

Spike Timing Dependent Adaptation for Mismatch Compensation

Katherine Cameron, Alan Murray
School of Engineering and Electronics
The University of Edinburgh
Edinburgh, EH9 3JL, UK

Email: k.cameron@ed.ac.uk, alan.murray@ee.ed.ac.uk

Steve Collins
Department of Engineering Science
University of Oxford
Oxford, OX1 3PJ, UK
Email: steve.collins@eng.ox.ac.uk

Abstract—This paper presents some circuitry for use within a visual-processing depth-recovery algorithm based upon spike timing. The accuracy of the depth calculation relies on a prediction which, when implemented in analogue VLSI, will be degraded by transistor mismatch. An adaptive circuit based on Spike Timing Dependent Plasticity (STDP) was designed to reduce the effects of VLSI process variations on the algorithm's performance. Simulation Results for the circuitry, designed using a $0.35\mu\text{m}$ process, are reported.

I. INTRODUCTION

Spike Timing Dependent Plasticity (STDP) is a neural algorithm operating not on spike-rate correlations but individual inter-spike timings. STDP therefore has the ability to adapt synaptic delays [1]-[3] to improve or “tune” network performance. The interesting properties of STDP have been investigated and successfully implemented in analogue VLSI [4]-[8]. Furthermore, the Monte Carlo simulation results reported in this paper come from a design which is an extension of our previous work, fabricated in $0.35\mu\text{m}$ silicon [9], which showed the ability to correct for mismatch effects.

The problems of transistor mismatch are exacerbated as device sizes are driven further into sub-micron geometries [10]. With an increasing number of devices on chip the demand for lower power consumption encourages sub-threshold design where the problem of mismatch is even greater [11]. As mismatch can effect circuit time constants, we have chosen to exploit STDP's ability to adapt delays and investigate the use of a time dependent spiking algorithm which compensates for the mismatch error within an exemplar neural system.

This paper will present the circuitry designed and Monte Carlo simulation of the circuit extraction results showing both the problems caused by mismatch and the benefits of the adaptation network. The circuitry will be fabricated using a $0.35\mu\text{m}$ process to verify the simulation results. Although these results are based upon an explicitly spike-driven depth-from motion algorithm, the techniques demonstrated are generic to any algorithm that is based upon inter-spike correlations. As a result, the results herein and the chip that implements the STDP process do not have a specific application. Rather, they aim to develop new forms of robust circuit for the noisy, mismatched and unreliable building nanoscale building blocks of future chips.

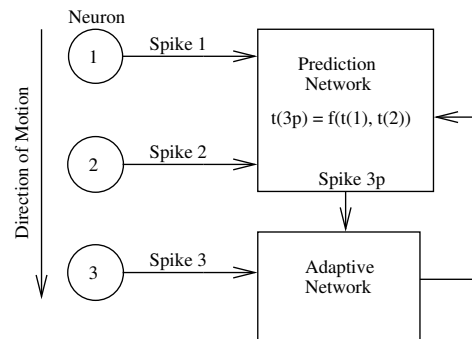


Fig. 1. **System Block Diagram.** Neurons 1 to 3 fire at predetermined intervals. Neuron 3p fires at a time predicted from the firing time of neurons 1 and 2. If the neurons fire at equal intervals, $t(3p) = t(2) + \{t(2) - t(1)\}$ where $t(1)$ is the firing time of spike 1, $t(2)$ for spike 2 etc. Spike 3p is the prediction of spike 3.

II. PREDICTIVE CIRCUITRY

We chose the vision-processing system described by Wörgötter et al. [12] to show STDP's mismatch compensation ability. It is a depth-from-motion algorithm which uses relative neural spike timings to recover depth information from radial flow-fields. The algorithm was found to be susceptible to noise and a prediction mechanism was added to make it more robust. The prediction uses the timing of two spikes to determine when a third spike will arrive. If the spike arrives within a small time window of the prediction it is treated as genuine. Timing mismatches are therefore potentially very damaging to this algorithm, rendering it a good candidate for this study.

We used a prediction network (figure 1) designed for the condition where the time between spikes 1 and 2 was equal to that between spikes 2 and 3. In this case the time of the predicted spike 3, spike 3p, is $t(2) + \{t(2) - t(1)\}$ where $t(1)$ is the time of spike 1, $t(2)$ is the time of spike 2 etc.

Preliminary attempts to compensate for mismatch error are described in [9]. The original design suffered from the problem that most of the mismatch was caused by a gain error, i.e. differing g_{ms} , but no gain correction could be applied. Instead an offset correction, equivalent to addition or subtraction of time, was applied which was only applicable to a fixed time delay, in that case 5ms. When the delay was changed to 1ms the circuit had to adapt to a different level of correction.

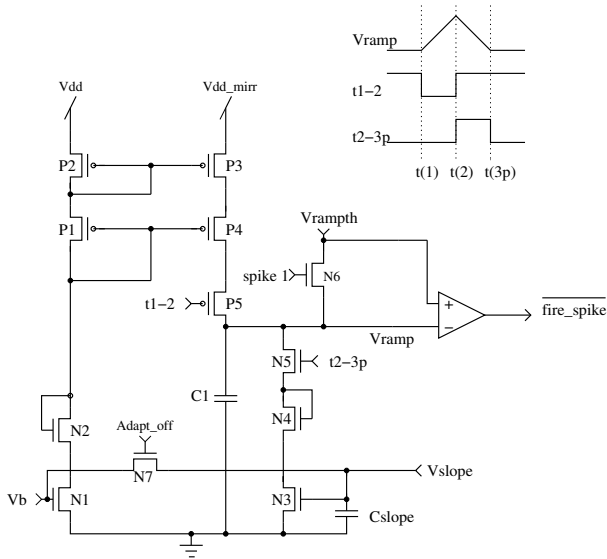


Fig. 2. **Spike prediction circuit.** C1 charges between $t(1)$ and $t(2)$ and discharges from $t(2)$ to $t(3p)$. If N7 is closed V_b equals V_{slope} and an accurate prediction will be given if the transistors have good matching. If that is not the case N7 can be opened and V_{slope} can be set at a different voltage to ensure equal charge and discharge times.

The predictive circuitry was re-designed to address this issue and is shown in figure 2. The sequence of operation is:

- $t(1)$ Spike 1 is fired and N6 acts as a switch which sets the initial value of V_{ramp} to V_{rampth} .
- $t(1)-t(2)$ P5 is on and C1 is charged through P3 and P4 which supplies a mirrored version of the current set by V_b .
- $t(2)$ P5 turns off while N5 becomes active.
- $t(2)-t(3p)$ C1 is discharged through N3 and N4 at a rate set by the voltage across C_{slope} . If the current source and sink are matched and V_b equals V_{slope} the time taken to discharge C1 should match the charging time.
- $t(3p)$ V_{ramp} is compared to V_{rampth} using a differential pair. When V_{ramp} crosses it, neuron 3p is “fired”. This ends the discharge period.

The bias voltage applied to N1 is now no longer permanently connected to N3. The closing of N7 allows the voltage across C_{slope} to be initialised to V_b but it can be open to enable different bias voltages to be set. The W/L ratio for transistors N3 and N4, and therefore N1 and N2, was altered to reflect the driving of N3 by the voltage across C_{slope} . The length of the transistors determines their response to the changing voltage V_{ramp} and was set at $3\mu m$. To reduce the capacitive coupling of V_{ramp} to V_{slope} it was important to keep the gate-drain parasitic capacitance to a minimum. The width chosen to reduce this effect was $0.6\mu m$.

If N7 is closed the prediction accuracy demands that the matching between transistors be good. Should mismatch occur, the discharging current can be greater or less than the charging current resulting in an early or late prediction. The mismatch could be eliminated by post-fabrication trimming but this

is expensive and time-consuming. Instead we minimise the effect of the mismatch by altering V_{slope} . The alteration is performed by the adaptive circuitry described in the next section based on the time difference between spikes 3 and 3p. Effectively, we use STDP to restore the near-perfect synchrony that has been corrupted by mismatch.

III. ADAPTIVE CIRCUIT DETAILS

Transistors N1 and N3 in figure 2 have sub-threshold gate voltages so any change in V_{slope} could potentially have a large impact on the discharging current. To ensure a small voltage change a switched capacitor technique is used as illustrated in figure 3. The HiLo circuit uses source followers to generate signals approximately 200mV above and below V_{slope} . If spike 3 arrives before the prediction, V_{slope} must be increased and the *up* signals, *up*, *up_t* and *up_id*, are active. If spike 3p arrives first, the *down* signals are active but the operation is essentially the same. Therefore only the process of increasing V_{slope} is described here.

A capacitor bank, C0-3, is used to implement the change in V_{slope} in a time dependent manner, as opposed to merely direction dependent. It is used in a similar way to that described in [13]. If a single capacitor were to be used then the change in V_{slope} would always be by a fixed amount, either up or down, irrespective of the accuracy of the prediction. This would lead to a choice between either very slow convergence and high accuracy or fast convergence with low accuracy. The inclusion of the capacitors is controlled by $CS1-3$ which are reset when the *up* signal is generated at the arrival of spike 3. The NOR and NOT gates marked S are “current starved”, by a sub-threshold biased transistor connected between VDD and the standard circuit configuration, and therefore have a slow rise time. After the *up* signal, the output of the NOR gate

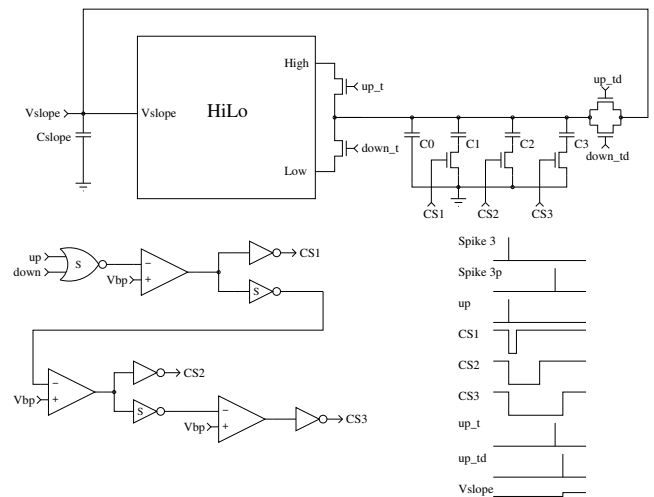


Fig. 3. **V_{slope} Update circuitry.** The circuit employs a switched capacitor technique to increase or decrease the voltage V_{slope} . The amount of change is determined by the signals $CS1-3$ which turn on over time, one after another. The logic gates marked S have a slow rise time and implement the delay between CS signals. The HiLo circuitry uses source followers to generate signals approximately $\pm 200mV$ around V_{slope} .

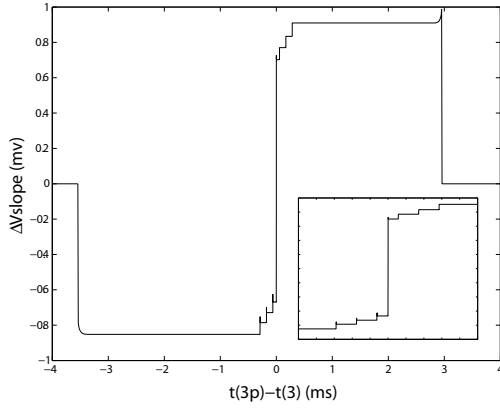


Fig. 4. **The change in V_{slope} vs prediction error.** The change in voltage of V_{slope} implemented by the circuitry shown in figure 3. The inset graph is of the range $-500\mu s - 500\mu s$ and shows the effect of switching in the capacitors C1-3 in figure 3.

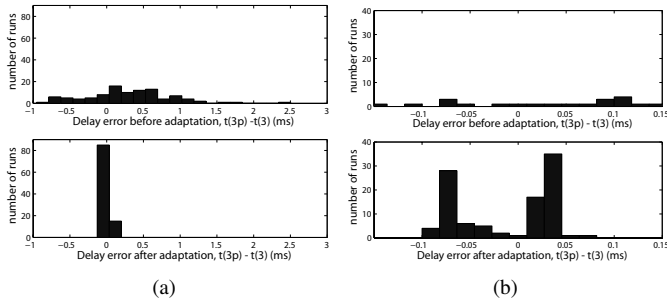


Fig. 5. **Distribution of prediction errors.** The time difference, $t(3p) - t(3)$, was measured before and after adaptation. The x-axis scale is set using the pre-adaptation range in (a) and the post-adaptation range in (b).

slowly rises until it reaches V_{bp} , in this case 2.5V. At this point CSI becomes high and the output of the first slow NOT gate starts to rise. With CSI high, charge can be stored on the connected capacitor. $CS2$ and $CS3$ become high approximately $100\mu s$ after the preceding CS signal. When spike $3p$ arrives $up.t$ is generated and the $High$ voltage from the $HiLo$ circuit is stored across all active capacitors. Therefore the amount of stored charge is related to the time between spike 3 and spike $3p$. $Up.td$ is generated slightly later at which point the charge is shared between the capacitor bank and C_{slope} resulting in a rise in V_{slope} . C_{slope} is a much larger capacitor than the others therefore the voltage change is much smaller than the 200mV difference generated by the $HiLo$ circuitry. Figure 4 shows the resulting change of V_{slope} with different time delays between the actual and predicted spikes. A windowing mechanism is used so that if a predicted spike arrives much earlier or later than the actual spike they are deemed to be unrelated. This results in the zero change at $|t(3p) - t(3)| > 3ms$. As the adaptive network drives V_{slope} towards a particular point it can be left active after calibration. It will then continue to compensate for any circuit drift caused by changes in environmental conditions or charge leakage from the capacitor. The only constraint is that more correct spikes are present than

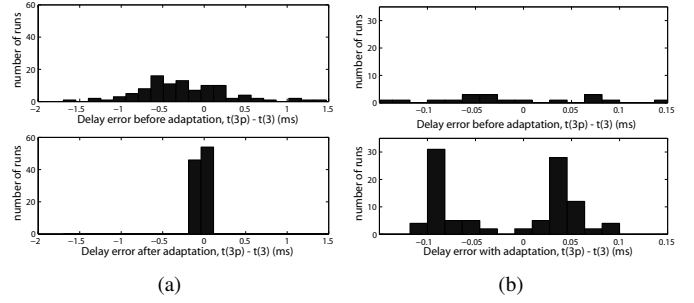


Fig. 6. **Distribution of prediction errors when $t(2) - t(1) = \frac{5}{6} \{t(3) - t(2)\}$.** The time difference, $t(3p) - t(3)$, was measured before and after adaptation. The x-axis scale is set using the pre-adaptation range in (a) and the post-adaptation range in (b).

incorrect ones.

IV. RESULTS

The circuitry was designed using the AMS $0.35\mu m$ C35 process. The testing was performed using Monte Carlo mismatch simulations of the circuit extraction in which the transistor and capacitor characteristics span the measured statistics of the fabrication process.

100 Monte Carlo simulations were run with a time delay between spikes of 5ms. The time difference, $t(3p) - t(3)$, was measured before and after adaptation. The results are shown in figure 5(a) and the improvement in the prediction is clear. Figure 5(b) shows the same data but this time the x-axis scale has been set to reflect the post-adaptation error range.

It can be seen that very few pre-adaptation results are as good as the full range of post-adaptation results. It is also a significant improvement on the results from our previous design reported in [9]. The two separate distributions visible in (b) are caused by the shape of the voltage change curve in figure 4. Around $t(3p) - t(3) = 0$ there is a swing of approximately 1.4mV. This means that when the prediction is close to $t(3)$ it tends to oscillate between “too fast” and “too slow” but is unable to get exactly to the zero point.

It is also possible to predict a time proportionate but not equal to $t(2) - t(1)$ by ratioing the widths of transistors N1 and N3 in figure 2. We plan to fabricate a prediction circuit with a 5:6 ratio, i.e a 5ms time delay will result in a prediction of 6ms, and the results of the extracted simulation are shown in figure 6.

While improving the prediction quality is important it was also an aim of this work to design a correction method that would work while the time delay between spikes 1 and 2 was varying, as it would when objects at different depths are being processed. Two spike trains were applied to both this circuit and the previous design [9] during Monte Carlo simulations. The first train had 25 sets of spikes and the time delay varied between 3 and 6ms. A set is defined as a spike 1, spike 2, spike 3 group. The second was a longer train of 33 sets with a larger delay range of 1 to 6ms. The time difference between every pair of spikes 3 and $3p$ was measured and the results of 5 simulations are shown in figure 7. In both cases the initial

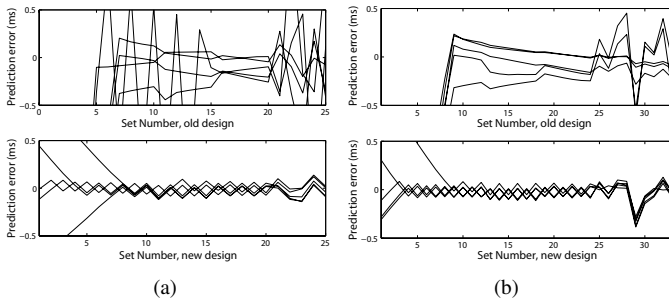


Fig. 7. **Prediction error during adaptation.** Spike trains were applied during 5 Monte Carlo simulations with varying time delays to the current circuitry and a previous design. (a) 25 spike sets were applied over a range of 3 - 6ms. (b) 33 spike sets were applied over a range of 1 - 6ms.

applied delays varied only a small amount around 5ms with the larger swings at the end of the simulation. Note also that the initial adaptation of the new circuitry can be seen as the error converges around 0. In figure 7(a) it is clear that the new circuitry provides a consistently good prediction error as the delay time changes. This is a significant improvement when compared to the results for the previous design. In (b) there is a less dramatic improvement but it is still quite clear. For the new design the only time an error is registered outside of the range $\pm 150\mu\text{s}$ is when the delay jumps from 1 to 6ms and the error has recovered by the next iteration. The difference between the results is partly due to the increased range of time delays and the number of spike sets applied. The 25 set simulation did not allow the original design to reach a particularly stable point before the delay began to vary. In one case this even resulted in a situation where no convergence took place. The 33 set simulation gave slightly more stability to the input and the old design results were better. While the gain correction provides a better result over changing time delays than the offset correction it is still not perfect. This is the result of the error in the prediction being composed of both a gain and an offset error.

V. CONCLUSION

We have shown through simulation that spike timing dependent algorithms can be used successfully to reduce the effects of device mismatch within the context of a neural processing system for visual scenes.

The circuitry generates a prediction for spike arrival from the timing of two previous spikes. This prediction can be degraded by device mismatch and the temporal difference between the prediction and actual spike is used to adjust circuit parameters until the prediction is accurate. The Monte Carlo simulation results indicate a superior performance under steady state input when compared to our previous design. It also solves much of the problem when the input has a varying time delay.

This result demonstrates a new use of spike-timing as a route to the production of reliable neural architectures implemented on “unreliable” substrates. We use the knowledge that timing relationships are present in input signals from

sensors to perform unsupervised, autonomous calibration of timing circuitry. We are led to suggest that the technique has wider implications for other forms of sensor and for all spiking neural systems fabricated in real, imperfect technologies. It demonstrates yet another useful feature of spike-based coding and may well be part of the evolutionary advantage exploited by the nervous system in its use of spiking neurons as a computational substrate.

As indicated in the introduction, these techniques are generic to any algorithms that use inter-spike correlations. In more general spike-driven architectures, where explicit correlations are not part of the process, correlated spike trains can be introduced to the network in an initial calibration phase, within which autonomous self-calibration occurs. Approaches of this nature will be vital to robust systems using deep-sub-micron devices, to avoid a heavily-supervised calibration methodology that will be both time-consuming and a waste of circuit area.

ACKNOWLEDGMENTS

The authors wish to acknowledge the helpful discussions with Vasin Boonsobhak.

REFERENCES

- [1] W. Gerstner, R. Kempter, J. L. van Hemmen, and H. Wagner, “Hebbian learning of pulse timing in the barn owl auditory system,” in *Pulsed Neural Networks*, W. Maass and C. M. Bishop, Eds. MIT Press, 1998, ch. 14, pp. 353–377.
- [2] H. Hünig, H. Glünder, and G. Palm, “Synaptic delay learning in pulse-coupled neurons,” *Neural Computation*, vol. 10, pp. 555–565, 1998.
- [3] C. W. Eurich, K. Pawelzik, U. Ernst, J. D. Cowan, and J. G. Milton, “Dynamics of self-organized delay adaptation,” *Physical Review Letters*, vol. 82, no. 7, pp. 1594–1597, 15th February 1999.
- [4] A. Bofill-i-Petit and A. F. Murray, “Learning temporal correlations in biologically-inspired aVLSI,” in *Proc. IEEE International Symposium on Circuits and Systems*, vol. 5, May 25–28 2003, pp. 817–820.
- [5] G. Indiveri, “Neuromorphic bistable VLSI synapse with spike-timing-dependent plasticity,” in *Advances in Neural Information Processing Systems*, vol. 15, 2003, pp. 1091–1098.
- [6] A. Bofill-i-Petit and A. F. Murray, “Synchrony detection by analogue VLSI neurons with bimodal STDP synapses,” in *Advances in Neural Information Processing Systems*, vol. 16. MIT Press, 2003.
- [7] G. Indiveri, E. Chicca, and R. Douglas, “A VLSI reconfigurable network of integrate-and-fire neurons with spike-based learning synapses,” in *Proc. European Symposium on Artificial Neural Networks*, April 28–30 2004, pp. 405–410.
- [8] A. Bofill-i-Petit and A. F. Murray, “Synchrony detection and amplification by silicon neurons with STDP synapses,” *IEEE Trans. Neural Networks*, vol. 5, no. 15, pp. 1296–1304, September 2004.
- [9] K. Cameron, V. Boonsobhak, A. Murray, and D. Renshaw, “Spike timing dependent plasticity (STDP) can ameliorate process variations in neuro-morphic VLSI,” *IEEE Trans. Neural Networks*, Accepted for Publication 2005, available at <http://www.see.ed.ac.uk/~s9727694/TNN0705.pdf>.
- [10] M. J. M. Pelgrom, A. C. J. Duinaijer, and A. P. G. Welbers, “Matching properties of MOS transistors,” *IEEE J. Solid-State Circuits*, vol. 24, no. 5, pp. 1433–1440, October 1989.
- [11] F. Forti and M. E. Wright, “Measurement of MOS current in the weak inversion region,” *IEEE J. Solid-State Circuits*, vol. 29, no. 2, pp. 138–142, February 1994.
- [12] F. Wörgötter, A. Cozzi, and V. Gerdes, “A parallel noise-robust algorithm to recover depth information from radial flow fields,” *Neural Computation*, vol. 11, pp. 381–416, 1999.
- [13] R. J. Vogelstein, U. Mallik, and G. Cauwenberghs, “Silicon spike-based synaptic array and address-event transceiver,” in *Proc. IEEE International Symposium on Circuits and Systems*, vol. V, 2004, pp. 385–388.