A biophysically-based neuromorphic model of spike rate- and timing-dependent plasticity

Citation


As Published

http://dx.doi.org/10.1073/pnas.1106161108

Publisher

National Academy of Sciences

Version

Final published version

Accessed

Tue Apr 26 12:23:39 EDT 2016

Citable Link

http://hdl.handle.net/1721.1/71924

Terms of Use

Article is made available in accordance with the publisher’s policy and may be subject to US copyright law. Please refer to the publisher’s site for terms of use.

Detailed Terms

The MIT Faculty has made this article openly available. Please share how this access benefits you. Your story matters.
A biophysically-based neuromorphic model of spike rate- and timing-dependent plasticity

Guy Rachmuth*1, Harel Z. Shouval1, Mark F. Bear3, and Chi-Sang Poon1,4

1Harvard-MIT Division of Health Sciences and Technology, Massachusetts Institute of Technology, Cambridge, MA 02139; 2Division of Engineering and Applied Sciences, Harvard University, Cambridge, MA 02138; 3Department of Neurobiology and Anatomy, University of Texas Medical School at Houston, Houston, TX 77030; and 4The Picower Institute for Learning and Memory and Department of Brain and Cognitive Sciences, Massachusetts Institute of Technology, Cambridge, MA 02139

Edited by* Leon N. Cooper, Brown University, Providence, RI, and approved October 3, 2011 (received for review May 24, 2011)

Current advances in neuromorphic engineering have made it possible to emulate complex neuronal ion channel and intracellular ionic dynamics in real time using highly compact and power-efficient complementary metal-oxide-semiconductor (CMOS) analog very-large-scale-integrated circuit technology. Recently, there has been growing interest in the neuromorphic emulation of the spike-timing-dependent plasticity (STDP) Hebbian learning rule by phenomenological modeling using CMOS, memristor or other analog devices. Here, we propose a CMOS circuit implementation of a biophysically grounded neuromorphic (iono-neuromorphic) model of synaptic plasticity that is capable of capturing both the spike rate-dependent plasticity (SRDP, of the Bienenstock-Cooper-Munro or BCM type) and STDP rules. The iono-neuromorphic model reproduces bidirectional synaptic changes with NMDA receptor-dependent and intracellular calcium-mediated long-term potentiation or long-term depression assuming retrograde endocannabinoid signaling as a second coincidence detector. Changes in excitatory or inhibitory synaptic weights are registered and stored in a non-volatile and compact digital format analogous to the discrete insertion and removal of AMPA or GABA receptor channels. The versatile Hebbian synapse device is applicable to a variety of neuromorphic, brain-machine interface, neurorobotics, neuromimetic computation, machine learning, and neural-inspired adaptive control problems.

iono-neuromorphic modeling | rate-based synaptic plasticity | silicon neuron | subthreshold microelectronics | VLSI circuit

Learning and memory are emergent animal behaviors governed by activity-dependent neuronal adaptation rules in response to changing environments. A putative neuronal mechanism of learning and memory is Hebbian synaptic plasticity (1)—the adaptive modification of excitatory synaptic strength following paired activation of the pre- and postsynaptic neurons. Two classical paradigms for the induction of Hebbian synaptic plasticity in the mammalian hippocampus and neocortex are rate-based plasticity (2–4) (herein referred to as spike-rate-dependent plasticity (SRDP)) and spike-timing-dependent plasticity (STDP) (5–7). The SRDP induction protocols control presynaptic firing rate in order to vary the sign and magnitude of synaptic plasticity (8): a high-frequency (20–100 Hz) train of presynaptic pulses results in long-term potentiation (LTP) of the synaptic strength, whereas a low-frequency (1–5 Hz) train results in long-term depression (LTD). These protocols are consistent with the theoretical learning rule (BCM rule) proposed by Bienenstock, Cooper, and Munro (9), in which the sign and magnitude of synaptic plasticity are controlled solely by postsynaptic activity as determined by presynaptic firing rate: low postsynaptic activity weakens synaptic efficacy and high postsynaptic activity strengthens it. By contrast, the STDP induction protocol stipulates that precise timing of pre- and postsynaptic activities determines the direction and strength of synaptic plasticity: repeated pairings of a presynaptic stimulus followed by a postsynaptic spike (prepost pairing, \( \Delta t > 0 \)) results in LTP, whereas reversing the order of pairing (postpre pairing, \( \Delta t < 0 \)) results in LTD (10). Mechanistically, both the SRDP and STDP induction protocols elicit NMDA receptor (NMDAR)-mediated intracellular calcium dynamics (4, 11, 12), which activate downstream processes that either up- or downregulate synaptic strength through the insertion or removal of individual excitatory AMPA receptor channels (13, 14). This common mechanistic link suggests a possible underlying interrelationship between these two seemingly distinct forms of Hebbian synaptic plasticity (15).

In an attempt to reconcile the SRDP and STDP rules, several computational models have been proposed (16–22). Currently, there is general agreement that LTP can result from both SRDP and STDP learning rules via postsynaptic NMDAR-mediated coincidence detection of presynaptic activities. However, the mechanism of timing-based LTD is less clear cut. Previous modeling studies have shown that a STDP learning rule involving a single postsynaptic coincidence detection mechanism as with LTD may induce LTD not only within the expected LTD window with postpre pairing (\( \Delta t < 0 \)) but also beyond the LTD window with prepost pairing (\( \Delta t > 0 \)) (17, 23–25). In order to robustly reproduce the canonical STDP curve with a single postsynaptic \( \Delta t < 0 \) LTD window, it has been proposed that a second coincidence detector may be required (23, 24). However, although many biophysical mechanisms and models of second coincidence detection for STDP have been proposed (25, 26), a general computational model that consistently unifies the SRDP and STDP rules based on an experimentally demonstrated second coincidence detector is currently lacking.

Previous theoretical analyses of the relationships between the SRDP and STDP rules were mostly based on numerical simulations of model equations on digital computers. Another approach to neural modeling is via direct emulation of neuronal dynamics on electronic devices such as complementary metal-oxide-semiconductor (CMOS) (27–29) or nanowire circuits (30–32); i.e., analog “neuromorphic” computation instead of digital model simulation. Recently, there has been growing interest in the neuromorphic modeling and implementation of the STDP learning rule using CMOS (32–42) or metal-oxide-metal circuits (43) or memristor-based nanodevices. Compared to conventional software-based computer modeling and simulation approaches, these neuromorphic electronic circuits have extremely small size (micro- to nanoscale) and low power requirements (\( \mu A \) to pA current per unit device with 0.5–5 V power supply) for large scale neural modeling and high speed simulation purposes. These capabil-
ities are critical for many real-time, portable/implantable neural computing applications such as neuroprosthesis, brain-machine interface, neurorobotics, neuromimetic computation, machine learning, or neural-inspired adaptive control (44). However, most such neuromorphic models emulate the temporally asymmetric STDP characteristic by direct phenomenological curve fitting (45) instead of biophysical modeling (26). It has been suggested that nonmechanistic phenomenological modeling of STDP may lead to many predictive failures especially when applied to other forms of synaptic plasticity (25). To our knowledge, none of the phenomenological neuromorphic STDP devices developed so far can reproduce the SRDP learning rule; hence, they are inflexible in responding to rate-based stimuli.

Another limitation of previous neuromorphic synaptic plasticity models is the difficulty of long term analog storage of synaptic weights using electrical capacitors (32, 46), which are volatile and bulky to implement on CMOS. Although compact nonvolatile long term memory of synaptic weights can potentially be achieved by using digital random-access memories (41) or advanced floating-gate (47, 48) or memristor technologies (49), these devices are not readily amenable to the biophysical modeling of NMDAR-mediated plasticity in a Hebbian synapse.

Here, we propose an iono-neuromorphic (i.e., biophysically grounded) CMOS circuit implementation of Hebbian synaptic plasticity that is capable of capturing both the NMDAR-dependent SRDP and STDP learning rules. Our iono-neuromorphic model is based on the hypothesis that retrograde endocannabinoid signaling and presynaptic NMDAR may provide a second coincidence detector of pre- and postsynaptic activity in addition to postsynaptic NMDAR (50–52). To emulate the underlying biophysical mechanisms, we employ a recently proposed wide-dynamic-range iono-neuromorphic CMOS circuit design approach that allows robust modeling of all types of voltage-dependent or ligand-gated ion channel and intracellular ionic dynamics on analog very-large-scale-integrated (aVLSI) circuits (53). We show that our iono-neuromorphic model readily reproduces LTP and LTD based on either the SRDP or STDP learning rules implemented on the same CMOS chip. The proposed iono-neuromorphic model of LTP and LTD lends itself readily to long term storage of synaptic weights in a nonvolatile digital format that is analogous to the discrete insertion and removal of AMPA or GABA receptor channels in real neurons, thus circumventing the limitations of analog memory.

Methods and Results

Iono-Neuromorphic Model of Postsynaptic NMDAR-Dependent LTP and LTD. Iono-Neuromorphic model of NMDA and AMPA channels. A “learning synapse” circuit model of an excitatory postsynaptic hippocampal dendritic spine compartment is designed as follows (Fig. 1A). A set of CMOS building block circuits biased in the subthreshold regime for robust iono-neuromorphic modeling [with wide input dynamic range to overcome device mismatch in subthreshold circuits (44)] are configured to emulate fast AMPA and slower NMDA channels, as described previously (53). The output currents are sent to a membrane node circuit that keeps the membrane potential $V_{\text{MEM}}$ near the resting potential $V_{\text{REST}}$ in the absence of stimulation (Fig. 1B). In response to a single presynaptic stimulation, excitatory $I_{\text{AMPA}}$ and inhibitory $I_{\text{NMDA}}$ impinge on the membrane capacitor ($C_{\text{MEM}}$) causing $V_{\text{MEM}}$ to generate an excitatory postsynaptic potential (EPSP) that relaxes towards $V_{\text{REST}}$ at a rate determined by the membrane time constant $\tau_{\text{MEM}} = \frac{C_{\text{MEM}}}{g_{\text{leak}}}$ (Fig. 1C). Importantly, several discrete AMPA channels carry excitatory postsynaptic current (EPSC) in parallel, and each channel is gated by a binary control variable $C_n$ (where $n = 1, 2, ..., N$) that determines whether a particular AMPA channel is active. Thus, the number of active AMPA channels encodes the synaptic weight.

NMDA channels—gated by both presynaptic glutamate and postsynaptic $V_{\text{MEM}}$ control over extracellular magnesium block—have slower dynamics, and encode coincident pre- and postsynaptic activities by $I_{\text{NMDA}}$ amplitude. Calcium influx via $I_{\text{NMDA}}$ (generated as its own current $I_{\text{Ca}^{2+}}$) is integrated on a current-voltage converter circuit to generate intracellular calcium ($[Ca^{2+}]$) signal (Fig. 1C). The calcium signal in turn activates downstream circuits that adjust the number of active AMPA channels ($C_n$ vector) according to a learning rule implemented by $[Ca^{2+}]$-dependent plasticity circuits.

The iono-neuromorphic synapse design is biologically intuitive and allows application of experimental manipulations to observe emergent behaviors. For example, the model is capable of modifying hippocampal silent synapses expressing only NMDA channels into expressing AMPA channels following an induction protocol (54). Additionally, the circuits allow tremendous flexibility in emulating synapses from various brain structures by simply tuning a small (1–4) set of parameters such as maximum conductance or activation dynamics of both excitatory and inhibitory channels.

Iono-Neuromorphic intracellular calcium-mediated plasticity model. Models of synaptic plasticity posit an important role of calcium in mediating downstream processing that results in expression of potentiation or depression of the synaptic weight. The learning rule implementation underlying on-chip synaptic plasticity is an adaptation of a biophysical model proposed by Shouval, et al. (17, 25) that relies on intracellular calcium dynamics to determine synaptic plasticity. The model computes the change in synaptic weight ($dw$) by evaluating:

$$dw = \eta([Ca]) \cdot \left(\Omega([Ca]) - \lambda w\right), \quad [1]$$

where $w$ is the present synaptic weight, $\Omega([Ca])$ is the calcium-

---

**Fig. 1.** (A) Simple synapse consisting of AMPA and NMDA channels, and calcium. (B) Circuit models of individual elements of the synapse, color coded with (A). (C) Circuit outputs in response to a presynaptic action potential (AP) input (AP$_{\text{Pre}}$). See also Fig. S1.
dependent update rule, $\eta([Ca^{2+}])$ is a monotonically increasing (supralinear) function of $[Ca^{2+}]$, that controls the learning rate, and $\lambda$ is a “forgetting” constant that assures that the synaptic weight reverses back from saturation if it is not maintained at its potentiated or depressed levels.

The $\Omega$ function employs LTP and LTD thresholds values ($\theta_{LTP}$ and $\theta_{LTD}$, respectively) to set potentiation and depression levels as a function of $[Ca^{2+}]$. By simply changing the $\theta_{LTP}$ and $\theta_{LTD}$ thresholds, several plasticity vs. $[Ca^{2+}]$ rules can be realized. For our purposes, we implemented the SRDP learning rule that has been observed experimentally in visual cortex and in hippocampus (55, 56). The rule postulates that for $\theta_{LTD} < [Ca^{2+}] < \theta_{LTP}$, the synaptic strength will depress; when $[Ca^{2+}] > \theta_{LTP}$, the synaptic strength potentiates. It is important to note that this rule was developed to explain SRDP results which involve long trains of stimulation that generate a constant output in response to $\lambda$ and $\theta$.

The $\Omega$ and $\eta$ functions are expressed mathematically using exponential functions, making them simple to implement using subthreshold-biased transistors. Thus, it is practical to use a VLSI technology to incorporate the learning rule into a large number of subthreshold-biased transistors. Therefore, it is practical to use a VLSI exponential functions, making them simple to implement using STDP protocols. Importantly, learning rules as seen in the hippocampus (Fig. 2B) and similar learning rules observed in the cerebellum (57). Importantly, LTD thresholds values ($\theta_{LTD}$) to set potentiation and depression levels $[Ca^{2+}]$. When $[Ca^{2+}] > \theta_{LTD}$, the synaptic strength will depress; when $[Ca^{2+}] > \theta_{LTD}$, the synaptic strength potentiates. It is important to note that this rule was developed to explain SRDP results which involve long trains of stimulation that generate a constant $[Ca^{2+}]$ level for a relatively long period of time. This assumption is not true for STDP protocols.

The $\Omega$ and $\eta$ functions are expressed mathematically using exponential functions, making them simple to implement using subthreshold-biased transistors. Thus, it is practical to use a VLSI technology to incorporate the learning rule into a large number of synapses. The $\Omega$ function computes an output signal $I_{\Omega}$ as a function of $[Ca^{2+}]$ signal from the calcium circuit (Fig. 2A). The $\Omega$ circuit is split into LTP and LTD sections. A cascade of differential pair circuits compare $[Ca^{2+}]$ and a threshold (either $\theta_{LTP}$ or $\theta_{LTD}$) and compute the output current for each section. These output currents are subtracted from each other, and the resultant current is sent to a comparator that generates a digital pulse when $I_{\Omega}$ falls below $\Theta_{\text{REST}}$.

The $\eta$ circuit is designed to mimic the calcium-dependent learning rate irrespective of the direction of plasticity. This circuit captures the fact that while both LTD and LTP can be induced using a 900 pulse train, longer stimulation trains are needed to generate LTD ($\sim$15 min at 1 Hz stimulation) compared to LTP (9 s at 100 Hz). Because we update the synaptic weight with enabling discrete AMPA channel circuits to fully turn on or off, we converted the $\eta$ circuit into a calcium-dependent digital $Enable$ signal. This approach allows the nonmetabolic synapse to modify its synaptic weight only during induction protocols that generate calcium influx that accumulates above a putative enable threshold, $\theta_e$. The circuit employs a transconductance circuit with a calcium-dependent bias current $I_{\text{BIAS-Ca}}$, which results in different charging rates based on a dynamic calcium level (Fig. 2C) and superlinear behavior of the $\eta$ function.

The circuit generates an output current $I_{\eta}$ proportional to the difference of $[Ca^{2+}]$ and a rest voltage $\eta_{\text{REST}}$ whenever $[Ca^{2+}] > V_{\eta_{\text{REST}}}$. $I_{\eta}$ is converted to a voltage signal $V_{\eta}$ via a comparator assuring monotonically increasing output as a function of the induction protocol. A leak transistor is included to keep $V_{\eta}$ near $V_{\eta_{\text{REST}}}$ for quiescent activity. $V_{\eta}$ is sent to a pulse-generator circuit that compares $V_{\eta}$ to the enable threshold $\theta_e$. The circuit resets itself whenever it goes high generating an $Enable$ pulse, and resets for a period of time $\Delta t$ (a time period that is not predictable due to the subthreshold biasing). The circuit therefore signals that an induction protocol has generated enough calcium to enable the synapse to update its synaptic weight (Fig. 2D). Because during an induction protocol, $[Ca^{2+}] > V_{\eta_{\text{REST}}}$ for a period of time, $V_{\eta}$ can generate several $Enable$ pulses during the induction protocol, allowing the synapse to update dynamically and then finally locking in a value when the $[Ca^{2+}]$ falls below $V_{\eta_{\text{REST}}}$.

This property is inherent in the model, suggesting that updating the synaptic weight is strongly dependent on the exact dynamics of the synaptic expression mechanism. In our circuits, the expression dynamics are a function of $[Ca^{2+}]$, according to:

$$\tau_{\text{update}} \sim \frac{\theta_e \cdot C_{\text{cap}}}{I_{\eta} \cdot [Ca^{2+}]}$$

where $C_{\text{cap}}$ is the capacitor of the $\eta$ circuit, and $\theta_e$ is a threshold voltage of the comparator. This equation suggests that induction protocols must endure at least $\tau_{\text{update}}$ seconds for the synaptic weight to update in an unsupervised manner. To simulate SRDT, we set $\theta_e$ to a level that requires calcium levels to remain elevated for at least several seconds or minutes before the $Enable$ signal is generated in order to reproduce experimental results (8).

Fig. 2. (A) $\Omega$ circuit. Input $[Ca^{2+}]$ signal used to compute output current $I_{\Omega}$. (B) $I_{\Omega} -$ $[Ca^{2+}]$, curve parameterized by $\theta_{LTP}$ and $\theta_{LTD}$. (C) $\eta$ circuit. Input $[Ca^{2+}]$ is used to both set the WR-TCA maximum current and as an input. The integrated $V_{\eta}$ is sent to a comparator that generates a digital pulse when $V_{\eta} > \theta_e$. (D) $\eta$ circuit output in response to $[Ca^{2+}]$ transient.
Digital storage of synaptic weights for LTP and LTD. In our model, synaptic weights are encoded iono-neuromorphically by a discrete set of digitally gated AMPA channels controlled by $C_n$, the updated synaptic strength. If $C_n$ is “HI” (LO), the nth AMPA conductance is activated (deactivated) and the synaptic weight is potentiated (depressed) compared to its previous state. To convert from an analog $\eta$ and $\Omega$ to a digital vector $C_n$, we set $\lambda = 1$, digitized $I_{CA}$ using a picocompand A/D converter (59, 60) and paired the Enable signal from the $\eta$ circuit to update a W-circuit (e.g., dw from Eq. 1) following a particular induction protocol.

The W-circuit updates the weight in two steps. First, an asynchronous digital finite state-machine ([FSM], see Fig. S1, Table S1 for details) uses the digitized $\Omega$ signal ($D_\Omega$) to compute future weight vector $C_n$, which are mapped to the inputs of dedicated D-flipflop (DFF) circuits that control individual AMPA channels. This vector then waits for an Enable (square pulse) command from the digitized $V_i$, which causes the $C_n$ vector to update the synaptic weight, and resets $V_{\text{syn}}$ back to zero.

The FSM provides a convenient abstraction of multiple nonlinear and interacting cellular and molecular mechanisms thought to be activated by postsynaptic Ca$^{2+}$ signal to drive synaptic plasticity; e.g., calcinurin driving LTD (61) or CaMKII driving LTP (62). The FSM can incorporate an arbitrary update rule as a function of $D_\Omega$. By modifying the ratios of mirrored $I_{\text{AMPA}}$ (see Fig. 1B), the synaptic weight can be varied in a nonlinear way. Interestingly, the inevitable mismatch across transistors due to CMOS process variability (63) provides a natural method for heterogeneous AMPA conductances across synapses. Thus the fabrication process may generate synapses with different maximum synaptic weight that, when applied in a large network, give rise to interesting neuronal computations.

CMOS prototype. As a proof of concept, the CMOS iono-neuromorphic Hebbian synaptic plasticity chip was prototyped using the AMI 1.5 $\mu$m process (Fig. S2A); use of deep-submicron processes will further reduce the chip size at a higher cost. Currently, the chip’s size was 4.6 mm x 4.6 mm with 116 I/O pins, although the circuits consumed less than 50% of the available area. The chip consisted of several ion-channel circuits (e.g., AMPA, NMDA, voltage-gated calcium, etc.), the $\eta$ and $\Omega$ circuits as well as digital storage circuits, totaling 400 transistors and 10 capacitors for implementation of both the NMDAR-dependent postsynaptic plasticity model and the endocannabinoid-dependent presynaptic plasticity model (see below). The transistor count reflects the relative complexity of our biophysically grounded iono-neuromorphic model compared to phenomenological models. Additional transistors were also needed in our wide-dynamic-range subthreshold CMOS circuit designs, which effectively mitigated the effects of transistor mismatch and significantly improved the robustness of aVLSI implementation (44, 53). The capacitors occupied the bulk of the chip area as some of them were relatively large (30 pF) in order to achieve a 1:1 electronic-to-biological time scale at nA current levels. The sizes of the capacitors are comparable to those required in phenomenological neuromorphic models. The circuits were interfaced with each other via external pins so that each circuit could be tested independently on a circuit board (Fig. S2B) to ensure proper operation. The chip consumed on the order of 100 mW of power (on a 5 V power supply) when only the analog circuits were operational (majority of the time). When undergoing a synaptic plasticity induction protocol, the activation of the digital portions of the circuit increased the power consumption only transiently. All circuits carried out simulations in real biological time.

Emulation of Postsynaptic NMDAR-Dependent SRDP Learning Rule.

The learning synapse circuit was tested using several well known induction protocols to draw direct comparisons with biological preparations. We first employed a pairing protocol used to identify the role of NMDAR-mediated calcium influx on synaptic plasticity. Similar to experimental protocols, $V_{\text{MEM}}$ was voltage clamped at $V_{\text{REST}}$ and the synaptic reversal potential $E_{\text{SYN}}$ ($-70$ mV, respectively) while presynaptic stimulation was applied at a rate of 1 Hz. The results reveal differences in $I_{\text{NMDA}}$, $I_{\text{CA}}$, and Ca$^{2+}$ accumulation depending on the clamping voltage (Fig. 3A). Because $I_{\text{NMDA}}$ consists of Ca and Na ions, its reversal potential is 0 mV. Therefore, when $V_{\text{MEM}}$ is clamped at $E_{\text{SYN}}$, $I_{\text{NMDA}}$ is near 0. However, $I_{\text{CA}}$, which has its own reversal potential $E_{\text{CA}} = 140$ mV, is substantially larger. When $V_{\text{MEM}} = E_{\text{SYN}}$-repeated stimulations resulted in elevated (Ca$^{2+}$) level above $\theta_{\text{LTP}}$ for a period longer than $\tau_{\text{update}}$ and eventually resulted in potentiation (Fig. 3B).

We next tested SRDP induction protocols aimed at showing NMDA channel activity as a Hebbian coincidence detector of pre- and postsynaptic activity and its control of synaptic plasticity. SRDP protocols assume that the mean AP firing rate is the main information transfer mode in neural networks, and so employ trains of stimulations. We used a standard protocol of 900 pulses at several presynaptic firing rates ranging from low- to high-frequency stimulations (4). The resultant [Ca$^{2+}$] accumulation was proportional to the input stimulus frequency such that for higher frequencies, [Ca$^{2+}$] levels were higher than for lower frequencies (Fig. 3C). The calcium conversion circuit (Fig. 1B) saturates at different voltages for different stimulation protocols, which are monotonically increasing as a function of frequency. The signals in Fig. 3C are raw signals directly from the calcium circuit. The $\Omega$ circuit was tuned to a standard BCM curve, which transitioned from LTD to LTP at a calcium level that was elicited by a 10 Hz stimulus train. The outputs were sent to the W-circuit and resulted in graded potentiation for presynaptic input rates $>$10 Hz, and depression for input rates $<$10 Hz (Fig. 3D). These results reproduce experimentally recorded observations.

The results show that classical induction protocols require only NMDAR-mediated calcium dynamics to express synaptic plasticity.
city. Similar to biological experiments, blocking on-chip NMDAR activation also blocked rate-based synaptic plasticity.

**Emulation of Postsynaptic NMDAR-Dependent STDP Learning Rule.** Recent studies show that when somatic APs back-propagate into the dendritic habitat and are repeatedly paired with presynaptic stimulation, they can induce synaptic plasticity depending on precise arrival sequence (12, 64). Biophysically, STDP appears to depend on NMDAR-mediated calcium influx, suggesting similar induction mechanisms as SRDP plasticity. If STDP models assume that maximum calcium levels drive the synaptic plasticity, the function of maximum calcium levels should be the same regardless of prepost or postpre pairings at +100 ms and a fast transition in the critical window between [−5 ms 5 ms] (Fig. 4A). However, experimental STDP results suggest that there may be two separate systems mediating the postpre and prepost portions of STDP (23, 24), which should be modeled as a discontinuity rather than a fast transition between the lowest depression and highest potentiation levels.

We subjected the learning synapse to an STDP stimulation protocol as described by Bi and Poo (12). In this protocol, 60 prepost or postpre pairings were applied at the synapse at a rate of 1 Hz [or lower (65)]. A standard presynaptic pulse activated the postsynaptic compartment. A back-propagating AP (bAP) generated by an integrate-and-fire neuron circuit (66) was sent to the postsynaptic compartment with a delay Δ of 1−100 ms, with Δ ≠ 0 for prepost stimulations, and Δ < 0 for postpre stimulations. In response to an STDP pairing, postsynaptic NMDAR-mediated [Ca]i is sent to the Ω and η circuits, and the [Ca]i dynamics are reflected in both IΔ and VMEP. Both the presence of glutamate and the voltage spike (by the backpropagating AP) is required for NMDA channel activation (and consequent calcium influx). Because a prepost stimulation should result in potentiation, we tuned the Ω parameters such that for any Δ > 0, [Ca]i elevated above both ΘTP and ΘLTD for a brief period of time, but pairings with increasing Δ generate [Ca]i signal with lower maximum amplitude (Fig. 4B). In the absence of another pairing, [Ca]i quickly decreased at a physiological rate to below ΘTP, remains above the ΘLTD levels for some time, and decreased back to baseline. Thus, to generate potentiation, the model predicts that either a downstream mechanism should “lock in” the fact that [Ca]i, crossed ΘTP (a low-pass filtering effect with a very long time constant), or that τupdate (Eq. 2) is at the milliseconds level. This fast-acting coincidence detector may be reflected in dendritic ion-channel activities (65, 67), or by downstream CaMKII autophosphorylation in response to transient calcium elevation (68). We tuned η circuit’s τupdate to be very short by modifying Θη, and reproduced the LTD portion of STDP window.

For LTD, however, the situation is much more complicated. For the artificial synapse circuit with only the NMDA channel as calcium source, STDP protocols did not display abrupt transition in the calcium level around Δt ~ 0 (Fig. 4B) after tuning the activation dynamics of the NMDA channel or modifying the width of bAP, as suggested previously (17). Moreover, in a model where [Ca]i is only generated by NMDA channels, the maximum [Ca]i reaches similar values for prepost pairing at +40 ms as it does for postpre pairings at −10 ms (Fig. 4B). Hence, based purely on NMDAR-initiated calcium dynamics there should be a prepost window for LTD, as previously hypothesized (17, 23−25).

If the prepost arrival difference is −10 ms, then the bAP arrives 10 ms before the glutamate, and so by the time the glutamate binds, VMEP has decayed significantly leading to a smaller fraction of NMDA channel activation. If the prepost arrival timing is +40 ms, the glutamate has arrived 40 ms before the bAP. With a binding time constant of 80 ms, some of the glutamate would have disassociated from the NMDA channels by the time the voltage spike arrives, leading to a similarly smaller activation of NMDA channels. Both situations lead to smaller calcium influx. In a model where only the maximum calcium level determines LTD or LTP, this result predicts that if LTD occurs at −10 ms timing, then LTD must occur at the +40 ms timing (due to the approximately the same maximum calcium level), as it is an element of the model itself irrespective of exact parameters. This prediction of an LTD window at +40 ms (or so) has been suggested by Shouval et al. (17, 25).

The underlying assumption of most plasticity models is that prolonged, moderate levels of calcium activate various intracellular phosphatases (such as calcineurin) and lead to LTD. The dependence on NMDAR activity implies coincident activity, but the postpre scenario requires another calcium source. Recent data suggest that voltage-gated L-type calcium channels (CaV1.2) contribute to calcium influx in postsynaptic neurons (69). We therefore added circuit models of CaV1.2, to the postsynaptic compartment, which together with presynaptic-activated NMDA channels combined to generate different [Ca]i dynamics in response to postpre and prepost stimulations. However, calcium influx via CaV1.2 is small compared with NMDAR-mediated ICa, and its addition was not enough to generate a large difference in the maximum [Ca]i levels. We added L-type calcium channels to the postsynaptic neuron to generate some calcium influx in response to a backpropagating AP which could potentially lead to a larger calcium level when the glutamate arrives for postpre pairings. In this scenario, the calcium level for a Δt = −10 ms pairing could be larger than Δt = +40 ms pairing, leading to the calcium generated by a −10 ms pairing crossing the LTD threshold, and calcium generated by a +40 ms pairing not crossing the LTD threshold. However, L-type calcium current is much smaller than NMDAR-mediated calcium influx, and so the additional calcium provided by the L-type channel does not significantly affect the postsynaptic [Ca]i.

By using the chip, a purely postsynaptic mechanism of coincidence detection cannot robustly express both LTD and LTP under both SRDP and STDP protocols with the same model parameters. The two major discrepancies for the prepost protocols are: (i) ΘTP and ΘLTD must be decreased significantly from SRDP levels, possibly implying separate downstream sensors of postsynaptic [Ca]i accumulation; and (ii) the learning rate τupdate must be modified to operate on the order of seconds for SRDP to possibly milliseconds to account for STDP. Tuning the η and Ω circuits allowed modeling of prepost LTD. However, the model did not generate a single postpre LTD behavior with the same set of parameters even with the addition of CaV1.2, as discussed previously (17, 23−25). We hypothesize that there may be a separate bAP-induced suppression of NMDA receptor-mediated EPSCs which sets the spike-timing window for LTD (64).

**Emulation of STDP Learning Rule with Retrograde Endocannabinoid Signaling.** The need for a second coincidence detector beyond...
the postsynaptic NMDA channel was recently discussed in several papers (23, 24, 50, 70). A second biophysical coincidence detector that accounts for postpre LTD may be endogenous cannabinoid (endocannabinoid) molecules that are released in response to postsynaptic Ca\(^{\text{V-L}}\) (51, 52, 71, 72) (Fig. 5A). The endocannabinoid signal acts as a transynaptic retrograde messenger and influences cannabinoid type 1 (CB1) receptors present on presynaptic neurons (73). In the most parsimonious model, the coincident activation of CB1 receptors and presynaptic NMDA autoreceptors results in long term reduction of neurotransmitter release.

To account for possible coincident detection via retrograde signaling, we designed a new artificial synapse model and included circuit models of: Ca\(^{\text{V-L}}\) as an additional postsynaptic calcium source, an endocannabinoid signal, CB1 receptors, presynaptic NMDA channels, and glutamate release mechanism. Postsynaptic Ca\(^{\text{V-L}}\) and NMDA channel circuits were connected in parallel to a single current-voltage converter circuit to generate postsynaptic \([\text{Ca}^{2+}]_{\text{PRE}}\). An endocannabinoid signal \((V_{\text{ENDO}})\) with an output proportional to \([\text{Ca}^{2+}]_{\text{CB1}}\) was sent to the presynaptic neuron and activated a CB1 signal \((V_{\text{CB1}})\) (Fig. 5B). This signal decayed back to baseline with first-order dynamics and a time constant of 40 ms. Because coincident presynaptic NMDAR and CB1 activity is required for a decrease in glutamate release, the presynaptic NMDAR circuit incorporated an additional subcircuit with inputs \(V_{\text{CB1}}\) and a parameter \(\theta_{\text{CB1}}\) that generated a positive signal when \(V_{\text{ENDO}} > 0\). This signal causes \(V_{\text{CB1}}\) to decrease from its baseline, affect \(\Delta_{\text{NMDA}}\)-PRE and thereby increasing \([\text{Ca}^{2+}]_{\text{PRE}}\) levels.

Because glutamate is released in discrete vesicles, a process that is conceptually similar to activating individual AMPA channels, for example), we employed a W-circuit with dedicated \(\Omega_{\text{Glu}}\) and \(\eta_{\text{Glu}}\) circuits that had \([\text{Ca}^{2+}]_{\text{PRE}}\) as their input. Similar to the description above for AMPA channels insertion and removal, \(\Omega_{\text{Glu}}\) and \(\eta_{\text{Glu}}\) were digitized and an FSM was used to determine the number of discrete glutamate vesicles released during the subsequent stimulation. The \(\theta_{\text{Glu}}\) threshold value was tuned so that during standard stimulation, the digital \(\eta_{\text{Glu}}\) circuit generated enable pulses at a high rate based on Eq. 2 (Fig. 5C). Because \([\text{Ca}^{2+}]_{\text{PRE}}\) is dynamic, we cannot predict the \(\tau_{\text{update}}\) with any degree of certainty. \(\tau_{\text{update}}\) was tuned so that low \([\text{Ca}^{2+}]_{\text{PRE}}\) levels (indicating no CB1 activity) resulted in a higher number of glutamate vesicles released. Higher levels of \([\text{Ca}^{2+}]_{\text{PRE}}\)

generated low glutamate release. Therefore, the amount of glutamate released is a (pseudo) stochastic event determined by \([\text{Ca}^{2+}]_{\text{PRE}}\).

We subjected the new learning synapse to an STDP stimulation protocol as described above, but used the same values of \(\theta_{\text{LTP}}, \theta_{\text{LTD}},\) and \(\eta_{\text{LTD}}\) in the postsynaptic \(\Omega\) and \(\eta\) circuits as those in SRDP simulations. During quiescent presynaptic activity, \([\text{Ca}^{2+}]_{\text{PRE}}\) remains low because L-type channels are inactive and there is no endocannabinoid signal. During prepost STDP protocols, LTP is induced by increases in \([\text{Ca}^{2+}]_{\text{PRE}}\) via pairing-specific activation of postsynaptic NMDAR and L-type calcium channels as before. In the prepost paradigm, the retrograde endocannabinoid signal resultant from the bAP is behind the presynaptic spike; hence, it always arrives at the presynaptic terminal before the postsynaptic NMDAR is no longer activated by the presynaptic spike, and decreases back to rest (with an assumed time constant of 40 ms) before the next pairing occurs. Because coincident detection at the presynaptic terminal requires paired activations of presynaptic NMDAR and CB1 receptor, the endocannabinoid signal alone does not affect glutamate release or the induction of LTD during prepost pairings (Fig. 5D).

In contrast, during postpre pairings \([\text{Ca}^{2+}]_{\text{PRE}}\) is modulated by \(V_{\text{ENDO}}\) and \(V_{\text{CB1}}\), and imparts its dynamics on the \(I_{\text{Glu}}\)-GLU signal. Because the \(\Omega_{\text{Glu}}\) circuit is biased towards decreasing glutamate release during postpre pairings, and the \(\eta_{\text{Glu}}\) circuit generates several enable signals during a single calcium transient, postpre pairings probabilistically bias the \(\Omega_{\text{Glu}}\) circuit towards generating a smaller amount of glutamate release for the imminent presynaptic stimulation. This strategy fits nicely with the fact that dozens of pairings (n=60–100) are required to generate STDP. During each postpre pairing, we simply decrease the probability of glutamate release. This smaller glutamate will reduce postsynaptic NMDAR and GLU, and its summation across the several pairings will cause the expression of LTD. Furthermore, the probability of releasing low glutamate decreases with increasing \(\Delta\) of the pairings. Fig. 5D shows several runs of the postpre STDP protocol and the resultant postsynaptic calcium signal.

**Fig. 5.** (A) An improved synapse model including presynaptic circuits of Ca\(^{\text{V-L}}\), NMDA autoreceptors and CB1 receptors. (B) Signals involved in transsynaptic communication; bAP; response of L-type VGCC, resulting \(V_{\text{ENDO}}\) and CB1 receptor activity. (C) Circuit design used to control glutamate release mechanism. (D) Chip results showing both LTD and LTP sections of STDP under retrograde endocannabinoid signaling. The blue line represents an average of several runs.
little or no effects on presynaptic-only induction of SRDP. Thus, all four induction protocols (low- and high-frequency presynaptic stimulations, prepost and postpre paired stimulations) can be easily emulated using the new synaptic model with the same set of parameter values for both SRDP and STDP protocols.

It is tempting to speculate that retrograde endocannabinoid signaling might modulate SRDP presynaptically when \([Ca^{2+}]_i\) becomes sufficiently high during rate-based stimulation, in a manner similar to our endocannabinoid-based STDP model. Consistent with this hypothesis, endocannabinoids have been shown to contribute to the LTD induced by low-frequency stimulation of the Schaffer collateral pathway in hippocampal CA1 neurons (74) and restrict the LTP induced by moderate high-frequency stimulation or maximum-intensity theta-burst stimulation of the same pathway (75). However, endocannabinoids did not affect LTP induced by robust and longer high-frequency stimulation (75). Instead, other studies showed that repetitive activation of the same pathway induces an endocannabinoid-mediated heterosynaptic LTD of nearby inhibitory inputs (76), and that priming these GABAergic receptors facilitates LTP of excitatory transmission (77, 78). Therefore, the effects of endocannabinoid signaling on SRDP are highly complex and may involve multiple interacting mechanisms that are not yet fully understood. Nevertheless, the ione-neuromorphic model presently proposed should be readily extendable to emulate endocannabinoid modulation of rate-based homosynaptic and heterosynaptic LTP and LTD of excitatory and inhibitory transmission and their interaction, when the requisite experimental data become more complete in future.

Discussion

The foregoing demonstrates our successful implementation and testing of an ione-neuromorphic circuit model of both SRDP and STDP learning rules on a miniature, low-power CMOS chip. The power consumption of our device is more than an order of magnitude lower than that of current memristor devices (which typically operate in the sub-\(\mu\)A current range) used to implement the phenomenological STDP rule (49). Our combined use of analog ione-neuromorphic modeling of NMDAR-dependent synaptic and intracellular calcium dynamics and retrograde endocannabinoid signaling allows robust on-chip simulations of bidirectional LTD and LTP induction based on either the STDP or SRDP learning rule. The proposed neurally-inspired digital storage of synaptic weights for long term maintenance of postsynaptic LTP and LTD emulating the insertion and removal of AMPA receptor channels in biological neurons provides an optimal mixed-signal hardware environment for reliable real-time simulation of Hebbian synaptic plasticity using power-efficient and compact aVLSI technology. Although not part of the present chip, a similar mixed-signal approach should be equally applicable to the digital implementation of long term maintenance of endocannabinoid-mediated presynaptic LTD, in that presynaptic neurotransmitter release is intrinsically quantal in nature and is up- or downregulated in discrete packets analogous to the discrete insertion/removal of postsynaptic AMPA channels.

In addition to these neurotechnological advances in ione-neuromorphic modeling and neural computation, the present work also has important implications in understanding the mechanisms of STDP from the perspective of computational neuroscience. Our simulation results support the notion that a second coincidence detector involving CB1 may be involved in the full expression of the canonical STDP curve characteristic of a prepost LTP window and a postpre LTD window without a second (prepost) LTD window, as suggested in several brain systems (50, 79–81). Clearly, endocannabinoid-dependent model of LTD is only one of several proposed models (25, 64, 65) that can be implemented using similar circuits as described here. Two interesting characteristics of the STDP rule is a dependence of the induction of LTD on the repetition frequency of prepost pairing (82) and a dominance of LTD over LTP in STDP integration of triplets or quadruplets of alternating prepost or postpre activities (83). Our model's demonstrated ability in reproducing both the SRDP and STDP rules is indicative of a frequency dependence of induction of LTP by STDP. It has been suggested that a triplet learning rule under certain assumptions can be mapped to a BCM rule (21); hence the present model relating the STDP and BCM learning rule biophysically should be compatible with the triplet STDP rule phenomenologically. While introducing more complex triplet curve-fit models including higher dimensional kernels such as proposed by Pfister and Gerstner (21) might be able to fit both the STDP and BCM rules, modeling of the actual biophysical processes is likely to fit more of the data, though this question will only be resolved by ongoing research. Although no attempt was made in this study to reproduce the timing-dependent integration of triplet and quadruplet stimuli, the relative robustness of the canonical STDP pairing protocol for the calcium-mediated induction of LTD with a single coincidence detector vis-à-vis the induction of LTD involving a second coincidence detector is consistent with a potentiation-dominance effect of STDP. For simplicity, our model does not include detailed biophysical descriptions of other synaptic processes (such as postsynaptic metabotropic glutamate receptors) for LTD and LTD induction and other intracellular protein signaling cascades involved in their maintenance (10, 26). These circuit models can be incorporated in future and can be optimized to generate a more comprehensive and space-efficient ione-neuromorphic CMOS synaptic plasticity system.

It should be noted that SRDP Hebbian synaptic plasticity rules are based on correlations of input and output signals with no feedback. Thus, synaptic modification algorithms follow an unsupervised learning rule. Reward-based and supervised learning rules, on the other hand, act to optimize some objective function by reducing error between actual and expected behaviors. These three learning rules are hypothesized to drive computations in various brain regions, and interact in an integrative manner (84). Thus far, unsupervised learning has received the bulk of attention due to its biophysical plausibility based on NMDAR-dependent mechanisms in heavily investigated hippocampal circuits. However, recent modeling studies suggest that the STDP rule may constitute a supervised learning rule if it is modulated by some global reward signal (85–88). Thus, both supervised and unsupervised learning can be accomplished by algorithms sharing the same biophysical substrates for SRDP and STDP Hebbian learning.

Recent studies have shown that inhibitory GABAergic synapses also undergo long term synaptic plasticity (10, 89). As with excitatory synaptic plasticity, postsynaptic \([Ca^{2+}]_i\) plays an important role in shaping inhibitory synaptic plasticity. For example, in neonatal rat hippocampus, calcium current through NMDA channels leads to LTD of GABA\(_\text{A}\) receptor channels while calci-
and the use of wide-dynamic-range circuit designs as proposed ability is mitigated by allowing sufficiently large device area constraint on circuit performance (63). Such hardware vulnerability is the intrinsic sensitivity to mismatch of CMOS transistor threshold voltage (94), which imposes a major neuromorphic models is the intrinsic sensitivity to mismatch of efficient, compact environment.

calcium entering via $\text{Ca}^{2+}$ results in LTP (90). A possible expression mechanism appears to be the insertion and deletion of individual GABA channels into postsynaptic membranes (91). Therefore, similar calcium-dependent inhibitory synaptic plasticity models may be constructed, but consist of a set of GABA receptors rather than glutamate receptors (Fig. 6). The expression of inhibitory synaptic weight adaptation is realized by activating or deactivating discrete GABA channels in response to $\text{Ca}^{2+}$ dynamics. Both excitatory and inhibitory synaptic plasticity can be modeled by our iono-neuromorphic design approach to significantly enhance the computational capacity of the on-chip system. Such iono-neuromorphic Hebbian learning systems may be applied to a variety of robotics, pattern recognition, machine learning, and nonlinear adaptive control problems (92, 93) in a power-efficient, compact environment.

A potential limitation of aVLSI implementation of iono-neuromorphic models is the intrinsic sensitivity to mismatch of CMOS transistor threshold voltage (94), which imposes a major constraint on circuit performance (65). Such hardware vulnerability is mitigated by allowing sufficiently large device area and the use of wide-dynamic-range circuit designs as proposed here (44, 53). Interestingly, the STDP algorithm itself can be used to correct for such circuit imperfections, making aVLSI implementations of the STDP learning rule relatively robust to device mismatch (95). Further improvements of iono-neuromorphic circuit performance include incorporating thermodynamically equivalent models of ion-channel kinetics (96) in our wide-dynamic-range circuit designs and the use of advanced CMOS processes that are optimized for subthreshold circuits operation with reduced sensitivity to transistor mismatch (97). At the same time, recent advent of near-nanoscale three-dimensional CMOS processes and integrated circuit technology will likely further decrease device dimensions and overall chip sizes in near future (44, 98). Such neurotechnology advances provide a new dimension for understanding how the brain works and for transitioning this knowledge to practical applications (99).

ACKNOWLEDGMENTS. The aVLSI chip was fabricated with the support of the MOSIS (Metal Oxide Semiconductor Implementation Service) Education Program. G.R. was a recipient of the National Defense Science and Engineering Fellowship. This work was supported by National Institutes of Health grants EB005460, RR028241 and HL067966.


