Learning Bidirectional Connections Between Areas With Standard Spike-Timing-Dependent Plasticity

M. Klein*, A. Cangelosi, and T. Wennekers

Centre for Robotics and Neural Systems, Plymouth University,
Portland Square, Plymouth PL4 8AA, UK
**E-mail: michael.q.klein@gmail.com*
www.plymouth.ac.uk

Spike-timing-dependent plasticity (STDP) is a learning algorithm that is simple, biologically plausible, and powerful. Hence, one would expect STDP (likely in combination with other learning algorithms) to be a key component in cortical models of higher cognitive functions, such as language comprehension or production. Such models would need to involve multiple brain areas and bidirectional links between representations in those different areas. However, STDP is an asymmetrical learning algorithm (in contrast to classical Hebbian learning, which is symmetrical). This makes the acquisition of bilateral connections between two neurons almost impossible and bilateral connections between representations very challenging. Here, we propose a solution based on specific connectivity patterns. Then, using numerical simulations, we show that our approach allows STDP to create strong bidirectional links between representations. Finally, we compare our architecture to neuroanatomical data.

Keywords: spike-timing-dependent plasticity; STDP; Hebbian learning; unsupervised learning; cortex; bilateral connections.

1. Introduction

It would be vastly inefficient for the brain to use a completely different set of representations for language production and language comprehension. Under the assumption that word *meaning* is represented somewhere in the brain (localised or distributed) and word *forms* are represented somewhere else in the brain (perhaps as some form of auditory-motor patterns), we can deduct that these two types of representations must have some sort of two-way connection. The reason is: not only must a word be able to activate its meaning (in language comprehension), but meaning must also be able to activate an appropriate word (in language production). This principle is illustrated in Seidenberg and McClelland's *triangular model*[1] and Plaut and Kello's model of phonology[2]. These classic connectionist mod-

els have great value in understanding distributed cognitive processing and the relation between neural processing and psycholinguistics. However, the flood of functional imaging data on language processing might, at this point, benefit significantly from explanatory models that take more neuroanatomical constraints into account. The increased understanding of physiological mechanisms of learning in the brain[3,4] would allow such models to use biologically plausible learning rules. Also, the increased computational power made available by highly parallel neuromorphic hardware such as SpiNNaker[5] is predicted to allow neural models with up to 1 billion spiking neurons to run in real time in the near future. Hence, modelling language processing in the brain with computational units that corresponds to the single neuron in the real brain moves within the realm of possibilities.

Most language processing has been demonstrated to take place in the cerebral cortex[6]. There is convincing evidence that connections in the cerebral cortex are formed by unsupervised learning of the Hebbian[7] type. STDP, being a form of Hebbian learning for spiking neurons, is increasingly well established as at least one of the learning rules of the cortex[4,8]. Hence, it would be a strong candidate for the learning rule of such a model. *Standard* STDP, however, has the well established problem of not being able to create bilateral connections between two neurons[9,10].

This is due to the very nature of STDP. If two neurons have synapses in both directions, any event that will potentiate the synapses in one direction will depress the synapse in the other direction (as will be explained in more detail in section 3). While it is possible to change parameters of the standard learning rules, specifically to increase the strength of potentiation so that it grossly out-weights the strength of depression, such a change will lead to an undesired increase of most weights, leading to strong links between every neuron in the long run. Therefore, to assure long-term stability of firing rates, parameters in standard STDP rules are typically chosen such that inhibition slightly dominates excitation[11].

There are, however, non-standard forms of STDP obtained by modifying the STDP learning rule. One of those has been demonstrated to allow the establishment of bilateral connections at high firing rates, while at low firing rates only unilateral connections are established[10]. We estimated that this approach is likely to be problematic in multi-area models where intermediate layers have no direct external input. Too high frequencies in those areas might lead to uncontrolled associations of neurons with all other neurons. Therefore, we would like to introduce an approach that uses the more established (frequency independent) standard STDP learning rule. We propose

an architecture that allows patterns to be associated between two brain areas by using sparse area to area connections. Hence, while representations in different brain areas are connected in both directions, hardly any pairs of neurons are. To completely activate representational patterns of an area recurrent within-area connections are used. Note that once a multi-layer assembly is learned, single spikes can nevertheless travel quickly through the layers for fast-forwards visual recognition[12]. The recurrent connections are then only needed to reactivate the complete assembly.

2. Spike-Timing Dependent Plasticity

Within the last decade, research on STDP has grown several theoretical branches. In general, the proposed plasticity rules share the feature that the weight changes are dependent on the temporal relation between the pre- and the post-synaptic spikes, specifically: the weight increases when the pre-synaptic spike precedes the post-synaptic spike and decreases if the post-synaptic spike precedes the pre-synaptic spike (although the opposite has been proposed as well). The closer the two spikes are in time, the stronger the increase or decrease of the weight.

They differ mainly along two dimensions: The first dimension is the way the weight change is (or is not) dependent on the current weight. This dependence can be additive[8] as in standard STDP, multiplicative[13] or exponential[14]. The second dimension is which pairs (or even triplets[10]) contribute to plasticity. The most prominent types of pairing being *all-to-all*[15] and *nearest neighbour*[16]. All-to-all means that in case of a spike event, the STDP learning rule is applied to that event in relation to all previous spiking events. Of course, a post-synaptic spike is only paired with all pre-synaptic spikes and a pre-synaptic spike is only paired with all post-synaptic spikes. *Nearest neighbour* means that the spike event is only paired with the last preceding spike.

The simulations reported in this paper are all based on what one may call *standard* STPD learning rule[8], i.e., additive weight dependence, and all-to-all spike pairing. Therefore, weight change Δw is calculated with equation 2.

$$\Delta w = \begin{cases} A_+ exp(\Delta t/\tau_+) & \text{if } \Delta t < 0 \\ -A_- exp(-\Delta t/\tau_-) & \text{if } \Delta t \geq 0 \end{cases} \quad (1)$$

Here A_+ and $-A_-$ are the learning rates for depression and excitation, respectively. Since this is standard additive STPD this learning rate is not

dependent on the previous weight in any way. Hence, unless the weight has an upper and lower bound, it can grow or decrease indefinitely. In our simulation, the upper limit of the weight is 1 and the lower limit is 0.

3. Difficulty of Creating Bilateral Connections with STDP

Note that when we talk about the difficulty in the formation of bilateral connections between two neurons, we do not talk about whether or not there are synapses in both directions, but about the weight of any synapses present. STDP makes it close to impossible to develop significantly strong excitatory weights in both directions. This is because STDP is basically a mechanism for detecting causality. If a neuron A fires before a neuron B the synapse connecting A to B is potentiated. The synapse has learned that the event E_A that caused A to fire preceded the event E_B that caused B to fire. The strength of the STDP learning rule is that it detects such a repetitive directed sequence of events. Note, however, that whenever A fires before B and the weight W_{AB} is potentiated, B also fires after A and the weight W_{BA} is depressed. This means that, if STDP is used as a learning algorithm and we have two neurons with synapses in both directions, only one synapse can develop a strong weight, while the other will be depressed.

There is, however, a parameter called the α-parameter which determines the relation between the weight increase and the weight decrease. It is defined in equation 2.

$$\alpha = \frac{A_- \tau_-}{A_+ \tau_+} \tag{2}$$

If the α parameter is set to a value greater than 1, the weights tend to go down, in particular if the firing of the pre- and post-synaptic neuron is random (noise)[11]. If the α parameter is set to a value smaller than 1, weights tend to increase. Hence, it is possible to set the α parameter of standard STDP to a value, in which the strength of potentiation in both directions grossly out-weights the strength of depression. In this way, strong bidirectional links can be formed between two neurons. Such a change, however, will in the long run lead to an undesired increase of most weights, leading to strong links between all neurons in the model. It can, therefore, be ruled out as a solution.

4. Our Approach

We have established that under a reasonable parameter setting, strong bilateral connections between two neurons cannot be established in a model that learns with standard STDP. However, it would of course be possible to connect two brain areas using two distinct hidden layers, as shown in Fig. 1A.

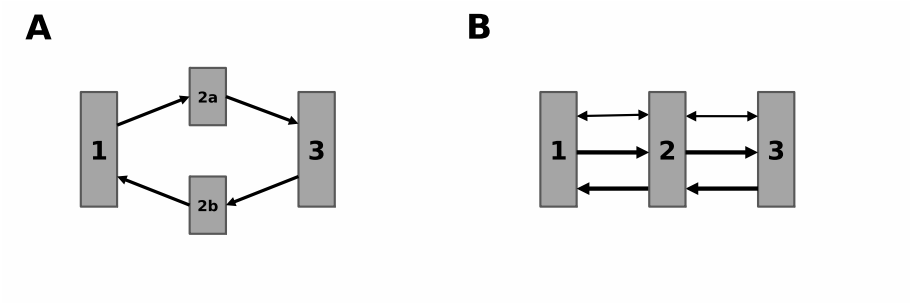

Fig. 1. A: If the intermediate layer is split into two *zones* 2a and 2b, then bilateral connections between two neurons would not be necessary to connect the two areas 1 and 3. B: Sparse probabilistic connections between areas lead to few bilaterally connected neuron pairs and many unilateral connections.

While this is certainly a valid way of connecting brain areas (and is definitely worth further exploration), two distinct hidden layers are needed. The key feature of such an architecture is the following: The outward connection of a neuron (the connection going to the next area) go to a different neuron than the inward connections (the connection coming from the next area). Consequently, there are no direct bidirectional links between neurons of one area with neurons of the other area.

This property can be approximated without the need for two distinct hidden layers by the usage of probabilistic sparse connections between two areas. If the connections are sparse and the probability of connections is symmetrical (i.e., the same probability in both directions), the chance of bidirectional synaptic links between neurons decreases faster than the probability of the connection in one direction (see Fig. 1B). This means that the number of connections in one direction can still be fairly high, while the percentage of connections that are bidirectional is very small.

This is computed in the following manner: with a connection probability of p_{sym} of a neuron n_A in area A to project to a neuron in area B and the

same probability of a neuron n_B in area B to project to a neuron in area A, the probability of neurons n_A and n_B to have a direct bilateral connection is p_{sym}^2. Consequently, the probability of having a unilateral connection in only one (but any) direction is $p_{sym} * (1 - p_{sym}) + p_{sym} * (1 - p_{sym})$, which is the sum of the probabilities of being connected in one, but not in the other direction. For example, if area A is connected to area B with $p = 0.25$ and area B is connected to area A with the same probability, then the probability of two neurons of these two brain areas to be bilaterally connected is only $p = 0.06$, while the probability of two neurons to be connected unilaterally is $p = 0.38$ (0.19 upwards and 0.19 downwards). Fig. 2 shows the relation between symmetrical probability and the probability of bilateral connections.

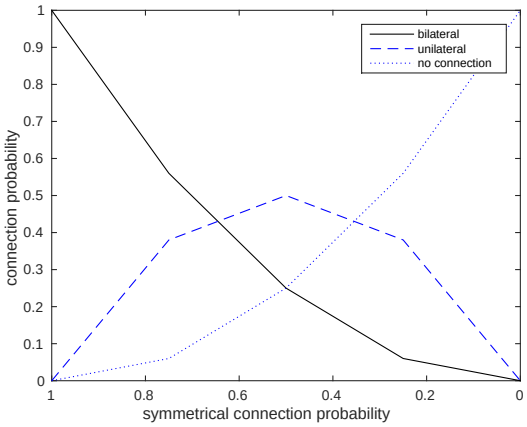

Fig. 2. Effect of symmetric probability of connections on probability of bilaterally connected neuron pairs. If two areas are connected with a symmetrical probability in both directions the probability of bilateral connections between two neurons (one in each area) decreases very fast. When the symmetric probability is in the mid to low range, the probability of bilateral connection is very low, while there are still sufficient unilateral connections. Note that x-axis of the graph runs from 1 to 0.

Hence, we can connect two brain areas bilaterally while almost no pairs of neurons have bilateral synaptic links. Of course, as a consequence of having very sparse area to area connections, the patterns need to be completed using recurrent connections within the area. If sufficient neurons of a pattern are activated in the area by connections from outside the area,

the pattern can be completed using the recurrent connections within the area.

5. Simulations

5.1. *Architecture*

While this approach does not need a hidden layer, models that use one are known to be more powerful. Hence, we ran our simulations using a three area model. Two of these areas receive outside input (henceforth called *input areas* or areas 1 and 3), while one area does not (the *hidden area* or area 2). Every area consists of a population of 625 excitatory neurons and a population of 625 inhibitory neurons. The function of the inhibitory neurons is homeostatic. They balance the overall activity. Using a more realistic ratio of excitatory to inhibitory cells[17] (which is known to be close to 4:1) would simple lead to higher inhibitory weights from the inhibitory to the excitatory population. Also, the combination of excitatory and inhibitory neurons within an area is know to lead to oscillations[18], which in turn stabilises the patterns within an area by means of recurrent connections. Therefore, every excitatory population has excitatory connections to the inhibitory populations, which in turn project back to the excitatory population of the same area. The excitatory neurons also have connections with the other excitatory neurons in the same area (recurrent connections). The probability of recurrent connections was held constant at $p = 0.95$ for this study.

Only excitatory layers have connections to other areas. We ran 10 simulations for each probability of layer-to-layer connectivity in this set:

$$P = \{0.01, 0.03, 0.05, 0.10, 0.15, 0.20, 0.25, \ldots 1.00\} \tag{3}$$

5.2. *Neuron Model*

We used current-based leaky integrate-and-fire neurons, with membrane potential v, the dynamics of which is given in equation 4.

$$\frac{dv}{dt} = -\frac{v}{\tau_m} + I\frac{1}{C} \tag{4}$$

Here, τ_m is the membrane time constant, C is the capacitance of the membrane and I is the input current to the neuron. The input I to the neuron,

caused by an incoming spike from another neuron is computed with an alpha function (equation 5).

$$I_s(t) = w\frac{e}{\tau_\alpha}te^{-t/\tau_\alpha} \tag{5}$$

Here, w is the weight (or peak value) and τ_α is the synaptic rise time.

A spike is emitted once the membrane potential v of the neuron reaches the threshold θ. Afterwards, the membrane potential remains at reset potential for an (absolute) refractory period of $0.5ms$.

5.3. *Training and Testing*

The models in our study were trained using pairs of patterns to the input areas of the model. Each pattern pair was presented 200 times for 1 second with 1 second break in between. The patterns were presented by means of Poisson neurons connected to the input areas spiking with a frequency of 80Hz. The patterns consisted of 625-dimensional random bit patterns. The percentage of bits that were set to one was 3%. Patterns were completely orthogonal. Random white noise with a frequency of 2 Hz is presented to all excitatory neurons of all areas at all times.

After training, the pattern is presented to area 1 only and we tested the reconstruction of the corresponding pattern in area 3. If the spiking activity of a neuron is above a threshold it is considered to be activated by the pattern in area 1. Neurons that are part of the pattern and are not activated are considered to be errors. The same holds for neurons that are activated, but are not part of the pattern.

5.4. *Results*

As expected, we found oscillation in all brain areas around 40Hz. These are caused by the interaction between the excitatory and the inhibitory layer (see Fig. 3). The excitatory connections (both between and within a layer) drive the activation up until the inhibitory neurons go above threshold which drives the activity down again. Once the activity is low, the inhibitory neurons no longer fire, which allows the excitatory neurons to drive the activation up again.

The representations in the middle layer (which are not determined by external input), gradually emerges (visible in Fig. 3C and D).

In simulations that used a high connection probability, bilateral connections between the patterns in the two input areas was not established.

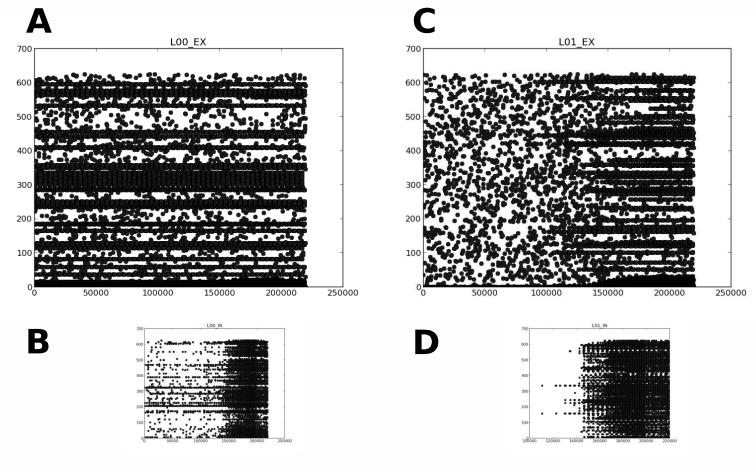

Fig. 3. A spike plot of the excitatory layer (A) and inhibitory layer (B) of one input layer. layer, as well as the excitatory layer (C) and inhibitory layer (D) of the hidden layer.

During the testing period, when the pattern was no longer applied to area 3, the pattern dissolves and a new, completely unrelated pattern emerges (visible in Fig. 4).

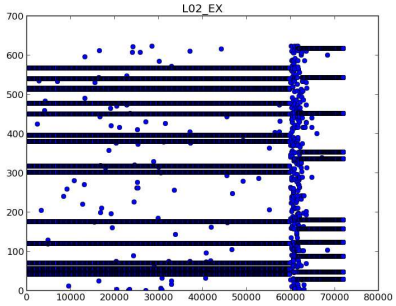

Fig. 4. A spike plot of area 3.

This can be explained by the weight patterns that emerge during training (Fig. 5). The Hinton-diagram shows weights of a neuron in the middle layer to layer 3. This neuron is connected to the input pattern in layer 1

by strong connections. If the connections had been established as desired, it would show strong links to the desired pattern in the next layer. Instead in shows exactly the opposite pattern. The weights that should be the strongest are in fact the weakest.

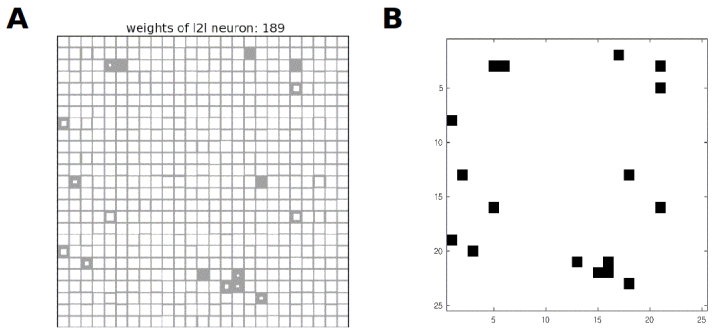

Fig. 5. (A) a Hinton-diagram (i.e., the weights in a matrix of white squares, where the size of the square represents the strength of the weight) of a neuron in the middle layer. (B) The target pattern that this neuron should activate. The weights show the opposite of the desired configuration.

As already sketched in section 3, this is happening, because when the pattern is applied to area 3, it generates a random pattern in hidden area 2 and connects with it with strong weights. While the connection in this direction grows stronger, the corresponding connections in the other directions grow weaker. When training stops, a strong link has been established from the patterns in area 3 to area 2, but no such link exists in the other direction (area 2 to area 3). Hence, during testing, the activation of a pattern in area 1 can activate a pattern in the hidden area, but no activation of the trained pattern in area 3 is possible from there.

While this finding agrees with our theoretical predictions, the more important result of this study is the behaviour of those models that had sparse between-area connections. For such a parameter setting, we found that bilateral connections between brain areas can indeed be learned with STDP. In contrast to the case reported above, the pattern was reconstructed through the learned weights, once the external stimulation was switched off. The simulations showed that (with the parameters used in this study) the performance peaks, between $p = 0.05$ and $p = 0.10$, being completely error free (see Fig. 6). The error reaches ceiling at $p = 0.35$. This is because of

the way the error is computed. To decide whether a neuron is part of an assembly the neuron with the highest firing frequency is detected and all neurons that fire within a certain percentile of that frequency are considered active. This works well unless there is no active assembly, because then all neurons are considered part of the assembly and therefore the error reaches maximum.

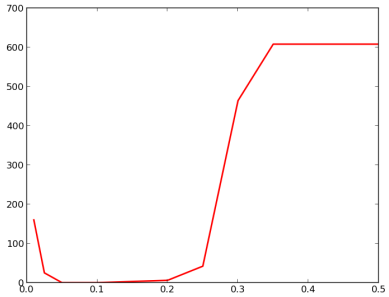

Fig. 6. The image shows the optimal connection probability (with other parameters being fixed).

We also tested the possibility of this architecture to form strong bilateral links between brain areas for 4, 5, and 6 area models. Pattern association was observed over multiple areas with a slight drop in performance for more layers. This will be reported in more detail elsewhere.

Given the small network size of the model used in the simulations, bilateral multi-area pattern association is limited to 3 patterns. Trying to learn more patterns lead to a fusion of patterns into one big neuron cluster. This is probably due to connections between the neurons of one patterns and neurons of the other patterns being potentiated by noise. Once the connections are above a certain threshold, they cannot be depressed anymore because any pre-synaptic spike leads to a postsynaptic spike, which causes the weight to increase until it reaches maximum. As soon as two patterns are connected by one strong synapse, they will co-activate and therefore all the neurons of one pattern will form strong synaptic links to the neurons of the other pattern.

6. Discussion

Our simulations clearly show that learning bi-directional associations between patterns in different brain areas is possible with standard STDP. However, it requires sparse area-to-area connectivity (5-10%). High recurrent within-area connectivity is needed to ensure pattern completion.

To compare these connectivity specifications with connectivity in the real brain is not straightforward, since there is no clear data about the percentage of neurons in one area that connect to another area. There is, however, a number of studies on this and closely related topics.

In a tracer study in the mouse Schüz et al.[19] finds that 55 to 70% of the stained axons were located in the near field of the neuron. Under the assumption that the density of synapses is constant along the axons, we can deduct that 55 to 70% of a neurons projections are inside the area. According to Schüz and Braitenberg[20], the vast number of cortico-cortical fibres projects to areas within the same lobe or to neighbouring areas in a different lobe, with the number of fibres of a certain length being inversely proportional to their length and no more than 2% going into the intra-hemispheric long-range bundle. Hence, the 30 to 45% of the synapses that are going to another brain area are distributed among all brain areas these neurons project to. Taking this data into account, a very low connection probability between areas and a high connection probability within areas seems biologically very realistic.

In terms of actual numbers of connections, Braitenberg and Schüz[17] estimate $9.2 \times 10^4/mm^3$ neurons and $7.2 \times 10^8/mm^3$ synapses in mouse cortex resulting in a ratio of 1:7826. Note that this does not mean that all synapses in that area are between neurons in that area, but it does give a good indication of the very high number of actually synapses per neuron. However, given that we have only 1250 neuron per brain area, we can certainly not model 7825 synapses per neuron.

One important aspect of our model that we like to point out is the within-area interaction between excitatory and inhibitory neurons. Not only does this keep the overall spiking activity in balance, but it also generates oscillations. These oscillations have the effect that at the input layers, the patterns can be applied to the model using Poisson neurons. No precise timing of the patterns are necessary. The oscillations naturally bring about the synchronous activation of neurons necessary for learning to the patterns. This is of the essence for the hidden layer where no external input can be used to control the timing.

It is possible that varying time delays[21] could also enable bidirectional communication between brain areas. In the real brain, there are time delays between spike events and post-synaptic potentials and these delays differ between individual connections. If two neurons are connected using a different timing, it might no longer be the case that every time the synapse in one direction is potentiated by STDP, the synapse in the opposite direction is depressed. The variation of delays is likely to be higher in long range connections between areas. Hence, it is worth studying time delays as a possible solution to the problem of bilateral association between different brain areas.

Also, it has to be said that the cortex is laminar and contains inter-laminar and intra-laminar circuitry. Therefore, it is not likely that the bidirectional connections between areas are directly between the same neurons. Instead, they might use a type a relay station, similar to the approach we have mentioned above, but have not pursuit in the simulation studies reported here. In the brain there is a difference between feed-forward and feed-back connections[22], while in our both directions have the same architecture. For a more realistic biological model, a difference between feed-back and feed-forward connections should be taken into account. One difference could be (and this is just speculation at this point) that the feed-back connections use a relay station. Consequently, feed-forward connections would be direct (and fast), while the feed-back connections would be poly-synaptic (and slower). The problem of direct bidirectional connections would not occur.

In this paper, we demonstrated that a documented difficulty of standard STDP can be solved without modifying the learning rule by an interaction of architectural considerations. It remains to be tested how far this approach scales up or generalises. Nevertheless, this research has solved a difficult problem and made an important step on the way to more complex STDP-based architectures and brought us closer to models of higher cognitive functions and language processing.

Acknowledgements

This work was supported under EPSRC Grant EP/J004561/1 (BABEL). We would like to thank two anonymous reviewers for very constructive comments.

References

1. M. S. Seidenberg and J. L. McClelland, A distributed, developmental model of word recognition and naming, *Psychological Review* **96**, 523 (1989).
2. D. C. Plaut and C. T. Kello, The emergence of phonology from the interplay of speech comprehension and production: A distributed connectionist approach, in *The emergence of language*, ed. B. MacWhinney (Mahwah, NJ: Erlbaum, 1999) pp. 381–415.
3. G. Q. Bi and M. M. Poo, Synaptic modifications in cultured hippocampal neurons: dependence on spike timing, synaptic strength, and postsynaptic cell type, *J. Neurosci.* **18**, 10464 (1998).
4. W. G. H. Markram and P. J. Sjöström, A history of spike-timing-dependent plasticity, *Frontiers in Synaptic Neuroscience* **3**, 1 (2011).
5. X. Jin, M. Lujan, L. A. Plana, S. Davies, S. Temple and S. Furber, Modeling spiking neural networks on spinnaker, *Computing in Science & Engineering* **12** (2010).
6. P. Indefrey and W. J. M. Levelt, The neural correlates of language production, in *The New Cognitive Neurosciences*, ed. M. Gazzaniga (Cambridge, MA: MIT Press, 2000) pp. 845–865, 2nd edn.
7. D. O. Hebb, *The Organization of Behaviour* (Wiley New York, 1949).
8. S. Song, K. D. Miller and L. F. Abbott, Competitive Hebbian learning through spike-timing dependent synaptic plasticity, *Nature Neuroscience* **3**, 919 (2000).
9. S. Song and L. F. Abbott, Column and map development and cortical re-mapping through spike-timing dependent plasticity, *Neuron* **32**, 339 (2001).
10. C. Clopath, L. Buesing, E. Vasilaki and W. Gerstner, Connectivity reflects coding: A model of voltage-based spike-timing-dependent-plasticity with homeostasis, *Nature Neuroscience* **13**, 344 (2010).
11. W. Gerstner and W. M. Kistler, *Spiking Neuron Models* (Cambridge University Press, New York, 2002).
12. S. J. Thorpe, D. Fize and C. Marlot, Speed of processing in the human visual system, *Nature* **381**, 520 (1996).
13. J. Rubin, D. D. Lee and H. Sompolinsky, Equilibrium properties of temporally asymmetric Hebbian plasticity, *Physical Review Letters* **86**, 364 (2001), PMID: 11177832.
14. R. Gütig, R. Aharonov, S. Rotter and H. Sompolinsky, Learning input correlations through nonlinear temporally asymmetric Hebbian plas-

ticity, *The Journal of Neuroscience* **23**, 3697 (2003).
15. W. Gerstner, R. Kempter, J. Van Hemmen and H. Wagner, A neuronal learning rule for sub-millisecond temporal coding, *Nature* **383**, 76 (1996).
16. E. M. Izhikevich, J. A. Gally and G. M. Edelman, Spike-timing dynamics of neuronal groups, *Cerebral Cortex* **14**, 933 (2004).
17. V. Braitenberg and A. Schüz, *Cortex: Statistics and Geometry of Neural Connectivity*, 2 edn. (Springer, 1998).
18. T. Wennekers and G. Palm, On the relation between neural modelling and experimental neuroscience, *Theory in Bioscience* **116**, 273 (1997).
19. A. Schüz, D. Chaimow, D. Liewald and M. Dortenman, Quantitative aspects of corticocortical connections: A tracer study in the mouse, *Cerebral Cortex* **16**, 1474 (2006).
20. A. Schüz and V. Braitenberg, The human cortical white matter: Quantitative aspects of cortico-cortical long-range connectivity, in *Cortical Areas: Unity and Diversity*, eds. A. Schüz and R. Miller (Taylor & Francis, London and New York, 2002)
21. E. M. Izhikevich, Polychronization: computation with spikes., *Neural Comput* **18**, 245 (2006).
22. A. Angelucci and P. C. Bressloff, Contribution of feedforward, lateral and feedback connections to the classical receptive field center and extra-classical receptive field surround of primate v1 neurons, *Prog Brain Res* **154**, 93 (2006).