Spike-timing-dependent plasticity with weight dependence evoked from physical constraints

Simeon A. Bamford[1,2], **Alan F. Murray**[3], **David J. Willshaw**[4]

[1]Neuroinformatics Doctoral Training Centre, University of Edinburgh.

[2]Complex Systems Modelling Group, Istituto Superiore di Sanità, Rome, simeon.bamford@iss.infn.it.

[3]Institute of Integrated Micro and Nano Systems, University of Edinburgh.

[4]Institute of Adaptive and Neural Computation, University of Edinburgh.

***Abstract*—Analogue and mixed-signal VLSI implementations of Spike-Timing-Dependent Plasticity (STDP) are reviewed. A circuit is presented with a compact implementation of STDP suitable for parallel integration in large synaptic arrays. In contrast to previously published circuits, it uses the limitations of the silicon substrate to achieve various forms and degrees of weight dependence of STDP. It also uses reverse-biased transistors to reduce leakage of a capacitance representing weight. Chip results are presented showing: various ways in which the learning rule may be shaped; how synaptic weights may retain some indication of their learned values over periods of minutes; and how distributions of weights for synapses convergent on single neurons may shift between more or less extreme bimodality according to the strength of correlational cues in their inputs.**

I. Introduction

A. Spike-Timing-Dependent Plasticity

Synaptic weight plasticity is a fundamental element of adaptability and memory in neural systems. STDP is an elaboration of Hebbian learning applicable in spiking neural systems, in which the modulation of synaptic weight is based on the relative timing of spikes produced by the pre-synaptic and post-synaptic neurons. Bi and Poo [8] observed that in cultured hippocampal neurons, the potentiation or depression of a synapse was dependent on the temporal order of induced pre- and post-synaptic activity. In this study [and in [36, 56]], pre-synaptic activity preceding post-synaptic activity cause potentiation and *vice versa*, (though in other studies the opposite temporal dependence has been observed [7] as well as symmetric temporal dependence [1]). Such STDP, as it has become known, was predicted prior to these observations in computational models [19] and has since been investigated extensively in computational neuroscience (the historical antecedents of STDP are traced in more detail by Morrison et al. [40] p. 481 and by Markram et al. [37]).

Song et al. [47] modelled STDP in a way which has been used in many subsequent studies, as follows. A pre-synaptic spike at time t_{pre} and a post-synaptic spike at time t_{post} modify the corresponding synaptic weight by $w \rightarrow w + F(\Delta t)$, where $\Delta t = t_{pre} - t_{post}$ and:

$$F(\Delta t) = \left\{ \begin{array}{ll} a_+.e^{(\frac{\Delta t}{\tau_+})}, & if \Delta t < 0 \\ -a_-.e^{(\frac{-\Delta t}{\tau_-})}, & if \Delta t \geq 0 \end{array} \right\} \quad (1)$$

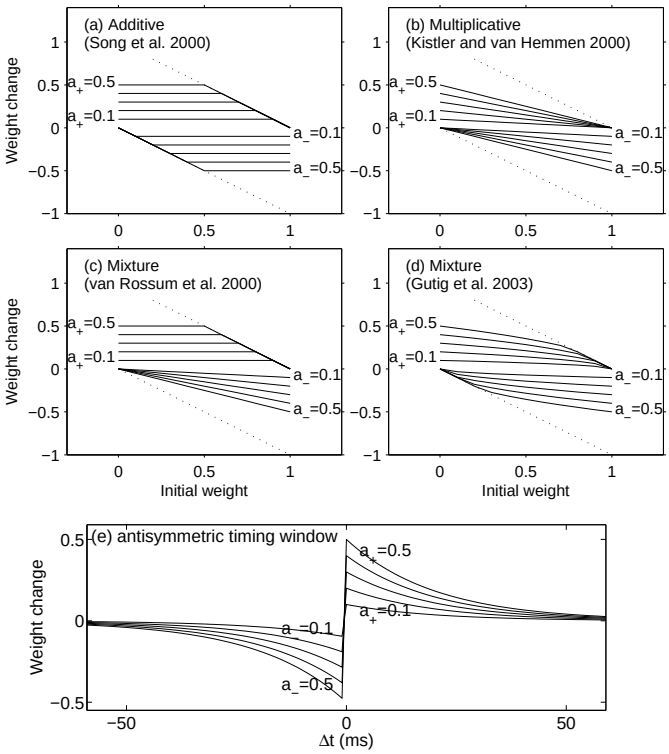


Figure 1: Theoretical models of STDP functions. Graphs (a)-(d) shows weight change (y axis) against initial weight (x axis). Lines link the effect of synaptic updates for the entire range of (normalised) initial weight for a particular magnitude of weight change; weight changes $a_{+/-}$ are swept through [0.1, 0.5]. Each graph shows potentiation (above) and depression (below). Dashed diagonal lines demark hard boundaries; for example the point 1,1 lies beyond the right-most dashed line, and is unreachable because a synapse with a weight of 1 is already at its maximum and cannot therefore experience a positive weight change. (a) Weight-independent learning rule [47]; (b) weight-dependent rule [30]; (c) mixture of weight-independent potentiation and weight-dependent depression [49]; (d) function which interpolates between weight-dependent and -independent rules, [22], for μ = 0.5 (see text for explanation). (e) Exponential decay of weight change against time difference between pre- and post-synaptic spikes; τ = 20 ms. In (a)-(d) all weight changes are maximised w.r.t. the exponential decay (i.e. at Δt = +/-0), and in (e) weight changes disregard any weight-dependent effects.

where $a_{+/-}$ are peak magnitudes of weight changes relative to the full weight range and $\tau_{+/-}$ are time constants for potentiation and depression respectively (experimental evidence suggests time constants of around 20 ms). This is cumulative for all pre- and post-synaptic spike pairs. *w* is bounded in the range $0 \leq w \leq 1$. The weight is then used to instantaneously increase the excitatory conductance of the neuronal membrane on the arrival of a spike and this conductance decays exponentially thereafter. Fig. 1(a) visualises this rule on a graph which plots peak weight change against initial weight; fig. 1(e) plots the more familiar view of weight change against Δt.

This model was used to show that in a neuron whose dendritic synapses implemented STDP, provided that $a_+\tau_+ < a_-\tau_-$,

the synaptic weights would diverge into a strong group and weak group (a bimodal distribution), with the effect that: (a) output spike rate was held within a narrow range relative to the range of mean input frequencies applied; (b) groups of synapses whose input spikes were more correlated, i.e. more likely to arrive within a narrow time window of each other, would be preferentially strengthened over synapses whose input spikes were less correlated. (a) is interesting as it is a form of homeostatic regulation and (b) is interesting as it allows unsupervised learning based on input correlations.

Other authors investigated multiplicative update rules, which assume linear attenuation of potentiation and depression as the upper and lower boundaries of weight, respectively, are approached. For example, Kistler and Van Hemmen [30] used the following function (using the same terms as equation 1, above):

$$F(\Delta t) = \left\{ \begin{array}{ll} (1-w)a_+.e^{(\frac{\Delta t}{\tau_+})}, & if \Delta t < 0 \\ -wa_-.e^{(\frac{-\Delta t}{\tau_-})}, & if \Delta t \geq 0 \end{array} \right\} \quad (2)$$

These rules are visualised in fig. 1(b). Van Rossum et al. [49] investigated a mixture of weight-independent potentiation and weight-dependent depression (fig. 1(c)), based on an experimental observation by Bi and Poo [8]. Where multiplicative rules are used, a unimodal distribution of synaptic weights results, but competition fails to achieve robust segregation of groups of synapses and the length of retention of any learnt correlations is greatly reduced [9].

Following this, Gutig et al. [22] developed a generalised STDP rule which allowed for a tunable degree of weight dependence, using the following function (again, adapted to use the same terms as above):

$$F(\Delta t) = \left\{ \begin{array}{ll} (1-w)^{\mu}a_+.e^{(\frac{\Delta t}{\tau_+})}, & if \Delta t < 0 \\ -w^{\mu}a_-.e^{(\frac{-\Delta t}{\tau_-})}, & if \Delta t \geq 0 \end{array} \right\} \quad (3)$$

Thus by changing the parameter μ, the model can capture a range of update rules between completely additive and completely multiplicative see fig. 1(d).

They then investigated various parameters for this model to determine under what conditions a bimodal distribution would result and to what extent such symmetry breaking would capture the correlational structure in the incoming activity. They found that there are constrained regions of the parameter space in which the correlational structure of the inputs can be captured by resulting synaptic weight distributions, and that the sensitivity of the outcome to differences in the amount of correlation between competing groups was maximised with $0 < \mu < 1$. Thus a degree of weight dependence can improve the ability of STDP to act as a correlation detection mechanism, and there is a trade-off between this effect and the reduction in the stability of learnt patterns.

There has been continued research and debate about the nature of STDP and its relevance as a candidate mechanism for memory and learning [40, 37]; a few points are worth mentioning: different temporal windows for plasticity have been mentioned above, and synapses at different locations on dendritic trees can have different temporal windows [32]; alternative models which update weight based on post-synaptic voltage rather than post-synaptic spike times can achieve similar performance and better matching to data [12, 15]; the relationship with prior knowledge about rate-based LTP and LTD is an active question, with some but not all observations of frequency dependence consistent with the models of STDP presented above; some models consider only the interaction of spike pairs nearest in time whereas others consider all possible pairings [40, section 4.1.2]; an STDP rule based on triplets of spikes rather than pairs can better match certain biological data [41]; the question of the underlying cellular mechanisms of STDP is unresolved - some models postulate internal synaptic variables, and these variables may be related to specific molecules [45]; even the nature of the synaptic weight is in question - it may be related to number of receptors or probability of vesicle release [45, 35], each hypothesis with different implications for frequency dependence. Notwithstanding all these possible elaborations, rules similar to the formalism of Song et al. [47] have been used to investigate: topographic map formation [46, 5]; the response to latency in inputs [23]; visual feature map learning [38]; receptive field reorganisation [53]; learning cross-modal spatial transformations [16]; the formation of synfire chains [29]; etc.

B. Neuromorphic implementations of STDP

Beyond its attractive computational properties, there are some practical reasons for STDP to be of interest to neuromorphic engineers: firstly, as it is based on spike timings it fits with a dominant paradigm of event-based neuromorphic circuitry for which much infrastructure has been developed; secondly, given a predominant design style in which synapse circuits are physically implemented contiguously with their post-synaptic neuron soma circuits, STDP is a form of Hebbian learning which only requires information which is anyway available physically at or close to the synapse circuit (i.e. the timings of pre- and post-synaptic spikes) and thus can have a local implementation which does not add to long-distance communication overheads. Due in part to these factors, there is the possibility of relatively compact circuit implementations. A number of STDP circuits have been published, and these are briefly reviewed in this section. Issues of weight dependence and weight stability are then specifically discussed in the following two sections respectively.

Häfliger et al. [26] produced a circuit in advance of the publication of the first biological evidence for STDP which nevertheless demonstrated some of its computational properties, and introduced some features common to later STDP circuits. Synaptic weight is represented by a voltage stored across a capacitor; this voltage is used to gate a transistor as part of a synapse circuit which passes charge to a neuron soma circuit and thus affects the efficacy of the synapse. Changes

are made to the weight capacitance upon the arrival of post-synaptic spikes, based on a value stored on another capacitor. This other capacitor is used to accumulate the effect of pre-synaptic spikes in the form of a leaky integrator. A differential pair then compares the value on the second capacitor to a threshold, in order to determine the direction and magnitude of the plasticity, which is actually applied as a current to the weight capacitance for the brief duration of a digital pulse representing the post-synaptic spike. Gordon and Hasler [21] implemented a qualitatively similar learning rule using floating gate transistors for storage and adaptation of synaptic weights (they have recently published a method for evoking STDP from floating gate transistor synapses [42]; Liu and Mockel [34] also achieved this, with a system that makes weight updates only after an accumulation of events, thus slowing the damage caused by hot-electron injection).

Indiveri et al. [27] presented a symmetrical circuit with two leaky integrators. One is augmented on the arrival of pre-synaptic spikes and decays towards one of the power rails, whereas the other is augmented in the opposite direction on the arrival of post-synaptic spikes and decays towards the other power rail. The first integrator therefore represents a *potential for potentiation*, which is then used, by means of gating a transistor, to increase the weight upon the arrival of a post-synaptic spike. The other integrator represents a *potential for depression*, which is used to decrease the weight upon the arrival of a pre-synaptic spike (*potential for potentiation* and *potential for depression* will be collectively referred to below as *potentials for plasticity*). The circuit therefore implements a learning rule with similarities to that described in equation 1, but (a) the magnitude terms $a_{+/-}$ were implemented by pulse lengths and then the term $a_+.e^{(\frac{\Delta t}{\tau_+})}$ and the corresponding depression term were limited in magnitude by further biases, creating time windows for plasticity which are not exponential decays; (b) in application, the voltage representing weight was limited to a range in which it could be used as a subthreshold bias in a synapse circuit, altering the interpretation of weight-change rules (this will be discussed below in section I-C).

Bofill-i Petit and Murray [11] was the first published circuit which (given a certain parametrisation) offers a close fit to equation 1. This is also based on the principle of two leaky integrators which modulate increments and decrements to a weight capacitor, and contains two notable elaborations: (a) given that the synapse circuits are intended to be laid out contiguously to their post-synaptic neuron circuit, the leaky integrator which accumulates the potential for depression based on post-synaptic spikes is implemented only once per neuron and the effect is distributed backwards to each of its dendritic synapses, for a possible space saving compared to having the integrator in each synapse circuit [27] (a current mode implementation and associated mirroring erodes the space saving, however); (b) the integrator of the potential for potentiation is modified to allow a tunable degree of weight dependence (this will be discussed below in section I-C).

Koickal et al. [31] presented another symmetrical circuit based on two leaky integrators, with the difference being that the leak and weight changes are both implemented by OTAs instead of single transistors. Charging of the weight capacitance is implemented with a unity gain buffer driving towards the instantaneous value of the leaky integrator representing the potential for plasticity, rather than towards a hard weight limit, introducing an unusual form of weight dependence in which, for example, a weak temporal correlation which should lead to a small potentiation according to the rules of equation 1 (the theoretical target rule stated in the paper), may in fact lead to depression if the synapse already has a high weight. Tanaka et al. [48] also used OTAs for weight changes but used them in open loop configuration in order to achieve a standard (antisymmetric) learning function, as well a symmetrical function [1].

Schemmel et al. [44] introduced STDP circuitry for use in a system with an intended speed-up factor of several orders of magnitude over real-time operation. Desiring high accuracy for weight changes and stable weight storage, they stored and refreshed weight digitally at each synapse with digital-to-analogue conversion (see also section I-D). The circuit which actually implements weight changes operates digitally at the periphery of the chip and is shared sequentially by all synapses; the advantage of locality and reduced communication is therefore sacrificed in this design for the advantages of compactness and the ability to programme different weight-change rules (the weight change rules are held as a pre-calculated look-up table). There is nevertheless local symmetrical circuitry for accumulation of plasticity events. This circuitry sacrifices the possibility of operating cumulatively over all pre- and post-synaptic spike pairs, in order to benefit from a dual mode of operation in which decay of a potential for plasticity occurs until a complementary spike arrives, switching the mode of the circuit to accumulation of plasticity concurrently with decay of the potential for the complementary form of plasticity.

The circuit of Arthur and Boahen [2] yet again uses complementary leaky integrators for accumulation of the potentials for plasticity, with the difference that these potentials are thresholded upon the arrival of the complementary spike to give all-or-nothing potentiation or depression events of binary synaptic weights (see also section I-D).

As a counterpoint to these examples in which analogue integration and decay of spike events may lead to either analogue [26, 11, 27, 31] or digital [2, 44] updates to a weight value, the implementation of Vogelstein et al. [51] uses an array of analogue synapses addressed by digital events but uses an off-chip microcontroller operating over RAM for a completely digital serial implementation of STDP. Other researchers such as Cassidy et al. [14] work on digital FPGA implementations of neural networks including STDP but the focus in this review is on analogue and mixed-signal implementations.

A parallel line of work pursued by Fusi et al. [18, 3, 39] has used alternative learning rules similar to some models mentioned in section I-A. Plasticity events occur on pre-synaptic spikes and their polarities are determined by the level of the post-synaptic neuron's membrane potential. This

is taken as indicative of the likely timing of a post-synaptic spike, since a neuron which has just spiked may have a low membrane potential whereas a neuron which is just about to spike may have a high membrane potential. In this way, spike-based correlation detection is recovered which has similar properties to the form of STDP presented in equation 1 but without the need to integrate post-synaptic spikes.

Finally, with the recent resurgence of interest in memristive devices has come the suggestion that a certain form of memristor which might become available in standard VLSI processes would have the ability, on its own, to behave as a synapse which implements STDP. Such a system has been modelled by Zamarreno-Ramos et al. [54] and a similar system has been tested for a single fabricated device by Jo et al. [28] (employing pulse-length modulation to define learning windows).

C. Weight dependence in neuromorphic STDP

The implementation of the different forms of weight dependence presented in section I-A in neuromorphic synapse circuits is now addressed. Zamarreno-Ramos et al. [54] showed that memristor-based synapses would deliver a weight-dependent form of STDP due to the physical limitations of memristors; this weight-dependence is evident though unnoted in the data of Jo et al. [28, figure 2a]. No one has yet presented a way to control the degree of weight dependence in memristor-based STDP. The CMOS STDP implementation of Bofill-i Petit and Murray [11] has weight-dependent potentiation and weight-independent depression, the (possibly unintentional) opposite of the learning rule observed by Bi and Poo [8] and modelled by Van Rossum et al. [49]. The amount of weight dependence is controllable and is achieved by using the weight voltage to generate a current which is used to modulate the magnitude of increments to the potential for potentiation. This requires 4 additional transistors and an additional continuously operated current. Although it is possible to explicitly engineer weight dependence like this, it may be more efficient if the natural limitations of transistors could be employed to provide weight dependence, as the natural limitations of the memristors of Jo et al. [28, figure 2a] seem to. The circuit which will be presented in section II takes just this approach, yet also allows the degree of weight-dependence to be controlled.

The bounding of the weight range should also be considered. Where a capacitance is used to model an analogue weight, the power rails which plasticity events are sourced from or sunken to provide hard limits to the weight. Thus there is no need to explicitly implement the bounding represented by the dashed lines in fig. 1 - such bounding comes for free as a physical limitation. Various implementations mentioned above utilise this effect [11, 48]. Other implementations could also utilise this effect but in practice have tighter bounds set by additional active bistability mechanisms [27, 39].

The floating-gate transistor implementations of STDP by both Ramakrishnan et al. [42] and Liu and Mockel [34] are weight dependent in a different manner to that modelled in equation 2, in which the magnitude of potentiation is related by a power law to the current weight rather than to the difference of the current weight from its maximum level; potentiation is thus facilitated rather than limited as weight increases. The application of the weight voltage from the circuit of Indiveri et al. [27] (also [39]) as a subthreshold bias in a synapse circuit effectively creates this same form of weight dependence, although with an exponential relationship of "weight" voltage to synaptic efficacy. For synapses with discrete STDP circuits, however, [11, 27, 31] (as opposed to single device synapse designs [28, 42]) it is possible to dissociate the profile of the "weight" voltage from its interpretation by downstream circuitry for creating synaptic currents. The linearised synapse circuit of Bofill-i Petit and Murray [11] could be used with the subthreshold STDP circuit of Indiveri et al. [27]; to do so would remove the aforementioned weight dependence, relating the weight voltage linearly to synaptic current. Given this assumption, the STDP would be almost but not quite weight-independent; the transistors which source and sink current to and from the weight voltage operate in the subthreshold region, in which their drain-source current is independent from their drain-source voltage, except in a narrow range in which $V_{ds} < 4KT$, i.e.$\approx$ 100 mV. Because there is a stack of 3 transistors leading from the weight capacitance to each power rail, the range in which this effect may be noticed is extended, and because one of these transistors operates in strong inversion, some weight-dependent reduction in plasticity in the direction of the boundaries may be observed throughout the entire range of weights. Complete weight dependence is not possible because of the inability of transistors to act as perfect current sources. This fact will be used in the circuit which will be presented in section II.

D. Weight stability in neuromorphic STDP

A major obstacle to the implementation of weight plasticity is that for long-term plasticity there needs to be a way of holding a continuously valued variable at a constant level. Synaptic weight is often implemented as a voltage across a capacitor, as in several of the implementations discussed in the previous section. However this is subject to leakage such that any learnt value will be lost over a period of milliseconds or seconds.

One solution (the aforementioned bistability mechanism [18, 27, 3, 39]) is to use weak positive feedback from an amplifier to drive the weight value either upwards or downwards away from a central threshold, yielding a distribution of weights which are bistable; they can take any value in the short term but in the long term they have only two stable states, potentiated or depressed.

Arthur and Boahen [2] went further, as noted above, modelling weights as having only two states even in the short term and using a static ram element to store this value, which could then be switched if an accumulation of either potentiating or depressing events surpassed a certain threshold. In a complementary approach, Häfliger and Riis [25] created an analogue memory element with many stable states, and

this was applied to synapses using an afore-mentioned STDP-like learning rule [24], though the implementation is bulky, requiring an amplifying element for each stable state; the circuit might be made space efficient by the use of time-multiplexing, however, along the lines of other multi-value stabilisation schemes [50].

An alternative solution is to use floating-gate technology; if synapse weights are stored as charge on floating gates then a learnt patterns of weights can be stable even in the absence of a power supply. Gordon and Hasler [21] and Ramakrishnan et al. [42], as noted above, used this technology for compact implementations of various spike-timing-dependent learning rules. Memristors are another candidate non-volatile storage technology which promises yet more compact synapses [28, 54]. Zhang et al. [55] used a further non-volatile technology (ionic electronic hybrid transistors) for a compact implementation of STDP. Their implementation is actually of a device which, together with surrounding pulse-handling circuitry, passes on a spike after a delay, where it is the delay that is subject to plasticity according to the relative timing of the incoming pulse to a further pulse; it remains to be seen if this mechanism can be used to deliver a more standard synaptic behaviour.

Another approach is to digitise the weight with a chosen level of accuracy; then it can be stored in standard memory. Such a digital memory has previously been used to periodically refresh local analogue storage of weights on capacitors [17]. More recently it has been used advantageously to allow synapses to become virtual rather than physical devices, so that one physical device can act as all the incoming synapses for a neuron, by sequentially receiving the weight information for each incoming spike and acting accordingly [20, 52]. This achieves a possible saving in area but with higher communication overheads.

Regarding the aforementioned approach of introducing bistability, Bofill-i Petit [10] argued that (weight-independent) STDP is inherently bistable and that this fact should be utilised to bypass the problem of volatile weight storage, rather than, for example, introducing explicit bistability. Such an approach is viable in situations where the input which leads to learnt patterns of weights is continuous. In the absence of such input, weights will converge on preferred levels due to leakage currents. The work presented in this paper is part of a project based on an alternative approach: by allowing weight distributions (which can change rapidly and are stored in volatile memory on capacitors) to influence network topology (which changes slowly and is stored in stable memory elements), features learnt from input can continue to influence the behaviour of a network long after the original input has occurred and the immediate memory trace has faded (the interested reader is directed to [6]). Notwithstanding this approach, steps are also taken to reduce leakage currents so as to maximise the lifetimes of memory traces stored on capacitors, as will be seen in section II.

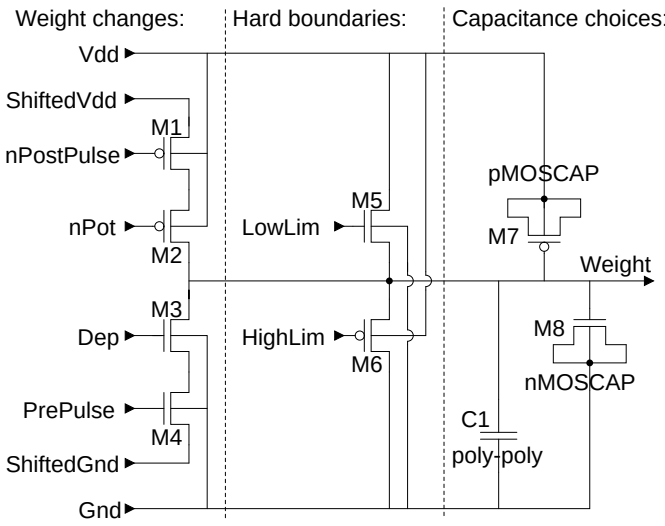


Figure 2: Proposed STDP circuit. M1-4 allow increments and decrements to the weight. Optional transistors M5-6 apply hard limits to the weight. C1 and M7-8 offer different choices for capacitive weight storage.

II. Circuit design

A. Proposed STDP circuit

The proposed STDP circuit follows the archetypal form apparent in many of the implementations reviewed in section I-B, by having a symmetrical design with leaky integrators for potentials for plasticity and by pulling up and down on a capacitance representing weight to implement plasticity events. Two novelties are presented. One is the application of a known principle for reducing weight leakage and thereby increasing the duration of learnt memories. The other is a method of introducing a variable amount of weight-dependence which does not require specific additional circuitry to do so but rather uses the limitations of CMOS to its advantage.

Fig. 2 shows the proposed circuit, which has been implemented in the AMS 0.35 μm 4-metal 2-poly process. The central weight voltage is stored across a capacitor. Three choices are shown for the implementation of this capacitance to the right of the graph, although it is intended that only one should be used in a given implementation; there is a poly-poly capacitor C1; then there are pMOSCAP (M7) and nMOSCAP (M8) structures, which benefit from greater capacitance for the same area (about $4\times$ more in the process used), but have a reduction in capacitance when the weight voltage gets close to their respective thresholds.

Incoming signals *Dep* and *nPot* represent the potentials for depression and potentiation respectively, which are created by leaky integrator circuits (the design of which is deferred to section II-B). The prefix *n* of *nPot* is to indicate that it is negatively defined, i.e. a low voltage represents a high potential for potentiation. Incoming signals *PrePulse* and *nPostPulse* are for digital pulses representing pre-synaptic and post-synaptic

spikes respectively. Although this circuit is intended to run at realistic biological time scales, these pulses are intended to be on the order of 10 ns in duration, somewhat shorter than the pulses used by e.g. Koickal et al. [31].

The stack of transistors M1-4 allow increments and decrements to the central weight voltage. The stack is sourced by *ShiftedVdd* and sinks to *ShiftedGnd*. These power rails are intended to be shifted inwards with respect to the outer power rails *Vdd* and *Gnd*; for example *Vdd* = 3.3 V, *ShiftedVdd* = 3.1 V, and *ShiftedGnd* = 0.2 V, all w.r.t. *Gnd*. The spike signals *PrePulse* and *nPostPulse,* however, switch between the outer power rails, such that when *PrePulse* is not active, M4 is negatively biased (likewise for M1 when *nPostPulse* is not active). Thus, in a manner suggested by Linares-Barranco and Serrano-Gotarredona [33], by sacrificing a few hundred mV of the whole voltage range, the leak from the *Weight* capacitance can be reduced to sub-pA levels, greatly increasing the lifespan of any synaptic weight changes.

When *PrePulse* is active, the current through M3-4 (and therefore the amount of depression) is dictated by two parameters: the values of *Dep* and *Weight* w.r.t. *ShiftedGnd*. *Dep* may vary from its resting level which is set around the nominal threshold of M3 up to *Vdd*. Thus transistor M3 will typically operate in strong inversion. *Weight* may vary between *ShiftedGnd* and *ShiftedVdd*; thus M3 may operate in either the saturated or linear region. M4, being gated by *Vdd*, will always operate in strong inversion and, if M3 and M4 are of the same dimensions, the drain voltage of M4 will not exceed $(ShiftedVdd - ShiftedGnd)/2$ and therefore M4 will not saturate, because $Vdd - Vt - ShiftedGnd$ will be greater than the drain-source voltage. For most values of *Dep*, the source voltage of M3 will stay close to *ShiftedGnd*; nevertheless the shifting of this source voltage according to the interaction between M3 and M4 complicates an analytical solution for the current through M3. However, to generalise, as *Weight* increases, M3 is more likely to be saturated, such that the current is only weakly dependent on *Weight* (through channel length modulation). As *Weight* decreases, M3 is more likely to pass into its linear region, in which the current will have a stronger dependence on *Weight*, being polynomially related with a negative squared term and a positive first-order term, according to the Sah equation. Raising *Dep* makes M3 more likely to operate in its linear region. Clearly, the same comments are valid for M1-2 in a complementary sense for parameters *nPot* and *Weight* whilst *nPostPulse* is active.

The effect on this circuit of potentiation and depression events was simulated and the results are shown in fig. 3. The effect of each type of event was simulated for a range of initial values for *Weight* and for a range of values of *nPot* and *Dep* respectively. It can be seen that as the initial level of *Weight* moves closer to its boundary in either direction, the amount of weight dependence of plasticity in that direction increases; this weight dependence is evoked from the physical constraint that transistors M2-3 cannot act as ideal current sources. This is qualitatively similar to the behaviour of the STDP model of Gutig et al. [22] for intermediate values of μ, as shown in fig. 1(d). One notable difference is that as *Dep* increases, the weight dependence becomes more pronounced; to be more precise, as *Dep* increases, the value for *Weight* at which transistor M3 reaches its boundary of saturation raises, such that strong weight dependence is apparent for a greater part of the entire range of *Weight*. This dependence of weight dependence on the potentials for plasticity is a behaviour not predicted or utilised by any of the models discussed in section I-A. However, it will be shown in section III-A that for a given distribution of a potential for plasticity, a certain weight dependence profile can be achieved.

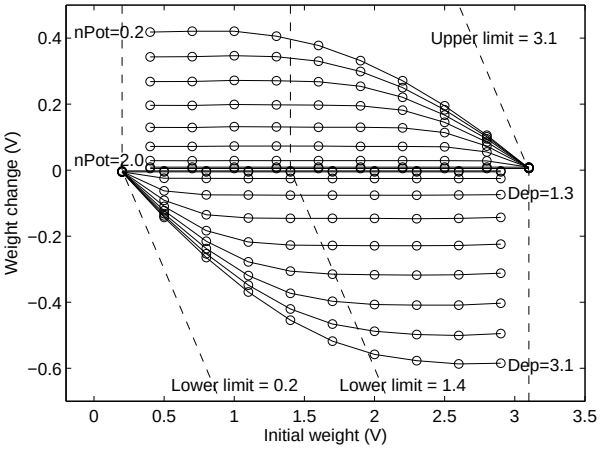


Figure 3: Simulated performance of STDP circuit. *Weight* was initialised at each of a set of values in the range 0.4-3.1 V. For each initial value of *Weight*, *nPot* was swept through a set of values in the range 0.2-2.9 V. For each of these conditions, a single active-low pulse was applied to *nPostPulse* for 10 ns and the resulting rise in *Weight* was recorded (the capacitance was a $10 \times 10\,\mu m$ poly-poly capacitor); these results are shown on the top half of the graph, with lines linking sets of results for each value of *nPot*. The bottom half of the graph shows a corresponding experiment with depression, where *Weight* was initialised in the range 0.2-2.9 V, *Dep* was set in the range 0.4-3.1 V and pulses were sent in to *PrePulse* for 2.5 ns (the difference in pulse width and the remaining differences in magnitude between potentiation and depression reflects a choice of width-length ratios for transistor pairs M1-M2 and M3-M4, which were optimised for implementing the parameters of a model which is not discussed in this paper). The outer boundary dashed lines mark the region of the graph which can be reached in normal operation. The left-most dashed line represents the lower boundary of 0.2 V imposed by setting *ShiftedGnd* at that voltage, and likewise for the upper limit of 3.1 V imposed by *ShiftedVdd*. The inner dashed line marked "Lower limit = 1.4" marks an arbitrary lower boundary for *Weight* which can be imposed if *LowLim* is set accordingly ($\approx$ 2 V). Simulation by Spectre within Cadence using AMS C35B4 process.

In fig. 3, the left- and right-most dotted lines show the boundary of achievable results for the normal range of operation, since *Weight* is bounded in the range 0.2-3.1 V by the shifted power rails. If however it were possible to raise the lower boundary for *Weight* without affecting the behaviour of the circuit then, at the expense of sacrificing some of the full

voltage range for *Weight*, the properties of weight dependence could be altered. On the graph is marked a hypothetical lower boundary of 1.4 V. If only the learning rule curves to the right of this boundary were used then the potentiation would be much more weight dependent than the depression, as can be seen by comparing the curves of the lines in this region of the graph; thus a learning rule similar to (in this case, the opposite of) Van Rossum et al. [49] could be recovered. This is the reason for "overflow" transistors M5-6. If the gate value of nMOS M5, marked as *LowLim* is raised from *Gnd* (where it has been assumed to be in the previous discussion), then as *Weight* passes below $LowLim - V_T$, M5 will turn on and restore *Weight* to $LowLim - V_T$. In order to use this mechanism it is necessary to make some assumptions about the speed of operation of the circuit, since the nominal threshold of the transistor does not provide a precise boundary; as *Weight* is raised above $LowLim - V_T$, M5 will continue to conduct, but will pass exponentially less current as *Weight* continues to rise. At some point, the remaining current will become irrelevant compared to the rate at which plasticity events change the weight in the course of normal operation, and *LowLim* should be set so that this occurs at the intended lower voltage level.

Between setting *LowLim* and *HighLim*, changing the range of the inner power rails *ShiftedVdd* and *ShiftedGnd*, and changing the average level achieved by *Dep* and *nPot* there are many ways to shape the learning function. These will be explored further in section III.

B. Further circuitry

The circuit described above has been fabricated on a test chip and its properties are explored below in section III. In addition, a version of this circuit was used in a multi-chip system intended to investigate topographic map formation and receptive field development through synaptic rewiring. The circuitry used in that system is described here in order to contextualise the circuit discussed above and to demonstrate practical usage. Section III-C presents results obtained from the multi-chip system.

Fig. 4 shows the synaptic circuitry relevant to STDP. In this system, weight is negatively defined, and thus w.r.t. the previous naming convention, *Weight*→*nWeight*, *nPot*→*Pot* and so on. Regarding the core circuitry, note that the potential of M6 to set a lower limit on synaptic efficacy is only partially realised in this system. In order to interpret the practical upper limit of *nWeight* which is achieved by the setting of *HighLim* as a synaptic efficacy of zero it would be necessary to have separate control of the source of M12, setting it so that when *nWeight* reaches that upper limit, the amount of current which passes through M13 upon a spike is so low w.r.t. the maximum synaptic efficacy as to be considered zero. Other implementations of the synaptic output could be considered which translate the range of values provided by *nWeight* more linearly into a synaptic output; indeed the downstream circuitry contains some modifications which go some way

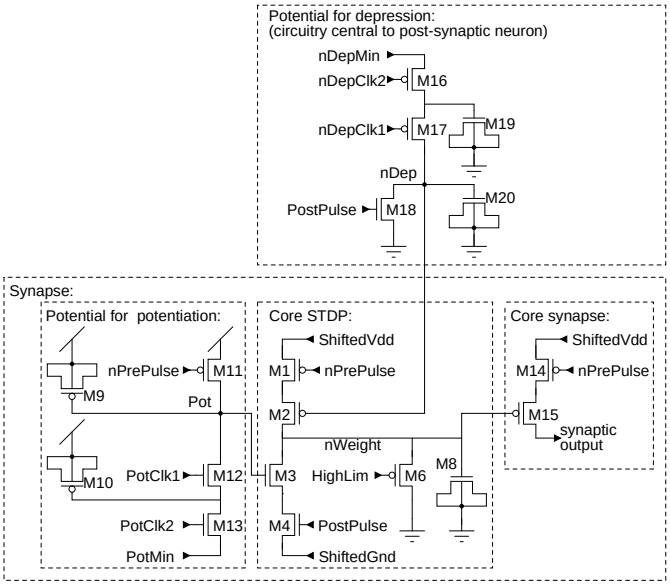


Figure 4: Synaptic STDP circuit. **Core STDP:** M1-8 is the core circuitry presented in fig. 2, where M5 is not present as a lower limit was considered unnecessary for the purposes of this system, and an nMOSCAP (M8) was chosen as the weight capacitance. **Potential for potentiation:** M9-13 implement a leaky integrator. M9 is a pMOSCAP which stores *Pot*; this is incremented (typically by several hundred mV) upon a brief pulse *nPrePulse*; the effect of one or several pre-synaptic spikes can be accumulated, parametrised by the pulse length. A switched-capacitor conductance (M10, M12-13) is used to leak *Pot* to *PotMin*. *PotClk1-2* provide non-overlapping pulses to control this leak, thus altering their frequency alters τ_+. Designing for clock rates of order 1 KHz allows ms-sensitive timing; clocks are ramped up and down slowly to minimise disruption to power rails. **Potential for depression:** M16-20 is a complementary leaky integrator of the same design; the resulting signal *nDep* is common to all the dendritic synapses of a neuron and thus is generated only once per neuron; its large capacitance and full range operation ensure that any capacitive coupling from other sources in its routing through the synaptic array have negligible effect. **Core synapse:** Upon the arrival of a pre-synaptic spike, M14 turns on, allowing M15 to deliver current to downstream circuitry, whose magnitude is modulated by *nWeight*. The downstream circuitry (not shown) consists of a further leaky integrator to produce a voltage representing synaptic conductance, which is used to pass current to a capacitor representing the neuron's membrane capacitance, as part of an implementation of an integrate-and-fire neuron.

towards linearising the effect of *nWeight* on synaptic efficacy, though these are not discussed here; the interested reader is directed to Bamford [4, chapter 3].

The leaks of the integrators are realised by switched capacitor conductances in this system. This particular design is however not essential to the novel circuitry presented in section II-A - a continuous leak could be substituted, as in Tanaka et al. [48]. The resting levels for the leaks, *nDepMin* and *PotMin* may be set around the nominal thresholds of M2-3 respectively providing truely exponential decays, or lower, thus reducing the residual weight change which can occur due to subthreshold currents on the arrival of isolated spikes, at the expense of giving the decay a slightly more

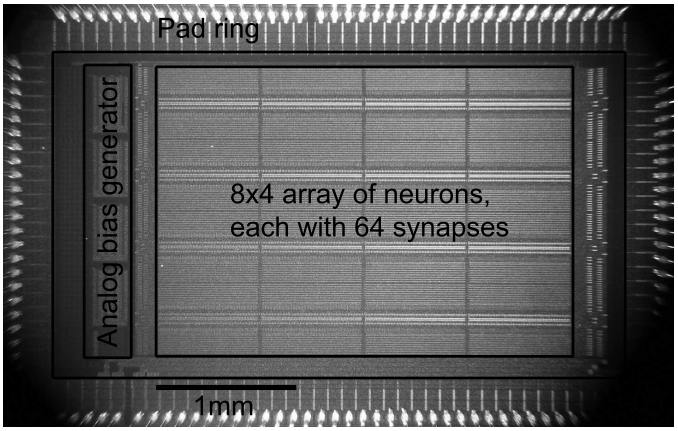


Figure 5: Die photo of 32 neuron $\times$ 64 synapse chip.

linear characteristic. Although time windows for plasticity are commonly modelled as antisymmetric exponential decays, this is not the only choice, either in biological modelling [43, figure 1] or in neuromorphic models [27] or for practical neuromorphic engineering purposes [13].

Following Bofill-i Petit and Murray [11], the leaky integrator for the potential for depression is situated in the central neuron circuitry and one copy of the circuit serves all dendritic synapses. By distributing a voltage rather than a current signal, the area requirement is reduced because there is no need for current mirrors at the synapses.

Although nPrePulse is shown as gating M1, M11 and M14, in fact these come from three separate pulse generators from global address-event delivery circuitry, so that their durations (always on the order of 10 ns) can be individually tuned, thus allowing the increments to *Pot*, the increments to *nWeight* and the magnitude of synaptic events to be controlled separately (likewise for the two separate applications of *PostPulse*, though each neuron contains its own pulse generators for these signals).

III. Results

The results in section III-A come from a test chip with a single copy of the circuit from fig. 2, where each of the three capacitive devices could be connected to or isolated from the *Weight* node by means of a transmission gate. Pulses were produced by a Xilinx Spartan 3 FPGA on an Opal Kelly XEM3010 integration board; analogue biases were produced by off-chip DACs and outputs were sampled by off-chip ADCs. The results in sections III-B and III-C come from a set up in which 8 chips were controlled by the FPGA and interconnected through AER. Each chip contained 32 neurons, where each neuron had 64 synapses (see fig. 5), containing the circuitry shown in fig. 4. The *nWeight* values of synapses could be sequentially buffered off-chip and sampled.

A. Weight dependence

Fig. 6(a) plots chip results for a sweep of potentiation and depression events similar to that in fig. 3, demonstrating good qualitative conformance to simulation, and showing the weight dependence which arises from the transistors' non-ideal behaviour.

Fig. 6(b) shows the effect of replacing the poly-poly capacitor with a pMOSCAP. The capacitance of a MOSCAP collapses if the gate-bulk voltage is in the region of its threshold voltage, therefore there is a bulge in the potentiation curves around 2.5-2.9 V as the effect of a potentiation event is to change the *Weight* voltage more whilst passing the same amount of charge. Depression events are larger in the same region. In fig. 6(c) an nMOSCAP was used instead. In this case the complementary effect occurs around 0.5 V, although it is less apparent on the graph because the threshold of an nMOS is closer to the power rail for the process used.

Fig. 7(a) shows the effect of using transistor M6 to impose a boundary. With *HighLim* set to 1.6 V, a *de facto* upper limit for *Weight* of 2.3 V is observed. As noted above, the precise position of the boundary can only be stated with reference to the time scale over which the system is used; for the results in this section, *Weight* was sampled $\approx$0.5 ms after the plasticity events. The effect of this upper limit is that there is much less weight dependence in potentiation than in depression. This is because for potentiation, the region in which strong weight dependence arises from transistor M2 operating in the linear region is almost completely excluded, whereas for depression, less of the region in which weak weight dependence arises from transistor M3 operating in saturation is included. Thus the learning rule used by Van Rossum et al. [49] is implemented.

Fig. 7(b) shows the results of the same experiment where a pMOSCAP is used. In this case the upper limit excluded most of the region in which the non-linearity of the capacitance has an effect, thus if a design is intended to operate in this region, it can benefit from the reduced area required without suffering from the non-linearity. In fact, a difference in behaviour between the two graphs can be observed as the beginning of the non-linearity is included, and the effect is to increase weight dependence of depression whilst reducing weight dependence of potentiation, since both potentiation and depression events have a slightly increased magnitude close to the boundary. As this may not be clear to the eye, normalised and rectified averages of the 5 sweeps with the greatest weight changes for each direction of plasticity from each graph were fitted to the function $y = x^{\mu}$ or $y = (1 - x)^{\mu}$ depending on the direction of plasticity, to find the parameter μ as in the formalism of Gutig et al. [22] (equation 3). Potentiation gives μ = 0.15 with poly-poly and μ = 0.09 with a pMOSCAP; depression gives μ = 0.54 with poly-poly and μ = 0.64 with a pMOSCAP. Thus the non-linearity of a MOSCAP can provide a desirable effect in this case.

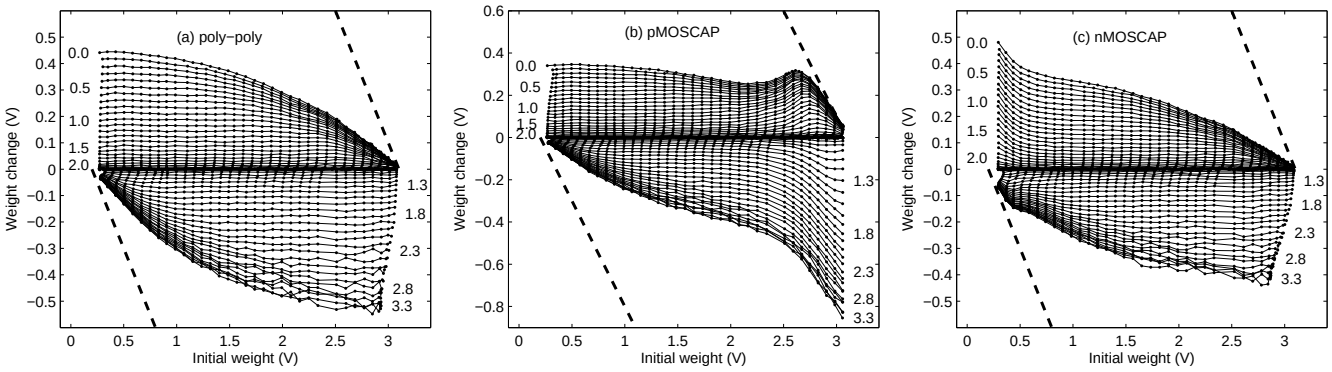


Figure 6: Weight dependence of plasticity: comparison of capacitor types. Each graph plots weight change against initial weight. Trials in which a negative-going pulse of ≈10 ns was applied to *nPostPulse* in the presence of an externally fixed level for *nPot* appear as data points in the top half of each graph (i.e. with positive weight change). data points which share the same level for *nPot* are linked together by lines. The beginnings of some of these lines are labelled with the actual voltage applied, e.g. 0.0, 0.5 etc. *nPot* was swept through the range 0.0-3.3 V with intervals of 0.05 V. Likewise, negative weight changes appear in the lower half of the graph, achieved by pulses to *PrePulse* of ≈2-3 ns in the presence of a certain *Dep*, which was swept through the same range and is similarly labelled. For each sweep, a range of initial weights was applied. The initial weights were set prior to each trial by lowering and raising *HighLim* and *LowLim* respectively; this method achieves different results in different regions of the whole voltage range, which is why inhomogeneities can be seen in the positions of initial weights. During each trial, *HighLim* and *LowLim* were set to 3.3 V and 0 V respectively. *ShiftedVdd* and *ShiftedGnd* were set to 3.1 V and 0.2 V, respectively, i.e. they were shifted inwards by 0.2 V with respect to *Vdd* and *Gnd*. Dashed lines mark theoretical limits for potentiation and depression. Each data point is the average of 5 trials; this averaging was to reduce the jitter apparent especially in the depression results due to the method used to create the short pulse. (a) The capacitor storing the weight value was a poly-poly capacitor; (b) the capacitor was pMOSCAP; (c) the capacitor was nMOSCAP.

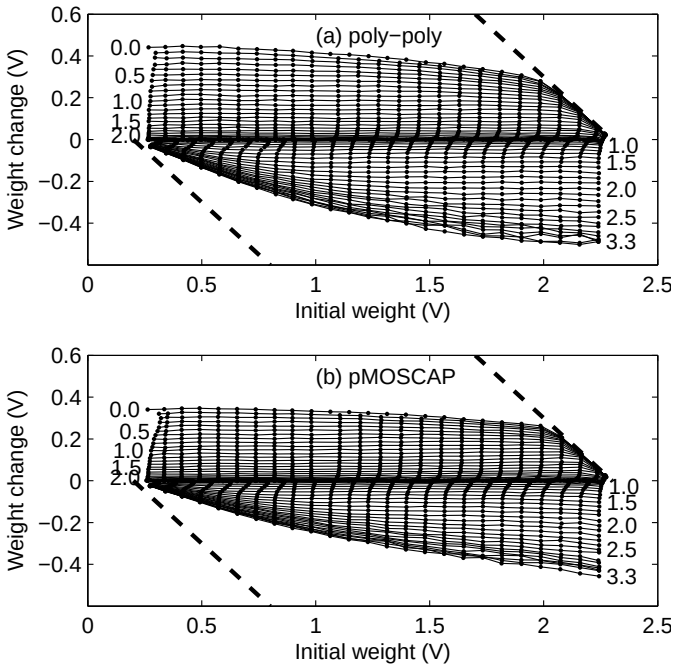
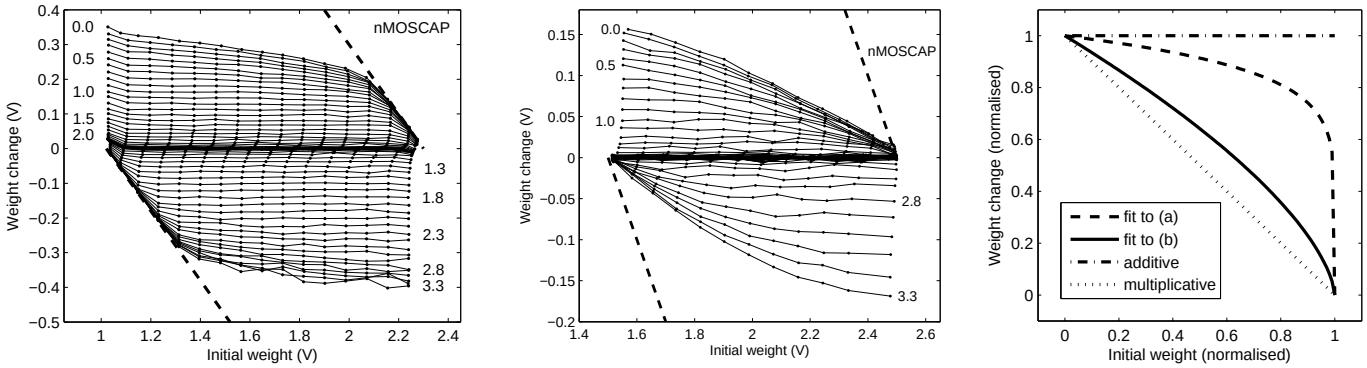


Figure 7: The effect of a shifted boundary. Results of weight sweeps as in fig. 6, except that *HighLim* was set to 1.6 V. Initial and final values for *Weight* were sampled ≈0.5 ms before and after plasticity events respectively. The right-most dashed line plots an upper limit of 2.3 V. (a) The capacitor storing the weight value was a poly-poly capacitor; (b) the capacitor was a pMOSCAP.

Fig. 8 compares the effect of the two different methods available for imposing boundaries. In fig. 8(a) *HighLim* and *LowLim* are both brought inwards; the effect is to achieve low weight dependence in both forms of plasticity. In fig. 8(b) *ShiftedVdd* and *ShiftedGnd* are both brought inwards. In this case, transistors M2 and M3 both operate mainly in their linear region and the effect is to achieve high weight dependence in both forms of plasticity. Although in both cases, the voltage range for *Weight* is reduced to a similar extent (≈1.3 V and 1 V respectively), the resulting learning rules are very different, with (a) approaching the additive learning rule in fig. 1(a) but (b) approaching the multiplicative rule in fig. 1(b). Clearly, neither the fully additive nor fully multiplicative case is achieved; each case represents a mixture which can be fitted to a point in the spectrum of mixtures offered by the formalism of Gutig et al. [22]. Performing these fits (method in the figure caption) gives values for μ of 0.13 and 0.64 respectively. These fitted functions are plotted in fig. 1(c), together with fully additive and fully multiplicative functions for comparison.

B. Weight stability

With *ShiftedGnd* and *ShiftedVdd* set to *Gnd* and *Vdd* (3.3 V) respectively for the circuit in fig. 4, *nWeight* discharges rapidly in the absence of plasticity events, at a rate (measured in the range 1.7 V down to 0.7 V) of ≈700 V/s; for an expected *nWeight* capacitance of ≈ 0.5 pF this implies a current of



(a) (b) (c)

Figure 8: Comparison of methods for imposing boundaries (nMOSCAPs were used in both experiments). (a) Experiment as in fig. 6, except that *HighLim* = 1.6 V and *LowLim* = 1.5 V; nominal lower and upper boundary lines plotted at 1 V and 2.3 V respectively (b) *ShiftedGnd* = 1.5 V and *ShiftedVdd* = 2.5 V; depression pulses were given the same duration as potentiation pulses ≈10 ns; (*HighLim* and *LowLim* were not used, i.e. they were set to the outer power rails). (c) In order to fit the data to the function in equation 3 in a way which respects its likely usage, Poisson spike trains were constructed for pre- and post-synaptic neurons, with 1000s of 1 Hz pre and post activity interspersed with 10 s of 100 Hz/1 Hz pre/post, 10 s of 1 Hz/100 Hz pre/post and 10 s of 100 Hz/100 Hz pre/post, in order to capture qualitatively the main firing regimes in which a synapse may find itself. These spike trains were used to construct traces for the potentials for plasticity, with $\tau_{+/-}$=20 ms, and the magnitudes of the increments and decrements to Dep and nPot respectively upon incoming spikes were set so that 5% of the potentials used for plasticity events would have been beyond the power rails had the limitation of the power rails not been present (increments and decrements were ≈1 V). The potentials were sampled at the moments of their respective plasticity events and this distribution was used to select a distribution of sweeps from the figures where each sweep was selected because the potential voltage which was used for the sweep was closest to that from the simulated sample. These sweeps were summed together for each direction of plasticity. These summed sweeps were then normalised in both dimensions, the data for depression was rectified and reversed and the two sweeps then summed together. Data points from fig. (a) affected by the imposed boundaries were excluded. The final sweep was fitted to the function $y = (1 - x)^{\mu}$ to yield fits to the data in (a) ($\mu = 0.13$) and (b) ($\mu = 0.64$); these are plotted, together with additive ($\mu = 0$) and multiplicative ($\mu = 1$) functions.

≈350 pA, larger even than the higher corner analysis provided by Linares-Barranco and Serrano-Gotarredona [33] for the same process. Discharging to *Gnd* is to be expected since a pMOSFET has a higher (negatively-defined) threshold with respect to *Vdd* than an nMOSFET of the same dimensions has with respect to *Gnd*; thus when a pMOSFET is gated by *Vdd* it is more deeply subthreshold than an nMOSFET gated by *Gnd*. When *ShiftedGnd* is raised to 0.2 V, and *ShiftedVdd* is lowered to 3.1 V, the rate at which *nWeight* changes is greatly reduced, as shown in fig. 9. Respectively raising and lowering these levels further does not decrease the rate of change any further, suggesting that reverse diode leakage to substrate becomes dominant at this point. Although learnt weight values begin to decay immediately, some trace of the learned memory is retained over several minutes.

C. Collective weight distributions

As described above in section II-B, the STDP circuit was used in a multi-chip system intended to investigate topographic map formation and receptive field development through synaptic rewiring. The role of STDP in that system was to detect correlations between neighbouring inputs in a topographic space, providing weight distributions which could then be used as a basis for synaptic rewiring. The general results of that system are presented in Bamford et al. [6], and it is those results to which the interested reader is directed as the basic demonstration of the ability of this circuit to act in the appropriate context as a correlation detector. In this paper some previously unpublished results from that system are presented which explore the weight distributions which it is capable of producing.

Spike streams were generated for two groups of virtual neurons and streamed to the multi-chip system. For each on-chip neuron, half of its dendritic synapses received spikes from each of these groups. For the first 240 s, spiking input consisted of simultaneous spikes from a randomly chosen 50% of the neurons in one group. after 20 ms there was another simultaneous set of spikes from neurons in the other input area. After a further 20ms this pattern repeated, with different randomly chosen neurons constituting the "flash" each time. Between these flashes there were Poisson-distributed spikes from all neurons in both input areas with the rate set so that the total spike rate per pre-synaptic neuron including spikes from flashes was 20 Hz. For a further 30 s afterwards spiking input consisted just of Poisson-distributed spikes from all neurons in both input areas with a fixed rate of 20 Hz. Thus, there was a longer period of input with intra-group but not inter-group correlations, followed by a shorter period of input with no correlations.

Time constants for STDP were $\tau_+ = \tau_- = 20$ ms, achieved by clock rates for *PotClk* and *nDepClk* of ≈ 750 Hz and 2100 Hz respectively (the clock rate for *nDep* being higher due to

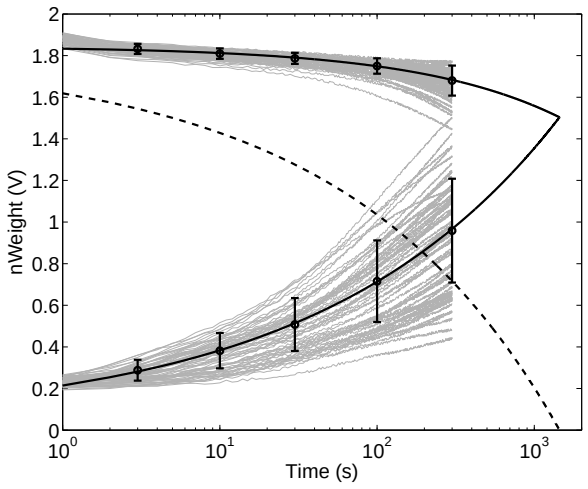


Figure 9: Weight stability. In all synapses in the multi-chip system, *nWeight* was initially set to *ShiftedGn*d = 0.2 V (whilst *ShiftedVdd* = 3.1 V and *HighLim* = 1.95 V). *nWeight* was then sampled for each synapse each second, up to 300s. A selection of these traces are shown for 64 synapses (grey lines). In a separate experiment, all *nWeight* voltages were initially raised to slightly beyond a nominal value for the upper boundary imposed by *HighLim* = 1.95 V. Sampling then proceeded as before and traces are shown for the same 64 synapses. Low synaptic weights rise at different rates whilst high weights fall at different rates, due to mismatch. The weight of each synapse asymptotes towards a resting level at which the currents into and out of the node balance. Simplistically, the trajectory of each *nWeight* can be modelled as an exponential decay. As an arithmetic average of exponential curves can be modelled with a power-law curve, best power-law fits of the rising and falling trends (over all synapses, not just those pictured) are shown and extended out to the point at which they cross over (solid black lines). Means and standard deviations are given at selected times for the weights of all synapses (not just those pictured) for each of the experiments, showing the quality of the fitting. While the rising and falling curve for an individual synapse will not cross but only converge, the distance between the rising and falling trend lines up to the crossing point (the dotted line) is indicative of the extent to which a weight will remain deflected away from its resting level after learning has occurred; thus the rate of reduction of this distance is indicative of the rate at which learnt memories decay. This distance decays to 90% of its maximum in 8 s, 75% in 43 s, 50% in 227 s and 0% in 24 mins. The relatively fast drop in *nWeight* around 1.9 V in the first few seconds is due to discharging through M6 in fig. 4.

greater capacitance on the output shared with 64 synapses); a_+ and a_- were adjusted *ad hoc* to achieve mid-range mean weights; synaptic events contributed an exponentially decaying kernel of increased membrane conductance with τ = 10 ms; ≈1.5 simultaneous spikes to maximum weight synapses would cause the input current to the IF neurons to peak at their rheobase current; neuron membranes decayed with τ = 20 ms.

Fig. 10 shows resulting weight distributions at specific points during this trial. The left column is a visualisation of the weight distributions over all the synapses of all the neurons. The central column is a supplementary view showing the mean weights for synapses for each neuron. The fact that these weights are distributed throughout the whole range indicates that the distributions seen in the left column are generally present within each neuron rather than being an aggregate effect across all neurons. During the trial, the weights, which were initially maximised, underwent depression until the neurons found their preferred level of activity (recalling the homeostatic regulation property mentioned in section I-A). This happened quickly, with the effect apparent even within the first second of the trial necessary to sample all the weights. The numbers to the right indicate the level of preference each neuron developed for one of the two groups of input neurons. For the preference of a neuron to be absolute, the distribution of its synaptic weights would have to be completely bimodal with all of the synapses connected to one group completely potentiated but all of the synapses connected to the other group completely depressed. During the intra-group correlations, a majority of neurons quickly developed strong group preferences, as indicated by the aforementioned group preference figures and the maps to the right (the presentation of the 256 neurons in a 16×16 grid is not important to this experiment as no topographic relations were imposed between neurons or inputs; the interested reader may wish to understand the complete context of these results in Bamford [4, chapter 6]). Strong preferences developed during the first 20-30 s. There was further slight improvement until 70 s after which the group preference measure did not change. After 240 s when the intra-group correlations in the input disappeared, the group preferences quickly collapsed until after 30 s there was not much evidence of them remaining.

During this process, the weight histograms to the left show that the strong correlational cues in the input caused strong bimodal divergence. Then, when the correlational cues were removed the average weight distribution reverted to a less clearly bimodal distribution.

D. Area and power

This section gives indicative figures for the circuit as laid out in the multi-chip system. The synaptic circuitry in fig. 4 (M1-15) had an area of ≈ 400 μm^2, of which ≈ 260 μm^2 is dedicated to transistors M8-9 which provide the capacitance for *nWeight* and *nPot* respectively. Of the area used by the core STDP circuitry consisting of M1-4, 6, and 8, more than half of this area is dedicated to M8. Thus using a MOSCAP instead of a poly-poly capacitor significantly reduced the area of the STDP circuit. Capacitances for *nWeight* (M8), *Pot* (M9) and *nDep* (M20) were designed at ≈ 0.5 pF, although for *nDep* there is a significant amount of parasitic capacitance due to it gating M2 64 times (one for each of the dendritic synapses of a neuron), raising this to ≈1.4 pF.

Considering energy costs of the STDP circuit, for the weight to change from its lowest to its highest level and back, *nWeight* must be completely charged and discharged; if boundaries are not restricted as in fig. 8, this would use ≈4 pJ (calculated using capacitances extracted from layout). For *Pot* to be

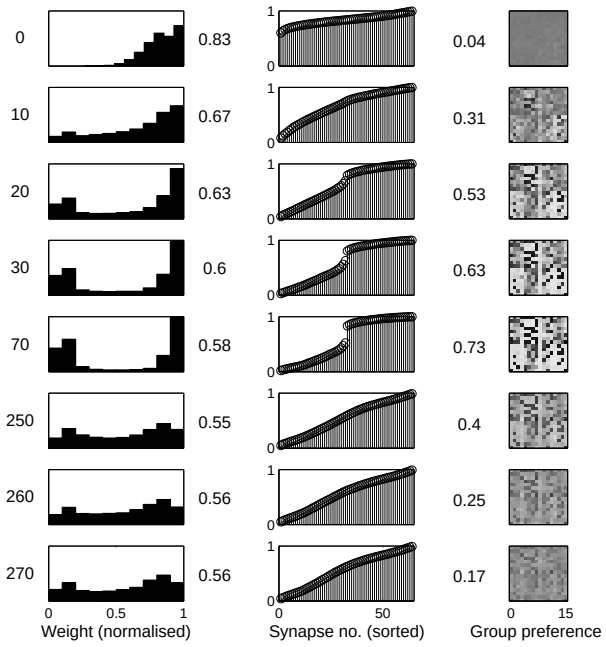


Figure 10: Relationship of weight distributions to detected correlations. Top to bottom: during the trial the weights of all 16384 synapses were sampled starting at the times (in seconds) labelled to the left of the leftmost column (a complete read-out took ≈1 s). Left: histogram of weights. The number to the right of each histogram is the mean normalised weight, where the range of *nWeight* is between *ShiftedGnd* = 0.2 V and the *de facto* maximum weight yielded by *HighLim* = 1.95 V. Centre: The mean weight (y axis) for the nth synapse (x axis) in each neuron, where the synapses in each neuron are sorted in ascending order of weights. Right: group preferences for neurons. Within each raster, each cell represents one on-chip neuron. The shade of the pixel gives the weight of the synapses connected to one of the groups as a proportion of the weight of all synapses on a scale from white to black. The number to the left of this visualisation of group preference is the mean of a related measure of group preference, defined as $\mathit{GroupPreference} = 2\left|\frac{\sum_i w_i^1}{\sum_i w_i} - \frac{1}{2}\right|$ where w_i is the weight of the *i*th synapse for the target neuron and w_i^1 is the weight of the *i*th synapse only for synapses with pre-synaptic neurons from one of the groups (an arbitrary choice). The measure therefore gives a value of 1 if group preferences are absolute, but 0 if there are no group preferences.

completely charged and discharged there is a similar energy cost. Additionally, every time there is a pre-synaptic spike, pulses must be generated for transistors M1 and M11. The cost of generating these pulses is shared amongst all axonal synapses of the pre-synaptic neuron, and the extra energy cost specific to a single synapse is the cost of completely charging and discharging the capacitance of these gates and the wires leading to them (≈ 300 fJ per synapse in this implementation). Similar costs are associated with transistors M4 and M18 upon post-synaptic spikes. For the switched capacitor implementation of exponential decays presented here there are also the costs of driving the clocks. The clocks are global and the additional cost per synapse is the cost of charging and discharging the gates M12-13 and the wires leading to them, which is ≈ 60 fJ (there is also an associated cost per neuron for *nDepClk*); a 1 KHz clock rate would mean a continuous power consumption of ≈ 60 pW per synapse.

IV. Discussion

A. Weight dependence

The STDP circuit presented bears comparison with several circuits which have been reviewed, most notably Bofill-i Petit and Murray [11] but without an explicit, active mechanism for weight-change reduction. Fig. 6(a) presented a weight dependence profile which is available from the circuit configuration of fig. 2. The weight dependence comes from physical constraints of the transistors, namely their inability to act as ideal current sources. The methods of controlling weight dependence then involve selecting regions of this profile by controlling the boundaries of the weight voltage. Fig. 8 presented the effects of two different ways to skew the learning rule towards or away from weight dependence. These two methods are: using overflow transistors to set boundaries (fig. 8a); and reducing the voltage range (fig. 8b). The fitted parameters show the effect on the learning rule in terms of the spectrum of possible learning rules presented by Gutig et al. [22], with the first method favouring weight independence and the second favouring weight dependence. These situations do not represent extremes of performance but are simply two examples. Greater weight independence could be achieved by reducing the average values for the potentials for plasticity; if a maximally additive rule were actually desired then it would be sufficient to limit *Dep* and *nPot* to levels such that transistors M2-3 in fig. 2 are operated in subthreshold, and to increase pulse durations accordingly. This has not been investigated in this paper however (it would simplify to a solution similar to that of Indiveri et al. [27]). A completely multiplicative rule cannot be achieved by this circuit, since the dependence of plasticity magnitude on *Weight* in the linear region is not linear. If strong weight dependence is required together with smaller plasticity events, this can be achieved to the extent that the duration of pulses can continue to be reduced; transistors could also be made longer up to the point at which compactness is compromised. The ability to control the degree and the form of weight dependence requires the ability to reinterpret the usable voltage range for *Weight* in subsequent circuitry; for example, for fig. 8a, it must be possible to interpret 1 V and 2.3 V as the minimum and maximum weight respectively, and interpolation should be linear.

The importance of MOSCAPs in reducing the area of the circuit was noted in section III-D, but their use affects the weight dependence profile, as shown in fig. 6(b-c). The region of the weight dependence profile affected by a MOSCAP can be avoided by either of the methods shown in fig. 8, and fig. 7 shows how the use of a pMOSCAP can actually help to provide a better match to the model of Van Rossum et al. [49].

Why might it be important to control weight dependence? It has been shown by Gutig et al. [22] that a moderate degree of weight dependence can improve the ability of STDP to act as a correlation detection mechanism. An intuition for why this may be the case is that with a completely additive learning rule there can be a great deal of pressure for the synapses to diverge into a bimodal distribution. In the absence of strong correlational cues as to which group of inputs should take control over the firing of the post-synaptic neuron, a bimodal distribution will form anyway based on spurious correlations [47, figure 2(a-b)]. Once such a distribution has formed its composition can be very stable [9], so that there is little chance of the composition of the winning group changing to favour correlations in the inputs. According to the analysis in Gutig et al. [22], as weight dependence of STDP increases, the effect on a stable distribution of weights of afferent synapses to a single post-synaptic neuron is that the poles of a bimodal distribution move closer together until a critical point is passed at which they form a unimodal distribution. There are many other factors that can influence the weight distribution, one of which is the strength of the correlational cues in the inputs. The results in fig. 10 show that the weights shift between a strongly bimodal distribution in the presence of strong correlational cues and a weakly bimodal distribution in their absence. To be clear, this is not a quantitative demonstration of increased sensitivity to correlations; it simply suggests that the behaviour of the system is skewed towards the bifurcation identified by Gutig et al. [22], with respect to the behaviour of a system with a fully additive learning rule, for which results such as those of Song et al. [47, figure 2(a-b)] may be expected. Regarding future work, one could consider further quantitative investigation, perhaps to the point of deriving formulae for STDP parameter setting by which optimal correlation detection performance could be achieved in given situations. However in the opinion of the authors, it may be ultimately more rewarding to search for complementary forms of homeostatic regulation which may automatically alter parameters to bring a system towards optimal performance.

B. Weight stability

The results in section III-B show that a straight-forward application of negative gate-source voltage for reducing subthreshold leakage current is effective in greatly reducing the speed at which a capacitor drifts towards a voltage level towards which it is predisposed. Issues of weight stability and weight dependence are treated together in this paper for two reasons. Firstly there is the practical reason that the same mechanism that reduces the leak from the capacitors (shifted inner power rails) can also be used to increase weight dependence, as shown in fig. 8(b). Beyond this, there is a trade-off between the weight stability offered by the inherent bistability of a weight-independent learning rule, and the increased sensitivity to correlations offered by the introduction of a degree of weight dependence. Even with the reduction of capacitor leakage, the synapses still lose their learnt weights over a time scale of only minutes, which is by no means biologically realistic. This nevertheless allows a much greater time scale in which synapses can accumulate evidence for correlations before this memory is either wiped away or consolidated by some other mechanism. In smaller processes leakage currents will become higher so that, although the technique of back-biasing the gate will continue to give an advantage w.r.t. a design such as Indiveri et al. [27], the time scale over which a weight can be retained for the same capacitance may be greatly reduced. Lower capacitance available in scaled-down circuits will also reduce time constants. Both of these scaling effects will make complementary stabilisation mechanisms more desirable. This technique for leakage reduction is compatible with some of the other mechanisms reviewed in section I-D; for example, a bistability mechanism could be added [18], or a multistability mechanism [25]; alternatively, weight distributions can be used to influence network topology, which can be stable over a much longer time scale [6].

C. Area and power

Energy consumption figures have been given in section III-D. For the core circuit, the energy consumption is based on the charging and discharging of the capacitance representing weight. However the rate and magnitude of weight changes is highly dependent on the parametrisation of and inputs to the system; for example, there are 3 orders of magnitude difference in values for $a_{+/-}$ used by Song et al. [47] and Young et al. [53], which would change this element of the power consumption by 3 orders of magnitude; given such uncertainties, energy comparisons such as those in Zhang et al. [55, table 1] may be misleading. Comparing to the subthreshold design of Indiveri et al. [27], although this circuit operates instead in strong inversion, this does not increase energy consumption since much higher currents are simply used for much shorter periods. Full range operation of *Weight* should increase overall energy usage by only a small factor but if this range is reduced as in fig. 8b then energy consumption is reduced *(Dep* and *nPot* operate across the greater part of the range in both designs). The choice of a switched-capacitor implementation for exponential decays of *nDep* and *Pot* is not ideal for power consumption, but this choice is incidental to the novel circuitry presented in this paper.

In comparison to Bofill-i Petit and Murray [11], which also provides for a tunable degree of weight dependence, this circuit does not require explicit circuitry for reducing weight change magnitude and therefore saves on the area and power costs associated with that circuitry. The overflow transistors have an energy cost where they are employed, since increments and decrements beyond the boundaries they impose are leaked away. This, however, occurs only when boundaries are exceed, rather than having a continuous cost (the bistability mechanism of Indiveri et al. [27] also imposes boundaries and has a continuous cost).

D. Mismatch

This circuit is affected by mismatch in the following ways. Transistors M2-3 are mismatched between synapses, altering

the magnitudes $a_{+/-}$; as the effect of these is dominated by high (super-threshold) gate voltages, the effect of mismatch can be expected to be lower than that in the corresponding transistors in the design of Indiveri et al. [27]. Transistors M5-6 are mismatched between synapses, affecting the level of any imposed boundaries. The timing of external pulse generators is based on transistor current sources, thus the pulse lengths can also be expected to vary, again affecting the magnitudes $a_{+/-}$. For the multi-chip implementation, the pre-synaptic pulses were in fact produced by a single generator for each chip, so calibration could be performed if desired. The post-synaptic pulses by contrast were produced by circuits specific to each neuron, and so there is a contribution to the mismatch of magnitudes $a_{+/-}$ which varies between neurons. As pulses need to be short, the transistors which control the timing may be biased above threshold, potentially reducing the effect of mismatch compared to a subthreshold design requiring long pulses. Mismatch in leak conductances, whether in the switched capacitor implementation presented here or in a continuous implementation [11, 27] affects time constants $\tau_{+/-}$. The cumulative effect of all these sources of mismatch is that the critical parameter $\beta = a_-\tau_-/a_+\tau_+$ identified by Song et al. [47] varies between synapses, with the effect that some are more prone to depression than others. For the multi-chip system, in an experiment in which β was 3.8, the coefficient of variation of β amongst synapses was 0.40 (based on weight change events recorded simultaneously from the same synapse from each of a set of 16 neurons distributed across 4 of the 8 chips, where the neurons were configured identically and given identical inputs (feed-forward only) so that any differences in their behaviour would be attributable to mismatch and a small amount of noise). It is difficult to assess the importance of this in the absence of clear applications for such circuitry, but it has been shown that the aggregate effect of STDP can be to ameliorate mismatch; Cameron et al. [13] demonstrated the use of STDP as an engineering solution, and Bamford [4, section 3.5.7 p. 88] gives a preliminary account of how the STDP circuits in the multi-chip system presented here can act to compensate for mismatch in downstream neural circuitry.

Conclusion

An analogue VLSI circuit has been presented with a compact implementation of STDP suitable for parallel integration in large synaptic arrays. In contrast to previously published STDP circuits, it uses the limitations of the silicon substrate to achieve various forms and degrees of weight dependence of STDP. It also uses reverse-biased transistors to reduce leakage of a capacitance representing weight. Chip results have been presented showing: various ways in which the learning rule may be shaped; how synaptic weights may retain some indication of their learned values over periods of minutes; and how distributions of weights for synapses convergent on single neurons may shift between more or less extreme bimodality according to the strength of correlational cues in their inputs.

Acknowledgements

The work on the multi-chip system was funded by EPSRC and MRC via the University of Edinburgh Doctoral Training Centre for Neuroinformatics. The test circuit was included on a chip produced at Istituto Superiore di Sanità for the ReNaChip EC project grant agreement no.: 216809. Adria Bofill-i-Petit and Giacomo Indiveri both passed on schematics and layout for their synapse designs; other circuitry used but not described here was acquired from others at INI Zurich. The authors are grateful to Katherine Cameron and to many people at the University of Edinburgh Institutes of Integrated Micro and Nano Systems and Adaptive and Neural Computation, and at the 2007 Telluride Neuromorphic Engineering Workshop.

References

[1] T Aihara, Y Abiru, Y Yamazaki, H Watanabe, Y Fukushima, and M Tsukada. The relation between spike-timing-dependent plasticity and Ca2+ dynamics in the hippocampal CA1 network. *Neuroscience*, 145:80–87, 2007.

[2] JV Arthur and K Boahen. Learning in silicon: timing is everything. In *Advances in Neural Information Processing Systems*, 2005.

[3] D Badoni, M Giulioni, and V Dante. An aVLSI recurrent network of spiking neurons with reconfigurable and plastic synapses. In *ISCAS*, 2006.

[4] SA Bamford. *Synaptic rewiring in neuromorphic VLSI for topographic map formation*. PhD thesis, University of Edinburgh (available at www.sim.me.uk/neural/thesis.pdf), 2009.

[5] SA Bamford, AF Murray, and DJ Willshaw. Synaptic rewiring for topographic map formation and receptive field development. *Neural Networks*, 23:517–527, 2010.

[6] SA Bamford, AF Murray, and DJ Willshaw. Large developing receptive fields using a distributed and locally reprogrammable address-event receiver. *Neural Networks, IEEE Transactions on*, 21:286–304, 2010.

[7] CC Bell, VZ Han, Y Sugawara, and K Grant. Synaptic plasticity in a cerebellum-like structure depends on temporal order. *Nature*, 387:278–281, 1997.

[8] GQ Bi and MM Poo. Synaptic modifications in cultured hippocampal neurons: dependence on spike timing, synaptic strength, and postsynaptic cell type. *J. Neurosci.*, 18:10464–10472, 1998.

[9] G Billings and MCW van Rossum. Memory retention and spike-timing-dependent plasticity. *Journal of Neurophysiology*, 101:2775–2788, 2009.

[10] A Bofill-i Petit. *An analogue VLSI study of temporally-asymmetric Hebbian learning*. PhD thesis, University of Edinburgh, 2005.

[11] A Bofill-i Petit and AF Murray. Synchrony detection and amplification by silicon neurons with STDP synapses. *Neural Networks, IEEE Transactions on*, 15:1296–1304, 2004.

[12] JM Brader, W Senn, and S Fusi. Learning real world stimuli in a neural network with spike-driven synaptic dynamics. *Neural computation*, 19(11):2881–2912, 2007.

[13] K Cameron, V Boonsobhak, A Murray, and D Renshaw. Spike Timing Dependent Plasticity (STDP) can ameliorate process variations in neuromorphic VLSI. *IEEE Transactions on Neural Networks*, 16:1626–1637, 2005.

[14] A Cassidy, AG Andreou, and J Georgiou. A combinational digital logic approach to STDP. In *IEEE International Symposium on Circuits and Systems (ISCAS)*, pages 673–676, 2011.

[15] C Clopath and W Gerstner. Voltage and spike timing interact in STDP - a unified model. *Frontiers in Synaptic Neuroscience*, 2, 2010.

[16] AP Davison and Y Fregnac. Learning cross-modal spatial transformations through spike-timing-dependent plasticity. *J. Neurosci.*, 26:5604–5615, 2006.

[17] S Eberhardt, T Duong, and A Thakoor. Design of parallel hardware neural network systems from custom analog VLSI building block chips. In *Artificial Neural Networks, International Joint Conference on*, volume 2, pages 183–190, 1989.

[18] S Fusi, M Annunziato, D Badoni, A Salamon, and DJ Amit. Spike-driven synaptic plasticity: theory, simulation, VLSI implementation. *Neural Computation*, 12: 2227–2258, 2000.

[19] W Gerstner, R Kempter, JL van Hemmen, and H Wagner. A neuronal learning rule for sub-millisecond temporal coding. *Nature*, 383:76–78, 1996.

[20] DH Goldberg, G Cauwenberghs, and AG Andreou. Probabilistic synaptic weighting in a reconfigurable network of VLSI integrate-and-fire neurons. *Neural Networks*, 14: 781–793, 2001.

[21] C Gordon and P Hasler. Biological learning modeled in an adaptive floating-gate system. *IEEE International Symposium on Circuits and Systems*, 2002.

[22] R Gutig, R Aharonov, S Rotter, and H Sompolinsky. Learning input correlations through nonlinear temporally asymmetric Hebbian plasticity. *Journal of Neuroscience*, 23(9):3697–3714, 2003.

[23] R Guyonneau, R Van Rullen, and SJ Thorpe. Neurons tune to the earliest spikes through STDP. *Neural Computation*, 17:859–879, 2005.

[24] P Häfliger. Adaptive WTA with an analog VLSI neuromorphic learning chip. *IEEE Transactions on Neural Networks*, 18(2):551–572, 2007.

[25] P Häfliger and HK Riis. A multi-level static memory cell. In *IEEE International Symposium on Circuits and Systems*, 2003.

[26] P Häfliger, M Mahowald, and L Watts. A spike-based learning neuron in analog VLSI. In MC Mozer, MI Jordan, and T Petsche, editors, *Advances in Neural Information Processing Systems*, volume 9, page 692, 1997.

[27] G Indiveri, E Chicca, and R Douglas. A VLSI array of low-power spiking neurons and bistable synapses with spike-timing-dependent plasticity. *IEEE Transactions on Neural Networks*, 17:211–221, 2006.

[28] SH Jo, T Chang, I Ebong, BB Bhadviya, P Mazumder, and W Lu. Nanoscale memristor device as synapse in neuromorphic systems. *Nano Letters*, 10:1297–1301, 2010.

[29] JK Jun and DZ Jin. Development of neural circuitry for precise temporal sequences through spontaneous activity, axon remodeling, and synaptic plasticity. *PLoS One*, 8, 2007.

[30] WM Kistler and JL Van Hemmen. Modeling synaptic plasticity in conjunction with the timing of pre- and postsynaptic action potentials. *Neural Computation*, 12: 385–405, 2000.

[31] TJ Koickal, A Hamilton, SL Tan, JA Covington, JW Gardner, and TC Pearce. Analog VLSI circuit implementation of an adaptive neuromorphic olfaction chip. *IEEE Trans. Circuits and Systems*, 54(1):60–73, 2007.

[32] JJ Letzkus, BM Kampa, and GJ Stuart. Learning rules for spike-timing-dependent plasticity depend on dendritic synapse location. *The Journal of Neuroscience*, 26: 10420–10429, 2006.

[33] B Linares-Barranco and T Serrano-Gotarredona. On the design and characterization of femtoampere current-mode circuits. *IEEE Journal of Solid-State Circuits*, 38 (8):1353–1363, 2003.

[34] S Liu and R Mockel. Temporally learning floating-gate VLSI synapses. In *International Symposium on Circuits and Systems*, pages 2154–2157, 2008.

[35] RC Malenka and MF Bear. LTP and LTD: an embarrassment of riches. *Neuron*, 44:5–21, 2004.

[36] H Markram, J Lubke, M Frotscher, and B Sakmann. Regulation of synaptic efficacy by coincidence of postsynaptic APs and EPSPs. *Science*, 275:213–215, 1997.

[37] H Markram, W Gerstner, and PJ Sjöström. A history of spike-timing-dependent plasticity. *Frontiers in synaptic neuroscience*, 3, 2011.

[38] T Masquelier and SJ Thorpe. Unsupervised learning of visual features through spike-timing-dependent plasticity. *PLoS Computational Biology*, 3:e31, 2007.

[39] S Mitra, S Fusi, and G Indiveri. Real–time classification of complex patterns using spike-based learning in neuromorphic VLSI. *IEEE Transactions on Biomedical Circuits and Systems*, 3(1):32–42, 2009.

[40] A Morrison, M Diesmann, and W Gerstner. Phenomenological models of synaptic plasticity based on spike timing. *Biological Cybernetics*, 98:459–478, 2008.

[41] J Pfister and W Gerstner. Triplets of spikes in a model of spike timingdependent plasticity. *The Journal of Neuroscience*, 26:9673–9682, 2006.

[42] S Ramakrishnan, PE Hasler, and C Gordon. Floating gate synapses with spike-time-dependent plasticity. *IEEE Transactions on Biomedical Circuits and Systems*, 5:244–252, 2011.

[43] PD Roberts and CC Bell. Spike-timing-dependent synaptic plasticity in biological systems. *Biological Cybernetics*, 87:392–403, 2002.

[44] J Schemmel, A Grubl, K Meier, and E Mueller. Implementing synaptic plasticity in a VLSI spiking neural network model. In *Neural Networks, International Joint Conference on*, 2006.

[45] W Senn, H Markram, and M Tsodyks. An algorithm for

modifying neurotransmitter release probability based on pre- and postsynaptic spike timing. *Neural Computation*, 13:35–67, 2001.

[46] S Song and LF Abbott. Cortical development and remapping through spike-timing-dependent plasticity. *Neuron*, 32:339–350, Oct 2001.

[47] S Song, KD Miller, and LF Abbott. Competitive Hebbian learning through spike-timing-dependent synaptic plasticity. *Nature*, 3:919–926, Sept 2000.

[48] H Tanaka, T Morie, and K Aihara. A CMOS spiking neural network circuit with symmetric/asymmetric STDP function. *IEICE Transactions on Fundamentals of Electronics, Communications and Computer Sciences*, 92:1690–1698, 2009.

[49] MCW Van Rossum, GQ Bi, and GG Turrigiano. Stable hebbian learning from spike-timing-dependent plasticity. *Journal of Neuroscience*, 20:8812–8821, 2000.

[50] E Vittoz, H Oguey, MA Maher, O Nys, E Dijkstra, and M Chevroulet. *Introduction to VLSI-Design of Neural Networks*, chapter Analog storage of adjustable synaptic weights, pages 47–63. Kluwer Academic Publ., 1991.

[51] RJ Vogelstein, F Tenore, R Philipp, MS Adlerstein, DH Goldberg, and G Cauwenberghs. Spike-timing-dependent plasticity in the address domain. In S. Thrun S. Becker and K. Obermayer, editors, *Advances in Neural Information Processing Systems*, number 15, Cambridge, MA, 2003. MIT Press.

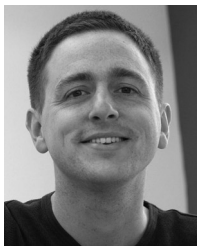
[52] RJ Vogelstein, U Mallik, JT Vogelstein, and G Cauwenberghs. Dynamically reconfigurable silicon array of spiking neurons with conductance-based synapses. *IEEE Transactions on Neural Networks*, 18:253–265, 2007.

[53] JM Young, WJ Waleszczyk, C Wang, MB Calford, B Dreher, and K Obermayer. Cortical reorganisation consistent with spike timing- but not correlation-dependent plasticity. *Nature Neuroscience*, 10:887–895, July 2007.

[54] Carlos Zamarreno-Ramos, Luis A. Camunas-Mesa, Jose A. Perez-Carrasco, Timothee Masquelier, Teresa Serrano-Gotarredona, and Bernabe Linares-Barranco. On spike-timing-dependent plasticity, memristive devices, and building a self-learning visual cortex. *Frontiers in Neuroscience*, 5, 2011.

[55] L Zhang, Q Lai, and Y Chen. Configurable neural phase shifter with spike-timing-dependent plasticity. *IEEE Electron Device Letters*, 31:716–718, 2010.

[56] LI Zhang, HW Tao, CE Holt, WA Harris, and MM Poo. A critical window for cooperation and competition among developing retinotectal synapses. *Nature*, 395: 37–44, 1998.



Simeon A. Bamford received a BA Hons in Artificial Intelligence in 1995 from the School of Cognitive and Computing Sciences at the University of Sussex. After an entrepreneurial career he returned to study and in 2009 received a PhD in Neuromorphic Engineering from the Neuroinformatics Doctoral Training Centre at the University of Edinburgh. He currently works with the Complex Systems Modelling Group at Istituto Superiore di Sanità, Rome.

Alan F. Murray is Professor of Neural Electronics and Head of the School of Engineering at the University of Edinburgh. He introduced the Pulse Stream method for analogue neural VLSI in 1985. Murray's interests are now in (a) biologically-inspired computational forms (particularly in VLSI hardware), where noise and overt temporal behaviour are important, and (b) direct interaction between silicon and real neuronal cells and networks. Murray is a Fellow of HEA, IET, IEEE and the Royal Society of Edinburgh and has published over 300 academic papers.

David J. Willshaw is Professor of Computational Neurobiology at the University of Edinburgh, UK. He has a career in neural networks and computational neuroscience stretching back over 30 years with over 100 scientific papers. In 1992 he was the recipient of the IEEE Neural Networks Council 1992 Pioneer Award. He has worked in a variety of research areas including associative memory, novel algorithms for combinatorial optimisation (where with Richard Durbin he developed the Elastic Net algorithm for the TSP) and the development of patterned nerve connections in the visual and neuromuscular systems, which is his current research interest. He is past Editor-in-Chief of the computational neuroscience journal Network: computation in neural systems. Since 1984 he has held long term research support from the UK Medical Research Council and the Wellcome Trust. In addition he is the grantholder of the Edinburgh Doctoral Training Centre in Neuroinformatics and Computational Neuroscience funded by UK Research Councils. Since 2002, this Centre has trained over 50 PhD students from the physical and informational sciences who are applying quantitative approaches to neuroscience and to neurally inspired computing.