Optical spike-timing-dependent plasticity with weight-dependent learning window and reward modulation

Quansheng Ren, Yaolin Zhang, Rui Wang, and Jianye Zhao*

School of Electronics Engineering and Computer Science, Peking University, Beijing 100871, China

**zhaojianye@pku.edu.cn*

Abstract: Optical spike-timing-dependent plasticity (STDP) synapses form the basis of learning in photonic neuromorphic system. In biological neural systems, STDP synapses generally have multiplicative boundary mechanisms, and can be modulated by a third factor such as dopamine. Analogously, we introduce a third factor into optical STDP: The current-injection of semiconductor optical amplifiers can be modified in an adaptive way according to local or global feedback signals. The local one is present synaptic weight, which elicits an optical weight-dependent STDP, while the global one is a reward signal. We demonstrate that the optical weight-dependent STDP can emulate the behavior of biological STDP synapses more closely, and can be seen as an intermediate configuration between additive and multiplicative STDP, which balances stability and competition among synapses. Simulation studies with scalable photonic neurons further show that optical STDP with reward modulation enables reward-based reinforcement learning.

OCIS codes: (070.4340) Nonlinear optical signal processing; (320.7085) Ultrafast information processing; (200.4260) Neural networks; (200.4700) Optical neural systems.

References and links

1. D. Monroe, "Neuromorphic computing gets ready for the (really) big time," Commun. ACM **57**, 13 (2014).
2. S. Song, K. D. Miller, and L. F. Abbott, "Competitive hebbian learning through spike-timing-dependent synaptic plasticity," Nat. Neurosci. **3**, 919–926 (2000).
3. P. J. Sjöström, E. A. Rancz, A. Roth, and M. Häusser, "Dendritic excitability and synaptic plasticity," Physiol. Rev. **88**, 769–840 (2008).
4. Q. Ren, K. M. Kolwankar, A. Samal, and J. Jost, "Stdp-driven networks and the C. elegans neuronal network," Physica A **389**, 3900–3914 (2010).
5. S. H. Jo, T. Chang, I. Ebong, B. B. Bhadviya, P. Mazumder, and W. Lu, "Nanoscale memristor device as synapse in neuromorphic systems," Nano Lett. **10**, 1297–1301 (2010).
6. A. N. Tait, M. A. Nahmias, Y. Tian, B. J. Shastri, and P. R. Prucnal, "Photonic Neuromorphic Signal Processing and Computing," in *Nanophotonic Information Physics: Nanointelligence and Nanophotonic Computing*, (Springer-Verlag, 2014).
7. S. Barbay, R. Kuszelewicz, and A. M. Yacomotti, "Excitability in a semiconductor laser with saturable absorber," Opt. Lett. **36**, 4476–4478 (2011).
8. F. Selmi, R. Braive, G. Beaudoin, I. Sagnes, R. Kuszelewicz, and S. Barbay, "Relative refractory period in an excitable semiconductor laser," Phys. Rev. Lett. **112**, 183902 (2014).
9. B. Romeira, J. Javaloyes, C. N. Ironside, J. M. L. Figueiredo, S. Balle, and O. Piro, "Excitability and optical pulse generation in semiconductor lasers driven by resonant tunneling diode photo-detectors," Opt. Express **21**, 20931–20940 (2013).

#246680 Received 24 Jul 2015; revised 11 Sep 2015; accepted 14 Sep 2015; published 17 Sep 2015

10. K. Kravtsov, M. P. Fok, D. Rosenbluth, and P. R. Prucnal, "Ultrafast all-optical implementation of a leaky integrate-and-fire neuron," Opt. Express **19**, 2133–2147 (2011).
11. M. A. Nahmias, B. J. Shastri, A. N. Tait, and P. R. Prucnal, "A leaky integrate-and-fire laser neuron for ultrafast cognitive computing," IEEE J. Sel. Top. Quantum Electron. **19**, 1800212 (2013).
12. M. P. Fok, Y. Tian, D. Rosenbluth, and P. R. Prucnal, "Pulse lead/lag timing detection for adaptive feedback and control based on optical spike-timing-dependent plasticity," Opt. Lett. **38**, 419–421 (2013).
13. R. Toole and M. P. Fok, "Photonic Implementation of a Neuronal Learning Algorithm based on Spike Timing Dependent Plasticity," in *Optical Fiber Communication Conference*, (Optical Society of America, 2015), p. W1K.6.
14. E. Izhikevich, "Solving the distal reward problem through linkage of stdp and dopamine signaling," Cereb. Cortex **17**, 2443 (2007).
15. M. A. Farries and A. L. Fairhall, "Reinforcement learning with modulated spike timing–dependent synaptic plasticity," J. Neurophysiol. **98**, 3648–3665 (2007).
16. J. Rubin, D. D. Lee, and H. Sompolinsky, "Equilibrium properties of temporally assymmetric hebbian plasticity," Phys. Rev. Lett. **86**, 364–367 (2001).
17. 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. **18**, 10464–10472 (1998).
18. J. Tegner and A. Kepecs, "Why neuronal dynamics should control synaptic learning rules," Adv. Neural Inf. Process. Sys. **14**, 285 (2001).
19. S. Friedmann, N. Frmaux, J. Schemmel, W. Gerstner, and K. Meier, "Reward-based learning under hardware constraints - using a risc processor embedded in a neuromorphic substrate," Front. Neurosci. **7**, 160 (2013).
20. M. Gilson and T. Fukai, "Stability versus neuronal specialization for stdp: Long-tail weight distributions solve the dilemma," PLoS One **6**, e25339 (2011).
21. F. Ginovart, J. C. Simon, and I. Valiente, "Gain recovery dynamics in semiconductor optical amplifier," Opt. Commun. **199**, 111–115 (2001).
22. B. J. Shastri, M. A. Nahmias, A. N. Tait, and P. R. Prucnal, "Simulations of a graphene excitable laser for spike processing," Opt. Quantum Electron. **46**, 1353–1358 (2014).
23. A. W. Fang, "A distributed feedback silicon evanescent laser," Opt. Express **16**, 4413–4419 (2008).
24. H. Park, A. W. Fang, R. Jones, O. Cohen, O. Raday, M. N. Sysak, M. J. Paniccia, and B. Je., "A hybrid algainas-silicon evanescent waveguide photodetector," Opt. Express **15**, 6044–6052 (2007).
25. P. J. Sjöström and W. Gerstner, "Spike-Timing Dependent Plasticity," Scholarpedia **5**, 1362 (2010).
26. J. K. Seamans and C. R. Yang, "The principal features and mechanisms of dopamine modulation in the prefrontal cortex," Prog. Neurobiol. **74**, 1–58 (2004).
27. H. S. Seung, "Learning in spiking neural networks by reinforcement of stochastic synaptic transmission," Neuron **40**, 1063–1073 (2004).
28. C. Eliasmith, T. C. Stewart, X. Choo, T. Bekolay, T. Dewolf, T. Charlie, Y. Tang, and D. Rasmussen, "A large-scale model of the functioning brain," Science **338**, 1202–1205 (2012).

1. Introduction

Neuromorphic systems [1] equipped with spike-timing-dependent plasticity (STDP) synaptic learning [2] have a wide range of applications in the fields of machine learning and adaptive control, and raise the prospect of brain-like artificial intelligence. STDP is a form of adaptive learning that determines the modification of synaptic weights based on temporal correlations between the spikes of pre- and postsynaptic neurons. It plays an important role in learning and memory in the brain, and underlies in the development and refinement of neuronal networks during development [3, 4]. Its learning dynamics can be artificially emulated by the nonlinear properties of certain electronic devices (eg. nanoscale memristor [5]), which has become a crucial factor for the rapidly growing field of neuromorphic engineering. Although electronics can exceed biological time scales, their computing speed is limited due to a bandwidth fan-in trade-off [6]. In recent years, several efforts shed light on the analogy between optical and biological neurons and pave the way to fast spike-time coding based optical systems with a speed several orders of magnitude faster than their biological or electronic counterparts [7–9]: The group of Barbay reported that semiconductor lasers with saturable absorber can exhibit subnanosecond neuronlike pulses [7]. Romeira et al. studied the excitability and optical pulse generation in semiconductor lasers driven by resonant tunneling diode photo-detectors experimentally and theoretically [9]. The group of Prucnal developed scalable photonic neurons [6, 10, 11], that

#246680 Received 24 Jul 2015; revised 11 Sep 2015; accepted 14 Sep 2015; published 17 Sep 2015

can operate up to a billion times faster than biological neurons, a speed that electronic neurons could never reach, and will be more energy efficient than electronic neurons in future. To equip photonic neurons with synaptic learning, Fok et al. proposed an optical STDP scheme [12, 13], where the fast physical mechanisms of a semiconductor optical amplifier (SOA) are utilized to determine the amount of weight change.

In biological neural systems, STDP synapses generally have boundary mechanisms to prevent synaptic weight from becoming too small or too large, and can also be modulated by a third factor such as dopamine, which brings reward-based reinforcement learning [14, 15]. There is a kind of boundary mechanisms for STDP rules. These boundary mechanisms can be categorized in two kinds depending on whether weight updating in STDP is additive or multiplicative [16]. For the additive STDP rule, the amount of weight change is independent on present weight size, thus hard weight bounds must be introduced to stabilize learning. While in the multiplicative scenario, the magnitude of weight increase scales inversely with present synaptic weight so that soft bounds can be achieved. Experimental data from neuroscience is not yet sufficient to support an additive rule, and the available evidence suggests that modifications are dependent on the weight [17]. The optical STDP scheme proposed by Fok et al. still lacks of a boundary mechanism as well as a neuromodulation mechanism. Moreover, in their scheme, the STDP module extracts presynaptic spikes after the artificial synapse, which differs from its biological counterpart and may cause excessive adjusting of synaptic weight. In a recurrent neural network, the unbounded and excessive weight increasing will lead to intensive spiking activity. Studies in computational neuroscience suggest that the additive STDP is simpler for the implementation of learning tasks, but is less robust than the multiplicative STDP [18]. On the other hand, without neuromodulation, optical STDP is not capable of providing reward learning, which connects synaptic plasticity closely to the learning of behaviors and is an indispensible issue for neuromorphic system [19].

Here, we introduce a third factor, or modulatory signal into the original scheme of optical STDP by varying the current-injection of SOAs. The bias current can be modified in an adaptive way according to local or global feedback signals: the local one is present synaptic weight, which elicits an optical weight-dependent STDP, while the global one is a reward signal calculated from difference between target output and actual output, which brings reward-based reinforcement learning. The optical weight-dependent STDP we proposed can be seen as an intermediate configuration between additive and multiplicative STDP [20], which can balance stability and competition among synapses.

2. The implementation of optical STDP with feedback signals

STDP is a sort of experimentally observed long-term synaptic plasticity that modifies the weights of synapses as a function of the precise temporal relations between pre- and postsynaptic spikes: an excitatory synapse is strengthened when the postsynaptic neuron fires shortly after the presynaptic one, and it is weakened when this temporal order is reversed. The amount of modification exponentially depends on the timing of the postsynaptic spike (t_{post}) relative to the presynaptic spike (t_{pre}), which is denoted as Δt ($\Delta t = t_{post} - t_{pre}$). The STDP function experimentally measured by Bi and Poo (1998) [17] and shown in Fig. 1(a) can be fitted to the following mathematical representation:

$$\Delta w(\Delta t) = \begin{cases} A_+ e^{-|\Delta t|/\tau_+} & \text{if } \Delta t > 0 \\ -A_- e^{-|\Delta t|/\tau_-} & \text{else} \end{cases}, \tag{1}$$

where Δw is the magnitude of weight changes, A and τ represent the amplitude and time constant respectively, and subscripts $+$ and $-$ identify the potentiation and depression window

#246680 Received 24 Jul 2015; revised 11 Sep 2015; accepted 14 Sep 2015; published 17 Sep 2015

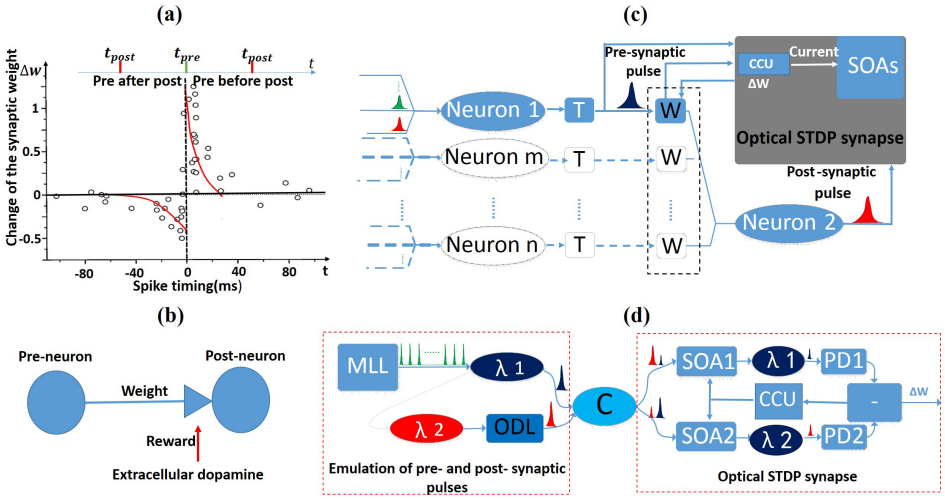

Fig. 1. (a) Experimental data of biological STDP as a function of the relative timing between presynaptic spike arrival and postsynaptic firing, schematically redrawn after Bi and Poo [17]. (b) The dynamics of a STDP synapse connecting two neurons is described by synapse strength (weight), and synaptic plasticity can be influenced by extracellular dopamine. (c) Two photonic neurons are connected with an optical STDP synapse. W - variable weight adjusting device, T - variable delay line, SOA - semiconductor optical amplifier, CCU - current control unit. (d) Optical implementation of weight-dependent STDP. MLL - mode-locked fiber ring laser, λ_1 / λ_2 - optical bandpass filters, ODL - variable optical delay line, PD - photodetector.

respectively. τ_+ and τ_- are 10s of ms for biological neurons. STDP is triggered by nearly coincident firing patterns on a millisecond timescale, while kinetics of synaptic plasticity can be also sensitive to changes in extracellular dopamine concentration, as illustrated in Fig. 1(b).

Figure 1(c) illustrates a basic unit of synaptic learning in photonic neuromorphic system, which consists of two photonic neurons and an optical STDP synapse. The presynaptic neuron receives a set of inputs, each of which is a spike train of picosecond (ps) width pulses. The neuron temporally integrates all input signals. If the integral value exceeds the threshold, an output spike will be formed and transferred to the postsynaptic neuron. Spike trains are simply represented by a sum of Dirac delta pulses centered on the respective spike times:

$$Spike(t) = \sum_{t_k \in S} \delta(t - t_k), \tag{2}$$

where $Spike(t)$ is the time sequence of pre- or postsynaptic spikes, S is the set of the spike times. Before the spike arrives at the postsynaptic neuron, its intensity is influenced by the synapse according to the synaptic weight. Electrical current perturbation $I_j(t)$ on a postsynaptic neuron j is equal to the linear sum of the presynaptic spikes:

$$I_j(t) = \sum_i w_{ji} Spike_i(t), \tag{3}$$

where w_{ji} represents the synaptic strength between pre-neuron i and post-neuron j. Notice that the postsynaptic neuron may also have many inputs from other neurons, thus the relation between presynaptic spikes and postsynaptic spikes may be causal or acausal: the STDP synapse is increased in weight if presynaptic spikes repeatedly occur before postsynaptic spikes within

#246680 Received 24 Jul 2015; revised 11 Sep 2015; accepted 14 Sep 2015; published 17 Sep 2015

a few hundred picoseconds (10s of ms for biological neurons), whereas the opposite temporal order elicits synaptic weakening. In this way, unsupervised learning is realized as following: the synapse is rewarded if its activity consistently causes or predicts the postsynaptic neuron's activity, while it is punished for repeated failure at predicting. In Fig. 1(c), the current-injection of SOAs is determined by the local feedback signal, i.e., the value of synaptic weight, through the current control unit (CCU), and simultaneously, the output power of the SOAs determines the resultant change in synaptic strength. Thus, the mutual feedback between the SOAs and synaptic weight is utilized to implement optical weight-dependent STDP. The current-injection of SOAs can be also determined by the global feedback signal, i.e., a reward signal calculated from difference between target output and actual output, to implement reward-based reinforcement learning, which is shown in Fig. 4(a).

Experimental implementation of our optical STDP scheme is illustrated in Fig. 1(d). A mode-locked fiber ring laser is used as the pulse source, and its output light is sliced with two bandpass filters. The resulting two ultrashort pulsed beams of wavelengths $\lambda_1 = 1550.1$ nm and $\lambda_2 = 1554.9$ nm are utilized as the pre- and postsynaptic spikes respectively. The relative time delay between the two beams is controlled by a variable optical delay line. The pre- and postsynaptic spikes are split by a 90/10 coupler such that larger portions of the former and the latter are injected respectively into the potentiation and the depression modules as the signal spikes of the SOAs, while smaller portions of the latter and the former are injected respectively into the potentiation and the depression modules as the probe spikes.

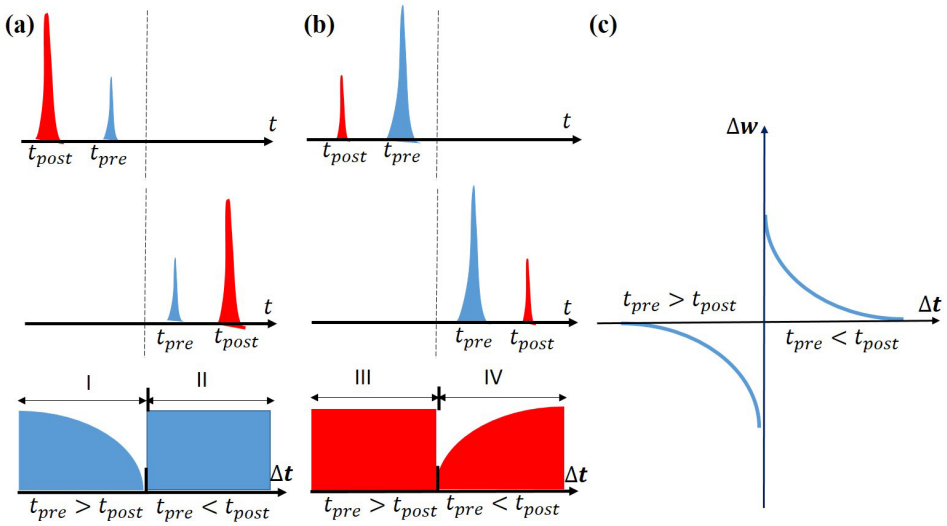

Fig. 2. (a)Formation of depression window. The blue and red colored pulses denote pre- and postsynaptic pulses respectively. The gain response colored in blue represents the light intensity of presynaptic pulse (in blue) detected by PD1 in Fig. 1(d). (b)Formation of potentiation window. The gain response colored in red represents the light intensity of postsynaptic pulse (in red) detected by PD2 in Fig. 1(d). (c)A linear subtraction of (a) and (b) results in STDP characteristic.

The potentiation and the depression modules have the same structure. The key device is an SOA (Kamelian SOA-NL-L1-C-FA), and its gain-recovery dynamics after saturation plays the role of the STDP learning window: Before the signal pulse enters the SOA, the carrier density corresponding to the SOA gain is kept at a constant value by the current-injection. After the pulse enters the SOA, the carrier density decreases due to stimulated emission. If the pulse

#246680 Received 24 Jul 2015; revised 11 Sep 2015; accepted 14 Sep 2015; published 17 Sep 2015

energy is enough, it will lead to a significant depletion of carriers. This process takes about a few picoseconds and is known as gain saturation. After the pulse has passed, the carrier density as well as the SOA gain begins to recover due to the injection of the carriers by the current I. The recovery time is typically several hundred picoseconds. The gain-recovery dynamics can be sampled by the probe pulse, whose energy is about one tenth of the signal pulse. Specifically, the probe pulse is extracted from the SOA output by using an optical bandpass filter, and then detected by a photodetector.

Figures 2(a) and 2(b) illustrate the mechanism for optical STDP. If the probe pulse enters the SOA before the signal pulse(region II and III), it is amplified by the SOA normally and not influenced by any carrier depletion. In this case, the output power remains a constant value P_{const}, which is used as a reference value. If the probe pulse enters the SOA after the signal pulse within several hundred picoseconds(region I and IV), it experiences the carrier depletion and its output power P_s is weaker than the normal case. The loss of the output power ΔP ($\Delta P = P_{const} - P_s$) is determined by the gain-recovery dynamics of SOA as well as the time interval between the signal pulse and the probe pulse. ΔP corresponds to the magnitude of weight increase ($\Delta w > 0$) or weight decrease ($\Delta w < 0$) in the potentiation or the depression modules respectively:

$$\Delta w(\Delta t, I) = \begin{cases} (P_{const} - P_s(\Delta t, I))/P_{const} & \text{if } \Delta t > 0 \\ (P_s(\Delta t, I) - P_{const})/P_{const} & \text{else} \end{cases}, \tag{4}$$

which implies a linear subtraction of the output power of both channels can get the final STDP curve, as illustrated in Fig. 2(c). $P_s(\Delta t, I)$ has a positive correlation with the carrier density ρ after saturation, which can be represented as a function of Δt and I, i.e., $\rho = \rho(\Delta t, I)$, determined by several non-linear partial differential coupled equations [21]. Experimental data of Eq. (4), as shown in Fig. 3(a), are in accordance with the experimentally and theoretically obtained gain recovery dynamics given in [21]. Especially, the gain-recovery dynamics highly depends on I: a stronger injection current I can expedite the carrier recovery process, which is equivalent to decreasing the height and width of the STDP learning window. As Eq. (1) shows, the STDP learning function of either potentiation or depression module can be determined by two parameters: A and τ, while Fig. 3(a) shows that Eq. (4) is equivalent to modulating A and τ by the injected-current I. Thus, we can utilize I to implement the modulatory signal of the optical STDP. It is modified in an adaptive way according to local feedback signal $I_{p/d}(w)$ or global feedback signals $I_{p/d}(Rwd)$, which can be realized by the CCU. $I_{p/d}(w)$ and $I_{p/d}(Rwd)$ are the modulatory signals responsible for weight-dependent boundary mechanism and reward-based reinforcement learning respectively. Besides, our optical STDP scheme also differs from that proposed in [12] in that the STDP module extracts presynaptic spikes before the weight adjusting device rather than after it, which avoids excessive adjusting of the synaptic weight.

3. Local feedback control and optical weight-dependent STDP

The local feedback signal is present synaptic weight w, which is obtained by integrating the weight change Δw and restricting normalized w within the interval $[0,1]$ for avoiding too large I. A variable attenuator is used as weight adjusting device, and the maximum and minimum settings of the attenuator correspond to hard bounds of weight updating. The positive and negative feedbacks $I_p(w) = 40mA + \alpha w$ and $I_d(w) = 40mA + \alpha(1-w)$ are implemented for the potentiation and the depression modules respectively, where $\alpha = 160mA$ is a scale factor. By this means, an optical weight-dependent STDP can be achieved. In our experiment, the repetition rate of the mode-locked fiber ring laser is 100 MHz. Therefore, the period of the pre- and postsynaptic spike pairs is about 10 ns, which is much larger than the SOA recovery time (about 25 ps@300mA). The relative time delay Δt between the pre- and postsynaptic spikes

#246680 Received 24 Jul 2015; revised 11 Sep 2015; accepted 14 Sep 2015; published 17 Sep 2015

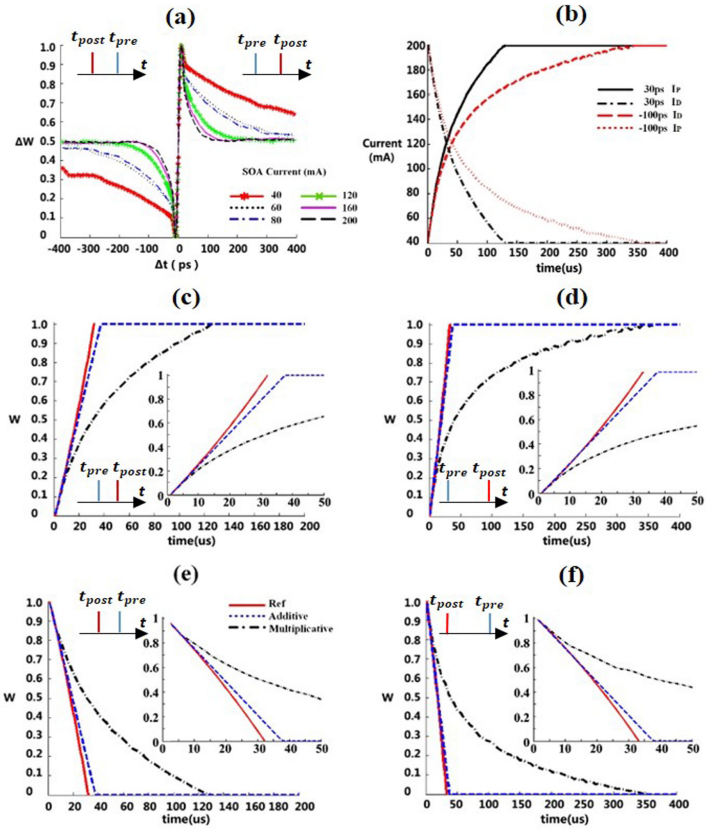

Fig. 3. (a) The measured learning window of Optical STDP with different SOA driving current. (b) The evolution of the SOA currents I_p and I_d for the potentiation and depression modules respectively in the cases of $\Delta t = 30ps$ and $\Delta t = -100ps$. (c) The evolution of the synaptic weight for the case of $\Delta t = 30$ ps, where presynaptic spikes repeatedly occur before postsynaptic spikes for 30 ps. Three different optical STDP schemes are studied: the optical STDP scheme like the one proposed in [12] (without boundary mechanism, full line), our optical STDP scheme without the feedback control of the SOA current (additive STDP, dashed line), and our optical STDP scheme with the feedback control of the SOA current (multiplicative STDP, dashdotted line). (d) The evolution of the synaptic weight for the case of $\Delta t = 100$ ps. (e) The evolution of the synaptic weight for the case of $\Delta t = -30ps$, where presynaptic spikes repeatedly occur after postsynaptic spikes for 30 ps. (f) The evolution of the synaptic weight for the case of $\Delta t = -100ps$.

can be adjusted within an interval [-400 ps, 400 ps]. Figure 3(a) shows the experimentally measured STDP learning window, which expresses the change of synaptic weight as a function of the relative timing of pre- and postsynaptic spikes, under different values of SOA current. As expected, the height and width of the STDP learning window decrease as the SOA current increases.

If repeated presynaptic spikes arrive before postsynaptic spikes within a few hundred picoseconds, the synaptic weight will continuously increase. In this case, the boundary mechanism determines how the synaptic weight approaches to the maximum value. Figure 3(c) shows the evolution of the synaptic weight for the case of $\Delta t = 30$ ps. Three different optical STDP

#246680 Received 24 Jul 2015; revised 11 Sep 2015; accepted 14 Sep 2015; published 17 Sep 2015

Table 1. Two-section DFB hybrid evanescant laser parameters

Param.	Value	Param.	Value
β	1×10^{-4}	η_c	0.4
λ	$850nm$	B_r	$1\times10^{-15}m^3s^{-1}$
V_a	$2.4\times10^{-18}m^3$	V_s	$2.4\times10^{-18}m^3$
Γ_a	0.06	Γ_s	0.05
τ_a	$1ns$	τ_s	$100ps$
g_a	$2.9\times10^{-12}m^3s^{-1}$	g_s	$14.5\times10^{-12}m^3s^{-1}$
n_{0a}	$1.1\times10^{24}m^{-3}$	n_{0s}	$0.89\times10^{24}m^{-3}$
τ_{ph}	$4.8ps$		

schemes are studied. When the optical STDP scheme like the one proposed in [12] is utilized, accelerated growth of the synaptic weight is observed, due to the absence of a boundary mechanism and the design of extracting presynaptic spikes after the weight adjusting device. When our optical STDP scheme is utilized without the feedback control of the SOA current, the synaptic weight grows at a constant rate until the maximum bound is reached, and then maintains at the maximum bound, which resembles the additive STDP characteristic with a hard bound. In the case where the feedback control of the SOA current is utilized, the synaptic weight grows more and more slowly when it approaches the maximum value, which coincides with the multiplicative STDP characteristic with a soft bound. When the synaptic weight reaches the maximum bound at last, it also maintains at the maximum bound. Thus, our scheme can be seen as an intermediate configuration between additive and multiplicative STDP [20], which can balance stability and competition among synapses. Figure 3(d) illustrates the evolution of the synaptic weight for the case of $\Delta t = 100$ ps. Since the relative time delay Δt between the pre- and postsynaptic spikes is larger compared with the previous case, the potentiation effect becomes weaker and the increasing of the synaptic weight becomes slower. On the other hand, if repeated presynaptic spikes arrive after postsynaptic spikes within a few hundred picoseconds, the synaptic weight will continuously decrease. Similar behaviors are shown in Figs. 3(e) and 3(f).

Figure 3(b) shows the evolution of the measured SOA currents in the potentiation and depression modules. In the case of $\Delta t = 30$ ps, during the synaptic weight approaching the maximum weight, the SOA current of the potentiation module I_p increases gradually to the preset maximum current, while the SOA current of the depression module I_d decreases gradually to the preset minimum current. This means when the synaptic weight approaches the maximum value, it is increasingly difficult for the potentiation, and is increasingly easy for the depression, which also prevents the synaptic weight to approach the maximum value. In the case of $\Delta t = -100$ ps, during the synaptic weight approaching the minimum weight, the SOA current of the potentiation module I_p decreases gradually to the preset minimum current, while the SOA current of the depression module I_d increases gradually to the preset maximum current. Thus, when the synaptic weight approaches the minimum value, it is increasingly difficult for the depression, and is increasingly easy for the potentiation, which also prevents the synaptic weight to approach the minimum value.

4. Global feedback control and optical weight-dependent STDP

In our optical STDP scheme, the global feedback signal is a reward signal *Rwd* calculated from difference between target output and actual output. To check the feasibility of reward learning using our optical STDP scheme, we examine it on the benchmark of reinforcement learning proposed by Farries and Fairhall in 2007 [15]. At the present stage, optical neuromorphic network on chip is unavailable, thus we numerically study a network consisted of 1,000 input units and a single output scalable photonic neuron proposed by Prucnal's group [11], which is

#246680 Received 24 Jul 2015; revised 11 Sep 2015; accepted 14 Sep 2015; published 17 Sep 2015

shown in Fig. 4(a). We use Matlab as the simulation tool. The spiking neuron model used in the simulation is based on a two-section DFB hybrid evanescant laser, proposed by Mitchell and Alexander in 2013 [22]. To simulate the device, a standard two-section laser rate equation is used:

$$\frac{dN_{ph}}{dt} = \Gamma_a g_a (n_a - n_{0a}) N_{ph} + \Gamma_s g_s (n_s - n_{0s}) N_{ph} - \frac{N_{ph}}{\tau_{ph}} + V_a \beta B_r n_a^2 \tag{5}$$

$$\frac{dn_a}{dt} = -\Gamma_a g_a (n_a - n_{0a}) \frac{N_{ph}}{V_a} - \frac{n_a}{\tau_a} + \frac{I_a + i_e(t)}{eV_a} \tag{6}$$

$$\frac{dn_s}{dt} = -\Gamma_s g_s (n_s - n_{0s}) \frac{N_{ph}}{V_s} - \frac{n_s}{\tau_s} + \frac{I_s}{eV_s}, \tag{7}$$

where $N_{ph}(t)$ is the total number of photons in the cavity, $n(t)$ is the number of carriers, Γ is the confinement factor, g is the differnential gain/loss, n_0 is the transparency carrier density, τ_p is the photon lifetime, V is the cavity volume, β is the spontaneous emission coupling factor, B_r is the Bimolecular recombination term, and $i_e(t)$ represents the electrical modulation in the gain provided by the photodetector system, which has been defined in Eq. (3). Subscripts a and s identify the active and absorber regions. We set the input currents to I_a = 2 mA and I_s = 0 mA for the gain and absorber regions. The output power of the laser is:

$$P_{out}(t) \approx \frac{\eta_c \Gamma_a}{\tau_p} \frac{hc}{\lambda} N_{ph}(t), \tag{8}$$

where η_c is the output power coupling coefficient, c is the speed of light, and λ is the wavelength of a single photon. The remaining realistic parameters can be found in [22–24] and summarized in Table 1. The simulation results of the scalable photonic neuron are shown in Fig. 4(b).

The weights of 1,000 synapses are modified according to the experimental data of our optical STDP obtained by the procedure mentioned above and shown in Fig. 3(a). Initial synaptic weights are chosen from a gaussian distribution. Simulations are composed of trials lasting 1000 ps. One half of the input units represent random background inputs, which are governed by Poisson processes at a rate of 5 GHz. For the remaining input units, in the first 100 ps and final 100 ps of a trial, they keep silent, but in the median time, they are governed by 800-ps spike trains truncated from independent Poisson spike trains with 5 GHz. Within a given simulation, these spike trains do not change across trials, but the actual time of each spike is randomly disturbed by a gaussian with a SD of 10 ps centered on each spike in each train. The actual and target output spike trains in the current trial ($Spike_{actual}(t)$ and $Spike_{target}(t)$) are smoothed by convolving the spike trains with a gaussian of unit height and SD 10 ps ($\mu = 0$, $\sigma = 10$), which can be regarded as a 'filtering process'. Their difference is regarded as the reward signal $Rwd(t) \in [0,1]$, which is a function of simulation time and can be mathematically represented as Eq. (9) shows:

$$Rwd(t) = Spike_{actual}(t) - Spike_{target}(t) = \Big(\sum_{t_k \in S_{actual}} \delta(t - t_k) - \sum_{t_k \in S_{target}} \delta(t - t_k)\Big) * e^{-\frac{(t-\mu)^2}{2\sigma^2}}, \tag{9}$$

where * is denoted as convolution, the maximum value of the $Rwd(t)$ represents the amplitude of spiking pulses. Then a function $temp(t) = e^{-3|Rwd(t)|}$ is introduced [15]. We further utilize an adaptation of the temporal difference algorithm for reinforcement learning, where the current-injection I is driven by $Rwd(t)$ according to the symbol of the difference $\Delta temp$ between $temp(t)$ and a running average (conducted over trials) of $< temp(t) >$ [15]: $\Delta temp = temp - < temp >$. If $\Delta temp > 0$, the depression module and the potentiation module works as normal,

#246680 Received 24 Jul 2015; revised 11 Sep 2015; accepted 14 Sep 2015; published 17 Sep 2015

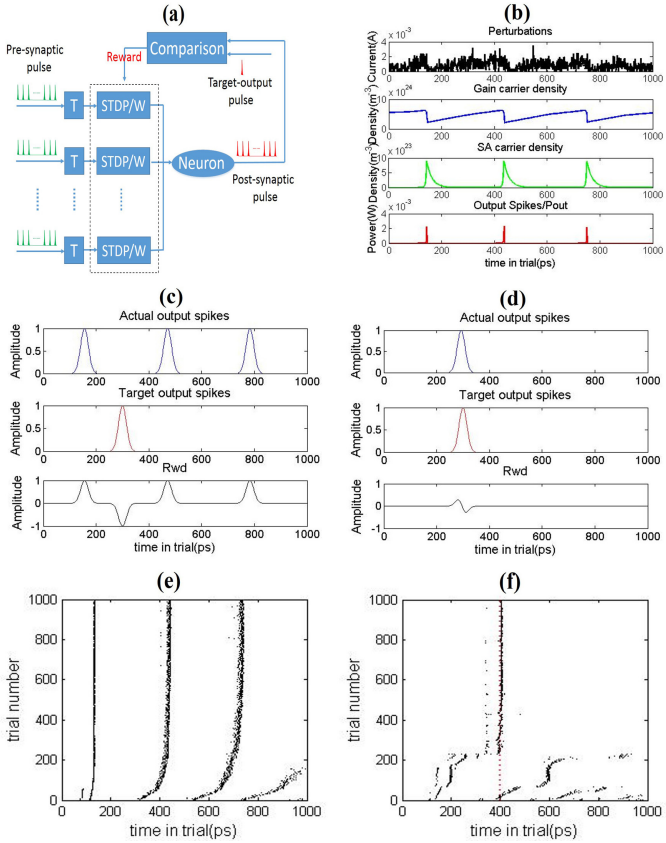

Fig. 4. (a) Structure of the network.The STDP/W module - As the STDP module shown in Fig. 1(c), it adjusts synaptic weight of each input neuron according to the present current value determined by *Rwd*. (b)A simulation of a DFB laser neuron with realistic parameters. Current perturbations from the input spikes (in black) modulate the gain (in blue). Enough excitatory activity leads to the saturation of the absorber to transparency (in green) and causes the release of a pulse (in red), followed by a short refractory period while the pump current recovers the carrier concentration back to its equilibrium value. (c)The simulation result of reward signal *Rwd* representing differences between the actual and desired output spikes when the target output is a single spike fired at 300ps. (d)The simulation result of reward signal *Rwd* with reinforcement learning completed. (e) Raster showing how response of output neuron changes under influence of the optical additive STDP without reward modulation. (f) Raster showing how the response of output neuron changes during reward learning when the target output is a single spike fired at 400ps.

and the control currents of both modules share the fomula: $I_{p/d}(Rwd) = 40mA + 40Rwd$. But if $\Delta temp < 0$, it should become increasingly easy for the depression, i.e., the control current $I_d(Rwd)$ should become smaller than normal, thus the formula of the bias current modulation $I_d(Rwd) = 80mA - 40Rwd$ is utilized; while it should become increasingly difficult for the potentiation, i.e., the control current $I_p(Rwd)$ should become larger than normal, thus the formula of the bias current modulation $I_p(Rwd) = 40mA + 120Rwd$ is adopted.

The goal of the above reward learning is to achieve a basic implementation of STDP-driven reinforcement learning using a simplest network structure, specifically, to explore the possibil-

#246680 Received 24 Jul 2015; revised 11 Sep 2015; accepted 14 Sep 2015; published 17 Sep 2015

ity of using our optical STDP to train a postsynaptic neuron to fire at specific time. Fig. 4(c) shows the corresponding simulation result of $Rwd(t)$ in learning period. The goal of reinforcement learning is to ensure the spiking-time of actual output spikes can match the desired ones, i.e., when the reinforcement is completed, *Rwd* will approach to zero, which has no further influence on synaptic strengths between pre- and post-neurons through the STDP rule. The 'filtering process' can ignore the tiny time gap between actual and desired output, which can avoid over-learning of synaptic weights. As shown in Fig. 4(d), when the actual output almost matchs the desired one, the reinforcement learning is also completed. For comparison, the effect of the optical STDP without reward modulation is shown in Fig. 4(e). The performance of the optical STDP with reward modulation is shown in Fig. 4(f). Initially, the network fires at fairly regular intervals with a mean rate of 3 GHz. The target output is a single spike fired at 400 ps of the trial. Under reward learning, the output neuron become to reliably fire at 400 ps and keep silent during most other time, which proves that the reward-based learning has been achieved under our current control mechanism with global feedback.

5. Conclusion

Stability is important in computational neuroscience, and reward-based reinforcement learning connecting synaptic plasticity closely to the learning of behaviors is also an indispensible issue for neuromorphic system. In this paper, we have proposed an optical STDP scheme with weight-dependent learning window and reward modulation. The gain-recovery dynamics of the SOA after saturation, highly depending on the current-injection, has been further utilized to establish a third factor, or modulatory signal for optical STDP. The semiconductor bias current is modified in an adaptive way according to the present synaptic weight or the reward signal. Our paper advances the design of optical STDP in the following four aspects:

First, we introduce a boundary mechanism for optical STDP synapses. Experimental data from neuroscience is not yet sufficient to support an additive rule, and the available evidence suggests that modifications are dependent on the weight [17]. The weight-dependent STDP we proposed can be seen as an intermediate configuration between additive and multiplicative STDP, which balances stability and competition among synapses.

Second, our STDP module extracts presynaptic spikes before the artificial synapses, which is more close to the biological counterpart. The biochemical mechanism of STDP depends on intracellular calcium transients [3, 25]: The release of neurotransmitters, such as glutamate, is triggered by the arrival of an action potential in the axon terminal. Then, postsynaptic NMDA receptors residing in the spine detect the coincidence of glutamate release due to the presynaptic spike and depolarization due to the postsynaptic spike, which results in a change of postsynaptic calcium. Downstream to the calcium influx is calmodulin, which can be utilized to distinguish between LTP and LTD-promoting calcium signals. Eventually, a kind of calcium/calmodulin-dependent enzyme is affected by the calcium transient, and has been hypothesized to encode synaptic weight. The temporal and causal sequence of this process indicates that, the NMDA-receptor-based coincidence deterction is anterior to the synaptic weight changing and encoding. Thus, the STDP module should extract presynaptic spikes before the artificial synapses. Otherwise, if the STDP module extracts presynaptic spikes after the artificial synapse, it will cause excessive adjusting of synaptic weight, as shown in the comparison results in Section 3.

Third, a neuromodulation mechanism is implemented to enable reward-based reinforcement learning. For biological synapses, STDP is triggered by nearly coincident firing patterns on a millisecond timescale, while kinetics of synaptic plasticity is sensitive to changes in extracellular dopamine concentration. Given the complex of the biochemical processes involved in STDP and the uncertainty about the precise levels of dopamine and intracellular signaling molecules, it is difficult to clearly articulate how dopamine affect signal transmission locally at single

#246680 Received 24 Jul 2015; revised 11 Sep 2015; accepted 14 Sep 2015; published 17 Sep 2015

synapses. However, it is reasonable to assume that the LTP and LTD components of STDP are modulated by dopamine the same way as they are in the classical LTP and LTD protocols, which has already been supported by experimental evidence [26] showing the important role of dopamine in altering synaptic weights. In fact, most numerical studies of the reinforcement learning with dopamine modulation of STDP [15] adopt this modeling strategy, that is, if extracellular dopamine is present, the synaptic change (positive or negative) due to a particular order of firing will be enhanced. Particularly, the work of the BrainScaleS project in [19] has demonstrated the suitability of their wafer-scale hardware system (the neuromorphic computing platform of the Human Brain Project) to emulate this kind of globally modulated STDP learning rule with the reward as feedback. In our scheme, the same characteristic is realized by introducing feedback control of SOA current-injection.

Finally, our reward-modulated optical STDP can be further utilized to enable a broader range of applications and to achieve a higher level function. For example, it can be introduced to last layers of photonic spiking deep neural nets. Reinforcement learning with reward modulation can be equivalent to computing an approximation to the gradient of the reward [27]. Thus, one can utilize unsupervised learning (may not change connection weights [28]) at the level of a single layer, to discover a representation that captures statistical regularities of the layer's input and extracts slightly higher-level features compared with the previous layer, while adopt our optical STDP scheme with reward modulation in the last one or two layers of photonic spiking deep neural nets, to achieve a higher level function, such as classification or recognition. Since the scale of the last layer is usually at least two order of magnitude less than lower layers, the implementation complexity of our optical STDP scheme does not really matter.

In conclusion, some traits of biological STDP synapses are utilizing to advance the design of optical STDP. Feedback control of SOA current-injection is introduced to achieve a weight-dependent learning window and reward-based reinforcement learning. Based on the optical STDP proposed in this paper, advanced cognitive computing with photonic neuromorphic systems is a step closer to reality.

Acknowledgments

The paper is supported by NSF China, Grant No. 61471010, 61104142 and 61371074.

#246680 Received 24 Jul 2015; revised 11 Sep 2015; accepted 14 Sep 2015; published 17 Sep 2015