entropy

Article

Modeling of a Neural System Based on Statistical Mechanics

Myoung Won Cho [1,*] and Moo Young Choi [2]

[1] Department of Global Medical Science, Sungshin Women's University, Seoul 01133, Korea
[2] Department of Physics and Astronomy and Center for Theoretical Physics, Seoul National University, Seoul 08826, Korea; mychoi@snu.ac.kr
* Correspondence: mwcho@sungshin.ac.kr; Tel.: +82-2-920-7280

Received: 28 August 2018; Accepted: 2 November 2018; Published: 5 November 2018

Abstract: The minimization of a free energy is often regarded as the key principle in understanding how the brain works and how the brain structure forms. In particular, a statistical-mechanics-based neural network model is expected to allow one to interpret many aspects of the neural firing and learning processes in terms of general concepts and mechanisms in statistical physics. Nevertheless, the definition of the free energy in a neural system is usually an intricate problem without an evident solution. After the pioneering work by Hopfield, several statistical-mechanics-based models have suggested a variety of definition of the free energy or the entropy in a neural system. Among those, the Feynman machine, proposed recently, presents the free energy of a neural system defined via the Feynman path integral formulation with the explicit time variable. In this study, we first give a brief review of the previous relevant models, paying attention to the troublesome problems in them, and examine how the Feynman machine overcomes several vulnerable points in previous models and derives the outcome of the firing or the learning rule in a (biological) neural system as the extremum state in the free energy. Specifically, the model reveals that the biological learning mechanism, called spike-timing-dependent plasticity, relates to the free-energy minimization principle. Basically, computing and learning mechanisms in the Feynman machine base on the exact spike timings of neurons, such as those in a biological neural system. We discuss the consequence of the adoption of an explicit time variable in modeling a neural system and the application of the free-energy minimization principle to understanding the phenomena in the brain.

Keywords: neural network model; statistical mechanics; free-energy minimization principle

1. Introduction

A neural network is a specific system to task with computing, memorizing, and thinking. The history of neural network modeling unfolded with the work of McCulloch and Pitts [1]. They introduced the notion of a formal neuron as a two-state threshold element and showed how a network of such elements can perform logical calculations. Their work paved the development of the artificial neuron models which are designed to mimic aspects of biological counterparts. Meanwhile, the first biological neuron model, introduced by Hodgkin and Huxley [2], expresses how the action potential in a neuron is initiated and propagates according to a set of nonlinear differential equations that approximate the electrical characteristics of a neuron. While artificial neurons usually take continuous values of activity on the discrete time domain, biological neurons give off discrete firing pulses on the continuous time domain. A firing-rate neuron model, which can be described as a low-pass filtered version of the biological neuron model, makes the bridge between the activity rule in the artificial neuron model and the firing dynamics in the biological neuron model. However, the firing-rate model, based on the neural activities on the time scale of tens of milliseconds or the moving average of firing rate, may not account for aspects of spike timings and correlations on the millisecond time scale.

The learning process is another important dynamics of a neural system. A neural system can memorize or learn via changing synaptic strengths, depending on neural activities. According to the basic mechanism for synaptic plasticity proposed by Hebb [3], an increase in synapse efficacy arises from the presynaptic cell's repeated and persistent stimulation of the postsynaptic cell. The resulting Hebbian theory attempts to explain associative learning, in which simultaneous activation of cells leads to a pronounced increase in the synaptic strength between those cells and provides a biological basis for errorless learning methods for education and memory rehabilitation. A firing-rate model, which is an artificial neuron model, describes the increase in the synaptic connection strength as a function of pre- and post-synaptic neural activities, e.g., as the product of the activities. On the other hand, series of experiments have revealed that the plasticity of a biological synapse occurs depending on the spike correlations on the millisecond time scale [4,5]. The learning behavior in a biological neural network could take after a learning rule in an artificial neuron model if spikes have weak correlations in the high-frequency regime; however, the biological learning mechanism, called the spike-timing-dependent plasticity (STDP) rule, can provide the system with more plentiful and useful learning behavior. One of them is the execution of competitive learning behavior, which is necessary for the development of a neural network with adaptive functions. Traditionally, inhibitory connections between neurons are believed to be essential for the execution of competitive learning behavior in a neural network. In contrast, the STDP rule can introduce competitive learning behavior to a neural network without inhibitory connections [6–9].

Meanwhile, neurons are inherently probabilistic in responding to external inputs, owing to the presence of noisy currents or chaotic behaviors associated with nonlinearity. Some neuron models describe the stochastic firing dynamics through differential equations with stochastic terms or Langevin dynamic equations; some other models do this through the Markov-chain Monte Carlo processes. Further, they describe the firing rule and/or the learning rule via statistical mechanism, and facilitate the use of statistical mechanics, which aims to explain the measurable properties of macroscopic systems on the basis of the behavior of the microscopic constituents of those systems. While an Ising-type model, well known to describe phase transitions in statistical mechanics, was first proposed by Little [10,11], a landmark was presented by Hopfield, who suggested a fruitful analogy between the asymptotic dynamics of such networks and equilibrium properties of random magnetic systems [12]. His work paved the road to the introduction of concepts and methods of statistical mechanics of random physical systems to the study of neural networks. Further, there are suggestions that the minimization of a free-energy may be the key principle in understanding how the brain works and how the brain structures form [13].

A statistical-mechanics-based model usually describes the firing rule and/or the learning rule as Markov-chain Monte Carlo processes of an energy function or gradient descents in a free energy. The free energy or the entropy of a neural model system is often defined intuitively as an objective quantity in information theory [13] (cf. Section 3.3). From the viewpoint of statistical physics, however, how to define a proper (free) energy in a complex and open system is far from a plain problem with an evident solution. One of the stringent problems is that an energy function should satisfy the invariance condition under the exchange of element positions, so that interaction strength between two elements should be symmetric. Some of the existing models solve the problem by assuming that all connections are symmetric (Sections 3.1 and 3.2), by treating asymmetric connections as a factor in external inputs (Section 3.4), or by estimating the entropy from the firing covariance in a Langevin dynamic equation (Section 3.4). If the energy and possible configurations of a system are well defined, the free energy of the system may be derived strictly from the internal energy and the entropy. Then, the minimization of the free energy brings about the minimization of the internal energy and/or the maximization of the entropy, where the dominance of either part is controlled by the temperature. The internal energy reaches its minimum at the extremum state, which is usually determined by recursive interactions between neurons or synaptic connections. The entropy relates to random fluctuations of firing states from the extremum state. Such random fluctuations in the firing dynamics and/or the learning dynamics is often adopted to avoid trapping in local minima during the learning process. On the other hand, some models make use of the entropy maximization principle as the key mechanism to invest a neuron model system with adequate competitive relationship between synaptic connections (Sections 3.3 and 3.4).

In the Feynman machine, the free energy is defined via the path integral, rather than from the Boltzmann distribution. Adopting asymmetric connections as well as asymmetric interactions between neurons in the time direction, the path integral formulation fulfils the invariance condition. Then, minimization of the free energy introduces a proper competitive learning rule to the Feynman machine. In particular, the biological synaptic plasticity rule is derived from the free-energy minimization principle. Attributes of firing dynamics or learning dynamics in a Feynman machine can resemble those in a conventional neuron model in a limit condition. However, the Feynman machine utilizes the exact timing of neural firings as the essential quantity in the implementation of computing or learning mechanism, such as biological neurons performing a function or communication through the use of the sensitivity for spike timings on the millisecond time scale [14,15].

2. Prologue

A firing-rate model expresses neural dynamics in terms of continuous variable $u_i(t)$, which represents the moving average of neural firings, i.e., the firing rate of neuron i. The time-dependent firing rate is often modeled as

$$\tau\frac{du_i}{dt} = -u_i + f(v_i), \tag{1}$$

where τ is the relaxation time and f is the activation function, with v_i being the scaled membrane potential of neuron i. If neuron i receives synaptic currents from neuron j via connection W_{ij} and receives external input h_i, the scaled membrane potential is given by the sum of synaptic currents and external inputs:

$$v_i = \sum_j W_{ij}u_i + h_i. \tag{2}$$

At the steady state, defined by $du_i/dt = 0$, the firing rate takes the simple form $u_i = f(v_i)$, which corresponds to the activity rule in an artificial neuron model. According to the simple Hebbian rule, the change of the connection strength is given by the product of the activities of pre- and post-synaptic neurons:

$$\Delta W_{ij} \propto u_i u_j. \tag{3}$$

Meanwhile, a two-state model or a spike-timing-based model expresses neural dynamics in terms of binary variable $\phi_i(t)$, which can only take values of 0 and 1. In the case of discrete time representation, $\phi_i(t)$ specifies the firing state of neuron i at time t; in the continuous time representation, it measures the number of firings in a small time interval. A biological synapse changes its strength according to the spike-timing-dependent plasticity (STDP) rule:

$$\Delta W_{ij} = \Omega_{ij}(t{-}t')\phi_i(t)\phi_j(t'), \tag{4}$$

where the STDP window $\Omega_{ij}(t)$ determines the change of the connection strength W_{ij} depending on the difference between the firing timings of post- and pre-synaptic neurons. On a long-time scale, the connection strength change is given by

$$\Delta W_{ij} = \int dt\, \Omega_{ij}(t)\, C_{ij}(t), \tag{5}$$

where $C_{ij}(t-t') = \langle \phi_i(t)\phi_j(t')\rangle$ is the cross-correlation function of neurons i and j. In general, the STDP window produces long-term potentiation (LTP) for pre-before-post pairings and long-term depression (LTD) for post-before-pre pairings (see Figure 1a). It often has a second interval of depression for pre-before-post pairings. The LTP part leads to an increase in the synaptic strength for a causally correlated neural firing pair. On the other hand, the LTD part, bringing about a decrease in the synaptic strength for a neural firing pair with weak causality, may introduce the competitive learning behavior in the absence of inhibitory connections. The even part and the odd part of the STDP window, $\Omega_{ij}^{(\pm)} \equiv \frac{1}{2}[\Omega_{ij}(t) \pm \Omega_{ij}(-t)]$, determine the changes in the sum of and in the difference between reciprocal connections, $W_{ij}^{(\pm)} \equiv \frac{1}{2}(W_{ij} \pm W_{ji})$, respectively, as follows:

$$\Delta W_{ij}^{(\pm)} = \int dt\, \Omega_{ij}(t)\, C_{ij}^{(\pm)}(t) = \int dt\, \Omega_{ij}^{(\pm)}(t)\, C_{ij}(t) \tag{6}$$

with $C_{ij}^{(\pm)} \equiv \frac{1}{2}[C_{ij}(t) \pm C_{ji}(t)] = \frac{1}{2}[C_{ij}(t) \pm C_{ij}(-t)]$. Namely, the change of the symmetric connection strength W_{ij}^{+} is determined by the even part of the STDP window. Heretofore, the symmetric connection strength is denoted by J_{ij}; a function of such symmetric connection strengths could serve as an energy function.

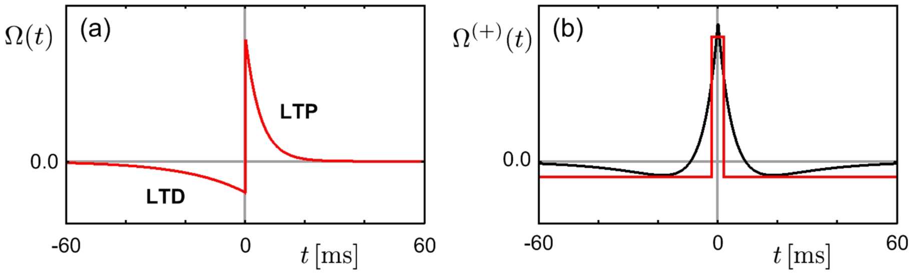

Figure 1. (**a**) Plot of a STDP window $\Omega(t)$, given by $\Omega(t) = Ae^{-t/2}$ for $t > 0$ and $-(A/4)e^{t/20}$ for $t < 0$ with a constant A. (**b**) Plot of the even part $\Omega^{(+)}$, given by $\Omega^{(+)} = A$ for $|t| \le 2$ and $-A/8$ otherwise (red line) and by $\Omega^{(+)} = \frac{1}{2}[\Omega(t) + \Omega(-t)]$ with $\Omega(t) = 1.5Ae^{-t/2}$ for $t > 0$ and $-(1.5A/4)e^{t/20}$ for $t < 0$ (black line).

3. Statistical-Mechanics-Based Models

3.1. Hopfield Network

A Hopfield network consists of a group of neurons together with symmetric connections between them. The state of neuron i is represented by the variable $\sigma_i \equiv 2\phi_i - 1$, which takes values ±1. The connection strengths between neurons are symmetric ($J_{ij} = J_{ji}$) and have no self-connection ($J_{ii} = 0$). In the learning process, the Hopfield network memorizes a set of p binary patterns $\{\sigma^{(n)}\}$ ($n = 1, 2, \ldots, p$) with $\sigma_i^{(n)} \in \{-1, 1\}$, which is achieved by setting the connections as the sum of outer products:

$$J_{ij} = \begin{cases} \eta \sum_{n=1}^{p} \sigma_i^{(n)} \sigma_j^{(n)} & \text{for } i \neq j \\ 0 & \text{for } i = j \end{cases} \tag{7}$$

with the constant η specifying the overall connection strength. In the firing process, the state of each neuron is updated according to the following rule

$$\sigma_i(t+1) = \begin{cases} +1 & \text{if } \sum_j J_{ij}\sigma_i(t) + B_i \geq 0 \\ -1 & \text{otherwise} \end{cases}, \tag{8}$$

where the constant B_i controls the bias of neuron i. Namely, the state of neuron i is updated as if the neuron had a threshold activation function with the threshold $-B_i$. The Hopfield network possesses the desired memory patterns $\{\sigma^{(n)}\}$ as its attractors, so that it usually flows to one of them depending on its initial state.

It is of interest that the update rule can be understood as the single spin-flip sampling in an Ising model system with the energy function

$$E = -\frac{1}{2}\sum_{i,j} J_{ij}\sigma_i\sigma_j - \sum_i B_i\sigma_i. \tag{9}$$

Whereas the deterministic update rule in Equation (8) corresponds to the zero-temperature dynamics, one may generalize it easily to finite temperatures by introducing probabilities. Specifically, adopting the importance sampling Markov process or the Monte Carlo method, we consider the probabilistic update in such a way that the probability of flipping σ_i is given by $P(\sigma_i \to -\sigma_i) \propto \exp(-\Delta E_i/T)$, where ΔE_i is the energy difference in flipping of σ_i. It is obvious that the important sampling process reduces to the deterministic update rule in Equation (8) as the temperature T approaches zero.

It is also of interest to note that only a single neuron is updated at each time step. Such asynchronous dynamics contrasts with the synchronous dynamics where all the neurons are updated simultaneously [11]. Neither dynamics provides a very realistic description of the dynamics in real networks, which presumably lies in between. As an attempt toward a more realistic description of the dynamics, a dynamic model working in continuous time but taking into account relevant time scales was also presented [16].

3.2. Boltzmann Machine

The Boltzmann machine, which is often referred to as a Monte Carlo version of the Hopfield network [17], has the energy function given by

$$E[\phi] = -\frac{1}{2}\sum_{i,j} J_{ij}\phi_i\phi_j - \sum_i B_i\phi_i, \tag{10}$$

where ϕ_i can only take values of 0 and 1 and $B_i = h_i + \theta_i$ with h_i being the external input to neuron i and θ_i the negative of the activation threshold in the system running freely without external inputs. The neurons are divided into "visible" and "hidden" ones. The external input h_i is applicable only if neuron i is a visible one and the system is in the training phase. In the case of the Metropolis importance sampling [18], the state of each neuron flips with the probability

$$P(\phi_i \to 1-\phi_i) = \min[e^{-\Delta E_i/T}, 1], \tag{11}$$

where ΔE_i is the energy difference in the flipping of ϕ_i. The expectation value of ϕ_i in the Monte Carlo process then obtains the form

$$\langle\phi_i\rangle = \frac{1}{Z}\mathrm{Tr}\,\phi_i\, e^{-E[\phi]/T}, \tag{12}$$

where $Z = \mathrm{Tr}\, e^{-E[\phi]/T}$ is the partition function with the trace standing for the summation over all configurations ($\mathrm{Tr} \equiv \prod_k \sum_{\phi_k=0,1}$). The expectation value can also be obtained from the derivative of the partition function with respect to the external source in the following way:

$$\langle \phi_i \rangle = \frac{T}{Z}\frac{\partial Z}{\partial B_i} = -\frac{\partial F}{\partial B_i}, \tag{13}$$

where $F = -T \log Z$ is the free energy of the system. Similarly, the cross-correlation function between two neurons i and j reads

$$\langle \phi_i \phi_j \rangle = \frac{1}{Z}\mathrm{Tr}\, \phi_i \phi_j \, e^{-E[\phi]/T} = \frac{T^2}{Z}\frac{\partial^2 Z}{\partial B_i \partial B_j}. \tag{14}$$

The learning rule for the Boltzmann machine is expressed originally in the form of a gradient descent process

$$\Delta J_{ij} \propto -\frac{\partial L}{\partial J_{ij}}, \tag{15}$$

where L is the relative entropy or Kullback–Leibler (KL) divergence given by

$$L = \mathrm{Tr}\, P(\phi) \log \frac{P(\phi)}{P_0(\phi)}. \tag{16}$$

The KL divergence measures the distance between probability distributions $P_0(\phi)$ and $P(\phi)$, vanishing if and only if the two distributions are identical. Here, $P(\phi) \propto e^{-E[\phi]/T}$ is the probability of the state $\phi = \{\phi_k\}$ of visible neurons when the system is in the training phase. $P_0(\phi) \equiv P(\phi)|_{h=0}$ is the corresponding probability for the network running freely with no external input. After some calculation, the change in a connection strength is obtained as

$$\Delta J_{ij} \propto C_{ij} - C_{0,ij}, \tag{17}$$

where the correlation function $C_{ij} \equiv \langle \phi_i \phi_j \rangle$ describes the average probability of two neurons both being in the *on* state with the environment clamping the states of the visual units, and $C_{0,ij} \equiv C_{ij}|_{h=0}$ is the corresponding probability for the network running freely without external inputs.

Interestingly, there is similarity between the learning rules based on the gradient descent in the KL divergence and in the free energy. Applying the density-matrix formalism with the correlation function taken as the density matrix $C = \{C_{ij}\}$, we express the free energy of the Boltzmann machine as $F = U - TS$ with the internal energy $U = -\frac{1}{2}\sum_{i,j} J_{ij} C_{ij} - \sum_i B_i \rho_i$ and the entropy $S = -\mathrm{tr}\, C \log C$, where $\rho_i \equiv \langle \phi_i \rangle$ stands for the firing probability of neuron i. With the condition $\sum_j J_{ij}\rho_j + B_i = 0$ in the extremum state, the gradient descent in the free energy takes the form

$$\Delta J_{ij} \propto C_{ij} + T\frac{\partial S}{\partial J_{ij}} - \frac{1}{2}\sum_{k,l} J_{kl}\left[\frac{\partial C_{kl}}{\partial J_{ij}} - 2\rho_k \frac{\partial \rho_l}{\partial J_{ij}}\right]. \tag{18}$$

Note that the derivative of the entropy with respect to the connection strength does not depend on external inputs.

Further, there is another learning rule for the Boltzmann machine, which is a simplified version of the STDP rule [9]. It has been developed to explore the emergent structure in a neural network running freely, motivated by the report that the STDP rule leads to the development of small-world and scale-free graphs in simulations [19,20]. As well known, small-world and scale-free properties are ubiquitous in complex networks [21,22]. Conventional models explain that a small-world network, characterized by short path lengths and high clustering, emerges as a result of randomly replacing a fraction of links on a lattice by new, randomly chosen links [23] and that a scale-free network emerges from stochastic growth in which new nodes are added continuously and attach themselves preferentially to existing nodes [24]. However, there are such empirical networks as the brain network, possessing small-world and/or scale-free properties, to which the conventional models are not applicable. Specifically, a neural network with a static number of neurons acquires the scale-free properties not as a result of the preferential attachment or rich-get-richer process in a growing network but as an equilibrium state of the connecting–disconnecting processes. In the biological learning rule, the connecting and the disconnecting processes are related to the LTP and the LTD parts of the STDP window, respectively.

In the model, neurons flip their states according to the importance sampling rule in Equation (11). Simultaneously, the strength of each connection, which can take values J and 0, changes with the probability

$$P(J_{ij} \to J - J_{ij}) = \frac{1}{\tau} \min\,[e^{-\Delta\mathcal{F}_{ij}/T}, 1], \tag{19}$$

where τ controls the ratio of time scales in the firing and the learning process and $\Delta\mathcal{F}_{ij}$ is the energy difference associated with flipping J_{ij}. The energy $\mathcal{F}$ in the learning process is given by

$$\mathcal{F} = -\frac{1}{2}\sum_{i,j}(C_{ij} - \eta\rho_i\rho_j + \mu)J_{ij}, \tag{20}$$

where the cross-correlation C_{ij} and the firing probability $\rho_i \equiv \langle\phi_i\rangle$ are measured in a moving time window (of about 100 steps). The competition strength η controls the contributions of interactions relative to those of independent activations; μ is the wiring propensity (which is opposite to the wiring cost). The value of J is chosen to be sufficiently larger than T and the value of τ to be a small number as usual. Figure 2 shows the log-log plots of the cumulative degree distribution in neural networks, which exhibit the typical features of scale-free graphs.

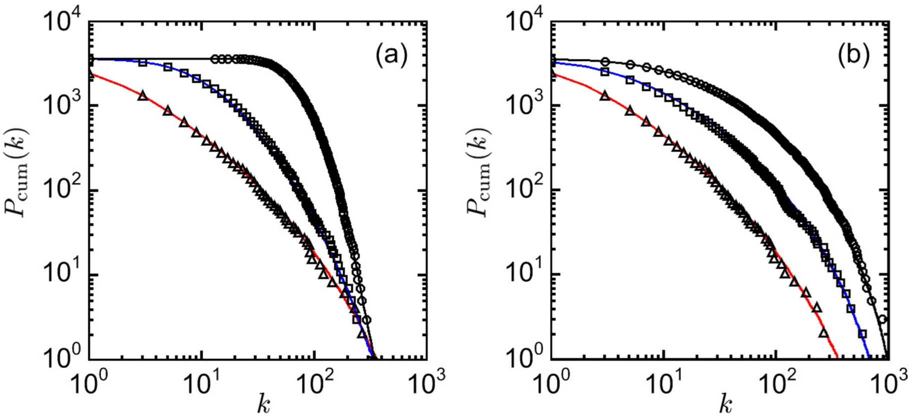

Figure 2. Log-log plots of the cumulative degree distribution $P_{\rm cum}(k) \equiv \sum_{k'=k}^{\infty} P(k')$, with the degree distribution $P(k)$ in a neural network, for different values of parameters: (**a**) competition strength η (= 20, 10, and 5 for triangles, squares, and circles, respectively); and (**b**) wiring propensity μ (= $17T$, $19T$, and $21T$ for triangles, squares, and circles, respectively). Lines are least-square fits of the data.

In the above, the connection strength has been taken to be a binary number for convenience in Monte Carlo simulations and in the discriminant of connected and disconnected neural pairs. In the case that the connection strength takes continuous values, the learning process could be expressed as the gradient descent in the energy function:

$$\Delta J_{ij} \propto -\frac{\partial \mathcal{F}}{\partial J_{ij}} = C_{ij} - \eta \rho_i \rho_j + \mu + \frac{1}{2} \sum_{k,l} J_{kl} \left[\frac{\partial C_{kl}}{\partial J_{ij}} - 2\eta \rho_k \frac{\partial \rho_l}{\partial J_{ij}} \right], \tag{21}$$

which is similar to the learning process in Equation (17) or Equation (18). The form of $\partial \mathcal{F}/\partial J_{ij}$ originates from an approximate of the even part of the STDP window (see Figure 1b). Note also that $\mathcal{F}$ often serves as a free energy rather than energy. Henceforth, it is demonstrated that the biological synaptic plasticity rule can be interpreted as the minimization process for a free energy (see Section 5).

3.3. Informatix Rule

The informatix rule does not employ the definition of the energy or free energy of a neural system. Nevertheless, in view of informatics, it provides an important hint as to how the entropy maximization principle admits a stochastic neural network with competitive learning behavior. Consider a neural network composed of two input–output layers and only feedforward connections from input to output neurons. Suppose that the activity of neuron i is given by $u_i = f(v_i)$ with the scaled membrane potential v_i given by Equation (2), where the feedforward connection W_{ij} is relevant only for $i \in \mathcal{G}_o$ and $j \in \mathcal{G}_i$ with $\mathcal{G}_i$ (or $\mathcal{G}_o$) denoting the set of input (or output) neurons, and the external input h_i is applicable only for $i \in \mathcal{G}_i$. According to the informatics, the information transfer is measured by the joint entropy

$$H(u) = -\int \mathcal{D}u P(u) \log P(u) \tag{22}$$

with $\int \mathcal{D}u \equiv \prod_k \int_{-\infty}^{\infty} du_k$. The joint entropy can be rewritten as

$$H(u) = \sum_{i \in \mathcal{G}_o} H(u_i) - I(u), \tag{23}$$

where $H(u_i)$ is the marginal entropy of the output u_i and $I(u)$ the mutual information of the outputs. They obtain the form

$$H(u_i) = -P(u_i) \log P(u_i) \tag{24}$$

and

$$I(u) = \int \mathcal{D}u \, P(u) \log \frac{P(u)}{\prod_{i \in \mathcal{G}_o} P(u_i)}. \tag{25}$$

In informatics, an ideal learning process is to maximize the joint entropy or the information transfer from input to output neurons. Provided that the input and the output layers have the same number of neurons, the derivative of the joint entropy with respect to the connection strength is given by

$$\frac{\partial H(u)}{\partial W_{ij}} = \frac{\partial I(u)}{\partial W_{ij}} = \frac{\partial}{\partial W_{ij}} \mathrm{tr} \log(\mathcal{J}), \tag{26}$$

where other terms are assumed to be independent of W and $\mathcal{J}_{ij} = \partial u_i / \partial u_j$ is the Jacobian of the information transfer from neuron j to neuron i. When there is no lateral connection between output neurons, the Jacobian would become $\mathcal{J}_{ij} = W_{ij}(\partial u_i/\partial v_i)$. Finally, the learning rule maximizing the joint entropy reads

$$\Delta W_{ij} = \frac{\partial}{\partial W_{ij}} \mathrm{tr} \log(W) + \frac{\partial}{\partial W_{ij}} \sum_k \log \frac{\partial u_k}{\partial v_k}, \tag{27}$$

where $\partial \mathrm{tr} \log(W)/\partial W_{ij}$ reduces to $(W^{-1})_{ji}$ with the inverse matrix W^{-1}. The learning rule described by Equation (27) corresponds to the competition mechanism in the PSL model or in the Feynman machine, as shown in the following sections. A further manipulation of this equation gives the informatix rule [25], which is in turn related to the independent component analysis (ICA), a popular algorithm for blind source separation [26].

3.4. Pseudo-Stochastic Learning Model

Consider a neural network composed of input–output layers, where output neurons have feedforward connections W from input neurons and lateral connections J with other output neurons. Neglecting the activation function f or assuming $u_i = v_i$, we write the time-dependent firing rate in the form

$$\tau \frac{du_i}{dt} = -u_i + \sum_j J_{ij} u_j + \sum_j W_{ij} u_j + h_i, \tag{28}$$

where the lateral connection J_{ij} is applicable only for $i, j \in \mathcal{G}_o$, the feedforward connection W_{ij} only for $i \in \mathcal{G}_o$ and $j \in \mathcal{G}_i$, and the external input h_i only for $i \in \mathcal{G}_i$.

At stationarity ($du_i/dt = 0$), Equation (28) gives the neural activity in the form

$$u_i = \begin{cases} h_i & \text{for } i \in \mathcal{G}_i \\ \sum_{k,l} K_{ik} W_{kl} h_l & \text{for } i \in \mathcal{G}_o, \end{cases} \tag{29}$$

where K_{ij} is the recursive lateral interaction, given by a component of the matrix $K \equiv (I - J)^{-1}$ with I being the identity. Namely, letting $D \equiv I - J$, we have $K_{ij} = D^{-1}_{ij} = (I + J + J^2 + \cdots)_{ij}$. Adopting the simple Hebbian rule in Equation (3), the feedforward connection changes as

$$\Delta W_{ij} \propto \sum_{k,l} K_{ik} W_{kl} h_l h_j. \tag{30}$$

In consideration of the learning process under varying external inputs, we write Equation (30) in the form

$$\Delta W_{ij} \propto (KWQ)_{ij}, \tag{31}$$

where $Q_{ij} \equiv \langle\langle h_i h_j \rangle\rangle$ measures the correlations of external inputs with $\langle\langle \cdot \rangle\rangle$ denoting the average in a long time period for varying external inputs. This is the correlation-based learning model for feature map formation [27]. A feature map formation model should have a competition mechanism to prevent output neurons from having the same features as neighbors. The correlation-based model achieves the competition mechanism through negative components in J or K originating from inhibitory lateral connections.

Figure 3 shows the emergent feature map, developed by the correlation-based model. It is noteworthy that the feature map displays the same characteristics as the feature map observed in the primary visual cortex (V1) area. In simulations, the recursive lateral interaction K has been modeled as a Mexican-hat-shaped function of distance. The output neurons would have the same features (i.e., ocular dominance) as others if the degree of the competition, controlled by k, is sufficiently small [28]. In general, the visual input correlation matrix is diagonalized owing to the symmetry properties of external inputs; the feedforward connections is then represented by low-dimensional feature vectors or spin-like variables. For example, the ocular dominance of visual cortex neurons, based on the difference between feedforward connections from left and right eyes, can be represented by Ising-type spin variables. Consequently, the feature map formation in a visual cortex area can be explained in terms of the energy of a spin-like model [28–32].

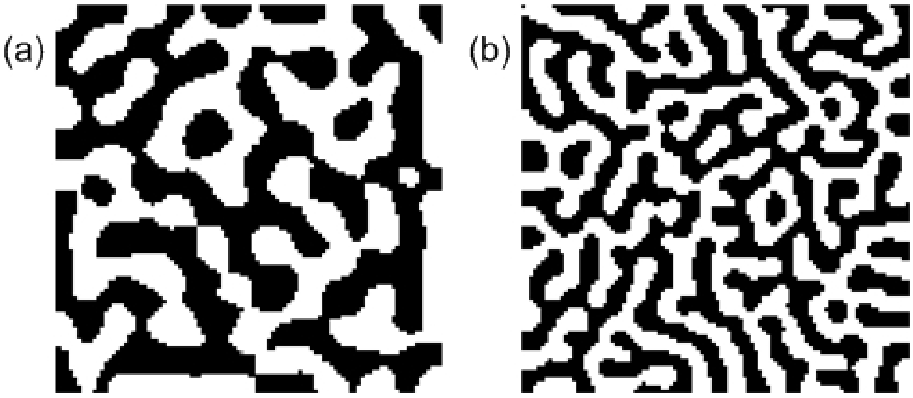

Figure 3. Ocular dominance map developed by the correlation-based learning model in Equation (31), where $K_{ij} = (1-kr_{ij}^2/\sigma^2)\exp(-r_{ij}^2/2\sigma^2)$ and $Q_{ij} = \delta_{ij} + \eta(1-\delta_{ij})$ with $\sigma = 2$, $\eta = 0.3$, and (**a**) $k = 0.3$ and (**b**) $k = 0.5$. The size of input and output layers are given by 2 and 100×100, respectively. The difference between the feedforward connections from the two input neurons, corresponding to the left and the right retina ganglion cells, becomes the ocular dominance of output neurons.

The firing activity and the learning rule in the correlation-based model can be derived through the use of statistical mechanics. Suppose that the activity of output neuron i is determined probabilistically, so that its expectation value is expressed as

$$\langle u_i \rangle = \frac{1}{Z} \int \mathcal{D}u \, u_i e^{-E[u]/T} \tag{32}$$

with the partition function $Z \equiv \int \mathcal{D}u \, e^{-E[u]/T}$ and a nonnegative constant T. Provided that J is a symmetric matrix, the energy function E is given by

$$E[u] = \frac{1}{2} \sum_{i,j \in \mathcal{G}_o} D_{ij} u_i u_j - \sum_{i \in \mathcal{G}_o} B_i u_i \tag{33}$$

with $B_i \equiv \sum_{k \in \mathcal{G}_i} W_{ik} h_k$. The expectation value is obtained from the derivative of the free energy $F = -T \log Z$ with respect to the external source:

$$\langle u_i \rangle = \frac{T}{Z} \frac{\partial Z}{\partial B_i} = -\frac{\partial F}{\partial B_i}. \tag{34}$$

Performing the Gaussian integral, we obtain the free energy as $F = U - TS$ with the internal energy

$$U = -\frac{1}{2} \sum_{i,j} K_{ij} B_i B_j = -\frac{1}{2} \sum_{i,j,k,l} K_{ij} W_{ik} W_{jl} h_k h_l \tag{35}$$

and the entropy

$$S = \frac{1}{2} \mathrm{tr} \log(G), \tag{36}$$

where the connected two-point function $G_{ij} \equiv \langle u_i u_j \rangle - \langle u_i \rangle \langle u_j \rangle$ obtains the form $G = TK$. In accord, the expectation value in Equation (34) reduces to

$$\langle u_i \rangle = \sum_k K_{ik} B_k = \sum_{k,l} K_{ik} W_{kl} h_l, \tag{37}$$

which agrees with Equation (29). Further, the learning rule in Equation (30) is obtained from the derivative of the free energy with respect to the connection strength:

$$\Delta W_{ij} \propto -\frac{\partial F}{\partial W_{ij}} = \sum_{k,l} K_{ik} W_{kl} h_l h_j. \tag{38}$$

Equations (34) and (38) demonstrate that both the firing process and the learning process in a neural network can be derived from the free energy of the system. The derivative of the free energy with respect to the external source yields the neural activity in an extremum state. In comparison with the derivative of the energy function, the derivative of the free energy includes the effects of recursive interactions between neurons. Then, the derivative of the free energy with respect to the connection strength offers a relevant learning rule of the system. Nevertheless, the derivative of the entropy in Equation (36) exerts effects neither on the firing process nor on the learning rule. It is plausible that the entropy, related to autonomous neural firings via thermal fluctuations, exerts no effects on the neural firing process, although the informatix rule allows that the entropy maximization induces a proper competition mechanism in a neural network without inhibitory connections.

On the other hand, the pseudo-stochastic learning (PSL) model suggests that the entropy maximization principle would exert meaningful effects on the learning rule when the entropy is obtained not from the Boltzmann distribution but from the neural firing correlations in a Langevin equation [33]. We introduce a noise term to Equation (28) and write

$$\tau \frac{du_i}{dt} = -u_i + \sum_j J_{ij} u_j + \sum_j W_{ij} u_j + h_i + \xi_i, \tag{39}$$

where ξ_i, referring to as endogenous neural firings via thermal fluctuations or external noisy currents, has the properties $\langle \xi_i \rangle = 0$, $\langle h_i \xi_j \rangle = 0$, and $\langle \xi_i \xi_j \rangle = \beta \delta_{ij}$ with constant β being proportional to the temperature T. The activities of input and output neurons in the steady state are given by $u_i = h_i + \xi_i$ and $u_i = \sum_j K_{ij} \left[\sum_k W_{jk} (h_k + \xi_k) + \xi_j \right]$, respectively, the expectation value of which agrees with Equation (29). The endogenous neural firings via thermal fluctuations exert no effects on the average over individual neural activities.

Meanwhile, the connected two-point function between output neurons becomes

$$G_{ij} = \beta \left[K K^\dagger + K W W^\dagger K^\dagger \right]_{ij}, \tag{40}$$

where $\beta K K^\dagger$ (or $\beta K W W^\dagger K^\dagger$) corresponds to the neural correlations originating from the autonomous firings of output (or input) neurons. With the substitution $G \approx T K W W^\dagger K^\dagger$ in Equation (36), the gradient flow of the free energy in Equation (38) leads to the learning rule

$$\Delta W_{ij} \propto \left[K W Q + T (W W^\dagger)^{-1} W \right]_{ij}, \tag{41}$$

where $(W W^\dagger)_{ij}$ is the inner product between incoming connection vectors onto output neurons i and j, so that $(W W^\dagger)^{-1} W = (I - I + W W^\dagger)^{-1} W \approx (I - W W^\dagger) W$ hinders output neurons from having the same feedforward connection pattern as others. Note that the second term in this equation corresponds to the first term in Equation (27) because $\frac{1}{2} \operatorname{tr} \log(W W^\dagger) = \operatorname{tr} \log(W)$ for a square matrix W.

Figure 4 presents a feature map, developed by the PSL model [33]. It has characteristics of a topographic map, such as the well-ordered connection distribution from retina (or LGN) cells to V1 neurons. In simulations, the lateral connections J (or K) have no negative values, but the entropy-originating term brings about adequate competition between output neurons.

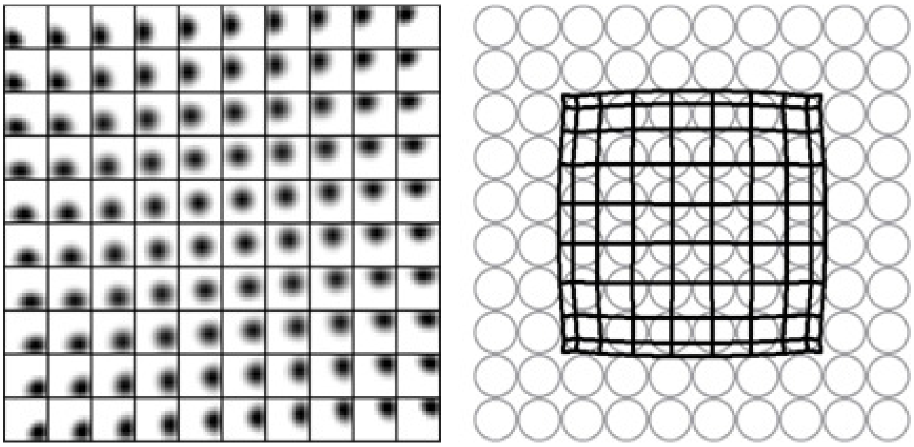

Figure 4. Topographic map developed by the PSL model in Equation (41). The **left** figure shows the receptive field (RF) of output neurons on the input layer, where the small boxes and the axes represent the connection strengths from input neurons to individual output neurons and the neuronal positions in the output layer, respectively. The **right** figure depicts the topographic map obtained from the connection strengths, where the gray circles and the mesh points represent, respectively, the positions of the input neurons and the RF centers of the output neurons on the input neurons. We have taken $K_{ij} = \exp(-r_{ij}^2)$, $Q_{ab} = \exp(-r_{ij}^2)$, and $T = 0.1$, along with the approximation $(WW^\dagger)^{-1} \approx I - WW^\dagger$ and the normalization condition $(WW^\dagger)_{ii} = 1$. Both the input and the output layers have the size 10×10.

4. Feynman Machine

The Feynman machine is a neural network model based on the free energy principle [34–37]. Both the firing and the learning rule in the model are derived from a single free energy, which is defined via the Feynman path integral rather than from the Boltzmann distribution. The typical form of a Feynman machine bases on an approximate solution of the firing dynamics equation in a biological neuron model:

$$C\frac{dV_i}{dt} = I_i^{\rm syn} + I_i^{\rm ion} + I_i^{\rm ext}, \tag{42}$$

where $V_i(t)$ is the membrane potential of neuron i at time t and C the capacitance per unit surface area. On the right-hand side, $I_i^{\rm syn}$, $I_i^{\rm ion}$, and $I_i^{\rm ext}$ represent the synaptic, ionic, and external stimulus currents, respectively. The synaptic current is given by

$$I_i^{\rm syn}(t) = \sum_j \int dt' g_{ij}^{\rm syn} \left[\mathcal{V}_{ij}^{\rm syn} - \mathcal{V}_i(t)\right] \alpha(t-t')\, \phi_j(t'), \tag{43}$$

where $g_{ij}^{\rm syn}$ is the maximum conductance per unit surface area (with $g_{ii}^{\rm syn} \equiv 0$), $\mathcal{V}_{ij}^{\rm syn} \equiv V_{ij}^{\rm syn} - V^{\rm eq}$ is the synaptic reversal potential $V_{ij}^{\rm syn}$ of the synaptic connection from neuron j to neuron i measured from the rest membrane potential $V^{\rm eq}$, and $\mathcal{V}_i(t) = V_i(t) - V^{\rm eq}$ is the membrane potential of neuron i from the rest potential. The function $\alpha(t)$ is typically modeled as $\alpha(t) = g(t)\exp[-g(t)]$ with $g(t) \equiv \max[(\tau_{\rm d} - t)/\tau_{\rm c}, 0]$, where $\tau_{\rm d}$ is the delay time and $\tau_{\rm c}$ the characteristic time. The ionic current is taken to be of a variety of form on the model-by-model basis.

It is rather formidable to obtain the exact solution of the firing dynamics equation in most models; however, in consideration of that $V_i(t)$ would converge to V^{eq} in the absence of external inputs and synaptic connections, we express the solution as a series in ϕ:

$$\mathcal{V}_i(t) = h_i(t) - \sum_j \int dt' \mathcal{D}^+_{ij}(t-t')\phi_j(t') + \mathcal{O}(\phi^2). \tag{44}$$

Here, $h_i(t)$ relates to the leaky integration of external stimulus current and $\mathcal{D}^+_{ij}(t) \approx -\lambda^{\text{syn}}(t)W_{ij}$ represents the synaptic interactions with other neurons, where

$$W_{ij} \equiv g^{\text{syn}}_{ij}\mathcal{V}^{\text{syn}}_{ij} = g^{\text{syn}}_{ij}(V^{\text{syn}}_{ij} - V^{\text{eq}}) \tag{45}$$

is the synaptic connection strength from neuron j to neuron i and the overall strength $\lambda^{\text{syn}}(t) = 0$ for $t < 0$. The exact form of $h_i(t)$ and $\lambda^{\text{syn}}(t)$ is obtained theoretically for simple biological neuron models such as the leaky integrating -firing model or via simulations [36]. The synaptic connection strength W_{ij} becomes excitatory for $V^{\text{syn}}_{ij} \geq V^{\text{eq}}$ and inhibitory otherwise. It is usually not the reversal potential V^{syn}_{ij} but the maximum conductance g^{syn}_{ij} (≥ 0) which is biologically plastic. Accordingly, the sign of a synaptic connection strength does not alter in the learning process.

While the variable $\phi_i(t)$ can take different values in different trials owing to stochasticity, the expectation value $\langle\phi_i(t)\rangle$ tends to increase with the membrane potential $\mathcal{V}_i(t)$. Considering that a neuron cannot fire again during the refractory period τ_{r}, we take the firing probability to be $\langle\phi_i(t)\rangle \propto \exp[T^{-1}\{\mathcal{V}_i(t) - \mathcal{V}^{\text{th}} + \int dt'\lambda^{\text{rep}}(t-t')\phi_i(t')\}]$ or

$$\langle\phi_i(t)\rangle \propto \exp\left[\frac{1}{T}\left\{-\sum_j \int dt'\, D^+_{ij}(t-t')\phi_j(t') + B_i(t)\right\}\right], \tag{46}$$

where $D^+_{ij}(t) = -\delta_{ij}\lambda^{\text{rep}}(t) + \mathcal{D}^+_{ij}(t)$ and $B_i(t) = h_i(t) - \mathcal{V}^{\text{th}}$ with $\mathcal{V}^{\text{th}} \equiv V^{\text{th}} - V^{\text{eq}}$ being the threshold membrane potential from the rest potential. Here, $-\lambda^{\text{rep}}(t)$ takes into account the presence of the refractory period, either diverging ($\to \infty$) for $0 < t \leq \tau_{\text{r}}$ or vanishing ($\to 0$) otherwise. Note that $D^+_{ij}(t)$, being applicable only for $t > 0$, represents the neural interactions along the ordinary time direction. The constant T, playing the role of temperature, measures the intensity of the noisy current [38,39].

Equation (46) makes it possible to calculate easily the firing probability of a neuron in a short time interval for given conditions. As an example, suppose that neuron i receives synaptic currents from other neurons in the absence of external stimulus currents. In the limit $T \to 0$, the firing condition of the neuron may be expressed as

$$\max\left[\sum_j \lambda^{\text{syn}}(t-t^*_j)W_{ij}\right] \geq \mathcal{V}^{\text{th}}, \tag{47}$$

where t^*_j represents the recent firing timing of neuron j [40]. The contour line in Figure 5e, obtained from $W_{ij} = g^{\text{syn}}_{ij}(V^{\text{syn}}_{ij} - V^{\text{th}})$ being constant in Equation (47), thus illustrates the condition of synaptic connection for a postsynaptic neuron to fire in response to the firing of a single presynaptic neuron. Note that biological neurons may not vary the activity in a short time because neural firing spikes have nearly the same form. Instead, they can perform a function or a communication through the use of the sensitivity for spike timings on the time scale of milliseconds [14,15]. Equation (47) describes how a biological neuron can have the sensitivity for exact timings of input spikes.

However, Equation (46) or Equation (47) does not allow one to predict the firing probability on a long-time scale during which neurons may interact recursively with one another. Such statistics of neural firings in a long time can only be obtained after the effects of all possible neural interaction ways are taken into account. The Feynman machine solves this problem by describing neural interactions in the form of a Feynman path integral. Specifically, the one-point function of neural firing states is given by

$$\langle \phi_i(t) \rangle = \frac{1}{Z} \mathrm{Tr}\, \phi_i(t)\, e^{A[\phi]/T} \tag{48}$$

with the partition function $Z \equiv \mathrm{Tr}\, e^{A[\phi]/T}$ as the normalization factor. Here, the trace stands for the summation over all configurations of firing states at the grid points in the spatiotemporal space. The action is given by

$$A[\phi] = -\frac{1}{2} \sum_{i,j} \int dt dt' D_{ij}(t-t')\phi_i(t)\phi_j(t') + \sum_i \int dt B_i(t)\phi_i(t) \tag{49}$$

with $D_{ij}(t) \equiv D^+_{ij}(t) + D^+_{ji}(t)$. Precisely speaking, the one-point function in Equation (48) contains the effects of interactions with firings at times later than the observation time, so that the firing probability in Equation (46) can be obtained after non-physical interactions are eliminated or the second quantization technique is adopted. The one-point function can be rewritten in the form

$$\langle \phi_i(t) \rangle = -\frac{\partial F}{\partial B_i(t)} \tag{50}$$

with the free energy $F = -T \log Z$. The connected two-point function $G_{ij}(t,t') \equiv \langle \phi_i(t)\phi_j(t') \rangle_c \equiv \langle \phi_i(t)\phi_j(t') \rangle - \langle \phi_i(t) \rangle \langle \phi_j(t') \rangle$ also obtains the form

$$G_{ij}(t,t') = -\frac{T\partial^2 F}{\partial B_i(t) \partial B_j(t')}. \tag{51}$$

The Feynman machine makes it possible to predict the firing probability or the cross-correlation on the millisecond time scale, with help of the methods in, e.g., liquid theory and statistical or quantum field theory [35]. For instance, after non-physical interactions are eliminated, the firing probability of neuron *i* at time *t* is expressed in the form

$$\begin{aligned} \langle \phi_i(t) \rangle = z_i(t) &+ \sum_j \int dt' z_i(t) z_j(t') f^+_{ij}(t-t') + \frac{1}{2} \sum_{j,k} \int dt' dt'' z_i(t) z_j(t') z_k(t'') \\ &\times \left[2 f^+_{ij}(t-t') f_{jk}(t'-t'') + f^+_{ij}(t-t') f^+_{ik}(t-t'') + f^+_{ij}(t-t') f^+_{ik}(t-t'') f_{jk}(t'-t'') \right] + \ldots, \end{aligned} \tag{52}$$

where $z_i(t) \equiv e^{B_i(t)/T}$ and $f_{ij}(t) \equiv f^+_{ij}(t) + f^+_{ji}(-t)$ with $f^+_{ij}(t) = e^{-D^+_{ij}(t)/T} - 1$ correspond, respectively, to the fugacity and the Mayer function in statistical physics. Some typical results predicted via the Feynman machine are presented in Figure 5, which displays results from simulations as well.

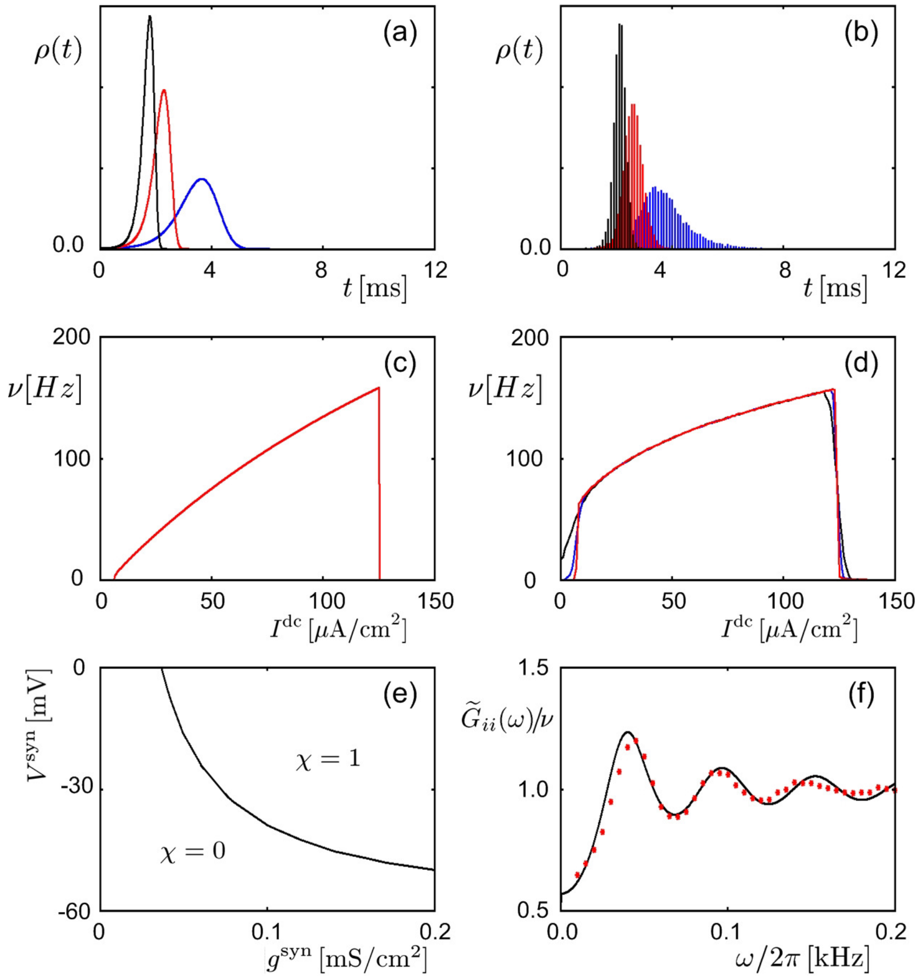

Figure 5. Neural firing statistics in the Hodgkin–Huxley model, predicted via the Feynman machine and obtained via simulations. (**a**,**b**) Firing probability $\rho(t)$ ($\equiv \langle \phi(t) \rangle$) of a neuron versus time t, activated by a presynaptic neuron fire at $t = 0$ in the presence of external noisy currents. Results from (**a**) theory and (**b**) simulations for three different values of the connection strength are presented in black, red, and blue [35]. (**c**,**d**) Mean firing rate $\nu \equiv \Delta^{-1} \int_0^{\Delta} dt\, \rho(t)$, with a long time interval Δ, of a neuron versus external direct current I^{dc} given additionally to activate the neuron. Shown are results from (**c**) theory and (**d**) simulations, for three different values of the noisy current intensity plotted in black, red, and blue lines in (**b**) [36]. (**e**) Plot of the contour line $\chi = 0.5$, obtained via simulations, on the parameter plane $(g^{\mathrm{syn}}, V^{\mathrm{syn}})$, where $\chi \equiv \int_0^{\tau_{\mathrm{r}}} dt\, \rho(t)$, with τ_{r} being the refractory period, is the total firing probability of a neuron activated by the synaptic current from a presynaptic neuron via a synapse of $(g^{\mathrm{syn}}, V^{\mathrm{syn}})$. Theoretically, the contour line consists of the points on which $g^{\mathrm{syn}}(V^{\mathrm{syn}} - V^{\mathrm{eq}})$ is constant [40]. (**f**) Frequency dependence of $\widetilde{G}_{ii}(\omega)/\nu$, where $\widetilde{G}_{ii}(\omega)$ is the Fourier transform of the autocorrelation function $G_{ii}(t)$. The black line represents theoretical results, whereas the red dots plot data points obtained via simulations [35].

5. Learning Principle in the Feynman Machine

Although the Feynman machine can adopt any kind of synaptic plasticity rule in its learning process, the minimization of the path-integral free energy serves as an ideal learning rule in the Feynman machine. Namely, the desirable change in a synaptic coupling strength is governed by the gradient flow of the free energy:

$$\Delta W_{ij} \propto -\frac{\partial F}{\partial W_{ij}}. \tag{53}$$

The Feynman machine learning principle is convinced by the fact that the firing states and the connection states constitute the coupled dynamics of a neural system with different attributes. The partition function $Z = \mathrm{Tr} e^{A[\phi]/T}$ can be rewritten in the form

$$Z = \prod_{i,t} \int_{-\infty}^{\infty} d\phi_i(t) \delta\big(\phi_i(t)(\phi_i(t)-1)\big) e^{A[\phi]/T} \equiv \int \mathcal{D}\phi \, \mathcal{W}_a[\phi] \, e^{A[\phi]/T}. \tag{54}$$

Accordingly, in consideration of varying external inputs, the partition function of a Feynman machine on a long time scale takes the form

$$Z_L = \int \mathcal{D}\phi \, \mathcal{D}W \, \mathcal{D}h \, \mathcal{W}_a[\phi] \, \mathcal{W}_s[W] \, \mathcal{W}_e[h] \, e^{A[\phi,W]/T}, \tag{55}$$

where $\mathcal{W}_a$, $\mathcal{W}_s$, and $\mathcal{W}_e$ are the weights or probability functions for neural activity, synaptic coupling strength, and external inputs, respectively. The connection state at the end of the learning process then obtains the form

$$\langle W_{ij} \rangle = \frac{1}{Z_L} \int \mathcal{D}\phi \mathcal{D}W \mathcal{D}h \mathcal{W}_a[\phi] \mathcal{W}_s[W] \mathcal{W}_e[h] W_{ij} e^{A[\phi,W]/T}. \tag{56}$$

With $F_L \equiv -\log Z_L$, the equilibrium condition $\partial F_L / \partial W_{ij} = 0$ brings on

$$\int \mathcal{D}h \mathcal{W}_e[h] \frac{\partial Z}{\partial W_{ij}} = 0, \tag{57}$$

which can be achieved by the gradient flow in Equation (53).

The Feynman machine needs an additional rule to restrict the synaptic coupling strengths within a range because the minimization of the free energy is achieved just by increasing the excitatory connection strengths and/or decreasing the inhibitory connection strengths as usual. One can define the weight function $\mathcal{W}_s[W]$ to control the synaptic coupling strength; however, the general form of $\mathcal{W}_s$, which works properly in extensive circumstances and bases on biological experiments, is yet to be suggested.

Remarkably, the typical form of the biological synaptic plasticity emerges from this ideal learning rule in the Feynman machine. With the STDP rule in Equation (5) substituted into Equation (53), the STDP window obtains the form

$$\Omega_{ij}(t) \propto -\frac{\partial^2 A}{\partial C_{ij}(t) \partial W_{ij}}. \tag{58}$$

A lengthy but straightforward calculation leads the STDP window to take the form

$$\Omega_{ij}(t) \propto \lambda^{\mathrm{syn}}(t) - \int dt' \Sigma_{ij}(t,t') \lambda^{\mathrm{syn}}(t'), \tag{59}$$

where $\Sigma_{ij}(t,t') \equiv T\partial^2 S / \partial C_{ij}(t) \partial D_{ij}(t')$ with the entropy $S \approx \frac{1}{2} \mathrm{tr} \log G$. The first term on the right-hand side of Equation (59), originating from the derivative of the internal energy $U = (1/2) \sum_{i,j} \int dt D_{ij}(t) C_{ij}(t)$, produces LTP for pre-before-post pairings. The second term, originating from the derivative of the entropy, produces LTD for post-before-pre parings and occasionally LTP for post-before-pre pairings. The resulting time dependence of $\lambda^{\mathrm{syn}}(t)$ and $\Omega(t)$ is shown in Figure 6. It is pleasing that these results indeed coincide with the form of the STDP window as well as the LTP part, observed in experiment.

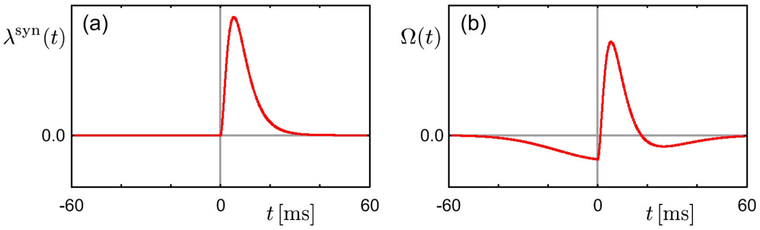

Figure 6. Plots of (**a**) $\lambda^{\text{syn}}(t)$ and (**b**) $\Omega(t)$, where $\lambda^{\text{syn}}(t)$ is taken from the solution of a leaky integrating–firing model, given by $C^{-1}\int_0^t dt' e^{\gamma(t'-t)}\alpha(t')$, and $\Sigma(t,t')$ is modeled simply as $A\exp[-(t+t')^2/2\sigma^2 - (t-t')^2/2\sigma^2]$ with $A = 0.02$ and $\sigma = 30$ ms.

Thus far, the learning process in a biological neural system has been predicted through estimating the cross-correlations of neural firings on given circumstances and producing it with a STDP window which is modeled rather coarsely from experimental observations [32,41,42]. On the other hand, the Feynman machine learning rule makes it possible to derive the cross-correlations, the STDP window, and their products from the derivatives of a free energy. For instance, the Feynman machine learning rule applied to the input–output layer system yields the development of a feedforward connection in the form [37]

$$\Delta W_{ij} \propto \int d\omega [\widetilde{K}(\omega) W \widetilde{Q}(\omega)]_{ij} - \beta[(WW^{\dagger})^{-1}W]_{ij}, \tag{60}$$

where $\widetilde{K}_{ij}(\omega)$ and $\widetilde{Q}_{ij}(\omega)$ are the Fourier transforms of the vertex two-point function $K_{ij}(t)$ between output neurons and the input correlation function $Q_{ij}(t-t') \equiv T^{-2}\int \mathcal{D}h\, \mathcal{W}_e[h] h_i(t) h_j(t') = T^{-2}\langle\langle h_i(t) h_j(t')\rangle\rangle$ in the spatiotemporal space.

This learning rule is capable of recognizing temporal patterns, such as a STDP-based learning model. It is not a firing-rate-based learning model but a spiking-timing-based one that can explain the development of selectivity for moving visual images, such as the directional selectivity in V1 [32]. If external inputs have weak correlations on the fast time scale so that $\widetilde{Q}(\omega)$ has a large peak at $\omega = 0$, Equation (60) reduces to a firing-rate-based learning model expressed in the form

$$\Delta W_{ij} \propto [\widetilde{K}(0) W \widetilde{Q}(0)]_{ij} - \beta[(WW^{\dagger})^{-1}W]_{ij}, \tag{61}$$

which corresponds just to Equation (41) in the PSL model.

6. Discussion

It is crucial in theoretical neuroscience to develop an abstract neuron model without losing the essential features of the biological neurons for information processing or learning. It is informative that Boolean algebra, rather than the real behavior of electronic devices, plays a key role in understanding how a computer operates. In a similar sense, the model neurons could be more significant than real neurons in understanding how the brain works. A neuron model explaining the firing activity and the learning rule through the concepts and principles in statistical mechanics could be helpful for revealing the essence of neural computing and learning mechanism.

The modeling based on statistical mechanics gives several advantages over other modeling methods. The firing activity and the learning rule are described by differential equations, the outcomes of which are inferred from the extremum states of appropriate functions. For example, the Hopfield network shows how to store desired patterns as the extremum states in a neural system. The probabilistic firing rule, based on important sampling, is regarded as a suitable way to avoid trapping in a local minimum during the learning process.

Nevertheless, it is an intriguing problem to probe the neural system via (equilibrium) statistical mechanics. The firing process in a biological neural network is usually governed by (non-integrable) nonlinear dynamical equations, and the learning process is usually taken under non-stationary external stimuli. A standard way to express the neural process in terms of statistical mechanism is to define the energy of the system. In view of statistical mechanics, this approach is based on some assumptions about the neural system, such as the ergodic hypothesis, indifference principle, entropy maximization principle, and so on. In particular, the assumption of ergodicity allows that a neural system is described by an energy function with no explicit time variable. Such an energy function further satisfies the invariance condition under the exchange of element positions, so that interaction strength between two elements should be symmetric. Therefore, with the connection strength between neurons regarded as the interaction strength between them, the energy-based neuron model assumes all connections to be symmetric and treat asymmetric connections as part of external inputs. Stochastic equations used to describe various phenomena subject to thermal fluctuations offer a possible way to describe the process in a neural system with asymmetric connections; the Langevin equation description of the firing-rate neuron model provides an example. The adoption of an explicit time variable could be another way to handle the problem. Namely, in the Feynman path integral approach, neural processes are described on the basis of an action, which is expressed as an integral in the spatiotemporal space. The action could satisfy the invariance condition if the interaction strength between neural pairs are given by a symmetric function in the spatiotemporal space or a Hermitian function in the momentum space.

An important aspect in a neural network model based on statistical mechanics is that the entropy maximization principle may be the origin of the competitive learning mechanism, which is indispensable for the learning process to prevent neurons from having the same features or functions with neighbors. In most neuron models, the implementation of the competition mechanism is based on the existence of inhibitory connections [27] or a normalization condition for neural activity [43], among others. Although not all statistical-mechanics-based models implement the competition mechanism through the maximization, the gradient descent in entropy does induce proper competitive learning behavior in a neural network. Particularly in a biological neural network, the competitive learning mechanism turns out to be based on the entropy maximization principle.

The statistics of neural firings could be obtained from the derivatives of the free energy with respect to external inputs if the free energy of a neural system is well defined. The learning rule in a statistical-mechanics-based model is often defined as the gradient descent in several quantities. However, the free energy is the most natural quantity to define an ideal learning process. The process in a neural system is basically given by the change in firing and connection states occurring on different dynamics scales. Namely, the neural process is described by dynamics coupled in variables with two different attributes, so that the extremum states of the free energy should relate to the outcomes of not only the firing but also the learning process.

A statistical-mechanics-based model also opens the possibility of understanding various phenomena in a neural system in terms of the general theory of phase transitions. The change in the features of rhythmic firings may be interpreted as a kind of phase transition phenomena. Further, the frame of connection structure could be altered in the learning process, depending on relevant conditions and temperature. In many learning models, the architecture is regarded as a fixed one and the change of connections is allowed without modifying the basic frame. Such models often suffer from the failure in the learning process when the architecture is not proper to handle given learning data. On the other hand, in a biological neural system, the architecture should not be fixed but modifiable flexibly; otherwise, a neural system, such as the brain, cannot operate steadily under various environmental conditions. It has been shown that a STDP-based learning model can modify the number and sizes of layers in a multilayer system, where the competition between connections causes extinction of less important connections and reorganization of the connection structure [8,44].

The Feynman machine allows one to interpret many aspects of the neural firing and learning processes in terms of general concepts and principles in statistical mechanics. While the Feynman machine has descended from conventional statistical-mechanics-based neuron models, it adheres to several attributes of neural processes. Among those, the explicit time variable, which has been neglected in conventional neural network models, plays a key role since the computing or learning mechanism in a biological neural network is often performed depending on exact firing timings of neurons. Unlike the original version which ignores such attributes of real neural processes as the effects of the variance in the membrane potential during neural firing, the Feynman machine describes neural dynamics in terms of the firing timing instead of the membrane potential, so that it can represent more realistic neural dynamics, as the relation between the membrane potential and firing timing in Equation (44) becomes more completely characterized. In particular, it has been demonstrated that an extended version of the Feynman machine with an additional interaction term is needed for figuring out neural firing dynamics in a synchronized state because the exact form of the membrane potential during neural firing may exerts effects on the phenomena [36].

Finally, we remark that the Feynman machine can be applied to explaining the emergence of computing ability not only in a neural system but also in a physical system. If the dynamics of the physical system is expressed in the form of a path integral involving several types of dynamical variables with the rate of change on different time scales and the changes of the variables optimize the path integral, the system can perhaps acquire the ability of computing, memorizing, or learning.

Author Contributions: M.W.C. conceived the research, developed the model, and wrote the paper. M.Y.C. supervised the analysis and wrote the paper.

Funding: This work was supported by the National Research Foundation of Korea through the Basic Science Program [Grant numbers: 2017R1A2B4010885 (MWC) and 2016R1D1A1A09917318 (MYC)].

Conflicts of Interest: The authors declare no conflict of interest.

References

1. McCulloch, W.; Pitts, W. A logical calculus of the ideas immanent in nervous activity. *Bull. Math. Biophys.* **1943**, *5*, 115–133. [CrossRef]
2. Hodgkin, A.L.; Huxley, A.F. A quantitative description of membrane current and its application to conduction and excitation in nerve. *J. Physiol.* **1952**, *117*, 500–544. [CrossRef] [PubMed]
3. Hebb, D.O. *The Organization of Behavior: A Neurophysiological Theory*; Wiley: New York, NY, USA, 1949.
4. Markram, H.; Lübke, J.; Frotscher, M.; Sakmann, B. Regulation of synaptic efficacy by coincidence of postsynaptic APs and EPSPs. *Science* **1997**, *275*, 213–215. [CrossRef] [PubMed]
5. Bi, G.-Q.; Poo, M.-M. Synaptic modifications in cultured Hippocampal neurons: Dependence on spike timing, synaptic strength, and postsynaptic cell type. *J. Neurosci.* **1998**, *18*, 10464–10472. [CrossRef] [PubMed]
6. Abbott, L.F.; Nelson, S.B. Synaptic plasticity: Taming the beast. *Nat. Neurosci.* **2000**, *3*, 1178–1183. [CrossRef]
7. Song, S.; Miller, K.D.; Abbott, L.F. Competitive Hebbian learning through spike-timing-dependent synaptic plasticity. *Nat. Neurosci.* **2000**, *3*, 919–926. [CrossRef] [PubMed]
8. Song, S.; Abbott, L.F. Cortical remapping through spike timing-dependent plasticity. *Neuron* **2001**, *32*, 1–20. [CrossRef]
9. Cho, M.W.; Choi, M.Y. Emergent topologies in activity-dependent self-organizing networks. *Europhys. Lett.* **2011**, *95*, 58005. [CrossRef]
10. Little, W.A. The existence of persistent states in the brain. *Math. Biosci.* **1974**, *19*, 101–120. [CrossRef]
11. Little, W.A.; Shaw, G.L. Analytical study of the memory storage capacity of a neural network. *Math. Biosci.* **1978**, *39*, 281–290. [CrossRef]
12. Hopfield, J.J. Neural networks and physical systems with emergent collective computational abilities. *Proc. Natl. Acad. Sci. USA* **1982**, *79*, 2554–2558. [CrossRef] [PubMed]
13. Friston, K. The free-energy principle: A rough guide to the brain? *Trends Cogn. Sci.* **2009**, *13*, 293–301. [CrossRef] [PubMed]
14. Gerstner, W.; Kempter, R.; van Hemmen, J.L.; Wagner, H. A neuronal learning rule for sub-millisecond temporal coding. *Nature* **1996**, *383*, 76–81. [CrossRef] [PubMed]
15. Cho, M.W.; Choi, M.Y. A model for the receptive field of retinal ganglion cells. *Neural Netw.* **2014**, *49*, 51–58. [CrossRef] [PubMed]
16. Choi, M.Y. Dynamic model of neural networks. *Phys. Rev. Lett.* **1988**, *61*, 2809–2812. [CrossRef] [PubMed]
17. Ackley, D.H.; Hinton G.; Sejnowski, T. A learning algorithm for Boltzmann machines. *Cogn. Sci.* **1985**, *9*, 147–169. [CrossRef]

18. Metropolis, N.; Rosenbluth, A.W.; Rosenbluth, M.N.; Teller, A.H. Equation of state calculations by fast computing machines. *J. Chem. Phys.* **1953**, *21*, 1087–1092, doi:10.1063/1.1699114 [CrossRef]
19. Palotai, Z.; Szirtes, G.; Lörincz, A. Emerging evolutionary features in noise driven STDP networks. In Proceedings of the 2004 IEEE International Joint Conference on Neural Networks (IEEE Cat. No.04CH37541), Budapest, Hungary, 25–29 July 2004; pp. 489–494.
20. Shin, C.-W.; Kim, S. Self-organized criticality and scale-free properties in emergent functional neural networks. *Phys. Rev. E* **2006**, *74*, 045101. [CrossRef] [PubMed]
21. Strogatz, S.H. Exploring complex networks. *Nature* **2001**, *410*, 268–276. [CrossRef] [PubMed]
22. Boccaletti, S.; Latora, V.; Moreno, Y.; Chavez, M.; Hwang, D.-U. Complex networks: Structure and dynamics. *Phys. Rep.* **2006**, *424*, 175–308. [CrossRef]
23. Watts, D.J.; Strogatz, S.H. Collective dynamics of 'small-world' networks. *Nature* **1998**, *393*, 440–442. [CrossRef] [PubMed]
24. Barabási, A.-L.; Albert, R. Emergence of scaling in random networks. *Science* **1999**, *509*, 509–512.
25. Bell, A.J.; Sejnowski, T.J. An information-maximization approach to blind separation and blind deconvolution. *Neural Comput.* **1995**, *7*, 1004–1034. [CrossRef]
26. Lee, T.-W. *Independent Component Analysis*; Springer: Boston, MA, USA, 1998; pp. 27–66.
27. Dayan, P.; Abbott, L.F. *Theoretical Neuroscience*; MIT Press: London, UK, 2001.
28. Cho, M.W.; Kim, S. Understanding visual map formation through vortex dynamics of spin Hamiltonian models. *Phys. Rev. Lett.* **2004**, *92*, 018101. [CrossRef] [PubMed]
29. Tanaka, S. Theory of self-organization of cortical maps: Mathematical framework. *Neural Netw.* **1990**, *3*, 615–640. [CrossRef]
30. Tanaka, S.; Shinbata, H. Mathematical model for self-organization of direction columns in the primate middle temporal area. *Biol. Cybern.* **1994**, *70*, 227–234. [CrossRef] [PubMed]
31. Cho, M.W.; Kim, S. Different ocular dominance map formation influenced by orientation preference columns in visual cortices. *Phys. Rev. Lett.* **2005**, *94*, 68701. [CrossRef] [PubMed]
32. Cho, M.W.; Choi, M.Y. Functional organization for direction preference in relation to orientation and ocular dominance maps. *J. Korean Phys. Soc.* **2009**, *55*, 2532–2536. [CrossRef]
33. Cho, M.W. Competitive learning behavior in a stochastic neural network. *J. Korean Phys. Soc.* **2015**, *67*, 1679–1685. [CrossRef]
34. Cho, M.W.; Choi, M.Y. Origin of the spike-timing-dependent plasticity rule. *Europhy. Lett.* **2016**, *115*, 38001. [CrossRef]
35. Cho, M.W. Calculation of precise firing statistics in a neural network model. *J. Korean Phys. Soc.* **2017**, *71*, 222–230. [CrossRef]
36. Cho, M.W. Study of collective synchronous dynamics in a neural network model. *J. Korean Phys. Soc.* **2018**, in press.
37. Cho, M.W. Learning process in a neural network model. *J. Korean Phys. Soc.* **2018**, in press.
38. Uhlenbeck, G.E.; Ornstein, L.S. On the theory of Brownian motion. *Phys. Rev.* **1930**, *36*, 823–841, doi:10.1103/PhysRev.36.823 [CrossRef]
39. Risken, H. *Fokker-Plank Equation: Method of Solution and Applications*; Springer: New York, NY, USA, 1989.
40. Cho, M.W. Study of neural response condition in a spiking neuron model. *New Phys. Sae Mulli.* **2016**, *66*, 786–792. [CrossRef]
41. Cho, M.W. Temporal Hebbian plasticity designed for efficient competitive learning. *J. Korean Phys. Soc.* **2014**, *64*, 1213–1219. [CrossRef]
42. Cho, M.W. Two symmetry-breaking mechanisms for the development of orientation selectivity in a neural system. *J. Korean Phys. Soc.* **2015**, *67*, 1661–1666. [CrossRef]
43. Piepenbrock, C. Natural Image Statistics for Cortical Orientation Map Development. In *Probabilistic Models of the Brain*; Rao, R.P.N., Olshausen, B.A., Lewicki, M.S., Eds.; MIT Press: Cambridge, UK, 2002; pp. 181–201.
44. Cho, M.W.; Choi, M.Y. Modulation of the boundary between hierarchically differentiated domains in a self-organizing neural system. *Europhy. Lett.* **2013**, *101*, 48004. [CrossRef]