

Received October 28, 2015, accepted December 13, 2015, date of publication December 17, 2015,
date of current version February 29, 2016.

Digital Object Identifier 10.1109/ACCESS.2015.2509005

A Review on Synergistic Learning

CHUNGUANG LI, (Senior Member, IEEE), AND YUKE LI

College of Information Science and Electronic Engineering, Zhejiang University, Hangzhou 310027, China

Corresponding author: C. Li (cgli@zju.edu.cn)

This work was supported in part by the National Natural Science Foundation of China under Grant 61171153 and Grant 61571392 and in part by the National Program for Special Support of Eminent Professionals.

ABSTRACT In neuroscience, it is widely believed that learning and memory are primarily based on synaptic plasticity which is a neural mechanism that modifies the strength of connections between neurons. As a counterpart in machine learning, the modification of connection strength (weight) endows artificial neural networks with a powerful learning capability to solve various problems. Independent of modification for synaptic strength, recent experimental results have revealed that a single neuron also has the ability to change its intrinsic excitability to fit the synaptic input. This mechanism is referred to as neuronal intrinsic plasticity (IP) in the literature. Computational learning rules for IP have been developed based on the hypothesis of information maximization with a stable response level. With the discovery of this novel plasticity mechanism, a series of studies has focused on how IP plays a role in biological neural systems and how they benefit the learning performance of artificial neural networks. In this review, corresponding research on synergies between IP and synaptic plasticity mechanisms is presented in both the computational modeling of biological neural systems and the applications of artificial neural networks, and this combination in artificial learning systems is defined as synergistic learning.

INDEX TERMS Intrinsic plasticity, synaptic plasticity, synergistic learning, homeostasis, sparse coding, artificial neural networks, machine learning.

I. INTRODUCTION

It is highly possible that brains are the most intelligent information processing systems in the world. For biological neural systems, it is a consensus that memory is stored in the strength of synapses between neurons and learning is a process to change the synaptic strength. Synaptic strength modification between neurons may depend on spike timing, firing rates or membrane voltage levels. In neuroscience, this modification is defined as *synaptic plasticity*. A sheer volume of work on this field has discovered a series of neurophysiological results and potential underlying processes related to synaptic modification. Computational modeling on synaptic plasticity has partially shed light on how synaptic modification gives rise to learning effects discovered in neural systems. From a bottom-up perspective, artificial neural networks (ANNs) are trying to capture the biological structures and underlying learning mechanisms of the real neural systems. As powerful bio-inspired machine learning techniques, ANNs are very competitive in solving a large set of learning problems such as regression, prediction, classification and clustering. In terms of the network construction, neurophysiological findings of simple cells and complex cells in the visual cortex provide a solid foundation for the hierarchical structures

of ANNs [1]. As for learning strategies of ANNs in machine learning, most learning algorithms depend on the connection strength (weight) modification between learning units and these algorithms are comparable to the biological synaptic plasticity mechanisms. Although synaptic plasticity rules in neuroscience and learning algorithms for ANNs in machine learning are usually developed in parallel, synaptic plasticity provides a solid biological background for learning strategies in applications.

Recent experimental studies have found that synaptic plasticity is not the unique plasticity mechanism in the neural systems during learning and development. Independent of synaptic strength modification, a single neuron also has the ability to change its intrinsic excitability to match its synaptic input [2]–[11]. This mechanism is referred to as *neuronal intrinsic plasticity (IP)*. With this IP mechanism, a single neuron can strengthen the excitability when its input is deprived for a period of time and weaken the excitability when the input is boosted. This is a strictly local process for a single neuron. Based on experimental results, a number of intrinsic plasticity rules have been developed to capture the learning properties of this novel mechanism. In computational modeling, it is hypothesized that IP is

trying to maximize the information capacity while keeping the mean firing activity of an individual neuron in a homeostatic level [12]–[14]. Maximizing the information capacity means maximizing the input-output mutual information and making the response tuning curve of the neuron match its synaptic input distribution. In other words, IP can provide an efficient representation of synaptic inputs with a constant metabolic cost.

Now that the newly discovered IP mechanism exists in biological neural systems with synaptic plasticity, it is an intriguing question how modification rules for intrinsic excitability of learning units will cooperate with weight modification in artificial learning systems. Recently some related work has provided some answers to this question but there is still relatively few studies contributing to the synergies between different plasticity mechanisms in computational modeling and machine learning. In this review, we refer to the combination of synaptic and intrinsic plasticity learning rules in artificial learning systems as “synergistic learning”. We are reviewing these studies together to get a better understanding of this novel topic and trying to link biological plasticity mechanisms in neuroscience to learning strategies in computational modeling and machine learning. In the remaining part of this review, we first successively describe experimental results and computational work on synaptic plasticity and intrinsic plasticity. We then present neural models combining synaptic plasticity with intrinsic plasticity for different computational tasks and machine learning algorithms integrating weight modification with intrinsic plasticity rules for training ANNs. Due to intrinsic plasticity is a long-term learning process that begins to happen from hours to days, we only focus on long-term synaptic plasticity as well and we do not introduce short-term synaptic plasticity (adaptation) here, which happens in seconds or minutes and is also an important part of plasticity.

II. SYNAPTIC PLASTICITY

A. SYNAPTIC PLASTICITY RULES IN NEUROSCIENCE

For biological neural systems, learning and memory are highly related to the synaptic strength modification between neurons. As the connectivity of the neural system is changing over time, brain functions and neural dynamics are gradually shaped. A large body of classical experimental results have unveiled synaptic plasticity mechanisms such as long-term potential and depression (LTP and LTD) [15]–[20]. Recent neurophysiological experiments continuously add evidence to synaptic plasticity. For example, in an experiment of mouse visual cortex [21], Ko et al found that in early life of mice the connection probability between neurons distinctly responsive to the natural movie is not significantly different to the connection probability between non-responsive neurons, which to some extent means the initial connectivity between neurons is relatively random. However, when the mice become mature, there is a significantly high connection probability for visually responsive neurons in comparison to the

connection probability of non-responsive neurons. This is a convincing result showing the synaptic modification during the primary visual cortex (V1) development.

A series of synaptic plasticity theories were proposed before relevant experiments were conducted to provide biological evidence. In general, the weight updating rules for synaptic plasticity are described as below,

$$w_{ji}(n+1) = w_{ji}(n) + \Delta w_{ji}(n), \quad (1)$$

where $w_{ji}(n)$ denotes the weight of a synapse linked from a presynaptic neuron i to a postsynaptic neuron j at time point n , and the modification term $\Delta w_{ji}(n)$ is the weight change from time point n to time point $n+1$. Hebbian learning is the most classical learning theory for synaptic plasticity in the neural systems. Hebbian learning theory claims that neurons fire together and wire together,

$$\Delta w_{ji} = \eta x_i y_j, \quad (2)$$

where x_i and y_j are the responses (firing rate) of a presynaptic neuron i and a postsynaptic neuron j , the modification Δw_{ji} for the synapse strength between these two neurons is determined by their positive response correlation, and η is a small positive learning rate. However, Hebbian learning rule is in essence unstable and the synaptic strength will be growing in an uncontrolled manner. Oja's rule is a famous updated version of Hebbian learning rule to keep the learning process bounded. This rule is derived by normalizing all the weights connected to a postsynaptic neuron j to have a unit length, $|\mathbf{w}_j| = (\sum_s w_{js}^2)^{\frac{1}{2}} = 1$. The normalization is shown as below,

$$w_{ji}(n+1) = \frac{w_{ji}(n) + \eta x_i(n) y_j(n)}{\left(\sum_s [w_{js}(n) + \eta x_s(n) y_j(n)]^2\right)^{\frac{1}{2}}}. \quad (3)$$

By expanding Eq. (3) into a Taylor series when η is far smaller than 1 and specifying that the neuron model is linear as $y_j = \sum_s w_{js} x_s$, we get an updated rule which has an additional weight decay term,

$$\Delta w_{ji} = \eta x_i y_j - w_{ji} y_j^2. \quad (4)$$

This rule keeps the Hebbian learning process bounded and evolving in a stable way. More importantly, an artificial neural network trained by Oja's rule can perform principal component analysis (PCA) for the input data [22]–[24]. Bienenstock-Cooper-Munro (BCM) theory is another famous Hebbian-like learning theory considering that the learning process is stabilized by a sliding threshold as a function of the average postsynaptic firing activity [25]–[28]. The original BCM rule for a linear neuron model is,

$$\begin{aligned} \Delta w_{ji} &= \eta(\phi(y_j)x_i - \epsilon w_{ji}), \\ \text{where } \phi(y_j) &= y_j(y_j - \theta_M), \\ \text{and } \theta_M &= E^2 \left[\frac{y_j}{y_0} \right]. \end{aligned} \quad (5)$$

$\phi(\cdot)$ is the postsynaptic activation function for weight modification and θ_M is a sliding threshold which is determined by the average activity of y_j for the postsynaptic neuron j .

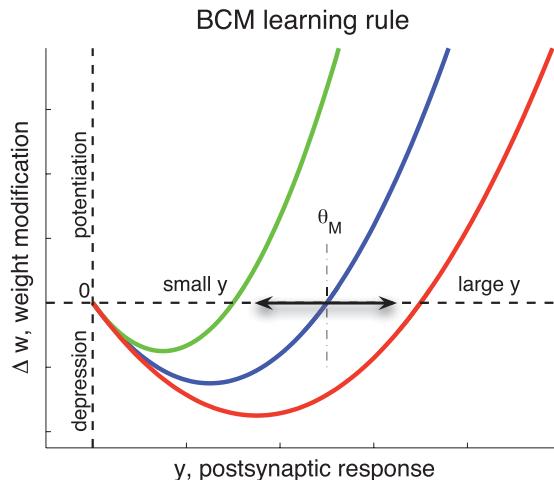


FIGURE 1. The BCM learning rule for synaptic plasticity. A sliding threshold θ_M depends on the average activity of the postsynaptic neuron. The green line indicates a learning curve with a relatively high average postsynaptic activity and the threshold is relatively low. The red line indicates a curve with a relatively low average activity and the threshold is relatively high. The synapse is undergoing Long-term potential (LTP) when the postsynaptic activity is higher than this threshold, and long-term depression (LTD) when it is lower.

$E^2[\cdot]$ denotes the mean squared average over time. ϵ is a small constant for weight decay and y_0 is a positive scaling constant. Fig. 1 shows three activation functions of the BCM theory for weight modification. Long-term potential (LTP) takes place when the postsynaptic response is higher than this threshold θ_M , and long-term depression (LTD) happens when it is lower. The BCM theory is gradually verified by experimental results from the biological basis [29]–[33].

These classical synaptic plasticity rules are rate-based rules and they are originally on the basis of theoretical hypotheses from a computational perspective rather than observation from biological experiments.

In recent decades, experimental results have found a novel synaptic plasticity mechanism relying on the relative firing timing between the presynaptic and postsynaptic neurons, which is referred to as spike-timing-dependent plasticity (STDP) [34]–[37]. During STDP, LTP happens when the presynaptic action potential precedes the postsynaptic action potential, while LTD happens for the reverse time order. The quantity of the synaptic modification is according to a time learning window (kernel). The basic STDP model is described as below [38]–[41],

$$\Delta w_{ji} = \sum_m \sum_n W(t_j^{(m)} - t_i^{(n)}),$$

where $W(\Delta t) = \begin{cases} A_+ \exp(-\Delta t/\tau_+) & \text{for } \Delta t > 0 \\ -A_- \exp(\Delta t/\tau_-) & \text{for } \Delta t < 0. \end{cases}$ (6)

t_j and t_i are the spiking times for the postsynaptic neuron j and the presynaptic neuron i , respectively. A_+ and A_- are modification amplitudes. τ_+ and τ_- are time constants for the learning window. W is the STDP time learning window and Fig. 2 shows a basic time learning window as an example.

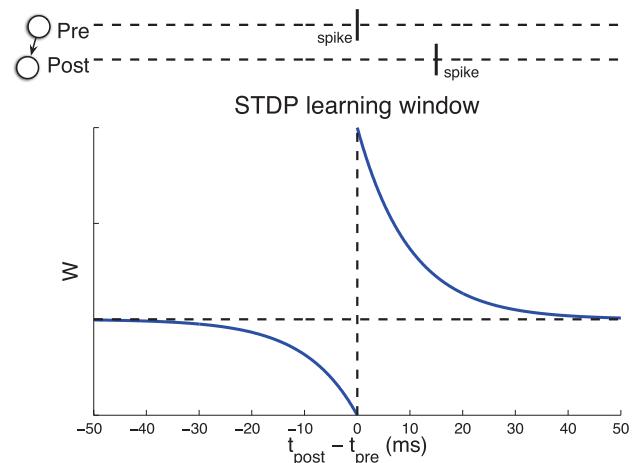


FIGURE 2. A basic STDP learning window (kernel) for synaptic plasticity is shown as an example. LTP happens when the presynaptic action potential precedes the postsynaptic action potential as shown at the top of the figure, while LTD happens for the reverse time order.

In this STDP model, the weight modification Δw is accumulated by all the spike pairs of presynaptic and postsynaptic neurons as an all-to-all learning method. The weight modification can also be calculated by the nearest neighbor spike pairs of presynaptic and postsynaptic neurons as a nearest neighbor learning method. STDP can be connected to Hebbian learning and BCM rules, thereby showing generality of synaptic plasticity mechanisms [42], [43]. In addition, STDP can also keep a stable correlation-based learning process [44], [45]. STDP has become an inspiring finding in neuroscience and a large amount of modeling work has shown how STDP is shaping brain functions and neural dynamics [42], [46].

TABLE 1. A brief summary of five important synaptic learning rules.

Rule Name	Comments
Hebbian	learning by response correlation between two neurons, unbounded, rate-based learning
Oja	an updated Hebbian learning rule with weight decay, bounded, rate-based learning
BCM	adjusting a sliding threshold for Hebbian-like learning, bounded, rate-based learning
STDP	learning by relative spike firing timing of two neurons, bounded, spike-based/causality-based learning
CaDP	probably a unified and fundamental synaptic learning rule

Nevertheless, STDP may not be perfect as the “first law” of synaptic plasticity. STDP has very different time learning windows in different neural systems and it can consider either all spike pairs or only neighbor spike pairs for learning. Calcium-dependent plasticity (CaDP) rule is trying to behave as a more fundamental rule for synaptic plasticity. This rule is based on the calcium control and it can unify different synaptic plasticity mechanisms depending on presynaptic frequencies, postsynaptic firing activities, or spike timing [47], [48]. A brief summary of these above-mentioned synaptic learning rules is presented in Table. 1.

B. WEIGHT MODIFICATION ALGORITHMS IN MACHINE LEARNING

In machine learning, the artificial neural network is a powerful technique in applications. In this section we will introduce general ideas of weight modification algorithms for training artificial neural networks to solve various problems. The weight modification of artificial neural networks is comparable to synaptic plasticity in neuroscience.

These above-mentioned synaptic plasticity rules in neuroscience can be applied to solving unsupervised learning problems such as principle component analysis (PCA), independent component analysis (ICA) and sparse coding [22]–[24], [27], [49]–[52]. It is believed that real neural systems are undergoing these computational processes to get efficient representation/codes for the neural stimuli. For these unsupervised learning algorithms, the weight set \mathbf{w}_j connected to a neuron j is a code/representation for the input data space and the response of that neuron y_j indicates how much the corresponding code will represent the input \mathbf{x} . The input stimulus can be reconstructed by the weighted summation of these codes,

$$\mathbf{x} \approx \sum_s y_s \mathbf{w}_s. \quad (7)$$

For different constraints for the weights \mathbf{w} , different types of representative components can be extracted. PCA and ICA are two important data analysis methods in engineering to extract a number of representative components underlying the observation data set. PCA is trying to get components having the highest variance and orthogonal to each other, and ICA is finding components independent to each other [51], [53]. Sparse coding is a representation strategy that input data are encoded by the firing of a small subset of neurons from a large number of available neurons [54], [55]. The weight set of a neuron connected to the input data is a basis function (as a receptive field or a Gabor filter). Each input is represented by a linear combination of these basis functions of sparsely firing neurons. The sparse coding theory indicates that when a stimulus is shown to the primary visual cortex (V1), simple cells will sparsely fire to encode this stimulus. Sparse coding is equivalent to independent component analysis under certain conditions [55]. Neurophysiological studies are in good agreement with computational work for the sparse coding theory. Sparse coding is an important representation theory connecting the neuroscience community to the computer vision community.

Another well-known unsupervised learning method is self-organizing map (SOM, or Kohonen map) [56], which is also a representation strategy. It is a class of artificial neural networks that can extract 2-dimensional representation for the input space of training data. SOM can be treated as a clustering task (similar to K-means) or a special case of vector quantization (VQ). The weight updating algorithm for training SOM utilizes local competition and it is in a different form of the above-mentioned synaptic plasticity rules.

As for supervised learning algorithms for artificial neural networks in machine learning, most of them depend on updating weights between learning units to build an input-output mapping. These learning algorithms for weight modification are usually derived from optimizing objective functions of errors and they are also usually different to the above-mentioned synaptic plasticity rules in neuroscience. These algorithms have been widely used and proved to be advantageous in regression, prediction, classification and so on. A typical supervised learning strategy for regression/prediction is shown as below,

$$E = \frac{1}{2} \sum_s (d_s - y_s^o)^2, \\ \Delta w_{ji} = -\eta \frac{\partial E}{\partial w_{ji}}, \quad (8)$$

where E is the mean squared error (MSE) function for measuring the distance between the desired output, d , and the output of the learning system, y^o , for all training samples. The modification of the weight is taking the negative direction of the gradient of E relative to the weight w . For a multilayer artificial neural network, the most successful learning algorithm is back-propagation (BP) training [57]. In this algorithm, the gradient in the multilayer network is calculated by chain rules to take advantage of feedback error signals (training signals). The BP algorithm is still the most favorable training method for deep convolutional neural networks and for the fine tuning stage of training deep boltzmann machines [58]–[60]. Although no valid biological foundations for error back-propagation have been found, biological findings of the retroaxonal signal can be treated as a similar signal back-propagation mechanism [61]. Underlying biological mechanisms with feedback signals for supervised learning/reinforcement learning have recently been discovered from the olfactory system of the fly and the sensory system of the electric fish [62], [63]. Although learning algorithms for training artificial learning systems are not necessarily as same as these synaptic learning rules presented in the previous section, there is a strong connection between biological systems and artificial systems for learning.

III. INTRINSIC PLASTICITY

A. EXPERIMENTAL RESULTS ON INTRINSIC PLASTICITY

In biological neural systems, the synaptic input of a neuron is changing over time due to unexpected external neural drive and learning results of synaptic plasticity. In the absence of homeostatic mechanisms, the synaptic input of a neuron may strongly increase or decrease for a long time and as a result the neural activity will be drifting to a extremely high or low level. From a perspective of metabolic cost, a restricted activity range of a neuron is really meaningful. It is comparable to that human beings keep a stable body temperature. The biological neural system will be out of work if the energy consumption rise to an extraordinarily high level and the average firing activity should be relatively low to

save energy. On the other hand, as the neural response is within a finite range, the synaptic input should be represented by the finite range of firing rates. However, as the input distribution is changing, a neuron can not represent its input perfectly when the input distribution does not match its response tuning curve. Considering both the energy consumption and the coding efficiency, it is mysterious by what mechanisms a single neuron can effectively represent its non-stationary synaptic inputs while keeping a stable response level.

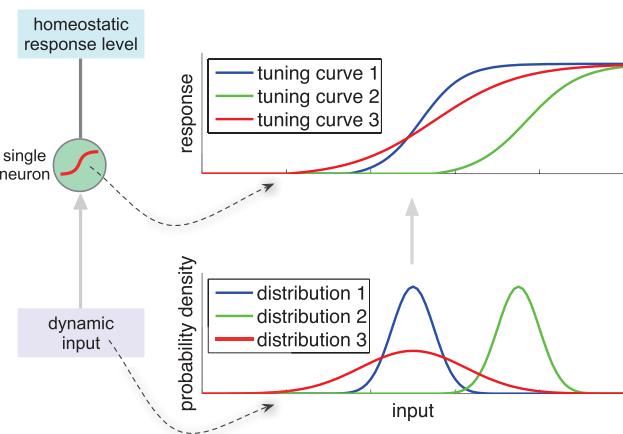


FIGURE 3. An interpretive graph of neuronal intrinsic plasticity. As the input distribution changes over time, the response tuning curve of a single neuron will be adapted to match the input for producing a stable output response level and preserving the ability of efficient coding. Three different response tuning curves in the top figure match three different input distributions in the bottom figure, respectively.

Fortunately, it has been found that single neurons also have the ability to change their intrinsic excitability over time to match different synaptic input levels. The intrinsic excitability of a neuron will increase when its synaptic input is deprived and decrease when the input is boosted significantly. The intrinsic excitability of a neuron can be described by its response tuning curve. This long-term learning ability is independent of the synapses and it is referred to as *neuronal intrinsic plasticity* (IP) [2]–[11]. Experimental results have shown that a single neuron can change the position and the slope of a response curve to match its sensitive region (linear part) to the center of the input distribution [9]. In this way the activity of the neuron can be bounded in a homeostatic range to encode the synaptic input more effectively. Therefore, the contradiction of energy saving and information preserving we mentioned above can be resolved by the IP mechanism. This experience-dependent modification of the the neuronal intrinsic excitability has been observed across brain areas of different animals and it is probably a ubiquitous phenomenon in biological neurons. An interpretive graph of neuronal intrinsic plasticity is shown in Fig. 3. As the input distribution of a single neuron changes in a biological neural system, the response tuning curve will be adapted to match different input distributions to produce a homeostatic response level and

preserve the ability of efficient coding as a result of neuronal intrinsic plasticity.

B. MODELING ON INTRINSIC PLASTICITY

It is still complicated to understand the underlying functions of the IP mechanism by experimental approaches. On the basis of experimental observations, computational modeling approaches can contribute to studying the learning properties of intrinsic plasticity. In modeling, it has been hypothesized that the IP mechanism is trying to make an individual neuron convey as much information as possible while keeping a homeostatic response level [13], [14]. Neurons in both primary visual cortex (V1) and the inferior temporal cortex (IT) show approximately exponential spike-count distributions in response to natural video sequences (spike-count rate is computed in various time windows) [64]. Since the exponential distribution has the highest entropy among all distributions of a non-negative random variable with a fixed mean [13], [65]–[68], it is convincing that visual cortical neurons can maximize their information capacity while keeping a fixed energy consumption with such exponential spike-count distributions as observed in experiments.

In order to have a better understanding of what role IP is playing in learning and memory, researchers have proposed several IP rules for computational modeling. In general, these rules bring the neural response distribution into a desired one with a relatively low activity level as observed in visual cortical neurons.

C. INTRINSIC PLASTICITY RULES FOR RATE-BASED NEURON MODELS

On the basis of information theory, Triesch used the Kullback-Leibler divergence (KL-divergence) to derive an IP rule for a rate-based (non-spiking) neuron model [13], [14], [69]. The classical rate-based neuron model is described by a sigmoid activation function as follows,

$$y = g(x) = \frac{1}{1 + \exp(-(ax + b))}, \quad (9)$$

where x denotes the total synaptic input and y denotes the output response in the form of the firing rate. g is the activation function to describe the response tuning curve. The parameter a is the gain of the activation function and the parameter b determines the threshold for generating firing activities. In other words, a is the slope and b is the bias of the activation function. The gain of the activation function controls the selectivity/discrimination ability, and the threshold controls the sensitivity of the neuron. With a large value of a and a small value of b , the neuron is more likely to fire. The IP learning rule based on this neuron model is obtained by minimizing the KL-divergence between the output distribution of the neuron and an exponential distribution,

$$\begin{aligned} \Delta a &= \eta \left[\frac{1}{a} + x - (2 + \frac{1}{\mu})xy + \frac{1}{\mu}xy^2 \right], \\ \Delta b &= \eta \left[1 - (2 + \frac{1}{\mu})y + \frac{1}{\mu}y^2 \right], \end{aligned} \quad (10)$$

where η is the learning rate and μ is the mean of the exponential distribution. As a learning result of this IP rule, the average firing activity of a neuron is kept in a low level. Based on this work, a related IP rule for a stochastically spiking neuron with refractoriness was presented [70]. The derivation of this IP rule is also related to minimizing the KL-divergence with respect to the parameters of the neuron model.

After the learning of Triesch's IP rule, the high density parts of the input distribution is matching the lower part of the sigmoid response curve. The lower part of a sigmoid curve is an insensitive region as its slope is less sharper than that of the linear part in the middle of the curve. Therefore, this learning result means that the adapted neuron greatly sacrifices its sensitivity and discriminative ability to reduce the output response level. However, experimental results show that a neuron can change its response curve to match the sensitive region of the curve to its input distribution [9]. Although Triesch's IP rule successfully keeps a neuron in a homeostatic state but it loses too much of the sensitivity and discriminative ability. Minimizing the firing rate and maximizing the sensitivity is contradictory to each other in this way. In [71], Li analyzed this problem in detail and provided a method to improve the IP rule. According to the experimental results in [64], the spike-count distributions are related to the window size used for estimation. Note that spike-count rate is an estimate for the probability of the instantaneous firing rate. Considering a spike train with a low average firing rate, a small window size can produce an exponential spike-count distribution described by a straight line in the decimal-logarithmic space, while a large window produces an exponential distribution with a droop-head (refer to [64, Fig. 3]). Li argued that if the firing rate of a neuron is low and the window size is small, there could be a divergence between the spike-count distribution and the firing-rate distribution, especially for the low-rate region. In this case, even if the *spike-count* distribution is exactly an exponential distribution, the *firing-rate* distribution may have a droop-head. By step-by-step reasoning, Li argued that the Weibull distribution is more appropriate to describe the firing-rate distributions in neurophysiology recordings. By using a Weibull distribution instead of an exponential distribution as the output target, the derived IP rule can make the sensitive region of the response curve properly match the high probability region of the input distribution while keeping the firing activity "low but not too low". This modified IP rule is shown as follows,

$$\begin{aligned}\Delta a &= \eta \left[\frac{1}{a} + \beta x(1-y) - xy - \frac{\beta}{\alpha^\beta} xy^\beta (1-y) \right], \\ \Delta b &= \eta \left[\beta(1-y) - y - \frac{\beta}{\alpha^\beta} y^\beta (1-y) \right],\end{aligned}\quad (11)$$

where α and β are two parameters of the probability density function (PDF) of a Weibull distribution. The Weibull distribution is exactly an exponential distribution when $\beta = 1$ and it looks like an exponential distribution with a droop-head

when β is slightly larger than 1. This rule is a generalized version of Triesch's IP rule and it resolves the contradiction of minimizing the firing rate and maximizing the sensitivity. In addition, if we take into account the unreliability of the synaptic transmission which may have extensive effects on neural systems [72]–[77], keeping a "low but not too low" average firing rate seems a more reasonable IP hypothesis in comparison to "as low as possible". Intuitively, the "not too low" firing rates may enhance reliability in biological neuronal network.

Actually, both Triesch's rule and Li's rule are closely related to Bell and Sejnowski's rule, which is a renowned information-maximization (infomax) rule for blind deconvolution/independent component analysis [78],

$$\begin{aligned}\Delta a &= \eta \left[\frac{1}{a} + x(1-2y) \right], \\ \Delta b &= \eta(1-2y).\end{aligned}\quad (12)$$

TABLE 2. A brief summary of different rate-based intrinsic plasticity rules.

Rule Name	Comments
Triesch's rule	maximizing information with an exponential output PDF
Li's rule	maximizing information with a Weibull output PDF, a generalization of Triesch's rule
Infomax rule	maximizing information without constraints, a generalization of above two rules

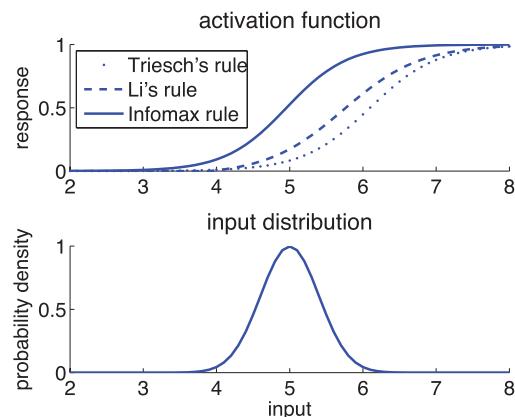


FIGURE 4. A comparison of learning results of three different rate-based IP rules.

This rule is based on maximizing the input-output mutual information of a single neuron without any constraints for the output distribution and thus has a simpler expression. As a learning result, the optimal output distribution is a uniform distribution in this case, and the neuron uses all its possible response levels equally and makes the steep parts of the activation function respond to the high density parts of the input probability density function. In other words, this information-maximization rule maximizes the discriminative ability of the neuron. Triesch's and Li's IP rules are

exactly equivalent to Bell and Sejnowski's infomax rule when neglecting the energy constraint. These three learning rules are compared in Table 2. The adapted activation functions trained by these three learning rules for the same input distribution are presented in Fig. 4. All these learning rules have a relatively straightforward interpretation since they are derived from minimizing the KL-divergence with respect to the gain and threshold of the activation function. Other activation functions such as tanh functions can also be applied in modeling and corresponding IP rules for those activation functions can be derived in a similar way.

D. INTRINSIC PLASTICITY RULES FOR SPIKING NEURON MODELS

For modeling the IP mechanism in a neural network, the rate-based model with a sigmoid activation function is far too simple to describe biological processes. As for spiking neuron models, Stemmler and Koch devised an IP rule for a Hodgkin-Huxley (HH) neuron [12]. This learning rule is realized by adjusting the voltage-dependent membrane conductances of the HH neuron to change the properties of voltage-gated ion channels. Experimental and theoretical studies indicate that neurons change their intrinsic excitability through the modification of voltage-gated ion channels [3], [4], [6], [7]. As the learning goes on, the spike-count distribution of the neuron will be a desired one. This rule can produce either exponential or uniform output distributions by modifying the constraint of the learning rule. Therefore, the output response of the neuron model can be either sparse or dense. To the best of our knowledge, this is a seminal study on modeling intrinsic plasticity. Although this learning rule is not derived from optimizing a specific objective function, it can be intuitively analyzed and works well according to simulation results.

However, it could be computationally prohibitive when simulating a large-scale network by using HH neurons. Relatively, the integrate-and-fire (IF) spiking neuron model is computationally effective to describe neural activities and it has been widely used in computational neuroscience [79], [80]. The basic subthreshold membrane potential dynamics of a single IF neuron is described by the following differential equation,

$$\tau_m \frac{dV(t)}{dt} = V_{\text{rest}} - V(t) + R_m I(t), \quad (13)$$

where $V(t)$ is the membrane potential of the single neuron, V_{rest} is the resting membrane potential, R_m is the membrane resistance, and $I(t)$ is the total injected current as the input of the neuron. τ_m is the membrane time constant and it is the product of the membrane resistance and the membrane capacity $\tau_m = R_m C_m$. The IF model does not describe the explicit biological processes essential for generating action potentials. In this model a firing time $t^{(f)}$ is defined by a threshold criterion in order to generate an action potential (spike),

$$t^{(f)} : V(t^{(f)}) = V_{\text{th}}.$$

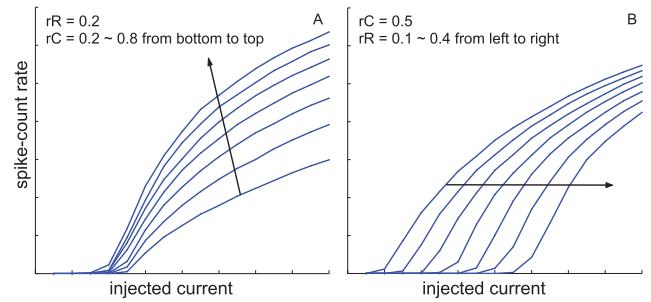


FIGURE 5. Tuning curve (spike-count rate versus constant injected current curve I) changes with different parameter settings. (A) the slope of the tuning curve becomes steeper as rC increases; (B) the threshold current of the tuning curve increases as rR grows.

When V reaches the threshold value V_{th} , an action potential is generated and the membrane potential is reset to V_{reset} , where it remains clamped for a refractory period. Savin and Triesch have developed an early IP rule for the IF neuron in a poster [81]. This IP rule is formulated as a set of simple proportional control laws for the parameters of the IF neuron. These control laws drive the first and second moments of the output distribution to match the corresponding moments of an exponential distribution. In this work, this rule is applied to solving simple ICA problems. Although the IF neuron is used, this IP rule is still rate-based and it is comparable to Triesch's rule for the rate-based neuron model.

Li and Li proposed a spike-based IP learning rule for the IF neuron model, which depends on spikes denoted by Dirac delta functions rather than using firing rates [82]. Since the IF neuron model is described by continuous-time differential equations, it is preferred to use continuous-time differential equations as well to describe the IP learning process of an IF neuron. The differential equations of the proposed IP learning rule are presented as follows,

$$\begin{aligned} \tau_{\text{ip}} \frac{drC}{dt} &= \frac{1}{rC} - yI + \beta(1-y)I, \\ \tau_{\text{ip}} \frac{drR}{dt} &= -rR + y - \beta(1-y), \end{aligned} \quad (14)$$

where $rC = \frac{1}{C_m}$ and $rR = \frac{1}{R_m}$, τ_{ip} denotes the relative integration resolution of the IF model and the IP learning rule, and the output y of the neuron in response to the input I is described by a summation of impulsive functions to denote spikes,

$$y = \epsilon \sum_f \delta(t - t^{(f)}), \quad (15)$$

where $\delta(t - t^{(f)})$ is a Dirac delta function representing a spike fired at time $t^{(f)}$ and ϵ indicates the strength of a spike. Using spikes directly is computationally more preferable and biologically more plausible for the IF neuron model. As shown in Fig. 5, the tuning curve of an IF neuron can be shifted due to the changing of two parameters, rR and rC . With this proposed spike-based learning rule, these two parameters of the IF neuron can be changed to adjust the

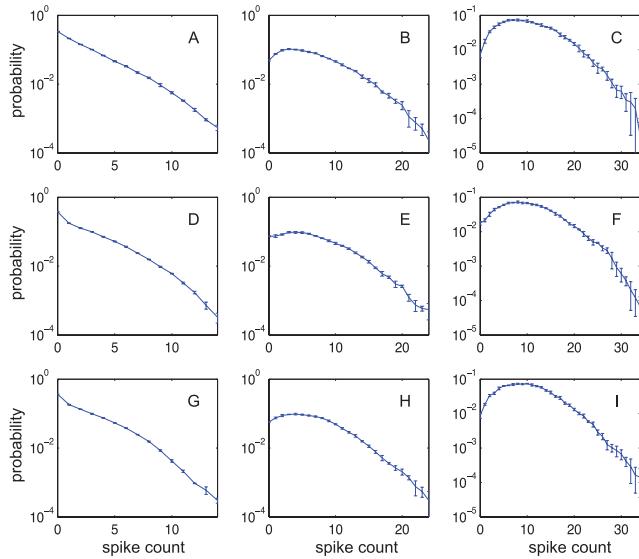


FIGURE 6. The average spike-count distributions of an adapted IF neuron with three Gaussian input distributions: pdf1 for (A,B,C), pdf2 for (D,E,F), and pdf3 for (G,H,I). The window size is adjusted so that the average number of spikes per window is the same for all neurons: two spikes for (A,D,G), six spikes for (B,E,H) and ten spikes for (C,F,I) per window. All distributions are plotted in the decimal-logarithmic space so that an exponential distribution appears as a straight line. The spike-count distributions depend on the window size used to calculate the rate. The spike-count distributions with a small window size (A,D,G) are approximately exponential, and those with a large window size (B,C,E,F,H,I) are with droop-heads, as Weibull distributions. Each spike-count distribution is obtained by averaging over 5 trials with the same 10-minute stimulus. Error bars are standard errors.

slope (selectivity) and bias (threshold) of the response curve. In this way, the average activity of an IF neuron is restricted in a homeostatic state while the spike-count distribution follows an approximate Weibull distribution (exponential distribution or probably with a droop-head due to a larger window size) as shown in Fig. 6. This rule keeps the properties of Li's rule for the rate-based neuron model and it makes an IF neuron track non-stationary inputs in a similar way of biological neurons [83], [84]. Although the IF model is a highly simplified model of biological neurons and does not specify different voltage-gated ion channels corresponding to those of biological neurons, this proposed IP rule implies a simple and direct spike-based mechanism for neuronal intrinsic plasticity. The intrinsic plasticity rules for spiking neurons models is compared in Table. 3.

TABLE 3. A brief comparison of intrinsic plasticity rules for spiking neuron models.

Method Name	Neuron Model	IP Model
Stemmler and Koch's	HH model	rate-based learning rules
Savin and Triesch's	IF model	rate-based learning rules
Li and Li's	IF model	spike-based learning rules

With IP learning rules designed for different kinds of neuron models such as rate-based neuron models with a nonlinear activation function, Hodgkin-Huxley (HH) models,

and integrated-and-fire (IF) models, it is possible to computationally study the underlying dynamics and functions of different scales of neural networks. In terms of efficient neural computation, simple neuron models like rate-based models can be used to build large-scale networks with a large number of neurons in hierarchical structures, while spiking neuron models like HH models are more suitable for studying individual neurons or small-scale networks. IF neuron models can be used in intermediate situations. IP rules for spiking neuron models can provide more specific information on biological dynamics and functions of intrinsic plasticity, and simple IP rules for rate-based models can facilitate the research on effects of intrinsic plasticity in large-scale modeling networks or applications.

IV. SYNERGISTIC LEARNING

A. SYNERGISTIC LEARNING IN COMPUTATIONAL MODELING

Intrinsic plasticity is a relatively novel field in comparison to synaptic plasticity. Most studies have focused on these two different neural mechanisms separately and relatively little work was contributing to illustrate the combinational effects of them. We have known that synaptic plasticity and intrinsic plasticity coexist in biological neural systems, it is interesting to study how these two different mechanisms are working together to complete computational tasks. In real biological neural systems, neurons are connected to each other in a complicated way, the modification produced by synaptic plasticity will affect the learning process of intrinsic plasticity, conversely, intrinsic plasticity will also have an influence on synaptic plasticity. Therefore, actually it is not a simple combination when these two plasticity mechanisms are working together.

Some modeling studies have tried to combine different plasticity mechanisms for sparse coding and independent component analysis (ICA). In [85], Földiak proposed a learning model for a simple ANN to form sparse coding representation. The learning units in this network are simple rate-based model neurons and the activation functions are also sigmoid functions as shown in Eq. (9). The training method is using Hebbian learning to update feedforward excitatory connections and an anti-Hebbian rule (Földiak's rule) to update local inhibitory connections. In addition, a sliding threshold of the activation function is used to keep the output response in a target level,

$$\Delta b_j = \eta(y_j - y_{\text{target}}), \quad (16)$$

where b_j is the threshold of the sigmoid activation function for the neuron j and y_{target} is the target firing rate for this neuron, which is specified before training. In fact, this usage of a sliding threshold could be earlier than the discovery of the intrinsic plasticity in biological experiments, but it can be treated as a simple rule to model the intrinsic plasticity process. Földiak's learning model is a classical ANN method for sparse coding. By using the same rate-based neuron model, Weber and Triesch applied Triesch's IP rule to a

generative model for sparse coding [86]. Triesch's IP rule enforces sparseness of the firing activities of the learning units. As for synaptic plasticity in this generative framework, the bottom-up and top-down connection weights are trained by local learning rules provided by the wake-sleep algorithm.

In [87], King et al used a spiking neuron model to drive sparse code formation. In this method, connections of the network are also trained by different forms of weight modification rules. The excitatory feedforward connections are trained by Oja's rule while the horizontal inhibitory connections are trained by the correlation measuring (CM) rule, which is a variant version of Földiak's rule. As for homeostatic spike-rate regulation, this method uses an adapting threshold of a step activation function to stabilize the output response in a target firing level over time,

$$\Delta\theta_j = \eta(\langle z_j \rangle - y_{\text{target}}), \quad (17)$$

where θ_j is the threshold to generate the spikes of the neuron j , $\langle z_j \rangle$ is the average firing rate calculated by the binary spike output z_j , and y_{target} is the target firing rate for this neuron. Although a spiking neuron model is used in this method, the learning rules in this method are all rate-based. As an important contribution of King's model, neurons are separated into an excitatory group and an inhibitory group, which is more biologically realistic. In biological neural systems, it could be impossible for a neuron to provide both excitation and inhibition to other neurons. Therefore, King's model is also called E-I Net (excitatory-inhibitory network). Fig. 7 shows sparse codes of excitatory neurons generated by King's E-I Net.

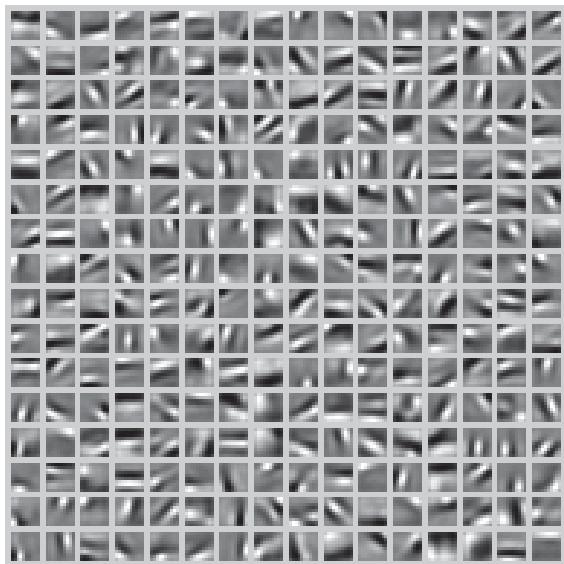


FIGURE 7. Sparse codes of excitatory neurons generated by King's E-I Net. These sparse codes are reproduced by King's Matlab codes. Földiak's model and Savin's model can also produce similar sparse codes from natural images.

Researchers have also used spike-based IP learning rules for sparse coding/ICA. In [70], Savin et al used a spiking

neuron model trained by spike-based rules for ICA. The neuron model in this work is a stochastically spiking neuron with refractoriness. The learning method for training the neural network combines spike-timing-dependent plasticity (STDP), synaptic scaling and intrinsic plasticity. The IP learning rule in this method is using spikes to modify the parameters of the neuron model in order to maximize the input-output mutual information under an exponential output distribution. The IP rule is much more complicated and the derivation of this rule is using the same idea of deriving Triesch' rule. In this work, both the synaptic plasticity rule and the intrinsic plasticity rule are based on spikes, which is different to the rate-based learning rules used in Földiak's model and King's model.

TABLE 4. A comparison of three different synergistic learning methods for sparse coding/ICA.

Method Name	Neuron Model	IP Model
Földiak's	rate-based model	rate-based learning rules
King's	spiking model	rate-based learning rules
Savin's	spiking model	spike-based learning rules

A summary of these learning methods for sparse coding/ICA in [70], [85], and [87] is presented in Table. 4. In these learning frameworks, sparse coding is mostly achieved by a combination of Hebbian-like rules for feedforward excitatory connections, anti-Hebbian rules for feedback/horizontal inhibitory connections, and intrinsic plasticity as a strictly local learning rule for the response tuning curves of individual neurons. Hebbian-like learning rule is trying to build the positive correlation between the input data and the output response while anti-Hebbian learning results in local competition to decorrelate the responses of different neurons and thus to drive a sparse response over the global network. The hemostatic state of the neural network is realized by the local learning of intrinsic plasticity. As the response of a single neuron is relatively low, the firing of the global network will be sparse when different neurons are decorrelated (unsynchronized). Therefore, it explains why these different rules can be working together in a neural network to form spare representations. In these studies, different forms of plasticity mechanisms are combined to generate sparse representation from natural scene stimuli. The results of these synergistic models, especially of the spiking neuron models, indicate that it is very likely that the sparse coding effect found in lower levels of the visual cortex is a combining result of different plasticity mechanisms. An schematic graph of a general model combining different plasticity mechanisms for sparse coding/ICA is shown in Fig. 8.

In local competition analysis (LCA) networks [52], the threshold of the activation function can also be adapted to speed up and stabilize the learning process. This network is designed as a biologically realistic model for sparse coding and a corresponding algorithm for calculating the responses

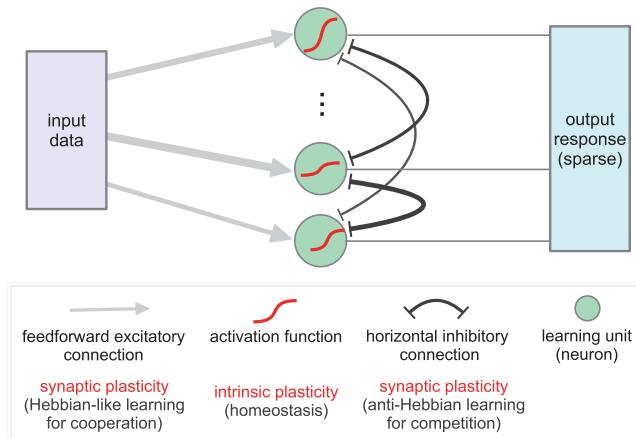


FIGURE 8. An general model combining different plasticity mechanisms for sparse coding/ICA. Three important mechanisms are utilized in this general model, (1) Hebbian-like learning for feedforward connections to form cooperation between the input and output of learning units, (2) anti-Hebbian learning for horizontal connections to produce competition within learning units so as to decorrelate the responses of learning units, and (3) intrinsic plasticity of single learning units for a homeostatic state of the whole network. As the response of a single learning unit is relatively low, the activity of the global network will be sparse with the decorrelation (unsynchronization) among different neurons introduced by anti-Hebbian learning.

of neurons is proposed. The learning of the feedforward connections in this method is realized by Oja's rules and it is as same as the algorithm derived by minimizing the objective function of mean squared error (MSE) which is the difference between the original input and its reconstruction from the sparse codes.

In [88], Lazar et al presented a self-organizing recurrent network (SORN) combining intrinsic plasticity and synaptic plasticity to learn spatio-temporal patterns in its input while maintaining its dynamics in a healthy regime for learning. In this method, the IP rule regulates a neuron's firing threshold to maintain a low average activity level. The synaptic rule is a simple model of STDP. This study is trying to make an effort to understand how different plasticity mechanisms interact to shape the brain's representation. In an earlier study, Lazar et al have used recurrent neural networks trained by different forms of plasticity for fading memory and time series prediction [89]. This work can also be treated as an application work.

Sparse coding, ICA and SOM are all trying to generate efficient representation for input data and they are intrinsically related to each other as unsupervised learning methods. Synergistic learning methods are also applied in other computational work in addition to sparse coding/ICA/SOM in neural networks.

In [90], Marković and Gros introduced the intrinsic plasticity rule into random recurrent neural networks to produce the self-sustained neural activity, which is a likely essential precondition to complex behavioral patterns. This work shows synaptic plasticity is insufficient to generate self-sustained activity for non-spiking neurons and nonsynaptic

mechanisms play an important role in the formation of complex patterns of neural activities [90], [91]. The simulation results have also indicated that self-organization of neural network dynamics into a critical regime could occur whenever intrinsic and synaptic plasticity are both present and relevant. Critical neural dynamics would then result from the interplay of the synergistic learning process. The intrinsic plasticity rule used in this work is derived by optimizing the information transfer between neural input and output, and basically it is not different from Triesch's rule.

Williams and Noble [92] and Williams [93] combined a simple intrinsic plasticity rule for adjusting the activation function threshold with a synaptic scaling rule to counteract the node saturation effect in continuous-time recurrent neural networks (CTRNNs), which is an effect that leaves neurons insensitive to input and stop signals from propagating. Both intrinsic plasticity and synaptic scaling are working as homeostatic processes. Simulation results in this study have shown that combining intrinsic plasticity and synaptic scaling can boost the node sensitivity and improve signal propagation. Thus it implies that homeostatic learning processes may prevent quiescence and hyper-excitation in biological neural systems.

Potentially, some interesting experimental results could be interpreted by the combination of synaptic plasticity and intrinsic plasticity. For example, in monocular deprivation experiments of mice and rodents, the response level of the neurons will decrease one day after the deprivation, but surprisingly, the response goes back to the original level after several days [94], [95]. Referring to the neurophysiological results and modeling of intrinsic plasticity, neurons have the ability to shift their response tuning curves to enhance the excitability when their inputs are deprived. It makes sense to conjecture that intrinsic plasticity plays a role in these visual deprivation experiments. In those experiments of monkeys for familiarity effects, the response of repeated stimuli will be reduced and the response tuning curve for these stimuli will be sharpened, resulting in a sparse representation of these familiar stimuli [96]–[99]. It is very likely that these results are produced by an information maximization process, which could be a combination of intrinsic and synaptic mechanisms. By using experimental approaches, it could be difficult to separate these two plasticity mechanisms in the above-mentioned experiments. By modeling with these two different plasticity mechanisms, we may be able to study the synergies between them.

B. SYNERGISTIC LEARNING IN MACHINE LEARNING

Since synaptic plasticity provides a biological background for weight modification rules in artificial neural networks, it is a straight research line to use IP learning rules in learning machines for solving problems such as regression, prediction, classification and clustering. Some unsupervised learning methods including sparse coding, ICA and SOM are discussed in the previous section, but to some extent these methods are originally modeling the learning process

of neural systems rather than providing effective algorithms for real-world applications. In this section we will focus on supervised learning algorithms in engineering combining ideas of intrinsic plasticity and synaptic plasticity.

Several studies have combined connection modification algorithms with rate-based IP rules or other modified versions for training artificial neural networks in machine learning. Simulation results of these studies have shown it is promising to utilize synergistic learning methods in applications.

In [100], Stein introduced an online learning algorithm with IP to improve the performance of reservoir networks and it has been shown that intrinsic plasticity and synaptic plasticity cooperate harmonically in reservoir networks. The intrinsic plasticity rule used in this work is Triesch's rule. Some convergence analysis is also provided in this work which is really important to verify that IP rules can be added into learning machines compatibly. In [101], Schrauwen et al used some other modified versions of Triesch's rule to improve the synergistic learning performance in [100]. These modified IP rules are trying to maximize the input-output mutual information with different output distributions. In [102], Neumann and Stein used an unsupervised scheme including the IP rule to pretrain the extreme learning machines (EML). The IP rule used in this method adapts the hidden layer analytically by a pseudo inverse technique instead of performing a computationally expensive gradient-descent.

As a recent study in [103], Li and Li have shown good performance of artificial neural networks trained by Bell and Sejnowski's infomax rule for adjusting activation functions of learning units and an error-entropy minimization (MEE) algorithm for updating weights. The MEE algorithm uses the error entropy criterion (EEC) to formulate the objective function for weight modification. EEC considers higher order statistical information of the error distribution and it is a good alternative of the mean squared error (MSE) criterion to get better performance of learning machines. This method is called information-theoretic learning [104]–[109]. The original Bell and Sejnowski's infomax rule in [78] is used for a neural network to solve ICA problems but in this study it is used as an IP rule for a single learning unit. As mentioned in the previous section, this rule adjusts the slope and the threshold of the activation function to maximize the mutual information between the input and output of a single neuron. As the IP rule adjusts the activation function of a single neuron without the energy constraint, all output levels tend to appear equivalently. Thus the input can be encoded much more efficiently and the discriminative ability of the neuron is maximized. In this learning paradigm, both weight modification and activation function modification are on the basis of information theory (optimizing the entropy).

In order to display the performance of the proposed information-theoretic synergistic learning algorithm, here we show a general class of feedforward neural networks (FNN) as an example. As shown in Fig. 9, this simple neural network is composed of an input layer, a single hidden layer and an output layer. The network input consisting of P external

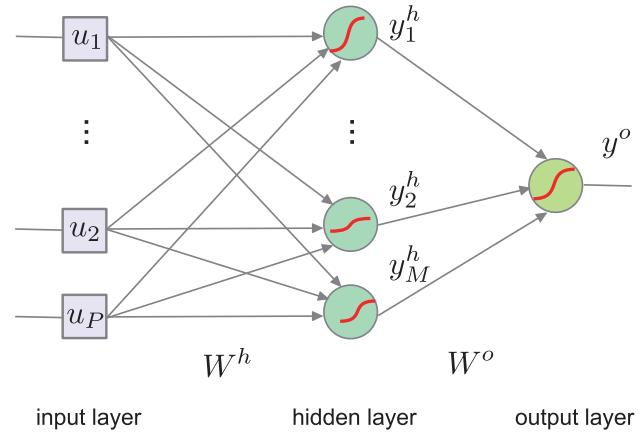


FIGURE 9. The construction of a simple feedforward neural network trained by the synergistic learning algorithm.

elements can be described by the $P \times 1$ vector \mathbf{u} . In the hidden layer, there are M neurons (processing elements). Each neuron in the hidden layer receives the weighted sum of the network input \mathbf{u} . The output of these neurons, \mathbf{y}^h , is described as \mathbf{y}^h , and is calculated by

$$\begin{aligned} \mathbf{y}^h &= \varphi(\mathbf{v}^h), \\ \mathbf{v}^h &= (\mathbf{W}^h)^T \mathbf{u}, \end{aligned} \quad (18)$$

where \mathbf{v}^h represents the input vector of the neurons in the hidden layer, and \mathbf{W}^h represents the $P \times M$ synaptic weight matrix connected from the input layer to the hidden layer. An element $w_{k,l}^h$ of this matrix represents the weight connection from the l th input node to the k th hidden neuron. For the output layer, we consider only one neuron, which receives the weighted sum of the output of hidden neurons, \mathbf{y}^h , and produces the network output, y^o . The calculation is described as

$$\begin{aligned} y^o &= \varphi(v^o), \\ v^o &= (\mathbf{W}^o)^T \mathbf{y}^h, \end{aligned} \quad (19)$$

where \mathbf{W}^o represents the $M \times 1$ synaptic weight matrix connected from the hidden layer to the output layer. An element w_i^o of this matrix represents the weight connection from the i th hidden neuron to the final output. The activation function of each neuron $\varphi(\cdot)$ is a nonlinear function and in this work the tanh activation function is used. The activation function of the k th neuron, $\varphi_k(\cdot)$, is

$$\begin{aligned} y_k &= \varphi_k(v_k; a_k, b_k) \\ &= \tanh(a_k v_k + b_k) \\ &= \frac{\exp(a_k v_k + b_k) - \exp(-a_k v_k - b_k)}{\exp(a_k v_k + b_k) + \exp(-a_k v_k - b_k)}, \end{aligned} \quad (20)$$

where v_k is the input of the k th neuron, y_k is the output of the k th neuron, a_k represents the sensitivity of the activation function and b_k is the bias, the corresponding information-maximization learning algorithm can be obtained as follows

$$\begin{aligned} a_k &= a_k + \eta_{IP} \left(\frac{1}{a_k} - 2E[v_k y_k] \right), \\ b_k &= b_k + \eta_{IP} (-2E[y_k]), \end{aligned} \quad (21)$$

where η_{IP} is the learning rate of intrinsic plasticity. For a training set including n_0 samples, the input-output pairs of the k th neuron, $[v_k(1), \dots, v_k(n_0)]^T$ and $[y_k(1), \dots, y_k(n_0)]^T$, are used to estimate the expected values $E[v_k y_k]$ and $E[y_k]$.

In terms of weight modification, the MEE algorithm is presented as below [103],

$$\Delta w = \eta \frac{1}{2n_0^2\sigma^2} \sum_{s=1}^{n_0} \sum_{t=1}^{n_0} (e(s) - e(t)) \cdot G_{\sigma\sqrt{2}}(e(t) - e(s)) \cdot \left(\frac{\partial e(t)}{\partial w} - \frac{\partial e(s)}{\partial w} \right), \quad (22)$$

where n_0 samples are presented in one weight updating step, G denotes the Gaussian kernel function and σ represents the kernel size for probability density function estimation [110]. The error signal e is defined as the difference between the desired output and the network output, $e = d - y^o$. For the output layer of a single-output FNN, the derivative of the error e with respect to the weight w_i^o in the $M \times 1$ matrix \mathbf{W}^o can be calculated as

$$\frac{\partial e}{\partial w_i^o} = -\varphi'(v^o)y_i^h, \quad (23)$$

where

$$\varphi'(v) = a(1 - \tanh^2(av + b)) = a(1 - y^2).$$

The derivative of the error e with respect to the weight $w_{k,l}^h$ in the matrix \mathbf{W}^h can be calculated from error back-propagation [111],

$$\frac{\partial e}{\partial w_{k,l}^h} = -\varphi'(v^o)w_k^o\varphi'(v_k^h)u_l. \quad (24)$$

By using Eq. (23) and Eq. (24), the weight modification algorithm in Eq. (22) can be computed.

The proposed synergistic learning algorithm for the FNN is summarized as follows:

Step 1 (Initialization): Choose a random set of small values for the $P \times M$ hidden layer weight matrix \mathbf{W}^h and the $M \times 1$ output layer weight matrix \mathbf{W}^o . Set $a = 1$ and $b = 0$ for each neuron.

Step 2 (Repetition): The epochwise training procedure begins with $n = 1$. Repeat the network calculations with the input vector $\mathbf{u}(n)$ as shown in Eq. (18), (19) and (20).

Step 3 (Weight Matrix Update): Update the weight matrices \mathbf{W}^o and \mathbf{W}^h by the weight updating algorithm based on error entropy maximization as shown in Eq. (22), (23) and (24).

Step 4 (Activation Function Update): Update the parameters a_k and b_k of the activation function φ_k of the neuron k using the intrinsic plasticity rule described in Eq. (21) with all sets of $v_k(n)$ and $y_k(n)$ of n_0 samples. By the batch version of the IP rule, the parameters a_k and b_k are only updated once during an epoch.

Step 5 (Return or Stop): If the stopping criterion is satisfied, the training procedure is stopped; otherwise, set $n = 1$ and return to Step 2.

As shown in Fig. 10, this information-maximization IP rule improves the ANNs in both the convergence speed and the

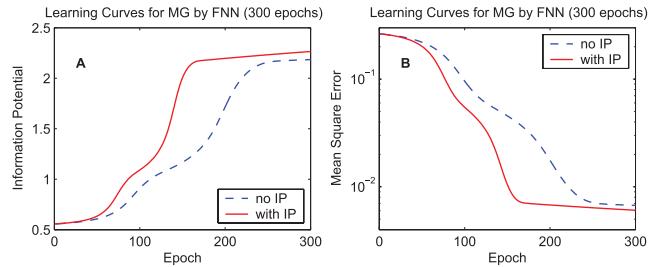


FIGURE 10. Learning performance of the synergies between information theoretic learning and infomax intrinsic plasticity. All these lines are 300-epoch learning curves of a 3-layer feedforward network trained with a data set of a Mackey-Glass chaotic time series ("MG"). The dashed lines denote the learning curves of the original MEE algorithm, and the solid lines denote the learning curves of the synergistic algorithm. (A) Learning curves of the quadratic information potential. (B) Learning curves of the mean squared error.

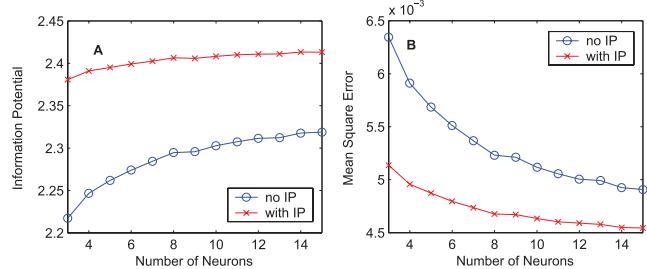


FIGURE 11. Relation between the training result and the number of hidden neurons of a 3-layer feedforward network. Training results are obtained after 1000 epochs for a data set of a Mackey-Glass chaotic time series ("MG"). The circle markers denote the results obtained by the MEE algorithm, and the cross markers denote the results obtained by the synergistic algorithm. (A) Results of the quadratic information potential. (B) Results of the mean squared error. The number of neuron units can be highly reduced with the infomax IP rule when the network still keeps good performance. This is an encouraging result of synergistic learning.

final training result. It shows that the discriminative ability of a neuron plays a critical role in improving the performance of ANNs. More importantly, due to the local nature of the IP mechanism and the simplicity of Bell and Sejnowski's infomax rule, this synergistic learning algorithm can be implemented with a relatively small increase in computational cost in comparison to the original information-theoretic learning algorithm. As shown in Fig. 11, a small-scale network with a few number of learning units can get satisfying performance with the assistance of the infomax IP rule. The number of neuron units can be highly reduced with the infomax IP rule as the network still keeps good performance. If the number of neuron units is fewer, the corresponding connections will be deleted, thereby reducing the calculation significantly. This is a very important computational feature of the synergistic learning. Keeping a relative low output response of a single neuron is important for sparse coding models in the last section, however, in terms of artificial neural networks for applications, maximizing the input-output mutual information seems more important and we can simply ignore the energy constraint.

In all these applications, the weight modification algorithms as synaptic plasticity rules are derived by

optimizing objective functions, which is a classical and intuitive method as mentioned in the previous section. However, these rate-based intrinsic plasticity rules are not directly related to the objective functions. If these parameters of activation functions are updated to optimize the objective functions, it is not surprising that the performance will be improved. Actually, this kind of learning rules for activation functions is usually computationally complex and it is not easy to derive them by optimizing the objective functions. Although the performance can be improved in this way, it will highly increase the computational cost. Note that the rate-based IP rules we have mentioned above are usually simple and can be easily implemented. Therefore, the rate-based IP rules will be beneficial with only a little additional computational cost.

V. CONCLUSION AND DISCUSSION

In this review, we have interpreted intrinsic plasticity and synaptic plasticity in biological neural systems and described corresponding learning rules in modeling. In fact, most current methods on modeling and applications only consider the weight modification, which is related to synaptic plasticity in biological neural systems. As a novel neural mechanism, intrinsic plasticity provides the biological background for adjusting the intrinsic excitability of the learning units in artificial neural systems. In the computational hypothesis, the IP mechanism is adjusting the response curve of a neuron to preserve the input information while maintaining an appropriate level of firing activity. The intrinsic plasticity rule can not only maintain the activities of the learning units in a homeostatic state but also boost the ability to represent the synaptic input. This is a fantastic effect for both the biological systems and the artificial information processing systems. Therefore, it seems demanding to study the synergies between IP and weight modification in artificial neural systems. We have presented how the IP rule can be combined with synaptic plasticity in computational modeling and machine learning to improve the performance in a synergistic way.

In computational modeling with synergistic learning, there is a consistent framework for sparse coding or independent component analysis. Hebbian-like learning algorithm establishes the positive correlation between the input and the learning units, anti-Hebbian learning deassociates the correlation between learning units and the intrinsic plasticity rule maintains sparse firing activities of the learning units. Each learning rule plays a different role and they work harmonically with each other for a unique goal.

In machine learning algorithms, an efficient representation of the input for each learning unit seems important to improve the performance. Inspired by connectionism, a large number of weight modification algorithms for learning machines have been developed and made a wonderful success in applications. Far less work has tried to apply intrinsic excitability modification of the learning units. According to these synergistic learning algorithms we have reviewed, intrinsic plasticity can be beneficial to improving the performance of the

learning machines. It is interesting to see how this excitability modification can be added to different network structures and be combined with different weight modification algorithms.

On the other hand, as we stated that the combination of intrinsic and synaptic plasticity behaves in a complex way and these two mechanisms are reciprocal to each other during the learning process, it is difficult to provide some accurate analysis for the properties of the synergistic learning process. Very limited studies have mathematically analyzed it and provided convergence conditions, restrictions, and theoretical insights for the combination.

In terms of keeping a homeostatic learning state, both Oja's rule and the BCM rule can maintain the synaptic plasticity process in a relatively stable way as we mentioned in the previous section. STDP as a spike-based biological process can also keep the modification of synapses restricted in a certain regime as it shares the properties of the BCM rule or a stable Hebbian rule [42], [44], [45]. On the other hand, synaptic scaling is a global modification mechanism to keep a homeostatic state for synaptic plasticity. During the modification of synaptic scaling, synapses are globally adjusted by the strength of other synapses. This process can be either multiplicative or subtractive. Although intrinsic plasticity is not the only mechanism counteracting the instability of neurons, it is still a really encouraging finding which is beyond synapses. Intrinsic plasticity controls the output response of neurons much more directly in comparison to synaptic processes. More importantly, intrinsic plasticity has the ability to keep a relatively high discrimination for synaptic inputs, and therefore provides efficient neural coding and signal propagation. It is really interesting that there are many different mechanisms cooperate with each other to keep the neural activities in a homeostatic regime, which in turn reflects that homeostasis is a key feature of the neural systems.

Inspired by the discoveries of different plasticity mechanisms in biological neural systems, we believe that synergistic learning can make a difference in computational neuroscience and machine learning, and learning results in modeling and applications are able to explore potential synergistic functions of different plasticity mechanisms in biological neural systems.

REFERENCES

- [1] D. H. Hubel and T. N. Wiesel, "Receptive fields and functional architecture of monkey striate cortex," *J. Physiol.*, vol. 195, no. 1, pp. 215–243, 1968.
- [2] E. Marder, L. F. Abbott, G. G. Turrigiano, Z. Liu, and J. Golowasch, "Memory from the dynamics of intrinsic membrane currents," *Proc. Natl. Acad. Sci. USA*, vol. 93, no. 24, pp. 13481–13486, 1996.
- [3] N. S. Desai, L. C. Rutherford, and G. G. Turrigiano, "Plasticity in the intrinsic excitability of cortical pyramidal neurons," *Nature Neurosci.*, vol. 2, no. 6, pp. 515–520, 1999.
- [4] W. Zhang and D. J. Linden, "The other side of the engram: Experience-driven changes in neuronal intrinsic excitability," *Nature Rev. Neurosci.*, vol. 4, no. 11, pp. 885–900, 2003.
- [5] G. Daoudal and D. Debanne, "Long-term plasticity of intrinsic excitability: Learning rules and mechanisms," *Learn. Memory*, vol. 10, no. 6, pp. 456–465, 2003.

- [6] M. Zhang, F. S. Hung, Y. Zhu, Z. Xie, and J.-H. Wang, "Calcium signal-dependent plasticity of neuronal excitability developed postnatally," *J. Neurobiol.*, vol. 61, no. 2, pp. 277–287, 2004.
- [7] R. H. Cudmore and G. G. Turrigiano, "Long-term potentiation of intrinsic excitability in LV visual cortical neurons," *J. Neurophysiol.*, vol. 92, no. 1, pp. 341–348, 2004.
- [8] R. Mozzachiodi and J. H. Byrne, "More than synaptic plasticity: Role of nonsynaptic plasticity in learning and memory," *Trends Neurosci.*, vol. 33, no. 1, pp. 17–26, 2010.
- [9] A. J. Watt and N. S. Desai, "Homeostatic plasticity and STDP: Keeping a neuron's cool in a fluctuating world," *Frontiers Synaptic Neurosci.*, vol. 2, Jun. 2010, Art. ID 5.
- [10] G. G. Turrigiano and S. B. Nelson, "Homeostatic plasticity in the developing nervous system," *Nature Rev. Neurosci.*, vol. 5, no. 2, pp. 97–107, 2004.
- [11] S. Kourrich, D. J. Calu, and A. Bonci, "Intrinsic plasticity: An emerging player in addiction," *Nature Rev. Neurosci.*, vol. 16, no. 3, pp. 173–184, 2015.
- [12] M. Stemmler and C. Koch, "How voltage-dependent conductances can adapt to maximize the information encoded by neuronal firing rate," *Nature Neurosci.*, vol. 2, no. 6, pp. 521–527, 1999.
- [13] J. Triesch, "A gradient rule for the plasticity of a neuron's intrinsic excitability," in *Artificial Neural Networks: Biological Inspirations*. New York, NY, USA: Springer, 2005, pp. 65–70.
- [14] J. Triesch, "Synergies between intrinsic and synaptic plasticity mechanisms," *Neural Comput.*, vol. 19, no. 4, pp. 885–909, 2007.
- [15] T. V. Bliss and T. Lømo, "Long-lasting potentiation of synaptic transmission in the dentate area of the anaesthetized rabbit following stimulation of the perforant path," *J. Physiol.*, vol. 232, no. 2, pp. 331–356, 1973.
- [16] T. V. Bliss and A. Gardner-Medwin, "Long-lasting potentiation of synaptic transmission in the dentate area of the unanaesthetized rabbit following stimulation of the perforant path," *J. Physiol.*, vol. 232, no. 2, pp. 357–374, 1973.
- [17] T. J. Teyler and P. DiScenna, "Long-term potentiation," *Annu. Rev. Neurosci.*, vol. 10, no. 1, pp. 131–161, 1987.
- [18] T. V. Bliss and G. L. Collingridge, "A synaptic model of memory: Long-term potentiation in the hippocampus," *Nature*, vol. 361, no. 6407, pp. 31–39, 1993.
- [19] R. C. Malenka and R. A. Nicoll, "Long-term potentiation—A decade of progress?" *Science*, vol. 285, no. 5435, pp. 1870–1874, 1999.
- [20] J. S. Haas, B. Zavala, and C. E. Landisman, "Activity-dependent long-term depression of electrical synapses," *Science*, vol. 334, no. 6054, pp. 389–393, 2011.
- [21] H. Ko et al., "The emergence of functional microcircuits in visual cortex," *Nature*, vol. 496, no. 7443, pp. 96–100, 2013.
- [22] E. Oja, "Simplified neuron model as a principal component analyzer," *J. Math. Biol.*, vol. 15, no. 3, pp. 267–273, 1982.
- [23] E. Oja, *Subspace Methods of Pattern Recognition*. New York, NY, USA: Wiley, 1983.
- [24] E. Oja, "Principal components, minor components, and linear neural networks," *Neural Netw.*, vol. 5, no. 6, pp. 927–935, 1992.
- [25] E. L. Bienenstock, L. N. Cooper, and P. W. Munro, "Theory for the development of neuron selectivity: Orientation specificity and binocular interaction in visual cortex," *J. Neurosci.*, vol. 2, no. 1, pp. 32–48, 1982.
- [26] N. Intrator and L. N. Cooper, "Objective function formulation of the BCM theory of visual cortical plasticity: Statistical connections, stability conditions," *Neural Netw.*, vol. 5, no. 1, pp. 3–17, 1992.
- [27] C. C. Law and L. N. Cooper, "Formation of receptive fields in realistic visual environments according to the Bienenstock, Cooper, and Munro (BCM) theory," *Proc. Nat. Acad. Sci. USA*, vol. 91, no. 16, pp. 7797–7801, 1994.
- [28] L. N. Cooper, N. Intrator, B. S. Blais, and H. Z. Shouval, *Theory of Cortical Plasticity*. Singapore: World Scientific, 2004.
- [29] S. M. Dudek and M. F. Bear, "Homosynaptic long-term depression in area CA1 of hippocampus and effects of N-methyl-D-aspartate receptor blockade," *Proc. Nat. Acad. Sci. USA*, vol. 89, no. 10, pp. 4363–4367, 1992.
- [30] R. M. Mulkey and R. C. Malenka, "Mechanisms underlying induction of homosynaptic long-term depression in area CA1 of the hippocampus," *Neuron*, vol. 9, no. 5, pp. 967–975, 1992.
- [31] A. Artola and W. Singer, "Long-term potentiation and NMDA receptors in rat visual cortex," *Nature*, vol. 330, no. 6149, pp. 649–652, 1987.
- [32] A. Kirkwood and M. F. Bear, "Hebbian synapses in visual cortex," *J. Neurosci.*, vol. 14, no. 3, pp. 1634–1645, 1994.
- [33] M. Mayford, J. Wang, E. R. Kandel, and T. J. O'Dell, "CaMKII regulates the frequency-response function of hippocampal synapses for the production of both LTD and LTP," *Cell*, vol. 81, no. 6, pp. 891–904, 1995.
- [34] C. C. Bell, V. Z. Han, Y. Sugawara, and K. Grant, "Synaptic plasticity in a cerebellum-like structure depends on temporal order," *Nature*, vol. 387, no. 6630, pp. 278–281, 1997.
- [35] H. Markram, J. Lübke, M. Frotscher, and B. Sakmann, "Regulation of synaptic efficacy by coincidence of postsynaptic APs and EPSPs," *Science*, vol. 275, no. 5297, pp. 213–215, 1997.
- [36] G. Q. Bi and M. M. Poo, "Synaptic modifications in cultured hippocampal neurons: Dependence on spike timing, synaptic strength, and postsynaptic cell type," *J. Neurosci.*, vol. 18, no. 24, pp. 10464–10472, 1998.
- [37] Y. Dan and M.-M. Poo, "Spike timing-dependent plasticity of neural circuits," *Neuron*, vol. 44, no. 1, pp. 23–30, 2004.
- [38] W. Gerstner, R. Kempter, J. L. van Hemmen, and H. Wagner, "A neuronal learning rule for sub-millisecond temporal coding," *Nature*, vol. 383, no. 6595, pp. 76–78, 1996.
- [39] R. Kempter, W. Gerstner, and J. L. van Hemmen, "Hebbian learning and spiking neurons," *Phys. Rev. E*, vol. 59, no. 4, pp. 4498–4514, 1999.
- [40] L. I. Zhang, H. W. Tao, C. E. Holt, W. A. Harris, and M.-M. Poo, "A critical window for cooperation and competition among developing retinotectal synapses," *Nature*, vol. 395, no. 6697, pp. 37–44, 1998.
- [41] S. Song, K. D. Miller, and L. F. Abbott, "Competitive Hebbian learning through spike-timing-dependent synaptic plasticity," *Nature Neurosci.*, vol. 3, no. 9, pp. 919–926, 2000.
- [42] E. M. Izhikevich and N. S. Desai, "Relating STDP to BCM," *Neural Comput.*, vol. 15, no. 7, pp. 1511–1523, 2003.
- [43] N. Caporale and Y. Dan, "Spike timing-dependent plasticity: A Hebbian learning rule," *Annu. Rev. Neurosci.*, vol. 31, pp. 25–46, Jul. 2000.
- [44] L. F. Abbott and S. B. Nelson, "Synaptic plasticity: Taming the beast," *Nature Neurosci.*, vol. 3, pp. 1178–1183, Nov. 2000.
- [45] M. C. van Rossum, G. Q. Bi, and G. G. Turrigiano, "Stable Hebbian learning from spike timing-dependent plasticity," *J. Neurosci.*, vol. 20, no. 23, pp. 8812–8821, 2000.
- [46] C. Clopath, L. Büsing, E. Vasiliaki, and W. Gerstner, "Connectivity reflects coding: A model of voltage-based STDP with homeostasis," *Nature Neurosci.*, vol. 13, no. 3, pp. 344–352, 2010.
- [47] H. Z. Shouval, M. F. Bear, and L. N. Cooper, "A unified model of NMDA receptor-dependent bidirectional synaptic plasticity," *Proc. Nat. Acad. Sci. USA*, vol. 99, no. 16, pp. 10831–10836, 2002.
- [48] H. Z. Shouval, "What is the appropriate description level for synaptic plasticity?" *Proc. Nat. Acad. Sci. USA*, vol. 108, no. 48, pp. 19103–19104, 2011.
- [49] A. Hyvärinen and E. Oja, "Independent component analysis by general nonlinear Hebbian-like learning rules," *Signal Process.*, vol. 64, no. 3, pp. 301–313, 1998.
- [50] A. Hyvärinen and E. Oja, "Independent component analysis: Algorithms and applications," *Neural Netw.*, vol. 13, nos. 4–5, pp. 411–430, 2000.
- [51] A. Hyvärinen, J. Karhunen, and E. Oja, *Independent Component Analysis*. New York, NY, USA: Wiley, 2004.
- [52] C. J. Rozell, D. H. Johnson, R. G. Baraniuk, and B. A. Olshausen, "Sparse coding via thresholding and local competition in neural circuits," *Neural Comput.*, vol. 20, no. 10, pp. 2526–2563, 2008.
- [53] I. Jolliffe, *Principal Component Analysis*. New York, NY, USA: Springer-Verlag, 2002.
- [54] B. A. Olshausen and D. J. Field, "Emergence of simple-cell receptive field properties by learning a sparse code for natural images," *Nature*, vol. 381, no. 6583, pp. 607–609, 1996.
- [55] B. A. Olshausen and D. J. Field, "Sparse coding with an overcomplete basis set: A strategy employed by V1?" *Vis. Res.*, vol. 37, no. 23, pp. 3311–3325, 1997.
- [56] T. Kohonen, "The self-organizing map," *Proc. IEEE*, vol. 78, no. 9, pp. 1464–1480, Sep. 1990.
- [57] D. E. Rumelhart, G. E. Hinton, and R. J. Williams, "Learning internal representations by error propagation," in *Parallel Distributed Processing: Explorations in the Microstructure of Cognition*, vol. 1, J. L. McClelland and D. E. Rumelhart, Eds. Cambridge, MA, USA: MIT Press, 1987, pp. 318–362.
- [58] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet classification with deep convolutional neural networks," in *Proc. Adv. Neural Inf. Process. Syst.*, 2012, pp. 1097–1105.

- [59] M. D. Zeiler and R. Fergus, "Visualizing and understanding convolutional networks," in *Computer Vision*. New York, NY, USA: Springer, 2014, pp. 818–833.
- [60] R. Salakhutdinov and G. E. Hinton, "Deep Boltzmann machines," in *Proc. Int. Conf. Artif. Intell. Statist.*, 2009, pp. 448–455.
- [61] K. D. Harris, "Stability of the fittest: Organizing learning through retroaxonal signals," *Trends Neurosci.*, vol. 31, no. 3, pp. 130–136, 2008.
- [62] S. X. Luo, R. Axel, and L. F. Abbott, "Generating sparse and selective third-order responses in the olfactory system of the fly," *Proc. Nat. Acad. Sci. USA*, vol. 107, no. 23, pp. 10713–10718, 2010.
- [63] A. Kennedy, G. Wayne, P. Kaifosh, K. Alviña, L. F. Abbott, and N. B. Sawtell, "A temporal basis for predicting the sensory consequences of motor commands in an electric fish," *Nature Neurosci.*, vol. 17, pp. 416–422, Feb. 2014.
- [64] R. Baddeley *et al.*, "Responses of neurons in primary and inferior temporal visual cortices to natural scenes," *Proc. Roy. Soc. London B, Biol. Sci.*, vol. 264, no. 1389, pp. 1775–1783, 1997.
- [65] E. T. Jaynes, "Information theory and statistical mechanics," *Phys. Rev.*, vol. 106, no. 4, pp. 620–630, 1957.
- [66] E. T. Jaynes, "Information theory and statistical mechanics. II," *Phys. Rev.*, vol. 108, no. 2, pp. 171–190, 1957.
- [67] S. Laughlin, "A simple coding procedure enhances a neuron's information capacity," *Zeitschrift Naturforschung C*, vol. 36, pp. 910–912, Jun. 1981.
- [68] W. B. Levy and R. A. Baxter, "Energy efficient neural codes," *Neural Comput.*, vol. 8, no. 3, pp. 531–543, 1996.
- [69] J. Triesch, "Synergies between intrinsic and synaptic plasticity in individual model neurons," in *Proc. Adv. Neural Inf. Process. Syst.*, 2004, pp. 1417–1424.
- [70] C. Savin, P. Joshi, and J. Triesch, "Independent component analysis in spiking neurons," *PLoS Comput. Biol.*, vol. 6, no. 4, 2010, Art. ID e1000757.
- [71] C. Li, "A model of neuronal intrinsic plasticity," *IEEE Trans. Auton. Mental Develop.*, vol. 3, no. 4, pp. 277–284, Dec. 2011.
- [72] C. Rosenmund, J. D. Clements, and G. L. Westbrook, "Nonuniform probability of glutamate release at a hippocampal synapse," *Science*, vol. 262, no. 5134, pp. 754–757, 1993.
- [73] N. A. Hessler, A. M. Shirke, and R. Malinow, "The probability of transmitter release at a mammalian central synapse," *Nature*, vol. 366, no. 6455, pp. 569–572, 1993.
- [74] C. Allen and C. F. Stevens, "An evaluation of causes for unreliability of synaptic transmission," *Proc. Nat. Acad. Sci. USA*, vol. 91, no. 22, pp. 10380–10383, 1994.
- [75] T. Branco and K. Staras, "The probability of neurotransmitter release: Variability and feedback control at single synapses," *Nature Rev. Neurosci.*, vol. 10, no. 5, pp. 373–383, 2009.
- [76] C. Li and Q. Zheng, "Synchronization of the small-world neuronal network with unreliable synapses," *Phys. Biol.*, vol. 7, no. 3, 2010, Art. ID 036010.
- [77] D. Guo and C. Li, "Signal propagation in feedforward neuronal networks with unreliable synapses," *J. Comput. Neurosci.*, vol. 30, no. 3, pp. 567–587, 2011.
- [78] A. J. Bell and T. J. Sejnowski, "An information-maximization approach to blind separation and blind deconvolution," *Neural Comput.*, vol. 7, no. 6, pp. 1129–1159, 1995.
- [79] W. Gerstner and W. M. Kistler, *Spiking Neuron Models: Single Neurons, Populations, Plasticity*. New York, NY, USA: Cambridge Univ. Press, 2002.
- [80] P. Dayan and L. E. Abbott, *Theoretical Neuroscience: Computational and Mathematical Modeling of Neural Systems*. Cambridge, MA, USA: MIT Press, 2001.
- [81] C. Savin and J. Triesch, "ICA with spiking neurons?" in *Proc. Comput. Syst. Neurosci.*, Salt Lake City, UT, USA, 2008.
- [82] C. Li and Y. Li, "A spike-based model of neuronal intrinsic plasticity," *IEEE Trans. Auton. Mental Develop.*, vol. 5, no. 1, pp. 62–73, Mar. 2013.
- [83] M. Díaz-Quesada and M. Maravall, "Intrinsic mechanisms for adaptive gain rescaling in barrel cortex," *J. Neurosci.*, vol. 28, no. 3, pp. 696–710, 2008.
- [84] M. Maravall, R. S. Petersen, A. L. Fairhall, E. Arabzadeh, and M. E. Diamond, "Shifts in coding properties and maintenance of information transmission during adaptation in barrel cortex," *PLoS Biol.*, vol. 5, no. 2, 2007, Art. ID e19.
- [85] P. Földiak, "Forming sparse representations by local anti-Hebbian learning," *Biol. Cybern.*, vol. 64, no. 2, pp. 165–170, 1990.
- [86] C. Weber and J. Triesch, "A sparse generative model of V1 simple cells with intrinsic plasticity," *Neural Comput.*, vol. 20, no. 5, pp. 1261–1284, 2008.
- [87] P. D. King, J. Zylberberg, and M. R. DeWeese, "Inhibitory interneurons decorrelate excitatory cells to drive sparse code formation in a spiking model of V1," *J. Neurosci.*, vol. 33, no. 13, pp. 5475–5485, 2013.
- [88] A. Lazar, G. Pipa, and J. Triesch, "SORN: A self-organizing recurrent neural network," *Frontiers Comput. Neurosci.*, vol. 3, Oct. 2009, Art. ID 23.
- [89] A. Lazar, G. Pipa, and J. Triesch, "Fading memory and time series prediction in recurrent networks with different forms of plasticity," *Neural Netw.*, vol. 20, no. 3, pp. 312–322, 2007.
- [90] D. Marković and C. Gros, "Intrinsic adaptation in autonomous recurrent neural networks," *Neural Comput.*, vol. 24, no. 2, pp. 523–540, 2012.
- [91] B. Siri, M. Quoy, B. Delord, B. Cessac, and H. Berry, "Effects of hebbian learning on the dynamics and structure of random networks with inhibitory and excitatory neurons," *J. Physiol.-Paris*, vol. 101, nos. 1–3, pp. 136–148, 2007.
- [92] H. Williams and J. Noble, "Homeostatic plasticity improves signal propagation in continuous-time recurrent neural networks," *Biosystems*, vol. 87, no. 2, pp. 252–259, 2007.
- [93] H. Williams, "Homeostatic plasticity in recurrent neural networks," in *Proc. 8th Int. Conf. Simulation Adapt. Behavior*, 2004, pp. 344–353.
- [94] K. B. Hengen, M. E. Lambo, S. D. Van Hooser, D. B. Katz, and G. G. Turrigiano, "Firing rate homeostasis in visual cortex of freely behaving rodents," *Neuron*, vol. 80, no. 2, pp. 335–342, 2013.
- [95] K. Whalley, "Synaptic plasticity: Balancing firing rates *in vivo*," *Nat. Rev. Neurosci.*, vol. 14, no. 12, pp. 820–821, 2013.
- [96] G. Rainer and E. K. Miller, "Effects of visual experience on the representation of objects in the prefrontal cortex," *Neuron*, vol. 27, no. 1, pp. 179–189, 2000.
- [97] D. J. Freedman, M. Riesenhuber, T. Poggio, and E. K. Miller, "Experience-dependent sharpening of visual shape selectivity in inferior temporal cortex," *Cerebral Cortex*, vol. 16, no. 11, pp. 1631–1644, 2006.
- [98] D. J. Freedman, "Familiarity breeds plasticity: Distinct effects of experience on putative excitatory and inhibitory neurons in inferior temporal cortex," *Neuron*, vol. 74, no. 1, pp. 8–11, 2012.
- [99] L. Woloszyn and D. L. Sheinberg, "Effects of long-term visual experience on responses of distinct classes of single units in inferior temporal cortex," *Neuron*, vol. 74, no. 1, pp. 193–205, 2012.
- [100] J. J. Steil, "Online reservoir adaptation by intrinsic plasticity for backpropagation-decorrelation and echo state learning," *Neural Netw.*, vol. 20, no. 3, pp. 353–364, 2007.
- [101] B. Schrauwen, M. Wardermann, D. Verstraeten, J. J. Steil, and D. Stroobandt, "Improving reservoirs using intrinsic plasticity," *Neurocomputing*, vol. 71, no. 7, pp. 1159–1171, 2008.
- [102] K. Neumann and J. J. Steil, "Batch intrinsic plasticity for extreme learning machines," in *Artificial Neural Networks and Machine Learning*. New York, NY, USA: Springer, 2011, pp. 339–346.
- [103] Y. Li and C. Li, "Synergies between intrinsic and synaptic plasticity based on information theoretic learning," *PLoS One*, vol. 8, no. 5, 2013, Art. ID e62894.
- [104] J. C. Principe, *Information Theoretic Learning: Renyi's Entropy and Kernel Perspectives*. New York, NY, USA: Springer-Verlag, 2010.
- [105] D. Erdogmus and J. C. Principe, "An error-entropy minimization algorithm for supervised training of nonlinear adaptive systems," *IEEE Trans. Signal Process.*, vol. 50, no. 7, pp. 1780–1786, Jul. 2002.
- [106] D. Erdogmus and J. C. Principe, "Comparison of entropy and mean square error criteria in adaptive system training using higher order statistics," in *Proc. ICA*, 2000, pp. 75–80.
- [107] S. Huang, C. Li, and Y. Liu, "Complex-valued filtering based on the minimization of complex-error entropy," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 24, no. 5, pp. 695–708, May 2013.
- [108] C. Li, P. Shen, Y. Liu, and Z. Zhang, "Diffusion information theoretic learning for distributed estimation over network," *IEEE Trans. Signal Process.*, vol. 61, no. 16, pp. 4011–4024, Aug. 2013.
- [109] P. Shen and C. Li, "Minimum total error entropy method for parameter estimation," *IEEE Trans. Signal Process.*, vol. 63, no. 15, pp. 4079–4090, Aug. 2015.
- [110] E. Parzen, "On estimation of a probability density function and mode," *Ann. Math. Statist.*, vol. 33, no. 3, pp. 1065–1076, 1962.
- [111] J. C. Principe, N. R. Euliano, and W. C. Lefebvre, *Neural and Adaptive Systems: Fundamentals Through Simulations With CD-ROM*, 1st ed. New York, NY, USA: Wiley, 1999.



CHUNGUANG LI (M'14–SM'14) received the M.S. degree in pattern recognition and intelligent systems and the Ph.D. degree in circuits and systems from the University of Electronic Science and Technology of China, Chengdu, China, in 2002 and 2004, respectively.

He is currently a Professor with the College of Information Science and Electronic Engineering, Zhejiang University, Hangzhou, China. His current research interests include statistical signal processing, machine learning, and computational neuroscience.



YUKE LI received the B.S. degree in electronics and information engineering from Jilin University, Changchun, China, in 2010, and the Ph.D. degree in electronics engineering from Zhejiang University, Hangzhou, China, in 2015. He is currently with the Alibaba Group. His current research interests include machine learning, computational neuroscience, and computer vision.

• • •