View metadata, citation and similar papers at core.ac.uk
brought to you by CORE
provided by UCL Discovery

Redundancy in synaptic connections enables neurons to learn optimally

Naoki Hiratani[1,2]* and Tomoki Fukai[1,3]

[1]Laboratory for Neural Coding and Brain Computation, RIKEN Center for Brain Science [2]Gatsby Computational Neuroscience Unit, University College London
[3]Department of Complexity Science and Engineering, University of Tokyo

Submitted to Proceedings of the National Academy of Sciences of the United States of America

Recent experimental studies suggest that, in cortical microcircuits of the mammalian brain, the majority of neuron-to-neuron connections are realized by multiple synapses. However, it is not known whether such redundant synaptic connections provide any functional benefit. Here, we show that redundant synaptic connections enable near-optimal learning in cooperation with synaptic rewiring. By constructing a simple dendritic neuron model, we demonstrate that with multisynaptic connections, synaptic plasticity approximates a sample-based Bayesian filtering algorithm known as particle filtering, and wiring plasticity implements its resampling process. Extending the proposed framework to a detailed single neuron model of perceptual learning in the primary visual cortex, we show that the model accounts for many experimental observations. In particular, the proposed model reproduces the dendritic position dependence of spike-timing-dependent plasticity, and the functional synaptic organization on the dendritic tree based on the stimulus selectivity of presynaptic neurons. Our study provides a novel conceptual framework for synaptic plasticity and rewiring.

synaptic plasticity | connectomics | synaptogenesis | dendritic computation

Introduction

Synaptic connection between neurons is the fundamental substrate for learning and computation in neural circuits. Previous morphological studies suggest that in cortical microcircuits, often several synaptic connections are found between the presynaptic axons and the postsynaptic dendrites of two connected neurons (1,2,3). Recent connectomics studies confirmed these observations in somatosensory (4), visual (5), and entorhinal (6) cortex, and also in hippocampus (7). In particular, in barrel cortex, the average number of synapses per connection is estimated to be around 10 (8). However, the functional importance of multisynaptic connections remains unknown. Especially, from a computational perspective, such redundancy in connection structure is potentially harmful for learning due to degeneracy (9,10). In this work, we study how neurons perform learning with multisynaptic connections and whether redundancy provides any benefit, from a Bayesian perspective.

Bayesian framework has been established as a candidate principle of information processing in the brain (11,12). Many results further suggest that not only computation, but learning process is also near optimal in terms of Bayesian for given stream of information (13,14,15), yet its underlying plasticity mechanism remains largely elusive. Previous theoretical studies revealed that Hebbian-type plasticity rules eventually enable neural circuits to perform optimal computation under appropriate normalization (16,17). However, these rules are not optimal in terms of learning, so that the learning rates are typically too slow to perform learning from a limited number of observations. Recently, some learning rules are proposed for rapid learning (18,19), yet their biological plausibility are still debatable. Here, we propose a novel framework of non-parametric near-optimal learning using multisynaptic connections. We show that neurons can exploit the variability among synapses in a multisynaptic connection to accurately estimate the causal relationship between pre- and postsynaptic activity. The learning rule is first derived for a simple neuron model, and then implemented in a detailed single neuron model. The derived rule is consistent with many known properties of dendritic plasticity and synaptic organization. In particular, the model explains a potential developmental origin of stimulus-dependent dendritic synaptic organization recently observed in layer 2/3 (L2/3) pyramidal neurons of rodent visual cortex, where presynaptic neurons having a similar receptive field to that of the postsynaptic neuron tend to have synaptic contacts at proximal dendrites (20). Furthermore, the model reveals potential functional roles of anti-Hebbian synaptic plasticity observed in distal dendrites (21,22).

Submission PDF

Results

A conceptual model of learning with multisynaptic connections

Let us first consider a model of two neurons connected with K numbers of synapses (Fig. 1A) to illustrate the concept of the proposed framework. In the model, synaptic connections from the presynaptic neuron are distributed on the dendritic tree of the postsynaptic neuron as observed in experiments (2,3). Although a cortical neuron receives synaptic inputs from several thousands of presynaptic neurons in reality, here we consider the simplified model to illustrate the conceptual novelty of the proposed framework. More realistic models will be studied in following sections.

The synapses generate different amplitudes of excitatory postsynaptic potentials at the soma mainly through two mechanisms. First, the amplitude of dendritic attenuation varies from synapse to synapse, because the distances from the soma are

Significance

Human and animals are capable of rapid learning from a small dataset, which is still difficult for artificial neural networks. Recent studies further suggest that our learning speed is nearly optimal given stream of information, but its underlying mechanism remains elusive. Here, we hypothesized that the elaborate connection structure between presynaptic axons and postsynaptic dendrites is the key element for this near-optimal learning, and derived a data-efficient rule for dendritic synaptic plasticity and rewiring from Bayesian theory. We implemented this rule in a detailed neuron model of visual perceptual learning, and found that the model well reproduces various known properties of dendritic plasticity and synaptic organization in cortical neurons.

Reserved for Publication Footnotes

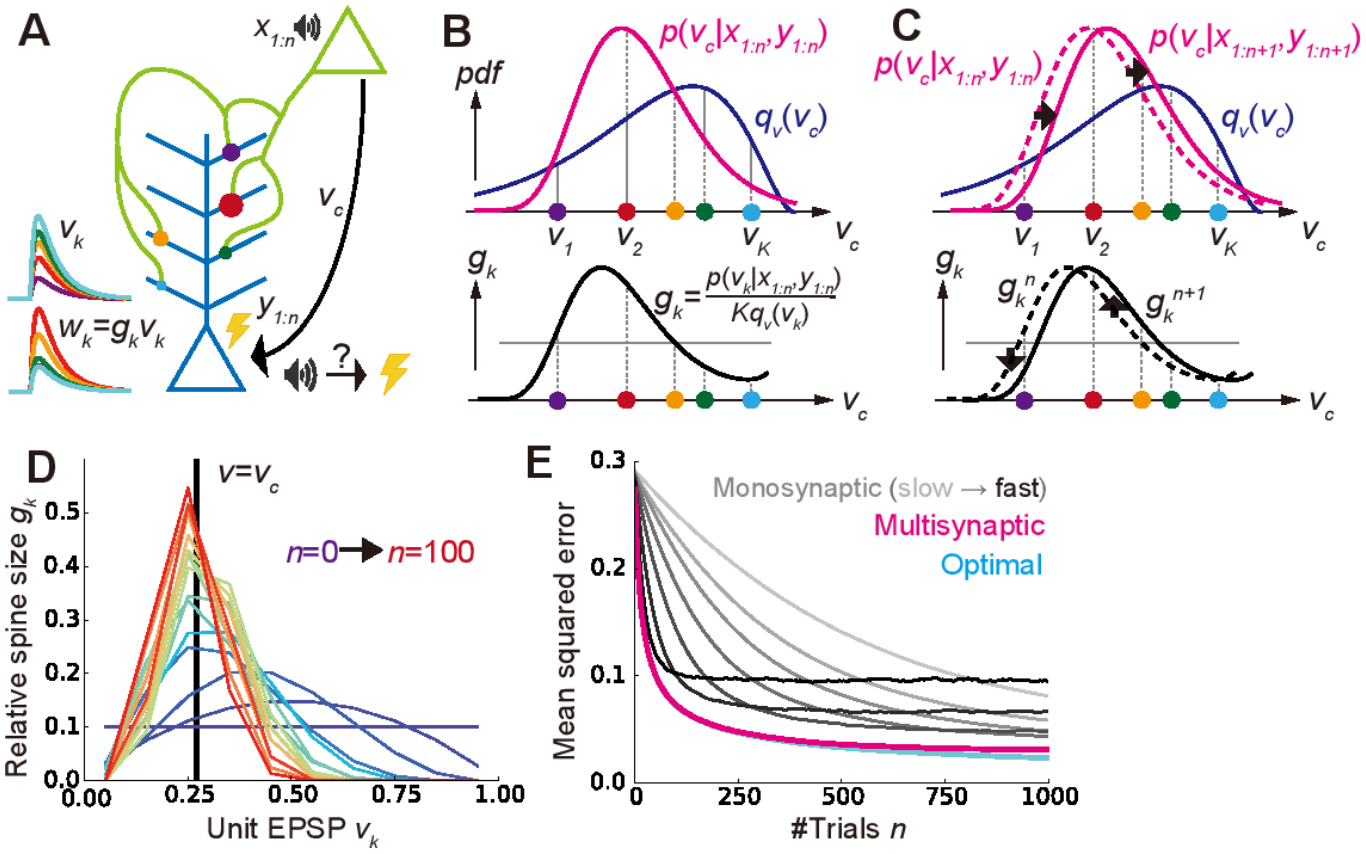

Fig. 1. A conceptual model of multisynaptic learning **A)** Schematic figure of the model consist of two neurons connected with K synapses. Curves on the left represent unit EPSP v_k (top) and the weighted EPSP $w_k=g_k v_k$ (bottom) of each synaptic connection. Note that synapses are consistently colored throughout Figure 1 and 2. **B)** Schematics of non-parametric representation of the probability distribution by multisynaptic connections. In both graphs, x-axes are unit EPSP, and the left (right) side corresponds to distal (proximal) dendrite. The mean over the true distribution $p(v_c|x_{1:n},y_{1:n})$ can be approximately calculated by taking samples (i.e. synapses) from the unit EPSP distribution $q_v(v)$ (top), and then taking a weighted sum over the spine size factor g_k representing the ratio $p(v_k|x_{1:n},y_{1:n})/q_v(v_k)$ (bottom). **C)** Illustration of synaptic weight updating. When the distribution $p(v_c|x_{1:n+1},y_{1:n+1})$ comes to the right side of the original distribution $p(v_c|x_{1:n},y_{1:n})$, a spine size factor g_k^{n+1} become larger (smaller) than g_k^{n} at proximal (distal) synapses. **D)** An example of learning dynamics at K=10 and q_v(v)=const. Each curve represents the distribution of relative spine sizes $\{g_k\}$, and the colors represent the growth of trial number. **E)** Comparison of performance among the proposed method, the monosynaptic rule, and the exact solution (see *A conceptual model of multisynaptic learning* in SI Appendix for details). The monosynaptic learning rule was implemented with learning rate η=0.01, 0.015, 0.02, 0.03, 0.05, 0.1, 0.2 (from gray to black), and the initial value was taken as . Lines were calculated by taking average over 10^4 independent simulations.

different (23,24). Let us denote this dendritic position dependence of synapse k as v_k, and call it as the unit EPSP, because v_k corresponds to the somatic potential caused by a unit conductance change at the synapse (i.e. somatic EPSP per AMPA receptor). As depicted in Figure 1A, unit EPSP v_k takes a small (large) value on a synapse at a distal (proximal) position on the dendrite. The second factor is the amount of AMPA receptors in the corresponding spine, which is approximately proportional to its spine size (25). If we denote this spine size factor as g_k, the somatic EPSP caused by a synaptic input through synapse k is written as $w_k = g_k v_k$. This means that even if the synaptic contact is made at a distal dendrite (i.e. even if v_k is small), if the spine size g_k is large, a synaptic input through synapse k has a strong impact at the soma (e.g. red synapse in Fig. 1A) or vice versa (e.g. cyan synapse in Fig. 1A).

On this model, we consider a simplified classical conditioning task as an example, though the framework is applicable for various inference tasks. Here, the presynaptic neuron activity represents the conditioned stimulus (CS) such as tone, and the postsynaptic neuron activity represents the unconditioned stimulus (US) such as shock. CS and US are represented by binary variables $x_n \in \{0,1\}$ and $y_n \in \{0,1\}$, where $x_n=1$ ($y_n=1$) denotes the presence of the CS (US), and subscript n stands for the trial number (Fig. 1A). Learning behavior of animals and human in such a conditioning can be explained by the Bayesian framework (26). In particular, in order to invoke an appropriate behavioral response, the brain needs to keep track of the likelihood of US given CS $v_c \equiv p(y_n=1|x_n=1)$, presumably by changing the synaptic weight between corresponding neurons. Thus, we consider supervised learning of the conditional probability v_c by multisynaptic connections, from pre- and postsynaptic activities representing CS and US, respectively. From finite trials up to n, this conditional probability is estimated as $\bar{v}_c^n = \int v_c' p(v_c' \mid x_{1:n}, y_{1:n}) dv_c'$, where $x_{1:n}=\{x_1,x_2,\ldots,x_n\}$ and $y_{1:n}=\{y_1,y_2,\ldots,y_n\}$ are the histories of input and output activities, and $p(v_c \mid x_{1:n}, y_{1:n})$ is the probability distribution of the hidden parameter v_c after n trials. Importantly, in general, it is impossible to get the optimal estimation of $\bar{v}_c^{n+1}$ directly from $\bar{v}_c^n$, because in order to calculate $\bar{v}_c^{n+1} = \int v_c' p(v_c' \mid x_{1:n+1}, y_{1:n+1}) dv_c'$, one first needs to calculate the distribution $p(v_c \mid x_{1:n+1}, y_{1:n+1})$ by integrating the previous distribution $p(v_c \mid x_{1:n}, y_{1:n})$ and the new observation at trial $n+1$: $\{x_{n+1}, y_{n+1}\}$. This means that for near-optimal learning, synaptic connections need to learn and represent the distribution $p(v_c \mid x_{1:n}, y_{1:n})$ instead of the point estimation $\bar{v}_c^n$. But, how can synapses achieve that? The key hypothesis of this paper is that redundancy in synaptic connections is the substrate for the non-parametric representation of this probabilistic distribution.

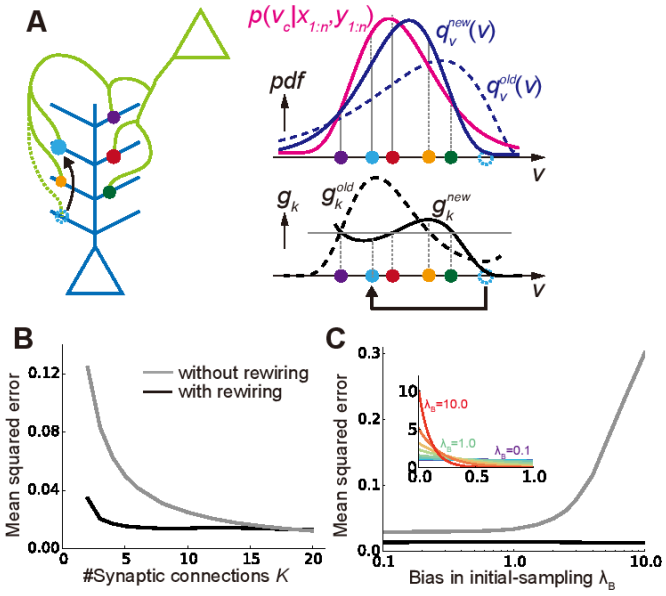

Fig. 2. Synaptic rewiring for efficient learningA) Schematic illustration of resampling. Dotted cyan circles represent an eliminated synapse, and the filled cyan circles represent a newly created synapse. **B, C)** Comparison of performance with/without synaptic rewiring at various synaptic multiplicity K **(B)**, and bias in initial-sampling λ_B **(C)**. For each bias parameter λ_B, the unit EPSP distribution $\{v_k\}$ was set as , as depicted in the inset. Lines are the means over 10^4 simulations.

Below, we show that dendritic summation over multisynaptic connections yields the optimal estimation from the given distribution $p(v_c \mid x_{1:n},y_{1:n})$, and dendritic-position-dependent Hebbian synaptic plasticity updates this distribution.

Dendritic summation as importance sampling

We first consider how dendritic summation achieves the calculation of the mean conditional probability $\bar{v}_c^n = \int v_c' p(v_c' \mid x_{1:n},y_{1:n}) dv_c'$. It is generally difficult to evaluate this integral by directly taking samples from the distribution $p(v_c \mid x_{1:n},y_{1:n})$ in a biologically plausible way, because the cumulative distribution changes its shape at every trial. Nevertheless, we can still estimate the mean value by using an alternative distribution as the proposal distribution, and taking weighted samples from it. This method is called importance sampling (27). In particular, here we can use the unit EPSP distribution $q_v(v)$ as the proposal distribution, because unit EPSPs $\{v_k\}$ of synaptic connections can be interpreted as samples depicted from the unit EPSP distribution q_v (Fig. 1B top). Thus, the mean $\bar{v}_c^n$ is approximately calculated as

$$\bar{v}_c^n = \int v_c' p(v_c' \mid x_{1:n},y_{1:n}) dv_c' \approx \frac{1}{K}\sum_{k=1}^{K} \frac{p(v_c=v_k \mid x_{1:n},y_{1:n})}{q_v(v_k)} v_k = \sum_k g_k^n v_k = \sum_k w_k^n, \quad [1]$$

where $g_k^n = \frac{p(v_c=v_k \mid x_{1:n},y_{1:n})}{Kq_v(v_k)}$. Therefore, if spine size g_k^n represents the relative weight of sample v_k, then dendritic summation over postsynaptic potentials $w_k^n \equiv g_k^n v_k$ naturally represents the desired value ($\bar{v}_c^n \approx \sum_k w_k^n$). For instance, if the distribution of synapses is biased toward proximal side (i.e. if the mean $\bar{v}_c^n$ is overestimated by the distribution of unit EPSPs as in Fig. 1B top), then synapses at distal dendrites should possess large spine sizes, while the spine sizes of proximal synapses should be smaller (Fig. 1B bottom).

Synaptic plasticity as particle filtering

In the previous section, we showed that redundant synaptic connections can represent probabilistic distribution $p(v_c=v_k \mid x_{1:n},y_{1:n})$ if spine sizes $\{g_k\}$ coincide with their importance $g_k^n = \frac{p(v_c=v_k \mid x_{1:n},y_{1:n})}{Kq_v(v_k)}$. But, how can synapses update their representation of the probabilistic distribution $p(v_c=v_k \mid x_{1:n},y_{1:n})$ based on a new observation $\{x_{n+1}, y_{n+1}\}$? Because $p(v_c=v_k \mid x_{1:n},y_{1:n})$ is mapped onto a set of spine sizes $\{g_k^n\}$ as in Equation 1, the update of the estimated distribution $p(v_k \mid x_{1:n},y_{1:n}) \to p(v_k \mid x_{1:n+1},y_{1:n+1})$ can be performed by the update of spine sizes $\{g_k^n\} \to \{g_k^{n+1}\}$. By considering particle filtering (28) on the parameter space (see *The learning rule for multisynaptic connections* in SI Appendix for details), we can derive the learning rule for spine size as

$$g_k^{n+1} = \frac{1+f(x_{n+1},y_{n+1};v_k)}{1+f(x_{n+1},y_{n+1};w^n)} g_k^n$$

$$f(x,y;v) \equiv (2v-1)x(2y-1)$$

This rule is primary Hebbian, because the weight change depends on the product of pre- and postsynaptic activity x_{n+1} and y_{n+1}. In addition to that, the change also depends on unit EPSP v_k. This dependence on unit EPSP reflects the dendritic position dependence of synaptic plasticity. In particular, for a distal synapse (i.e. for small v_k), the position-dependent term $(2v_k\text{-}1)$ takes a negative value (note that $0 \le v_k < 1$), thus yielding an anti-Hebbian rule as observed in neocortical synapses (21,22).

For instance, if the new data $\{x_{n+1}, y_{n+1}\}$ indicates that the value of v_c is in fact larger then previously estimated, then the distribution $p(v_c \mid x_{1:n+1},y_{1:n+1})$ shifts to the right side (upper panel of Fig. 1C). This means that the spine size g_k^{n+1} becomes larger then g_k^n at synapses on the right side (i.e. proximal side), whereas synapses get smaller on the left side (i.e. distal side; bottom panel of Fig. 1C). Therefore, pre- and postsynaptic activity causes LTP at proximal synapses induces LTD at distal synapses as observed in experiments (21,22). The derived learning rule (Eq. 2) also depends on the total EPSP amplitude $w^n \equiv \sum_k w_k^n \equiv \sum_k g_k^n v_k$. This term reflects a normalization factor possibly modulated through redistribution of synaptic vesicles over the presynaptic axon (29). A surrogate learning rule without this normalization factor will be studied in a later section.

We performed simulations by assuming that the two neurons are connected with ten synapses with the uniform unit-EPSP distribution (i.e. $q_v(v)$ = const.). At an initial phase of learning, the distribution of spine size $\{g_k^n\}$ has a broad shape (purple lines in Fig. 1D), and the mean of distribution is far away from the true value ($v=v_c$). However, the distribution is skewed around the true value as evidence is accumulated through stochastic pre- and postsynaptic activities (red lines in Fig. 1D). Indeed, the estimation performance of the proposed method is nearly the same as that of the exact optimal estimation, and much better than the standard monosynaptic learning rules (Fig. 1E; see *Monosynaptic learning rule* in SI Appendix for details).

Synaptogenesis as resampling

As shown above, weight modification in multisynaptic connections enables a near optimal learning. However, to represent the distribution accurately, many synaptic connections are required (gray line in Fig. 2B), while the number of synapses between an excitatory neuron pair is typically around five in the cortical microcircuits. Moreover, even if many synapses are allocated between presynaptic and postsynaptic neurons, if the unit EPSP distribution is highly biased, the estimation is poorly performed (gray line in Fig. 2C). We next show that this problem can be avoided by introducing synaptogenesis (30) into the learning rule.

In the proposed framework, when synaptic connections are fixed (i.e. when $\{v_k\}$ are fixed), some synapses quickly become useless for representing the distribution. For instance, in Figure 2A, (dotted) cyan synapse is too proximal to contribute for the representation of $p(v_c \mid x,y)$. Therefore, by removing the

Submission PDF

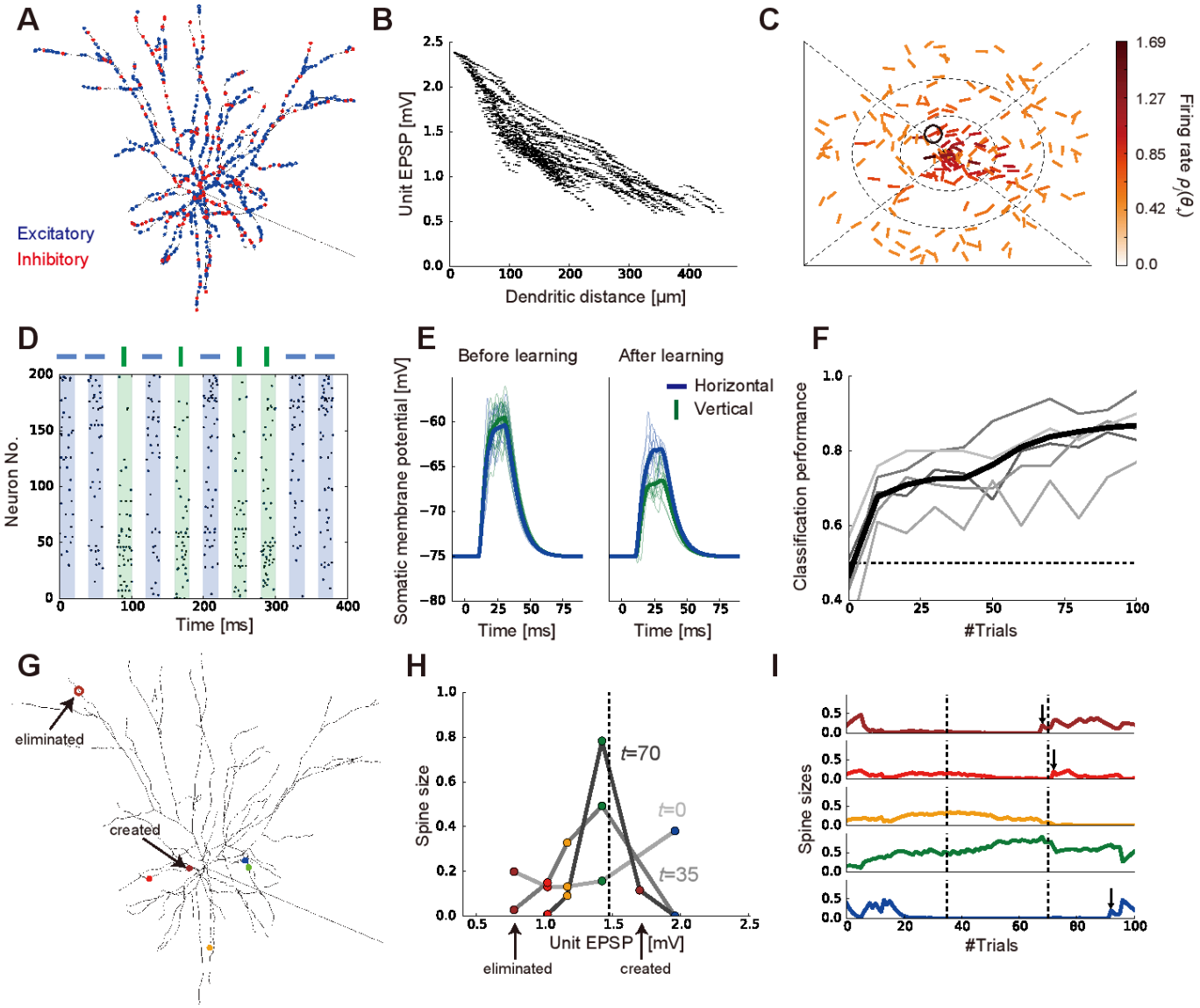

Fig. 3. A detailed model of multisynaptic learning with multiple presynaptic neuronsA) Morphology of the detailed neuron model. Blue and red points on the dendritic trees represent excitatory and inhibitory synaptic inputs, respectively. **B)** Dendritic position dependence of unit EPSP. Each dot represents a synaptic contact on the dendritic tree. **C)** An example of the visual selectivity patterns of presynaptic neurons. Position and angle of each bar represent the receptive field (RF) and the orientation selectivity of each presynaptic neuron, where the RF was defined relative to the RF of the postsynaptic neuron (the central position). Colors represent the firing rates of presynaptic neurons when a horizontal bar stimulus is presented at the RF of the postsynaptic neuron. Here, the firing rates were evaluated as the expected number of spikes within 20ms stimulus duration (see *Stimulus selectivity* in SI Appendix for details). The black circle shows the selectivity of the representative neuron depicted in **G-I**. **D)** Examples of input spike trains generated from the horizontal (target) and vertical (non-target) stimuli. Presynaptic neurons were sorted by their stimulus preference. Note that in the actual simulations, variables were initialized after each stimulation trial. See *Task configuration* in SI Appendix for details of the task. **E)** Somatic responses before and after learning. Thick lines represent the average response curves over 100 trials and thin lines are trial-by-trial responses. **F)** The average learning curves over 50 simulations (black line) and examples of learning curves (gray lines). **G-I)** An example of learning dynamics under the multisynaptic rule (see Results for details).

cyan synapse and creating a new synapse at a random site, on average, the representation becomes more effective (Fig. 2A). Importantly, in our framework, spine size factor g_k is proportional to the informatic importance of the synapse by definition, thus optimal rewiring is achievable simply by removing the synapse with the smallest spine size. Ideally, the new synapse should be sampled from the underlying distribution of $\{g_k\}$ for an efficient rewiring (31), yet it is not clear if such a sampling is biologically plausible; hence below we consider a uniform sampling from the parameter space. Although here we assumed simultaneous elimination and creation of synaptic contacts for simplicity, the strict balance between elimination and creation is not necessary, as will be shown later in the detailed neuron model.

By introducing this resampling process, the model is able to achieve high performance robustly. With rewiring, a small error is achieved even when the total number of synaptic connections is just around three (black line in Fig. 2B). In contrast, more than 10 synapses are required for achieving the same performance without rewiring (gray line in Fig. 2B). Similarly, even if the initial distribution of $\{v_k\}$ is poorly taken, with rewiring the neuron can achieve a robust learning (black line in Fig. 2C), whereas the performance highly depends on the initial distribution of the synapses in the absence of rewiring (gray line in Fig. 2C).

Recent experimental results suggest that the creation of new synapses is clustered at active dendritic branches (32). Correspondingly, by sampling new synapses near large synapses, per-

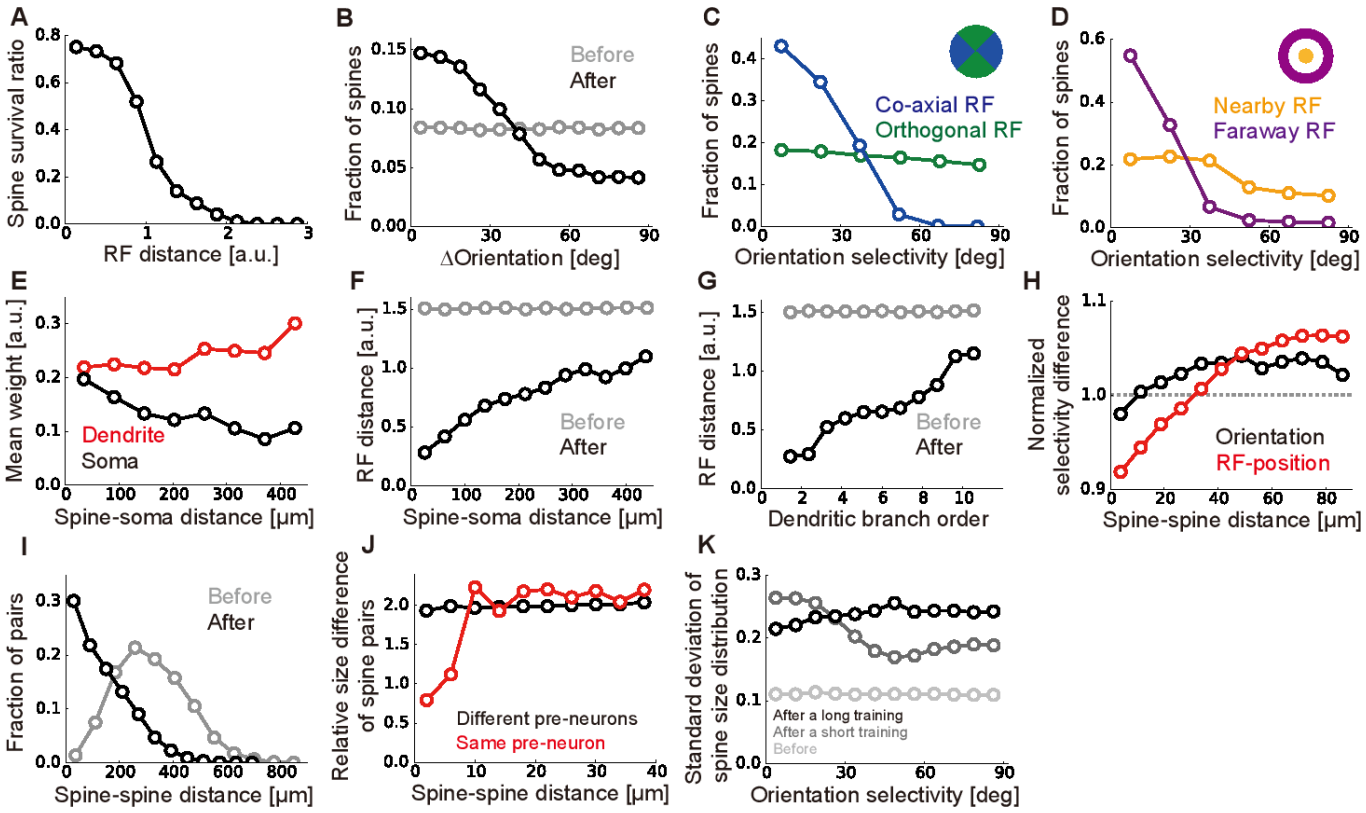

Fig. 4. Synaptic organization on the dendrite by the multisynaptic learning ruleA) Survival ratio of spines with different receptive field (RF) distances from the postsynaptic neuron. **B)** Fraction of spines having various orientation selectivity before and after learning. **C, D)** Fraction of spines survived after learning, calculated for different orientation selectivity at co-axial/orthogonal RFs **(C)**, and at nearby/faraway RFs **(D)**. We defined the RF of presynaptic neuron j being orthogonal if , and co-axial otherwise. The RF of neuron j was defined as nearby if r_j<0.5, but faraway if r_j>1.0 (see *Stimulus selectivity* in SI Appendix). **E)** Relationship between the dendritic distance and the relative weight at the dendrite g_k and the soma $g_k v_k/v_{max}$. **F)** Relationship between the dendritic distance of a spine from the soma and its RF distance in the visual space. **G)** The same as **F**, but calculated for the dendritic branch order, not the dendritic distance. **H)** Dependence of normalized RF difference (red), and normalized orientation difference (black) on the between-spine distance were calculated for two synapses projected from different neurons. We used the Euclidean distance in the visual field for RF distance between presynaptic neurons i and j, and the normalization was taken over all synapse pairs. **I)** Distributions of dendritic distance between synapses projected from the same presynaptic neuron before and after learning. **J)** Relative spine size difference between spines projected from the same presynaptic neuron or different neurons calculated for pairs with different spine distance. The relative size difference between spine i and j was defined as $|\log(g_i/g_j)|$. **K)** Standard deviation (SD) of spine size distribution at various orientation selectivity for synapses from presynaptic neurons with nearby RFs (r_j<0.5). The distributions for short and long training were taken after learning from 10 and 1000 samples, respectively. All panels were calculated by taking averages over 500 independently simulated neurons, and the learning was performed from 1000 training samples.

formance becomes better given a large number of samples (SI Appendix, Fig. S1A; see *Uniform and multinomial sampling* in SI Appendix), though this difference almost disappears under an explicit normalization (SI Appendix, Fig. S1B).

Detailed single neuron model of learning from many presynaptic neurons

In the previous sections, we found that synaptic plasticity in multisynaptic connections can achieve non-parametric near-optimal learning in a simple model with one presynaptic neuron. To investigate its biological plausibility, we next extend the proposed framework to a detailed single neuron model receiving inputs from many presynaptic neurons. To this end, we constructed an active dendritic model using NEURON simulator (33) based on a previous model of L2/3 pyramidal neurons of the primary visual cortex (34). We randomly distributed 1000 excitatory synaptic inputs from 200 presynaptic neurons on the dendritic tree of the postsynaptic neuron, while fixing synaptic connections per presynaptic neuron at K=5 (Fig. 3A; see *Morphology* in SI Appendix for the details of the model). We assumed that all excitatory inputs are made on spines, and each spine is projected from only one bouton for simplicity. In addition, 200 inhibitory synaptic inputs were added on the dendrite to keep the excitatory/inhibitory (E/I) balance (35). We first assigned a small constant conductance for each synapse, and then measured the somatic potential change, which corresponds to the unit EPSP in the model. As observed in cortical neurons (23), input at a more distal dendrite showed larger attenuation at the soma, though variability was quite high across branches (Fig. 3B).

Next, we consider a perceptual learning task in this neuron model. Each excitatory presynaptic neuron was assumed to be a local pyramidal neuron, modeled as a simple cell having a small receptive field (RF) and a preferred orientation in the visual space (Fig. 3C). Axonal projections from each presynaptic neuron were made onto five randomly selected dendritic branches of the postsynaptic neuron regardless of the stimulus selectivity, because visual cortex of mice has a rather diverse retinotopic structure (36). In this setting, the post-neuron should be able to infer the orientation of the stimulus presented at its RF from the presynaptic inputs, because cells having similar RFs or orientation selectivity are often co-activated (37,38). Thus, we consider a supervised learning task in which the postsynaptic neuron has to learn to detect a horizontal grading, not a vertical grading, from stochastic presynaptic spikes depicted in Figure 3D. In reality, the modulation of lateral connections in L2/3 is arguably guided by the feedforward inputs from layer 4 (39,40). However, for simplicity, we instead introduced an explicit supervised signal to

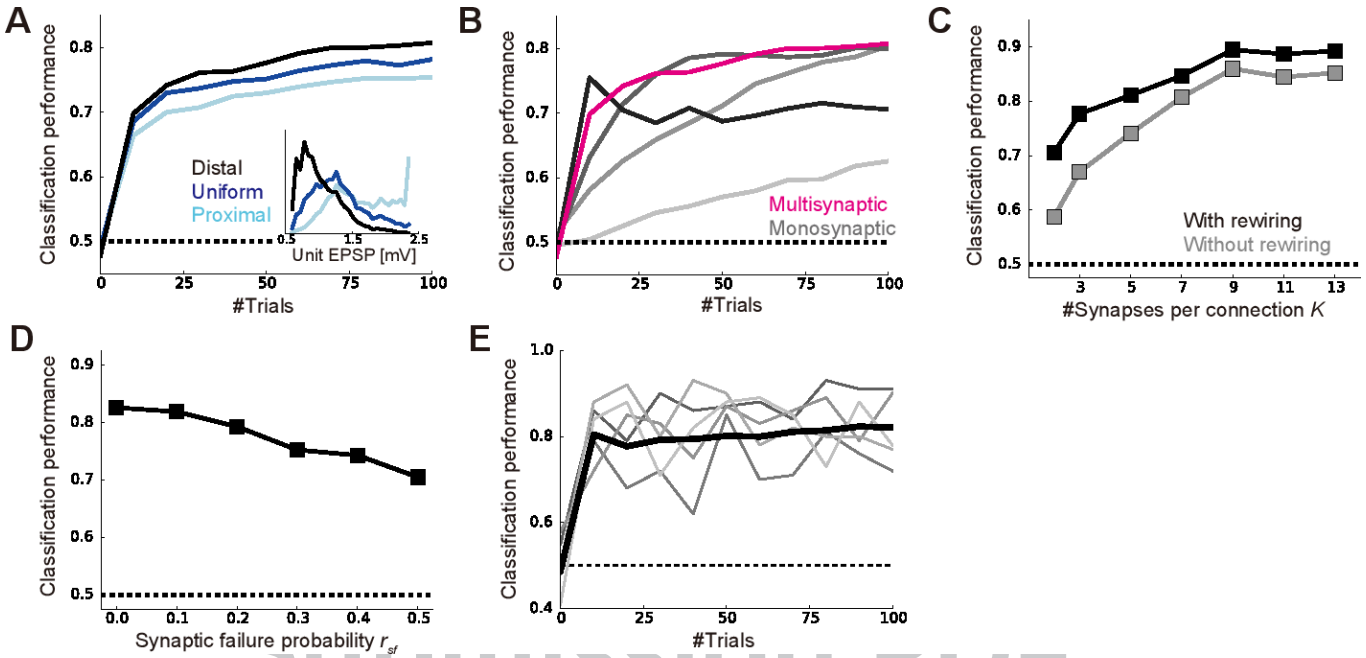

Fig. 5. Dynamics of the multisynaptic learning rule under various conditions A) Learning dynamics under various initial synaptic distributions. The inset represents the unit EPSP distributions when synaptic connections are biased toward the distal dendrite (black), unbiased (blue), and biased toward the proximal (light blue). **B)** Comparison with the monosynaptic learning. We set the learning rate as η_w=0.03, 0.1, 0.3, 1.0, from light gray to black lines. To keep the E/I balance, the inhibitory weight was set to γ_I=2.0 for η_w=1.0, and γ_I=1.25 for the rest. The magenta line is the same as the black line in **A**. **C)** Classification performance after learning with different numbers of synapses per connection with or without rewiring. For the E/I balance, the inhibitory weights were chosen as γ_I=2.0, 1.2, 0.75, 0.6, 0.5, 0.4, 0.3, when the number of synapses per connections were K=2, 3, 5, 7, 9, 11, 13, respectively. **D)** The performance after learning with various synaptic failure probabilities. Both in panel **C** and **D**, the performance was calculated after 1000 trials. **E)** Learning dynamics under the surrogate rule. Thin gray lines represent examples. All panels were calculated by taking the means over 50 simulations.

the postsynaptic neuron. In this formulation, we can directly apply the rule for synaptic plasticity and rewiring introduced in the previous section (see *The learning rule for the detailed model* in SI Appendix). In the rewiring process, a new synaptic contact was made on one of the branches on which the presynaptic neuron initially had at least one synaptic contact, to mimic the axonal spatial constraint. Here, in addition to the rewiring by the proposed multisynaptic rule, we implemented elimination of synapses from uncorrelated presynaptic neurons, to better replicate developmental synaptic dynamics.

Initially, the postsynaptic somatic membrane potential responded similarly to both horizontal and vertical stimuli, but the neuron gradually learned to show a selective response to the horizontal stimulus (Fig. 3E). After 100 trials, the two stimuli became easily distinguishable by the somatic membrane dynamics (Fig. 3E and F; see *Performance evaluation* in SI Appendix for details). Next, we examined how the proposed mechanism works in detail. To this end, we focused on a presynaptic neuron circled in Figure 3C, and tracked the changes in its synaptic projections and spine sizes (Fig. 3G-I). Because the neuron has a RF near the postsynaptic RF, and its orientation selectivity is nearly horizontal, the total synaptic weight from this neuron should be moderately large after learning. Indeed, the Bayesian optimal weight was estimated to be around 1.5 mV in the model (vertical dotted line in Fig. 3H), under the assumption of linear dendritic integration. Overall, the unit EPSPs of the majority of synapses were initially around 1.0-1.5 mV, while smaller or larger unit EPSPs were rare due to dendritic morphology (Fig. 3B). To counterbalance this bias toward the center, we initialized the spine sizes in a U-shape (light gray line in Fig. 3H). In this way, the prior distribution of the total synaptic weight becomes roughly uniform (see also Fig. 1B). After a short training, the most proximal spine (the blue one) was depotentiated, whereas spines with moderate unit EPSP sizes were potentiated (yellow and green ones on dark gray line in Fig. 3H). This is because, the expected distribution of the weight from this presynaptic neuron shifted to the left side (i.e. to a smaller EPSP) after the training, and this shift was implemented by reducing the spine size of the proximal synapse, while increasing the sizes of others (as in Fig. 1C, but here the change is to the opposite direction). Note that, the most distal spine (the brown one) was also depressed here, as the expected distribution got squeezed toward the center. Finally, after a longer training, the expected distribution became more squeezed, hence all but the green spine were depotentiated (black line in Fig. 3H). Moreover, the most distal synapse was eliminated because its spine size became too small to make any meaningful contribution to the representation, and a new synapse was created at a proximal site (open and closed brown circles in Fig. 3G, respectively) as explained in Figure 2A. This rewiring achieves a more efficient representation of the weight distribution on average. Indeed, the new brown synapse was potentiated subsequently (top panel in Fig. 3I). Note that, in this example, red and blue synapses were also rewired shortly after this moment (vertical arrows above red and blue traces in Fig. 3I).

The model reproduces various properties of synaptic organization on the dendrite

While we confirmed that the proposed learning paradigm works well in a realistic model setting, we further investigated its consistency with experimental results. We first calculated spine survival ratio for connections from different presynaptic neurons. As suggested from experimental studies (20,39), more synapses survived if the presynaptic neuron had a RF near the postsynaptic RF after learning (Fig. 4A). Likewise, synapses having similar orientation selectivity to the postsynaptic neuron showed higher survival rates (Fig. 4B) as indicated from previous observations (5,39). However, this orientation dependence was evident only for projections from neurons with a RF in the direction of the postsynaptic orientation selectivity (blue line in Fig. 4C), and the

spines projected from neurons with orthogonal RFs remained to have uniform selectivity even after learning (green line in Fig. 4C), as reported in a recent experiment (20). In contrast, both connections from neurons with nearby and faraway RFs showed clear orientation dependence, though the dependence was more evident for the latter in the model (Fig. 4D). The consistencies with the experimental results (Fig. 4A-D) support the legitimacy of our model setting, though they were achieved by the elimination of uncorrelated spines, not by the multisynaptic learning rule per se.

We next investigated changes in dendritic synaptic organization generated by the multisynaptic learning. Overall, the mean spine size was slightly larger at distal dendrites (red line in Fig. 4E), but this trend was not strong enough to compensate the dendritic attenuation (black line in Fig. 4E), being consistent with previous observations in neocortical pyramidal neurons (41). Importantly, neurons with RFs faraway from the postsynaptic RF likely formed synaptic projections more on distal dendrites than on proximal ones (Fig. 4F), and at higher dendritic branch orders than at lower ones (Fig. 4G), as observed previously (20). This is because, in the proposed learning rule, if pre- and postsynaptic neurons have similar spatial selectivity, synaptic connections are preferably rewired toward proximal positions (Fig. 3G), and vice versa (Fig. 2A). Moreover, nearby spines on the dendrite showed similar RF selectivity even if multisynaptic pairs (i.e., synapse pairs projected from the same neuron) were excluded from the analysis (red line in Fig. 4H), due to the dendritic position dependence of presynaptic RFs. On the other hand, similarity between nearby spines was less significant in orientation selectivity (black line in Fig. 4H), as observed previously in rodent experiments (20,42). These results suggest a potential importance of developmental plasticity in somatic-distance dependent synaptic organization.

In the model, the position of a newly created synapse was limited to the branches where the presynaptic neuron initially had a projection, to roughly reproduce the spatial constraint on synaptic contacts. As a result, although there are many locations on the dendrite where the unit EPSP size is optimal for a given presynaptic neuron, only few of them are accessible from the neuron, hence synapses from the same presynaptic neuron may form clusters there. Indeed, by examining changes in multisynaptic connection structure, we found that the dendritic distance between two spines projected from the same presynaptic neuron became much shorter after learning (Fig. 4I), creating clusters of synapses from the same axons. This result suggests that clustering of multisynaptic connections observed in the experiments (6) is possibly caused by developmental synaptogenesis under a spatial constraint. Furthermore, as observed in hippocampal neurons (7), two synapses from the same presynaptic neuron had similar spine sizes if the connections were spatially close to each other, but the correlation in spine size disappeared if they were distant (red line in Fig. 4J). On the other hand, spine sizes of two synapses from different neurons were always uncorrelated regardless of the spine distance (black line in Fig. 4J).

Lastly, we studied the spine size distribution. In the proposed framework, the mean spine size does not essentially depend on presynaptic stimulus selectivity due to normalization, but the variance may change. In particular, the spine size variance is expected to be small if the presynaptic activity is highly stochastic, because the distribution of spine sizes stays nearly uniform in this condition, while the spine size variance should increase upon accumulation of samples. Indeed, in the initial phase of learning, the variance of spine size went up for projections from neurons with horizontal orientation selectivity (gray line Fig. 4K), though the spine size variance from other presynaptic neurons caught up eventually (black line Fig. 4K). In this regard, a recent experimental study found higher variability in postsynaptic density (PSD) areas for projections from neurons sharing orientation preference with the postsynaptic cell, though the data was from adult, not from juvenile mice (5).

The multisynaptic rule robustly enables fast learning

The correspondence with experiment observations discussed in the previous section supports the plausibility of our framework as a candidate mechanism of synaptic plasticity on the dendrites. Hence, we further studied the robustness of learning dynamics under the proposed multisynaptic rule. Below, we turn off the spine elimination mechanism that is not compensated by creation, as this process affects the learning dynamics.

In the proposed model, if the initial synaptic distribution on the dendrite $q_v(v)$ is close to the desired distribution $p_v(v)$, spine size modification is in principle unnecessary. In particular, the optimal EPSPs of most presynaptic neurons are small in our L2/3 model (Fig. 3C); hence most synaptic contacts should be placed on distal branches on average. Indeed, when the initial synaptic distribution was biased toward the distal side, improvement in classification performance became faster (black vs blue lines in Fig. 5A). This result suggests that the synaptic distribution on the postsynaptic dendrite may work as a prior distribution.

We next compared the learning performance with the standard monosynaptic learning rule in which the learning rate is a free parameter (see *Monosynaptic rule for the detailed model* in SI Appendix). If the learning rate is chosen at a small value, the neuron took a very large number of trials to learn the classification task (light gray line in Fig. 5B). On the other hand, if the learning rate is too large, the learning dynamics became unstable and the performance dropped off after a dozen trials (black line in Fig. 5B). Therefore, the learning performance was comparable with the multisynaptic rule only in a small parameter region ($\eta_w \sim 0.1$). By contrast, in the multisynaptic rule, stable fast learning was achievable without any fine-tuning (magenta line in Fig. 5B).

As expected from Figure 2, the proposed learning mechanism worked well even if the number of synapses per connection was small (Fig. 5C). Without rewiring, the classification task required seven synapses per connection for an 80% success rate, but three was enough with rewiring (Fig. 5C). Moreover, the learning performance was robust against synaptic failure (Fig. 5D). Although local excitatory inputs to L2/3 pyramidal cells have a relatively high release probability (43), the stochasticity of synaptic transmission at each synapse may affect learning and classification. We found that even if the half of presynaptic spikes were omitted at each synapse (see *Task configuration* in SI Appendix for details), the classification performance was still significantly above the chance level (Fig. 5D). Does the presynaptic stochasticity only add up noise? This was likely the case when the release probability was kept constant because the variability in the somatic EPSP height grows with the variance of $\{g_k\}$ in this scenario (SI Appendix, Fig. S2A; see *Presynaptic stochasticity* in SI Appendix for details). On the other hand, if matching exists between presynaptic release probability and the postsynaptic spine size as often observed in experiments (44,45), the Fano factor of the somatic EPSP height decreased as the performance went up (SI Appendix, Fig. S2B), because g_k can be jointly represented by the pre- and postsynaptic factors. This result indicates that the variability in somatic EPSP may encode the uncertainty in the synaptic representation.

In the proposed model, competition was assumed among synapses projected from the same presynaptic neuron, but it is unclear if homeostatic plasticity works in such a specific manner. Thus, we next constructed a surrogate learning rule that only requires a global homeostatic plasticity. In this rule, the importance of a synapse was not compared with other synapses from the same presynaptic neuron, but was compared with a hypothesized standard synapse (see *The surrogate learning rule* in SI Appendix). When the unit EPSP size of the standard synapse was chosen

Submission PDF

appropriately, the surrogate rule indeed enabled neuron to learn the classification task robustly and quickly (Fig. 5E). Overall, these results support the robustness and biological plausibility of the proposed multisynaptic learning rule.

Discussion

In this work, first we have used a simple conceptual model to show: (i) Multisynaptic connections provide a non-parametric representation of probabilistic distribution of the hidden parameter using redundancy in synaptic connections (Fig. 1AB); (ii) Updating of probabilistic distribution given new inputs can be performed by a Hebbian-type synaptic plasticity when the output activity is supervised (Fig. 1C-E); (iii) Elimination and creation of spines is crucial for efficient representation and fast learning (Fig. 2A-C). In short, synaptic plasticity and rewiring at multisynaptic connections naturally implements an efficient sample-based Bayesian filtering algorithm. Secondly, we have demonstrated that the proposed multisynaptic learning rule works well in a detailed single neuron model receiving stochastic spikes from many neurons (Fig. 3). Moreover, we found that the model reproduces the somatic-distance dependent synaptic organization observed in the L2/3 of rodent visual cortex (Fig. 4F and G). Furthermore, the model suggests that the dendritic distribution of multisynaptic inputs provides a prior distribution of the expected synaptic weight (Fig. 5A).

Experimental predictions

Our study provides several experimentally testable predictions on dendritic synaptic plasticity, and the resultant synaptic distribution. First, the model suggests a crucial role of developmental synaptogenesis in the formulation of presynaptic selectivity-dependent synaptic organization on the dendritic tree (Fig. 4F and G), observed in the primary visual cortex (20). More specifically, we have revealed that the RF-dependence of synaptic organization is a natural consequence of the Bayesian optimal learning under the given implementation. Evidently, retinotopic organization of presynaptic neurons is partially responsible for this dendritic projection pattern, as a neuron tends to make a projection onto a dendritic branch near the presynaptic cell body (8,46). However, a recent experiment reported that RF-dependent global synaptic organization on the dendrite is absent in the primary visual cortex of ferrets (47). This result indirectly supports the non-anatomical origin of the dendritic synaptic organization, as a similar organization is arguably expected in ferrets if the synaptic organization is purely anatomical.

Our study also predicts developmental convergence of synaptic connections from each presynaptic neuron (Fig. 3G and Fig. 4I). It is indeed known that in adult cortex, synaptic connections from the same presynaptic neuron are often clustered (4,6). Our model interprets synaptic clustering as a result of an experience-dependent resampling process by synaptic rewiring, and predicts that synaptic connections are less clustered in immature animal. In particular, our result suggests that synaptic clustering occurs in a relatively large spatial scale ($\sim$100μm; as shown in Fig 4I), not in a fine spatial scale ($\sim$10μm). This may explain a recent report on the lack of fine clustering structure in the rodent visual cortex (5).

Furthermore, our study provides an insight on the functional role of anti-Hebbian plasticity at distal synapses (21,22). Even if the presynaptic activity is not tightly correlated with the postsynaptic activity, that does not mean the presynaptic input is not important. For instance, in our detailed neuron model, inputs from neurons having a RF faraway from the postsynaptic RF still helps the postsynaptic neuron to infer the presented stimulus (Fig. 3). More generally, long-range inputs are typically not correlated with the output spike trains, because the inputs usually carry contextual information (48), or delayed feedback signals (49), yet play important moduratory roles. Our study indicates that anti-Hebbian plasticity at distal synapses prevents these connections from being eliminated, by keeping the synaptic connection strong. This may explain why modulatory inputs are often projected to distal dendrites (48,49), though active dendritic computation shuold also be crucial especially in case of Layer 5 or CA1 pryramidal neurons (24).

Related works

Previous theoretical studies often explain synaptic plasticity as stochastic gradient descent on some objective functions (17,40,50,51), but these models require fine-tuning of the learning rate for explaining near-optimal learning performance observed in humans (13,14) and rats (15), unlike our model. Moreover, in this study, we proposed synaptic dynamics during learning as a sample-based inference process, in contrast to previous studies in which sample-based interpretations were applied for neural dynamics (52).

The relationship between presynaptic stochasticity and the achievement level of learning has been studied before, yet the previous models required an independent tuning of pre- and postsynaptic factors (53,54). On the other hand, in our framework, the experimentally observed pre-post matching (44,45) is enough to approximately represent the uncertainty in learning performance by variability in the somatic membrane dynamics (SI Appendix, Fig. S2). It is known that presynaptic stochasticity can self-consistently generate a robust Poisson-like spiking activity in a recurrent network of leaky integrate-and-fire neurons (55). Hence, the uncertainty information reflected in the somatic membrane dynamics can be transmitted to downstream neurons via asynchronous spiking activity.

On the anti-Hebbian plasticity at distal synapse, previous modeling studies have revealed its potential phenomenological origins (56), but its functional benefits, especially optimality, have not been well investigated before. Particle filtering is an established method in machine learning (28), and has been applied to artificial neural networks (57), yet its biological correspondence had been elusive. A previous study proposed importance sampling as a potential implementation of Bayesian computation in the brain (58). In particular, they found that the oblique effect in orientation detection is naturally explained by sampling from a population with biased orientation selectivity. However, sampling was performed only in neural activity space, not in the synaptic parameter space unlike our model, and the underlying learning mechanism was not investigated either.

Previous computational studies on dendritic computation have been emphasizing the importance of active dendritic process (24), especially for performing inference from correlated inputs (59), or for computation at terminal tufts of cortical layer 5 or CA1 neurons (40). Nevertheless, experimental studies suggest the summation of excitatory inputs through dendritic tree is approximately linear (60,61). Indeed, we have shown that a linear summation of synaptic inputs is suitable for implementing importance sampling. Moreover, we have demonstrated that even in a detailed neuron model with active dendrites, a learning rule assuming a linear synaptic summation works well.

Material and Methods

In the conceptual model, $p(x_n=1)$ was set at 30%, and the conditional probability v_c was randomly chosen from [0,1) at each simulation (not at each trial). Except for Figure 2B, the number of connections was kept at $K = 10$. In the detailed single neuron model, we constructed a model of L2/3 pyramidal neuron using NEURON simulator (33), based on a previous model (34). Further details are given in SI Appendix.

Code availability

The simulation codes for the detailed neuron model are available at ModelDB (http://modeldb.yale.edu/225075 with access code "1234").

Acknowledgements

Submission PDF

The authors thank to Peter Latham for discussions and comments on the manuscript. This work was partly supported by CREST, JST (JPMJCR13W1 to TF) and Grants-in-Aid for Scientific Research (KAKENHI) from MEXT (no 15H04265, 16H01289, and 17H06036 to TF).

1. Deuchars J, West DC, Thomson AM (1994) Relationships between morphology and physiology of pyramid-pyramid single axon connections in rat neocortex in vitro. *J Physiol* 478(Pt 3):423–435.
2. Markram H, Lübke J, Frotscher M, Roth A, Sakmann B (1997) Physiology and anatomy of synaptic connections between thick tufted pyramidal neurones in the developing rat neocortex. *J Physiol* 500(2):409–440.
3. Feldmeyer D, Egger V, Lübke J, Sakmann B (1999) Reliable synaptic connections between pairs of excitatory layer 4 neurones within a single 'barrel' of developing rat somatosensory cortex. *J Physiol* 521(1):169–190.
4. Kasthuri N, et al. (2015) Saturated Reconstruction of a Volume of Neocortex. *Cell* 162(3):648–661.
5. Lee W-CA, et al. (2016) Anatomy and function of an excitatory network in the visual cortex. *Nature* 532(7599):370–374.
6. Schmidt H, et al. (2017) Axonal synapse sorting in medial entorhinal cortex. *Nature* 549(7673):469–475.
7. Bartol TM, et al. (2015) Nanoconnectomic upper bound on the variability of synaptic plasticity. *eLife*:e10778.
8. Gal E, et al. (2017) Rich cell-type-specific network topology in neocortical microcircuitry. *Nat Neurosci* 20(7):1004–1013.
9. Watanabe S (2001) Algebraic Analysis for Nonidentifiable Learning Machines. *Neural Comput* 13(4):899–933.
10. Amari S, Park H, Ozeki T (2006) Singularities Affect Dynamics of Learning in Neuromanifolds. *Neural Comput* 18(5):1007–1065.
11. Knill DC, Pouget A (2004) The Bayesian brain: the role of uncertainty in neural coding and computation. *Trends Neurosci* 27(12):712–719.
12. Körding KP, Wolpert DM (2006) Bayesian decision theory in sensorimotor control. *Trends Cogn Sci* 10(7):319–326.
13. Behrens TEJ, Woolrich MW, Walton ME, Rushworth MFS (2007) Learning the value of information in an uncertain world. *Nat Neurosci* 10(9):1214–1221.
14. Lake BM, Salakhutdinov R, Tenenbaum JB (2015) Human-level concept learning through probabilistic program induction. *Science* 350(6266):1332–1338.
15. Madarasz TJ, et al. (2016) Evaluation of ambiguous associations in the amygdala by learning the structure of the environment. *Nat Neurosci* 19(7): 965–972.
16. Soltani A, Wang X-J (2010) Synaptic computation underlying probabilistic inference. *Nat Neurosci* 13(1):112–119.
17. Nessler B, Pfeiffer M, Buesing L, Maass W (2013) Bayesian Computation Emerges in Generic Cortical Microcircuits through Spike-Timing-Dependent Plasticity. *PLoS Comput Biol* 9(4):e1003037.
18. Aitchison L, Latham PE (2014) Bayesian synaptic plasticity makes predictions about plasticity experiments in vivo. *ArXiv14101029 Q-Bio*.
19. Gütig R (2016) Spiking neurons can discover predictive features by aggregate-label learning. *Science* 351(6277):aab4113.
20. Iacaruso MF, Gasler IT, Hofer SB (2017) Synaptic organization of visual space in primary visual cortex. *Nature* 547(7664):449–452.
21. Letzkus JJ, Kampa BM, Stuart GJ (2006) Learning Rules for Spike Timing-Dependent Plasticity Depend on Dendritic Synapse Location. *J Neurosci* 26(41):10420–10429.
22. Sjöström PJ, Häusser M (2006) A Cooperative Switch Determines the Sign of Synaptic Plasticity in Distal Dendrites of Neocortical Pyramidal Neurons. *Neuron* 51(2):227–238.
23. Stuart G, Spruston N (1998) Determinants of Voltage Attenuation in Neocortical Pyramidal Neuron Dendrites. *J Neurosci* 18(10):3501–3510.
24. Segev I, London M (2000) Untangling Dendrites with Quantitative Models. *Science* 290(5492):744–750.
25. Matsuzaki M, Honkura N, Ellis-Davies GCR, Kasai H (2004) Structural basis of long-term potentiation in single dendritic spines. *Nature* 429(6993):761–766.
26. Courville AC, Daw ND, Touretzky DS (2006) Bayesian theories of conditioning in a changing world. *Trends Cogn Sci* 10(7):294–300.
27. Robert C, Casella G (2013) *Monte Carlo Statistical Methods* (Springer Science & Business Media).
28. Doucet A, Godsill S, Andrieu C (2000) On sequential Monte Carlo sampling methods for Bayesian filtering. *Stat Comput* 10(3):197–208.
29. Staras K, et al. (2010) A Vesicle Superpool Spans Multiple Presynaptic Terminals in Hippocampal Neurons. *Neuron* 66(1):37–44.
30. Holtmaat A, Svoboda K (2009) Experience-dependent structural synaptic plasticity in the mammalian brain. *Nat Rev Neurosci* 10(9):647–658.
31. Douc R, Cappe O (2005) Comparison of resampling schemes for particle filtering. *ISPA 2005. Proceedings of the 4th International Symposium on Image and Signal Processing and Analysis, 2005.*, pp 64–69.
32. Yang G, et al. (2014) Sleep promotes branch-specific formation of dendritic spines after learning. *Science* 344(6188):1173–1178.
33. Hines ML, Carnevale NT (1997) The NEURON Simulation Environment. *Neural Comput* 9(6):1179–1209.
34. Smith SL, Smith IT, Branco T, Häusser M (2013) Dendritic spikes enhance stimulus selectivity in cortical neurons in vivo. *Nature* 503(7474):115–120.
35. Froemke RC (2015) Plasticity of Cortical Excitatory-Inhibitory Balance. *Annu Rev Neurosci* 38(1):195–219.
36. Bonin V, Histed MH, Yurgenson S, Reid RC (2011) Local Diversity and Fine-Scale Organization of Receptive Fields in Mouse Visual Cortex. *J Neurosci* 31(50):18506–18521.
37. Simoncelli EP, Olshausen BA (2001) Natural Image Statistics and Neural Representation. *Annu Rev Neurosci* 24(1):1193–1216.
38. Geisler WS, Perry JS, Super BJ, Gallogly DP (2001) Edge co-occurrence in natural images predicts contour grouping performance. *Vision Res* 41(6):711–724.
39. Ko H, et al. (2013) The emergence of functional microcircuits in visual cortex. *Nature* 496(7443):96–100.
40. Urbanczik R, Senn W (2014) Learning by the Dendritic Prediction of Somatic Spiking. *Neuron* 81(3):521–528.
41. Williams SR, Stuart GJ (2003) Role of dendritic synapse location in the control of action potential output. *Trends Neurosci* 26(3):147–154.
42. Jia H, Rochefort NL, Chen X, Konnerth A (2010) Dendritic organization of sensory input to cortical neurons in vivo. *Nature* 464(7293):1307–1312.
43. Branco T, Staras K (2009) The probability of neurotransmitter release: variability and feedback control at single synapses. *Nat Rev Neurosci* 10(5):nrn2634.
44. Harris KM, Stevens JK (1989) Dendritic spines of CA1 pyramidal cells in the rat hippocampus: serial electron microscopy with reference to their biophysical characteristics. *J Neurosci* 9(8):2982–2997.
45. Loebel A, Bé J-VL, Richardson MJE, Markram H, Herz AVM (2013) Matched Pre- and Post-Synaptic Changes Underlie Synaptic Plasticity over Long Time Scales. *J Neurosci* 33(15):6257–6266.
46. Markram H, et al. (2015) Reconstruction and Simulation of Neocortical Microcircuitry. *Cell* 163(2):456–492.
47. Scholl B, Wilson DE, Fitzpatrick D (2017) Local Order within Global Disorder: Synaptic Architecture of Visual Space. *Neuron* 96(5):1127–1138.e4.
48. Bittner KC, et al. (2015) Conjunctive input processing drives feature selectivity in hippocampal CA1 neurons. *Nat Neurosci* 18(8):1133–1142.
49. Manita S, et al. (2015) A Top-Down Cortical Circuit for Accurate Sensory Perception. *Neuron* 86(5):1304–1316.
50. Pfister J-P, Toyoizumi T, Barber D, Gerstner W (2006) Optimal Spike-Timing-Dependent Plasticity for Precise Action Potential Firing in Supervised Learning. *Neural Comput* 18(6):1318–1348.
51. Hiratani N, Fukai T (2016) Hebbian Wiring Plasticity Generates Efficient Network Structures for Robust Inference with Synaptic Weight Plasticity. *Front Neural Circuits* 10:41.
52. Orbán G, Berkes P, Fiser J, Lengyel M (2016) Neural Variability and Sampling-Based Probabilistic Representations in the Visual Cortex. *Neuron* 92(2):530–543.
53. Aitchison L, Latham PE (2015) Synaptic sampling: A connection between PSP variability and uncertainty explains neurophysiological observations. *ArXiv150504544 Q-Bio*.
54. Costa RP, et al. (2017) Synaptic Transmission Optimization Predicts Expression Loci of Long-Term Plasticity. *Neuron* 96(1):177–189.e7.
55. Moreno-Bote R (2014) Poisson-Like Spiking in Circuits with Probabilistic Synapses. *PLOS Comput Biol* 10(7):e1003522.
56. Graupner M, Brunel N (2012) Calcium-based plasticity model explains sensitivity of synaptic changes to spike pattern, rate, and dendritic location. *Proc Natl Acad Sci* 109(10):3991–3996.
57. Freitas JFG de, Niranjan M, Gee AH, Doucet A (2000) Sequential Monte Carlo Methods to Train Neural Network Models. *Neural Comput* 12(4):955–993.
58. Shi L, Griffiths TL (2009) Neural Implementation of Hierarchical Bayesian Inference by Importance Sampling. *Advances in Neural Information Processing Systems 22*, pp 1669–1677.
59. Ujfalussy BB, Makara JK, Branco T, Lengyel M (2015) Dendritic nonlinearities are tuned for efficient spike-based computations in cortical circuits. *eLife*:e10056.
60. Cash S, Yuste R (1999) Linear Summation of Excitatory Inputs by CA1 Pyramidal Neurons. *Neuron* 22(2):383–394.
61. Hao J, Wang X, Dan Y, Poo M, Zhang X (2009) An arithmetic rule for spatial summation of excitatory and inhibitory inputs in pyramidal neurons. *Proc Natl Acad Sci* 106(51):21906–21911.

Submission Draft