshold is 3 (Jörntell and Ekerot 2006, Billings, Piasini et al. 2014). Tonic Golgi cell inhibition of glomeruli means that only stronger signals are transmitted. Assuming contact by mossy fibres on granule cells is at random, granule cells randomly sample mossy fibre frequencies. For those that meet the input threshold the sample size is 3. The number of granule cells that receive contact from 3 active mossy fibres[14] is large, so that if the mossy fibre rates in each sample are averaged (giving the average rate received by that granule cell), the frequency distribution of the sample means is (near) normal by the central limit theorem, even if mossy fibre input to the population is modest, and regardless of the physical distribution of active mossy fibres and range and frequency distribution of the rates they fire at.

The distribution is only approximately normal because the number of mossy fibre rates in a sample is small. This raises the question: why don't granule cells take a bigger sample? A possible reason is that a low number of inputs to granule cells is necessary for optimised pattern separation and decorrelation (Cayco-Gajic, Clopath

[14] In a mutually regulating strip connected overhead by parallel fibres.

et al. 2017). However, a sample size of 3-5 is sufficient to transform the shape of the original distribution (the frequency distribution of mossy fibre rates) into a recognisably (if untidy) bell-like distribution of the sample means.[15] The number of samples in Figure 6.3 is 4,500, an approximate estimate (derived from Figure 6.2) of the number of granule cells that receive 3 or 4 inputs.

Only samples that contain strong signals cause a granule cell to fire, because excitatory inputs must individually be strong enough to win the glomerular competition (and collectively they must be strong enough to depolarise the cell). These form the 'top slice' of the distribution of the sample means. Thus the mean of input rates to the subset of granule cells that fire has a normalised frequency distribution (Figure 6.3): meaning, the bandwidth and shape of the distribution are fixed regardless of the range and frequency distribution of mossy fibre rates, and of the number and binary pattern of active mossy fibres (providing inputs to the system).

Also, the bandwidth is much narrower than the range of mossy fibre rates. Because the mean of the sample means is equal to the mean of the sampled distribution (that is, of mossy fibre rates), again by the central limit theorem, and the distribution of the sample means is normal, granule cell rates vary with the mean of mossy fibre rates (because a normal distribution follows the mean).

[15] Another possible reason may be that a larger number of short dendrites would increase the probability that the same mossy fibre would be resampled.

A key function of Golgi cells (in this contention) is to make regulation of the density of active parallel fibres (discussed in the last subsection) independent of mossy fibre rates as well as of the number and binary pattern of active mossy fibres providing input to the system. Generally stronger (say) mossy fibre rates might be expected to mean fewer inputs to a granule cell are needed to make it fire, so that more should fire (because it is more probable), other things being equal. Instead (we claim that) an increase in mossy fibre frequency, to the extent it makes more granule cells fire, has a self-rectifying effect on the granule cell input threshold. That effect is mediated by a shift in the probability distribution that predicts the number of active parallel fibres which contact a Golgi cell, increasing the probability of stronger inhibition of granule cells, with proportionate downward pressure on the number that receive 3 competitive excitatory inputs, and therefore on the number that fire.

This creates a loop of mutually regulating probabilities which fix the input threshold as well as regulating (density of) parallel fibre activity, and can be quantified and modelled to give the regulated fraction of active parallel fibres. Because the self-regulated density of parallel fibre activity is independent of firing rates, output rates vary faithfully with input rates (where input and output here mean input and output of recoding).

FIGURE 6.3

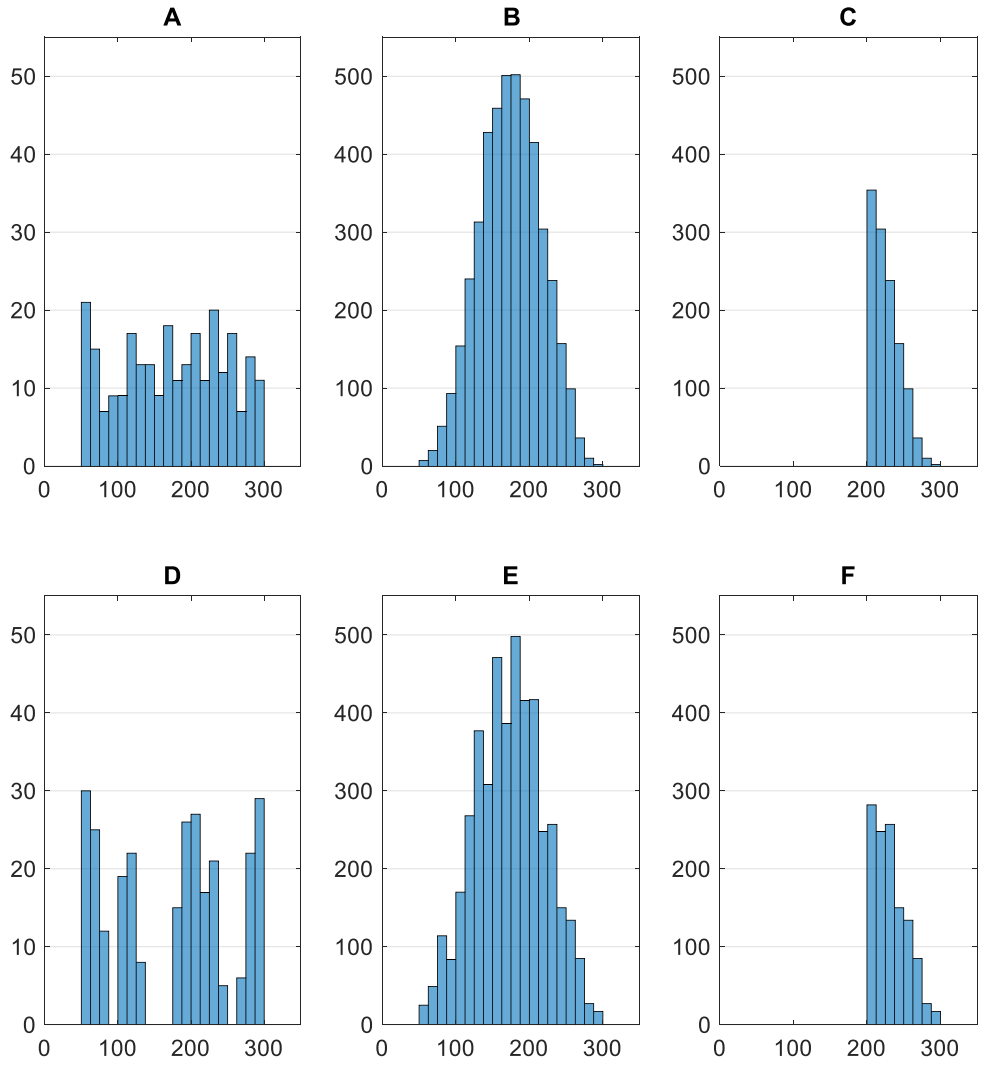


FIGURE 6.3

Random sampling by granule cells of mossy fibre rates. Panels **A** and **D** show a simulation of randomly distributed firing rates of a population of 300 mossy fibres, in the range 50-300 Hz, thought to be the range for activation under behaviour. The probability of firing is manipulated in B to create a discontinuous distribution. 300 is the average number of mossy fibres that are active in Figure 6.2 out of the population of

2,000 that innervate a mediolateral row of 20 mossy fibre terminal cluster fields (a 200 µm x 3 mm strip: a ‘beam’).

In panels **B** and **E** the A and D data, respectively, are each randomly sampled 4,500 times, sample size 3. Each sample represents the input to a granule cell. A minimum of 3 is needed for a granule cell to fire (but is not enough on its own). The number of samples (4,500) is our previous estimate of the number of granule cells in a beam that receive contact from either 3 or 4 active mossy fibres (the large majority 3, hence the sample size). The sample means (the mean of the rates received by each granule cell) are plotted as a frequency distribution. The arbitrary distribution of mossy fibre rates in A and the discontinuous distribution in D are both converted to an approximately normal distribution by the central limit theorem.

A subset of this group – which receive high mean rates of excitatory input – fire. The distribution of the sample means for this group is shown in panels **C** and **F**. Each top slice contains 1,200 granule cells, the average parallel fibre activity estimated in Figure 6.2. The bandwidth of this group is narrower than the mossy fibre range and the shape of the distribution is independent of the mossy fibre distribution. The range of the top slice, but not the bandwidth, varies with the mean of mossy fibre rates (because the mean of the sample means approximates the mean of the sampled population).

It is a further function of Golgi cells is that this adds a second layer of randomisation to decorrelation (the first layer is provided by random contact by mossy fibres on granule cells). As a result, patterns are uniformly dense, because active parallel

fibres are randomly distributed in the sagittal plane. Moreover, because the regulated level of parallel fibre activity is standard (because it is independent of input variables), density is ubiquitous and time-invariant. Fixed density[16] has important functional roles. Some of these are discussed after the next section.

6.2.4 Linear translation of mossy fibre rates → granule cell rates

A final function of Golgi cells – by preserving a stable input threshold – is that it means granule cell firing rates do not receive a variable effect from the number of excitatory inputs they receive. To repeat: three inputs – the probable input threshold (Jörntell and Ekerot 2006, Bengtsson and Jörntell 2009, Billings, Piasini et al. 2014) – are not alone sufficient to make a granule cell fire. Only a subset fire, made up, broadly speaking, of those receiving the strongest aggregate convergent input, where 'strong' refers to the sum of input rates. The average of input rates is a reliable measure of the sum because almost all (granule cells that fire) receive 3 inputs, because the input threshold is fixed and 4 inputs is improbable (and the number of granule cells that receive 4 inputs that are also competitive is very low indeed: Figure 6.2C).

[16] This does not mean there is a uniform distribution or fixed density of active granule cells in the underlying granular layer, or even that a fixed fraction is active of the local number that meet the input threshold.

The mossy fibre-granule cell relay is adapted to transmit short duration mossy fibre action potentials received at high frequency, and impressively adapted (in a number of ways) for high frequency transmission (Ritzau-Jost, Delvendahl et al. 2014 p.152, Delvendahl and Hallermann 2016). Also, the glomerular competition means that in theory the postsynaptic effect of a mossy fibre 'winner' is undiluted, or more weakly diluted, so that a granule cell dendrite is activated at a level proportional to the input firing rate. Moreover, other variables are controlled – each granule cell dendrite receives a single input equidistant from the soma. Granule cells 'have a relatively linear and uncomplicated conversion of depolarisation level to spike rate' (Bengtsson and Jörntell 2009 p.2393, citing Jorntell and Ekerot 2006 and D'Angelo et al 1998), such that the granule cell firing rate is reported to vary faithfully with input signal frequency (Rancz, Ishikawa et al. 2007). Thus, for a number of reasons including the fixed granule cell input threshold, granule cell rates reliably reflect (selected, by recoding) mossy fibre rates.

6.2.5 Isolation of operational variables

Because the density of parallel fibre activity is independent of mossy fibres rates, and translation to granule cell rates is faithful, density of the binary pattern of active parallel fibres is independent of the rates they fire at. As a result, Purkinje cells receive no different effect from the number or binary pattern of inputs they receive, because these are functionally invariant, the first because it is constant and the second because it is random, and because parallel fibre contact is exclusively on

spines on distalmost branchlets of the Purkinje cell arbour. The only functional difference (between binary patterns) is that each is specific to a different number and pattern of mossy fibre inputs (and permutation of mossy fibre rates, since that too is coded in the binary pattern of parallel fibre activity). Similarly, functional variation of the range, bandwidth, permutation and frequency distribution of mossy fibre rates are removed from recoded internal – that is, parallel fibre – activity at the scale of input to a Purkinje cell.

Thus in addition to sparseness of recoded parallel fibre activity, the granular layer causes the number of functional variables expressed in parallel fibre activity to be dramatically cut down to (1) input specificity of recoded patterns; and (2) the adjustable range (as opposed to bandwidth) of the top slice, expressed as granule cell rates. It also procures that these vary independently. This permits those variables to be used in different functions without interference by the execution of either function with the operation of the other, despite being closely functionally integrated. This suggests a challenge to the idea that synaptic weights are a filter that controls output rates, because it obviates the assumption (by learning models) that rates are controlled by learned patterns.

The isolation of rates – by removing an effect of other variables on transmission – is consistent with the reported linear rate coding in the cerebellum (for example Rancz, Ishikawa et al. 2007, Jelitai, Puggioni et al. 2016), and the consistently-reported linear relationship of firing rates and task-related parameters. ‘The firing rate of many

cerebellar neurons is a linear function of task related parameters…[and this] has been found at all levels of the cerebellar circuit' (Raymond and Medina 2018 p.239, who provide references). We note in this connection that temporal coding is reported to be absent at the parallel fibre-Purkinje cell synapse (van Beugen, Gao et al. 2013), seemingly leaving rates to control the simple spike rate, and through the simple spike rate the firing rate of nuclear projection neurons that carry the output of the circuit.

6.2.6 Pattern memory is stored at circuit level

Assuming regulation is ubiquitous, so that the pattern of parallel fibre activity is randomly and therefore uniformly distributed in the sagittal plane, input to a microzone is evenly distributed along its full length. As a result, a climbing fibre volley (microzones are defined by their climbing fibre input, so that climbing fibre input is received as a broadside by the entire Purkinje cell population by definition) is inevitably paired with convergent and uniformly dense parallel fibre input all the way down. Pattern memory trained with a conditioning protocol is therefore stored across a whole microzone – memory is stored and presumably therefore expressed at circuit level rather than at (what learning models propose is) the level of single Purkinje cells.

Finally, 'density' of active parallel fibres is shorthand for the proportion that are active, rather than number per unit area. We mention this distinction because a stable

number per unit area would mean the variable size of the Purkinje cell arbour (which varies substantially between peaks and furrows of the folded cerebellar cortex) (Eccles, Ito et al. 1967) would affect the number of active cells that make contact – so that the number would be a variable. A stable proportion does not have this effect because the size and shape of the Purkinje cell arbour and the thickness of the underlying granular layer co-vary (Eccles, Ito et al. 1967) – the granular layer is thinner in furrows and thicker in peaks, commensurate with the larger Purkinje cell arbour size in peaks.

6.3 Binary memory of cerebellar circuits

6.3.1 Introduction

Near coincidence of climbing fibre and parallel fibre stimulation (therefore ‘paired’) induces long-term depression of the parallel fibre-Purkinje cell synapse (Hansel, Linden et al. 2001, Ito 2001, Qiu and Knopfel 2009) (‘trained’). Learning theories of the cerebellum propose that, following training, Purkinje cells acquire a learned response to input in a known pattern as a result of incremental synaptic weight adjustments trained under climbing fibre tuition (Albus 1971, Fujita 1982, Brunel, Hakim et al. 2004, Dean, Porrill et al. 2010). In this way learning displaces the naïve response to input rates, contributing to a learned effect on motor output. Modified synaptic weights determine the response to a particular set of input signals, which in

that sense the system remembers. Adjustments are made by an algorithm and depend on the model. Pattern memory and control of output coding are inseparable aspects of the learned response, both driven by a single binary pattern of active input cells (and permutation of input values).

We disagree (irreconcilably). We propose that pattern matching is at circuit level, rather than the level of individual Purkinje cells. There is no memory of individual patterns. Discrimination is at the level of, and between, the whole class of known patterns, and the residual class of all other patterns. Memories are not stored as graded synaptic weights. The response does not discriminate between patterns within a class. There is no graded or intermediate response to a partial match. A fraction of the Purkinje cells which innervate a nuclear group is sufficient to veto output of the circuit – if any part of a known pattern (received all along a microzone) is a mismatch, it blocks the response of the whole circuit. This is not to suggest that the output of the circuit is binary, but that the mechanisms of pattern detection and control of nuclear rates are separate. The function of a determination – match or not – is to select which circuits have output and when, but does not control the nuclear rate. The response to a match is permissive – permitting but not coding output. Overlap of stored patterns is well tolerated and useful.

6.3.2 A small number of Purkinje cells is sufficient for strong contact on a whole nuclear group

Substantially all of the output of the cerebellar cortex which converges on a functional group of nuclear cells is from the same microzone or the same functional but dispersed group of microzones which form part of a multizonal circuit (Pantò, Zappalà et al. 2001, Apps and Garwicz 2005). Purkinje cells fire spontaneously at robust rates (Häusser and Clark 1997, Raman and Bean 1999, Cerminara and Rawson 2004, Zhou, Lin et al. 2014), and individually make powerful contact on each of their nuclear targets via 24-36 boutons, each containing multiple synaptic densities (Telgkamp, Padgett et al. 2004, Person and Raman 2012a). 70 out of 86 boutons examined with electron micrographs of mouse medial and lateral cerebellar nuclei had multiple synaptic densities (Telgkamp, Padgett et al. 2004). Out of 10 boutons reconstructed from a single slice all 10 had multiple synapses with an average of 9.2 $\pm$ 1.3 densities per bouton. Purkinje cells outnumber nuclear cells by over 10 to 1 (11:1 in mice, for example) and each Purkinje cell makes contact on 4 or 5 nuclear cells, with convergence of 30-50:1 (in rats, both referenced in Person and Raman 2012a).

The strong firing of Purkinje cells and their individually strong contact on nuclear cells mean that a single Purkinje cell may significantly impact on firing of its targets (Pedroarena and Schwarz 2003). Modulation of nuclear cell firing requires the ‘substantial co-modulation of a large proportion of the PCs [Purkinje cells] that

innervate the cell' (Bengtsson, Ekerot et al. 2011, abstract). We assume that output of a microzone is not topographically organised – so that a single Purkinje cell may contact at random any 4 or 5 nuclear cells (for convenience we use 5) in the nuclear target group. ‘To date, there is no evidence to support [the idea] that different PCs [Purkinje cells] of the microzone control specific CN [cerebellar nuclei] cells within the micro-group [associated group of nuclear cells]’ (Bengtsson and Jorntell in Apps, Hawkes et al. 2018 p. 663).

Assuming 9 synapses on a nuclear cell per Purkinje cell bouton, and contact by a Purkinje cell on 5 nuclear cells, 5 spontaneously active Purkinje cells may inhibit at high frequency as much as half a nuclear group, making around 216-324 synapses on each nuclear cell. The activation of many synapses across several boutons would represent substantial inhibitory drive, from only 2.5% of the Purkinje cells in the microzone. Adding 5 more active Purkinje cells would result in inhibition by 10 Purkinje cells of 5-50 nuclear neurons representing 10%-100% of the target nuclear group, at 216-324 synapses per cell if input is distributed to 100% of the group, or up to 2,160-3,240 synapses in the improbable event that input converges on just 10% of the nuclear group, and so on. These examples are intended to illustrate that a small fraction (2.5-5%) of the population of Purkinje cells in a microzone may powerfully inhibit nuclear firing across a substantial proportion of the target group. How substantial is illustrated in Figure 6.4.

FIGURE 6.4

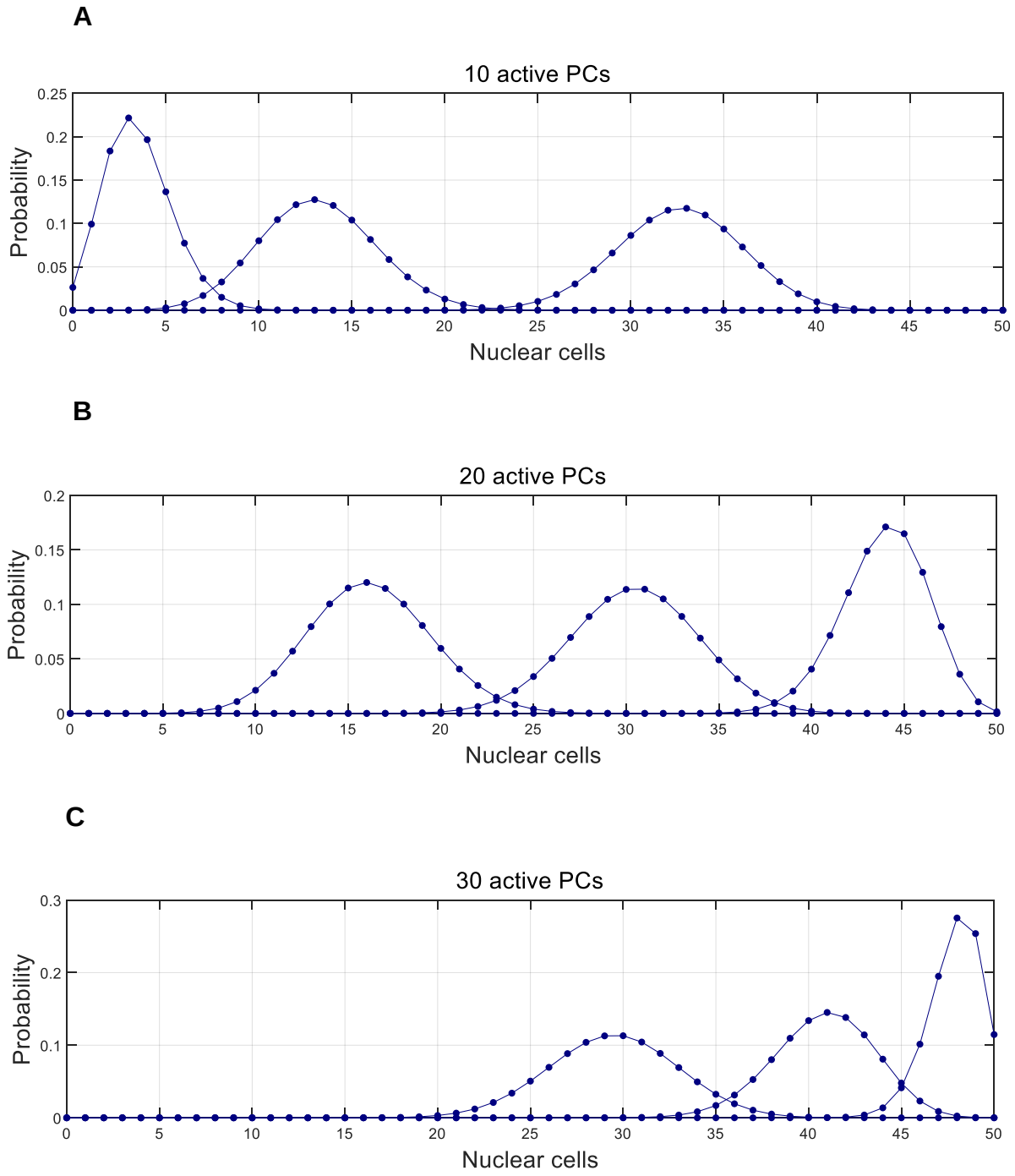


FIGURE 6.4

A fraction of Purkinje cells afferent to a nuclear group is sufficient to make strong contact on the whole of the group. A – C: The probability that the number of

nuclear cells on the x axis (in the range 1-50) receives contact from 1 or more (right peak), 2 or more (centre), and 3 or more (left) out of 10, 20, or 30 (A - C respectively) active Purkinje cells. (y axes are scaled to the data.) The significant feature is how near the right the peaks are, and the narrowness of the majority of the area under the curve (indicating a narrow range of the most likely number of nuclear cells). In B (20 active Purkinje cells), for example, it is likely that all but a small number of nuclear cells receive contact from at least 1 Purkinje cell and the majority receive input from at least 2.

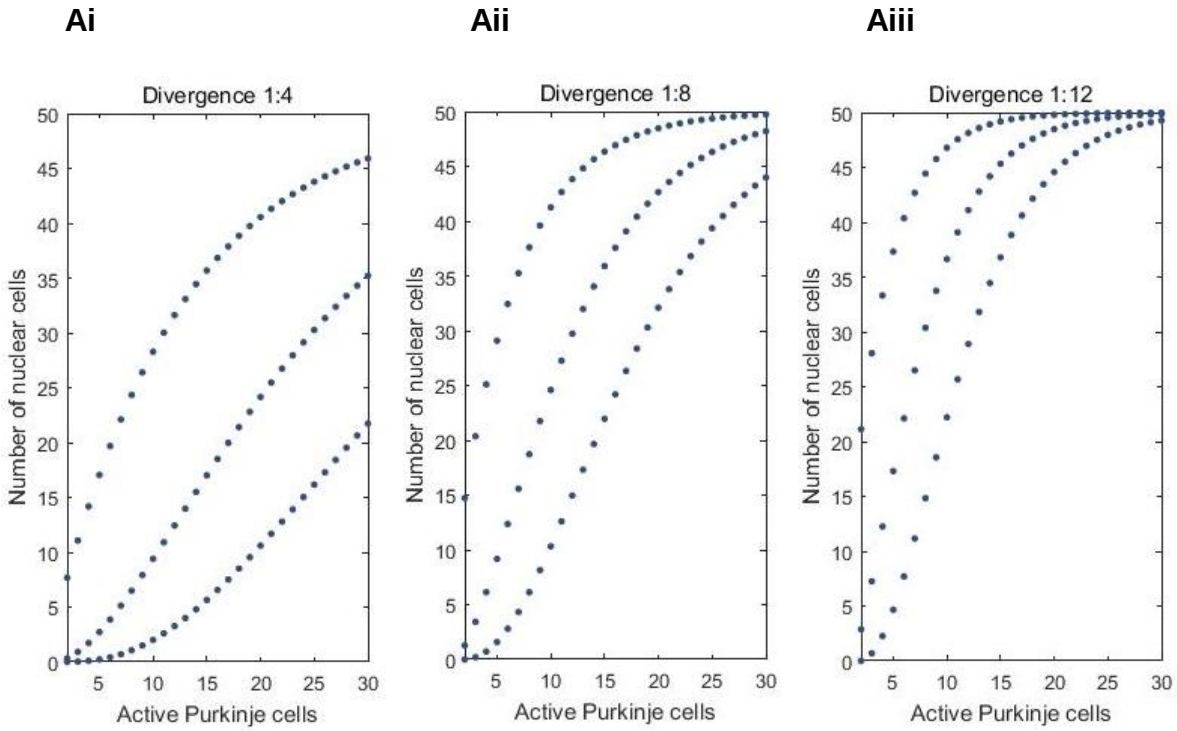
6.3.3 Functional significance

The anatomy of contact by Purkinje cells on the output cells of the cerebellar circuit allows us to draw inferences about its function, assuming random distribution of contact by a Purkinje cell within a nuclear group.

1) The inhibition of a nuclear group reaches functional saturation efficiently, that is, inhibition by Purkinje cells of nuclear cells is broadly equally distributed even at low numbers of active Purkinje cells.
2) A modest number of spontaneously active Purkinje cells is enough to strongly inhibit the whole of a nuclear group. The Purkinje cells do not have to be clustered together but can be any group of the necessary minimum number. However, it is likely that a (mediolateral) row of Purkinje cells that spans a microzone behaves as a group because they sample the same parallel fibre activity.

3) The coordinated suppression of Purkinje cell firing across the whole population of a microzone is necessary in order to disinhibit a nuclear cell group, because less than full coordination means that each nuclear cell is at high risk of receiving strong inhibition.
4) Most of the inhibition of a nuclear cell is functionally supernumerary for most of the time. This is necessary in order that it can be any handful of Purkinje cells which blocks output of the circuit if Purkinje cell rates are not co-modulated across the whole group.

FIGURE 6.5



Bi - iv

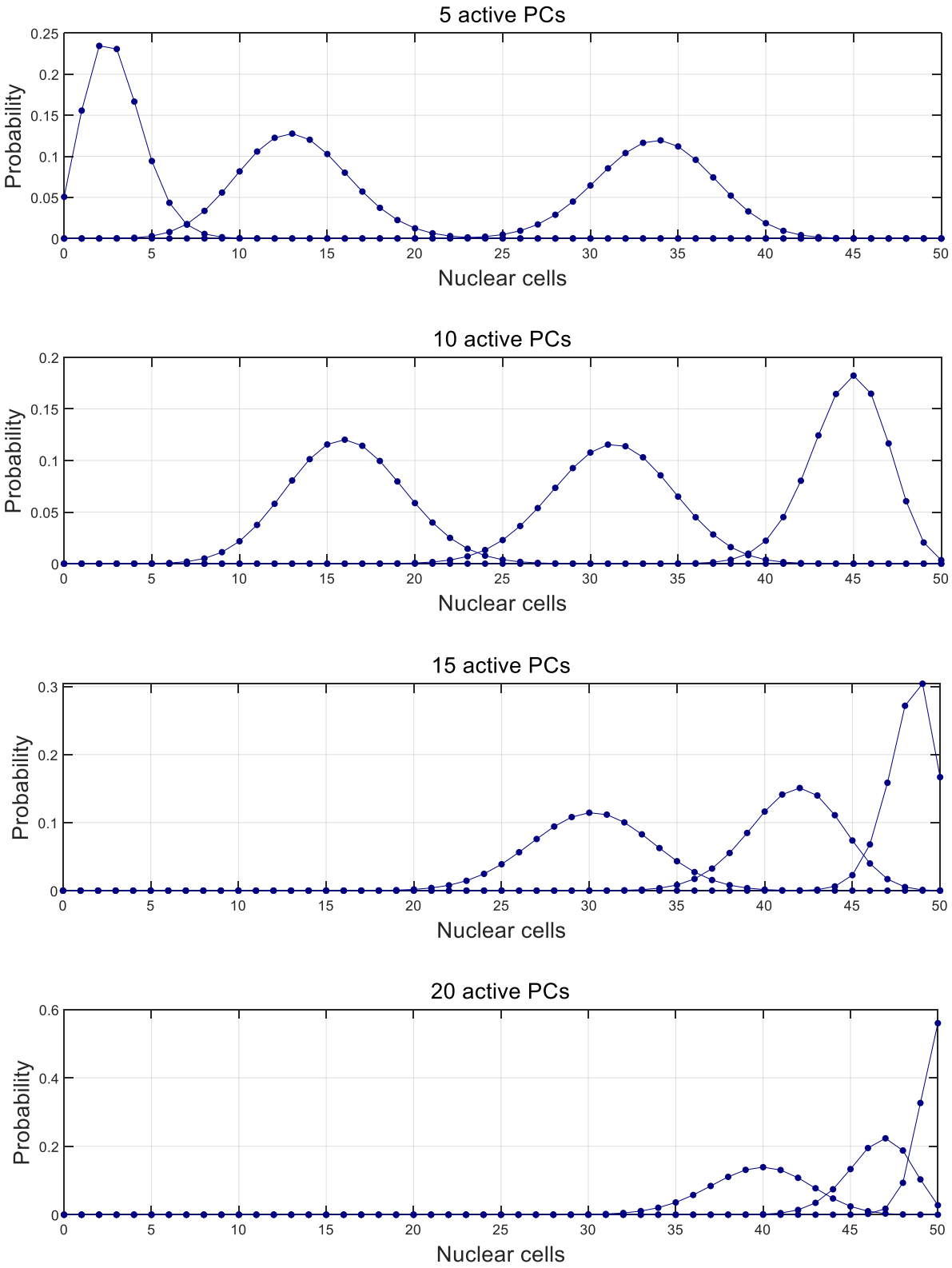

5 active PCs
Probability
Nuclear cells
10 active PCs
Probability
Nuclear cells
15 active PCs
Probability
Nuclear cells
20 active PCs
Probability
Nuclear cells

FIGURE 6.5

The effect of increasing divergence of Purkinje cells onto nuclear cells. Deep nuclei contain excitatory (presumed glutamatergic) interneurons that fire spontaneously and which are normally inhibited by Purkinje cells (Uusisaari and Knopfel 2008) but released from inhibition in the conditioned response. This may effectively increase divergence of Purkinje cells onto nuclear cells because a Purkinje cell has an effect on more than only the nuclear cells it contacts directly. **Ai – iii:** The expected number of nuclear cells, out of a group of 50, which receive contact from at least 1 Purkinje cell (top curve), at least 2 (middle curve) and at least 3 (bottom curve), out of a number x which are active. A steeper curve indicates that fewer active Purkinje cells are needed to inhibit a nuclear group. With divergence of 1:4 (first panel) around 20 active Purkinje cells are needed for an expected 40 out of 50 nuclear cells to receive contact from at least 1 and only around half a nuclear group receives contact from at least 2, whereas with divergence of 1:12 (third panel) and 20 active Purkinje cells, all nuclear cells receive contact from at least one active Purkinje cell and almost all from at least 2. **Bi – iv:** The probability that a number of nuclear cells, x, receives contact from 1 or more (right peak), 2 or more (centre), and 3 or more (left) out of 5, 10, 15 or 20 (Bi - iv respectively) active Purkinje cells, with divergence of 1:10. Compare Figure 6.4 A and B to (in this figure) Bii and Biv respectively.

6.3.4 Adaptations that abet a strong effect by a handful of Purkinje cells

Efficiency of the inhibition of a nuclear group by a low number of Purkinje cells may be increased by nuclear interneurons (see Figure 6.5), and by the distribution of inhibitory and excitatory contact on nuclear projection neurons.

Contact by Purkinje cells on nuclear cells ‘is characterised by preferential targeting of cell somata rather than dendrites’ (Uusisaari and De Schutter 2011, p.3443), while the majority (75%) of excitatory inputs are distal (de Zeeuw and Berrebi 1996). Purkinje cell synapses are therefore positioned to block an effect of excitatory input to nuclear cells. This powerful inhibitory veto is strengthened by inhibition of excitatory interneurons (Uusisaari and Knopfel 2008) because it weakens or silences intrinsic firing of interneurons, so that the tonic effect of intrinsic Purkinje cell activity is both to directly inhibit nuclear cells and to block tonic excitation. The significance of a strong inhibitory bottleneck is that it increases the potency of inhibition by a Purkinje cell of its nuclear targets, so that fewer are need to be potent.

6.3.5 The form of pattern memory

Our previous estimate (in section 2) of the density of parallel fibre activity (number active per unit area in the sagittal plane) would mean that a Purkinje cell receives

contact from several hundreds at a time (n = ~900, assuming a Purkinje cell span of 300 μm). Because activity is evenly and randomly distributed at the scale of input to a Purkinje cell (section 2), training teaches an evenly and randomly distributed pattern of synapses. Pattern density is regulated, uniform and ubiquitous. This consistency is undisturbed by folding of the cerebellar cortex because the size and shape of the Purkinje cell arbour and the thickness of the granular layer are adjusted so that the number of parallel fibres that contact a Purkinje cell is unaffected (to reason from the unaffected number that intersect the Purkinje cell dendritic field) (Eccles, Ito et al. 1967).

The narrow range of the regulated density of active parallel fibres causes a predictable result of overlap of stored patterns. The amount of overlap is predicted by the number stored, and the same for all patterns. The proportion of synapses which also belong to 1 other pattern, and to 2, and 3 and so on is also predictable (Figure 6.6) and also the same for all stored patterns. As more patterns are stored the ratio of total trained to total untrained (with paired input) synapses[17] shifts. Moreover the split is the same for all Purkinje cells trained to the same number of patterns and therefore all Purkinje cells in the same microzone (and all microzones in the same circuit), because climbing fibre signals are received in a volley across the whole population, and are always paired.

[17] References to untrained synapses do not mean there is no plastic effect on transmission of experience but that they have not been trained with paired input.

An estimated 80-85% of parallel fibre-Purkinje cell synapses are strongly long-term depressed, to the extent that there is ‘no detectable somatic response’ to granule cell stimulation (Isope and Barbour 2002 p.9676). This is consistent with a high estimate of ‘electrically silent’ synapses made by parallel fibres activated by cutaneous stimulation (Ekerot and Jörntell 2001). The Isope and Barbour detection threshold could in theory leave room for compound responses (Boris Barbour, private correspondence dated 7 December 2018) i.e. it does not conclusively rule this out. However, if so, this is likely to be weak and (not less important) it would not be pattern-specific but generic, that is, the same for all trained patterns, because random overlap of uniformly dense patterns means they all overlap in the same predictable proportions.

Because trained synapses are functionally inert there is no effect of unrelated memories on the response to other patterns. Trained synapses are (functionally) the ‘same’ weight whether they are part of one pattern or several. The function of parallel fibre-Purkinje cell synaptic depression is to render transmission functionally negligible (in this view). Graded weights are not necessary for this task – in fact graded transmission would impair function.

FIGURE 6.6

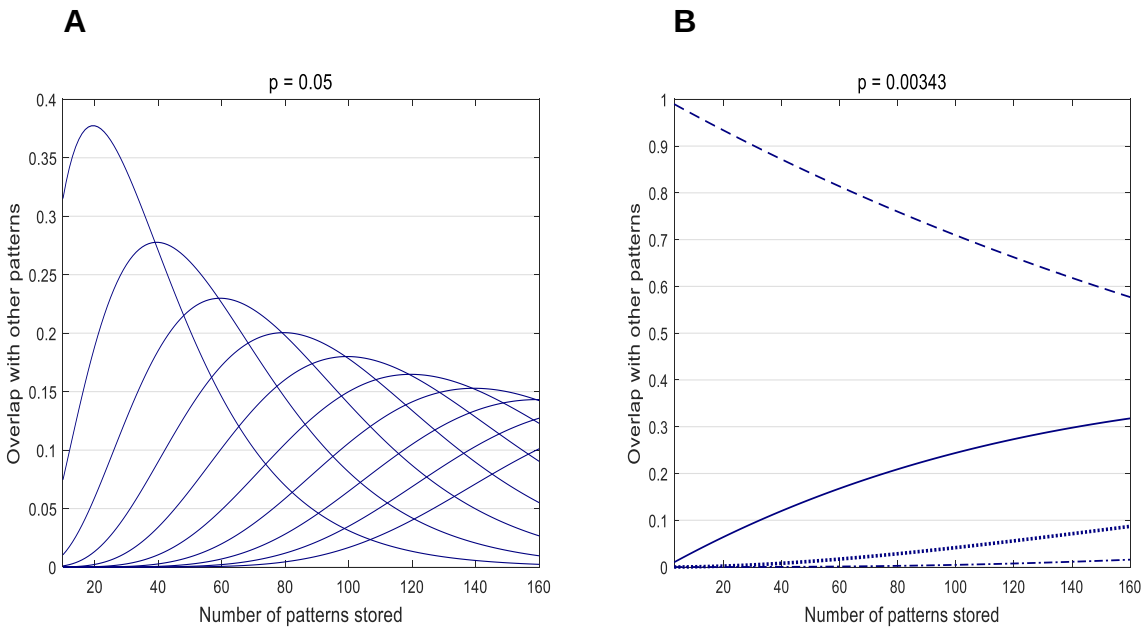


FIGURE 6.6

The changing relative proportions of a pattern which overlap with 1 other pattern, and with 2, and 3 and so on as more patterns are stored is given by a binomial distribution where the solution is the proportion of each pattern (the same for all of them) that overlaps with k other patterns, $n + 1$ is the total number of patterns stored, and p is the fraction of parallel fibres that are active (in reality a constant, because the level of parallel fibre activity is regulated). In (A) a pattern contains 5% of the total number of parallel fibres. The left peak is the proportion of synapses which also belong to one other pattern, the next one to the right is the proportion that also belong to 2 other patterns, the next is the proportion that belong to 3 other patterns, and so on, up to 10 other patterns. The x axis starts at 10 for convenience in running the calculations. (B) assumes 1,200 out of 350,000 parallel fibres are active, the estimated (in section 2)

number at internally regulated levels of activity, so $p = 0.00343$. The pattern is greatly stretched as if pulled from the right. (The purpose of A, with a higher, non-physiological number active, is to show the relationship of the relative proportions of overlap not evident only from the range in B.) The dashed line is the fraction of a pattern that does not overlap with others. The solid line is the proportion of synapses which also belong to one other pattern. The dotted line is the proportion that also belong to 2 other patterns, and the dots and dashes are the proportion that overlap with 3 other patterns. Overlap with 4 and higher numbers is either extremely low or absent in the displayed range.

6.3.6 Discussion

So: a nuclear group is held under tonic inhibition unless there is a coordinated reduction in the firing rate of the entire population of Purkinje cells that innervate it. Put another way, a microzone must receive a match along its full length to modulate nuclear firing. A pattern that does not meet this condition does not lift the nuclear inhibitory blockade.

A failure to meet that threshold does not trigger a graded response but no response at all, regardless of how good or bad the match is overall, or which parts are good and which parts are bad. There is no proportionate (or any other) response to a partial match. This is important because the response to a partial match would be

arbitrary. In this sense, counter to traditional expectation, pattern matching controls the response to input that a circuit does *not* recognise, and not patterns it does. That response is the ongoing and functionally unweakened inhibition of nuclear cells that carry the output of the circuit. This is the default state.

The supernumerary inhibition by Purkinje cells of nuclear cells and strong individual contact, necessary to permit a small handful of (any) Purkinje cells to block output, creates a need (and so can account) for specialist adaptations of Purkinje cell-nuclear cell contact which permit nuclear cells to maintain a baseline rate of firing even under heavy inhibitory bombardment, which remains very sensitive to dynamics of the simple spike rate and can be adjusted in either direction (Telgkamp and Raman 2002, Pedroarena and Schwarz 2003, Telgkamp, Padgett et al. 2004, Turecek, Jackman et al. 2017).

We stress that we do not propose an all-or-nothing response of individual Purkinje cells or suggest that Purkinje cells must be silenced to modulate firing of nuclear cells. Merely if any (more than a low minimum) are not in step with the modulation of the rest, they are sufficient to block an effect on (the whole of) a nuclear group. If only a small part of a pattern is a mismatch (and regardless of which part), it is sufficient to block output.

Contrary to traditional learning models, the effect of training is to *eliminate* variation of the post-synaptic effect of known patterns, because input in a known pattern is received exclusively at severely depressed synapses, and this is true for all known patterns. The pattern that cues output does not also code it.

All unknown patterns are equal in having no effect through 85% of synapses and otherwise addressing a random sample of operational synapses. That is, pattern memory is binary – the response indicates only if a pattern is known or not, and does not discriminate individually between known patterns, or between unknown patterns. There is no need for graded synaptic weights, because learned patterns do not control the output firing rate (we claim). The role of pattern matching in the conditioned response is instead permissive – it selects which circuits have output and when, but does not code output.

'Binary' here means something very different from the perceptron meaning. There, a trained Purkinje cell is limited to two response, say 1 or 0 (Brunel, Hakim et al. 2004). Both responses are learned i.e. the result of iterative adjustment of synaptic weights. All patterns have an effect on the postsynaptic Purkinje cell which either falls above or below a threshold. The size of both classes is limited by storage capacity. Learning uses an algorithm to adjust synaptic weights. The correct response following training requires a repeat of a learned permutation of input rates. Our meaning is *none* of these things. The response does not control Purkinje cell rates (output rates are controlled separately), there is no algorithm, pair-trained synaptic

weights are invariant (all silent), and unknown patterns do not need learning, so the size of the class of unknown patterns is unlimited. Also, the response (mediated by the pattern matching function) of the postsynaptic cell is triggered by the binary pattern of inputs (some on, some off) only. The permutation of rates does not affect the response.

6.4 Stellate cell networks

6.4.1 Background

As noted, learning models propose that the cerebellum implements a supervised learning algorithm which uses iterative adjustment of parallel fibre synaptic weights. Purkinje cells in this way acquire a learned response driven by input in a remembered pattern, displacing the naïve response to input rates (Albus 1971, Fujita 1982, Brunel, Hakim et al. 2004, Dean, Porrill et al. 2010).

Following training with a conditioning protocol (or direct stimulation which mimics a conditioning protocol) Purkinje cells respond to the conditioned stimulus with a reduction or pause in firing (Jirenhed, Bengtsson et al. 2007, Rasmussen, Jirenhed et al. 2008). Purkinje cells are interleaved with and receive contact from inhibitory interneurons (Palay and Chan-Palay 1974, Paula-Barbosa, Tavares et al. 1983, Sultan and Bower 1998) which in turn receive contact from parallel fibres. It is thought the conditioned response of Purkinje cells is abetted by feed-forward

inhibition of Purkinje cells by molecular layer interneurons (MLIs) driven by parallel fibres, following training.

It has been reported that ‘MLIs encode locomotion-dependent changes in GC [granule cell] input with linear changes in firing rate’ (Jelitai, Puggioni et al. 2016 p.6) (counting spikes in 200 ms bins to calculate frequency). (The method of measurement in this study did not permit a distinction between stellate cells and basket cells.) But what restricts feed-forward inhibition to the learned response, and what chooses the source of input signals to control it?

6.4.2 Bad memory

Pair training potentiates the parallel fibre-stellate cell synapse (Jörntell and Ekerot 2003, Rancillac and Crépel 2004, Smith and Otis 2005, Jörntell and Ekerot 2011). Potentiation is reversed by unpaired, parallel-fibre-only input. Assuming the high proportion of silent parallel fibre-Purkinje cell synapses (Isope and Barbour 2002) is because they have received endogenous training with paired input, we would expect a similar proportion of parallel fibre-stellate cell synapses to be potentiated under tuition of the same climbing fibre signals (and we can confirm mathematically that anatomical differences don’t affect the proportion).

Because trained synapses are potentiated, the high proportion should mean that – on the face of it – feed-forward inhibition is indiscriminate, that is, driven by known and unknown patterns alike. Put another way, stellate cells are pattern blind. We are able

to confirm this by calculating the probability distribution for input to operational synapses with a known and with an unknown pattern of parallel fibre activity (see deeper level section). Learned blindness of interneurons is a problem the cerebellum has therefore been obliged to solve. In fact, the position is different for stellate cells at superficial and deeper levels of the molecular layer because input they receive from parallel fibres is topographically stratified. The main focus of the following discussion is on deeper-stratum stellate cells but superficial level is discussed first.

6.4.3 Superficial level

MLI morphology varies continuously (Palay and Chan-Palay 1974, Paula-Barbosa, Tavares et al. 1983) with the depth of the cell body in the molecular layer. Among other things, deeper cells have longer main axons, more and longer collaterals, and drop more descending collaterals that terminate in the Purkinje cell layer (Paula-Barbosa, Tavares et al. 1983, Sultan and Bower 1998). MLIs at deeper level are networked, so that they receive tonic inhibition – MLIs fire intrinsically (Häusser and Clark 1997, Ruigrok, Hensbroek et al. 2011). By contrast, connections between superficial stellate cells, whose axons wander around without leaving their dendritic field (Palay and Chan-Palay 1974), are less frequent or absent, so that superficial cells receive weak or no contact from each other.

Pattern blindness raises the question, what is learning for at the parallel fibre-stellate cell synapse? Evidence comes from the C3 region of the cerebellar cortex, which is involved in forelimb movement. Stellate cells in this region respond (by firing)

exclusively to stimulation of an associated discrete region of the body surface, or receptive field, (Jörntell and Ekerot 2002, Ekerot and Jörntell 2003, Jörntell and Ekerot 2003, Jörntell and Ekerot 2011), and not to stimulation of other fields. It is thought that this stimulation drives paired input which trains potentiation of the parallel fibre-stellate cell synapse. Input to stellate cells evoked by stimulation of other fields is to untrained synapses. Transmission at an 'untrained' synapse is not weak but nil (Jörntell and Ekerot 2003, even at several 100 Hz: Henrik Jorntell, private correspondence dated 31 March 2017).

The interneurons used in these studies were activated by cutaneous stimulation, effectively making it a selection requirement that they were outer level, although they were not expressly selected for their depth. This is because there is, in the C3 region of the cerebellar cortex in adult cats, 'a specific depth distribution of granule cells depending on the type of input they received' (Jörntell and Ekerot 2006, p.11795, Quy, Fujita et al. 2011). Input triggered by cutaneous stimulation is received by superficial granule cells. Superficial granule cells prevalently bifurcate in the outer level of the molecular layer (Palay and Chan-Palay 1974, Zhang and Linden 2012), where they contact outer level stellate cells.

The wiring of input to C3 circuits effectively extends modular cerebellar circuit wiring to the body surface. Stimulation evokes mossy fibre and climbing fibre signals that travel by different pathways to converge on the same vertical slice of the cerebellar cortex (Garwicz, Ekerot et al. 1998, Garwicz, Jörntell et al. 1998). Climbing fibres

terminate on Purkinje cells while mossy fibres terminate on granule cells in the subjacent granular layer. On the whole, there is ‘a close correspondence [of terminal fields] between inputs conveyed by climbing fibres to the molecular layer and those conveyed by mossy fibres to the underlying granular layer’ (Apps and Hawkes 2009 p.677). Thus learning is induced with a single peripheral stimulus which evokes paired input – not just the same type of stimulus but the same event. This ‘is a feature that seems to be observed across species and other parts of the cerebellar cortex (for example, crus II in the rat)’ (Apps and Garwicz 2005, p.305).

Thus at outer level paired patterns of parallel fibre activity are not perfectly random, being made up of the subset of parallel fibres that are the axons of local granule cells in the underlying strip of the granular layer. As a result, stellate cells at this level are very good at discriminating between activity of local granule cells and signals originating elsewhere, in other circuits, triggered by stimulation of other fields.

6.4.4 Sub-superficial level

At deeper level, where input is not sourced in the same way, pattern separation in this way is not possible. Deeper-lying stellate cells receive contact from the parallel fibres of deeper-lying granule cells, which in the C3 region receive signals triggered by limb movements (Jörntell and Ekerot 2006), presumably by activation of receptors and nerve endings in muscles and tendons – proprioceptors. A pattern of parallel

fibre activity at this level has no special relationship with the source of climbing fibre signals, does not have the same trigger as climbing fibre signals, and terminates in a disorganised way, mixing signals from different sources.

Because we (claim to) know that the density of active parallel fibres is in a fixed narrow range (derived in section 2), we can calculate the probability distribution for input to a stellate cell – the probability that given any particular pattern of parallel fibre activity a stellate cell receives contact from 0 active cells, from 1, from 2 and so on.

A Purkinje cell receives an estimated 175,000 parallel fibre synapses (Napper and Harvey 1988), one each from around the same number of granule cells (Harvey and Napper 1991), while a roughly equal number of parallel fibres pass through the same space without making contact. A deeper-level stellate cell dendritic field (in adult rats) extends around 110-130 μm in both horizontal and vertical directions (Palay and Chan-Palay 1974, pp.217-221). Assuming a Purkinje cell and stellate cell arbour size of 300 x 300 μm and 120 x 120 μm respectively, some 350,000/~6 or ~60,000 parallel fibres pass through a stellate cell dendritic field, assuming parallel fibres are evenly distributed. If a stellate cell receives contact from 1,000 parallel fibres, around 1 in every 60 parallel fibres which pass through a field makes contact. 1 in ~200 are active (derived from the estimated proportion that are active), so around 300. There is accordingly a probability of 1/60 = ~0.01667 that an active cell makes contact. The probability of contact by k (out of 300) active cells is therefore given by

$$\frac{300!}{k!\,(300-k)!} * 0.01667^{k} * (1-0.01667)^{300-k}$$

(2)

This gives the Table A probabilities for the range k = 0 to 10 where k is the number of active fibres that contact a stellate cell and p is the probability of that number given a random pattern of parallel fibre activity. There is accordingly a high probability (p = 0.9074) that the number lies between 2 and 8. This agrees closely with the reported number, indicated by 'two to eight substantial EPSPs [excitatory postsynaptic potentials]' (Jörntell and Ekerot 2003 p.9628).

An unknown pattern of input to a stellate cell may be to all trained synapses, all untrained, or a mixture (with all untrained the least likely). The probability of n inputs at trained synapses is the probability of n inputs to any synapse reduced by the probability that at least one input is to an untrained synapse (so: 1 – *P*(none)) but increased by the sum of the probabilities that a higher total number of inputs, a, is reduced by a number, b, to untrained synapses, such that $a - b = n$. For example, the odds that there are two inputs to trained synapses is increased by the product of the probabilities that there are three inputs in total, and any one is to an untrained synapse. So the probability of n inputs at trained synapses with a random pattern is:

$$P(n) - P(n)(1 - 0.85^n) + \sum_{y=1}^{z} P(n+y)\left(\frac{(n+y)!}{y!\,n!} * 0.15^y * (1-0.15)^n\right) \quad (3)$$

where $P(n)$ is the probability of n inputs (to any synapse, derived in Table A) and $n + z$ is the maximum number of inputs with more than insignificant odds (so for example around 10 in Table B because higher numbers have a very low probability).

Table A

k	0	1	2	3	4	5	6	7	8	9	10
p	.006	.033	.083	.14	.176	.177	.147	.105	.065	.036	.018

Table B

n	0	1	2	3	4	5	6	7	8	9	10
p	.013	.059	.12	.182	.195	.16	.117	.070	.035	.014	.003

FIGURE 6.7

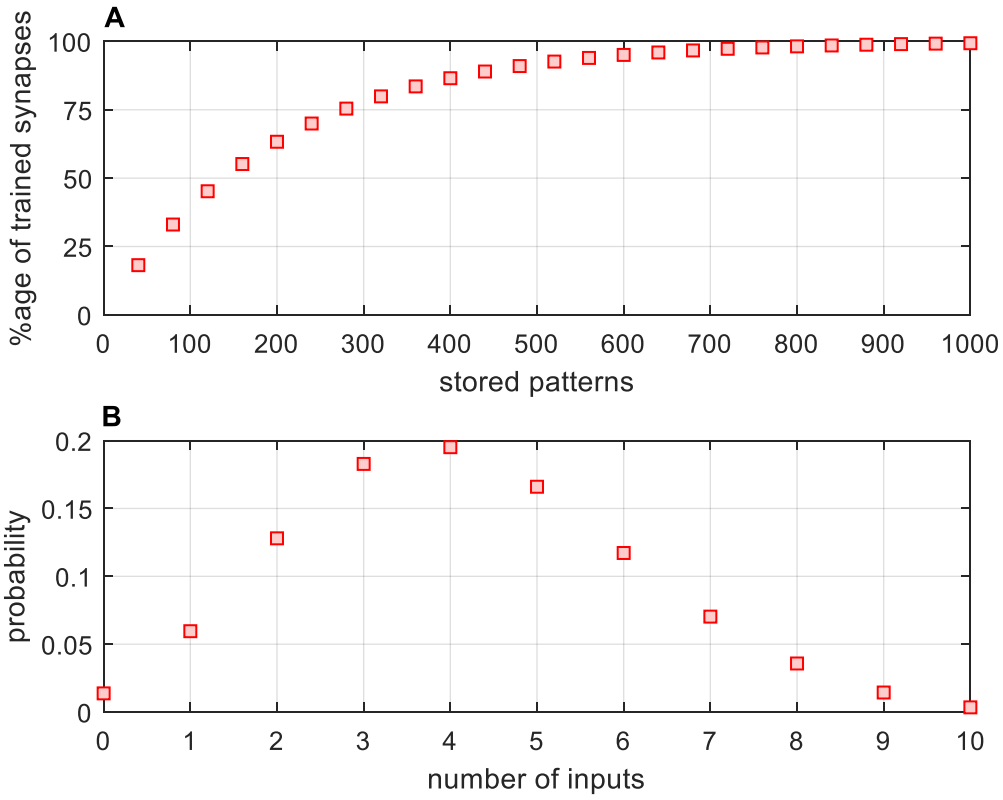


FIGURE 6.7

A: Estimate of the storage capacity of a stellate cell assuming that a total of 1,000 parallel fibres make contact. Data are thinned for clarity. The number that make contact per pattern is the weighted average of the Table A probability distribution. If 85% of synapses are trained – predicted by the reported proportion for the parallel fibre-Purkinje cell synapse – a stellate cells stores around 370 patterns. **B:** The probability distribution for the number of active parallel fibres that make contact on a stellate cell at trained (therefore operational) synapses, assuming 85% of synapses are trained, with an unknown pattern of input – Table B in graph form. The number is in the range 0 – 10 (a higher number is very unlikely). Each number of inputs has an associated probability.

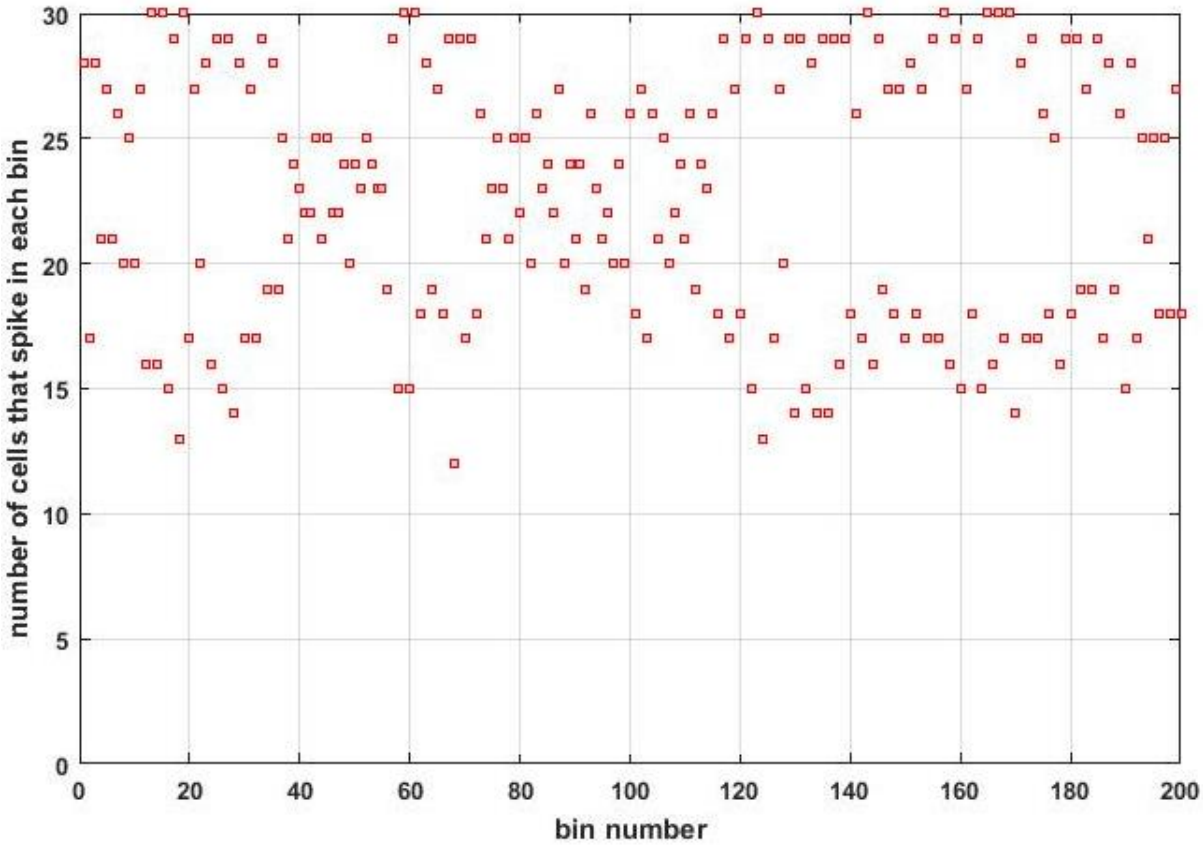
This argues: 1) The number of inputs received at operational synapses depends much more on chance than on learning. 2) Synaptic weights are not pattern specific, because stored patterns overlap very substantially (and the effect on transmission is not even generic, because pattern size is small). 3) Sub-superficial MLIs are in any case pattern blind, so unable to have an effect that requires them to store and remember patterns. 4) The fixed, regulated level of parallel fibre input to sub-superficial MLIs is insufficient to drive a learned response, predicting a different or additional source. This is consistent with evidence that a number of mechanisms severely weaken an effect of isolated excitatory input to stellate cells on dendritic signalling (Häusser and Clark 1997, Abrahamsson, Cathala et al. 2012, Tran-Van-Minh, Abrahamsson et al. 2016), arguing that a somatic effect is blocked at the low numbers a stellate cell receives (Table B) at regulated levels of parallel fibre activity. It is also supported by a model of stellate cell network activity that replicates stellate cell spiking under inhibitory input only (Figure 6.8).

6.4.5 Control of stellate cells in the conditioned response

What, then, elevates the stellate cell rate during the conditioned response? A candidate to contribute to this function is nucleocortical feedback (Houck and Person 2015, Gao, Proietti-Onori et al. 2016). Feedback is carried by collaterals of nuclear cell axons that project back to the granular layer where they terminate as mossy fibres, forming a closed circuit (Gao, Proietti-Onori et al. 2016). Nuclear cells fire intrinsically even under heavy tonic inhibition by Purkinje cells, partly as a result of

short-term depression at stable afferent rates (Telgkamp and Raman 2002, Pedroarena and Schwarz 2003, Telgkamp, Padgett et al. 2004). Nuclear cells remain very sensitive, however, to the dynamics of the simple spike rate (Telgkamp and Raman 2002), such that their firing is transiently modulated by a *change* of the rate, which can be in either direction (Pedroarena and Schwarz 2003, Baumel, Jacobson et al. 2009). So that a drop in the Purkinje cell rate, such as the learned pause seen following training with a conditioning protocol, causes them to fire at a transiently elevated rate.

FIGURE 6.8



17 Hz (top), 24.2 Hz (bottom)

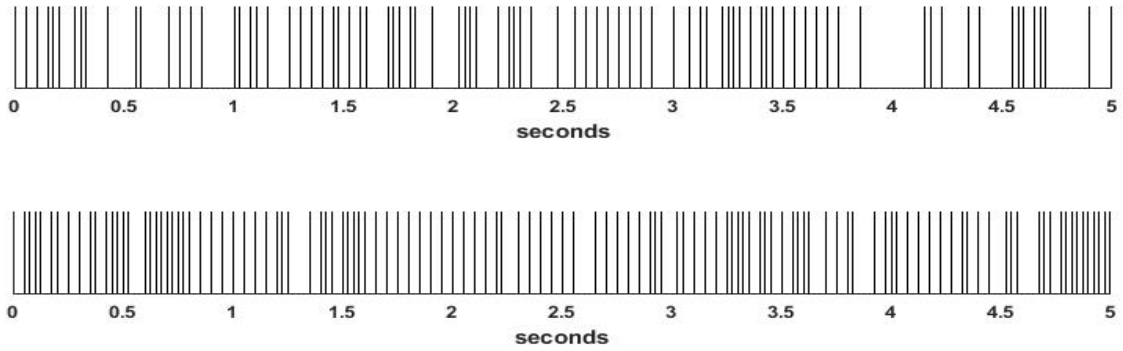


FIGURE 6.8

Simulation of a sub-superficial stellate cell network. The graph shows the number of deeper-stratum stellate cells (out of 30 which contact a simulated Purkinje cell) that are active in each of two hundred 25 ms bins (so for 5 seconds). Cells were modelled individually, each receiving contact from a number of their neighbours determined by probability that varies with the depth of the cell body in the molecular layer, to reflect depth-dependent anatomy. The un-networked, intrinsic firing rate of a stellate cell is around 40 Hz (hence 25 ms bins). A cell either spikes in a bin or is silent, depending on the number of cells that make contact on it that spike in the bin before. This reflects (in discrete bins for convenience) the unpredictable delay reported to be caused by the erratically-variable time and number of recent input spikes. The firing pattern of individual cells generated by this model (examples are shown under the graph) is a good match with the reported firing pattern (compare Häusser and Clark 1997 figs 1 and 3). The model replicates stellate cell spiking under inhibitory input only, suggesting that an effect of excitatory input is normally blocked at low numbers of inputs, so at homeostatically regulated density of parallel fibre activity.

Closed-circuit feedback thus provides a targeted and timely injection of extra mossy fibre input to a microzone in the conditioned response, following training. This does not create a positive feedback loop (we suggest, contrary to the reported interpretation) because the dominant effect is to cause an uplift in the number of local granule cells that meet firing threshold – as opposed to an increase in granule cell firing rates. The reason is that many more cells that received only 2 inputs before – and so did not fire – now receive an extra input and thus reach the input threshold (the minimum necessary to fire, illustrated and discussed in Figure 6.9), than the number that received three and now receive a supernumerary fourth input. The sum of excitatory inputs received by networked stellate cells during the conditioned response is thus temporarily elevated. Elevation is sufficient to overcome the tonic somatic inhibitory blockade. As a result, deeper-stratum stellate cells become phasically sensitive to excitatory input, and inhibit Purkinje cells at elevated, supra-tonic rates.

Mossy fibre input to the cerebellum is vertically topographically organised (reported for the C3 region: Jörntell and Ekerot 2006, Quy, Fujita et al. 2011). Input triggered by cutaneous stimulation is received by superficial granule cells, and deeper-lying granule cells receive signals triggered by limb movements (Jörntell and Ekerot 2006), suggesting it is driven by activation of internal receptors in muscles and joints: proprioception. Topography is preserved in the molecular layer, with ‘granule cells in the inner granule cell layer giving rise to PFs [parallel fibres] in the inner molecular layer and granule cells in the outer granule cell layer giving rise to PFs in the outer

molecular layer' (Zhang and Linden 2012, p.122), as a ‘prevalent rule’ (Palay and Chan-Palay 1974, p.66).

Nucleocortical feedback terminates mainly in the superficial granular layer (Gao, Proietti-Onori et al. 2016), therefore prevalently exciting granule cells whose axon divides in the superficial molecular level. Ascending axon contact on stellate cells is unconfirmed but would mean contact was confined below superficial level to the same closed circuit. At superficial level, discrimination of stellate cells between parallel fibre signals arising in that circuit, in the subjacent granular layer, and in other circuits, is excellent, as discussed (Garwicz, Jörntell et al. 1998, Jörntell and Ekerot 2002, Jörntell and Ekerot 2003), such that feedback would excite stellate cells in the same circuit and have no effect at all in other circuits.

Because stellate cells at deeper level have poor pattern memory control in the conditioned response, when they are briefly responsive, can be by parallel fibre signals which do not need to be in a known configuration (memory is unnecessary either of the binary pattern of activity or the permutation of rates). Input to a microzone in the control window is distinct from the pattern which is used for pattern matching. Signals used in pattern matching are spread across a longer time window (the learning window for synaptic plasticity), and learned by repetition. Signals that control simple spike rates acting through stellate cells – and therefore nuclear rates – are in a time-adapted (because the intrinsic Purkinje cell response is time-adapted:

Johansson, Jirenhed et al. 2014)[18] and narrower window, and are not constrained to a repeating configuration. Possibly the windows overlap and it is likely that both sets of signals are from substantially the same set of (movement-related) sources. However, signals activity in the control window represents more up-to-date information, transmitted by wide-diameter, myelinated, fast-transmitting neurons (Loeb and Mileusnic 2015). It is not rigidly confined to the same set of parallel fibres every time, allowing control of Purkinje cell rates by conditions on *this* occasion not the last time or previous occasions, when the learned pattern was memorised.[19]

6.4.6 Linear translation of granule cell rates → Purkinje cell rates

MLIs reflect 'granule cell input with linear changes in firing rate' (Jelitai, Puggioni et al. 2016 p.6). The effect of inhibition by MLIs of Purkinje cells is rate dependent and linear, such 'that locomotion-dependent modulation of the balance between excitation and inhibition [of Purkinje cell dendrites] generates depolarising or hyperpolarising dendritic V_m [dendritic membrane voltage] changes that linearly transform into bidirectional modulation of PC SSp [Purkinje cell simple spike] output' (Jelitai, Puggioni et al. 2016 p.9).

[18] Purkinje cells acquire an intrinsic plastic response to known input, even with MLIs pharmaceutically blocked Johansson, F., D. A. Jirenhed, A. Rasmussen, R. Zucca and G. Hesslow (2014). "Memory trace and timing mechanism localized to cerebellar Purkinje cells." Proc Natl Acad Sci U S A **111**(41): 14930-14934.. The response is an adaptively-timed, mGluR7-dependent Johansson, F., H. A. Carlsson, A. Rasmussen, C. H. Yeo and G. Hesslow (2015). "Activation of a Temporal Memory in Purkinje Cells by the mGluR7 Receptor." Cell Rep **13**(9): 1741-1746. transient fall in their firing rate.

[19] Also, the spatial pattern of parallel fibre activity is sensitive to mossy fibre rates, so the timing of the window opened by pattern matching is rate sensitive.

FIGURE 6.9

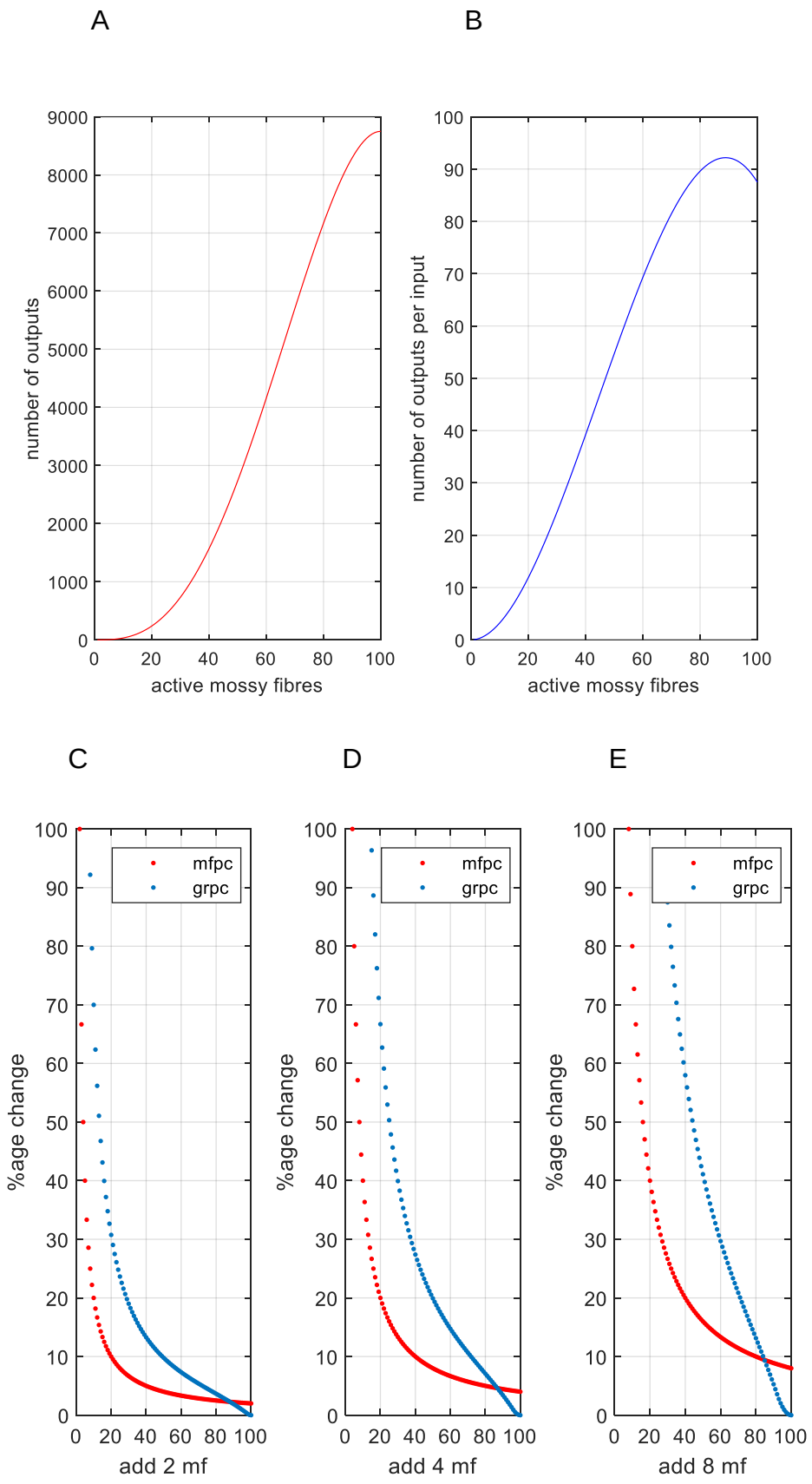

A
B
number of outputs
active mossy fibres
number of outputs per input
active mossy fibres
C
D
E
mfpc
grpc
%age change
add 2 mf
add 4 mf
add 8 mf

FIGURE 6.9

Closed circuit feedback in the conditioned response provides extra mossy fibre input to trained circuits. As discussed (in section 2) the number of granule cells that meet the input threshold in a region with the dimensions of a mossy fibre terminal cluster field can be calculated from the number of active mossy fibres. **A:** The number of inputs (active mossy fibres) to a cluster field plotted against the number of outputs (granule cells that meet the input threshold). To fire, a granule cell needs mossy fibre input to either 3 or all 4 dendrites. Modulation by Golgi cells is disregarded. **B:** Nucleocortical feedback adds to the number of mossy fibre inputs received in the superficial granular layer. The number of outputs per input increases as the number of inputs rises. That is, a single mossy fibre (either more or less) has a larger effect if it is one of a larger number of inputs. This relationship breaks down – the curve peaks – at the high end of the range because an extra input is more likely to be to the fourth dendrite of a granule cell which already fires. **C-E:** The percentage change in the number of active granule cells (blue data) caused by the addition of an extra 2, 4 and 8 active mossy fibres (C, D and E respectively) to a number that are already active, x. The red data show the percentage change in the number of mossy fibres. For most of the range a change to the mossy fibre number causes a larger percentage change to the granule cell number. Blue values are lower than red values at very high values of x because at those levels, many granule cells receive 4 inputs, and so still receive enough input to fire if one is removed.

Thus, in the control window, when the balance of excitation and inhibition tips strongly towards inhibition (because direct input to Purkinje cells is received at inert

synapses, and an effect of excitatory input to MLIs is unblocked), there is linear translation of granule cell to simple spike rates. This is in conflict with the perceptron and adaptive filter models where output is controlled by learning and not by input rates.

6.4.7 Selection of movement-related input signals to control output rates

The longer and more highly collateralised axons of deeper-stratum stellate cells mean that the concentration of synaptic contact by stellate cells on Purkinje cells increases with depth. So assuming the strength of the inhibition of Purkinje cells increases with the concentration of contact, it increases with depth. Therefore, when the sub-superficial block of feed-forward inhibition is suspended in the conditioned response, stellate cells at that level dominate control of the simple spike rate. This has the result that there is depth-dependent selection of parallel fibre signals that control the simple spike rate in the conditioned response, so that control is by mossy fibres received by those granule cells, in the corresponding stratum of the granular layer. In the C3 zone those signals are movement related (Jörntell and Ekerot 2006).

Thus performance-sensitive, phase-locked mossy fibre signals may dominate control of feed-forward inhibition with learned timing provided by training with the endogenous equivalent of a conditioning protocol driven by a movement cycle. Movement-sensitivity may not be limited to signals generated internally. Some skin

receptors may provide ‘cutaneous proprioception’ during movement (Jörntell and Ekerot 2006), meaning: distortion of the skin during movement, especially cyclic movement, may cause some mechanoreceptors to generate a time-varying signal.

6.5 Control of output rates

6.5.1 Background

Smooth and precise control of movement at variable speeds, thought to be a function of the cerebellum, is dependent on incremental modulation of motor output on a short timescale (milliseconds). The large majority of cerebellar circuits (i.e. non-floccular circuits) include a discrete group of deep nuclear cells which contain excitatory projection neurons which carry the output of the circuit.[20] Nuclear cells receive heavy inhibitory input from Purkinje cells. Purkinje cells are organised functionally into groups of hundreds of cells called microzones (Oscarsson 1979, Ozden, Sullivan et al. 2009, Ramirez and Stell 2016) whose output is channelled down onto a much smaller number of nuclear cells.

[20] Unless otherwise stated, ‘nuclear cell/neuron’ is used as shorthand for these cells (although deep nuclei also contain other cell types), and circuit as a synonym of microcomplex, including a multizonal microcomplex (a circuit that contains more than one microzone) Apps, R. and M. Garwicz (2005). "Anatomical and physiological foundations of cerebellar information processing." Nat Rev Neurosci **6**(4): 297-311.

Purkinje cells fire intrinsically (Häusser and Clark 1997, Raman and Bean 1999, Cerminara and Rawson 2004) at robust rates (Zhou, Lin et al. 2014). Nuclear neurons, which also fire spontaneously (Person and Raman 2012b, Mercer, Palarz et al. 2016), maintain a baseline rate of firing even under heavy inhibitory bombardment. The baseline rate is independent of inhibitory frequency at stable rates but remains very sensitive to dynamics of the simple spike rate and can be adjusted in either direction (Telgkamp and Raman 2002, Pedroarena and Schwarz 2003, Telgkamp, Padgett et al. 2004, Turecek, Jackman et al. 2017), so that a fall in the simple spike rate causes transient elevation of the nuclear rate, and vice versa.

Following training with a conditioning protocol or afferent stimulation that mimics a conditioning protocol (Jirenhed, Bengtsson et al. 2007, Rasmussen, Jirenhed et al. 2008), Purkinje cells respond with a transient reduction and sometimes a full pause in firing. The conditioned fall in Purkinje cell rates may in some circuits gate mossy fibre collateral input to nuclear cells. However, not all mossy fibres send collateral projections to deep nuclei. Only 1 out of 15 mossy fibres originating in dorsal column nuclei, for example, had a collateral which terminated in a deep nucleus (Quy, Fujita et al. 2011), and there is generally light collateral input to the dentate nucleus, such that ‘suppression of PCs [Purkinje cells]…plays the primary role in generating output from DN [the dentate nucleus]’ (Ishikawa, Tomatsu et al. 2014). Thus the Purkinje cell rate may in some circuits be the main and even sole controller of nuclear output cells, while in others it contributes control (Ishikawa, Tomatsu et al. 2014, Jirenhed and Hesslow 2016). Optogenetically controlled pauses in Purkinje cell firing produce robust motor output (Heiney, Kim et al. 2014, Lee, Mathews et al. 2015).

This appears to raise a number of problems. (1) Spike timing of Purkinje cells is individually unpredictably erratic, so on the face of it poorly suited to smoothly graded control of nuclear rates. This is seen, for example, in recordings from single cells across the step cycle of a mouse (Sauerbrei, Lubenov et al. 2015). (2) It is thought that both sensory information and motor commands are predominantly coded as firing rates (Delvendahl and Hallermann 2016). At least two spikes are needed to infer a so-called instantaneous rate but a reliable 'reading' requires temporal integration over a longer period. How are rates coded on a shorter, behavioural timescale? (3) Purkinje cell-mediated inhibitory postsynaptic currents are very fast, with time constants of around 2.5 ms (Person and Raman 2012b, Mercer, Palarz et al. 2016), suggesting a short window. An optimum of 3 ms has been estimated for Purkinje cell targets that control eye movement (Payne, French et al. 2019). Without integrating over a longer period how is a smooth response generated at short time scales relevant for behavioural control? (4) There is no evidence of internal organisation of the output of a microzone (Bengtsson and Jorntell in Apps, Hawkes et al. 2018 p. 663). A Purkinje cell makes contact on an average of 4-5 nuclear cells, and a nuclear cell receives contact from a random sample of around 30-50 Purkinje cells (Person and Raman 2012a). Moreover each nuclear cell receives contact from a different and variable number of Purkinje cells. On the face of it, independent Purkinje cell rates would simultaneously drive different nuclear rates with no obvious rationale. (5) What is the purpose of the functional organisation of Purkinje cells and of nuclear cells into groups?

We propose possible answers in outline, in the short form of a physiological argument with computational support.

6.5.2 A functional argument for convergence

Extracellular recordings from Purkinje cells show a phase-dependent increase and decrease in their firing rate during locomotion (in cats: Armstrong and Edgley 1984, Armstrong and Edgley 1988, Edgley and Lidierth 1988). However, individually, Purkinje cells spike erratically during a step cycle (in mice: Sauerbrei, Lubenov et al. 2015). The exact spiking pattern of a single cell can be widely different in each step (Sauerbrei, Lubenov et al. 2015 Figure 2A rasters), but gives a smooth curve averaged across many steps. The given explanation of such variation is that it reflects sensitivity to variables that change in each step – exact weight distribution between feet, joint angles, centre of gravity and so on. However, Purkinje cell firing is variable even under tightly controlled conditions – for example when responding to perfectly sinusoidal head rotation in a VOR paradigm (Guo, Ke et al. 2014).

An alternative (and our) explanation is that the erratic pattern (but smooth averaged rate) of Purkinje cell spiking is the expression of a smoothly-modulated moment-to-moment probability that a Purkinje cell spikes. As noted, each nuclear cell receives contact from 30-50 Purkinje cells (Person and Raman 2012a). Accordingly a nuclear cell receives at any moment the sum of spikes discharged by 30-50 Purkinje cells

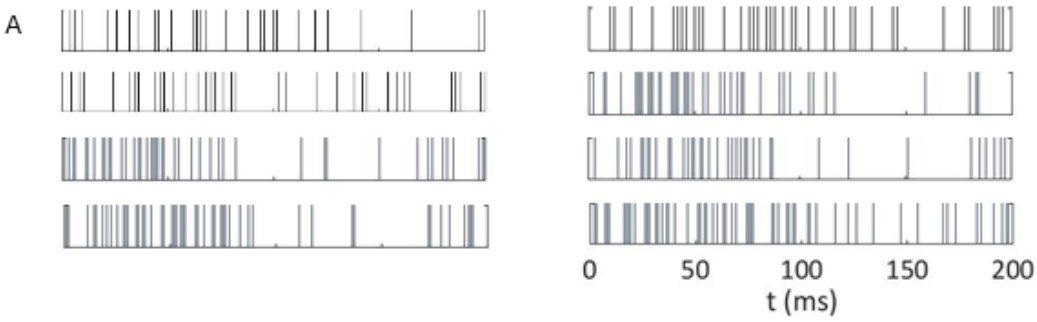
(each making heavy contact, discussed below). The individually erratic spiking pattern of Purkinje cells can be replicated with a changing probability that oscillates sinusoidally between minimum and maximum instantaneous firing rates of 50-200 Hz, mimicking the range in the mouse (Sauerbrei, Lubenov et al. 2015), such that it would give a sinusoidal spike count when averaged over many trials. Accordingly if Purkinje cells afferent to a nuclear cell spike with synchronised time-varying probability, so that at any moment each has an equal, phase-locked probability that it spikes (but with an independent outcome), the sum of spikes counted in 1 ms bins behaves like the average, yielding a smoothed curve across a cycle (Figure 6.10). It follows that millisecond synchronisation of the probability across a functional population of Purkinje cells would mean the nuclear group that receive the output of a microzone all receive the same smoothed collective – in that sense 'averaged' – rate.

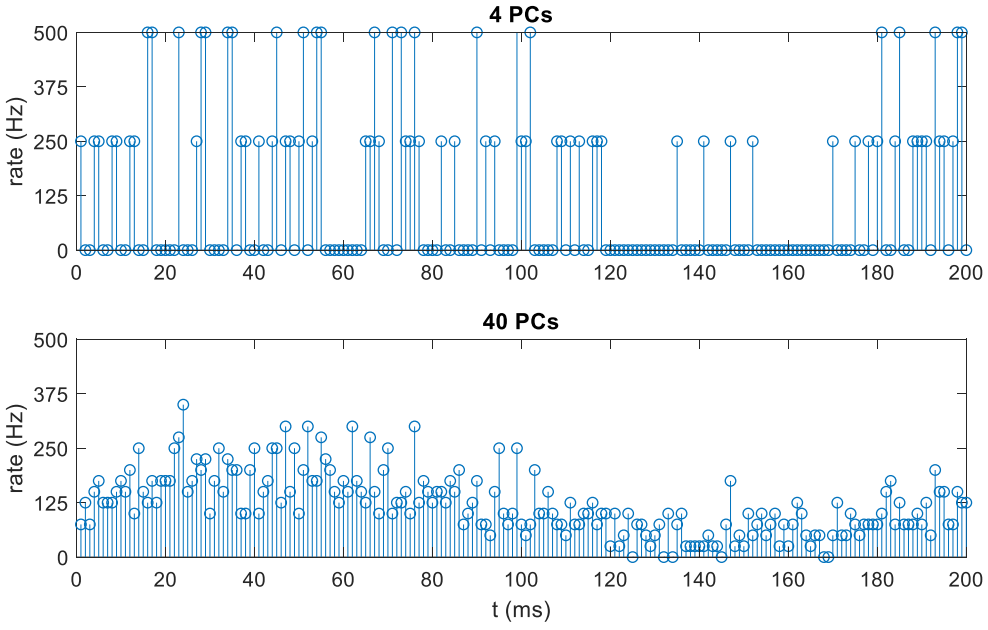
Averaging alone (at these numbers) means the influence of chance on the moment-to-moment total is insufficient to eliminate random variability. However, integration, even with a short window, traps variation within a modest and reliable range (Figure 6.10). It is still not perfectly smooth; we propose that nuclear interneurons make a further contribution, discussed later.

An important feature is that averaging (combined with integration) gives a smoothed curve even at low afferent rates. The sometimes lengthy full pause seen in individual Purkinje cells disappears from the collective count, so that Purkinje cell control of

nuclear cells can be continuous, permitting smooth control despite long interspike intervals in any one Purkinje cell.

FIGURE 6.10





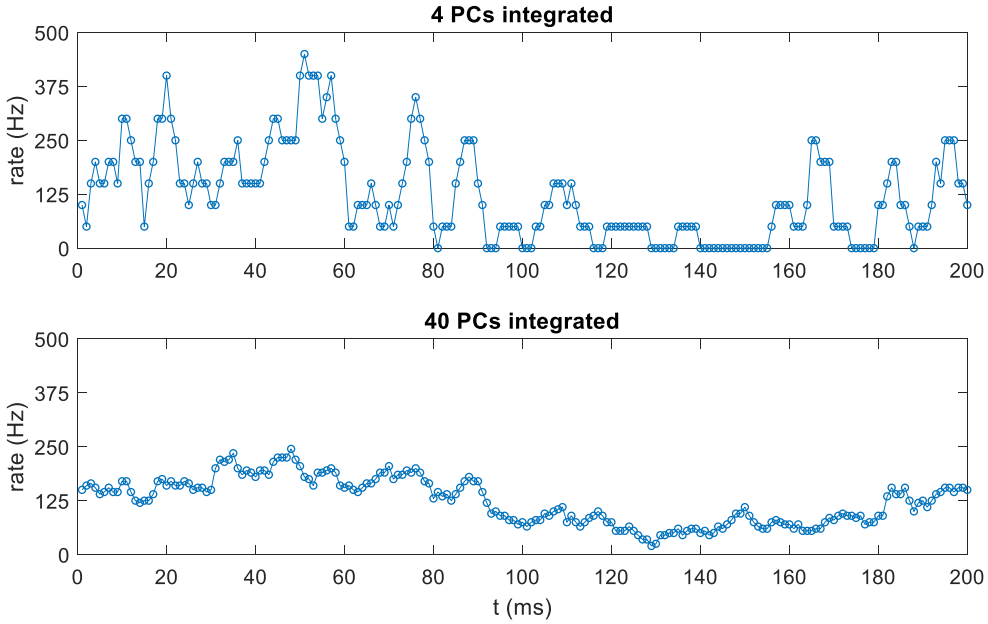


FIGURE 6.10

A: Simulated spiking by 8 Purkinje cells across a mouse step cycle. Each Purkinje cell fires spikes with a changing probability that oscillates sinusoidally between minimum and maximum extrapolated firing rates of 50-200 Hz, mimicking the range in the mouse, such that it would give a sinusoidal spike count when averaged over many trials. In the landscape panels spikes are generated with the same changing probability with time, and counted in 1 ms bins, and plotted as 'instantaneous' firing rates (the average rate generated by the probability of a spike at that point in the cycle) across a single step cycle. **1st and 3rd panels:** convergence of 4 Purkinje cells onto a nuclear cell; **2nd and 4th panels:** convergence of 40 Purkinje cells. In the 3rd and 4th panels the firing rate reflects a 5 ms rolling integration window. In the second panel (high convergence without integration) the rate still fluctuates rapidly in large steps, even in consecutive bins. The combination of 40:1 convergence plus 5 ms integration in the

4th panel reduces moment-to-moment variation to a modest range predicted by the probability of Purkinje cell spiking.

For high resolution rate coding spike-triggered somatic depolarisation – contact by Purkinje cells on nuclear cells is preferentially at the soma (Uusisaari and De Schutter 2011) – must have a short time constant. This is so that the response is limited to an effect of only recent spikes. There must also be a short integration window, partly for the same reason and partly so that timely spikes drive a temporally-specific effect.

A problem with short time constants is that an equally short interspike interval is sufficient to mean they don't sum. Even with a shorter interval, depolarisation is not smoothly varying but a series of peaks. Purkinje cell simple spike rates in the mouse vary in the range 0-250 Hz across the step cycle (Sauerbrei, Lubenov et al. 2015). So even at the highest rate, an integration window of say 3 ms – estimated to be optimal for floccular control of eye movement (Payne, French et al. 2019) – would contain either one peak or none. This would mean that even a high afferent rate is converted into a time-varying somatic signal. Put another way, rate coding is lost in translation even at a constant afferent rate.

An ‘averaged’ rate provided in real time by convergence presents rate coded information as a continuously-varying signal so that nuclear cells can have a short integration window without any of these problems. As far as we know coordinated firing probabilities for functionally grouped Purkinje cells has not been reported. But an averaged rate (approximated by binning spikes across the population of recorded cells) has been shown to have a linear, rapidly-translated relationship with eye movement (Payne, French et al. 2019).

6.5.3 Randomised simple spike timing is an intended consequence of circuit wiring

This appears superficially similar to a strategy of improving the signal to noise ratio. Generally, neurons respond over multiple trials with different spike timing to dynamic stimulation (Mainen and Sejnowski 1995, de Ruyter van Steveninck, Lewen et al. 1997, Schreiber, Fellous et al. 2004). ‘To what extent this neural variability contributes to meaningful processing (as opposed to being meaningless noise) is the fundamental question of neuronal coding’ (Faisal, Selen et al. 2008 p.293).

The functional and necessary asynchrony of spike timing within a Purkinje cell group is not simply a failure to synchronise, we suggest. Convergence of erratically spiking Purkinje cells is the *solution*, rather than a problem, and brought about by design.

This may be a function of stellate cells. Inhibition by stellate cells causes regular intrinsic firing of Purkinje cells to become erratic (Häusser and Clark 1997, Jelitai, Puggioni et al. 2016), and selective silencing of molecular layer interneurons causes regularity to increase (and leads to locomotor deficits: Jelitai, Puggioni et al. 2016). Stellate cells themselves fire erratically in vitro (Häusser and Clark 1997, Carter and Regehr 2002, Ruigrok, Hensbroek et al. 2011) and even more so in vivo (Jörntell and Ekerot 2002, Barmack and Yakhnitsa 2008). Stellate cells form planar networks that lie between Purkinje cells (Palay and Chan-Palay 1974, Sultan and Bower 1998). Networked cells make only a few synaptic contacts on each other (Trigo, Sakaba et al. 2012) at the soma (Lemkey-Johnston and Larramendi 1968).

Purkinje cells also form planar networks. Purkinje cells inhibit each other through recurrent collaterals (Chan-Palay 1971) which extend sagittally, usually in both directions, forming an arbour which is severely flattened in the sagittal plane, in other words aligned along the microzone. Contact on Purkinje cells is typically on near neighbours (Witter, Rudolph et al. 2016). Each Purkinje cell receives contact from an estimated 5 to 10 other Purkinje cells (Witter, Rudolph et al. 2016), mainly at the soma (Palay and Chan-Palay 1974). The similarities with stellate cell networks – inhibitory, planar networks and contact at the soma from a low number of other cells – suggest that the recurrent inputs may contribute to irregular Purkinje cell firing.

6.5.4 Functional organisation into cell groups: a reason for microzones and nuclear groups

There is a general consensus that the cerebellum has modular organisation that may be at the level of microzones (Apps, Hawkes et al. 2018). Convergence of Purkinje cells onto nuclear cells provides a reason for Purkinje cells to be the organised into functional groups. Divergence increases effective convergence without needing more Purkinje cells, so is energetically efficient and saves space. But as there must be at least enough nuclear cells to diverge onto, it is also a reason for the organisation of nuclear cells, too, into groups.

Functional divergence (and therefore minimum nuclear group size) may be increased through excitatory nuclear interneurons. Deep nuclei contain excitatory (presumed glutamatergic) interneurons (Uusisaari and Knopfel 2008). They are intrinsically active but normally held under inhibitory restraint by Purkinje cells from which they are released in the conditioned response, presumably exciting nuclear projection neurons. Accordingly, a Purkinje cell may by this route contribute to control, and the conditioned response, of more nuclear cells than only the ones it contacts directly.

6.5.5 Nuclear interneurons enhance averaging without a longer integration window

Why don't Purkinje cells simply diverge onto all nuclear cells in a group? Why interpose interneurons? The individually strong (Person and Raman 2012b) predominantly somatic (Uusisaari and De Schutter 2011) contact of a Purkinje cell on a nuclear cell may impose a limit on the number of synapses for which space is available (at least for significant contact). Each Purkinje cell contributes on average around 30 boutons per nuclear cell (range 24-36) (Person and Raman 2012a), most containing multiple synaptic densities (Telgkamp, Padgett et al. 2004).

There may be in addition a functional reason. We note earlier that smoothing of the averaged rate of inhibition by convergence and a short integration window is imperfect. If smoothness is a function of convergence, why isn't the convergence ratio higher? In fact, simulating higher than physiological convergence has a strongly diminishing effect on the averaged rate curve. Doubling and even trebling the physiological average of around 40:1 has little further effect. The reason is that the signal to noise ratio increases as the square root of the number of averaged data. In other words, the improvement falls off as the convergence ratio increases. As a result there is little to be gained from higher convergence.

FIGURE 6.11

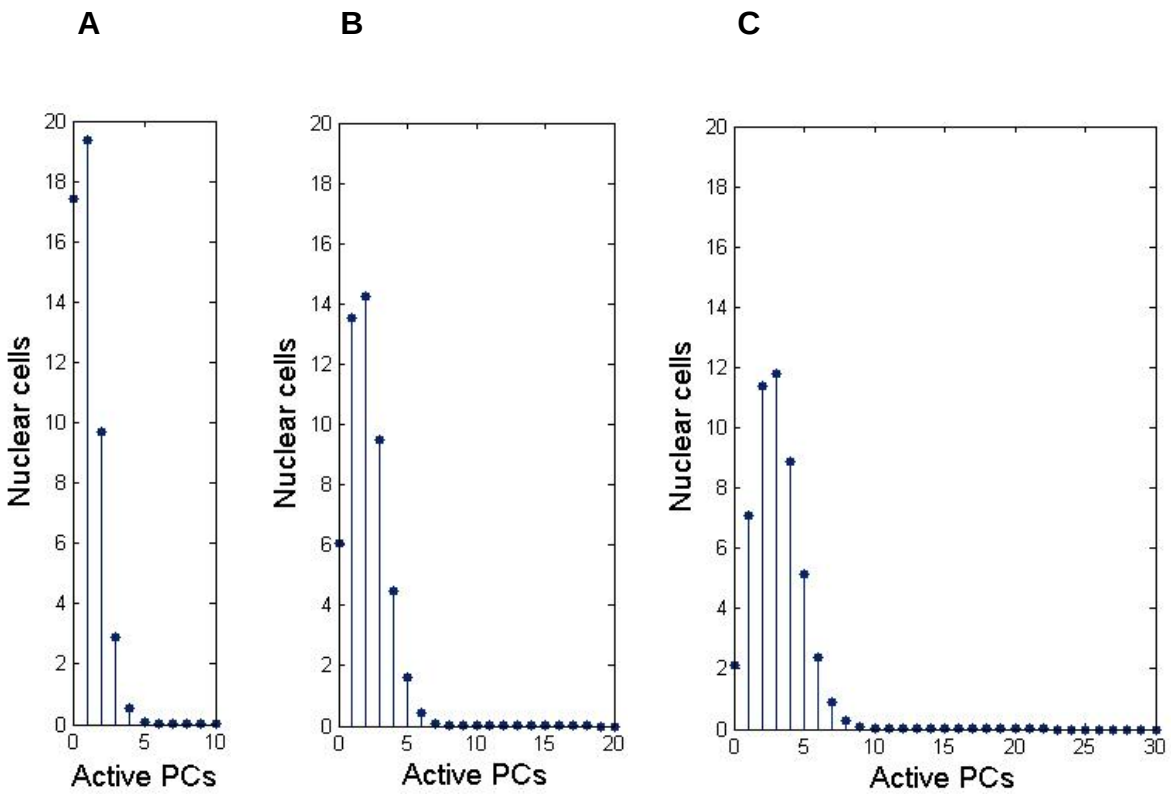


FIGURE 6.11

Stem diagrams showing the expected number of nuclear cells receiving contact from 0 Purkinje cells, from 1, from 2 and so on, out of a subset of 10 (A), 20 (B) and 30 (C), respectively, that are randomly selected from the afferent population (of hundreds) that converge onto a nuclear group. With 10 Purkinje cells, a probable ~35% of nuclear cells receive no contact at all, suggesting that 10 Purkinje cells may not be enough to veto output of a nuclear group (without help from nuclear interneurons). In B (20 Purkinje cells) it is likely that all but a small number of nuclear cells receive contact from at least 1 Purkinje cell and the majority receive input from at least 2. Purkinje cells make individually strong contact on each of their targets, so that is sufficient that firing of only a fraction of the population of a microzone is not coordinated with the rest for a

high probability that a nuclear cell is strongly affected, and for all nuclear cells to be at the same high risk. With a subset of 30 Purkinje cells (C) only 2 nuclear cells receive no contact and over 80% receive contact from 2 or more. The data do not show higher functional divergence through interneurons, which should mean fewer Purkinje cells are necessary for a similarly strong impact.

However, a higher bisynaptic divergence ratio through interneurons duly increases the bisynaptic convergence ratio, other things being equal. Bisynaptic contact adds a second layer of integration – Purkinje cells onto interneurons and interneurons onto nuclear cells – with concurrent (as opposed to sequential) integration windows. This adds smoothness (in simulations) without needing a longer integration period. A smooth averaged rate has the additional benefit of driving out random variability of the rate curve.

6.5.6 Evidence of coordinated firing of functionally-grouped Purkinje cells

Is there evidence of synchronisation of spike probability (but not spike timing) at microzone level? Sets of Purkinje cells located in the same sagittal plane show coordinated simple spike pauses (Ramirez and Stell 2016) which generate motor output (Heiney, Kim et al. 2014, Lee, Mathews et al. 2015). Synchronisation of spike timing during behaviour is reported to be confined to paired recordings from Purkinje

cells that are on-beam, and to be absent in recordings from paired Purkinje cells that are off-beam (Heck, Thach et al. 2007), so spike timing is not synchronised sagittally. Modulation of nuclear cell firing requires the co-modulation of a large proportion of Purkinje cells that innervate the cell (Bengtsson, Ekerot et al. 2011). That is, a few Purkinje cells whose firing is not modulated can block an effect of the rest, consistent with strong individual contact by a single Purkinje cells of each of its targets (Telgkamp, Padgett et al. 2004, Person and Raman 2012a, Person and Raman 2012b). This would suggest that coordination of Purkinje cell firing across the whole population of a microzone is necessary for graded modulation of nuclear firing, because less than full coordination means that each nuclear cell is at high risk of strong inhibition from an out-of-step cell (Figure 6.11).

Synchronisation of spike probability among functionally-grouped Purkinje cells would neatly explain random sampling by nuclear cells of Purkinje cell rates. It is functionally immaterial which Purkinje cells contact a particular nuclear cell, and that nuclear cells each receive a different sample. This would provide a rationale for functional organization of Purkinje cells that operates at microzone level, and not at the level of individual Purkinje cells. Clearly, this is in conflict with the expectation that that ‘individual Purkinje cells most probably require specific error signals and learn heterogenously’ (Zang and De Schutter 2019 p.3), anticipated by learning models (Albus 1971, Fujita 1982).

Possibly, circuit design on these principles is a driver for modular organisation at microzone level, so that circuit design can be (partly) explained as a mechanism of turning rate coded sensory control of motor output into finely incremental modulation of nuclear firing in a short integration window, at all afferent rates. This enables Purkinje cell-mediated inhibitory postsynaptic currents to have short time constants (necessary for fine resolution) without sacrificing a continuously time-varying effect on the postsynaptic cell.

6.5.7 A possible mechanism of Purkinje cell synchrony

There are several moot but possible mechanisms of coordination. One possibility is to configure circuits so that Purkinje cells in a microzone all receive at any moment what is functionally speaking the same parallel fibre input, meaning the same randomly distributed frequency distribution of parallel fibre rates from the same number of active cells. We argue earlier that granule cell rates are confined to a narrow range with a fixed bandwidth (and other variables are controlled) such that control of firing of a Purkinje cell is by the only remaining functional variable, the adjustable range. How does the cerebellum contrive that the whole Purkinje cell population of a microzone receives, at any moment, the same range?

The range varies with the mean of mossy fibre rates received by the underlying – therefore mediolateral – strip of the granular layer (argued earlier). Mossy fibre

terminal branches end in clusters of terminals (Wu, Sugihara et al. 1999, Sultan and Heck 2003, Shinoda and Sugihara 2013) that are aligned in the sagittal direction (Sultan 2001). Coupled with high divergence onto granule cells, this has the result of randomly intermingling mossy fibre rates in the sagittal direction. If rates are perfectly intermingled, granule cells in a sagittal band all randomly sample the same pool of rates. A mediolateral strip of granule cells therefore receives a mixture of rates approaching a representative sample of each band it crosses, and two strips that cross the same sagittal bands contain an identical row of samples (if we assume samples are representative of the sampled band). Therefore, if all strips that supply parallel fibres to a microzone cross the same pattern of sagittal bands, they all, at any moment, contain an identical mediolateral row of such samples (like a row of squares on a chess board but without defined or fixed boundaries).

Parallel fibre activity in each strip is therefore driven, moment-by-moment, by functionally identical mossy fibre rates (and the variables that differ between samples – the number and binary pattern of active mossy fibres and permutation of rates they each fire at – are not material *to this function*). In reality, identical is probably an overstatement because it is unlikely that rates in a band are perfectly intermingled, or therefore that an intersection is fully representative of the band as a whole, but it may nonetheless be one of the strategies that contributes to coordination of granule cell rates (and this is a proposal of the model summarised here in outline).

6.5.8 Purkinje cell averaged rates → nuclear rates

As noted, the conditioned fall in Purkinje cell rates may in some circuits gate mossy fibre collateral input to nuclear cells, but also exerts control of its own over nuclear rates, which may in some circuits dominate control (Ishikawa, Tomatsu et al. 2014, Jirenhed and Hesslow 2016). Nuclear cells are insensitive to steady afferent rates (see also Turecek, Jackman et al. 2017), but 'respond to increasing, as well as decreasing, changes in PC [Purkinje cell] firing rate with immediate modification of their output firing' (Pedroarena and Schwarz 2003 p.713). There are a number of adaptations that make synaptic transmission fast and reliable. These include 'large boutons, glial ensheathment, GABA transporters confined to astrocytes, [and] multiple release sites' (Telgkamp, Padgett et al. 2004 p.123). Spillover confinement to multisynaptic Purkinje cell boutons, with transporters confined to astrocytic processes at the bouton perimeter, procures spillover-mediated transmission reported to have a high response probability of postsynaptic receptors (Telgkamp, Padgett et al. 2004). Thus spillover-mediated transmission mitigates unpredictable single synapse neurotransmitter release, so transmission is reliable and precisely timed as well as fast.

The amount of the increase or decrease in the nuclear rate is likely to be proportionate to the rate of change of the simple spike rate (assuming that firing amongst a population of Purkinje cells that contact a nuclear cell is asynchronous:

Indira Raman, personal correspondence dated 4 December 2018).[21] Averaged Purkinje cell rates across a step cycle describe a repeating wave form (Sauerbrei, Lubenov et al. 2015). A downslope, where the averaged rate is falling, elevates nuclear rates, and vice versa. What controls the rate of change of the averaged rate in this phase? (The next two paragraphs recap some work referenced earlier.)

MLIs reflect 'granule cell input with linear changes in firing rate' (Jelitai, Puggioni et al. 2016 p.6). The effect of inhibition by MLIs of Purkinje cell firing is rate dependent and linear, such 'that locomotion-dependent modulation of the balance between excitation and inhibition [of Purkinje cell dendrites] generates depolarising or hyperpolarising dendritic V_m [dendritic membrane voltage] changes that linearly transform into bidirectional modulation of PC SSp [Purkinje cell simple spike] output' (Jelitai, Puggioni et al. 2016 p.9). This is a challenge for learning models, which propose that output (following training) is learned, that is, training displaces the naïve response to input rates by making iterative adjustment of parallel fibre synaptic weights (Albus 1971, Fujita 1982, Brunel, Hakim et al. 2004, Dean, Porrill et al. 2010). If instead translation is linear as reported, and control is mediated by MLIs, what controls MLI rates?

The higher concentration of synaptic contact by stellate cells on Purkinje cells at greater depth (to reason from morphology) has the (proposed) result that there is

[21] Professor Raman considers firing may not be asynchronous, but that in theory perfectly asynchronous firing would probably have this relationship with nuclear rates.

depth-dependent selection of parallel fibre signals that control the simple spike rate in the conditioned response, which deeper-stratum parallel fibres dominate. Mossy fibre input to the cerebellum is vertically topographically organised (reported for the C3 region: Jörntell and Ekerot 2006, Quy, Fujita et al. 2011). Deeper-lying granule cells receive signals triggered by limb movements (Jörntell and Ekerot 2006), suggesting it is driven by activation of internal receptors in muscles and joints: proprioception. Since topography is preserved in the molecular layer (Palay and Chan-Palay 1974, Zhang and Linden 2012), control of MLIs that dominate control of Purkinje cells is by mossy fibre signals from this source.

The control of Purkinje cell firing by proprioception, and linear translation of firing rates, argue that output (of C3 circuits) has a linear, rapidly-translated relationship with movement, where input to a circuit is provided by movement to which the circuit contributes control (provided input is proportionate to movement parameters). Moreover timing is phase-locked, so also effectively movement selected.

To return to the question raised earlier regarding ‘interference’ from lateral inhibition by MLIs, coordination of Purkinje cell firing means it does not matter which stellate cells they receive inhibition from provided the size (area in the sagittal plane) of the afferent network of stellate cells that contact a Purkinje cell is uniform. Indeed an area increased by the range of lateral inhibition should mean the network receives a more representative sample of parallel fibre rates.

6.6 Somatotopy

What is the evidence that the body location that provides the bulk of the input to a circuit receives the output of the circuit?

A repeated finding of mapping studies is that cerebellar circuits have a complex relationship with the coordinates they represent. Mapping mossy fibre input has long been reported to generate a patchy mosaic i.e. discontinuous representation of body parts (but somatotopy within the shards: the 'fractured somatotopy' of Shambes, Gibson et al. 1978). One reason is that circuits receive (as mossy fibre input) information from more than one source. In the C3 region ‘each microzone receives input from several mossy fibre receptive fields’ (Apps and Garwicz 2005 p.305). Also, it is common for signals originating from the same external source to be sent to more than one circuit (for mossy fibres: Voogd, Pardoe et al. 2003, Pijpers, Apps et al. 2006, for climbing fibres: Fujita and Sugihara 2013).

Mapping by output (stimulating sites in the cerebellum to see which muscles contract) does not yield a more orderly topographical organisation than the fractured input map (stimulating the body surface or muscle and mapping the response in the cerebellum). An example in humans has been provided by electrical stimulation (60 Hz for 2 seconds) of the posterior cerebellum in patients in surgery for tumours (Mottolese, Richard et al. 2013). The evoked movements are strictly ipsilateral but

otherwise ‘the same body part can be represented in different sectors, different body parts can be embodied in a single region and the [cerebellar] cortical size of a body part is proportional to its functional importance rather than its physical size’ (ibid. p.337).

Nonetheless there is order. The response of climbing fibres to somatosensory stimuli has allowed microzones to be somatotopically mapped i.e. matched with the receptive field (region of the body surface) to which they are responsive (Garwicz, Ekerot et al. 1998, Garwicz, Jörntell et al. 1998). In cats, ‘climbing fibres in adjacent microzones are activated from adjacent skin areas, forming a detailed somatotopic map of the ipsilateral forelimb skin, particularly in distal parts’ (Ekerot and Jörntell 2003 p.101). Mossy fibre input representing stimulation of the body surface is normally sent to the same circuit as the climbing fibre signal (evoked by the same stimultation) (Garwicz, Jörntell et al. 1998, Voogd, Pardoe et al. 2003, Odeh, Ackerley et al. 2005, Pijpers, Apps et al. 2006, Apps and Hawkes 2009). Evidence is inconclusive but indicates that regions of the cerebellar cortex which map to particular body parts receive input from regions of the pons which in turn receive input from regions of the cerebral cortex which map to the same body parts (Odeh, Ackerley et al. 2005). A study of the relationship of the receptive fields of microzones in the C1, C3 and Y cerebellar cortical zones and movements controlled by anterior interpositus sites to which output of those microzones had been mapped, showed ‘a general specificity of the input-output relationship’ (Ekerot, Jorntell et al. 1995 p.365).

‘Perhaps nowhere else in the history of ideas has there been a more striking pattern of reliance on metaphors than in the history of reflection about the brain’ (Daugman 1993 p.23). Mechanical analogies for the brain have been with us for centuries. Other analogies include hydraulics, electronics and telegraph metaphors (Hodgkin and Huxley borrowed the mathematics developed for signal propagation in coaxial cables to model the generation of action potentials), and thermodynamics. The computer metaphor, and network models imported from statistics, have been so widely adopted the lines have become blurred between language used by analogy and proposals that are intended to be literal (see Werner 2011 for a review). References to computation, for example, are used in a way that suggests there is thought to be no need for them to be qualified.

There is a widespread and well-funded belief that artifical intelligence may provide insight into brain function by analogy. In the cerebellum, it has become a presumption that the cerebellum implements an algorithm that converts input values (by which input signals are represented) into output, and a presumption of learning models that output is a learned response to a remembered set of input values. This may be an entrenched mistake to which this thesis is an alternative.

6.7 Experimental tests

There are issues testing network models generally because even the best

experimental methods of measuring activity don't allow individually accurate spatial triangulation of defined populations of active cells, or simultaneously-recorded, precisely-timed measurements of events on a biological timescale. An example is the finding that 2/3 of granule cells may be active simultaneously (Giovannucci, Badura et al. 2017), which would clearly cause problems for models which predict that granule cell coding is sparse.

Another issue for the present proposal is that there is (in the present proposal) no overarching model that provides a quantifiable description of the function of the cerebellum. This is because there is no physiological equivalent of an overarching mechanism. There are, rather, solutions for different local problems around the circuit (possibly reflecting the piecemeal evolution of the circuit and/or its adaptation to different functions) whose functions are integrated. Put another way, the proposal is not one model but several. Accordingly there is no singular test of a quantified parameter of a central operating principle, because there is no central principle.

However, anatomical tests are available. Questions that might be asked include: How many Golgi cells innervate a cluster field? Do each of the dendrites of a granule cell receive inhibition from a different Golgi cell? Do granule cells make ascending contact on stellate cells and/or on Purkinje cells? Is the amount (or probability) of contact by granule cells higher locally than for distant granule cells? Do excitatory nuclear interneurons contact output-carrying nuclear projection neurons? How many excitatory nuclear interneurons does a Purkinje cell make contact on, and how many

output cells does each of those contact, and how many output cells in total does a Purkinje cell indirectly influence in this way? Is there in fact internal organisation of the output of a microzone, or do Purkinje cells make contact at random within a nuclear group?

Physiological tests include: If granule cell ascending axons make synaptic contact on stellate cells and/or Purkinje cells, can those synapses be trained? If so, in what direction and under what conditions? Is Purkinje cell spiking individually erratic recorded simultaneously from Purkinje cells in the same microzone? Are ‘instantaneous’ Purkinje cell rates coordinated? If so, only in the conditioned response or also outside, at other times? What is the relationship between the Purkinje cell firing rate and nuclear output cell rates by the indirect pathway, mediated by excitatory nuclear interneurons? Is the nuclear output cell rate proportionate to the derivative of a change in the coordinated, afferent, averaged Purkinje cell rate?

APPENDICES

APPENDIX 1

Marr's 'A theory of the cerebellar cortex'

Marr proposed that patterns of active parallel fibres that are repeatedly paired with climbing fibre input to the same Purkinje cell 'at about the same time' induce strengthening of the corresponding pattern of active parallel fibre synapses. Climbing fibres thus provide an instruction signal that teaches facilitation of those synapses. The result is that a Purkinje cell will respond to the subsequent repeat of a learned pattern but not respond to an unlearned pattern, because part of these will be received by unmodified, weak synapses. Thus Purkinje cells learn to respond selectively to granule cell representations of mossy fibre activity.

Marr argues that recoding in the granular layer, of mossy fibre input into granule cell activity, sparsens and separates patterns of internal activity, so that a Purkinje cell can store more patterns than if it received mossy fibre signals directly. Recoding turns the proportion of active mossy fibres into a lower proportion of active granule cells (termed expanding state space). This is important because storage is limited by the requirement that an unlearned pattern does not address a sufficient number of facilitated synapses to initiate a positive response.

Sparseness/separation is facilitated by Golgi cells. Golgi cells occupy compartments, one per compartment (Marr incorrectly thought Purkinje cells outnumbered Golgi cells by an order of magnitude). By sampling direct mossy fibre input to a compartment, and parallel fibre activity that passes through, a Golgi cell 'estimates' overall granule cell activity (in a row of like compartments whose granule cells contribute to parallel fibre activity). Direct mossy fibre input provides an initial estimate revised if necessary by sampling parallel fibre activity. This is used to adjust the rate of Golgi cell inhibition received by granule cells in that compartment in order to adjust the number of mossy fibre inputs needed to make a granule cell fire (a codon), and therefore the number that fire. This keeps the number generally low. Mossy fibre input to basal dendrites and parallel fibre input to apical dendrites are separately summed and the stronger controls the Golgi cell firing rate and therefore local codon size. Separate summation and control by one or the other is not evidenced and the mechanism is not specified.

Marr assumes synapses are either 'totally modified' or silent, and that patterns of parallel fibre activity are spatially random. For a Purkinje cell to fire, the proportion of a learned pattern received at modified synapses must be 'close to 1'. The threshold is a fraction and not a fixed number because, despite the efforts of Golgi cells, activity levels of parallel fibres, though bounded, will *not* be constant. Setting the threshold at a fixed number of signals would mean the necessary proportion varies with pattern size – that is, the threshold would vary depending on the number of active cells in a pattern. Stored patterns overlap – this follows because they are random. Marr envisages an unknown but large proportion of modified synapses.

To meet threshold, the number of active parallel fibres which make contact at modified synapses must exceed the minimum number necessary to prevail over feed-forward inhibition driven by the same pattern of parallel fibre activity. Inhibition is provided by MLIs. The threshold to drive a response is set at the fraction of active parallel fibres that must be received at modified synapses. The necessary number is proportionate to the total number that are active, because a higher total means a higher number drives inhibition. There is no learning at the parallel fibre-MLI synapse. Marr assumes that stellate cells and basket cells, and 'intermediate forms', can be treated as functionally the same. He does not explain how tolerance of this arrangement would affect output.

Marr allows that operation in this way might lead you to expect that each Purkinje cell would be partnered with a single MLI of the same size and shape, so it would sample the same parallel fibre activity and sample enough to be representative. He argues (without evidence) that the actual morphology is more efficient, and that 40 MLIs (the number that contact a Purkinje cell) have a total dendritic field sufficient to sample parallel fibre activity. He gives the reason that mossy fibres give rise to multiple terminals distributed sagittally, so that the mixture of mossy fibre signals that drives parallel fibre input to the MLIs that laterally inhibit a Purkinje cell is much the same as the one that drives parallel fibre input to that Purkinje cell.

Learning is at the level of single Purkinje cells and Purkinje cell firing is treated as output (on the assumption that the CNS has the means of converting a Purkinje cell signal into a motor command). The function of learning is to cue output in response to recoded input patterns (and not to control output rates). Purkinje cell rates are controlled by parallel fibre rates – Marr does not include an effect of synaptic weight on transmission, or allow for variation of the effect on Purkinje cell firing of variation of the number of active parallel fibres in an input pattern. He does not simulate recoding or explain how the compartmentalised codon adjustment combines with the same computation in other compartments for a joint effect on parallel fibre activity. Nor does he suggest how adjustable codon size would impact on transfer of rate (or other) coded information from mossy fibres to granule cells, or explain how un-normalised (modally diverse, eclectically sourced) mossy fibre data are translated into granule cell rates.

Albus's 'Theory of cerebellar function'

Like Marr, Albus (Albus 1971) proposes that patterns of recoded activity are learned under climbing fibre instruction, and stored as parallel fibre-Purkinje cell synaptic modification (and the unit of learning is a Purkinje cell). Albus's central proposal is that the operation of the cerebellum may be analogous to a perceptron algorithm. The perceptron algorithm was originally developed as a neural model of image recognition (Rosenblatt 1961). In Albus's proposal, mossy fibres equate to sensory cells, granule cells provide association cells, Purkinje cells are the response cells,

and the parallel fibre-Purkinje cell synapse provides adjustable weights. The algorithm teaches incremental adjustment of synaptic weights. If the response cell fires when it shouldn't, active synapses are weakened; if it doesn't fire when it should, active synapses are strengthened. Synaptic modification causes incremental and cumulative adjustments to the postsynaptic effect of a signal. Following training, inputs (thus adjusted) to the classifier (the Purkinje cell) are summed and compared to a threshold. The response depends on whether the sum is over or under threshold, and accordingly binary. All patterns are classified in this way into one or other of two classes. All patterns of both classes must be learned. Albus does not include a proposal of a physiological mechanism that implements the algorithm.

The restriction, as Albus sees it, of the application of this idea to neural function is set by limits on the reliability of classification. This is limited in turn by tolerance of overlap of learned patterns (that is, the number that can be learned before learning more causes errors of classification). Albus proposes that the large number of granule cells improves reliability provided the fraction of active granule cells is confined to a low percentage. The low percentage (Albus proposes 1%) means a given synapse participates in fewer patterns. 'The restriction that only 1% of the association cells are allowed to be active for any input pattern means that any association cell participates in only 1% of all classifications'.

Like Marr, Albus proposes that Golgi cells regulate parallel fibre activity by adjusting the number of mossy fibre inputs needed to make a granule cell fire. Albus argues

mathematically that the overall level of activity (i.e., spikes per second) of parallel fibres is nearly constant, but not what the level is, or otherwise how regulation would work physiologically. He proposes 1% for modelling convenience, that is, because learning on a perceptron model would be reliable and learning would take relatively few lessons.

Albus proposes that that a climbing fibre signal is equivalent to the unconditioned stimulus in a conditioning study, the pause in simple spike firing after a complex spike equates to the unconditioned response, a mossy fibre pattern shortly following a complex spike equates to the conditioned stimulus, and therefore that the corresponding pattern of active parallel fibre-Purkinje cell synapses is weakened as a result of training, so that this also drives a simple spike pause, the conditioned response. Firing of a Purkinje cell in the learning window – the climbing fibre induced simple spike pause – is, in this proposal, an error signal. This equates to the perceptron training algorithm because if a response cell/Purkinje cell fires when it should not fire, active parallel fibre synapses are weakened; if it does not fire when it should not do so, no adjustment is made.

As tuition trains only weight loss at the parallel fibre-Purkinje cell synapse, weight gain needs a different mechanism. This, or the equivalent, is provided by plasticity at the parallel fibre-interneuron synapse, Albus argues, subject to the requirement that learning is in different directions at synapses on superficial and deeper interneurons. Synapses on superficial stellate cells, which inhibit Purkinje cells only on beam, are

strengthened. This facilitates conditioned depression of the simple spike rate because a known pattern addresses strengthened synapses, driving feed-forward inhibition. At deeper level, at synapses on deeper stellate cell and basket cells, training weakens transmission. Weakening provides bidirectional control of Purkinje cells rates because input to interneurons in a learned pattern causes a phasic reduction of inhibition of their targets – the equivalent of bi-directional parallel fibre-Purkinje cell plasticity. The model does not explain how parallel fibre-interneuron synaptic adjustment is confined only to the 'immediate vicinity of the Purkinje cell' under instruction, or how instruction teaches weight change in different directions at superficial and deeper level.

Albus does not propose how climbing fibre signals are controlled – how does an instruction signal know what to teach? Nor is it clear how a binary response is to be reconciled with physiological data showing that simple spike firing rates encode dynamic and analog features of movements.

APPENDIX 2

Anatomical reasoning that contact on each of the dendrites of a granule cell is likely to be from a different mossy fibre

Mossy fibres terminate as clusters of terminals. An estimated 100 mossy fibres terminate in a region the size of cluster field. A mossy fibre gives rise to an estimated 7-8 terminals per cluster (Sultan and Heck 2003). Terminals intermingle. Granule cells have a small number of non-branching, short dendrites (15 μm, soma diameter 5-8 μm, so that a granule cell fits comfortably inside a cluster field). If we assume that each dendrite of a granule cell has an equal chance of contact with each of 7 x 100 mossy fibre terminals, the probability of an n^{th} contact with the same mossy fibre is:

$$P_n = \left[\prod_{x=2}^{n}\left(\frac{m-x+1}{100m-x+1}\right)\right] \times \frac{y!}{x!\,(y-x)!}$$

where *m* is the number of terminals per cluster and *y* is the number of granule cell dendrites. The probability that a granule cell receives contact to each of its dendrites from a different mossy fibre is thus 1 minus the sum of the probabilities that it receives contact to any two or more dendrites from the same mossy fibre, or ~0.947. By this calculation around 95% of granule cells receive contact to each of their

dendrites from a different mossy fibre, (and between 1 and 2 in a thousand receive contact from the same mossy fibre to 3 dendrites).

APPENDIX 3

The area of the granular layer that drives input to a microzone is larger than a microzone by an order of magnitude

In physiological conditions the amount of change to a parallel fibre pattern is not the result of changing input only to a single cluster field but of changing input on beam for 3 mm in both directions. A parallel fibre span of 6 mm is reported for cats, chickens and monkeys (Brand, Dahl et al. 1976, Mugnaini 1983); 5 mm is reported in rats (Harvey and Napper 1988, Harvey and Napper 1991). Assuming a microzone length of 15 mm (an estimate for the C3 region in cats: Dean, Porrill et al. 2010), the area of the granular layer receiving mossy fibre input that may contribute to a volley of parallel fibre input to a microzone is 15 x 6 = 90 mm^2. The area of a microzone in the sagittal plane, which receives the sum of recoding in that region, is 0.3 x 15 = 4.5 mm^2, a ratio of 20:1.

APPENDIX 4

Convergent estimates of the number of granule cells in a cluster field

There are an estimated 1.92×10^6 granule cells per µl (in rats: Harvey and Napper 1988). The number of cluster fields which fit in that volume is $(1{,}000/150)^2 \times (1{,}000/200) = \sim 222$, assuming cluster field dimensions of 150 x 150 x 200 µm (Sultan and Heck 2003). The number of granule cells in a cluster field is therefore 1,920,000/222 = 8,649. This is a good match with the estimate obtained in the text: if 100 mossy fibres each contribute 7 terminals to a cluster field (Sultan and Heck 2003) so that a cluster field contains 700 terminals, then the number of granule cells should be around 8,750 (700 terminals multiplied by 50 dendrites per terminal divided by 4 dendrites per granule cell).

APPENDIX 5

Excitatory and inhibitory input to a glomerulus vary independently

Timing theories are based on the idea that direct mossy fibre input to Golgi cells excites feed-forward inhibition of mossy fibre/granule cell transmission, so that a mossy fibre provides precision timing of its own transmission (by restricting it to a narrow time window). ‘Temporally precise inhibitory input [by Golgi cells to granule cells] that narrows the window for the temporal summation of discrete mossy fiber inputs...forms the basis of a variety of contemporary cerebellar models’ (Duguid, Branco et al. 2015 p.13099). This idea is under fire (ibid) and always suffered from a number of problems. For example, probably less than one in ten mossy fibre terminals make contact on a Golgi cell (Hámori 1992), so that more than 90% of mossy fibre terminals provide no direct drive to Golgi cells, and those that do are not individually sufficient for an effect (because an estimated minimum of 4 mossy fibres are necessary to modulate firing: Kanichay and Silver 2008). Also, a Golgi cell receives contact from an estimated 40 mossy fibres (Kanichay and Silver 2008), a fraction of the number that terminate in its dendritic field.

It is not known what proportion of the granule cells within the axonal field of a Golgi cell are inhibited by that cell. It was at one time thought all of them (Eccles, Ito et al. 1967, Marr 1969, Albus 1971, Albus 1989), with ‘relatively little if any

convergence...[by Golgi cells on] glomeruli' (Eccles, Ito et al. 1967 p.27). However, this is difficult to reconcile with other evidence. The ratio of Purkinje cells to Golgi cells has been estimated at 1:1.5 in man, 1:1.9 in cats and 1:3.3 in rats (Lange 1974). It may be as low as 1:1 (Ito 1984, Llinás and Negrello 2015). The axonal range of a Golgi cell is 650 +/- 179 µm in the direction of the long axis of a microzone by 180 +/- 40 µm in the mediolateral direction (in mice: Barmack and Yakhnitsa 2008). So that even using the most conservative (1:1) estimated ratio of Golgi cell to Purkinje cell numbers, the density of Golgi cell soma must mean that axonal fields overlap (because the axonal range is longer in both directions – especially sagittally – than the distance between cell bodies).

In theory this means that a granule cell could receive up to four rates of Golgi cell inhibition, one to each dendrite. It is thought that glomerulus probably receives input from a single Golgi cell (Eccles, Ito et al. 1967). It 'can be assumed that only one Golgi cell axon enters a glomerulus' (D'Angelo, Solinas et al. 2013 p.10). However, this is unconfirmed and has been criticised on the grounds that it 'seems to stem from the ultrastructural study of the glomerulus made by (Jakab and Hamori, 1988) [who estimated 1-2] Golgi cell axons per glomerulus...[but] investigated only 2 glomeruli' (Bengtsson, Geborek et al. 2013 p.2 in pdf online).

So a mossy fibre may make no contact on *any* of the Golgi cells which inhibit its terminals. Even where it does, and a feed-forward loop exists, a single mossy fibre is insufficient for an effect on Golgi cell rates (Kanichay and Silver 2008), so that in the

event of an effect there would be interference from (a variable number of) other excitatory inputs (minimum 3, but not improbably more).

APPENDIX 6

Higher probability of contact on a Golgi cell by local than by distant granule cells

Assuming a sagittal span of 300 µm for the apical dendrites of a Golgi cell (Eccles, Ito et al. 1967), an estimated 350,000 parallel fibres pass through a Golgi cell territory of which ~1,200 that are not local (D'Angelo, Solinas et al. 2013) make contact (1 in ~300, on apical dendrites). In contrast ~1 in 15 local granule cells make contact (about 800, on basal and apical dendrites (Cesana, Pietrajtis et al. 2013)) out of, at a rough approximation, 12,000 in a 150 x 300 µm region of the granular layer).

APPENDIX 7

Probability distribution for input to a stellate cell assuming 500 parallel fibres make contact

In the main text we use an estimate of 1,000 for the number of parallel fibres that make contact on a stellate cell. If we had used a lower estimate of the number of parallel fibres which contact a stellate cell – 1,000 is at the high end of the range of estimates – the probable number of inputs would be lower. For example, if we assume that a stellate cell receives contact from 500 parallel fibres, we get

k	0	1	2	3	4	5	6	7	8	9	10
p	.0813	.2049	.2573	.2147	.1339	.0666	.0275	.0097	.0029	.0008	.0002

where k is the number of inputs and p is the probability of that number, giving a weighted average for *p* of 2.5416. So if 500 make contact the most likely number of inputs (the number out of 500 that are active) is 2, and p = ~0.85 that the number is between 1 and 4.

APPENDIX 8

Examples from the literature

It has been reported that stellate cells are sensitive to excitatory inputs, even in low numbers. We take some examples and in each case show that the appearance of that effect is a result of the (entirely rigorous and perfectly sound) procedures that were used to collect the data.

The examples are:

1. ‘Small numbers of coincident excitatory quanta reliably and rapidly’ trigger firing (Carter and Regehr 2002 p.1309).

   The effect of excitatory quanta was measured with $GABA_A$ receptors blocked by an antagonist. Unblocked inhibitory input was found, by contrast, to strongly suppress firing, entirely blocking it, indicating that ‘inhibitory quanta tightly control the influence of excitatory quanta on stellate cell firing’ (Carter and Regehr 2002 p.1313).

2. Firing is driven by 2-8 EPSPs in C3 stellate cells (Jörntell and Ekerot 2003).

   Data were obtained from stellate cells in the C3 region of the cerebellar cortex. These were selected because they responded to cutaneous stimulation.

Cutaneous signals are received by granule cells that preferentially divide, and contact stellate cells, at outer level (Jörntell and Ekerot 2006, Zhang and Linden 2012). Stellate cells at outer level are not functionally networked (we have argued), so lack tonic inhibition.

3. Distance-dependent attenuation of passively-transmitted dendritic signals is relatively modest, ~50% at a distance of ~50 µm (Abrahamsson, Cathala et al. 2012).

   Measurements were made with $GABA_A$, NMDA and glycine receptors all blocked (Abrahamsson, Cathala et al. 2012, p.1170) so that, like (1) and (2), tonic inhibition of stellate cells is absent.

4. A pharmacological block of ionotropic glutamate receptors causes a (modest but clear) drop in the stellate cell firing rate in slices even with most parallel fibre input probably removed by slice preparation (Häusser and Clark 1997).

   This was thought to be possibly the result of blocking what was left of background, spontaneous excitatory input (relevance: if we are correct there should be no effect of background synaptic input). The reduction of the stellate cell firing rate may have been the result of blocking an effect of ambient extracellular glutamate acting through ionotropic glutamate receptors that stellate cells have since been reported to express extrasynaptically (Szapiro and Barbour 2007), as opposed to blocking background synaptic input.

APPENDIX 9

Giovannucci et al 2017

It was reported in 2017 that a seemingly far larger proportion of granule cells is active during learned movements than predicted by theory (Giovannucci, Badura et al. 2017), using a blink reflex conditioning protocol. An implication is that coding is not sparse, counter to the prediction of traditional theories (Marr 1969, Albus 1971) and also counter to our prediction of the regulated density of parallel fibre activity [ref]. It was estimated that as many as 2/3 of sampled granule cells may be active, when measured over the course of a whole action (but still on a sub-second time scale (Giovannucci, Badura et al. 2017).

On the face of it, the results are difficult to reconcile with the idea that the cerebellum stores a memory of patterns because even a very low number of such very dense patterns would make pattern discrimination unworkable. Assuming active parallel fibres are randomly distributed at the scale of input to a Purkinje cell, a memory (stored as depressed synapses) of only two patterns would mean that 2/3 of each pattern overlaps with the other, at ~44% of the synapses, and only ~11% of synapses are left untrained. Storing three patterns would mean that ~89% of each of them overlaps with at least one of the other two, and ~4% of the total number of

synapses are left untrained. Four patterns would leave a little over 1% of synapses untrained, and storing five would leave around 0.4% untrained, and so on.

However, temporal resolution in the Giovannucci study was ~200 ms, approximately the duration of the learned blink response (and roughly the length of the step cycle in mice) (around 200-300 ms: Sarnaik and Raman 2018). That is, temporal resolution of imaging is the duration of a whole movement *cycle*. Two of the authors of the Giovannucci study have since pointed out that the protocol did not combine 'imaging and behavioural studies with concomitant electrophysiological recordings [as it would need to in order] to establish the level of fine-temporal dimensions of granule cell activity' (Badura and De Zeeuw 2017 p.R417).

The estimated activity, then, may represent more than a single pattern of granule cell activation. In addition, the recordings may reflect reafferent sensory inputs (and potentially corticonuclear feedback), possibly from the anterior interpositus (Giovannucci, Badura et al. 2017), which contains cells that increase their firing during the conditioned blink response (Freeman and Steinmetz 2011, Heiney, Wohl et al. 2014). They include activations at the time of (as well as leading) the blink, and may include sensory signals triggered by blinking. Their spatial resolution did not permit confidence that recordings were only from blink circuits (nor, indeed, is the blink driven by activations only in blink circuits but, rather, by activations across the large region that provides parallel fibre innervation to them).

The significance is that a count made in this way of the proportion of sampled granule cells that are active does not represent the number on a functionally finer spatial and temporal scale, or therefore the density of parallel fibre activity. We consider now in turn briefly theoretical reasons for the same conclusion, and then possible methodological reasons that the count itself may have been too high, as opposed to reasons that a high count with these methods does not represent the functional number.

The density of activity in a sample of granule cells does not equate to the density of parallel fibre activity. The microzone-bounded termination footprint of feedback crosses at right angles a mediolateral row of cluster fields that provides parallel fibres that pass through the intersection, which we use to model regulation of parallel fibre activity. The granular layer at the intersection contains around 2.5% of the population of granule cells that provide parallel fibres that pass overhead, so that even if the local count was as high as the headline 2/3 figure, it would still represent around only around 1.6% of the parallel fibre total. Thus, at least in theory, strong activity of granule cells measured somatically does not exclude sparse activity of parallel fibres.

A functional reason to query a high proportion of active granule cells is that it would suggest synaptic learning is redundant because (effectively, functionally) all synapses would be trained in the same direction, following training to a number of

coded patterns in only low single figures. This is contradicted experimentally (Jörntell and Ekerot 2003, Jörntell and Ekerot 2011) and unsatisfying intellectually (because a role of learning in the cerebellum is probable).

Methodological reasons that the count may have been too high (in addition to the reasons that a high count may not mean coding is dense) are: (1) On a 200 ms timescale, ~30% of sampled granule cells were active (following training) (Sam Wang, private correspondence dated 4 January 2017), half the headline figure (which, as we understand it, was counted over a longer period); (2) somatic calcium may be present at elevated but sub-threshold levels, for example, if afferent mossy fibres are active in sub-threshold numbers (if, as reported, unsynchronised input is filtered out: Jörntell and Ekerot 2006); and (3) Somatic calcium may not only be caused by granule cell spiking. ‘Smoking gun proof’ is lacking that there is no leakage into the soma (Andrea Giovannucci, private correspondence dated 2 December 2018). (On the point of whether spiking is sufficient to account for calcium see Giovannucci, Badura et al. 2017 Supplementary Figure 2.) (4) On the same point, another source of elevated somatic calcium may be sub-threshold signals generated mid-granule-cell-axon by stimulation of extrasynaptic $GABA_A$ receptors by GABA spillover, which causes depolarisation that can spread 100s of μm back to the soma (Pugh and Jahr 2011, Pugh and Jahr 2013).

REFERENCES

Abrahamsson, T., L. Cathala, K. Matsui, R. Shigemoto and D. A. Digregorio (2012). "Thin dendrites of cerebellar interneurons confer sublinear synaptic integration and a gradient of short-term plasticity." Neuron **73**(6): 1159-1172.

Albus, J. S. (1971). A Theory of Cerebellar Function.

Albus, J. S. (1971). "A theory of cerebellar function." Mathematical Biosciences **10**(1-2): 25-61.

Albus, J. S. (1989). The Marr and Albus theories of the cerebellum-two early models of associative memory. Digest of Papers. COMPCON Spring 89. Thirty-Fourth IEEE Computer Society International Conference: Intellectual Leverage.

Apps, R. and M. Garwicz (2005). "Anatomical and physiological foundations of cerebellar information processing." Nat Rev Neurosci **6**(4): 297-311.

Apps, R. and R. Hawkes (2009). "Cerebellar cortical organization: a one-map hypothesis." Nat Rev Neurosci **10**(9): 670-681.

Apps, R., R. Hawkes, S. Aoki, F. Bengtsson, A. M. Brown, G. Chen, T. J. Ebner, P. Isope, H. Jorntell, E. P. Lackey, C. Lawrenson, B. Lumb, M. Schonewille, R. V. Sillitoe, L. Spaeth, I. Sugihara, A. Valera, J. Voogd, D. R. Wylie and T. J. H. Ruigrok (2018). "Cerebellar Modules and Their Role as Operational Cerebellar Processing Units." Cerebellum.

Armstrong, B. D. and R. J. Harvey (1966). "Responses in the inferior olive to stimulation of the cerebellar and cerebral cortices in the cat." J Physiol **187**(3): 553-574.

Armstrong, D. M. and S. A. Edgley (1984). "Discharges of Purkinje cells in the paravermal part of the cerebellar anterior lobe during locomotion in the cat." The Journal of physiology **352**: 403-424.

Armstrong, D. M. and S. A. Edgley (1988). "Discharges of interpositus and Purkinje cells of the cat cerebellum during locomotion under different conditions." J Physiol **400**: 425-445.

Armstrong, D. M. and J. A. Rawson (1979). "Activity patterns of cerebellar cortical neurones and climbing fibre afferents in the awake cat." The Journal of physiology **289**: 425-448.

Badura, A. and C. I. De Zeeuw (2017). "Cerebellar Granule Cells: Dense, Rich and Evolving Representations." Curr Biol **27**(11): R415-r418.

Barbour, B. (1993). "Synaptic currents evoked in Purkinje cells by stimulating individual granule cells." Neuron **11**(4): 759-769.

Barmack, N. H. and V. Yakhnitsa (2008). "Functions of interneurons in mouse cerebellum." J Neurosci **28**(5): 1140-1152.

Baumel, Y., G. A. Jacobson and D. Cohen (2009). "Implications of functional anatomy on information processing in the deep cerebellar nuclei." Front Cell Neurosci **3**: 14.

Bazzigaluppi, P., J. R. De Gruijl, R. S. van der Giessen, S. Khosrovani, C. I. De Zeeuw and M. T. de Jeu (2012). "Olivary subthreshold oscillations and burst activity revisited." Front Neural Circuits **6**: 91.

Bengtsson, F., C. F. Ekerot and H. Jorntell (2011). "In vivo analysis of inhibitory synaptic inputs and rebounds in deep cerebellar nuclear neurons." PLoS One **6**(4): e18822.

Bengtsson, F., P. Geborek and H. Jörntell (2013). "Cross-correlations between pairs of neurons in cerebellar cortex in vivo." Neural networks : the official journal of the International Neural Network Society **47**: 88-94.

Bengtsson, F. and G. Hesslow (2006). "Cerebellar control of the inferior olive." Cerebellum **5**(1): 7-14.

Bengtsson, F. and H. Jörntell (2009). "Sensory transmission in cerebellar granule cells relies on similarly coded mossy fiber inputs." Proc Natl Acad Sci U S A **106**(7): 2389-2394.

Bengtsson, F., P. Svensson and G. Hesslow (2004). "Feedback control of Purkinje cell activity by the cerebello-olivary pathway." Eur J Neurosci **20**(11): 2999-3005.

Bidoret, C., A. Ayon, B. Barbour and M. Casado (2009). "Presynaptic NR2A-containing NMDA receptors implement a high-pass filter synaptic plasticity rule." Proc Natl Acad Sci U S A **106**(33): 14126-14131.
Billings, G., E. Piasini, A. Lorincz, Z. Nusser and R. A. Silver (2014). "Network structure within the cerebellar input layer enables lossless sparse encoding." Neuron **83**(4): 960-974.
Blenkinsop, T. A. and E. J. Lang (2006). "Block of inferior olive gap junctional coupling decreases Purkinje cell complex spike synchrony and rhythmicity." J Neurosci **26**(6): 1739-1748.
Braitenberg, V. and R. P. Atwood (1958). "Morphological observations on the cerebellar cortex." J Comp Neurol **109**(1): 1-33.
Braitenberg, V., D. Heck and F. Sultan (1997). "The detection and generation of sequences as a key to cerebellar function: experiments and theory." Behav Brain Sci **20**(2): 229-245; discussion 245-277.
Brand, S., A. L. Dahl and E. Mugnaini (1976). "The length of parallel fibers in the cat cerebellar cortex. An experimental light and electron microscopic study." Exp Brain Res **26**(1): 39-58.
Brickley, S. G., S. G. Cull-Candy and M. Farrant (1996). "Development of a tonic form of synaptic inhibition in rat cerebellar granule cells resulting from persistent activation of GABAA receptors." J Physiol **497 (Pt 3)**: 753-759.
Brochu, G., L. Maler and R. Hawkes (1990). "Zebrin II: a polypeptide antigen expressed selectively by Purkinje cells reveals compartments in rat and fish cerebellum." J Comp Neurol **291**(4): 538-552.
Brunel, N., V. Hakim, P. Isope, J. P. Nadal and B. Barbour (2004). "Optimal information storage and the distribution of synaptic weights: perceptron versus Purkinje cell." Neuron **43**(5): 745-757.
Buisseret-Delmas, C. and P. Angaut (1993). "The cerebellar olivo-corticonuclear connections in the rat." Prog Neurobiol **40**(1): 63-87.
Carter, A. G. and W. G. Regehr (2002). "Quantal events shape cerebellar interneuron firing." Nat Neurosci **5**(12): 1309-1318.
Cayco-Gajic, N. A., C. Clopath and R. A. Silver (2017). "Sparse synaptic connectivity is required for decorrelation and pattern separation in feedforward networks." Nat Commun **8**(1): 1116.
Cayco-Gajic, N. A. and R. A. Silver (2019). "Re-evaluating Circuit Mechanisms Underlying Pattern Separation." Neuron **101**(4): 584-602.
Cerminara, N. L., H. Aoki, M. Loft, I. Sugihara and R. Apps (2013). "Structural basis of cerebellar microcircuits in the rat." J Neurosci **33**(42): 16427-16442.
Cerminara, N. L., E. J. Lang, R. V. Sillitoe and R. Apps (2015). "Redefining the cerebellar cortex as an assembly of non-uniform Purkinje cell microcircuits." Nat Rev Neurosci **16**(2): 79-93.
Cerminara, N. L. and J. A. Rawson (2004). "Evidence that climbing fibers control an intrinsic spike generator in cerebellar Purkinje cells." J Neurosci **24**(19): 4510-4517.
Cesana, E., K. Pietrajtis, C. Bidoret, P. Isope, E. D'Angelo, S. Dieudonne and L. Forti (2013). "Granule cell ascending axon excitatory synapses onto Golgi cells implement a potent feedback circuit in the cerebellar granular layer." J Neurosci **33**(30): 12430-12446.
Chadderton, P., T. W. Margrie and M. Hausser (2004). "Integration of quanta in cerebellar granule cells during sensory processing." Nature **428**(6985): 856-860.
Chan-Palay, V. (1971). "The recurrent collaterals of Purkinje cell axons: a correlated study of the rat's cerebellar cortex with electron microscopy and the Golgi method." Z Anat Entwicklungsgesch **134**(2): 200-234.
Chan-Palay, V. and S. L. Palay (1972). "The stellate cells of the rat's cerebellar cortex." Z Anat Entwicklungsgesch **136**(2): 224-248.
Christian, K. M. and R. F. Thompson (2003). "Neural substrates of eyeblink conditioning: acquisition and retention." Learn Mem **10**(6): 427-455.
Coesmans, M., J. T. Weber, C. I. De Zeeuw and C. Hansel (2004). "Bidirectional parallel fiber plasticity in the cerebellum under climbing fiber control." Neuron **44**(4): 691-700.
Crill, W. E. (1970). "Unitary multiple-spiked responses in cat inferior olive nucleus." J Neurophysiol **33**(2): 199-209.

D'Angelo, E., S. Solinas, J. Mapelli, D. Gandolfi, L. Mapelli and F. Prestori (2013). "The cerebellar Golgi cell and spatiotemporal organization of granular layer activity." Front Neural Circuits **7**: 93.
Daugman, J. G. (1993). Brain metaphor and brain theory. Computational Neuroscience. S. E.L., Cambridge, MA: MIT Press.
De Gruijl, J. R., P. Bazzigaluppi, M. T. G. de Jeu and C. I. De Zeeuw (2012). "Climbing fiber burst size and olivary sub-threshold oscillations in a network setting." PLoS computational biology **8**(12): e1002814-e1002814.
De Gruijl, J. R., T. M. Hoogland and C. I. De Zeeuw (2014). "Behavioral correlates of complex spike synchrony in cerebellar microzones." J Neurosci **34**(27): 8937-8947.
de Ruyter van Steveninck, R. R., G. D. Lewen, S. P. Strong, R. Koberle and W. Bialek (1997). "Reproducibility and variability in neural spike trains." Science **275**(5307): 1805-1808.
de Zeeuw, C. I. and A. S. Berrebi (1996). "Individual Purkinje cell axons terminate on both inhibitory and excitatory neurons in the cerebellar and vestibular nuclei." Ann N Y Acad Sci **781**: 607-610.
De Zeeuw, C. I., D. R. Wylie, P. L. DiGiorgi and J. I. Simpson (1994). "Projections of individual Purkinje cells of identified zones in the flocculus to the vestibular and cerebellar nuclei in the rabbit." J Comp Neurol **349**(3): 428-447.
Dean, P., J. Porrill, C. F. Ekerot and H. Jörntell (2010). "The cerebellar microcircuit as an adaptive filter: experimental and computational evidence." Nat Rev Neurosci **11**(1): 30-43.
Delvendahl, I. and S. Hallermann (2016). "The Cerebellar Mossy Fiber Synapse as a Model for High-Frequency Transmission in the Mammalian CNS." Trends Neurosci **39**(11): 722-737.
Dieudonné, S. (1998). "Submillisecond kinetics and low efficacy of parallel fibre-Golgi cell synaptic currents in the rat cerebellum." J Physiol **510 (Pt 3)**: 845-866.
Duguid, I., T. Branco, P. Chadderton, C. Arlt, K. Powell and M. Hausser (2015). "Control of cerebellar granule cell output by sensory-evoked Golgi cell inhibition." Proc Natl Acad Sci U S A **112**(42): 13099-13104.
Duguid, I., T. Branco, M. London, P. Chadderton and M. Hausser (2012). "Tonic inhibition enhances fidelity of sensory information transmission in the cerebellar cortex." J Neurosci **32**(32): 11132-11143.
Ebner, T. J. and J. R. Bloedel (1984). "Climbing fiber action on the responsiveness of Purkinje cells to parallel fiber inputs." Brain Res **309**(1): 182-186.
Ebner, T. J., Q. X. Yu and J. R. Bloedel (1983). "Increase in Purkinje cell gain associated with naturally activated climbing fiber input." J Neurophysiol **50**(1): 205-219.
Eccles, J. C., M. Ito and J. Szentágothai (1967). The cerebellum as a neuronal machine. Berlin, New York etc., Springer-Verlag.
Eccles, J. C., R. Llinás and K. Sasaki (1964). "Golgi cell inhibition in the cerebellar cortex." Nature **204**: 1265-1266.
Eccles, J. C., R. Llinás and K. Sasaki (1966). "The mossy fibre-granule cell relay of the cerebellum and its inhibitory control by Golgi cells." Exp Brain Res **1**(1): 82-101.
Edgley, S. A. and M. Lidierth (1988). "Step-related discharges of Purkinje cells in the paravermal cortex of the cerebellar anterior lobe in the cat." J Physiol **401**: 399-415.
Ekerot, C. F. and H. Jörntell (2001). "Parallel fibre receptive fields of Purkinje cells and interneurons are climbing fibre-specific." Eur J Neurosci **13**(7): 1303-1310.
Ekerot, C. F. and H. Jörntell (2003). "Parallel fiber receptive fields: a key to understanding cerebellar operation and learning." Cerebellum **2**(2): 101-109.
Ekerot, C. F., H. Jorntell and M. Garwicz (1995). "Functional relation between corticonuclear input and movements evoked on microstimulation in cerebellar nucleus interpositus anterior in the cat." Exp Brain Res **106**(3): 365-376.
Faisal, A. A., L. P. J. Selen and D. M. Wolpert (2008). "Noise in the nervous system." Nature reviews. Neuroscience **9**(4): 292-303.

Freeman, J. H. and C. A. Rabinak (2004). "Eyeblink Conditioning in Rats Using Pontine Stimulation as a Conditioned Stimulus." Integrative physiological and behavioral science : the official journal of the Pavlovian Society **39**(3): 180-191.
Freeman, J. H. and A. B. Steinmetz (2011). "Neural circuitry and plasticity mechanisms underlying delay eyeblink conditioning." Learn Mem **18**(10): 666-677.
Fujita, H. and I. Sugihara (2013). "Branching patterns of olivocerebellar axons in relation to the compartmental organization of the cerebellum." Front Neural Circuits **7**: 3.
Fujita, M. (1982). "Adaptive filter model of the cerebellum." Biol Cybern **45**(3): 195-206.
Gao, Z., M. Proietti-Onori, Z. Lin, M. M. Ten Brinke, H. J. Boele, J. W. Potters, T. J. Ruigrok, F. E. Hoebeek and C. I. De Zeeuw (2016). "Excitatory Cerebellar Nucleocortical Circuit Provides Internal Amplification during Associative Conditioning." Neuron **89**(3): 645-657.
Garwicz, M., R. Apps and J. R. Trott (1996). "Micro-organization of olivocerebellar and corticonuclear connections of the paravermal cerebellum in the cat." Eur J Neurosci **8**(12): 2726-2738.
Garwicz, M. and C. F. Ekerot (1994). "Topographical organization of the cerebellar cortical projection to nucleus interpositus anterior in the cat." The Journal of physiology **474**(2): 245-260.
Garwicz, M., C. F. Ekerot and H. Jorntell (1998). "Organizational principles of cerebellar neuronal circuitry." News in Physiological Sciences **13**: 28-34.
Garwicz, M., H. Jörntell and C. F. Ekerot (1998). "Cutaneous receptive fields and topography of mossy fibres and climbing fibres projecting to cat cerebellar C3 zone." J Physiol **512 (Pt 1)**: 277-293.
Gilbert, P. F. and W. T. Thach (1977). "Purkinje cell activity during motor learning." Brain Res **128**(2): 309-328.
Giovannucci, A., A. Badura, B. Deverett, F. Najafi, T. D. Pereira, Z. Gao, I. Ozden, A. D. Kloth, E. Pnevmatikakis, L. Paninski, C. I. De Zeeuw, J. F. Medina and S. S. Wang (2017). "Cerebellar granule cells acquire a widespread predictive feedback signal during motor learning." Nat Neurosci.
Guo, C. C., M. C. Ke and J. L. Raymond (2014). "Cerebellar encoding of multiple candidate error cues in the service of motor learning." J Neurosci **34**(30): 9880-9890.
Hámori, J. (1992). Anatomy and neurochemical anatomy of the cerebellum. Cerebellar degenerations : clinical neurobiology. A. Plaitakis, Kluwer Academic.
Hansel, C. (2009). "Reading the clock: how Purkinje cells decode the phase of olivary oscillations." Neuron **62**(3): 308-309.
Hansel, C. and D. J. Linden (2000). "Long-term depression of the cerebellar climbing fiber--Purkinje neuron synapse." Neuron **26**(2): 473-482.
Hansel, C., D. J. Linden and E. D'Angelo (2001). "Beyond parallel fiber LTD: the diversity of synaptic and non-synaptic plasticity in the cerebellum." Nat Neurosci **4**(5): 467-475.
Harvey, R. J. and R. M. Napper (1988). "Quantitative study of granule and Purkinje cells in the cerebellar cortex of the rat." J Comp Neurol **274**(2): 151-157.
Harvey, R. J. and R. M. Napper (1991). "Quantitative studies on the mammalian cerebellum." Prog Neurobiol **36**(6): 437-463.
Häusser, M. and B. A. Clark (1997). "Tonic synaptic inhibition modulates neuronal output pattern and spatiotemporal synaptic integration." Neuron **19**(3): 665-678.
Heck, D. H., W. T. Thach and J. G. Keating (2007). "On-beam synchrony in the cerebellum as the mechanism for the timing and coordination of movement." Proceedings of the National Academy of Sciences **104**(18): 7658.
Heiney, S. A., J. Kim, G. J. Augustine and J. F. Medina (2014). "Precise control of movement kinematics by optogenetic inhibition of Purkinje cell activity." J Neurosci **34**(6): 2321-2330.
Heiney, S. A., M. P. Wohl, S. N. Chettih, L. I. Ruffolo and J. F. Medina (2014). "Cerebellar-Dependent Expression of Motor Learning during Eyeblink Conditioning in Head-Fixed Mice." The Journal of Neuroscience **34**(45): 14845.
Hesslow, G. and M. Ivarsson (1996). "Inhibition of the inferior olive during conditioned responses in the decerebrate ferret." Exp Brain Res **110**(1): 36-46.

Houck, B. D. and A. L. Person (2015). "Cerebellar Premotor Output Neurons Collateralize to Innervate the Cerebellar Cortex." J Comp Neurol **523**(15): 2254-2271.
Hoxha, E., F. Tempia, P. Lippiello and M. C. Miniaci (2016). "Modulation, Plasticity and Pathophysiology of the Parallel Fiber-Purkinje Cell Synapse." Front Synaptic Neurosci **8**: 35.
Husson, Z., C. V. Rousseau, I. Broll, H. U. Zeilhofer and S. Dieudonne (2014). "Differential GABAergic and glycinergic inputs of inhibitory interneurons and Purkinje cells to principal cells of the cerebellar nuclei." J Neurosci **34**(28): 9418-9431.
Ishikawa, T., S. Tomatsu, Y. Tsunoda, J. Lee, D. S. Hoffman and S. Kakei (2014). "Releasing dentate nucleus cells from Purkinje cell inhibition generates output from the cerebrocerebellum." PLoS One **9**(10): e108774.
Isope, P. and B. Barbour (2002). "Properties of unitary granule cell-->Purkinje cell synapses in adult rat cerebellar slices." J Neurosci **22**(22): 9668-9678.
Ito, M. (1984). The cerebellum and neural control. New York, Raven Press.
Ito, M. (1989). "Long-term depression." Annu Rev Neurosci **12**: 85-102.
Ito, M. (2001). "Cerebellar long-term depression: characterization, signal transduction, and functional roles." Physiol Rev **81**(3): 1143-1195.
Jakab, R. L. and J. Hámori (1988). "Quantitative morphology and synaptology of cerebellar glomeruli in the rat." Anat Embryol (Berl) **179**(1): 81-88.
Jelitai, M., P. Puggioni, T. Ishikawa, A. Rinaldi and I. Duguid (2016). "Dendritic excitation-inhibition balance shapes cerebellar output during motor behaviour." Nat Commun **7**: 13722.
Jirenhed, D. A., F. Bengtsson and G. Hesslow (2007). "Acquisition, extinction, and reacquisition of a cerebellar cortical memory trace." J Neurosci **27**(10): 2493-2502.
Jirenhed, D. A. and G. Hesslow (2016). "Are Purkinje Cell Pauses Drivers of Classically Conditioned Blink Responses?" The Cerebellum **15**(4): 526-534.
Johansson, F., H. A. Carlsson, A. Rasmussen, C. H. Yeo and G. Hesslow (2015). "Activation of a Temporal Memory in Purkinje Cells by the mGluR7 Receptor." Cell Rep **13**(9): 1741-1746.
Johansson, F., D. A. Jirenhed, A. Rasmussen, R. Zucca and G. Hesslow (2014). "Memory trace and timing mechanism localized to cerebellar Purkinje cells." Proc Natl Acad Sci U S A **111**(41): 14930-14934.
Jörntell, H. and C. F. Ekerot (2002). "Reciprocal bidirectional plasticity of parallel fiber receptive fields in cerebellar Purkinje cells and their afferent interneurons." Neuron **34**(5): 797-806.
Jörntell, H. and C. F. Ekerot (2003). "Receptive field plasticity profoundly alters the cutaneous parallel fiber synaptic input to cerebellar interneurons in vivo." J Neurosci **23**(29): 9620-9631.
Jörntell, H. and C. F. Ekerot (2006). "Properties of somatosensory synaptic integration in cerebellar granule cells in vivo." J Neurosci **26**(45): 11786-11797.
Jörntell, H. and C. F. Ekerot (2011). "Receptive Field Remodeling Induced by Skin Stimulation in Cerebellar Neurons in vivo." Front Neural Circuits **5**: 3.
Jörntell, H. and C. Hansel (2006). "Synaptic memories upside down: bidirectional plasticity at cerebellar parallel fiber-Purkinje cell synapses." Neuron **52**(2): 227-238.
Kanichay, R. T. and R. A. Silver (2008). "Synaptic and cellular properties of the feedforward inhibitory circuit within the input layer of the cerebellar cortex." J Neurosci **28**(36): 8955-8967.
Kitamura, K. and M. Häusser (2011). "Dendritic calcium signaling triggered by spontaneous and sensory-evoked climbing fiber input to cerebellar Purkinje cells in vivo." J Neurosci **31**(30): 10847-10858.
Kondo, S. and A. Marty (1998). "Synaptic currents at individual connections among stellate cells in rat cerebellar slices." J Physiol **509 (Pt 1)**: 221-232.
Lange, W. (1974). "Regional differences in the distribution of golgi cells in the cerebellar cortex of man and some other mammals." Cell Tissue Res **153**(2): 219-226.

Lee, K. H., P. J. Mathews, A. M. Reeves, K. Y. Choe, S. A. Jami, R. E. Serrano and T. S. Otis (2015). "Circuit mechanisms underlying motor memory formation in the cerebellum." Neuron **86**(2): 529-540.
Lemkey-Johnston, N. and L. M. Larramendi (1968). "Types and distribution of synapses upon basket and stellate cells of the mouse cerebellum: an electron microscopic study." J Comp Neurol **134**(1): 73-112.
Llinás, R., R. Baker and C. Sotelo (1974). "Electrotonic coupling between neurons in cat inferior olive." J Neurophysiol **37**(3): 560-571.
Llinás, R. and J. P. Welsh (1993). "On the cerebellum and motor learning." Curr Opin Neurobiol **3**(6): 958-965.
Llinás, R. and Y. Yarom (1981). "Electrophysiology of mammalian inferior olivary neurones in vitro. Different types of voltage-dependent ionic conductances." The Journal of physiology **315**: 549-567.
Llinás, R. R. and M. N. Negrello. (2015). "Cerebellum." Scholarpedia, 10(1):4606 Retrieved 06/01, 2017, from http://www.scholarpedia.org/article/Cerebellum.
Llinás, R. R., K. D. Walton and E. J. Lang (2004). 7. Cerebellum. The synaptic organization of the brain. G. M. Shepherd. Oxford, Oxford University Press.
Loeb, G. E. and M. Mileusnic (2015). "Proprioceptors and models of transduction." Scholarpedia **10**: 12390.
Long, M. A., M. R. Deans, D. L. Paul and B. W. Connors (2002). "Rhythmicity without synchrony in the electrically uncoupled inferior olive." J Neurosci **22**(24): 10898-10905.
Lou, J. S. and J. R. Bloedel (1992). "Responses of sagittally aligned Purkinje cells during perturbed locomotion: relation of climbing fiber activation to simple spike modulation." J Neurophysiol **68**(5): 1820-1833.
Mainen, Z. F. and T. J. Sejnowski (1995). "Reliability of spike timing in neocortical neurons." Science **268**(5216): 1503-1506.
Marr, D. (1969). "A theory of cerebellar cortex." J Physiol **202**(2): 437-470.
Maruta, J., R. A. Hensbroek and J. I. Simpson (2007). "Intraburst and interburst signaling by climbing fibers." J Neurosci **27**(42): 11263-11270.
Mathy, A., S. S. Ho, J. T. Davie, I. C. Duguid, B. A. Clark and M. Hausser (2009). "Encoding of oscillations by axonal bursts in inferior olive neurons." Neuron **62**(3): 388-399.
McCormick, D. A. and R. F. Thompson (1984). "Cerebellum: essential involvement in the classically conditioned eyelid response." Science **223**(4633): 296-299.
Medina, J. F. and M. D. Mauk (2000). "Computer simulation of cerebellar information processing." Nat Neurosci **3 Suppl**: 1205-1211.
Mercer, A. A., K. J. Palarz, N. Tabatadze, C. S. Woolley and I. M. Raman (2016). "Sex differences in cerebellar synaptic transmission and sex-specific responses to autism-linked Gabrb3 mutations in mice." eLife **5**: e07596.
Midtgaard, J. (1992). "Membrane properties and synaptic responses of Golgi cells and stellate cells in the turtle cerebellum in vitro." J Physiol **457**: 329-354.
Mitchell, S. J. and R. A. Silver (2000a). "GABA spillover from single inhibitory axons suppresses low-frequency excitatory transmission at the cerebellar glomerulus." J Neurosci **20**(23): 8651-8658.
Mitchell, S. J. and R. A. Silver (2000b). "Glutamate spillover suppresses inhibition by activating presynaptic mGluRs." Nature **404**(6777): 498-502.
Mottolese, C., N. Richard, S. Harquel, A. Szathmari, A. Sirigu and M. Desmurget (2013). "Mapping motor representations in the human cerebellum." Brain **136**(Pt 1): 330-342.
Mugnaini, E. (1983). "The length of cerebellar parallel fibers in chicken and rhesus monkey." J Comp Neurol **220**(1): 7-15.
Najafi, F. and J. F. Medina (2013). "Beyond "all-or-nothing" climbing fibers: graded representation of teaching signals in Purkinje cells." Front Neural Circuits **7**: 115.

Napper, R. M. and R. J. Harvey (1988). "Number of parallel fiber synapses on an individual Purkinje cell in the cerebellum of the rat." J Comp Neurol **274**(2): 168-177.
Odeh, F., R. Ackerley, J. G. Bjaalie and R. Apps (2005). "Pontine maps linking somatosensory and cerebellar cortices are in register with climbing fiber somatotopy." J Neurosci **25**(24): 5680-5690.
Oscarsson, O. (1979). "Functional units of the cerebellum - sagittal zones and microzones." Trends in Neurosciences **2**: 143-145.
Ozden, I., D. A. Dombeck, T. M. Hoogland, D. W. Tank and S. S. Wang (2012). "Widespread state-dependent shifts in cerebellar activity in locomoting mice." PLoS One **7**(8): e42650.
Ozden, I., M. R. Sullivan, H. M. Lee and S. S. Wang (2009). "Reliable coding emerges from coactivation of climbing fibers in microbands of cerebellar Purkinje neurons." J Neurosci **29**(34): 10463-10473.
Ozol, K., J. M. Hayden, J. Oberdick and R. Hawkes (1999). "Transverse zones in the vermis of the mouse cerebellum." J Comp Neurol **412**(1): 95-111.
Palay, S. L. and V. Chan-Palay (1974). Cerebellar cortex : cytology and organization. Berlin, Springer.
Palkovits, M., E. Mezey, J. Hámori and J. Szentágothai (1977). "Quantitative histological analysis of the cerebellar nuclei in the cat. I. Numerical data on cells and on synapses." Exp Brain Res **28**(1-2): 189-209.
Pantò, M. R., A. Zappalà, R. Parenti, M. F. Serapide and F. Cicirata (2001). "Corticonuclear projections of the cerebellum preserve both anteroposterior and mediolateral pairing patterns." European Journal of Neuroscience **13**(4): 694-708.
Paula-Barbosa, M. M., M. A. Tavares, C. Ruela and H. Barroca (1983). "The distribution of stellate cell descending axons in the rat cerebellum: a Golgi and a combined Golgi-electron microscopical study." J Anat **137 (Pt 4)**: 757-764.
Payne, H. L., R. L. French, C. C. Guo, T. B. Nguyen-Vu, T. Manninen and J. L. Raymond (2019). "Cerebellar Purkinje cells control eye movements with a rapid rate code that is invariant to spike irregularity." Elife **8**.
Pedroarena, C. M. and C. Schwarz (2003). "Efficacy and short-term plasticity at GABAergic synapses between Purkinje and cerebellar nuclei neurons." J Neurophysiol **89**(2): 704-715.
Person, A. L. and I. M. Raman (2012a). "Synchrony and neural coding in cerebellar circuits." Frontiers in Neural Circuits **6**: 97.
Person, A. L. and I. M. Raman (2012b). "Purkinje neuron synchrony elicits time-locked spiking in the cerebellar nuclei." Nature **481**(7382): 502-505.
Pichitpornchai, C., J. A. Rawson and S. Rees (1994). "Morphology of parallel fibres in the cerebellar cortex of the rat: an experimental light and electron microscopic study with biocytin." J Comp Neurol **342**(2): 206-220.
Pijpers, A., R. Apps, J. Pardoe, J. Voogd and T. J. Ruigrok (2006). "Precise spatial relationships between mossy fibers and climbing fibers in rat cerebellar cortical zones." J Neurosci **26**(46): 12067-12080.
Pijpers, A., J. Voogd and T. J. Ruigrok (2005). "Topography of olivo-cortico-nuclear modules in the intermediate cerebellum of the rat." J Comp Neurol **492**(2): 193-213.
Porrill, J., P. Dean and J. V. Stone (2004). "Recurrent cerebellar architecture solves the motor-error problem." Proceedings. Biological sciences **271**(1541): 789-796.
Pugh, J. R. and C. E. Jahr (2011). "Axonal GABAA receptors increase cerebellar granule cell excitability and synaptic activity." J Neurosci **31**(2): 565-574.
Pugh, J. R. and C. E. Jahr (2013). "Activation of Axonal Receptors by GABA Spillover Increases Somatic Firing." The Journal of Neuroscience **33**(43): 16924-16929.
Qiu, D. L. and T. Knopfel (2009). "Presynaptically expressed long-term depression at cerebellar parallel fiber synapses." Pflugers Arch **457**(4): 865-875.
Quy, P. N., H. Fujita, Y. Sakamoto, J. Na and I. Sugihara (2011). "Projection patterns of single mossy fiber axons originating from the dorsal column nuclei mapped on the aldolase C compartments in the rat cerebellar cortex." J Comp Neurol **519**(5): 874-899.

Raman, I. M. and B. P. Bean (1999). "Ionic currents underlying spontaneous action potentials in isolated cerebellar Purkinje neurons." J Neurosci **19**(5): 1663-1674.
Ramirez, J. E. and B. M. Stell (2016). "Calcium Imaging Reveals Coordinated Simple Spike Pauses in Populations of Cerebellar Purkinje Cells." Cell Reports **17**(12): 3125-3132.
Ramnani, N. (2006). "The primate cortico-cerebellar system: anatomy and function." Nat Rev Neurosci **7**(7): 511-522.
Rancillac, A. and F. Crépel (2004). "Synapses between parallel fibres and stellate cells express long-term changes in synaptic efficacy in rat cerebellum." J Physiol **554**(Pt 3): 707-720.
Rancz, E. A., T. Ishikawa, I. Duguid, P. Chadderton, S. Mahon and M. Hausser (2007). "High-fidelity transmission of sensory information by single cerebellar mossy fibre boutons." Nature **450**(7173): 1245-1248.
Rasmussen, A. (2019). "Graded error signals in eyeblink conditioning." Neurobiol Learn Mem.
Rasmussen, A., D. A. Jirenhed and G. Hesslow (2008). "Simple and complex spike firing patterns in Purkinje cells during classical conditioning." Cerebellum **7**(4): 563-566.
Rasmussen, A., D. A. Jirenhed, R. Zucca, F. Johansson, P. Svensson and G. Hesslow (2013). "Number of spikes in climbing fibers determines the direction of cerebellar learning." J Neurosci **33**(33): 13436-13440.
Raymond, J. L. and J. F. Medina (2018). "Computational Principles of Supervised Learning in the Cerebellum." Annu Rev Neurosci **41**: 233-253.
Ritzau-Jost, A., I. Delvendahl, A. Rings, N. Byczkowicz, H. Harada, R. Shigemoto, J. Hirrlinger, J. Eilers and S. Hallermann (2014). "Ultrafast action potentials mediate kilohertz signaling at a central synapse." Neuron **84**(1): 152-163.
Rosenblatt, F. (1961). "Principles of neurodynamics: perceptrons and the theory of brain mechanisms." Report no. 1196-G-8, reproduced by the Armed Services Technical Information Agency.
Ruigrok, T. J. (1997). "Cerebellar nuclei: the olivary connection." Prog Brain Res **114**: 167-192.
Ruigrok, T. J. (2011). "Ins and outs of cerebellar modules." Cerebellum **10**(3): 464-474.
Ruigrok, T. J., R. A. Hensbroek and J. I. Simpson (2011). "Spontaneous activity signatures of morphologically identified interneurons in the vestibulocerebellum." J Neurosci **31**(2): 712-724.
Ruigrok, T. J. and J. Voogd (1990). "Cerebellar nucleo-olivary projections in the rat: an anterograde tracing study with Phaseolus vulgaris-leucoagglutinin (PHA-L)." J Comp Neurol **298**(3): 315-333.
Sarnaik, R. and I. M. Raman (2018). "Control of voluntary and optogenetically perturbed locomotion by spike rate and timing of neurons of the mouse cerebellar nuclei." Elife **7**.
Sauerbrei, B. A., E. V. Lubenov and A. G. Siapas (2015). "Structured Variability in Purkinje Cell Activity during Locomotion." Neuron **87**(4): 840-852.
Schreiber, S., J. M. Fellous, P. Tiesinga and T. J. Sejnowski (2004). "Influence of ionic conductances on spike timing reliability of cortical neurons for suprathreshold rhythmic inputs." J Neurophysiol **91**(1): 194-205.
Sejnowski, T. J. (1977). "Storing covariance with nonlinearly interacting neurons." J Math Biol **4**(4): 303-321.
Shadmehr, R. (2017). "Distinct neural circuits for control of movement vs. holding still." J Neurophysiol **117**(4): 1431-1460.
Shambes, G. M., J. M. Gibson and W. Welker (1978). "Fractured somatotopy in granule cell tactile areas of rat cerebellar hemispheres revealed by micromapping." Brain Behav Evol **15**(2): 94-140.
Shinoda, Y. and I. Sugihara (2013). Axonal trajectories of single climbing and mossy fiber neurons in the cerebellar cortex and nucleus. Handbook of the Cerebellum and Cerebellar Disorders. M. Manto, D. L. Gruol, J. D. Schmahmann, N. Koibuchi and F. Rossi, Springer Reference**:** 437-467.
Shinoda, Y., I. Sugihara, H. S. Wu and Y. Sugiuchi (2000). "The entire trajectory of single climbing and mossy fibers in the cerebellar nuclei and cortex." Prog Brain Res **124**: 173-186.

Simpson, J. I., D. R. Wylie and C. I. De Zeeuw (1996). "On climbing fiber signals and their consequence(s)." Behavioral and Brain Sciences **19**(3): 384-398.
Smith, S. L. and T. S. Otis (2005). "Pattern-dependent, simultaneous plasticity differentially transforms the input-output relationship of a feedforward circuit." Proc Natl Acad Sci U S A **102**(41): 14901-14906.
Soler-Llavina, G. J. and B. L. Sabatini (2006). "Synapse-specific plasticity and compartmentalized signaling in cerebellar stellate cells." Nat Neurosci **9**(6): 798-806.
Steinmetz, J. E., C. G. Logan, D. J. Rosen, J. K. Thompson, D. G. Lavond and R. F. Thompson (1987). "Initial localization of the acoustic conditioned stimulus projection system to the cerebellum essential for classical eyelid conditioning." Proc Natl Acad Sci U S A **84**(10): 3531-3535.
Streng, M. L., L. S. Popa and T. J. Ebner (2018). "Complex Spike Wars: a New Hope." The Cerebellum **17**(6): 735-746.
Stuart, G., N. Spruston, B. Sakmann and M. Hausser (1997). "Action potential initiation and backpropagation in neurons of the mammalian CNS." Trends Neurosci **20**(3): 125-131.
Sugihara, I., H. Fujita, J. Na, P. N. Quy, B. Y. Li and D. Ikeda (2009). "Projection of reconstructed single Purkinje cell axons in relation to the cortical and nuclear aldolase C compartments of the rat cerebellum." J Comp Neurol **512**(2): 282-304.
Sugihara, I., E. J. Lang and R. Llinás (1993). "Uniform olivocerebellar conduction time underlies Purkinje cell complex spike synchronicity in the rat cerebellum." The Journal of physiology **470**: 243-271.
Sultan, F. (2000). "Exploring a critical parameter of timing in the mouse cerebellar microcircuitry: the parallel fiber diameter." Neurosci Lett **280**(1): 41-44.
Sultan, F. (2001). "Distribution of mossy fibre rosettes in the cerebellum of cat and mice: evidence for a parasagittal organization at the single fibre level." Eur J Neurosci **13**(11): 2123-2130.
Sultan, F. and J. M. Bower (1998). "Quantitative Golgi study of the rat cerebellar molecular layer interneurons using principal component analysis." J Comp Neurol **393**(3): 353-373.
Sultan, F. and D. Heck (2003). "Detection of sequences in the cerebellar cortex: numerical estimate of the possible number of tidal-wave inducing sequences represented." J Physiol Paris **97**(4-6): 591-600.
Szapiro, G. and B. Barbour (2007). "Multiple climbing fibers signal to molecular layer interneurons exclusively via glutamate spillover." Nat Neurosci **10**(6): 735-742.
Telgkamp, P., D. E. Padgett, V. A. Ledoux, C. S. Woolley and I. M. Raman (2004). "Maintenance of high-frequency transmission at purkinje to cerebellar nuclear synapses by spillover from boutons with multiple release sites." Neuron **41**(1): 113-126.
Telgkamp, P. and I. M. Raman (2002). "Depression of inhibitory synaptic transmission between Purkinje cells and neurons of the cerebellar nuclei." J Neurosci **22**(19): 8447-8457.
ten Brinke, M. M. (2017). On Olivocerebellar Activity and Function. Doctoral dissertation, Erasmus.
Teune, T. M., J. van der Burg, J. van der Moer, J. Voogd and T. J. Ruigrok (2000). "Topography of cerebellar nuclear projections to the brain stem in the rat." Prog Brain Res **124**: 141-172.
Tran-Van-Minh, A., T. Abrahamsson, L. Cathala and D. A. DiGregorio (2016). "Differential Dendritic Integration of Synaptic Potentials and Calcium in Cerebellar Interneurons." Neuron **91**(4): 837-850.
Trigo, F. F., T. Sakaba, D. Ogden and A. Marty (2012). "Readily releasable pool of synaptic vesicles measured at single synaptic contacts." Proc Natl Acad Sci U S A **109**(44): 18138-18143.
Turecek, J., S. L. Jackman and W. G. Regehr (2017). "Synaptotagmin 7 confers frequency invariance onto specialized depressing synapses." Nature **551**: 503.
Uusisaari, M. and E. De Schutter (2011). "The mysterious microcircuitry of the cerebellar nuclei." J Physiol **589**(Pt 14): 3441-3457.
Uusisaari, M. and T. Knopfel (2008). "GABAergic synaptic communication in the GABAergic and non-GABAergic cells in the deep cerebellar nuclei." Neuroscience **156**(3): 537-549.

van Beugen, B. J., Z. Gao, H.-J. Boele, F. E. Hoebeek and C. I. De Zeeuw (2013). "High Frequency Burst Firing of Granule Cells Ensures Transmission at the Parallel Fiber to Purkinje Cell Synapse at the Cost of Temporal Coding." Frontiers in Neural Circuits **7**: 95.
van Kan, P. L., A. R. Gibson and J. C. Houk (1993). "Movement-related inputs to intermediate cerebellum of the monkey." J Neurophysiol **69**(1): 74-94.
Vervaeke, K., A. Lorincz, Z. Nusser and R. A. Silver (2012). "Gap junctions compensate for sublinear dendritic integration in an inhibitory network." Science (New York, N.Y.) **335**(6076): 1624-1628.
Voogd, J., J. Pardoe, T. J. Ruigrok and R. Apps (2003). "The distribution of climbing and mossy fiber collateral branches from the copula pyramidis and the paramedian lobule: congruence of climbing fiber cortical zones and the pattern of zebrin banding within the rat cerebellum." J Neurosci **23**(11): 4645-4656.
Wadiche, J. I. and C. E. Jahr (2005). "Patterned expression of Purkinje cell glutamate transporters controls synaptic plasticity." Nat Neurosci **8**(10): 1329-1334.
Wall, M. J. and M. M. Usowicz (1997). "Development of action potential-dependent and independent spontaneous GABAA receptor-mediated currents in granule cells of postnatal rat cerebellum." Eur J Neurosci **9**(3): 533-548.
Werner, G. (2011). "Letting the brain speak for itself." Front Physiol **2**: 60.
Witter, L., S. Rudolph, R. T. Pressler, S. I. Lahlaf and W. G. Regehr (2016). "Purkinje Cell Collaterals Enable Output Signals from the Cerebellar Cortex to Feed Back to Purkinje Cells and Interneurons." Neuron **91**(2): 312-319.
Wolpert, D. M. and R. C. Miall (1996). "Forward Models for Physiological Motor Control." Neural Netw **9**(8): 1265-1279.
Wolpert, D. M., R. C. Miall and M. Kawato (1998). "Internal models in the cerebellum." Trends Cogn Sci **2**(9): 338-347.
Wu, H. S., I. Sugihara and Y. Shinoda (1999). "Projection patterns of single mossy fibers originating from the lateral reticular nucleus in the rat cerebellar cortex and nuclei." J Comp Neurol **411**(1): 97-118.
Wylie, D. R., C. I. De Zeeuw, P. L. Digiorgi and J. I. Simpson (1994). "Projections of individual purkinje cells of identified zones in the ventral nodulus to the vestibular and cerebellar nuclei in the rabbit." The Journal of Comparative Neurology **349**(3): 448-463.
Zang, Y. and E. De Schutter (2019). "Climbing Fibers Provide Graded Error Signals in Cerebellar Learning." Frontiers in Systems Neuroscience **13**(46).
Zhang, W. and D. J. Linden (2012). "Calcium influx measured at single presynaptic boutons of cerebellar granule cell ascending axons and parallel fibers." Cerebellum **11**(1): 121-131.
Zheng, N. and I. M. Raman (2010). "Synaptic inhibition, excitation, and plasticity in neurons of the cerebellar nuclei." Cerebellum **9**(1): 56-66.
Zhou, H., Z. Lin, K. Voges, C. Ju, Z. Gao, L. W. Bosman, T. J. Ruigrok, F. E. Hoebeek, C. I. De Zeeuw and M. Schonewille (2014). "Cerebellar modules operate at different frequencies." Elife **3**: e02536.
Zucca, R., A. Rasmussen and F. Bengtsson (2016). "Climbing Fiber Regulation of Spontaneous Purkinje Cell Activity and Cerebellum-Dependent Blink Responses(1,2,3)." eNeuro **3**(1).