Homeostatic Fault Tolerance in Spiking Neural Networks utilizing Dynamic Partial Reconfiguration of FPGAs

Anju P. Johnson†, Junxiu Liu*, Alan G. Millard†, Shvan Karim*, Andy M. Tyrrell†,
Jim Harkin*, Jon Timmis†, Liam McDaid* and David M. Halliday†
†Department of Electronic Engineering, University of York, York YO10 5DD, UK
∗School of Computing and Intelligent Systems, Ulster University, Derry BT48 7JL, UK
Email: {anju.johnson, alan.millard, andy.tyrrell, jon.timmis, david.halliday}@york.ac.uk
Email: {j.liu1@, haji_karim-s@email., jg.harkin@, lj.mcdaid@}ulster.ac.uk

***Abstract*—We present a novel methodology that addresses the problem of faults in synapses of a spiking neural network using astrocyte regulation, inspired by recovery processes in the brain. Since Field Programmable Gate Arrays (FPGAs) are widely used for neural network applications, we aim to achieve fault tolerance in an astrocyte-neuron unit implemented on an FPGA. A fault is considered as a reduction in transmission probability of a synapse, leading to reduced spiking activity. Our novel repair mechanism exploits Dynamic Partial Reconfiguration (DPR) of the FPGA Clock Management Tiles (CMTs) to increase the clock frequency of neurons with reduced synaptic input, which restores the firing rate to pre-fault levels. We demonstrate the repair methodology on a spiking neural network implemented on an FPGA. The system maintains effective functional behavior with a loss of up to 99% of the original synaptic inputs to a neuron. Our repair mechanism has minimal hardware overhead with the tuning circuit (repair unit) which consumes only 0.8215% of the complete design and therefore supports scalable implementations. Additionally, the overall architecture has a minimal impact on power consumption ($1.371W$). The work opens up a novel way to utilize the capabilities of modern hardware to mimic homeostatic self-repair behavior achieving fault recovery.**

***Index Terms*—Fault Tolerance, Self-Repair, Spiking Neural Network, Astrocyte, Homeostasis, Field Programmable Gate Array, Dynamic Partial Reconfiguration, Bio-inspired Engineering.**

I. Introduction

FPGAs are frequently used to implement artificial neural networks as they combine computing capability, logic resources and memory capacity in a single device [1]. Also, FPGA allows neural networks to be evolved on hardware and new topologies/networks executed faster [2]. In this research, we focus on SRAM-based FPGAs since it is the most commonly used reconfigurable platform. SRAM-based FPGAs are prone to hardware failures such as *Single Event Upsets* (SEUs) [3]. This creates an issue for dependability for safety critical applications.

The present work is based on the inspiration derived from robust biological systems, which can detect and correct a range of errors. For instance, the human brain is continuously adapting a changing environment. The mechanisms that monitor excitation and maintain the functional properties of neurons are by definition homeostatic [4]. In this work, we demonstrate homeostasis using the dynamic reconfiguration properties of clock management cores in an FPGA. Dynamic Partial Reconfiguration (DPR) is an FPGA-specific technological advancement which aims at modifying the existing circuit mapped on the FPGA without needing to turn off the circuit functioning in other parts of the FPGA. Various works have demonstrated the possibility of fault tolerance in FPGAs via DPR [5]. As a variant to the classical DPR, we use Dynamic clock alteration, an alternate DPR technique to establish the task of fault tolerance. This work is the first report of an application of DPR-based clocking schemes for neural networks targeting fault tolerance. Various researchers have demonstrated fault tolerance in hardware implementations of neural networks [6]–[8]. Compared to these works, the work proposed in this paper demonstrates higher fault tolerance and the methodology is feasible in the presence of at least one healthy synapse. Some recent works also suggests the use of learning mechanisms to recover faults in synapses [9], [10].

Astrocytes have been shown to coexist with neurons where these cells communicate with synapses and neurons, thereby regulating synaptic activity [11]. We employ FPGAs to implement the astrocyte-neuron based self-repairing unit, which considers faults as a condition that results in a silent or near silent neuron caused by low transmission probability (PR) of a synapse. Faults in synapses that lead to reduced transmission probability may be due to an external cause such as sensor failures or internal faults such as SEUs in synaptic connections. Repair is defined as the ability of the system to restore firing rates. The proposed mechanism maintains constant neural activity by increasing the clock rate for the faulty neurons.

The rest of the paper is organized as follows. Section II describes the background required for better understanding of the paper. Section III presents the proposed idea of neuronal self-tuning for homeostatic regulation of firing rates. Section V presents experimental results establishing the effectiveness of the proposed scheme. Finally, the paper concludes in Section VI.

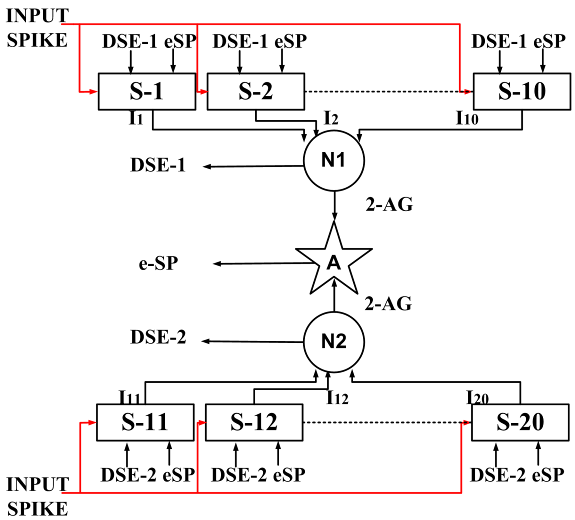

Fig. 1: **Basic unit for self repair mediated by an astrocyte** Two neurons N_1 and N_2, each receive 10 synaptic inputs ($S-1$ to $S-10$ and $S-11$ to $S-20$). A represents the astrocyte connected to N_1 and N_2. The signals $DSE-1$ and $DSE-2$ are local to synapses connected to N_1 and N_2 respectively, whereas eSP is a global signal associated with all synapses connected to A.

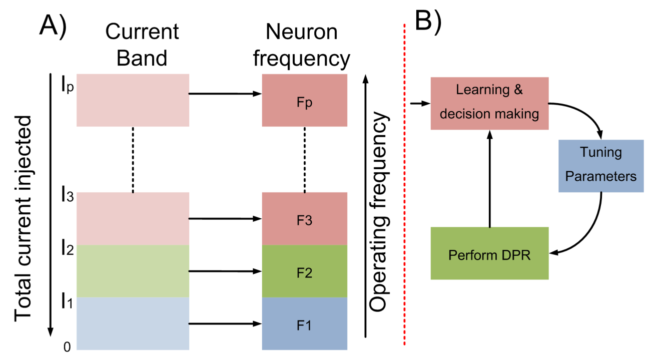

Fig. 2: **Illustration of proposed self-tuning methodology** (A) The maximum injected current falls at a time slot Δt under one of the current band $I_i - I_j$. The current falling in each bands are mapped to corresponding operating frequencies of the neural clock. As the maximum injected current falls in higher order bands, corresponding mapped operating frequency of the neuron decreases. (B) The neural self-tuning is performed following three phases, namely, (1) monitoring the maximum current injected to the neuron and making a decision based on observed maximum current, (2) modeling of DCM tuning parameters, and (3) performing DPR.

II. Background

A. Reduced Model of Bio-inspired Self Repair Unit

The detailed hardware model of the astrocyte-neuron self-repairing unit is presented in [12]. In this work, we use a simplification of this model which has greater than 90% hardware efficiency compared to [12], and at the same time achieves the same level of fault repair [7]. This model simplifies the complex chemical processes inside an astrocyte by retaining the key features of direct negative feedback and indirect positive feedback in the self-repairing unit shown in Fig. 1. The architecture consists of two neurons (N_1 and N_2) and a common astrocyte (A). Each neuron is associated with a set of synapses. In our experiments we use 10 synapses for each neuron. The neurons are provided by input Poisson spike trains. In addition to the spike inputs, the synapses receive direct signaling (DSE) from the associated neuron and indirect signaling (eSP) from the associated astrocyte. There is no spike transmission between N_1 and N_2. The synapses associated with the two neurons are influenced by the common signal eSP. The synapse processes the signals DSE and eSP, and makes a decision on the current to be injected into the neuron. More details of this model is presented in [7].

B. Dynamic Partial Reconfiguration of Clock Generation Unit

DPR in clock management tiles of the FPGA provides a way for generating custom clocks on the fly depending on the requirements of applications. The usual techniques to generate such custom clocks is to use some clock generation circuitry such as the *Phase Locked Loop* (PLL) module or the *Digital Clock Manager* (DCM) module. The relation between the input and output clock signals is given by

$$F_{CLKFX} = F_{CLKIN} \times \frac{M}{D} \tag{1}$$

Where F_{CLKIN} is the input clock signal to the DCM, F_{CLKFX} the corresponding synthesized clock signal, M is a multiplication factor and D is a division factor. The DPR capability of the FPGA allows modification of the M and D values during runtime to synthesize different clock frequencies. By controlling these parameters, various clock frequencies can be synthesized on-the-fly. For more details, see [13].

III. Astrocyte neuron network incorporating DPR based self-tuning

In addition to the reduced model discussed in section II-A, the proposed architecture consists of two more components: (a) A dynamically reconfigurable clock management unit, one for each neuron in the system, (b) A global clock management unit for generating the clock frequencies of components in the architecture other than the neurons.

The working of the proposed system can be summarized as follows: All synapses associated with a neuron are excitatory in nature and they inject a constant amount of current (I_{inj}) to the neuron. Based on the probabilistic nature of the synapse, the total current injected to the neuron varies with time. Considering a small duration for observation, the maximum current injected to the neuron remains fairly constant in the absence of synaptic faults. In the case of synaptic failures, the maximum current injected to the neuron diminishes based on the percentage of synaptic failures. All neurons in the

TABLE I: **Current bands to clock frequency mapping for neural self tuning: values derived empirically**

Percentage of synaptic Fault	I_{max} range	DCM Parameters		Neuron Clock frequency(MHz)
		M	D	
$[0-70)\%$	$(10.I_{inj}-4.I_{inj})$	2	2	100
$[70-80)\%$	$(4.I_{inj}-2.I_{inj})$	3	2	133
$[80-100)\%$	$(2.I_{inj}-0)$	3	1	200

system monitor the maximum current injected for a duration Δt. Based on this observation, the neurons decides whether or not to initiate a dynamic partial reconfiguration. This allows the neuron to maintain a constant firing rate if the total injected current reduces due to synaptic failures.

The self-repairing hardware paradigm presented in Fig. 2, shows three phases of the hardware cycle required to perform neuronal self-tuning. The first phase is the learning and decision-making phase. The neuron learns the maximum current injected into it. Based on the maximum current injected in each duration, neuron decides whether or not to perform a DPR. To illustrate the self-tuning concept we first consider the case where x out of 100 synapses associated with neuron N_1 are faulty (PR=0.0). The maximum current that can flow to neuron N_1 (in the absence of an astrocyte) at any time during the existence of a fault is $(100-x)I_{inj}$. The neuron monitors the total injected current to obtain a baseline measurement. Based on the maximum injected current, the neuron makes a decision whether or not to undergo an operating frequency change. If the maximum injected current in slot Δt_i varies from that in slot Δt_{i-1}, a frequency change is desired. In the second phase, the neuron formalizes the DCM tuning parameters. The details and range of tuning parameters are discussed in section II-B. The final phase is to perform DPR. The neuron writes the DPR parameters to the reconfiguration ports. This initiates a DPR at its associated clock management unit.

We illustrate the proposed idea by dividing the input current into three bands. The presence of an astrocyte is sufficient to establish a repair if the fault in one of the neurons in a two neuron system is up to 70%. Beyond this fault level, the firing rate drastically reduces. Our approach tries to establish a homeostatic regulation of firing rate beyond 70% faulty synapses. Based on the experimental observation, we have determined the required operating frequencies of the neuron in the presence of faults higher than 70%. This is depicted in Table I.

IV. Application

Our application of neural self-tuning is in robot navigation. For instance, SNN based fault tolerance finds application in robots working in noisy environments, in which, the inputs to sensors are weak. This leads to low input signals– a condition similar to low transmissions in synapses. Also, hardware faults in synapses can also be recovered by this technique. The presence of astrocyte in SNNs achieving fault tolerance in the presence of synaptic failure has been demonstrated in [6]. In this work, the robot car cannot complete the straight line moving task under the fault rate of 80% or higher. The work proposed in this paper demonstrates higher fault tolerance and the methodology is feasible in the presence of at least one healthy synapse. Hence DPR based neural tuning is a promising solution for robotics applications demanding fault tolerance. More details of this work is presented in [14] with detailed applications.

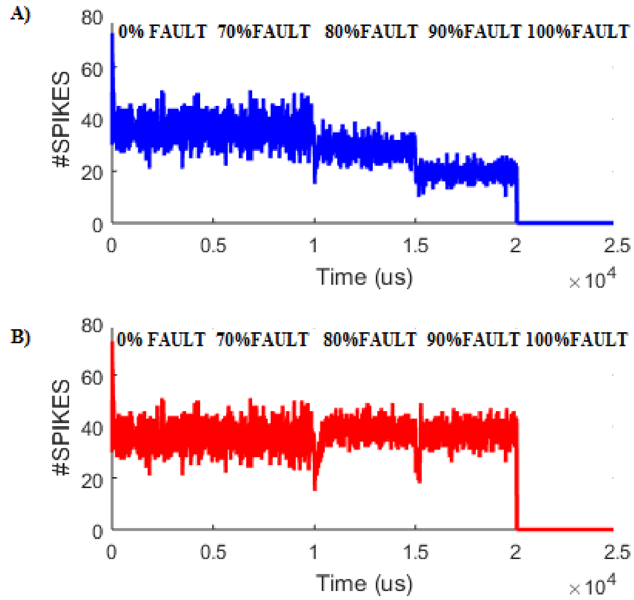

Fig. 3: **Network in Fig. 1 with fault levels** $(70-100)\%$ **in neuron** N_1**.** (A) absence of dynamic partial reconfiguration of clock management cores (Astrocyte is present) (B) with the dynamic partial reconfiguration of clock management cores (Astrocyte and DPR).

V. Experimental Results

The hardware architectures support for homeostatic regulation of neuronal firing rate was designed using Verilog HDL. The designs were synthesized and implemented using Xilinx ISE 14.7 CAD software.

A. Simulation Results to Demonstrate the Proposed Diagnostic and Repair Process

The proposed architecture was simulated using the Xilinx Isim simulator. Fig. 3 shows the homeostatic regulation of firing rate. In our experiments, we introduced faults (by lowering transmission probability of synapse) of 70% at time $500\mu s$, 80% at time $1000\mu s$, 90% at time $1500\mu s$ and 100% at time $2000\mu s$. As demonstrated in Fig. 3(A), the network faces a loss in firing rate in case of faults higher than 70% when using a Astrocyte only repair mechanism. We were able to achieve a complete recovery of firing rates as long as a single synapse is non-faulty. This is depicted in Fig. 3(B). We can observe a dip in firing rate at the start of each repair. This demonstrates the time required for establishing DPR.

TABLE II: **Hardware utilization of the two neuron self-repairing unit**

Resource	Slice	Slice Reg	LUT	DSP	DCM	PLL
Neuron network	3139	1537	10403	20	0	0
Tuning circuitry	26	36	37	0	2	1
Total	3165	1573	10440	20	2	1

TABLE III: **Pearson Correlation Coefficient**

No fault vs 70%fault	No fault vs 80%fault	No fault vs 90%fault
0.999995	0.999995	0.999997

B. Hardware Results on Xilinx Virtex-V FPGA

The proposed methodology is implemented on the Xilinx Virtex-V FPGA board. Recovery of firing rates in the proposed methodology, implemented on the FPGA is monitored using the Xilinx ChipScope Pro analyzer. Power estimation of the circuits was carried out using Xilinx XPower Analyzer and delay estimation using Xilinx Timing Analyzer. Estimated total on-chip power dissipation of the overall architecture is $1.371W$. Table II reports the hardware resource footprint of the proposed model. As evident from these reports, the proposed neural tunability for homeostatic regulation of neural firing rate can be implemented with reduced hardware overhead and power consumption.

C. Statistical Comparison

In our experiments, we incorporated multiple faults in the synapses of the SANN system. We have used two ways to compare the spiking activity of the system. One method is by using *Pearson correlation coefficient* (Pearson's r) [15]. Using Pearson's r we compare the timings of spike generation of the system subjected to various grades of fault. Table III reports the correlation between the spike times generated. From this measure, it is evident that spike times generated by the system have strong linear dependency (reported values are close to 1) with each other. Secondly, we analyse the histograms of spike frequencies subjected to faults of various grades (histograms not shown). The average spiking activity of the neuron connected to faulty synapses for all test cases were centred around mean 37 spikes, showing that the spikes generated are analogous. We also observe a reduction in standard deviation between the spike intervals as clock frequency increases. This shows that the neuron fires more regular as its input frequency increases. This is straight forward and finds explanation from jittery behaviour of Xilinx DCM module [16] and also LIF neuron model.

VI. Conclusion

In this paper, a novel methodology for homeostatic regulation of neuronal firing rate is presented. In order to achieve a complete recovery in the presence of a range of faults, we utilize the DPR capability of clock management modules in the FPGA. Beyond the capabilities of previous homeostatic regulation of neural firing rate, a full recovery is achievable in our design. The proposed design is appropriate for FPGA-based applications running in environments that induce faults in systems, where reliability is critical. This work opens new directions in bio-inspired research.

VII. Acknowledgements

The work is part of the SPANNER project and is funded by EPSRC grant(EP/N007050/1, EP/N00714X/1). Additionally, the authors would like to acknowledge the platform grant(EP/K040820/1) funded by EPSRC.

References

[1] C. Edwards, "Growing Pains for Deep Learning," *Communications of the ACM*, vol. 58, no. 7, pp. 14–16, Jul. 2015.

[2] S. Merchant, G. D. Peterson, S. K. Park, and S. G. Kong, "FPGA Implementation of Evolvable Block-based Neural Networks," in *IEEE Congress on Evolutionary Computation (CEC)*. IEEE, 2006, pp. 3129–3136.

[3] N. Jing, J.-Y. Lee, Z. Feng, W. He, Z. Mao, and L. He, "SEU Fault Evaluation and Characteristics for SRAM-based FPGA Architectures and Synthesis Algorithms," *ACM Transactions on Design Automation of Electronic Systems (TODAES)*, vol. 18, no. 1, p. 13, 2013.

[4] G. W. Davis and I. Bezprozvanny, "Maintaining the Stability of Neural Function: a Homeostatic Hypothesis," *Annual Review of Physiology*, vol. 63, no. 1, pp. 847–869, 2001.

[5] J. J. Davis and P. Y. K. Cheung, "Achieving Low-overhead Fault Tolerance for Parallel Accelerators with Dynamic Partial Reconfiguration," in *2014 24th International Conference on Field Programmable Logic and Applications (FPL)*, Sep. 2014, pp. 1–6.

[6] J. Liu, J. Harkin, L. McDaid, D. M. Halliday, A. M. Tyrrell, and J. Timmis, "Self-Repairing Mobile Robotic Car using Astrocyte-Neuron Networks," in *International Joint Conference on Neural Networks (IJCNN) (in press)*. IEEE, Jul. 2016, pp. 1379–1386.

[7] A. P. Johnson, D. M. Halliday, A. G. Millard, A. M. Tyrrell, J. Timmis, J. Liu, J. Harkin, L. McDaid, and S. Karim, "An FPGA-based Hardware-Efficient Fault-Tolerant Astrocyte-Neuron Network," in *2016 IEEE Symposium Series on Computational Intelligence (SSCI)*, Dec. 2016, pp. 1–8.

[8] M. Krcma, Z. Kotasek, and J. Lojda, "Implementation of fault tolerant techniques into FPNNs," in *IEEE International Conference on Field-Programmable Technology (FPT)*, Dec. 2016, pp. 297–298.

[9] J. Liu, L. McDaid, J. Harkin, J. Wade, S. Karim, A. P. Johnson, A. G. Millard, D. M. Halliday, A. M. Tyrrell, and J. Timmis, "Self-Repairing Learning Rule for Spiking Astrocyte-Neuron Networks (accepted)," in *Proceedings of the 9th International Conference on Neural Information Processing (ICONIP)*, 2017.

[10] A. P. Johnson, J. Liu, A. G. Millard, S. Karim, A. M. Tyrrell, J. Harkin, J. Timmis, L. McDaid, and D. M. Halliday, "Homeostatic Fault Tolerance in Spiking Neural Networks utilizing Dynamic Partial Reconfiguration of FPGAs (accepted)," in *31th International Conference on VLSI Design and 17th International Conference on Embedded Systems (VLSID)*, Jan. 2018.

[11] B. Stevens, "Neuron-astrocyte Signaling in the Development and Plasticity of Neural Circuits," *Neurosignals*, vol. 16, no. 4, pp. 278–288, 2008.

[12] J. Liu, J. Harkin, L. Maguire, L. McDaid, J. Wade, and M. McElholm, "Self-repairing hardware with astrocyte-neuron networks," in *2016 IEEE International Symposium on Circuits and Systems (ISCAS)*, May 2016, pp. 1350–1353.

[13] *Virtex-5 FPGA User Guide UG 190 (v 5.4)*, Xilinx Inc, [Online]. Available: http://www.xilinx.com/support/documentation/user_guides/ug190.pdf, Mar. 2012, accessed: 2017-06-30.

[14] A. P. Johnson, J. Liu, A. G. Millard, S. Karim, A. M. Tyrrell, J. Harkin, J. Timmis, L. J. McDaid, and D. M. Halliday, "Homeostatic Fault Tolerance in Spiking Neural Networks: A Dynamic Hardware Perspective," *IEEE Transactions on Circuits and Systems I: Regular Papers*, vol. PP, no. 99, pp. 1–13, Jul. 2017.

[15] J. Benesty, J. Chen, Y. Huang, and I. Cohen, *Pearson Correlation Coefficient*. Springer Berlin Heidelberg, 2009, pp. 1–4.

[16] A. P. Johnson, R. S. Chakraborty, and D. Mukhopadhyay, "An Improved DCM-based Tunable True Random Number Generator for Xilinx FPGA," *IEEE Transactions on Circuits and Systems II: Express Briefs*, 2016.