

# A computational study of synaptic mechanisms of partial memory transfer in cerebellar vestibulo-ocular-reflex learning

Naoki Masuda · Shun-ichi Amari

Received: 24 October 2006 / Revised: 19 May 2007 / Accepted: 25 May 2007 / Published online: 7 July 2007  
© Springer Science + Business Media, LLC 2007

**Abstract** There is a debate regarding whether motor memory is stored in the cerebellar cortex, or the cerebellar nuclei, or both. Memory may be acquired in the cortex and then be transferred to the cerebellar nuclei. Based on a dynamical system modeling with a minimal set of variables, we theoretically investigated possible mechanisms of memory transfer and consolidation in the context of vestibulo-ocular reflex learning. We tested different plasticity rules for synapses in the cerebellar nuclei and took robustness of behavior against parameter variation as the criterion of plausibility of a model variant. In the most plausible scenarios, mossy-fiber nucleus-neuron synapses or Purkinje-cell nucleus-neuron synapses are plastic on a slow time scale and store permanent memory, whose content is passed from the cerebellar cortex storing transient memory. In these scenarios, synaptic strengths are potentiated when the mossy-fiber afferents to the nuclei are active during a pause in Purkinje-cell activities. Furthermore, assuming that mossy fibers create a limited variety of signals compared to parallel fibers, our model shows partial memory transfer from the cortex to the nuclei.

**Keywords** Cerebellum · Motor learning · Synaptic plasticity

## 1 Introduction

The cerebellum is involved in various types of motor learning. It is composed of the cerebellar cortex and several cerebellar nuclei. In particular, the vestibular nucleus (VN) provides motor commands for the vestibulo-ocular reflex (VOR). Two main pathways link vestibular mossy fiber (MF) signals to VN output (Fig. 1). One that directly relays the MFs to the VN is termed the direct pathway. The pathway via MFs, granule cells (GCs), parallel fibers (PFs), which are axons of GCs, and Purkinje cells (PCs) in the flocculo-nodular lobes of the cerebellar cortex, is termed the indirect pathway. Because PCs, which are the sole output from the cerebellar cortex, are GABAergic, firing rates of the VN are suppressed by activating this pathway. Human cerebellar nuclei contain  $5 \times 10^5$  neurons and are targeted by  $1.5 \times 10^7$  PCs. The indirect pathway contains interneurons, such as basket, stellate, and Golgi cells, which implement local feedforward and feedback loops. The cerebellar cortex has  $5 \times 10^{10}$  neurons, a considerable part of which stems from the GCs; they account for half the neurons in the whole brain. Because of this enormous number, the indirect pathway is endowed with a fan-out and fan-in structure. Each PC receives synaptic contacts from  $10^5$ – $10^6$  PFs. Each GC collects signals from 4–5 MFs. Another anatomical feature of the indirect pathway is that climbing fibers (CFs) from the inferior olive (IO) in the medulla innervate on PCs. Each CF contacts 1–15 PCs, and each PC receives just a single CF input. CFs are considered to convey somatosensory, visual, and cerebral-cortical information.

Based on the large number of GCs, Marr first conjectured that the indirect pathway operates as a perceptron with a large storage (Marr 1969). Albus suggested long-term depression (LTD) rather than long-term potentiation (LTP) of PF-PC synaptic weights should occur, with the

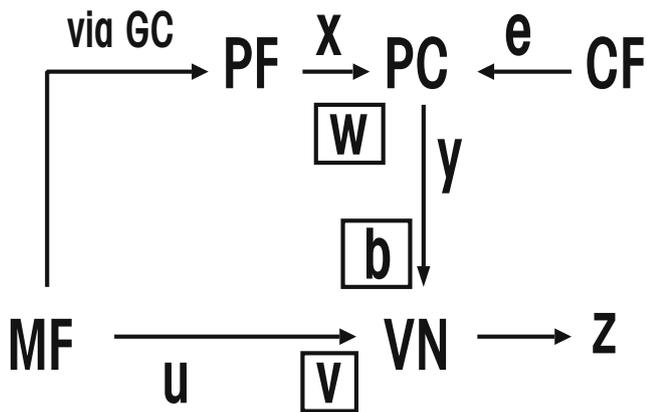
---

**Action Editor:** Erik De Schutter

---

N. Masuda (✉) · S.-i. Amari  
Amari Research Unit, RIKEN Brain Science Institute,  
2-1 Hirosawa, Wako, Saitama 351-0198, Japan  
e-mail: masuda@mist.i.u-tokyo.ac.jp

S.-i. Amari  
e-mail: amari@brain.riken.jp



**Fig. 1** Architecture of the VOR learning model. The *variables surrounded by squares* are synaptic weights

CFs carrying error-correcting signals (Albus 1971). Experimental work confirmed existence of both forms of plasticity at PF-PC synapses in the flocculus lobe. LTD can be induced when afferents to the PCs from PFs and ones from the CFs are simultaneously active (Ito et al. 1982b; Sakurai 1987; Hansel et al. 2001; Ito 2001; Boyden et al. 2004); visual and vestibular signals meet there. The depression lasts for at least one hour after these associative inputs are removed. LTP occurs with PF activity not paired with CF activities (Sakurai 1987; Hansel et al. 2001; Ito 2001; Boyden et al. 2004). The necessity of the flocculus for motor learning and its retention is supported by evidence that acute flocculus shutdown impairs memory consolidation in VOR (Ito et al. 1982a; Nagao and Kitazawa 2003).

However, the cerebellar cortex is not the only site where visual and vestibular signals converge to induce synaptic plasticity. An alternative potential memory site is the cerebellar nuclei. Synaptic plasticity may occur exclusively in the cerebellar nuclei, with the flocculus being just a relay (Miles and Lisberger 1981; Lisberger 1988). Several lines of evidence support this idea. First, flocculus shutdown after three days of VOR learning does not impair acquired motor memory (Luebke and Robinson 1994). Second, PC activities are plastic in response to vestibular signals but in the direction opposite to the direction predicted by the motor learning scheme based on LTD of PF-PC synapses (Miles and Lisberger 1981; Lisberger 1988; du Lac et al. 1995). Third, LTP occurs in the vestibular and interpositus nuclei when peduncles are electrically stimulated (Racine et al. 1986).

Relative contributions of the two mechanisms to motor learning remain to be studied further. The two putative memory sites may cooperate in VOR learning and also in eyelid conditioning (Lisberger 1988; du Lac et al. 1995; Raymond et al. 1996; Mauk 1997; Mauk and Donegan 1997; Raymond and Lisberger 1998; Ito 2001; Nagao and Kitazawa 2003; Boyden et al. 2004). Related to this issue is

the distinction between transient (hours) and permanent (days or longer) plasticity. Plasticity of PF-PC synapses may operate on a short timescale, whereas synaptic plasticity at the cerebellar nuclei may operate in the long term. This concert implies that transient memory in the cerebellar cortex is eventually transferred to the cerebellar nuclei (trigger-and-storage model) (Perrett and Mauk 1995; Raymond et al. 1996; Mauk 1997; Mauk and Donegan 1997), and that control signals embedded in the PC firing elicit permanent plasticity in the cerebellar nuclei. The acquired memory will not be disturbed if the cerebellar cortex is lesioned after a sufficiently long learning period during which memory transfer occurs. Recent VOR experiments with cats (Kassardjian et al. 2005) and mice (Shutoh et al. 2006) support this.

Numerical simulations for VOR (Peterson et al. 1991) and eyelid conditioning (Medina and Mauk 1999; Medina et al. 2000) suggest that LTP in the cerebellar nuclei is induced by MF activities during a pause in PC activities. We study linear firing rate models of the cerebellar circuitry with a minimal set of variables to clarify the essential mechanism of memory transfer in VOR adaptation. In this work, robustness against parameter variation is adopted as the principle of model selection. The theoretical tool used for implementing this criterion is fast-slow analysis, which benefits from dissociation between two timescales of synaptic plasticity, one in PF-PC synapses and the other in VN synapses. Transient firing dynamics with multiple timescales were proposed to be a mechanism of cerebellar learning (Lisberger and Sejnowski 1992). Here we consider multiple timescales of synaptic plasticity, not of transient neural dynamics with fixed synapses.

We identify which plasticity rules for VN synapses are consistent with experimental results and efficiency of memory transfer. Specifically, we consider five plasticity rules listed in Table 1. We show that two plasticity rules guided by the PC signal robustly reproduce transfer of the acquired VOR gain from the cerebellar cortex to the VN. Then we briefly address the issue of contrastive memory capacity of two memory sites, which is predicted from numerous GCs, which may endow the cerebellar cortex with high computational capacity (Marr 1969; Albus 1971; Yamazaki and Tanaka 2005) compared to the nuclei. We show that the PC-driven plasticity rule is also robust in realizing partial memory transfer.

## 2 Models

### 2.1 Linear rate model of VOR

The VOR stabilizes the retinal image using information from vestibular organs. We developed a linear rate coding

**Table 1** Five hypothetical plasticity rules for the VN synapses

| Synapse | Rule      | Model         | Results           | References  |
|---------|-----------|---------------|-------------------|---|
| MF–VN   | CF-driven | Section 2.3.1 | Sections 3.1, 3.4 | Racine et al. 1986; Pugh and Raman 2006; Peterson et al. 1991; Medina and Mauk 1999; Medina et al. 2000 |
|         | Hebbian   | Section 2.3.2 | Sections 3.2, 3.4 |   |
| PC–VN   | PC-driven | Section 2.3.3 | Sections 3.3, 3.4 | Morishita and Sastry 1996; Aizenman et al. 1998; Ouardouz and Sastry 2000                               |
|         | Hebbian   | Section 2.4.1 | Section 3.2       |   |
|         | PC-driven | Section 2.4.2 | Section 3.3       |   |

model of the VOR (Fig. 1). We assume that the vestibular signal is composed of a single frequency component. The modeled circuit receives sinusoidal vestibular signal  $\sin\omega t$ . Through VOR learning, the output of the circuit approaches a desired output. Target output is assumed to have gain  $R$  (initial gain is  $R_0=1$ ) and phase  $\theta$  with respect to the vestibular signal. The target phase corresponds to appropriate response timing acquired via phase adaptation (Kramer et al. 1995). The desired output is  $R \sin(\omega t + \theta)$ .

Hair cells code the time derivative of the vestibular signal, whose phase is different from that of the vestibular signal by  $\pi/2$ . Hair cell signals are received by MFs. We assume that phase elements are created by MFs with heterogeneous delay lines. We represent  $m$  MF signals by a time-dependent vector  $\mathbf{u}=(u_1, u_2, \dots, u_m)^t$ , where  $t$  means the transpose. Modulation of firing rates of the  $i$ -th MF in response to sinusoidal vestibular signals is  $u_i=f(\omega t - \psi_i)$ , where the response function  $f$  has period  $2\pi$ . We assume that phase  $\psi_i$  with respect to the vestibular signal is uniformly distributed on  $[-\Delta_\psi, \Delta_\psi]$ . The ability of MFs to create delay elements is likely to be constrained compared to that of PFs, which benefit from feedback circuitry comprising inhibitory interneurons. We model this factor by a small phase dispersion  $\Delta_\psi$ .

The MF signals are also propagated to GCs and then to PFs. PFs form negative feedback circuits in combination with Golgi cells and plastic synapses. These circuits are assumed to enrich the variety of timing elements created by the MFs (De Schutter and Bjaalie 2001), which is theoretically useful in the eyelid conditioning (Perrett et al. 1993; Medina and Mauk 1999; Medina et al. 2000; Yamazaki and Tanaka 2005) and the VOR (Davies and Melvill Jones 1976; Fujita 1982; Raymond and Lisberger 1998). Without modeling the mechanism of phase broadening explicitly, we express  $n$  PF firing rates by  $\mathbf{x}=(x_1, x_2, \dots, x_n)^t$  where  $x_i(t)=f(\omega t + \phi_i)$ . The phase lead  $\phi_i$  is uniformly distributed on  $[0, 2\pi]$  and reflects broadening of the response time distribution due to the feedback loop. We have replaced the abundance of PFs ( $n \gg m$ ) by the ability of the PFs to create any phase elements ( $[0, 2\pi]$ ) in comparison to the MFs whose phase elements are confined

in  $[-\Delta_\psi, \Delta_\psi]$ . Therefore, the numbers  $m$  and  $n$  of synapses do not matter in the following.

In this work, we consider two types of firing rate inputs to the VN and to the PCs: (a) time and frequency independent constant value ( $f(a)=1$ ), and (b) sinusoidal rate ( $f(a)=\sin(a)$ ). In (b), we ignore spontaneous discharges of these signals for theoretical tractability (Goldberg and Fernandez 1971). The simpler form (a) disregards dynamics of firing rates and timing learning but serves to graphical understanding of dynamics of synaptic weights and memory transfer. The form (b) allows analytical understanding of simultaneous learning of gain and timing, although firing rates are allowed to have negative values. PCs receive PF signals  $\mathbf{x}$ . For simplicity, we represent the firing rate of the PC population by a single variable  $y$ . The PF-PC synaptic weights are represented by  $\mathbf{w}=(w_1, w_2, \dots, w_n)^t$ . The PC firing rate is given by

$$y = \mathbf{w}^t \mathbf{x} + y_0, \tag{1}$$

where  $y_0$  is the spontaneous discharges of the PC, which is known to occur at 40–70 Hz (Thach 1968; Armstrong and Rawson 1979; Akemann and Knöpfel 2006). With (a), we interpret each  $w_i$  to be composed of two synapses associated to the motor output to the opposite directions. These two synapses are assumed to counteract, and we express  $w_i=w_{iL}-w_{iR}$  ( $w_{iL}, w_{iR} \geq 0$ ). For simplicity, we have supposed that the reference direction is leftward, and  $L$  and  $R$  correspond to the leftward and rightward directions, respectively. The portion of the PC signal  $w_i x_i = w_{iL} x_i - w_{iR} x_i$ , is the signed summation of  $w_{iL} x_i \geq 0$  corresponding to the leftward motor output and  $w_{iR} x_i \geq 0$  corresponding to the rightward motor output. Although we work with  $w_i$  rather than with  $w_{iL}$  and  $w_{iR}$ , such an interpretation is necessary to ensure that VOR learning is launched by LTD of PF-PC synapses (see the next section).

A bundle of MF collaterals also projects onto the VN without passing through the cerebellar cortex. MF-VN synaptic weights are represented by  $\mathbf{v}=(v_1, v_2, \dots, v_m)^t$ . The output of the modeled circuit is represented by a single variable  $z$ , that is, the mean firing rate of the nuclear

neurons. Because the PC population inhibits the VN, we have

$$z = \mathbf{v}^t \mathbf{u} - b\mathbf{y} + z_0 = \mathbf{v}^t \mathbf{u} - b\mathbf{w}^t \mathbf{x} - b\mathbf{y}_0 + z_0, \quad (2)$$

where  $z_0$  is the spontaneous discharges of the nuclear neurons, which is known to occur at 40–60 Hz (Thach 1968; LeDoux et al. 1998; Roland and Jaeger 2005), and  $b$  is the PC-VN synaptic weight. Similar to the case of the PF-PC synapses, we interpret each  $v_i$  to be composed of  $v_{iL}$  corresponding to movement in the reference direction and  $v_{iR}$  corresponding to the opposite direction.

## 2.2 Learning rule for CF-driven PF–PC plasticity

It is assumed that CFs carry the error signal

$$e = R \sin(\omega t + \theta) - z. \quad (3)$$

PF-PC synaptic weights are depressed when the PF and CF afferents to the PCs are simultaneously activated (Ito et al. 1982b). A positive (negative)  $e$  implies that the circuit output should be tilted to the left (right) to erase the error at time  $t$ . We assume that a positive (negative)  $e$  elicits LTD of the PF-PC synapses driving the movement to the left (right) at time  $t$ . Consequently, the LTD term is equal to  $d\mathbf{w}_L/dt = -\eta_1 e \mathbf{x}$  when  $e > 0$  and  $d\mathbf{w}_R/dt = -\eta_1 (-e) \mathbf{x}$  when  $e < 0$ , where  $\eta_1$  is the learning rate. In this way, the CF signal, which is equal to  $e$  when  $e > 0$  and  $-e$  when  $e < 0$ , is ensured to be nonnegative. With  $f(a) = 1$  (form (a)),  $d\mathbf{w}_L/dt$  and  $d\mathbf{w}_R/dt$  appositely become negative because the elements of  $\mathbf{x}$  are unity. In this case, the LTD rule for  $\mathbf{w}_L$  and  $\mathbf{w}_R$  are summed up into one term:  $d\mathbf{w}/dt = d\mathbf{w}_L/dt - d\mathbf{w}_R/dt = -\eta_1 e \mathbf{x}$ . In this way, we justify a possible negative value of  $e$ . For parsimony, we work with  $\mathbf{w}$  and arbitrarily signed  $e$  instead of  $\mathbf{w}_L$ ,  $\mathbf{w}_R$ , and the nonnegative CF signal. With  $f(a) = \sin(a)$  (form (b)),  $d\mathbf{w}_L/dt$  and  $d\mathbf{w}_R/dt$  may not represent LTD because  $\mathbf{x}$  can be negative. Form (b) is used only for a formal but preliminary analysis of simultaneous gain and phase learning. For both forms of  $f(a)$ , the LTD effect is represented in the Widrow–Hoff form (Dayan and Abbott 2001, p. 320):  $e \mathbf{x} = (\partial e^2 / \partial \mathbf{w}) / 2$ , where we note that  $e$  is a function of  $\mathbf{w}$  based on Eqs. (2) and (3). Therefore, LTD lessens the error  $e^2$ .

In the absence of  $e$ , PC-PF synapses experience LTP (Sakurai 1987; Coesmans et al. 2004), which is non-associative and realizes memory decay. We implement LTP by an additional component  $\eta_2 \mathbf{x}$ , which expresses subtractive normalization counteracting the LTD term (Dayan and Abbott 2001, p. 290). To limit the amount of synaptic plasticity, we assume a multiplicative normalization term  $\eta_3 n \mathbf{w}$ . The factor  $n$  restricts the size of each synaptic weight to scale as  $1/n$  so that the PC firing rate

saturates. The learning rule for the PF–PC synapses becomes

$$\begin{aligned} \frac{d\mathbf{w}}{dt} = & -\eta_1 [R \sin(\omega t + \theta) - \mathbf{v}^t \mathbf{u} + b\mathbf{w}^t \mathbf{x} + b\mathbf{y}_0 - z_0] \mathbf{x} \\ & + \eta_2 \mathbf{x} - \eta_3 n \mathbf{w}. \end{aligned} \quad (4)$$

The quantity inside [] is equal to the error  $e$ .

Synaptic plasticity occurs much more slowly than vestibular signals whose timescale is  $\omega^{-1}$ . Therefore, we take the average over a sinusoidal cycle to obtain

$$\begin{aligned} \frac{d\mathbf{w}}{dt} = & -\eta_1 [R \langle \sin(\omega t + \theta) \mathbf{x} \rangle - \langle \mathbf{x} \mathbf{u}^t \rangle \mathbf{v} + \langle \mathbf{x} \mathbf{x}^t \rangle b \mathbf{w} \\ & + (b\mathbf{y}_0 - z_0) \langle \mathbf{x} \rangle] + \eta_2 \langle \mathbf{x} \rangle - \eta_3 n \mathbf{w}, \end{aligned} \quad (5)$$

where  $\langle \rangle$  indicates the time average. Note that  $\langle \mathbf{x} \mathbf{u}^t \rangle$  and  $\langle \mathbf{x} \mathbf{x}^t \rangle$  are respectively an  $n$  by  $m$  matrix and an  $n$  by  $n$  matrix representing correlation between signals. Learning proceeds based on signal average, signal correlation, and signal-target correlation.

Initially, there is no specific target gain and phase to be acquired, and the synaptic weights rest in stationary values. Then the VOR gain is  $R = R_0$  ( $= 1$ ). The corresponding synaptic weights  $(\mathbf{w}, \mathbf{v}) = (\mathbf{w}_0, \mathbf{v}_0)$  satisfy

$$e = R_0 \sin(\omega t + \theta) - \mathbf{v}_0^t \mathbf{u} + b\mathbf{w}_0^t \mathbf{x} + b\mathbf{y}_0 - z_0 = 0. \quad (6)$$

By setting  $d\mathbf{w}/dt = 0$  in Eq. (5), we derive

$$\begin{aligned} \eta_2 \langle \mathbf{x} \rangle = & \eta_1 [R_0 \langle \sin(\omega t + \theta) \mathbf{x} \rangle - \langle \mathbf{x} \mathbf{u}^t \rangle \mathbf{v}_0 + \langle \mathbf{x} \mathbf{x}^t \rangle b \mathbf{w}_0 \\ & + (b\mathbf{y}_0 - z_0) \langle \mathbf{x} \rangle] + \eta_3 n \mathbf{w}_0 = \eta_3 n \mathbf{w}_0, \end{aligned} \quad (7)$$

which yields

$$\begin{aligned} \frac{d\mathbf{w}}{dt} = & -\eta_1 [R \langle \sin(\omega t + \theta) \mathbf{x} \rangle - \langle \mathbf{x} \mathbf{u}^t \rangle \mathbf{v} + \langle \mathbf{x} \mathbf{x}^t \rangle b \mathbf{w} \\ & + (b\mathbf{y}_0 - z_0) \langle \mathbf{x} \rangle] - \eta_3 n (\mathbf{w} - \mathbf{w}_0) \\ = & -\eta_1 [(R - R_0) \langle \sin(\omega t + \theta) \mathbf{x} \rangle - \langle \mathbf{x} \mathbf{u}^t \rangle (\mathbf{v} - \mathbf{v}_0) \\ & + \langle \mathbf{x} \mathbf{x}^t \rangle (b \mathbf{w} - \mathbf{w}_0)] - \eta_3 n (\mathbf{w} - \mathbf{w}_0). \end{aligned} \quad (8)$$

In words, we have erased a learning rate  $\eta_2$  ( $> 0$ ) by using the initial condition.

## 2.3 Learning rules for MF–VN plasticity

The synapses connecting the MFs and the cerebellar nuclei can be potentiated (Racine et al. 1986; Pugh and Raman 2006). Enhanced excitability of the deep cerebellar nucleus in response to tetanic stimulation (Aizenman and Linden 2000) and of the VN via VOR learning (Shutoh et al. 2006), and increases in the synapse number in the interpositus nucleus after eyelid conditioning (Kleim et al. 2002) also support learning-related strengthening of this connection. The organizing principles of nuclear plasticity

and nucleus specificity are only partially understood (Mauk 1997; Mauk and Donegan 1997; Medina and Mauk 1999; Hansel et al. 2001; Ito 2001). We examine three LTP rules for MF–VN synapses termed (1) CF-driven, (2) Hebbian, and (3) PC-driven rules (Medina and Mauk 1999). We assume that LTD occurs in the absence of learning (Medina and Mauk 1999), which is modeled by  $-\eta_5 \mathbf{u} - \eta_6 m \mathbf{v}$ . The first term is consistent with the recent evidence that burst stimulation of MFs induce LTD of these synapses (Zhang and Linden 2006). The second term limits the amount of plasticity of each synapse to the order of  $1/m$ . Then the VN signal saturates because there are  $m$  MF–VN synapses. We fix the PC–VN synaptic weight when considering the MF–VN plasticity. Accordingly, we set  $b=1$  in this subsection.

### 2.3.1 CF-driven MF–VN plasticity

The IO sends excitatory collaterals to the deep cerebellar nuclei (du Lac et al. 1995; Hansel et al. 2001; Ito 2001). A recent anatomical study is negative about the existence of CF collaterals to the VN (Sugihara et al. 2004). However, CF collaterals to the cerebellar nuclei in general may be important in motor learning schemes involving memory transfer. Let us examine consequences of such potential collaterals. The CF-driven learning dictates that LTP proceeds when hypothetical CF afferents to the VN, which are assumed to carry an error signal, and MF afferents  $\mathbf{u}$  to the VN are simultaneously active. This MF–VN plasticity is formalized in a manner similar to the PF–PC plasticity, which is also CF-driven. By averaging over a sinusoidal cycle, we obtain the learning rate:

$$\begin{aligned} \frac{d\mathbf{v}}{dt} &= \langle \eta_4 [R \sin(\omega t + \theta) - \mathbf{v}^t \mathbf{u} + \mathbf{w}^t \mathbf{x} + y_0 - z_0] \mathbf{u} \\ &\quad - \eta_5 \mathbf{u} - \eta_6 m \mathbf{v} \rangle \\ &= \eta_4 [R \langle \sin(\omega t + \theta) \mathbf{u} \rangle - \langle \mathbf{u} \mathbf{u}^t \rangle \mathbf{v} + \langle \mathbf{u} \mathbf{x}^t \rangle \mathbf{w} + (y_0 - z_0) \langle \mathbf{u} \rangle] \\ &\quad - \eta_5 \langle \mathbf{u} \rangle - \eta_6 m \mathbf{v}, \end{aligned} \tag{9}$$

where we set  $b=1$ . In the initial steady state with no error signal, Eq. (9) in combination with Eq. (6) yields

$$\begin{aligned} \eta_5 \langle \mathbf{u} \rangle &= \eta_4 [R_0 \langle \sin(\omega t + \theta) \mathbf{u} \rangle - \langle \mathbf{u} \mathbf{u}^t \rangle \mathbf{v}_0 + \langle \mathbf{u} \mathbf{x}^t \rangle \mathbf{w}_0 \\ &\quad + (y_0 - z_0) \langle \mathbf{u} \rangle] - \eta_6 m \mathbf{v}_0 = -\eta_6 m \mathbf{v}_0. \end{aligned} \tag{10}$$

Substituting Eq. (10) into Eq. (9) yields the final form of the CF-driven rule:

$$\begin{aligned} \frac{d\mathbf{v}}{dt} &= \eta_4 [(R - R_0) \langle \sin(\omega t + \theta) \mathbf{u} \rangle - \langle \mathbf{u} \mathbf{u}^t \rangle (\mathbf{v} - \mathbf{v}_0) \\ &\quad + \langle \mathbf{u} \mathbf{x}^t \rangle (\mathbf{w} - \mathbf{w}_0)] - \eta_6 m (\mathbf{v} - \mathbf{v}_0). \end{aligned} \tag{11}$$

### 2.3.2 Hebbian MF–VN plasticity

The Hebbian LTP occurs when the MF firing rate  $\mathbf{u}$  and the VN firing rate  $z = \mathbf{v}^t \mathbf{u} - \mathbf{w}^t \mathbf{x} - y_0$  are simultaneously large. Therefore, with  $b=1$ , we obtain

$$\begin{aligned} \frac{d\mathbf{v}}{dt} &= \langle \eta_4 (\mathbf{v}^t \mathbf{u} - \mathbf{w}^t \mathbf{x} - y_0 + z_0) \mathbf{u} - \eta_5 \mathbf{u} - \eta_6 m \mathbf{v} \rangle \\ &= \eta_4 [\langle \mathbf{u} \mathbf{u}^t \rangle \mathbf{v} - \langle \mathbf{u} \mathbf{x}^t \rangle \mathbf{w} - (y_0 - z_0) \langle \mathbf{u} \rangle] - \eta_5 \langle \mathbf{u} \rangle - \eta_6 m \mathbf{v}. \end{aligned} \tag{12}$$

Because the initial steady state corresponds to  $(\mathbf{w}, \mathbf{v}) = (\mathbf{w}_0, \mathbf{v}_0)$ , we have

$$\frac{d\mathbf{v}}{dt} = [\eta_4 \langle \mathbf{u} \mathbf{u}^t \rangle - \eta_6 m] (\mathbf{v} - \mathbf{v}_0) - \eta_4 \langle \mathbf{u} \mathbf{x}^t \rangle (\mathbf{w} - \mathbf{w}_0). \tag{13}$$

### 2.3.3 PC-driven MF–VN plasticity

In the PC-driven rule, LTP of MF–VN synapses is guided by the PC activity  $y$ . Small  $y$  accompanied by a large MF activity should logically induce LTP, which is supported by *in vitro* experiments (Pugh and Raman 2006). We let the LTP term equal to  $(y_{\max} - y) \mathbf{u}$ , where  $y_{\max}$  is the maximum PC firing rate. Then, we obtain

$$\begin{aligned} \frac{d\mathbf{v}}{dt} &= \langle \eta_4 (y_{\max} - \mathbf{w}^t \mathbf{x} - y_0) \mathbf{u} - \eta_5 \mathbf{u} - \eta_6 m \mathbf{v} \rangle \\ &= \eta_4 [(y_{\max} - y_0) \langle \mathbf{u} \rangle - \langle \mathbf{u} \mathbf{x}^t \rangle \mathbf{w}] - \eta_5 \langle \mathbf{u} \rangle - \eta_6 m \mathbf{v}, \end{aligned} \tag{14}$$

which leads to

$$\frac{d\mathbf{v}}{dt} = -\eta_4 \langle \mathbf{u} \mathbf{x}^t \rangle (\mathbf{w} - \mathbf{w}_0) - \eta_6 m (\mathbf{v} - \mathbf{v}_0). \tag{15}$$

## 2.4 Learning rules for PC–VN plasticity

The synapses connecting the PCs and the deep cerebellar nuclei are plastic (Morishita and Sastry 1996; Aizenman et al. 1998; Ouardouz and Sastry 2000) and might be the locus of permanent memory, although their plasticity is currently unknown to be input-specific (Hansel et al. 2001; Ito 2001). Here we study the theoretical possibility of memory transfer from PF–PC synapses to PC–VN synapses.

Now the PC–VN synapse weight by  $b$  is plastic, and the MF–VN synapses  $\mathbf{v}$  are kept constant. The postsynaptic inhibitory input from the PC to the VN is equal to  $by$ , with  $y$  given by Eq. (1). PF–PC plasticity obeys the CF-driven rule as before:

$$\frac{d\mathbf{w}}{dt} = -\eta_1 [R \sin(\omega t + \theta) - \mathbf{v}_0^t \mathbf{u} + by - z_0] \mathbf{x} - \eta_3 n (\mathbf{w} - \mathbf{w}_0). \tag{16}$$

### 2.4.1 Hebbian PC–VN plasticity

We deal with two rules for PC–VN plasticity. The first is the Hebbian rule. Logically speaking, LTP (LTD) of the PF–PC synapses, which results in reduction (enhancement) of the circuit output, is equivalent to LTP (LTD) of the PC–VN synapses. In the deep cerebellar nucleus, LTP of the corresponding synapses is induced when nucleus neurons show rebound excitation, and this plasticity has a postsynaptic origin (Aizenman et al. 1998; Ouardouz and Sastry 2000). However, specific plasticity rules are unknown. Therefore we assume that the VN neurons show similar rebound excitation, and that LTP occurs if the inhibitory PC input to the VN is strong enough to induce large hyperpolarization of the VN neurons and hence more rebound excitation. If postsynaptic excitation is reduced, the PC–VN synapse is assumed to undergo LTD, following the evidence for the deep cerebellar nucleus (Aizenman et al. 1998). We model the effect of rebound excitation on the Hebbian plasticity as follows:

$$\frac{db}{dt} = \eta_4 y (\mathbf{v}_0^t \mathbf{u} + by + z_0 - z_{\text{ref}}) - \eta_6 b. \quad (17)$$

The first Hebbian term in Eq. (17) consists of the presynaptic firing rate  $y$  and the amount of postsynaptic excitation. Because the nucleus neurons would experience large rebound depolarization after the release of large hyperpolarization, PC–VN synapses are assumed to be potentiated more when  $by$  is larger. The threshold  $z_{\text{ref}}$  divides the LTP and LTD regimes. The second term implements saturation of the synaptic weight.

### 2.4.2 PC-driven PC–VN plasticity

The second rule is the PC-driven rule for which the PC–VN synapse is potentiated when the two afferents to the VN, that is, the MF and PC signals are both strong. The plasticity is modeled as follows:

$$\frac{db}{dt} = \eta_4 \mathbf{v}_0^t \mathbf{u} (y - y_{\text{ref}}) - \eta_6 b, \quad (18)$$

Because  $\eta_4 \gg \eta_6$ , LTP occurs when the PC firing rate is larger than a reference level  $y_{\text{ref}}$ . LTD occurs otherwise. Because we deal with form (a) in which MF signals are constant ( $f(a)=1$ ), this LTP rule can be also regarded as an input-driven rule (LTP when the PC signal is strong).

## 2.5 Fast–slow analysis of learning dynamics

We assume that the rate of LTD of the PF–PC synapses is much higher than the rate of LTP:  $\eta_1 \gg \eta_2, \eta_3$ . We also assume that memory decay is sufficiently slow relative to associative

plasticity; otherwise VOR learning would be too leaky to be functional. Therefore  $\eta_1 \gg \eta_2, \eta_3$  and  $\eta_4 \gg \eta_5, \eta_6$ . Excitability of the VN increases slowly in accordance with the slow timescale of VOR learning (Shutoh et al. 2006). Therefore, the plasticity of VN synapses is assumed to be much slower (learning on a timescale of 8–12 h and forgetting on a timescale of 10 days) than that of PF–PC synapses (learning on 0.5 h and forgetting on 1 day):  $\eta_1 \gg \eta_4$  and  $\eta_2, \eta_3 \gg \eta_5, \eta_6$ .

These relations allow us to simplify the analysis of synaptic dynamics based on fast–slow analysis. In other words, we decompose synaptic dynamics into the fast dynamics of  $\mathbf{w}$  and the slow dynamics of  $\mathbf{v}$  or  $b$ . The PF–PC synaptic weights  $\mathbf{w}$  first evolve with  $\mathbf{v}$  or  $b$  almost fixed until they reach a quasistationary state ( $d\mathbf{w}/dt \approx 0$  in Eq. (8)). This is the fast PF–PC synaptic plasticity. Then, keeping  $d\mathbf{w}/dt=0$ , slow adaptation of  $\mathbf{v}$  or  $b$  occurs through MF–VN (PC–VN) synaptic plasticity. These dynamics can be graphically understood by drawing a time course of the synaptic weights in the space spanned by  $\mathbf{w}$  and  $\mathbf{v}$  or  $b$ . A trajectory of  $(\mathbf{w}, \mathbf{v})$  or  $(\mathbf{w}, b)$  initially approaches the fast nullcline defined by  $d\mathbf{w}/dt=0$ . Then, the slow dynamics drives the trajectory toward the crossing of the fast nullcline and the slow nullcline defined by  $d\mathbf{v}/dt=0$  or  $db/dt=0$ . The crossing corresponds to the equilibrium synaptic weights.

## 2.6 Measurement of acquired memory

We assume that the acquired memory is stored in modified synaptic weights. Synaptic plasticity does not necessarily lead to improved performance. We have to constrain ourselves to the synaptic modifications that are responsible for changes in the gain and the phase. To quantify this, we evaluate how much of the acquired response is lost if we reset synaptic weights to the initial values. The circuit output is linear in terms of the MF–VN and PC–VN contributions. Therefore, in the case of the MF–VN plasticity, we measure how much the output of each pathway independently contributes to the acquired output of the whole circuit. Accordingly, we set the PC–VN synaptic weight  $b=1$ . The memory stored in the PF–PC synapses is

$$\begin{aligned} & (\mathbf{v}^t \mathbf{u} - \mathbf{w}^t \mathbf{x} - y_0 + z_0) - (\mathbf{v}_0^t \mathbf{u} - \mathbf{w}_0^t \mathbf{x} - y_0 + z_0) \\ & = (-\mathbf{w}^t + \mathbf{w}_0^t) \mathbf{x}, \end{aligned} \quad (19)$$

and that stored in the MF–VN synapses is

$$\begin{aligned} & (\mathbf{v}^t \mathbf{u} - \mathbf{w}^t \mathbf{x} - y_0 + z_0) - (\mathbf{v}_0^t \mathbf{u} - \mathbf{w}^t \mathbf{x} - y_0 + z_0) \\ & = (\mathbf{v}^t - \mathbf{v}_0^t) \mathbf{u}. \end{aligned} \quad (20)$$

For PC–VN plasticity, the memory stored in the PF–PC synapses is

$$\begin{aligned}
 & (\mathbf{v}^t \mathbf{u} - b(\mathbf{w}^t \mathbf{x} + y_0) + z_0) - (\mathbf{v}^t \mathbf{u} - b(\mathbf{w}_0^t \mathbf{x} + y_0) + z_0) \\
 & = b(-\mathbf{w}^t + \mathbf{w}_0^t) \mathbf{x}, \tag{21}
 \end{aligned}$$

and that stored in the MF–VN synapses is

$$\begin{aligned}
 & (\mathbf{v}_0^t \mathbf{u} - b(\mathbf{w}^t \mathbf{x} + y_0) + z_0) - (\mathbf{v}_0^t \mathbf{u} - b_0(\mathbf{w}^t \mathbf{x} + y_0) + z_0) \\
 & = (-b + b_0) \mathbf{w}^t \mathbf{x}. \tag{22}
 \end{aligned}$$

### 2.7 Parameters for numerical simulations

We set the learning rates equal to  $\eta_1=1$ ,  $\eta_3=0.1$ ,  $\eta_4=0.1$ , and  $\eta_6=0.01$ . Let us note that the values of  $\eta_2$  and  $\eta_5$  are automatically determined from the initial conditions (e.g. Eq. (7)). Initial synaptic weights are  $w_0=1$ ,  $v_0=1$ , and  $b_0=1$ . Spontaneous firing of nucleus cells persists even if excitatory inputs to the nuclei are blocked (Aizenman and Linden 1999). Therefore the spontaneous firing rates of the PC ( $y_0$ ) and the VN neuron ( $z_0$ ) should satisfy  $-y_0+z_0>0$ , and we set  $y_0=0.5$  and  $z_0=1.5$ . These parameter values assure that the steady-state gain is unity ( $R_0=1$ ). To assure that the same condition holds for the PC–VN plasticity rules, we set  $z_{\text{ref}}=4$ . For the gain-only learning, we use one PF–PC synapse ( $n=1$ ) and one MF–VN synapse ( $m=1$ ).

## 3 Results

Plasticity of the PF–PC synapses is always driven by concurrent PF and CF firing. We analyze memory transfer of gain learning with five learning rules in the VN explained above, namely, the CF-driven MF–VN synaptic plasticity, the Hebbian MF–VN plasticity, the Hebbian PC–VN plasticity, the PC-driven MF–VN plasticity, and the PC-driven PC–VN plasticity. Then we examine memory transfer of simultaneous gain-phase learning for the PC-driven MF–VN plasticity.

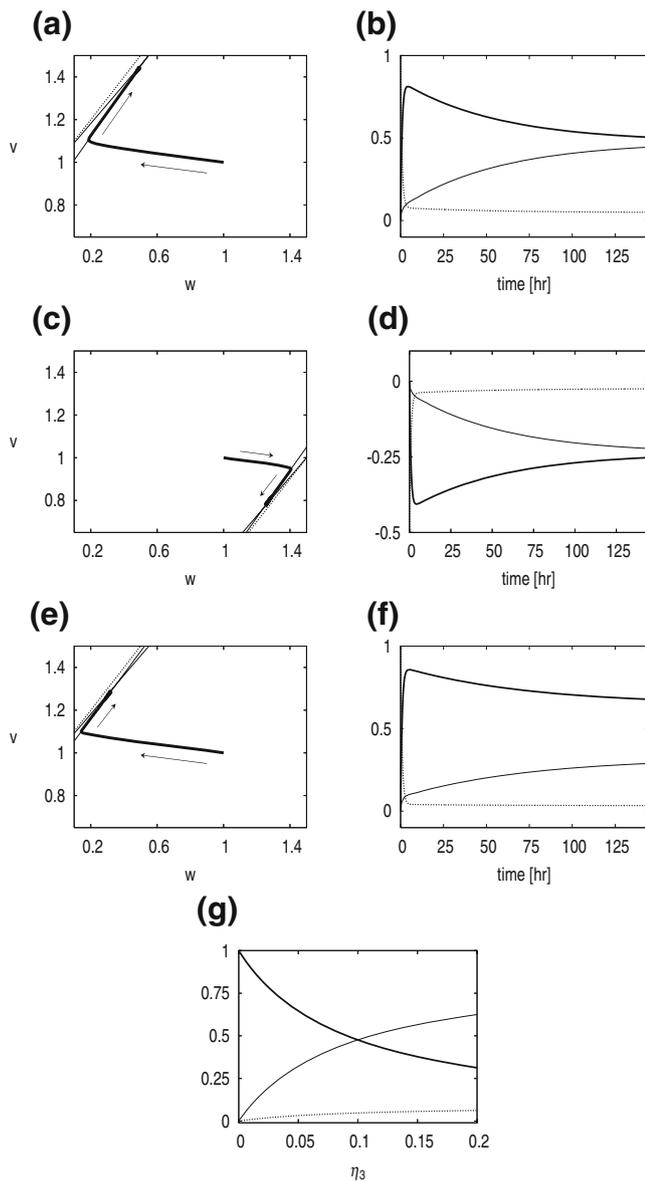
### 3.1 Transfer of gain information with CF-driven MF–VN plasticity

We first ignore response timing and search for a robust mechanism of memory transfer by tracking the evolution of synaptic weights. Then, the vestibular input is constant in time. In accordance we set  $\omega=0$  and  $\theta=\pi/2$  so that the desired output is equal to  $R$ . The response profile of the neurons is assumed to be  $f(a)=1$ , hence the firing rate of a MF (PF) is constant and equal to  $u_i=1$  ( $x_i=1$ ). As a result,

all the MFs and PFs receive the identical input and have the same responsiveness. Therefore, we ignore synaptic specificity and look at the mean synaptic strength of the PF–PC synapses and that of the MF–VN synapses. Equivalently, we represent each ensemble of synapses by single variables  $v$  and  $w$  ( $m=n=1$ ). We will simulate how a target VOR gain is learned by the PF–PC synapse  $w$  and transferred to the MF–VN synapse  $v$ .

The time course of synaptic weights is obtained through fast–slow analysis (see Section 2.5). The details of calculations are shown in Appendix A. A typical time course of synaptic weights based on Eqs. (8) and (11) is depicted in Fig. 2(a) for the adaptation to a larger gain  $R>R_0$  (thick solid line). The PF–PC synapse first experiences LTD to decrease the error shown by the dotted line in Fig. 2(b). Accordingly, the contribution of the PF–PC synapse to the system output transiently increases, as shown by the thick solid line in Fig. 2(b). The trajectory of the synaptic weights (thick solid line in Fig. 2(a)) approaches the fast nullcline (thin solid line), which is close to the error-free line (dotted line). This is the fast dynamics. Then, LTP of the MF–VN synapse ensues to replace the memory stored in the PF–PC synapse (thin solid line in Fig. 2(b)). This process accompanies LTP of the PF–PC synapse along the fast nullcline (the part of the thick solid line overlapping with a thin solid line in Fig. 2(a)). The LTP of the PF–PC synapse here is nonassociative and decreases the amount of memory stored in the PF–PC synapse. This is the slow dynamics. In the long run, the memory is stored mainly in the MF–VN synapse, not in the PF–PC synapse. The overall learning error has been small since the fast learning was completed. This scenario also applies when  $R<R_0$  (Fig. 2(c) and (d)). In this case, net LTP of the PF–PC synapse (LTD of the PF–PC synapses responsible for the motor output opposite to the reference direction; see Sections 2.1 and 2.2 for the interpretation) first occurs. Then LTD of the MF–VN synapse (LTP of the MF–VN synapses contributing to the motor output in the direction opposite to the reference direction) follows.

Perfect learning would imply a trajectory that converges on the  $e=0$  line, that is,  $R-v+w+y_0-z_0=0$  (dotted lines in Fig. 2(a) and (c)). The actual error after sufficient training time is small, as shown in the Fig. 2 and assured by the calculations in Appendix A. The circuit is capable of both error suppression and memory transfer. However, the CF-driven plasticity has deficiencies. First, because  $\eta_1 \gg \eta_3$  and  $\eta_4 \gg \eta_6$  (memory decay is ignored during VOR learning), the two nullclines are close (compare two thin solid lines in Fig. 2(a), similar for Fig. 2(c)). Mathematical details are included in Appendix A. This implies that the position of the equilibrium synaptic weights obtained as a crossing of the two nullclines is sensitive to values of the learning rates, whose choice is rather arbitrary in the absence of data.



**Fig. 2** Memory transfer with the CF-driven MF–VN plasticity. **(a, b)** Gain-up learning ( $R=2$ ) and **(c, d)** gain-down learning ( $R=0.5$ ). **(a)** and **(c)** show time courses of synaptic weights (*thick solid lines*) and the nullclines (*thin solid lines*). The *dotted lines* represent the error-free situations. **(b)** and **(d)** show the time courses of the signed amount of memory stored in the PF–PC (*thick solid lines*) and MF–VN (*thin solid lines*) synapses together with the error (*dotted lines*). **(e, f)** Gain-up learning ( $R=2$ ) when a learning rate is modified to  $\eta_3=0.05$ , with all the other parameters unchanged. The amount of memory transfer (*thin solid line in (f)*) is much reduced compared with the case of  $\eta_3=0.1$  (*thin solid line in (b)*). The effect of modulating  $\eta_3$  is summarized in **(g)** for gain-up learning. Even though the learning error (*dotted line*) stays small, the amount of memory in the PF–PC synapse (*thick solid line*) and that in MF–VN synapse (*thin solid line*) are sensitive to  $\eta_3$

Indeed, if we set  $\eta_3=0.05$  instead of  $\eta_3=0.1$ , which was used in Fig. 2(a–d), the degree of memory transfer considerably changes (compare Fig. 2(e) and (f) with (a

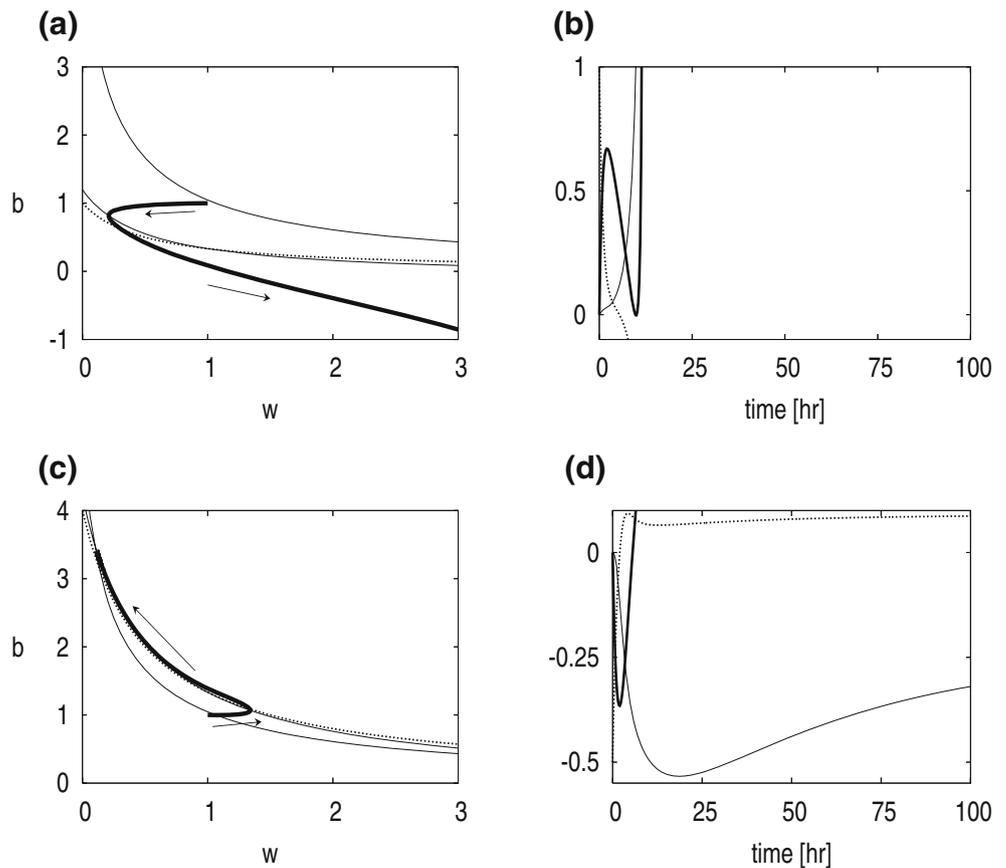
and (b)). Memory transfer is not robust against variation of  $\eta_3$  (Fig. 2(g)), as well as variation of other parameters, as mathematically shown in Appendix A (Eqs. (29) and (30)). The equilibrium can be even organized so that PF–PC synapses experience net LTP for gain-up learning, which is against experimental data (Ito et al. 1982b). Second, memory transfer takes long time (compare Fig. 2(b,d,f) with the results for other plasticity rules shown in Figs. 4(b,d) and 5(b,d)). This is because proximity of the two nullclines (thin solid lines in Fig. 2(a,c,e)) indicates that, on the fast nullcline ( $dw/dt=0$ ), evolution of the MF–VN synaptic weight slows down to a large extent ( $dv/dt \approx 0$ ). Third, the flocculus is not necessary with this rule because the VN directly receives the error signal. But at least short-time VOR (Ito 2001) and HOKR (Shutoh et al. 2006) learning needs the flocculus.

### 3.2 Transfer of gain information with Hebbian plasticity

With Hebbian MF–VN plasticity, synaptic weights evolve according to Eqs. (8) and (13). In Appendix A, it is shown that the amount and the direction of plasticity in the PF–PC and MF–VN synapses sensitively depend on parameter values. Similar to the CF-driven MF–VN plasticity, memory transfer realized by the Hebbian MF–VN plasticity is not robust against parameter variation.

With Hebbian PC–VN plasticity, synaptic weights evolve according to Eqs. (16) and (17). A time course of the synaptic weights are shown in Fig. 3(a) for  $R>R_0$ . The two nullclines (thin solid lines) do not cross each other, and the synaptic weights diverge, which is unrealistic. When  $R<R_0$  (Fig. 3(c)), the trajectory converges onto the fast nullcline quickly and moves toward an equilibrium. However, the two nullclines have similar curvatures, and the position of the equilibrium is sensitive to parameter values for the same reason as that for the CF-driven and Hebbian MF–VN plasticity rules. Both for  $R>R_0$  and  $R<R_0$ , memory transfer does not occur in a robust manner (Fig. 3(b) and (d)). An intuitive explanation is as follows. After a sufficient learning period, the VN signal stabilizes at a steady level with a steady strength of rebound activities for which LTP and LTD balance. In this steady state, the VN receives the corresponding amount of inhibition ( $by=b(\mathbf{w}'\mathbf{x}+y_0)$ ) from the PC. The steady state is modulated by  $b(\mathbf{w}'\mathbf{x}+y_0)$ . The PF–PC synaptic plasticity is guided by the CF signal  $e=R-z=R-\mathbf{v}'\mathbf{u}+b(\mathbf{w}'\mathbf{x}+y_0)-z_0$ , in which  $b(\mathbf{w}'\mathbf{x}+y_0)$  is the only adjustable quantity again. Note that the MF–VN synaptic weight  $\mathbf{v}$  is fixed. The net inhibition  $b(\mathbf{w}'\mathbf{x}+y_0)$  that stabilizes the PF–PC synaptic weight does not generally stabilize the PC–VN synaptic weight and vice versa. Even if the equilibrium happens to exist as in Fig. 3(d), it is not robust against parameter variation.

**Fig. 3** Memory transfer with the CF-driven PC–VN plasticity. The target gain  $R=2$  (**a, b**) and  $R=0.5$  (**c, d**). Time courses of the synaptic weights (**a, c**) and the memory stored in each synapse (**b, d**) are shown. See the caption of Fig. 2 for legends



### 3.3 Transfer of gain information with PC-driven plasticity

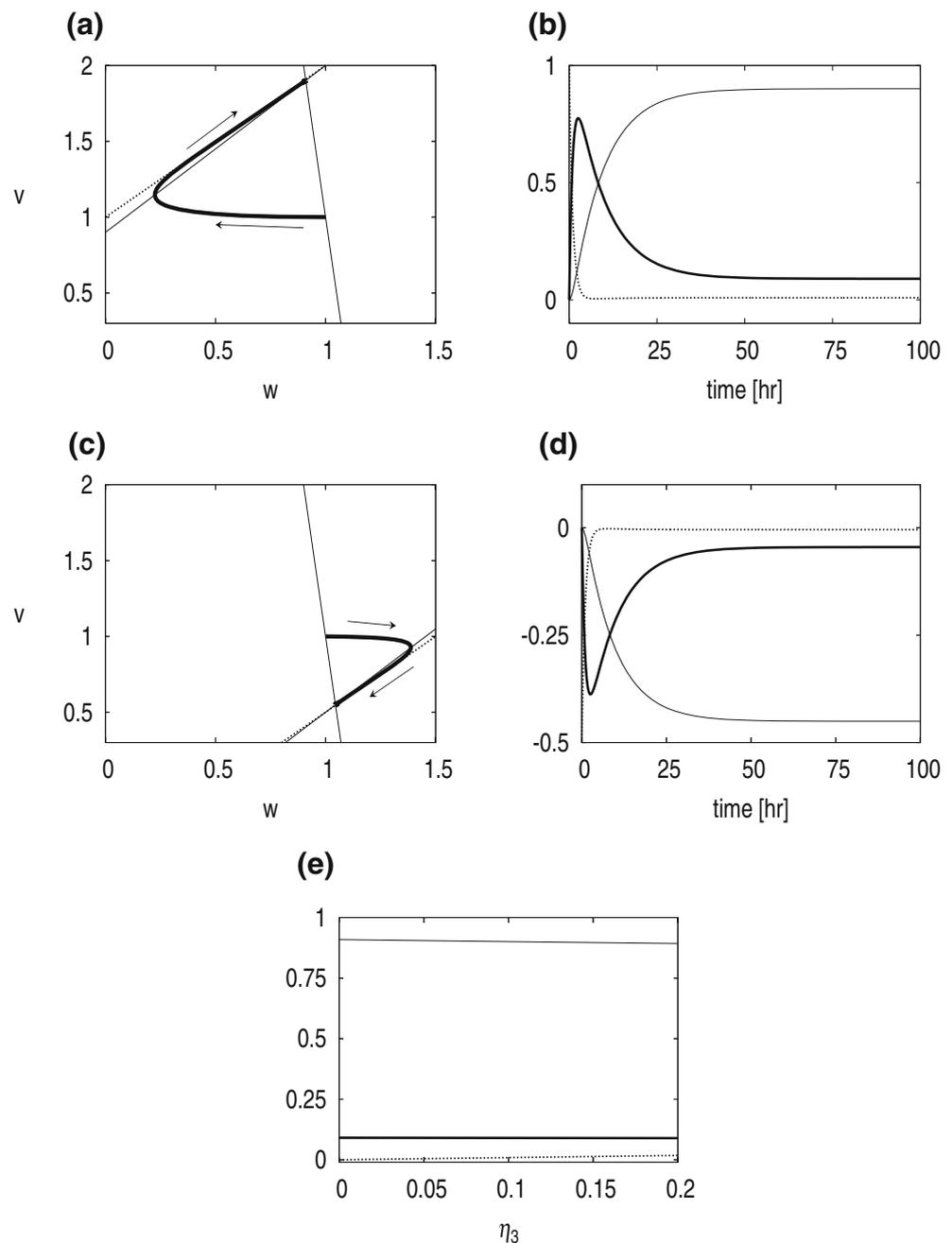
With PC-driven MF–VN plasticity, typical synaptic time courses are drawn in Fig. 4(a) and (c) for  $R>R_0$  and  $R<R_0$ , respectively. The corresponding time courses of the amount of memory in each synapse are shown in Fig. 4(b) ( $R>R_0$ ) and Fig. 4(d) ( $R<R_0$ ). This rule generates robust memory transfer. In contrast to the CF-driven and Hebbian plasticity rules, the slow nullcline has a large negative slope and is separated from the fast nullcline (thin solid lines in Fig. 4(a,c)). Consequently, when the equilibrium (crossing of two thin solid lines) is reached after learning,  $R>R_0$  ( $R<R_0$ ) duly results in LTD (LTP; see Sections 2.1 and 2.2 for the interpretation of this LTP as LTD of another set of PF–PC synapses) of the PF–PC synaptic weight  $w$  and LTP (LTD) of the MF–VN synaptic weight  $v$ . At the same time, initial LTD (LTP) of  $w$  is partially compensated by subsequent LTP (LTD) of  $w$ . The memory transfer is robust because the crossing of two nullclines does not move much when the value of a learning rate is slightly changed. Indeed, Fig. 4(e) shows that the degree of memory transfer hardly depends on  $\eta_3$  in contrast to the results for the CF-driven plasticity (Fig. 2(g)). Similar robustness with respect

to other parameters is guaranteed by mathematical results in Appendix A (Eqs. (37) and (38)). In fact, the amount of memory eventually stored in the PF–PC synapse is small:  $O(\eta_6/\eta_4)$  (Eq. (37) in Appendix A), whereas the amount moved to the MF–VN synapse is large:  $O(1)$  (Eq. (38)). A byproduct of the two separate nullclines is that transfer happens more rapidly than for the CF-driven and Hebbian rules.

With PC-driven PC–VN plasticity, the synaptic weights evolve as represented by thick solid lines in Fig. 5(a) ( $R>R_0$ ) and Fig. 5(c) ( $R<R_0$ ) (see Appendix B for details). The PF–PC synaptic weight  $w$  and the PC–VN synaptic weight  $b$  both experience net LTD (LTP) but on different timescales when  $R>R_0$  ( $R<R_0$ ). As shown in Fig. 5(b) and (d), error is reduced (dotted lines) and the memory is transferred from the PF–PC synapse (thick solid lines) to the PC–VN synapse (thin solid lines). Because the fast and slow nullclines (thin solid lines in Fig. 5(a,c)) are sufficiently far from each other, memory transfer is robust against parameter variation, as is the case with the PC-driven MF–VN plasticity.

We conclude that the PC-driven rules, but not the CF-driven or the Hebbian rules, provides robust mechanisms of

**Fig. 4** Memory transfer with the PC-driven MF–VN plasticity. The target gain  $R=2$  (**a, b**) and  $R=0.5$  (**c, d**). Time courses of the synaptic weights (**a, c**) and the memory stored in each synapse (**b, d**) are shown. **(e)** The effect of modulating  $\eta_3$  for gain-up learning. In contrast to CF-driven MF–VN plasticity (Fig. 3(g)), the amounts of memory stored in the PF–PC (thick solid line) and MF–VN (thin solid line) synapses are robust against the parameter variation. See the caption of Fig. 2 for legends



memory transfer. Our results confirm and extend the prediction of schematic models (Lisberger 1988; Perrett and Mauk 1995; Raymond et al. 1996; Mauk 1997) and numerical simulations (Medina and Mauk 1999; Medina et al. 2000, 2001; Peterson et al. 1991), which focused on MF–nucleus plasticity.

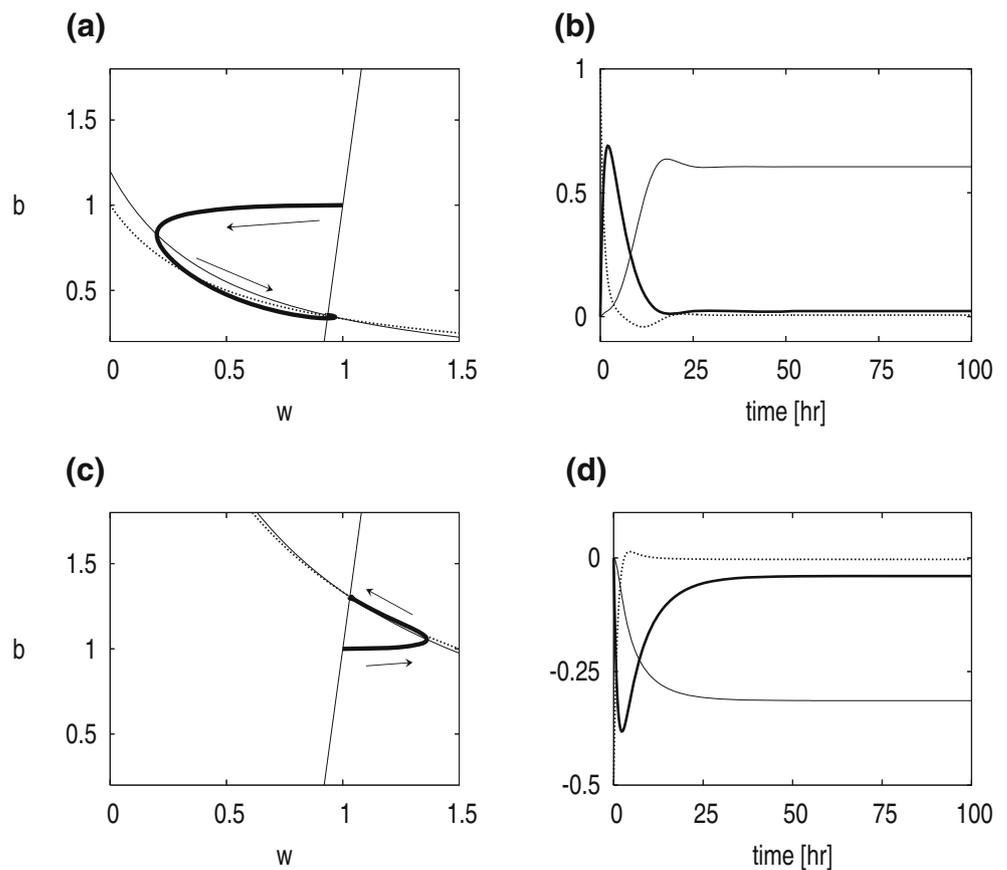
### 3.4 Transfer of gain and phase information

Proper response timing as well as gain has to be learned in VOR adaptation. Consequently, we incorporate learning of

timing, or equivalently phase, of the response. Now, the desired output is  $R \sin(\omega t + \theta)$  with nonzero  $\omega$ . We assume that the stimulation-induced neuronal responses are sinusoidal ( $f(a) = \sin(a)$ ), so that the circuit output may acquire the desired output. The firing rates of MFs and PFs are respectively  $u_i(t) = \sin(\omega t + \psi_i)$  and  $x_i(t) = \sin(\omega t + \phi_i)$ . We consider MF–VN plasticity because its analysis is tractable compared to PC–VN plasticity.

A successful transfer of the gain and the phase to the VN is schematically shown in Fig. 6 for  $R_0=0$ ,  $R=1$ . MFs and PFs generate sinusoidal firing rates with dispersed phase

**Fig. 5** Memory transfer with the PC-driven PC–VN plasticity. The target gain  $R=2$  (**a, b**) and  $R=0.5$  (**c, d**). Time courses of the synaptic weights (**a, c**) and the memory stored in each synapse (**b, d**) when PC–VN synapses store permanent memory. The error-free line (dotted lines in **(a)** and **(c)**) is expressed as  $b(w+y_0)=v_0-R=0$ . See the caption of Fig. 2 for legends



elements as shown in Fig. 6(a). Each trace corresponds to the signal of one fiber. In a very early stage of learning, the PC signal (thick solid line in Fig. 6(b)) is adapting to the desired output (thick dashed line) by PF–PC synaptic plasticity. Because the PC inhibit the VN, the PC signal has the sign opposite to the desired output. The effective contribution of the PC to the circuit output is shown by the thin dashed line. The MF–VN synapses are plastic on a slower timescale. Thus the postsynaptic input to the VN directly from the MFs, which is controlled by the MF–VN synaptic weights, is tiny (thin solid line). The circuit output  $z$  (thin dashed line) is approximately equal to the sign-flipped PC signal. After some time, the PF–PC synapses maintain most of the VOR memory, as is represented by the enhanced PC signal (thick solid line in Fig. 6(c)). The difference between the desired output (thick dashed line) and  $z$  (thin dashed line) has become small. The MF–VN contribution to  $z$  is growing but still small (thin solid line). After sufficient time, the MF–VN synaptic contribution (thin solid line in Fig. 6(d)) accounts for a large portion of the circuit output, whereas the PF–PC synaptic contribution decays (thick solid line). The error has been small since much earlier time (Fig. 6(c) and (d)). If

$\Delta\psi < \pi$ , the variety of phase leads generated by MFs is limited. Then the MF afferent to the VN may not be timed to the phase of the desired output so that the phase acquired by the MF–VN synapses may deviate from the desired phase (imagine an arbitrary horizontal shift of the thin solid line in Fig. 6(d)). We analyze the effect of  $\Delta\psi$  on memory transfer.

As shown in Appendix C (Eq. (45)), we can get away the effect of  $\omega$  by time averaging, which is justified because synaptic plasticity occurs much more slowly than the modulation of the vestibular signal parameterized by  $\omega$ . The amount of permanent memory stored in the MF–VN synapses quantified by  $(\mathbf{v}^t - \mathbf{v}_0^t)\mathbf{u}$  (see Section 2.6) is reexpressed as  $(\mathbf{v}^t - \mathbf{v}_0^t)\mathbf{u} \equiv r_D (R - R_0) \sin(\omega t + \theta_D)$  by using a suitable gain  $r_D$  and phase  $\theta_D$ . If  $\theta_D$  is close to  $\theta$ , memory transfer is achieved properly, and  $r_D$  represents the portion of the target gain acquired by the MF–VN synapses. Similarly, the portion of the memory remaining in the PF–PC synapses is expressed as  $(-\mathbf{w}^t + \mathbf{w}_0^t) \mathbf{x} \equiv r_I (R - R_0) \sin(\omega t + \theta_I)$  by suitable  $r_I$  and  $\theta_I$ .

Detailed calculations in Appendix C conclude that the CF-driven and Hebbian MF–VN plasticity rules do not support robust memory transfer of not only the target gain,

which was analyzed before, but also the target phase. Therefore, we focus on the PC-driven MF–VN plasticity.

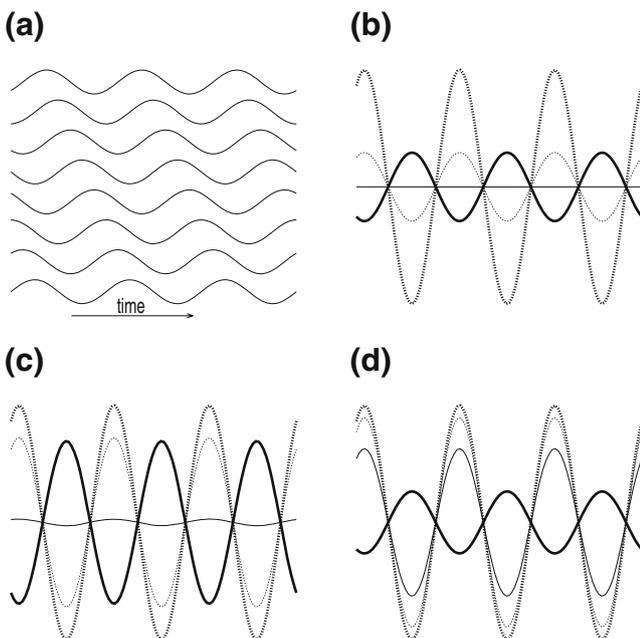
With the PC-driven rule, the calculations in Appendix C show

$$r_D = \sqrt{\frac{\cos^2 \theta}{\left[1 + \frac{4\eta_1\eta_6 + 16\eta_3\eta_6}{\eta_1\eta_4 A_c(\Delta_\psi)}\right]^2} + \frac{\sin^2 \theta}{\left[1 + \frac{4\eta_1\eta_6 + 16\eta_3\eta_6}{\eta_1\eta_4 A_s(\Delta_\psi)}\right]^2}}, \quad (23)$$

$$r_I = 4\eta_1\eta_6 \sqrt{\frac{\cos^2 \theta}{[\eta_1\eta_4 A_c(\Delta_\psi) + 4\eta_1\eta_6 + 16\eta_3\eta_6]^2} + \frac{\sin^2 \theta}{[\eta_1\eta_4 A_s(\Delta_\psi) + 4\eta_1\eta_6 + 16\eta_3\eta_6]^2}}, \quad (25)$$

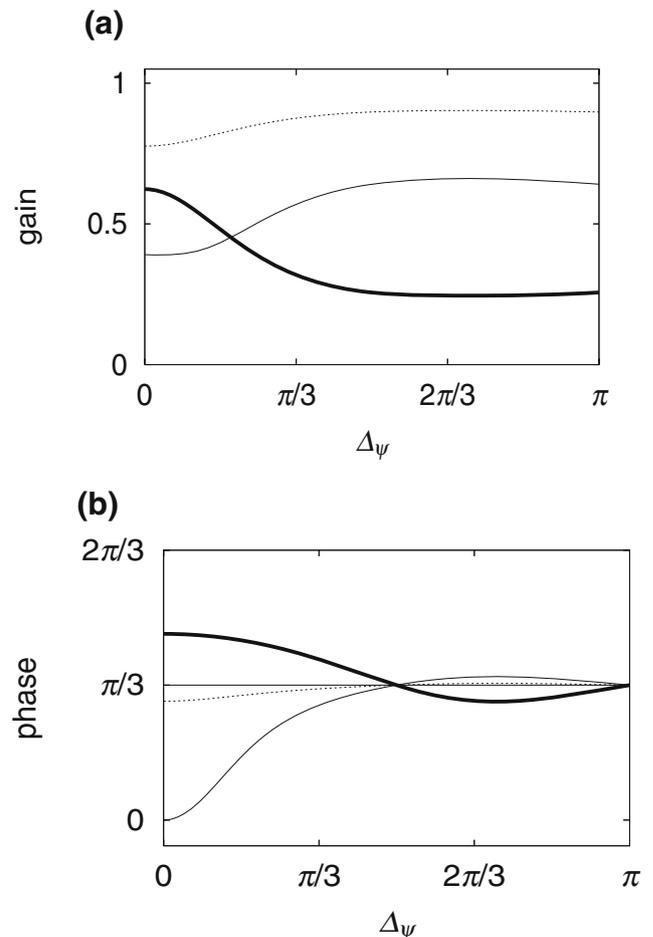
$$\theta_I = \tan^{-1} \left[ \frac{\eta_1\eta_4 A_c(\Delta_\psi) + 4\eta_1\eta_6 + 16\eta_3\eta_6}{\eta_1\eta_4 A_s(\Delta_\psi) + 4\eta_1\eta_6 + 16\eta_3\eta_6} \tan \theta \right], \quad (26)$$

where  $A_c(\Delta_\psi) = 1 + \sin 2\Delta_\psi / 2\Delta_\psi$  and  $A_s(\Delta_\psi) = 1 - \sin 2\Delta_\psi / 2\Delta_\psi$ . As a function of  $\Delta_\psi$ , these four variables together with the gain and the phase of the circuit output  $z$  are plotted in Fig. 7(a) and (b). We set  $R_0 = 1$ ,  $R = 2$ , and  $\theta = \pi/3$ . When the phase leads generated by MFs are dispersed enough to cover the desired phase lead ( $\Delta_\psi \geq \theta = \pi/3$ ), the



**Fig. 6** Schematics of simultaneous gain and phase learning. MFs and PFs generate periodic signals with various phases as shown in (a). The target output (thick dotted line), the circuit output (thin dashed line), the PC signal (thick solid line), and the total postsynaptic input from the MFs to the VN (thin solid line) are shown for (b) very early, (c) early, and (d) late stages of learning

$$\theta_D = \tan^{-1} \left[ \frac{1 + \frac{4\eta_6(\eta_1 + 4\eta_3)}{\eta_1\eta_4 A_c(\Delta_\psi)}}{1 + \frac{4\eta_6(\eta_1 + 4\eta_3)}{\eta_1\eta_4 A_s(\Delta_\psi)}} \tan \theta \right], \quad (24)$$



**Fig. 7** Simultaneous gain and phase learning with the PC-driven MF–VN plasticity. The curves are obtained using Eqs. (23–26). The gain of the circuit output additional to the baseline level  $R_0$  (dotted lines),  $r_D$  (thick solid lines), and  $r_I$  (thin solid lines) are shown in (a). The phase of the circuit output (dotted lines),  $\theta_D$  (thick solid lines), and  $\theta_I$  (thin solid lines) are shown in (b). The target gain and the target phase are  $R = 2$  and  $\theta = \pi/3$ , respectively

desired gain and phase are first acquired by the PF–PC synapses and then transferred to the MF–VN synapses, as depicted in Fig. 6. Eventually, the signal via the MF–VN pathway characterized by  $r_D$  (thin solid line in Fig. 7(a)) and  $\theta_D$  (thin solid line in Fig. 7(b)) approximates the desired output. Note that  $r_D=1$  with  $\theta_D=\theta$  implies perfect memory transfer. Substituting  $\Delta_\psi=\pi$  into Eqs. (24) and (26) leads to  $\theta_D=\theta$  and  $r_D=\eta_1\eta_4/(\eta_1\eta_4+4\eta_1\eta_6+16\eta_3\eta_6)$ ; the phase transfer is perfect in this case. Because we assumed  $\eta_1\gg\eta_3$ ,  $\eta_4\gg\eta_6$ , the gain information is also transferred sufficiently ( $r_D\cong 1$ ,  $r_I=O(\eta_6/\eta_4)\ll 1$ ; see Appendix C). Transfer is possible because transient information on  $R$  and  $\theta$ , which is stored in the PF–PC synapses and expressed by the PC signal ( $r_I$  and  $\theta_I$ ), serves as an effective teacher for learning of MF–VN synapses.

If  $\Delta_\psi<\theta$ , the target response phase  $\theta$  cannot be generated by the MF afferents to the VN. Then, transfer of the phase information to the MF–VN synapses will be deteriorated. Indeed,  $\theta_D$  deviates from  $\theta$  when  $\Delta_\psi$  is small (thin solid line in Fig. 7(b)). Mathematically, if  $A_s(\Delta_\psi)=O(\eta_3\eta_6/\eta_1\eta_4)$ , namely,  $\Delta_\psi=O((\eta_3\eta_6/\eta_1\eta_4)^{1/2})$ ,  $\theta_D$  is not close to  $\theta$ . Particularly, as  $\Delta_\psi$  tends to zero,  $r_D=\cos\theta$  and  $\theta_D=0$  are approached. When  $\Delta_\psi$  is small, the PC signal, which is plastic owing to the PF–PC synapses, rescues this mismatch. The gain information ( $r_I$ ) and the phase ( $\theta_I$ ) acquired by the PF–PC synapses are plotted by thick solid lines in Fig. 7(a) and (b), respectively. When  $\Delta_\psi$  is not large ( $\Delta_\psi<\pi/2$ ), Eqs. (24) and (26) assure  $\theta_I>\theta>\theta_D$ . In this case, the PF–PC synapses learn to compensate the phase lag caused by insufficient phase learning by the MF–VN synapses. As  $\Delta_\psi$  decreases,  $\theta_D$  decreases, and  $\theta_I$  increases, which implies pronounced phase compensation.

In summary, the analysis of simultaneous transfer of the gain and phase information adds another support that the PC-driven synaptic plasticity is more preferable to the CF-driven and Hebbian plasticity.

## 4 Discussion

### 4.1 Summary

Our analysis illuminated how the flocculus storing transient memory and the VN, which is a putative site of permanent memory, can cooperate in VOR learning. For robust memory transfer with plastic MF–VN synapses, plasticity of these synapses must depend on negatively correlated activities of PCs and MF input to the VN. These results are consistent with experimental (Perrett et al. 1993) and numerical (Medina and Mauk 1999; Medina et al. 2000) studies of eyelid conditioning and with in vitro recording (Pugh and Raman 2006). We also showed that PC–VN synapses instead of MF–VN synapses may store permanent memory

if LTP occurs when PCs and MFs are simultaneously active. The role of PC–VN synapses in VOR learning has been neglected in previous literature (Medina and Mauk 1999; Kassardjian et al. 2005). We also briefly examined the extended model in which the target gain is transferred to the VN, whereas timing information may remain in the cerebellar cortex. We showed that PC-driven MF–VN plasticity, which turned out to be successful in gain learning, is also consistent with this scenario. Repetitively used information such as the response gain may be robustly transferred to the nuclei, whereas more specific (and hence less frequently repeated) information such as response timing may remain in PF–PC synapses.

Strength of the current work is given by its analytical basis. Using simple firing-rate models and qualitative criteria, we were able to select plausible synaptic mechanisms of memory consolidation. In the context of eyelid conditioning, Medina, Mauk and coworkers presented numerical models to explain memory transfer (Medina and Mauk 1999; Medina et al. 2000, 2001). They concluded that the PC-driven MF–VN plasticity is preferable to the CF-driven and Hebbian rules. In their models, the learning rates determine equilibrium synaptic weights and seem to control the possibility of memory transfer. Values of the learning rates are not necessarily accessible in experiments. Our model uses the ratios of the learning rates and the values of other model parameters just for qualitative purposes. With robustness against parameter variation used as the model selection criterion, fast–slow analysis led to the conclusion that the PC-driven plasticity in the MF–VN or PC–VN synapses is the best. Also from a memory capacity perspective, our gain-and-phase model opts for the PC-driven synaptic plasticity. Suppose the MF–VN plasticity. With the PF–PC synaptic weights  $\mathbf{w}$  fixed after transient learning has completed, the input to the VN is  $-y=-\mathbf{w}^T\mathbf{x}-y_0=R\sin(\omega t+\theta)-\mathbf{v}^T\mathbf{u}-z_0$ . This is the error signal for the VN because the VN output without eventual contribution of the PC signal is equal to  $\mathbf{v}^T\mathbf{u}+z_0$ , which should be equal to  $R\sin(\omega t+\theta)$  after perfect memory transfer. Therefore,  $-y$  guides error-correcting learning in the VN, as the CF signal  $e$  does for PF–PC synapses. The same argument applies to PC–VN plasticity as well.

We neglected noise. Actually, the PC-driven plasticity rules are also preferred to others in terms of robustness against dynamical noise, which is a byproduct of the fast–slow analysis. Dynamical noise plays a role similar to random perturbation of model parameters. As a result, parameter-sensitive steady synaptic weights, which originate from an undesirable plasticity rule, are also sensitive to noise. By contrast, for the parameter-insensitive plasticity rules, noise does not alter the steady synaptic weights very much.

In the absence of the error signal from the CF, the memory that the model has acquired will fade away, which

agrees with the behavioral results (Shutoh et al. 2006). This type of forgetting occurs no matter whether noise is present or not, and its rate is determined by the decay rate of the VN synaptic weights.

The model is limited in several aspects. First, our model is a linear rate code model. We discarded detailed time courses of neural responses except that the firing rate of each neuron can be modulated sinusoidally with phase shifts. Actually, PCs produce simple and complex spikes. Nuclear neurons also own intricate firing modes such as rebound depolarization (Aizenman et al. 1998; Aizenman and Linden 1999; Ouardouz and Sastry 2000; Pugh and Raman 2006), which may be also true for VN neurons. Transient firing rates may be relevant to cerebellar memory (Lisberger and Sejnowski 1992). Learning that depends on such spiking or transient properties is beyond the scope of our model. Second, we assumed that the MFs create a family of phase leads. Evidence suggests that the PFs produce timing elements as we assumed (Perrett et al. 1993), but timing specificity of MFs projecting to the VN is not well known. Our model is obviously an oversimplification in this regard. Third, the vestibular input was assumed to be a single sinusoid with a sufficiently large frequency. Vestibular inputs composed of multiple sinusoids could lead to crosstalk of different frequency components. A slow sinusoidal input would make the target gain depending on the input frequency because of the elasticity of the oculomotor plant. Implementing these detailed features is warranted for future work.

#### 4.2 Implications for savings

A part of the acquired memory may remain over time to facilitate learning at a later time. This phenomenon is termed savings. Let us suppose that the training period and the resting period alternate with a total period of one day. The amount of error decreases during training periods, but it increases during the rest periods in which the error signal through CFs is absent. Performance improves more rapidly and progressively in later training sessions because the memory partially remains even after a rest period. This enhancement continues until the performance saturates (Shutoh et al. 2006). The cerebellar cortex may be responsible for transient learning and forgetting, and construction of permanent memory in the nucleus may underlie savings (Medina et al. 2001; Masuda and Amari 2006). Our model is capable of savings. Memory transfer occurs in both training and resting periods, which is different from sleep learning. In sleep learning, motor skills improve through sleeping, and contents and stages of

sleep influence the degree of memory consolidation and reconsolidation (Stickgold et al. 2001; Walker et al. 2003). In our framework, memory consolidation can proceed in the awake and sleep states alike.

#### 4.3 Memory capacity of two pathways

The capacity of permanent versus transient memory sites in VOR is an unresolved issue. The indirect pathway composed of GCs, PFs, and PCs has been thought to be computationally more powerful than the direct pathway via MF–VN synapses because of a large number of GCs and PFs:  $n \cong 1.5 \times 10^7$  (Marr 1969; Albus 1971). Such fan-out-fan-in architecture is not found in the direct pathway. The number of vestibular ganglion cells is 1650 in monkeys (Nagao et al. 1997) and 190 in rats (Osanaï et al. 1999). Only 5% (monkey) to 10% (rats) among them monosynaptically project to the VN (S. Nagao, private communication), which implies  $m = 20\text{--}80$ .

The two pathways differ not only in the number of neurons but also in the complexity of circuits. Local feedforward and feedback loops involving inhibitory interneurons are abundant in the indirect pathway (Ito 2001). To our current knowledge, the direct pathway is devoid of such loops, which may limit its representation ability. In experiments (Perrett et al. 1993) and simulation studies (Peterson et al. 1991; Raymond et al. 1996; Mauk and Donegan 1997; Medina et al. 2000) of eyelid conditioning and VOR learning, the timing and the context information are not transferred to the cerebellar nuclei, whereas the gain information is. This suggests that the VN may consolidate the information on the VOR gain, whereas the cerebellar cortex may store timing information.

As a preliminary step to address this issue, we translated the potentially different memory capacity of the two pathways into the difference in the diversity of response timing of the neurons. Specifically, the PFs were assumed to create any phase elements, whereas MF collaterals to the VN were assumed to create constrained phase elements.

#### 4.4 Site of permanent memory

Earlier numerical work (Mauk 1997; Mauk and Donegan 1997; Medina and Mauk 1999; Medina et al. 2000, 2001; Peterson et al. 1991) suggests that permanent memory converges to the MF-nucleus synapses. These synapses indeed show LTP (Racine et al. 1986; Pugh and Raman 2006) and LTD (Zhang and Linden 2006). Based on our theory, the essential factor is that permanent learning is driven by concurrent PC and MF activities. Some electrophysiological (Aizenman and Linden 2000; Shutoh et al.

2006) and anatomical (Kleim et al. 2002) evidence on nucleus plasticity is not specific about which synapses learn. The synapses connecting PC and the deep cerebellar nucleus also undergo LTP (Aizenman et al. 1998; Ouardouz and Sastry 2000) and LTD (Morishita and Sastry 1996; Aizenman et al. 1998). These data suggest a postsynaptic origin of PC–VN plasticity. Our theory predicts that memory transfer with PC–VN plasticity is robust with the PC-driven plasticity. However, it is not robust with the Hebbian plasticity, which is more closely tied to the postsynaptic plasticity mechanism observed in *in vitro* recordings.

The permanent memory capacity is limited by the number of MF–VN or PC–VN synapses. The number of MF–VN synapses was estimated to be  $m=20–80$  in rats and monkeys. To roughly evaluate the number of PC–VN synapses, we disregard the animal specificity and note that there are about 700 PCs in the flocculus of mice (S. Nagao, private communication). The ratio of the number of PCs to that of the VN neurons is 26:1 in cats (Palkovits et al. 1977), and simple application of this ratio provides a rough estimate of 30 VN neurons. If a majority of the PCs is responsible for VOR adaptation, the number of involved synapses (from 700 PC cells to 30 VN cells) is much larger than the number of the MF–VN synapses (from 80 MFs to 30 VNs). Then, PC–VN synapses would have a larger storage than MF–VN synapses do. Combined with behavioral experiments addressing the memory capacity, such anatomical consideration may help understand the locus of permanent memory.

#### 4.5 Memory transfer beyond cerebellar motor learning

Parallelism of transient and permanent memory is a general organizing principle throughout the brain. A human imaging study shows that the cerebellum is involved in early stages of motor sequence learning. Instead of the cerebellar nuclei, the basal ganglia and the frontal lobe are involved in permanent learning. Furthermore, delayed recall requires activation of the primary motor cortex, the premotor cortex, and the parietal cortex (Penhune and Doyon 2002, 2005). By contrast, the primary motor cortex of humans is involved in early but not late stages in skill learning (Muellbacher et al. 2002). Then, the motor memory has to be transferred to somewhere else for consolidation. Another example is the basal ganglia pathway. In motor sequence learning of monkeys, the anterior caudate nucleus and the anterior putamen contribute to early learning, and the middle-posterior putamen to longer memory (Miyachi et al. 1997, 2002). In the saccadic task, memory of rewarded association tasks is first stored in the basal ganglia, and its signal is used as an effective

teacher for slower learning sites in the frontal cortex (Pasupathy and Miller 2005). Another famous example is the interaction between the cerebral cortex and the hippocampus. Hippocampus is believed to store transient memory, and only the important information is reinstated in the cerebral cortex (McClelland et al. 1995). Also, memory transfer is found in fear conditioning in the amygdala (Repa et al. 2001; Medina et al. 2002). Even though we emulated VOR learning, our framework is general in the sense that we adopted a linear rate coding model and fast–slow analysis. Adequate modifications of our model may be useful for understanding different neural systems in which transient memory and permanent memory are intertwined.

**Acknowledgments** We thank S. Nagao, T. Knöpfel, and C. Rockland for helpful discussions and carefully reading the manuscript. This work is supported by Special Postdoctoral Researchers Program of RIKEN.

#### Appendix A: Fast–slow analysis with static MF and PF firing and MF–VN plasticity

When the vestibular signal is static, we set  $\omega = 0$ ,  $\theta = \pi/2$ , and hence set the vestibular signal  $\sin(\omega t + \theta) = 1$ . Given that the MF and PF firing rates are static with  $f(a) = 1$ , we have  $\langle \sin(\omega t + \theta) \mathbf{u} \rangle_i = \langle \sin(\omega t + \theta) \mathbf{x} \rangle_i = \langle \mathbf{u} \rangle_i = \langle \mathbf{x} \rangle_i = \langle \mathbf{u} \mathbf{x}^t \rangle_{i,j} = \langle \mathbf{x} \mathbf{u}^t \rangle_{j,i} = \langle \mathbf{x} \mathbf{x}^t \rangle_{i,j} = 1$ . Here,  $\langle \mathbf{u} \rangle_i$ , for example, is time average of the  $i$ -th MF signal. Because the PC–VN synapse is assumed to be static under MF–VN plasticity, we set  $b = 1$ . Let us assume  $m = n = 1$  (and quit bold notations) and perform fast–slow analysis. Similar analysis works for general  $m$  and  $n$ .

In an early stage of learning, the PF–PC synapse  $w$  evolves much faster than the MF–VN synapse  $v$  does. The fast nullcline defined by setting  $dw/dt = 0$  in Eq. (8) is given by

$$v = v_0 + R - R_0 + \frac{\eta_1 + \eta_3}{\eta_1} (w - w_0). \tag{27}$$

On a short timescale,  $w$  and  $v$  converge onto this line.

##### A.1 CF-driven MF–VN plasticity

For the CF-driven learning, the slow nullcline to which  $w$  and  $v$  converge in a long run is given by setting  $dv/dt = 0$  in Eq. (11):

$$v = v_0 + \frac{\eta_4}{\eta_4 + \eta_6} (R - R_0) + \frac{\eta_4}{\eta_4 + \eta_6} (w - w_0). \tag{28}$$

A trajectory of the synaptic weights in the  $w$ – $v$  plane approaches somewhere on the fast nullcline (Eq. (27)) and

then slides along it toward the equilibrium obtained as the crossing of the two nullclines. The crossing is given by

$$w^* = w_0 - \frac{\eta_1 \eta_6 (R - R_0)}{\eta_1 \eta_6 + \eta_3 \eta_4 + \eta_3 \eta_6}, \tag{29}$$

$$v^* = v_0 + \frac{\eta_3 \eta_4 (R - R_0)}{\eta_1 \eta_6 + \eta_3 \eta_4 + \eta_3 \eta_6}. \tag{30}$$

The error at the equilibrium is

$$e^* = R - v^* + w^* + y_0 - z_0 = \frac{\eta_3 \eta_6 (R - R_0)}{\eta_1 \eta_6 + \eta_3 \eta_4 + \eta_3 \eta_6}, \tag{31}$$

which is small given  $\eta_1 \gg \eta_3$ .

### A.2 Hebbian MF–VN plasticity

For the Hebbian learning, the  $w$ -nullcline is given by Eq. (27), and the  $v$ -nullcline derived from Eq. (13) becomes

$$v = v_0 + \frac{\eta_4}{\eta_4 - \eta_6} (w - w_0). \tag{32}$$

The equilibrium is given by

$$w^* = w_0 + \frac{\eta_1 (\eta_4 - \eta_6) (R - R_0)}{\eta_1 \eta_6 - \eta_3 \eta_4 + \eta_3 \eta_6}, \tag{33}$$

$$v^* = v_0 + \frac{\eta_1 \eta_4 (R - R_0)}{\eta_1 \eta_6 - \eta_3 \eta_4 + \eta_3 \eta_6}, \tag{34}$$

$$e^* = \frac{\eta_3 (-\eta_4 + \eta_6) (R - R_0)}{\eta_1 \eta_6 - \eta_3 \eta_4 + \eta_3 \eta_6}. \tag{35}$$

Because the relative magnitudes of  $\eta_1 \eta_6$  and  $\eta_3 \eta_4$  are indecisive, the sign of  $\eta_1 \eta_6 - \eta_3 \eta_4 + \eta_3 \eta_6$  is indefinite.

### A.3 PC-driven MF–VN plasticity

For the PC-driven learning, the slow nullcline derived from Eq. (15) becomes

$$v = v_0 - \frac{\eta_4}{\eta_6} (w - w_0), \tag{36}$$

and the equilibrium is given by

$$w^* = w_0 - \frac{\eta_1 \eta_6 (R - R_0)}{\eta_1 \eta_4 + \eta_1 \eta_6 + \eta_3 \eta_6}, \tag{37}$$

$$v^* = v_0 + \frac{\eta_1 \eta_4 (R - R_0)}{\eta_1 \eta_4 + \eta_1 \eta_6 + \eta_3 \eta_6}, \tag{38}$$

$$e^* = \frac{\eta_3 \eta_6 (R - R_0)}{\eta_1 \eta_4 + \eta_1 \eta_6 + \eta_3 \eta_6}. \tag{39}$$

## Appendix B: Fast–slow analysis for PC-driven PC–VN plasticity

For PC-driven PC–VN plasticity, we can analytically obtain the equilibrium solutions. The initial synaptic weights are  $w=w_0$ ,  $v=v_0$ , and  $b=b_0$ , which are chosen to yield the initial gain  $R_0$ . This condition together with Eqs. (16) and (18) provides the relations:

$$R_0 - v_0 + b_0(w_0 + y_0) - z_0 = 0, \tag{40}$$

$$\eta_4(w_0 + y_0 - y_{ref}) - \eta_6 b_0 = 0. \tag{41}$$

Because we assumed  $\eta_1 \gg \eta_3$  and  $\eta_4 \gg \eta_6$ , there are two equilibria ( $w^*$ ,  $b^*$ ) given by

$$\left( w_0 - \frac{\eta_6 (R - v_0)}{\eta_4 (w_0 + y_0)}, \frac{v_0 - R}{w_0 + y_0} \right) \tag{42}$$

and

$$\left( -y_0 + \frac{\eta_6 (R - v_0 - z_0)}{\eta_4 (w_0 + y_0)}, b_0 - \frac{\eta_4 (w_0 + y_0)}{\eta_6} + \frac{R - v_0 - z_0}{w_0 + y_0} \right). \tag{43}$$

In both solutions, the final error is

$$e^* = \frac{\eta_6 (R - v_0 - z_0) (R - R_0)}{\eta_4 (w_0 + y_0)^2}, \tag{44}$$

which is small. We discard the second solution because it is unstable.

## Appendix C: Gain and phase learning with dynamic neural responses and MF–VN plasticity

With  $f(a)=\sin(a)$ , we obtain  $\langle \mathbf{x} \rangle = \langle \mathbf{u} \rangle = 0$ ,  $\langle \sin(\omega t + \theta) \mathbf{x} \rangle_i = \cos(\theta - \phi_i)/2$ ,  $\langle \mathbf{x} \mathbf{u}^t \rangle_{i,j} = \langle \mathbf{u} \mathbf{x}^t \rangle_{j,i} = \cos(\psi_j - \phi_i)/2$ , and  $\langle \mathbf{x} \mathbf{x}^t \rangle_{i,j} = \cos(\phi_j - \phi_i)/2$ . The signals with closer phase leads are more correlated. Then, Eq. (8) reads

$$\begin{aligned} \frac{dw_i}{dt} = & -\frac{\eta_1}{2} \left[ (R - R_0) \cos(\theta - \alpha_i) \right. \\ & - \sum_{j=1}^m \cos(\psi_j - \alpha_i) (v_j - v_{0j}) \\ & \left. + \sum_{j=1}^n \cos(\alpha_j - \alpha_i) (w_j - w_{0j}) \right] \\ & - \eta_3 n (w_i - w_{0i}). \end{aligned} \tag{45}$$

C.1 CF-driven MF–VN plasticity

With the CF-driven learning rule, Eq. (11) reads

$$\begin{aligned} \frac{dv_i}{dt} = & \frac{\eta_4}{2} \left[ (R - R_0) \cos(\theta - \psi_i) \right. \\ & - \sum_{j=1}^m \cos(\psi_j - \psi_i)(v_j - v_{0j}) \\ & \left. + \sum_{j=1}^n \cos(\alpha_j - \psi_i)(w_j - w_{0j}) \right] \\ & - \eta_6 m (v_i - v_{0i}). \end{aligned} \tag{46}$$

In terms of the order parameters defined by

$$W_c = \sum_{i=1}^n \cos \alpha_i (w_i - w_{0i}), \quad W_s = \sum_{i=1}^n \sin \alpha_i (w_i - w_{0i}),$$

$$V_c = \sum_{i=1}^m \cos \psi_i (v_i - v_{0i}), \quad V_s = \sum_{i=1}^m \sin \psi_i (v_i - v_{0i}),$$

the equilibrium for the desired output  $R \sin(\omega t + \theta)$  is obtained by solving

$$\begin{aligned} W_c^* = & -\frac{\eta_1}{2\eta_3 n} \left[ (R - R_0) \cos \theta \sum_{i=1}^n \cos^2 \phi_i \right. \\ & + (R - R_0) \sin \theta \sum_{i=1}^n \cos \phi_i \sin \phi_i \\ & + (W_c^* - V_c^*) \sum_{i=1}^n \cos^2 \phi_i \\ & \left. + (W_s^* - V_s^*) \sum_{i=1}^n \cos \phi_i \sin \phi_i \right], \end{aligned}$$

$$\begin{aligned} W_s^* = & -\frac{\eta_1}{2\eta_3 n} \left[ (R - R_0) \cos \theta \sum_{i=1}^n \cos \phi_i \sin \phi_i \right. \\ & + (R - R_0) \sin \theta \sum_{i=1}^n \sin^2 \phi_i \\ & + (W_c^* - V_c^*) \sum_{i=1}^n \cos \phi_i \sin \phi_i \\ & \left. + (W_s^* - V_s^*) \sum_{i=1}^n \sin^2 \phi_i \right], \end{aligned}$$

$$\begin{aligned} V_c^* = & \frac{\eta_4}{2\eta_6 m} \left[ (R - R_0) \cos \theta \sum_{i=1}^m \cos^2 \psi_i \right. \\ & + (R - R_0) \sin \theta \sum_{i=1}^m \cos \psi_i \sin \psi_i \\ & + (W_c^* - V_c^*) \sum_{i=1}^m \cos^2 \psi_i \\ & \left. + (W_s^* - V_s^*) \sum_{i=1}^m \cos \psi_i \sin \psi_i \right], \end{aligned}$$

$$\begin{aligned} V_s^* = & \frac{\eta_4}{2\eta_6 m} \left[ (R - R_0) \cos \theta \sum_{i=1}^m \cos \psi_i \sin \psi_i \right. \\ & + (R - R_0) \sin \theta \sum_{i=1}^m \sin^2 \psi_i \\ & + (W_c^* - V_c^*) \sum_{i=1}^m \cos \psi_i \sin \psi_i \\ & \left. + (W_s^* - V_s^*) \sum_{i=1}^m \sin^2 \psi_i \right]. \end{aligned}$$

The phase leads  $\varphi_i$  and  $\psi_i$  are assumed to be distributed uniformly on  $[0, 2\pi]$  and  $[-\Delta_\psi, \Delta_\psi]$ , respectively. Assuming that  $n$  and  $m$  are large, we have, for example,

$$\sum_{i=1}^n \cos^2 \alpha_i = \frac{n}{2\pi} \int_0^{2\pi} \frac{1 + \cos 2\alpha}{2\alpha} d\alpha = \frac{n}{2}. \tag{47}$$

For notational convenience, we write  $A_c(\Delta_\psi) = 1 + \sin 2\Delta_\psi / 2\Delta_\psi$  and  $A_s(\Delta_\psi) = 1 - \sin 2\Delta_\psi / 2\Delta_\psi$ . We note that  $1 \leq A_c \leq 2$ , and  $A_s$  decreases in  $\Delta_\psi$ . Particularly,  $A_s \cong 2\Delta_\psi^2 / 3$  becomes small as  $\Delta_\psi \rightarrow 0$ . Then, the equilibrium is given by

$$V_\alpha^* = \frac{\eta_3 \eta_4 (R - R_0) \cos \theta A_\alpha(\Delta_\psi)}{\eta_1 \eta_6 + 4\eta_3 \eta_6 + \eta_3 \eta_4 A_\alpha(\Delta_\psi)}, \tag{48}$$

$$W_\alpha^* = -\frac{\eta_1 \eta_6 (R - R_0) \cos \theta}{\eta_1 \eta_6 + 4\eta_3 \eta_6 + \eta_3 \eta_4 A_\alpha(\Delta_\psi)}, \tag{49}$$

where  $\alpha = c$  or  $s$ . The amount of the memory stored in the MF–VN synapses is given by

$$\begin{aligned} & \sum_{i=1}^m \sin(\omega t + \psi_i)(v_i^* - v_{0i}^*) \\ & = V_c^* \sin \omega t + V_s^* \cos \omega t = r_D (R - R_0) \sin(\omega t + \theta_D), \end{aligned} \tag{50}$$

where

$$r_D = \sqrt{\frac{\cos^2 \theta}{\left[1 + \frac{\eta_1 \eta_6 + 4\eta_3 \eta_6}{\eta_3 \eta_4 A_c(\Delta_\psi)}\right]^2} + \frac{\sin^2 \theta}{\left[1 + \frac{\eta_1 \eta_6 + 4\eta_3 \eta_6}{\eta_3 \eta_4 A_s(\Delta_\psi)}\right]^2}}, \quad (51)$$

$$\theta_D = \tan^{-1} \left[ \frac{1 + \frac{\eta_1 \eta_6 + 4\eta_3 \eta_6}{\eta_3 \eta_4 A_c(\Delta_\psi)}}{1 + \frac{\eta_1 \eta_6 + 4\eta_3 \eta_6}{\eta_3 \eta_4 A_s(\Delta_\psi)}} \tan \theta \right]. \quad (52)$$

If MFs create any delay elements ( $\Delta_\psi = \pi$ ,  $A_c(\Delta_\psi) = A_s(\Delta_\psi) = 1$ ), Eq. (52) results in  $\theta_D = \theta$ , that is, the perfect phase learning by the MF–VN synapses. However,  $r_D = \eta_3 \eta_4 / (\eta_1 \eta_6 + \eta_3 \eta_4 + 4\eta_3 \eta_6)$  can be considerably smaller than the ideal value ( $= 1$ ) because  $\eta_1 \eta_6$  may be as large as  $\eta_3 \eta_4$  in general. The MF–VN synapses learn the desired gain only for a restricted parameter range, and gain transfer is not robust against parameter variation. This is in line with the result of the gain-only theory.

Furthermore, the discrepancy between  $\theta_D$  and  $\theta$  cannot be ignored when  $\Delta_\psi = O((\eta_1 \eta_6 / \eta_3 \eta_4)^{1/2})$  i.e. when  $\Delta_\psi$  is of the order of  $(\eta_1 \eta_6 / \eta_3 \eta_4)^{1/2}$ . Because  $\eta_1 \gg \eta_3$  and  $\eta_4 \gg \eta_6$ , one cannot tell without additional information whether  $\eta_1 \eta_6 \gg \eta_3 \eta_4$ ,  $\eta_1 \eta_6 \cong \eta_3 \eta_4$ , or  $\eta_1 \eta_6 \ll \eta_3 \eta_4$ . Because  $\eta_1 \eta_6 / \eta_3 \eta_4$  is not necessarily small, transfer can degrade even for a large  $\Delta_\psi$ .

## C.2 Hebbian MF–VN plasticity

With the Hebbian rule, Eq. (13) reads

$$\frac{dv_i}{dt} = \frac{\eta_4}{2} \left[ \sum_{j=1}^m \cos(\psi_j - \psi_i) (v_j - v_{0j}) - \sum_{j=1}^n \cos(\alpha_j - \psi_i) (w_j - w_{0j}) \right] - \eta_6 m (v_i - v_{0i}). \quad (53)$$

By solving the equilibrium in combination with Eq. (45), we have

$$V_\alpha^* = \frac{\eta_1 \eta_4 (R - R_0) \cos \theta A_\alpha(\Delta_\psi)}{-4\eta_3 \eta_4 A_\alpha(\Delta_\psi) + 4\eta_1 \eta_6 + 16\eta_3 \eta_6}, \quad (54)$$

$$W_\alpha^* = \frac{\eta_1 (R - R_0) \cos \theta [\eta_4 A_\alpha(\Delta_\psi) - 4\eta_6]}{-4\eta_3 \eta_4 A_\alpha(\Delta_\psi) + 4\eta_1 \eta_6 + 16\eta_3 \eta_6}, \quad (55)$$

which leads to

$$r_D = \sqrt{\frac{\cos^2 \theta}{\left[\frac{4\eta_1 \eta_6 + 16\eta_3 \eta_6}{\eta_1 \eta_4 A_c(\Delta_\psi)} - \frac{4\eta_3}{\eta_1}\right]^2} + \frac{\sin^2 \theta}{\left[\frac{4\eta_1 \eta_6 + 16\eta_3 \eta_6}{\eta_1 \eta_4 A_s(\Delta_\psi)} - \frac{4\eta_3}{\eta_1}\right]^2}}, \quad (56)$$

$$\theta_D = \tan^{-1} \left[ \frac{\frac{\eta_1 \eta_6 + 4\eta_3 \eta_6}{\eta_1 \eta_4 A_c(\Delta_\psi)} - \frac{\eta_3}{\eta_1}}{\frac{\eta_1 \eta_6 + 4\eta_3 \eta_6}{\eta_1 \eta_4 A_s(\Delta_\psi)} - \frac{\eta_3}{\eta_1}} \tan \theta \right]. \quad (57)$$

When MFs create any phase leads ( $\Delta_\psi = \pi$ ), Eq. (57) implies perfect transfer of the target phase ( $\theta_D = \theta$ ). However,  $r_D = \eta_1 \eta_4 / (4\eta_1 \eta_6 + 16\eta_3 \eta_6 - 4\eta_3 \eta_4)$ , derived from Eq. (56), is indefinite because  $4\eta_1 \eta_6 + 16\eta_3 \eta_6 - 4\eta_3 \eta_4$  can take an arbitrary value. Consequently, unrealistic phenomena such as overlearning ( $r_D > 1$ ) can arise in the model. Regarding phase learning, the error is prominent when  $\Delta_\psi$  is as small as  $\Delta_\psi = O((\eta_6 / \eta_4)^{1/2})$  (recall  $\eta_4 \gg \eta_6$ ), which is suitable.

## C.3 PC-driven MF–VN plasticity

With the PC-driven learning, Eq. (15) reads

$$\frac{dv_i}{dt} = -\frac{\eta_4}{2} \sum_{j=1}^n \cos(\alpha_j - \psi_i) (w_j - w_{0j}) - \eta_6 m (v_i - v_{0i}). \quad (58)$$

We derive

$$V_\alpha^* = \frac{\eta_1 \eta_4 (R - R_0) \cos \theta A_\alpha(\Delta_\psi)}{\eta_1 \eta_4 A_\alpha(\Delta_\psi) + 4\eta_1 \eta_6 + 16\eta_3 \eta_6}, \quad (59)$$

$$W_\alpha^* = -\frac{4\eta_1 \eta_6 (R - R_0) \cos \theta}{\eta_1 \eta_4 A_\alpha(\Delta_\psi) + 4\eta_1 \eta_6 + 16\eta_3 \eta_6}, \quad (60)$$

which in combination with Eq. (50) yields  $r_D$  (Eq. (23)) and  $\theta_D$  (Eq. (24)). The amount of the memory stored in the PF–PC synapses is represented by

$$-W_c^* \sin \omega t - W_s^* \cos \omega t = r_I (R - R_0) \sin(\omega t + \theta_I), \quad (61)$$

which yields  $r_I$  (Eq. (25)) and  $\theta_I$  (Eq. (26)).

## References

Aizenman, C. D., & Linden, D. J. (1999). Regulation of the rebound depolarization and spontaneous firing patterns of deep nuclear

- neurons in slices of rat cerebellum. *Journal of Neurophysiology*, 82, 1697–1709.
- Aizenman, C. D., & Linden, D. J. (2000). Rapid, synaptically driven increases in the intrinsic excitability of cerebellar deep nuclear neurons. *Nature Neuroscience*, 3, 109–111.
- Aizenman, C. D., Manis, P. B., & Linden, D. J. (1998). Polarity of long-term synaptic gain change is related to postsynaptic spike firing at a cerebellar inhibitory synapse. *Neuron*, 21, 827–835.
- Akemann, W., & Knöpfel, T. (2006). Interaction of Kv3 potassium channels and resurgent sodium current influences the rate of spontaneous firing of Purkinje neurons. *Journal of Neuroscience*, 26, 4602–4612.
- Albus, J. S. (1971). A theory of cerebellar function. *Mathematical Biosciences*, 10, 25–61.
- Armstrong, D. M., & Rawson, J. A. (1979). Activity patterns of cerebellar cortical neurons and climbing fibre afferents in the awake cat. *Journal of Physiology*, 289, 425–448.
- Boyden, E. S., Katoh, A., & Raymond, J. L. (2004). Cerebellum-dependent learning: The role of multiple plasticity mechanisms. *Annual Review of Neuroscience*, 27, 581–609.
- Coemans, M., Weber, J. T., De Zeeuw, C. I., & Hansel, C. (2004). Bidirectional parallel fiber plasticity in the cerebellum under climbing fiber control. *Neuron*, 44, 691–700.
- Dayan, P., & Abbott, L. F. (2001). Theoretical neuroscience—computational and mathematical modeling of neural systems. MIT.
- Davies, P., & Melvill Jones, G. (1976). An adaptive neural model compatible with plastic changes induced in the human vestibulo-ocular reflex by prolonged optical reversal of vision. *Brain Research*, 103, 546–550.
- De Schutter, E., & Bjaalie, J. G. (2001). Coding in the granular layer of the cerebellum. *Progress in Brain Research*, 130, 279–296.
- du Lac, S., Raymond, J. L., Sejnowski, T. J., & Lisberger, S. G. (1995). Learning and memory in the vestibulo-ocular reflex. *Annual Review of Neuroscience*, 18, 409–441.
- Fujita, M. (1982). Adaptive filter model of the cerebellum. *Biological Cybernetics*, 45, 195–206.
- Goldberg, J. M., & Fernandez, C. (1971). Physiology of peripheral neurons innervating semicircular canals of the squirrel monkey. I. Resting discharge and response to constant angular accelerations. *Journal of Neurophysiology*, 34, 635–660.
- Hansel, C., Linden, D. J., & D'Angelo, E. (2001). Beyond parallel fiber LTD: the diversity of synaptic and nonsynaptic plasticity in the cerebellum. *Nature Neuroscience*, 4, 467–475.
- Ito, M. (2001). Cerebellar long-term depression: characterization, signal transduction, and functional roles. *Physiological Reviews*, 81, 1143–1195.
- Ito, M., Jastreboff, P. J., & Miyashita, Y. (1982a). Specific effects of unilateral lesions in the flocculus upon eye movements in albino rabbits. *Experimental Brain Research*, 45, 233–242.
- Ito, M., Sakurai, M., & Tongroach, P. (1982b). Climbing fibre induced depression of both mossy fibre responsiveness and glutamate sensitivity of cerebellar Purkinje cells. *Journal of Physiology*, 324, 113–134.
- Kassardjian, C. D., Yao-Fang, T., Chung, J. Y. J., Heskin, R., Peterson, M. J., & Broussard, D. M. (2005). The site of a motor memory shifts with consolidation. *Journal Neuroscience*, 25, 7979–7985.
- Kleim, J. A., Freeman Jr., J. H., Bruneau, R., Nolan, B. C., Cooper, N. R., Zook, A. et al. (2002). Synapse formation is associated with memory storage in the cerebellum. *Proceedings of the National Academy of Sciences of the United States of America*, 99, 13228–13231.
- Kramer, P. D., Shelhamer, M., & Zee, D. S. (1995). Short-term adaptation of the phase of the vestibulo-ocular reflex (VOR) in normal human subjects. *Experimental Brain Research*, 106, 318–326.
- LeDoux, M. S., Hurst, D. C., & Lorden, J. F. (1998). Single-unit activity of cerebellar nuclear cells in the awake genetically dystonic rat. *Neuroscience*, 86, 533–545.
- Lisberger, S. G. (1988). The neural basis for learning of simple motor skills. *Science*, 242, 728–735.
- Lisberger, S. F., & Sejnowski, T. J. (1992). Motor learning in a recurrent network model based on the vestibulo-ocular reflex. *Nature*, 360, 159–161.
- Luebke, A. E., & Robinson, D. A. (1994). Gain changes of the cat's vestibulo-ocular reflex after flocculus deactivation. *Experimental Brain Research*, 98, 379–390.
- Marr, D. (1969). A theory of cerebellar cortex. *Journal of Physiology*, 202, 437–470.
- Masuda, N., & Amari, S. (2006). Modeling memory transfer and savings in cerebellar motor learning. *Advances in Neural Information Processing Systems*, 18, 859–866 (Y. Weiss, B. Scholkopf, J. Platt Eds.).
- Mauk, M. D. (1997). Roles of cerebellar cortex and nuclei in motor learning: Contradictions or clues? *Neuron*, 18, 343–346.
- Mauk, M. D., & Donegan, N. H. (1997). A model of Pavlovian eyelid conditioning based on the synaptic organization of the cerebellum. *Learning & Memory*, 3, 130–158.
- McClelland, J. L., McNaughton, B. L., & O'Reilly, R. C. (1995). Why there are complementary learning systems in the hippocampus and neocortex: Insights from the successes and failures of connectionist models of learning and memory. *Psychological Review*, 102, 419–457.
- Medina, J. F., Garcia, K. S., & Mauk, M. D. (2001). A mechanism for savings in the cerebellum. *Journal of Neuroscience*, 21, 4081–4089.
- Medina, J. F., Garcia, K. S., Nores, W. L., Taylor, N. M., & Mauk, M. D. (2000). Timing mechanisms in the cerebellum: Testing predictions of a large-scale computer simulation. *Journal of Neuroscience*, 20, 5516–5525.
- Medina, J. F., & Mauk, M. D. (1999). Simulations of cerebellar motor learning: Computational analysis of plasticity at the mossy fiber to deep nucleus synapse. *Journal of Neuroscience*, 19, 7140–7151.
- Medina, J. F., Repa, J. C., Mauk, M. D., & LeDoux, J. E. (2002). Parallels between cerebellum- and amygdala-dependent conditioning. *Nature Review Neuroscience*, 3, 122–131.
- Miles, F. A., & Lisberger, S. G. (1981). Plasticity in the vestibulo-ocular reflex: A new hypothesis. *Annual Review Neuroscience*, 4, 273–299.
- Miyachi, S., Hikosaka, O., & Lu, X. (2002). Differential activation of monkey striatal neurons in the early and late stages of procedural learning. *Experimental Brain Research*, 146, 122–126.
- Miyachi, S., Hikosaka, O., Miyashita, K., Kárádi, Z., & Rand, M. K. (1997). Differential roles of monkey striatum in learning of sequential hand movement. *Experimental Brain Research*, 115, 1–5.
- Morishita, W., & Sastry, B. R. (1996). Postsynaptic mechanisms underlying long-term depression of GABAergic transmission in neurons of the deep cerebellar nuclei. *Journal of Neurophysiology*, 76, 59–68.
- Muellbacher, W., Ziemann, U., Wissel, J., Dang, N., Kofler, M., Facchini, S. et al. (2002). Early consolidation in human primary motor cortex. *Nature*, 415, 640–644.
- Nagao, S., Kitamura, T., Nakamura, N., Hiramatsu, T., & Yamada, J. (1997). Differences of the primate flocculus and ventral paraflocculus in the mossy and climbing fiber input organization. *Journal of Comparative Neurology*, 382, 480–498.
- Nagao, S., & Kitazawa, H. (2003). Effects of reversible shutdown of the monkey flocculus on the retention of adaptation of the horizontal vestibulo-ocular reflex. *Neuroscience*, 118, 563–570.

- Osanai, R., Nagao, S., Kitamura, T., Kawabata, I., & Yamada, J. (1999). Differences in mossy and climbing afferent sources between flocculus and ventral and dorsal paraflocculus in the rat. *Experimental Brain Research*, *124*, 248–264.
- Ouardouz, M., & Sastry, B. R. (2000). Mechanisms underlying LTP of inhibitory synaptic transmission in the deep cerebellar nuclei. *Journal of Neurophysiology*, *84*, 1414–1421.
- Palkovits, M., Mezey, É., Hámori, J., & Szentágothai, J. (1977). Quantitative histological analysis of the cerebellar nuclei in the cat. I. numerical data on cells and on synapses. *Experimental Brain Research*, *28*, 189–209.
- Pasupathy, A., & Miller, E. K. (2005). Different time courses of learning-related activity in the prefrontal cortex and striatum. *Nature*, *433*, 873–876.
- Penhune, V. P., & Doyon, J. (2002). Dynamic cortical and subcortical networks in learning and delayed recall of timed motor sequences. *Journal of Neuroscience*, *22*, 1397–1406.
- Penhune, V. B., & Doyon, J. (2005). Cerebellum and M1 interaction during early learning of timed motor sequences. *Neuroimage*, *26*, 801–812.
- Perrett, S. P., Luis, B. P., & Mauk, M. D. (1993). Cerebellar cortex lesions disrupt learning-dependent timing of conditioned eyelid responses. *Journal of Neuroscience*, *13*, 1708–1718.
- Perrett, S. P., & Mauk, M. D. (1995). Extinction of conditioned eyelid responses requires the anterior lobe of cerebellar cortex. *Journal of Neuroscience*, *15*, 2074–2080.
- Peterson, B. W., Baker, J. F., & Houk, J. C. (1991). A model of adaptive control of vestibuloocular reflex based on properties of cross-axis adaptation. *Annual New York Academy Science*, *627*, 319–337.
- Pugh, J. R., & Raman, I. M. (2006). Potentiation of mossy fiber EPSCs in the cerebellar nuclei by NMDA receptor activation followed by postinhibitory rebound current. *Neuron*, *51*, 113–123.
- Racine, R. J., Wilson, D. A., Gingell, R., & Sunderland, D. (1986). Long-term potentiation in the interpositus and vestibular nuclei in the rat. *Experimental Brain Research*, *63*, 158–162.
- Raymond, J. L., Lisberger, S. G., & Mauk, M. D. (1996). The cerebellum: A neuronal learning machine? *Science*, *272*, 1126–1131.
- Repa, J. C., Muller, J., Apergis, J., Desrochers, T. M., Zhou, Y., & LeDoux, J. E. (2001). Two different lateral amygdala cell populations contribute to the initiation and storage of memory. *Nature Neuroscience*, *4*, 724–731.
- Roland, N. C., & Jaeger, D. (2005). Coding of tactile response properties in the rat deep cerebellar nuclei. *Journal of Neurophysiology*, *94*, 1236–1251.
- Raymond, J. L., & Lisberger, S. G. (1998). Neural learning rules for the vestibulo-ocular reflex. *Journal of Neuroscience*, *18*, 9112–9129.
- Sakurai, M. (1987). Synaptic modification of parallel fibre-Purkinje cell transmission in *in vitro* guinea-pig cerebellar slices. *Journal of Physiology*, *394*, 463–480.
- Shutoh, F., Ohki, M., Kitazawa, H., Itohara, S., & Nagao, S. (2006). Memory trace of motor learning shifts transsynaptically from cerebellar cortex to nuclei for consolidation. *Neuroscience*, *139*, 767–777.
- Stickgold, R., Hobson, J. A., Fosse, R., & Fosse, M. (2001). Sleep, learning, and dreams: Off-line memory reprocessing. *Science*, *294*, 1052–1057.
- Sugihara, I., Ebata, S., & Shinoda, Y. (2004). Functional compartmentalization in the flocculus and the ventral dentate and dorsal group y nuclei: an analysis of single olivocerebellar axonal morphology. *Journal of Compensation and Benefits*, *470*, 113.
- Thach, W. T. (1968). Discharge of Purkinje and cerebellar nuclear neurons during rapidly alternating arm movements in the monkey. *Journal of Neurophysiology*, *31*, 785–797.
- Walker, M. P., Brakefield, T., Hobson, J. A., & Stickgold, R. (2003). Dissociable stages of human memory consolidation and reconsolidation. *Nature*, *425*, 616–620.
- Yamazaki, T., & Tanaka, S. (2005). Neural modeling of an internal clock. *Neural Computation*, *17*, 1032–1058.
- Zhang, W., & Linden, D. J. (2006). Long-term depression at the mossy fiber-deep cerebellar nucleus synapse. *Journal of Neuroscience*, *26*, 6935–6944.