



Real-World-Time Simulation of Memory Consolidation in a Large-Scale Cerebellar Model

Masato Gosui¹ and Tadashi Yamazaki^{1,2,3*}

¹ Department of Communication Engineering and Informatics, Graduate School of Informatics and Engineering, The University of Electro-Communications, Tokyo, Japan, ² Neuroinformatics Japan Center, RIKEN Brain Science Institute, Saitama, Japan, ³ Artificial Intelligence Research Center, National Institute of Advanced Industrial Science and Technology, Ibaraki, Japan

We report development of a large-scale spiking network model of the cerebellum composed of more than 1 million neurons. The model is implemented on graphics processing units (GPUs), which are dedicated hardware for parallel computing. Using 4 GPUs simultaneously, we achieve realtime simulation, in which computer simulation of cerebellar activity for 1 s completes within 1 s in the real-world time, with temporal resolution of 1 ms. This allows us to carry out a very long-term computer simulation of cerebellar activity in a practical time with millisecond temporal resolution. Using the model, we carry out computer simulation of long-term gain adaptation of optokinetic response (OKR) eye movements for 5 days aimed to study the neural mechanisms of posttraining memory consolidation. The simulation results are consistent with animal experiments and our theory of posttraining memory consolidation. These results suggest that realtime computing provides a useful means to study a very slow neural process such as memory consolidation in the brain.

OPEN ACCESS

Edited by:

Arjen Van Ooyen,
VU University Amsterdam,
Netherlands

Reviewed by:

Andreas Knoblauch,
Albstadt-Sigmaringen University,
Germany
Christian Tetzlaff,
Max Planck Institute for Dynamics and
Self-Organization, Germany

*Correspondence:

Tadashi Yamazaki
fina15@neuralgorithm.org

Received: 04 September 2015

Accepted: 18 February 2016

Published: 03 March 2016

Citation:

Gosui M and Yamazaki T (2016)
Real-World-Time Simulation of
Memory Consolidation in a
Large-Scale Cerebellar Model.
Front. Neuroanat. 10:21.
doi: 10.3389/fnana.2016.00021

Keywords: cerebellum, model, memory consolidation, optokinetic response, realtime simulation, graphics processing unit

1. INTRODUCTION

Memory formation has two stages: memory acquisition and memory consolidation (Dudai, 2004). A single session of training forms a type of memory which is fragile and persists only a short period up to minutes to hours. This phase is called memory acquisition. After the training, the learned memory, a short-term memory, decays spontaneously and quickly within a day. Meanwhile, repeated training with a sufficient rest between training sessions gradually form another type of memory, a long-term memory, which is robust and persists for days and weeks. This phase is called memory consolidation. Memory consolidation occurs after training but not during training. That is, when we take a rest after training, the brain still continues working to consolidate the learned memory. This posttraining memory consolidation is thought to be the basis of spacing effect (Ebbinghaus, 1885), in which a massed training is inferior to repeated training to form a robust long-term memory, even if the total training time is equal. Therefore, it is important to study how the brain works after training as well as during training to elucidate the memory mechanisms and our behaviors.

In cerebellar motor learning, both memory formation and consolidation occur within the cerebellum. In gain adaptation of vestibulo-ocular reflex (VOR) and optokinetic response (OKR), parallel fiber-Purkinje cell (PF-PC) synapses in the cerebellar cortex store short-term memory,

whereas mossy fiber-vestibular nuclear cell (MF-VN) synapses in the brain stem store long-term memory (Kassardjian et al., 2005; Shutoh et al., 2006). OKR is an oculomotor reflex in which the eye moves to the same direction of the visual world's movement to reduce the slip of the retinal image. In OKR adaptation, the amplitude of eye movement, called gain, changes by training. By a single 1-h training, the gain increases quickly, which corresponds to memory acquisition. After the training, the gain decreases naturally to the original level within a day. By repeating the 1-h training everyday, the gain increases gradually throughout 1 week (Shutoh et al., 2006), which represents memory consolidation. Moreover, injection of muscimol, a γ -Aminobutyric acid (GABA) receptor agonist, to the cerebellar cortex immediately after the training disrupts memory consolidation (Okamoto et al., 2011), indicating that training alone is not sufficient for memory consolidation. Accumulating evidence suggests that posttraining memory consolidation of OKR gain takes the following steps (Shutoh et al., 2006). By a single 1-h training, PF-PC synapses undergo LTD induced by conjunctive activation of PFs and the CF innervated to the same PCs (Ito, 1989), and thereby the OKR gain increases. After the training, PFs gradually recover from the LTD, which erase the memory of learned OKR gain in the cortex. On the other hand, because inhibition exerted by PCs to VN is weakened due to the LTD, the VN is depolarized tonically. This depolarization, combined with presynaptic MF activation, induces LTP at MF-VN synapses (McElvain et al., 2010; Person and Raman, 2010), and thereby forming the memory of OKR gain in the nucleus. In this way, while the cortical memory is erased gradually after the training, the nuclear memory forms simultaneously as a long-term memory, as if the learned cortical memory is transferred to the nucleus and consolidated there.

We have proposed a theory of the cerebellar posttraining memory consolidation in OKR adaptation (Yamazaki et al., 2015). The theory captures an essence of the macroscopic dynamics of synaptic mechanisms underlying the posttraining memory consolidation. On the other hand, the theory does not provide insights on mesoscopic cellular/synaptic dynamics on the posttraining memory consolidation. For example, the theory does not tell us about spatiotemporal spike patterns of individual neurons. To study the detailed cellular/synaptic dynamics, an elaborated, realistic cerebellar model is necessary. A problem of such elaborated models, however, is that they would spend too much computational time. Typically, computer simulation of large-scale spiking network models is 10–100 times slower than the real-world time (Nageswaran et al., 2009). This means, if we wanted to carry out a computer simulation of memory consolidation for 1 week, and the computer simulation was 100 times slower than real time, the simulation would spend about 2 years in total to complete. This is practically impossible.

In this study, we adopted high-performance computing (HPC) technology to solve these problems. We used graphics processing units (GPUs) to calculate equations of neurons in parallel, which could speed up the numerical simulation drastically. Specifically, we built a very large-scale spiking network model of the cerebellum composed of 1 million neurons, which is a model of 1 mm³ of cats' cerebellum. Moreover,

owing to the parallel computing on GPUs, we were able to conduct the computer simulation fast enough to complete a very long computer simulation in a practical time. Eventually, we achieved realtime simulation, which means that computer simulation of the cerebellar activity for 1 s completes within 1 s in the real-world time (Igarashi et al., 2011; Yamazaki and Igarashi, 2013). This is essential for computer simulation of the cerebellar posttraining memory consolidation, because the memory consolidation takes days or even weeks. Using the present cerebellar model, we performed computer simulation of long-term OKR adaptation of training for 5 days, and obtained qualitatively the same results with experiments (Shutoh et al., 2006) and our previous theoretical model (Yamazaki et al., 2015). We also examined the detailed spike patterns of neurons, which was abstracted and therefore ignored in our theory.

2. MATERIALS AND METHODS

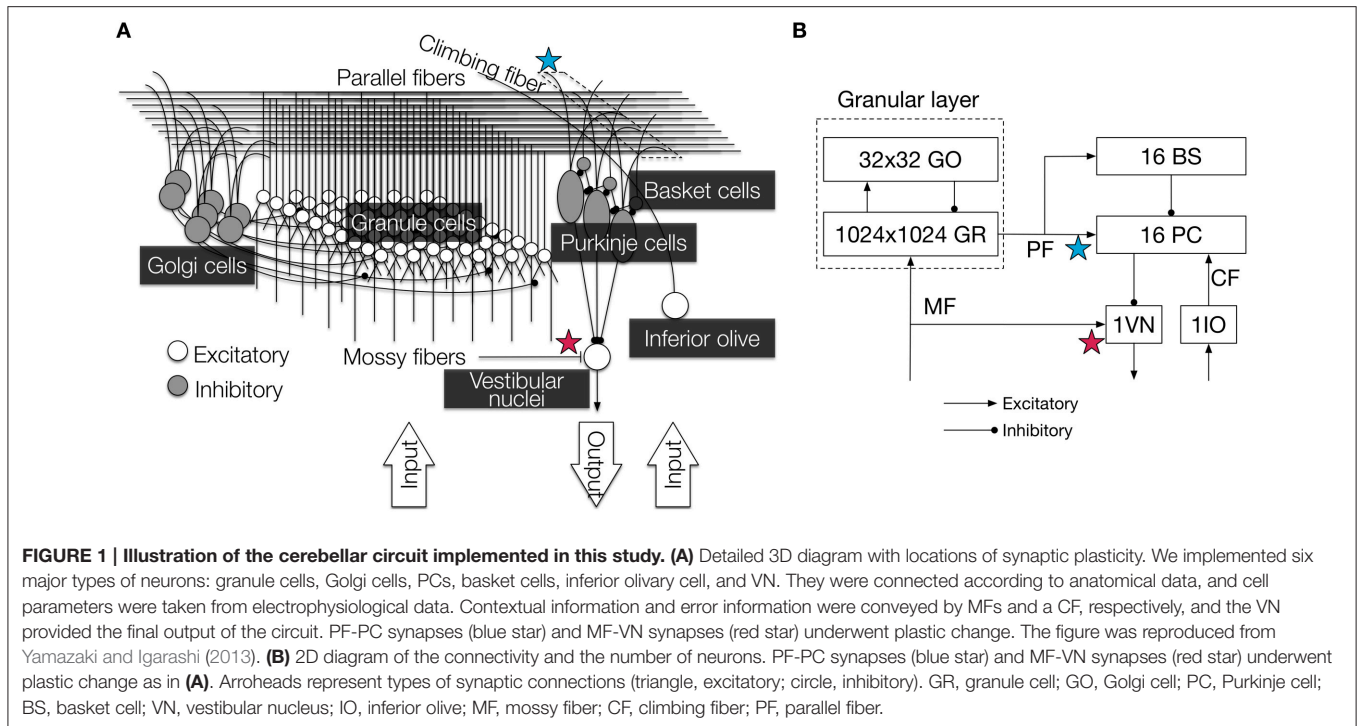
2.1. Model

Our cerebellar model is built based on a 1 mm³ of the cerebellar corticonuclear microcomplex (**Figure 1**) of cats, which is thought to be a functional module of the cerebellum (Ito, 1984, 2012). The original model had 100,000 granule cells, which is 10 times smaller than cats' cerebellum (Ito, 1984), and was already reported elsewhere (Yamazaki and Tanaka, 2007; Yamazaki and Nagao, 2012; Yamazaki and Igarashi, 2013). In this study, we extended the previous model as follows. First, the present model includes 1 million granule cells, thereby the model includes the same number of neurons with 1mm³ of the cats' cerebellum. Second, the present model has synaptic plasticity at mossy fiber-vestibular nuclear cell (MF-VN) synapses, as well as parallel fiber-Purkinje cell (PF-PC) synapses. Except the number of granule cells and MF-VN synaptic plasticity, the previous and present models are the same. Therefore, we summarize the model specification only briefly below. The details are found in our previous papers (Yamazaki and Tanaka, 2007; Yamazaki and Nagao, 2012; Yamazaki and Igarashi, 2013).

The present model is composed of 1,048,576 (= 1024 × 1024) granule cells, 1024 Golgi cells, 16 PCs, 16 basket cells, 1 inferior olivary cell, and 1 VN, connected according to cats anatomical data (Yamazaki and Tanaka, 2007). Neurons are modeled as conductance-based integrate-and-fire units (Gerstner et al., 2014):

$$C \frac{du}{dt} = -g_{\text{leak}}(u(t) - E_{\text{leak}}) - g_{\text{ex:AMPA}}(t)(u(t) - E_{\text{ex}}) - g_{\text{ex:NMDA}}(t)(u(t) - E_{\text{ex}}) - g_{\text{inh:GABA}}(t)(u(t) - E_{\text{inh}}) - g_{\text{ahp}}(t)(u(t) - E_{\text{ahp}}) + I_{\text{ext}}(t), \quad (1)$$

where $u(t)$ is the membrane potential at time t , C is the capacitance, g_{leak} and E_{leak} are the conductance and reversal potential of the leak current, respectively, $g_{\text{ex:AMPA}}(t)$, $g_{\text{ex:NMDA}}(t)$, $g_{\text{inh:GABA}}(t)$ are synaptic conductances of excitatory α -amino-3-hydroxy-5-methyl-4-isoxazolepropionic acid (AMPA) and N -methyl- D -aspartic acid (NMDA), and inhibitory GABA synapses, E_{ex} and E_{inh} are reversal potentials, $g_{\text{ahp}}(t)$ and E_{ahp} are the conductance and reversal potential of after-hyper



polarization, respectively, and $I_{\text{ext}}(t)$ is an external current. When $u(t)$ exceeds a threshold θ_{spike} at time t , the neuron elicits a spike at time t . Cell parameters are taken from turtles and rodents electrophysiological data (Yamazaki and Tanaka, 2007). The values used in this study are summarized in **Table 1**. Synaptic conductance $g_x(t)$ for type x is calculated as a convolution of presynaptic spike events with an exponential kernel as

$$g_x(t) = \sum_j w_j \cdot \bar{g}_x \sum_{f \in S_j} \exp_x \left(- (t - t^{(f)}) \right) \Theta(t - t^{(f)}), \quad (2)$$

where \bar{g}_x is the peak conductance, w_j is the synaptic weight which is constant, S_j is the set of spikes elicited by presynaptic cell j , $t^{(f)}$ is the spike time for the f th spike, $\exp_x(t)$ is the exponential kernel, and $\Theta(t)$ is the Heaviside step function. The exponential kernels used in the present study are summarized in **Table 2**, whereas the synaptic weights are shown in **Table 3**.

The model has two distinct synaptic plasticity sites. One is PF-PC synapses, which undergo long-term depression (LTD) by conjunctive activation of granule-cell axons called parallel fibers (PFs) and a climbing fiber (CF) innervating to the same PC (Ito, 1989), and long-term potentiation (LTP) as well by sole activation of PFs (Lev-Ram et al., 2003). We modeled these bidirectional plasticity as follows:

$$\begin{aligned} \tau_w \frac{dw_{ij}}{dt} &= -w_{ij}(t) + x_{ij}(t) \\ \tau_x \frac{dx_{ij}}{dt} &= -x_{ij}(t) - c_{\text{LTD}} \sum_{s=0}^{50\text{ms}} \text{PF}_{ij}(t-s) \text{CF}(t) + c_{\text{LTP}} \text{PF}_{ij}(t), \end{aligned} \quad (3)$$

TABLE 1 | Summary of cell parameters.

Parameter	Neuron type					
	GR	GO	PC	BS	VN	IO
θ_{spike} (mV)	-35.0	-52.0	-55.0	-55.0	-38.8	-50.0
C (pF)	3.1	28.0	107.0	107.0	122.3	10.0
g_{leak} (nS)	0.43	2.3	2.32	2.32	1.63	0.67
E_{leak} (mV)	-58.0	-55.0	-68.0	-68.0	-56.0	-60.0
$\bar{g}_{\text{ex:AMPA}}$ (nS)	0.18	45.5	0.7	0.7	50.0	1.0
$\bar{g}_{\text{ex:NMDA}}$ (nS)	0.025	30.0	-	-	25.8	-
E_{ex} (mV)	0	0	0	0	0	0
\bar{g}_{inh} (nS)	0.028	-	1.0	-	30.0	0.18
E_{inh} (mV)	-82.0	-	-75.0	-	-88.0	-75.0
\bar{g}_{ahp} (nS)	1.0	20.0	0.1	0.1	50.0	1.0
E_{ahp} (mV)	-82.0	-72.7	-70.0	-70.0	-70.0	-75.0
τ_{ahp} (ms)	5.0	5.0	5.0	2.5	2.5	10.0
I_{ext} (nA)	-	-	0.25	0.1	0.8	-

GR, granule cell; GO, Golgi cell; PC, Purkinje cell; BS, basket cell; VN, vestibular nuclear neuron; IO, inferior olivary neuron; -, nonexistent.

where $w_{ij}(t)$ is the synaptic weight between PC i and PF j , τ_w and τ_x are time constants where $\tau_w \ll \tau_x$, $x_{ij}(t)$ is an internal variable, c_{LTD} and c_{LTP} are constants, $\text{PF}_{ij}(t)$ is 1 if PF j on PC i elicits a spike at time t and 0 otherwise, and $\text{CF}(t)$ is 1 if the climbing fiber elicits a spike at time t and 0 otherwise. The term $\sum_{s=0}^{50\text{ms}} \text{PF}_{ij}(t-s) \text{CF}(t)$ means that PFs that elicit spikes 0–50 ms earlier than the time when the climbing fiber elicits a spike undergo LTD. If n spikes travel along a PF during 50 ms, the weight change becomes n times c_{LTD} . Transmission delay

TABLE 2 | Summary of synaptic functions.

Neuron type	Equation
GR	$\text{exp}_{\text{ex:AMPA}}(t) = e^{-t/1.2}$ $\text{exp}_{\text{ex:NMDA}}(t) = e^{-t/52.0}$ $\text{exp}_{\text{inh}}(t) = 0.43 \times e^{-t/7.0} + 0.57 \times e^{-t/59.0}$
GO	$\text{exp}_{\text{ex:AMPA}}(t) = e^{-t/1.5}$ $\text{exp}_{\text{ex:NMDA}}(t) = 0.33 \times e^{-t/31.0} + 0.67 \times e^{-t/170.0}$
PC	$\text{exp}_{\text{ex:AMPA}}(t) = e^{-t/8.3}$ $\text{exp}_{\text{inh}}(t) = e^{-t/10.0}$
BS	$\text{exp}_{\text{ex:AMPA}}(t) = e^{-t/8.3}$
VN	$\text{exp}_{\text{ex:AMPA}}(t) = e^{-t/9.9}$ $\text{exp}_{\text{ex:NMDA}}(t) = e^{-t/30.6}$ $\text{exp}_{\text{inh}}(t) = e^{-t/42.3}$
IO	$\text{exp}_{\text{ex:AMPA}}(t) = e^{-t/10.0}$

Abbreviations as in Table 1.

TABLE 3 | Summary of synaptic weights.

Presynaptic neuron	Postsynaptic neuron					
	MF	GR	GO	PC	BS	VN
MF	-	4.0	-	-	-	0.2
GR	-	-	0.00004	0.00075	0.00015	-
GO	-	10.0	-	-	-	-
PC	-	-	-	-	-	0.05
BS	-	-	-	5.3	-	-
IO	-	-	-	1.0	-	-

Abbreviations as in Table 1.

of PF spikes might be essential for plasticity (Knoblauch et al., 2014). The conduction velocity of PFs has been experimentally estimated as 0.24 m/s (Vranesic et al., 1994). This results in the transmission delay of 1 mm PF is maximally about 4.2 ms, which could be negligible as long as we assume 50 ms time window for LTD. Therefore, we did not model transmission delays of PF spikes. On the other hand, we do not exactly describe the biological counterpart of x_{ij} . A potential interpretation of x_{ij} would be intracellular concentration of some kinases involving PKC-MAPK positive feedback loop, which plays an essential role in maintenance of induced LTD (Kuroda et al., 2001). The initial values of w and x were set at 1.0 and 0.0, respectively.

The other plasticity is MF-VN synapses, which undergo bidirectional plasticity by a modified Hebbian mechanism. The original equation was proposed by our previous theoretical model (Yamazaki et al., 2015) based on Zhang and Linden (2006); Person and Raman (2010); McElvain et al. (2010) as follows:

$$\tau_v \frac{dv}{dt} = -v(t) \langle \text{MF}(t) \rangle + \langle \text{MF}(t) (\text{VN}(t) - \theta(t)) \rangle, \quad (4)$$

where τ_v is time constant, $v(t)$ is the synaptic weight at MF-VN synapses at time t , $\text{MF}(t)$ is the activity of MFs, $\text{VN}(t)$ is the activity of VN, $\langle \cdot \rangle$ is the temporal average over a certain time window (we assumed 6 s), and $\theta(t)$ is a running average

TABLE 4 | Summary of learning parameters.

Parameter	Value
τ_w (min)	20.0
τ_x (min)	240.0
ρ_{LTD}	0.005
ρ_{LTP}	0.1

of $\text{VN}(t)$, namely $\theta(t) = \langle \text{VN}(t) \rangle$. The left-hand side represents the temporal increment of $v(t)$. The 1st term in the right-hand side represents LTD by sole activation of MFs, and 2nd term represents the Hebbian mechanism, where the weight change correlates with the correlated activity of pre- and postsynaptic neurons. Here, the term $\theta(t)$ acts as a threshold; only when the postsynaptic neuron is activated strongly to exceed $\theta(t)$, the synapses undergo LTP, otherwise LTD or no change. In this way, $\theta(t)$ determines the direction of synaptic change. Moreover, the value of $\theta(t)$ itself changes temporally depending on the temporal history of $\text{VN}(t)$. Higher $\theta(t)$ value makes the synapse harder to undergo LTP. The initial value of v was set at 1.0. The parameters for w and v are summarized in Table 4.

As far as we have tested, the general network dynamics does not change so largely over a wide range of parameter settings. We have found three points that are necessary to achieve robust learning. First, granule-Golgi cell recurrent network should be tuned so as to generate the population code of granule cells robustly. Second, basket cell \rightarrow PC synaptic connections should not be so strong; otherwise, PCs would be silent completely. Third, PC \rightarrow VN synaptic connections should not be so strong; otherwise, VN would be silent completely. If we satisfy these three points, the network, as far as we have tested, works robustly.

2.2. Simulation Paradigm

We conducted computer simulation of long-term OKR gain adaptation as in Shutoh et al. (2006). Specifically, we repeated a 1-h simulated OKR training followed by 23-h rest 5 times corresponding to 5-days training. In each OKR training, simulated optokinetic stimulus is fed to MFs, and retinal slip is fed to a CF. Both optokinetic stimulus and retinal slip are modeled as Poisson spikes with the following firing rate:

$$\begin{aligned} f_{\text{MF}_{\text{train}}}(t) &= \overline{\text{MF}_{\text{train}}} \left(1 + \sin \frac{2\pi t}{T} \right) \quad (\text{for MFs}) \\ f_{\text{CF}_{\text{train}}}(t) &= \overline{\text{CF}_{\text{train}}} \left(1 + \sin \frac{2\pi t}{T} \right) \quad (\text{for a CF}), \end{aligned} \quad (5)$$

where $f_{\text{MF}_{\text{train}}}(t)$ and $f_{\text{CF}_{\text{train}}}(t)$ are the firing rate of MFs and a CF, respectively, $\overline{\text{MF}_{\text{train}}}$ and $\overline{\text{CF}_{\text{train}}}$ are the mean activity of MFs and a CF, which are set at 15 spikes/s and 1.5 spikes/s, respectively. T is a period of a cycle of optokinetic stimulus, which is assumed to be rotated sinusoidally in front of animal subjects. We set $T = 6$ s consistently with the experiments (Shutoh et al., 2006). Because one cycle is 6 s, daily 1-h training consists of 600 cycles of simulated optokinetic stimulus. On the other hand, after training,

we assumed that both MFs and a CF elicited spikes spontaneously with the following firing rate:

$$\begin{aligned} f_{\text{MF}_{\text{rest}}}(t) &= \overline{\text{MF}_{\text{rest}}} \quad (\text{for MFs}) \\ f_{\text{CF}_{\text{rest}}}(t) &= \overline{\text{CF}_{\text{rest}}} \quad (\text{for a CF}), \end{aligned} \quad (6)$$

where $\overline{\text{MF}_{\text{rest}}}$ and $\overline{\text{CF}_{\text{rest}}}$ are set at 5 spikes/s and 1 spikes/s, respectively.

Once we define the firing rate of MF and CF as above, and assume that the activity of a simulated neuron (e.g., firing rate) reflects the strength of input signals to the neuron almost linearly as in the case of integrate-and-fire neurons used in this study (Gerstner et al., 2014), we could estimate the activity of VN as a linear sum of excitatory MF activity and inhibitory PC activity. The PC activity could be estimated as a linear sum of PF activity and basket cell activity, and further by solely MF activity. By substituting the MF and VN activities for $\text{MF}(t)$ and $\text{VN}(t)$ in Equation (4), we could obtain the following simplified equation for v . The detailed derivation is found in our previous paper (Yamazaki et al., 2015).

$$\tau_v \frac{dv}{dt} = -w(t)v + w_c, \quad (7)$$

where $w(t)$ is the average synaptic weight of all PF-PC synapses, and w_c is a constant that defines the initial weight of PF-PC synapses, namely, 1.0. We used Equation (7) rather than Equation (4) for simplicity to update $v(t)$.

2.3. Data Analysis

We conducted computer simulation of the 5-days OKR training, and obtained spike data of all individual neurons and synaptic weight data of PF-PC synapses and MF-VN synapses. The total simulation time was $5 \times 24 \times 60 \times 60 \times 1000 = 4.32 \times 10^8$ ms, with temporal resolution of 1 ms.

We analyzed how the OKR gain changed before and after training for each day. To do so, before training for each day, we fed 10 cycles of simulated optokinetic stimulus to the network and obtained the spike data of VN. We made a spike histogram with bin size of 100 ms, fitted the data with a cosine function with the period of 6 s, and calculated the modulation amplitude. We defined the modulation amplitude as the OKR gain before training. We made the same procedure to obtain the OKR gain after training for each day.

We also examined how the granule cells transmit mossy fiber signals robustly against noise in Poisson spike trains. Granule cells must produce almost identical spike pattern in response to the same optokinetic stimulus with different noise across cycles; otherwise, learning at Purkinje cells would fail. To quantify the reproducibility of the granule cell spike pattern in response to the same simulated optokinetic stimulus, we calculated the reproducibility index at time t defined as the normalized cross correlation as follows:

$$R(t) = \frac{\sum_j z_j^{(i)}(t) z_j^{(i+1)}(t)}{\sqrt{\sum_j z_j^{(i)}(t)} \sqrt{\sum_j z_j^{(i+1)}(t)}}, \quad (8)$$

where $z_j^{(i)}(t)$ is the activity of granule cell j at cycle i of simulated optokinetic stimulus at time t , which was calculated by convolution of the spikes with a causal exponential:

$$z_j^{(i)}(t) = \sum_{f \in S_j^{(i)}} \exp\left(-\frac{t-t^{(f)}}{\tau}\right) \Theta(t-t^{(f)}), \quad (9)$$

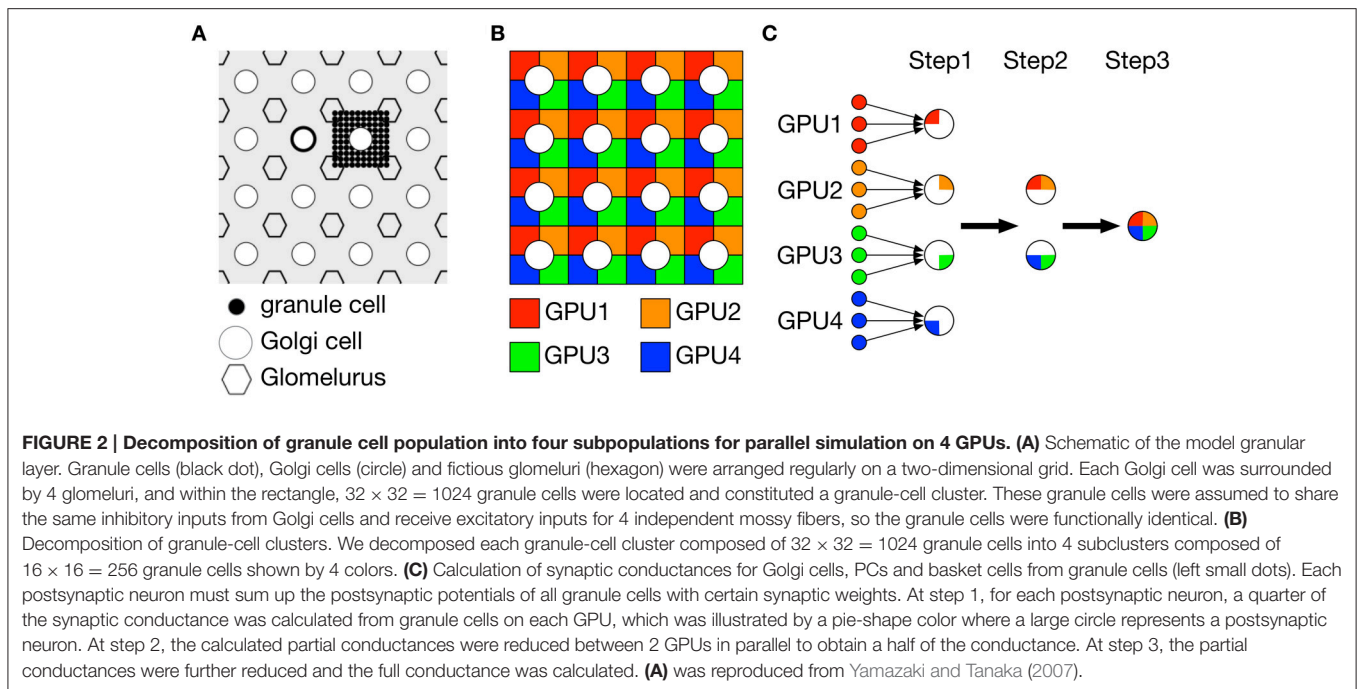
where $S_j^{(i)}$ is the set of spikes elicited by granule cell j at cycle i , $t^{(f)}$ is the spike time for the f th spike, τ is the time constant of 8.3 ms, and $\Theta(t)$ is the Heaviside step function. Intuitively, $z_j^{(i)}(t)$ is a temporal trace of EPSPs of PF j on a PC at cycle i , and $\tau = 8.3$ ms is the time constant of AMPA receptor-mediated PF-EPSPs at a PC (Llano et al., 1991). We calculated the average and standard deviation of the reproducibility index among 10 pairs of two successive cycles.

2.4. Numerical Method

All equations that govern the network dynamics are solved numerically. Specifically, differential equations describing membrane potentials are solved by 2nd-order Runge-Kutta method with temporal resolution (Δt) of 1 ms. The simulation program is written in C with CUDA (Common Unified Device Architecture) (NVIDIA, 2015) and most of the calculation is made on GPUs.

In our previous study (Yamazaki and Igarashi, 2013), we used only 1 GPU (NVIDIA GeForce GTX580) to simulate 100,000 granule cells in realtime. On the other hand, the present model has 10 times more granule cells, which makes computer simulation far slower than realtime. The most time-consuming part is to calculate synaptic conductances of Golgi cells, basket cells and PCs, where these cells receive excitatory inputs from granule cells via PFs. Due to the large number of granule cells, the calculation spends too much time. To address this issue, we decomposed the granular layer network composed of granule cells and Golgi cells into 4 identical subnetworks and calculated the dynamics in parallel on 4 GPUs (2 boards of NVIDIA GeForce GTX TITAN Z, each contains 2 GPUs). In the following, we explain how to decompose the network and calculate the conductance on 4 GPUs.

Figure 2A illustrates a part of the granular layer of our model. The granular layer is composed of 1024×1024 granule cells and 32×32 Golgi cells arranged regularly on a two-dimensional grid. Granule cells are further divided as 32×32 clusters, where each cluster consists of 32×32 granule cells. Due to short dendrites of granule cells, we assumed that the granule cells in the same cluster shared inhibitory inputs from the same Golgi cells. On the other hand, granule cells receive 4 excitatory MF inputs. We assumed that granule cells receive 4 MF inputs independently of the other granule cells. This structure allows us to decompose the granular layer network into 4 identical subnetworks composed of 512×512 granule cells and 32×32 Golgi cells, where granule cells are further divided into 32×32 clusters in which each cluster contained $16 \times 16 = 256$ granule cells as shown in **Figure 2B**. We conducted simulation of each subnetwork on a GPU, thereby we employed 4 GPUs for



simulation of 4 subnetworks. In each subnetwork, we calculated quarter of synaptic conductance for each Golgi cell from granule cells in the same subnetwork. We then exchanged the partial conductances across subnetworks over GPUs and obtained the full conductance by summing up the partial values (Figure 2C). This is made by direct memory exchange between 2 GPUs called peer access, which is much faster than conventional memory exchange via CPUs. Because calculation of synaptic conductance is linear, our split-reduction method over 4 GPUs provides the same result with the conventional method. The same method is used to calculate synaptic conductances of basket cells and PCs as well.

3. RESULTS

3.1. Simulation Time

First, we measured how the simulation time was accelerated by using multi GPUs. Using only 1 GPU, we found that computer simulation of the cerebellar activity for 6 s, corresponding to 1 cycle of simulated optokinetic stimulus, spends 17.7 s. Using 2 GPUs, 9.10 s are spent. Finally, using 4 GPUs, we achieved 5.33 s for 6 s simulation, indicating realtime simulation. Therefore, we used 4 GPUs for further simulation.

3.2. Long-term OKR Gain Change

We conducted computer simulation of long-term OKR adaptation for 5 days. For each day, we performed a simulated 1-h OKR training. During the training, MFs convey simulated optokinetic stimuli, whereas a CF conveys simulated retinal slip error signals. After the training, both MFs and the CF elicit Poisson spikes spontaneously with a constant firing rate, respectively.

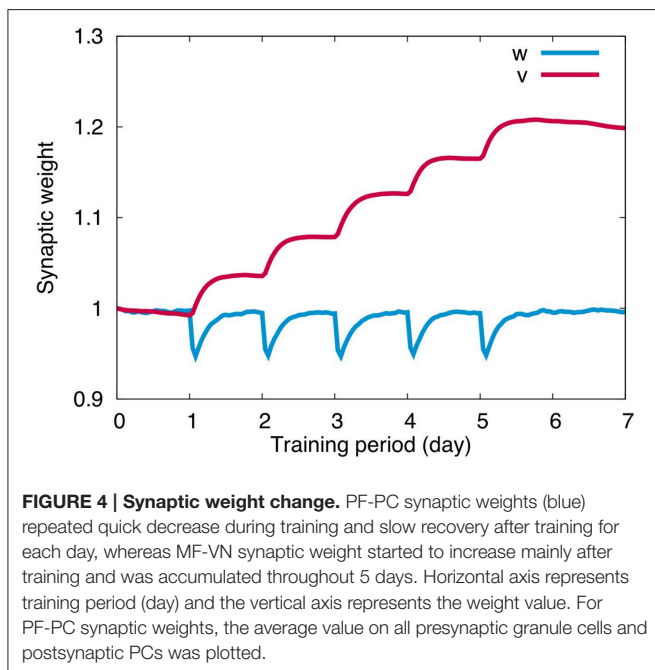
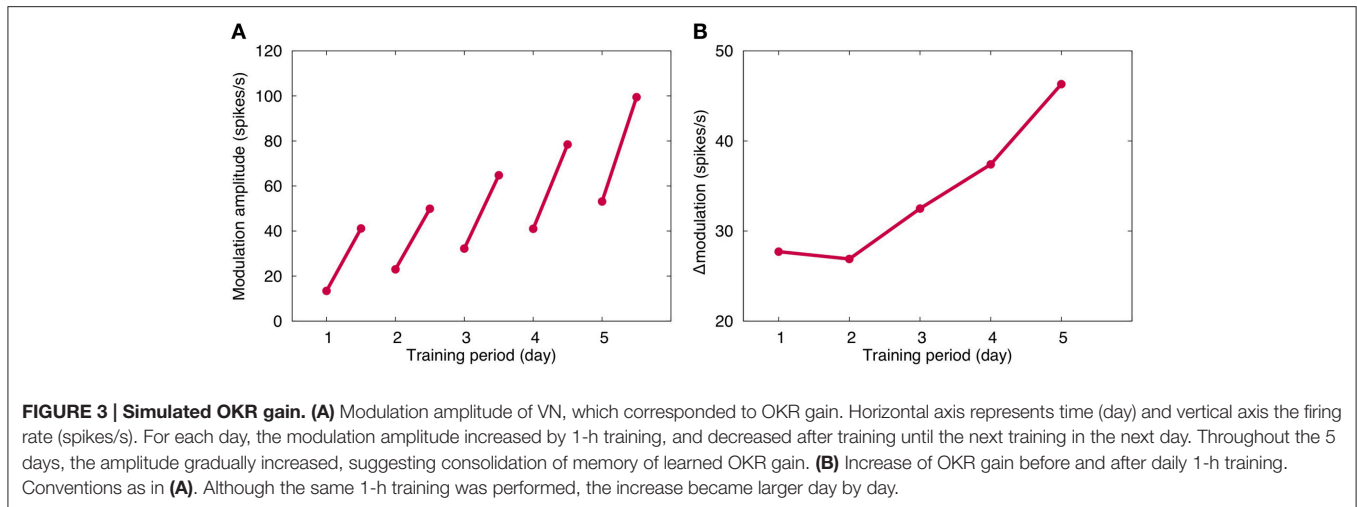
Figure 3A plots the OKR gain obtained in our long-term OKR training simulation for 5 days. By daily 1-h training, the OKR gain increases during training, and after the training, the learned OKR gain almost disappears. This indicates memory acquisition of OKR gain. On the other hand, throughout the 5 days, OKR gain gradually increases, indicating memory consolidation. The present numerical result is qualitatively consistent with previous experimental and theoretical results (Shutoh et al., 2006; Yamazaki et al., 2015).

Figure 3B plots the daily increment of learned OKR gain by 1-h training. The increment becomes larger day by day, suggesting that repeated daily training accelerates the memory acquisition. This result is consistent with previous experiments (Shutoh et al., 2006).

3.3. Change of Synaptic Weights

Figure 4 plots the change of weights at PF-PC synapses (w) and MF-VN synapses (v) throughout the 5 days training. For w , we calculated the average of all PF-PC synaptic weights with respect to PFs and PCs. Similarly for v , we calculated the average of all MF-VN synaptic weights with respect to MFs. PF-PC synapses undergo LTD during training, and slowly return to the original weight value after training spontaneously. PF-PC synapses repeat the same temporal change 5 times for 5 days, suggesting that PF-PC synapses store only short-term memory of OKR gain for hours. On the other hand, MF-VN synapses change little during training, and slowly increase after training. The synaptic weight accumulates every day after training, suggesting that MF-VN synapses store long-term memory of OKR gain.

The overall dynamics is as follows. First, memory of OKR gain is formed in the cerebellar cortex by PF-PC LTD during training. Second, after training, learned cortical memory is



decayed slowly and disappears completely by the next day, and finally, during the slow decay of the cortical memory, memory is formed in the vestibular nucleus by MF-VN LTP, as if the cortical memory is transferred to the nucleus and consolidated. The present numerical result is consistent with the previous theoretical results (Yamazaki et al., 2015).

3.4. Change of Eye Movement Trajectory

So far, both the current numerical and previous theoretical studies show qualitatively the same results. A benefit of our numerical study is that we could obtain detailed data of individual neurons such as membrane potential and spike trains with a fine temporal resolution of 1 ms, which were abstracted in our theoretical model (Yamazaki et al., 2015).

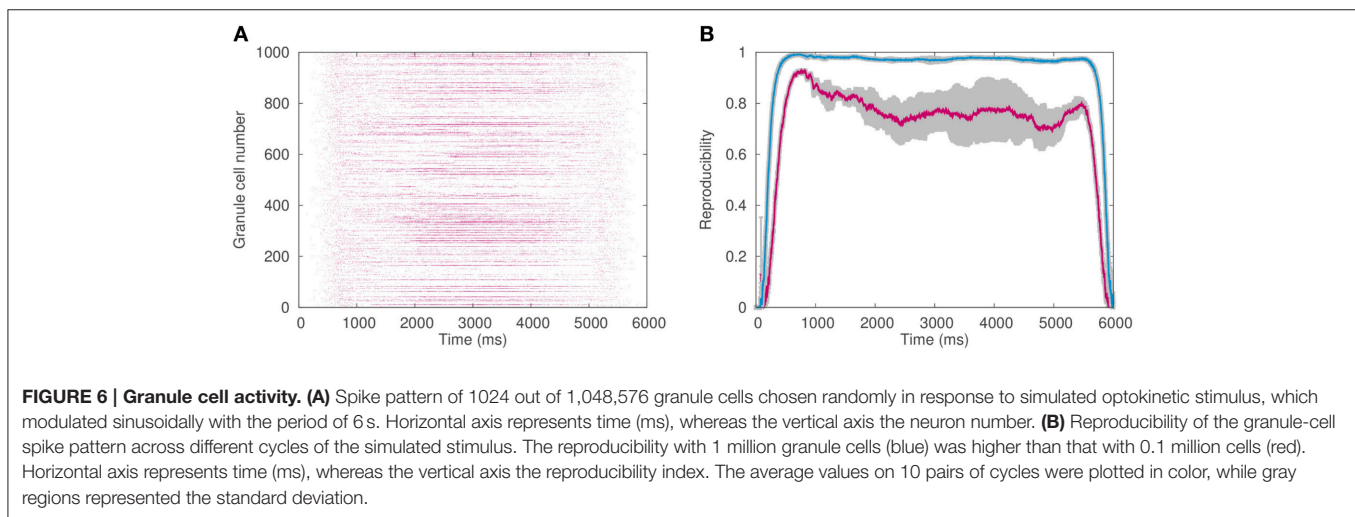
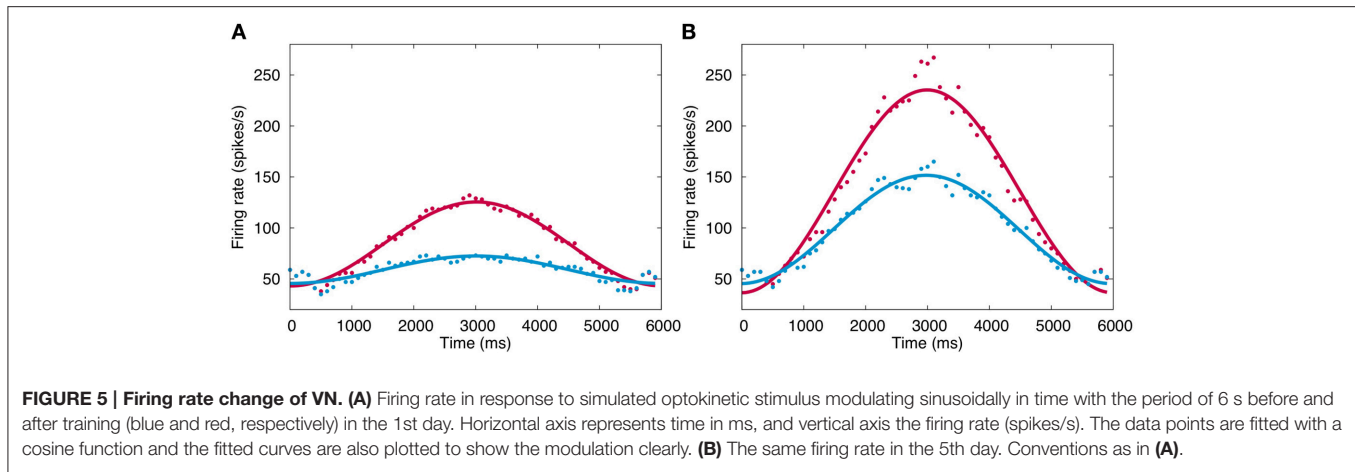
Figure 5 plots the firing rate of VN in response to simulated sinusoidal optokinetic stimulus before and after training at the 1st day (A) and the 5th day (B). The firing rate modulates sinusoidally as the input signals. The modulation amplitude increases by daily 1-h training, and the amplitude also increases gradually throughout 5 days. On the other hand, the baseline firing rate does not change largely from 30–50 spikes/s. Here, the modulation amplitude of VN represents the OKR gain (Shutoh et al., 2006), suggesting that the OKR gain becomes larger by repeated daily training. These results also suggest that realtime simulation allows us to study both macroscopic behaviors of a neural network such as OKR gain, and mesoscopic dynamics of individual neurons in the neural network such as a membrane potential and spike trains.

3.5. Robust Signal Transmission by the Enormous Number of Granule Cells

Granule cells must transmit information conveyed by mossy fibers to Purkinje cells and interneurons faithfully against input noise, otherwise, learning at Purkinje cells would fail. In OKR, mossy fibers convey information on visual world movement, and granule cells produce a spatiotemporal spike pattern that represents the stimulus reliably. For this purpose, the almost identical spike pattern of granule cells must be produced across cycles of the optokinetic stimulus.

Here, we examined how the enormous number of granule cells help them to transmit mossy fiber information faithfully and robustly against input noise. Specifically, we calculated the reproducibility index (Equation 8) that quantifies the reproducibility of the spike pattern of granule cells across cycles of the simulated optokinetic stimulus on different cycles, while changing the number of granule cells in the network.

Figure 6A plots an example of the spike pattern of granule cells, whereas **Figure 6B** plots the reproducibility. As can be seen, the reproducibility is better when 1 million granule cells were employed than 0.1 million granule cells. This result suggests that a functional role of the enormous number of granule cells is



robust transmission of mossy fiber signals to PCs against input noise.

4. DISCUSSION

4.1. Understanding Memory Consolidation Mechanisms

Memory consolidation is a slow process that takes days and weeks. To study the neural mechanisms of memory consolidation, two ways are possible: either conducting experiments or making theoretical models. A theoretical model is a mathematical description of a specific phenomenon. To make such model, we ignore most of experimental details and capture the essence of the phenomenon. For example, in our theoretical model of posttraining memory consolidation in the cerebellum (Yamazaki et al., 2015), we abstracted all detailed physiology of individual neurons, detailed anatomical structure, and detailed input stimuli. This provides a clear view of how the memory consolidates after training, but we still do not know the detailed neuronal process during the memory consolidation. Large-scale, realistic spiking network models are appropriate for

this purpose, but the computational time would be problematic instead.

HPC technology solves this problem. The advantage is two-folds. First, the technology allows us to build a larger-scale model composed of more neurons and synapses with more detailed morphology and biophysical properties than conventional models. Very large-scale functional brain models have been built (Izhikevich and Edelman, 2008; Eliasmith et al., 2012). Notably, The Blue Brain Project and Human Brain Project attempt to build a realistic whole brain model, and they recently published a very detailed cortical microcolumn model (Markram et al., 2015). Second, the technology allows us to carry out computer simulations much faster than that on a single-threaded CPU. The latter makes the above-mentioned long-term computer simulation possible in a reasonable time. For instance, if the computer simulation runs in real time, a simulation of memory consolidation for 1 week completes in 1 week. In our study, we adopted GPUs. Using our large-scale, detailed spiking network model of the cerebellum implemented on multi GPUs, we were able to simulate the detailed temporal dynamics of individual neurons, while observing the slow memory consolidation process

simultaneously. The present study is, as far as the author knows, a first demonstration of a very long time computer simulation of an elaborated spiking network model for days. We were able to replicate our previous theoretical results (Yamazaki et al., 2015), and further examined detailed neuronal and synaptic dynamics during memory consolidation. In cerebellar motor learning, location of motor memory and the role of LTD at PF-PC synapses have been a matter of debate for more than 30 years (Mellvill-Jones, 2000). The present study could provide an answer from the modeling view point.

4.2. Realtime Simulation and the Programming

The present cerebellar model consists of more than 1 million spiking neurons. In general, computer simulation of such large-scale model takes very long time. The simulation could be 10–100 times slower than the real-world time (Nageswaran et al., 2009). However, owing to HPC technology, we were able to conduct computer simulation in realtime, where simulation of cerebellar activity for 1 s completes within 1 s in the real-world time. This allowed us to conduct a complete computer simulation of long-term OKR adaptation training for 5 days in a practical time.

We used 4 GPUs simultaneously to perform realtime simulation of 1 million neurons. To do so, we had to write the simulation program in C with CUDA, a platform for GPU computing, and employed some parallel algorithms to use GPUs efficiently. Specifically, we used some algorithms to compute synaptic conductances of Golgi cells, basket cells and PCs that receive excitatory inputs from many granule cells. This is quite technical and difficult, and so there should be a more simple way to adopt the power of parallel computing in neuroscience. One potential way would be to develop a neural simulator primarily designed for GPUs and some other accelerators. Naveros et al. (2014) has reported development of such a spiking neuron network simulator on a GPU. Some groups have used the software for realtime robot control (Garrido et al., 2013; Casellato et al., 2015).

Realtime simulation is only a milestone, and we expect even faster computer simulation. Scalability, however, would be a problem. Generally speaking, using more GPUs would employ more latency for communications and overhead of communication operations, which could easily be a bottle neck.

4.3. Advantages of Large-Scale Models over Theoretical Models

Although the present study reproduced qualitatively the same results with our previous theoretical model (Yamazaki et al., 2015), some results are slightly different. First, the MF-VN synaptic weight in the present model tends to decay spontaneously, whereas that in the theoretical model did not. This is because in the theoretical model, the decay term was canceled out and removed by a mathematical treatment. In experiments (Shutoh et al., 2006), the learned long-term OKR gain almost vanishes after 2 weeks from the last training, suggesting that it is natural for the synaptic weight to decay spontaneously. Second, the increase of modulation amplitude

of VN before and after the 1-h training gradually becomes larger throughout 5 days in the current study (**Figure 2B**), whereas the change is constant in the theoretical model. The same experiments demonstrate that the increase becomes larger gradually day by day. This result suggests that the present large-scale model captures the detailed dynamics of long-term OKR gain adaptation better than the theoretical model.

Moreover, the present model allows us to study the detailed temporal dynamics of individual neurons with a fine temporal resolution. We were able to obtain detailed spike data of PCs and VN, and analyzed the firing patterns as in **Figure 6**. This is an advantage of an elaborated spiking network model over theoretical models, which abstract detailed temporal dynamics of individual neurons. We will be able to go into the details of molecular mechanisms of memory acquisition and consolidation (Abel and Lattal, 2001; Ito, 2002), if the HPC technology advances further.

We were also able to examine how the number of neurons could affect the stability of the network dynamics. In the present model, we incorporated more than 1 million granule cells, because the cats' cerebellum has 1 million granule cells per 1 mm³ (Ito, 1984). The cerebellar granule cells constitute the largest population in the whole brain (Azevedo et al., 2009). A question arises: why does the cerebellum have such an enormous number of granule cells? A theoretical study has demonstrated that incorporating more granule cells makes the network more reliable for controlling hardware robots (Pinzon-Morales and Hirata, 2015). In the present study, we demonstrated that the enormous number of granule cells makes signal transmission from MFs to PFs more robust as in **Figure 6**.

In summary, combination of large-scale, detailed spiking network models with HPC technology for realtime simulation will provide a strong means to study mesoscopic, detailed neural mechanisms for macroscopic behavioral phenomenon that could take very long time for days and weeks such as memory formation.

4.4. Data Sharing

We will release the source code of the model used in this study under an open-source license upon publication, to facilitate open collaboration and ensure scientific reproducibility, on Cerebellar Platform (<https://cerebellum.neuroinf.jp/>).

AUTHOR CONTRIBUTIONS

TY designed research; MS and TY performed research; MS and TY analyzed data; TY and MS wrote the paper.

FUNDING

JSPS Kakenhi Grant Number (26430009) and UEC Tenure Track Program (6F15).

ACKNOWLEDGMENTS

We would like to thank Soichi Nagao, Jun Igarashi, and Daisuke Miyamoto for helpful discussions.

REFERENCES

- Abel, T., and Lattal, K. (2001). Molecular mechanisms of memory acquisition, consolidation and retrieval. *Curr. Opin. Neurobiol.* 11, 180–187. doi: 10.1016/S0959-4388(00)00194-X
- Azevedo, F. A., Carvalho, L. R., Grinberg, L. T., Farfel, J. M., Ferretti, R. E., Leite, R. E., et al. (2009). Equal numbers of neuronal and nonneuronal cells make the human brain an isometrically scaled-up primate brain. *J. Comp. Neurol.* 513, 532–541. doi: 10.1002/cne.21974
- Casellato, C., Antonietti, A., Garrido, J. A., Ferrigno, G., D'Angelo, E., and Pedrocchi, A. (2015). Distributed cerebellar plasticity implements generalized multiple-scale memory components in real-robot sensorimotor tasks. *Front. Comput. Neurosci.* 9:24. doi: 10.3389/fncom.2015.00024
- Dudai, Y. (2004). The neurobiology of consolidations, or, how stable is the engram? *Ann. Rev. Psychol.* 55, 51–86. doi: 10.1146/annurev.psych.55.090902.142050
- Ebbinghaus, H. (1885). *Memory: A Contribution to Experimental Psychology*. New York, NY: Dover.
- Eliasmith, C., Stewart, T. C., Choo, X., Bekolay, T., DeWolf, T., Tang, Y., et al. (2012). A large-scale model of the functioning brain. *Science* 338, 1202–1205. doi: 10.1126/science.1225266
- Garrido, J. A., Luque, N. R., D'Angelo, E., and Ros, E. (2013). Distributed cerebellar plasticity implements adaptable gain control in a manipulation task: a closed-loop robotic simulation. *Front. Neural Circuits* 7:159. doi: 10.3389/fncir.2013.00159
- Gerstner, W., Kistler, W. M., Naud, R., and Paninski, L. (2014). *Neuronal Dynamics: From Single Neurons to Networks and Models of Cognition*. Cambridge: Cambridge University Press. doi: 10.1017/cbo9781107447615
- Human Brain Project (2013). Available online at: <https://www.humanbrainproject.eu/>
- Igarashi, J., Shouno, O., Fukai, T., and Tsujino, H. (2011). Real-time simulation of a spiking neural network model of the basal ganglia circuitry using general purpose computing on graphics processing units. *Neural Netw.* 24, 950–960. doi: 10.1016/j.neunet.2011.06.008
- Ito, M. (1984). *The Cerebellum and Neural Control*. New York, NY: Raven Press.
- Ito, M. (1989). Long-term depression. *Ann. Rev. Neurosci.* 12, 85–102. doi: 10.1146/annurev.ne.12.030189.000505
- Ito, M. (2002). The molecular organization of cerebellar long-term depression. *Nat. Rev. Neurosci.* 3, 896–902. doi: 10.1038/nrn962
- Ito, M. (2012). *The Cerebellum: Brain for the Implicit Self*. Upper Saddle River, NJ: FT Press.
- Izhikevich, E. M., and Edelman, G. M. (2008). Large-scale model of mammalian thalamocortical systems. *Proc. Natl. Acad. Sci. U.S.A.* 105, 3593–3598. doi: 10.1073/pnas.0712231105
- Kassardjian, C. D., Tan, Y. F., Chung, J. Y., Heskin, R., Peterson, M. J., and Broussard, D. M. (2005). The site of a motor memory shifts with consolidation. *J. Neurosci.* 25, 7979–7985. doi: 10.1523/JNEUROSCI.2215-05.2005
- Knoblauch, A., Hauser, F., Gewaltig, M.-O., Körner, E., and Palm, G. (2014). Does spike-timing-dependent synaptic plasticity couple or decouple neurons firing in synchrony? *Front. Comput. Neurosci.* 6:55. doi: 10.3389/fncom.2012.00055
- Kuroda, S., Schweighofer, N., and Kawato, M. (2001). Exploration of signal transduction pathways in cerebellar long-term depression by kinetic stimulation. *J. Neurosci.* 21, 5693–5702.
- Lev-Ram, V., Mehta, S. B., Kleinfeld, D., and Tsien, R. Y. (2003). Reversing cerebellar long-term depression. *Proc. Natl. Acad. Sci. U.S.A.* 100, 15989–15993. doi: 10.1073/pnas.2636935100
- Llano, I., Marty, A., Armstrong, C. M., and Konnerth, A. (1991). Synaptic- and agonist-induced excitatory currents of purkinje cells in rat cerebellar slices. *J. Physiol. (Lond.)* 434, 183–213. doi: 10.1113/jphysiol.1991.sp018465
- Markram, H., Muller, E., Ramaswamy, S., Reimann, M. W., Abdellah, M., Sanches, C. A., et al. (2015). Reconstruction and simulation of neocortical microcircuitry. *Cell* 163, 456–492. doi: 10.1016/j.cell.2015.09.029
- McElvain, L. E., Bagnall, M. W., Sakatos, A., and du Lac, S. (2010). Bidirectional plasticity gated by hyperpolarization controls the gain of postsynaptic firing responses at central vestibular nerve synapses. *Neuron* 68, 763–775. doi: 10.1016/j.neuron.2010.09.025
- Mellvill-Jones, G. (2000). “Posture,” in *Principles of Neural Science, 4th Edn.*, eds E. Kandel, J. Schwartz, and T. Jessell (New York, NY: McGraw-Hill), 816–831.
- Nageswaran, J. M., Dutt, N., Krichmar, J. L., Nicolau, A., and Veidenbaum, A. (2009). A configurable simulation environment for the efficient simulation of large-scale spiking neural networks on graphics processors. *Neural Netw.* 22, 791–800. doi: 10.1016/j.neunet.2009.06.028
- Naveros, F., Luque, N. R., Garrido, J. A., Carrillo, R. R., Anguita, M., and Ros, E. (2014). A spiking neural simulator integrating event-driven and time-driven computation schemes using parallel cpu-gpu co-processing: a case study. *IEEE Trans. Neural Netw. Learn. Syst.* 26, 1567–1574. doi: 10.1109/TNNLS.2014.2345844
- NVIDIA (2015). *CUDA Programming Guide 7.0*.
- Okamoto, T., Shirao, T., Shutoh, F., Suzuki, T., and Nagao, S. (2011). Post-training cerebellar cortical activity plays an important role for consolidation of memory of cerebellum-dependent motor learning. *Neurosci. Lett.* 504, 53–56. doi: 10.1016/j.neulet.2011.08.056
- Person, A. L., and Raman, I. M. (2010). Deactivation of L-type Ca current by inhibition controls LTP at excitatory synapses in the cerebellar nuclei. *Neuron* 66, 550–559. doi: 10.1016/j.neuron.2010.04.024
- Pinzon-Morales, R.-D., and Hirata, Y. (2015). A realistic bi-hemispheric model of the cerebellum uncovers the purpose of the abundant granule cells during motor control. *Front. Neural Circuits* 9:18. doi: 10.3389/fncir.2015.00018
- Shutoh, F., Ohki, M., Kitazawa, H., Itohara, S., and Nagao, S. (2006). Memory trace of motor learning shifts transsynaptically from cerebellar cortex to nuclei for consolidation. *Neuroscience* 139, 767–777. doi: 10.1016/j.neuroscience.2005.12.035
- The Blue Brain Project (2005). Available online at: <http://bluebrain.epfl.ch/>
- Vranesic, I., Iijima, T., Ichikawa, M., Matsumoto, G., and Knöpfel, T. (1994). Signal transmission in the parallel fiber-purkinje cell system visualized by high-resolution imaging. *Proc. Natl. Acad. Sci. U.S.A.* 91, 13014–13017. doi: 10.1073/pnas.91.26.13014
- Yamazaki, T., and Igarashi, J. (2013). Realtime cerebellum: a large-scale spiking network model of the cerebellum that runs in realtime using a graphics processing unit. *Neural Netw.* 47, 103–111. doi: 10.1016/j.neunet.2013.01.019
- Yamazaki, T., and Nagao, S. (2012). A computational mechanism for unified gain and timing control in the cerebellum. *PLoS ONE* 7:e33319. doi: 10.1371/journal.pone.0033319
- Yamazaki, T., Nagao, S., Lennon, W., and Tanaka, S. (2015). Modeling memory consolidation during posttraining periods in cerebellovestibular learning. *Proc. Natl. Acad. Sci. U.S.A.* 112, 3541–3546. doi: 10.1073/pnas.1413798112
- Yamazaki, T., and Tanaka, S. (2007). A spiking network model for passage-of-time representation in the cerebellum. *Eur. J. Neurosci.* 26, 2279–2292. doi: 10.1111/j.1460-9568.2007.05837.x
- Zhang, W., and Linden, D. (2006). Long-term depression at the mossy fiber-deep cerebellar nucleus synapse. *J. Neurosci.* 26, 6935–6944. doi: 10.1523/JNEUROSCI.0784-06.2006

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2016 Gosui and Yamazaki. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.