

Astrocyte to Spiking Neuron Communication using Networks-on-Chip Ring Topology

George Martin, Jim Harkin, Liam J. McDaid, John J. Wade and Junxiu Liu
Computing and Intelligent Systems
Ulster University, Magee Campus
Northern Ireland
(E-mail: martin-g11@email.ulster.ac.uk, {jg.harkin, lj.mcdaid, jj.wade, j.liu1} @ulster.ac.uk).

Fearghal Morgan
Bio-Inspired Electronics and Reconfigurable Computing
Research Group
Electrical and Electronic Engineering, National University
of Ireland
Galway, Republic of Ireland
fearghal.morgan@nuigalway.ie

Abstract— Hardware faults are becoming more frequent due to geometric scaling, reducing the reliability and lifespan of devices. Current fault-tolerant approaches use redundancy or a central controller to detect and/or repair faults. However, these methods are also susceptible to faults. Astrocytes have been shown to facilitate biological self-repair in silent or near silent neurons in the brain by increasing the Probability of Release (PR) in healthy synapses. Astrocytes modulate synaptic activity, which leads to increased or decreased PR. To date, this has been proven with computational modelling and therefore the next step is to replicate this self-repair process in hardware to provide self-repairing systems. One of the key challenges for hardware neuro-glia networks is the facilitation of scalable communication between interacting neurons and astrocyte cells. This paper contributes a low-level Networks-on-Chip (NoC) ring topology for astrocyte to neuron/synapse communication which provide a scalable solution to this interconnect challenge. This work extends our current FPGA-based Hierarchical Networks-on-Chip (HNoC). FPGA results demonstrates that the new ring topology provides a good trade-off between low area/interconnect wiring overhead and communication speed for the relatively slow-changing data between astrocyte and neurons.

Keywords— *Networks-on-chip, astrocyte, neuro-glia, spiking neural networks, self-repair, FPGA, ring-topology.*

I. INTRODUCTION

With integrated circuit component geometric scaling, the likelihood of faults increases [1]. Researchers are considering self-repair as a desirable remedy to maintain functionality and to increase the operational lifetime of an electronic system [2]. Current fault tolerant mechanisms are based on coarse grained redundancy and employ the use of a central repair-decision agent to either find faults or to correct them. Triple Mode Redundancy (TMR) is generally employed in mission critical systems which cannot be repaired after deployment [3]. The TMR process of triplicating critical components, along with a comparator (voter) element, increases area overhead. Although the system may endure faults or the loss of critical components, it relies too heavily on spare parts and the use of the comparator (voter). Other methods include online detection or correction and autonomous self-repair [4]. The key weaknesses of existing approaches are the limited granularity at which repairs can be

implemented (i.e. gate, component level) and the lack of a distributed repair-decision mechanism (when the central repair-decision agent is compromised, self-repair is no longer possible).

Recent research has shown that biological traits such as fine grained repair and distributed repair-decision making are performed in the brain via astrocyte networks [5], [6]. Computational models of neuro-glia networks and repair have been successfully captured and applied to spiking neural networks (SNNs) [7]. This work demonstrates that a breakdown in a neurons firing activity, due to failed synapses (fine-grained level), can be repaired to near pre-fault firing rate by re-strengthening the neuron's remaining synaptic connections. The biological mechanism which controls the repair decision has been identified as the astrocyte (a type of glia cell) highly distributed within networks of neurons [7]. In traditional SNNs, synapses are connected to spiking neurons, in neuro-glia networks, additional direct connectivity exists between each astrocyte and multiple neurons. This bi-directional coupling provides an indirect communication mechanism from the astrocyte to the neurons during spiking events. While SNN spiking-based communication is discrete, astrocytic communication is continuous.

There has been an increasing interest in the implementation of astrocyte cells within both neuromorphic circuitry [8], [9] and within digital hardware devices [10]–[13]. One of the key challenges in progressing neuro-glia networks within a hardware SNN is the implementation of a scalable astrocytic network. Implementing scalable dense hardware neural and hardware astrocyte networks together, supporting the required network data rates presents a significant challenge. Networks-on-Chip (NoCs) have emerged as a common approach to connecting large networks of processing elements within a single chip [14]–[16]. The NoC uses routers and packet-based communication [17] and have been regarded as a suitable interconnect for hardware SNNs [18]–[20]. For example, the Hierarchical Network on Chip (HNoC) [1] paradigm developed by the authors demonstrates scalable interconnect for hardware SNNs. This paper presents a NoC-based solution to the connection of both neurons and astrocytes, and for information exchanges within a neuro-glia network.

The lowest HNoC communication level (level 1) is the *neuron NoC* facility, which consists of a node router connected in a point-to-point (direct connection) topology. HNoC level 2 is a *tile NoC* facility which connects 10 node routers via a tile router. HNoC level 3 is the *cluster NoC* facility which connects four tile routers. The *cluster NoC* facility has the ability to communicate with other *cluster NoC* facilities using north, east, south and west connection points. Each cluster facility supports the connection of 400 neurons, and communicates with other facilities using up and downstream communications [1].

This paper contributes a new ring topology extension to the previously reported HNoC [1], and for the first time facilitates astrocyte to neuron/synapse communication and information exchange within a neuro-glia network. Results illustrate that the new ring topology provides a good trade-off between low area/interconnect wiring overhead and communication speed for the relatively slow-changing data between astrocytes and neurons/synapses.

The structure of the paper is as follows: Section II reviews reported self-repair strategies including current neuro-glia networks. Section III presents the HNoC level 1 ring-topology for astrocyte-neuron communication support. Section IV presents results on scalability and section V concludes the paper.

II. BIOLOGY TO HARDWARE

Faults in integrated circuit components can be classified soft or hard faults. Soft faults, the most common type of fault, are temporary. They are caused by radiation or power fluctuations, and can be repaired or corrected by resetting or reconfiguring the device. Hard faults are caused by physical defects in either a component or the silicon interconnect; by wear-out or during the device manufacturing process, rendering the afflicted areas useless [21], [22].

A. Biological Self-repair

Recent research has shown that biological traits such as fine grained repair and distributed repair-decision making are performed in the brain via astrocyte networks [5], [6]. While neurons communicate using spike events, astrocytes communicate with neurons and other astrocytes bi-directionally using different chemical signaling pathways, supporting the uptake and release of glio-transmitters. Astrocytes can also modulate synaptic transmission [5] increase or decrease PR (Probability of Release) in associated synapses.

Astrocytes contain receptors which are activated when a spike event or action potential occurs. A spike event triggers the release of glutamate (Glu) from the presynaptic axon into the cleft. The glutamate binds to receptors on the postsynaptic side causing depolarisation of the postsynaptic neuron. At high levels of depolarisation, endocannabinoids, 2-AG (2-arachidonyl glycerol) are synthesised and subsequently released from the postsynaptic neuron which is taken up by the astrocyte. This causes oscillations of Ca^{2+} (calcium) within the astrocyte which in turn causes the release of glutamate or glio-transmitters; this is an indirect feedback mechanism from the astrocyte to the neuron(s) which allows the astrocyte to communicate with the neuron (see Fig. 1). There are two feedback signaling pathways, namely:

a) The indirect feedback via the astrocyte referred to as e-SP (Endocannabinoid-mediated Synaptic Potentiation) which strengthens the PR.

b) The direct feedback referred to as DSE (Depolarization-induced Suppression of Excitation), which causes a decrease of PR associated with synapses.

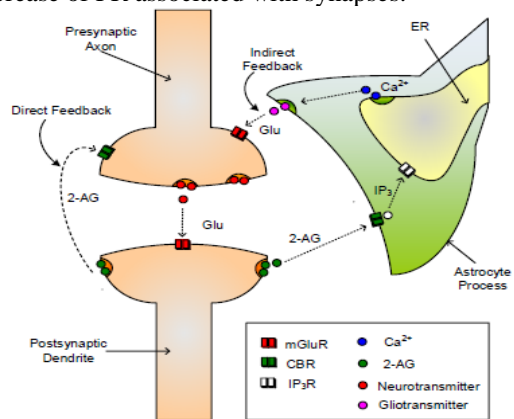


Fig. 1. A Presynaptic Axon and the Postsynaptic dendrite during a spike [5].

This signaling behavior has been modelled in previous work [5] and is the mechanism by which repair decisions are communicated at network level. This is referred to as low-level communication when the neurons and synapses interact with the astrocyte. Inter-astrocyte interactions are defined as high-level communications.

B. Neuro-glia Networks

Computational models of such repair have been successfully captured and applied to spiking neural networks to demonstrate repair of neuron firing activity [7]. For example, when an active neuron suddenly stops firing, while its input activity is maintained, it is deemed faulty. This is modelled as a rapid drop in PR at its associated synaptic sites. These faulty neurons are referred to as silent or near silent neurons. Work has shown that astrocytes can detect faulty synapses (fine-grained level) associated with silent neurons [5], and by subsequently increasing the weights on surrounding healthy synapses, they restore the neuron to its original functionality i.e. the increasing PR restarts the learning process which potentiates weights on healthy synapses. Fig. 2 illustrates two neurons firing with Astrocyte feedback. N1 and N2 depicts neurons, A1 an astrocyte, e-SP and DSE are excitatory and suppressive feedback signals. C1 and C2 contain tripartite synapses. In (A) both neurons are firing however in (B) N2 has stopped firing. Although DSE from N2 has stopped the astrocyte e-SP feedback is still active due to N1 still remaining active. This leads to an increase in PR and weights of the remaining healthy synapses of C2 and a restoration of N2 firing activity to pre fault levels of firing activity [7] illustrates two neurons firing, when one neuron stops, the excitatory signal provided by the astrocyte (e-Sp) is maintained by the nearby neuron.

Therefore, by increasing the PR in the remaining healthy synapses, the neuron functionality is restored to its pre fault level of activity. The increased complexity of the signaling between the astrocyte, synapses and neurons provides the capability to sense and repair synaptic connections, where the astrocyte regulates the degree of repair. Astrocytes are also connected via intracellular signaling routes (gap junctions) which allow the secondary messenger IP₃ (inositol trisphosphate) to pass through, thereby allowing astrocytes to communicate with each other providing a distributed repair-decision making capability. At an abstract level, one can view

astrocytes as a high level network which exercises plasticity over neural networks, with interactions between both networks occurring via the direct and indirect signaling pathways.

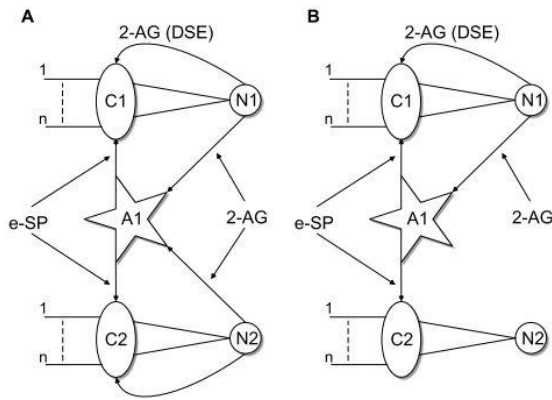


Fig. 2. Astrocyte feedback N1 and N2 depicts neurons and A1 an astrocyte, e-SP and DSE are excitatory and suppressive feedback signals.

Progress has been made in modelling the astrocyte process [23] and its interactions with Spiking Neural Networks (SNNs) [7]. These models are limited by the computational resources and the performance of simulations. SNNs have also been implemented using FPGA hardware and the level of parallelism exhibited by hardware, improves the SNN performance over that of software models along with lower power and area overhead. Therefore, it is timely to explore hardware emulation as hardware models are now more readily available, and thus it is possible to employ self-repair on a SNN using a bio-inspired model of astrocytes.

C. From Biology to Hardware

Recent work has implemented astrocyte cells in neuromorphic systems [8], [9] and digital circuits [10]–[13] with the aim of exploring their behaviour. None have considered the challenge of facilitating scalable interconnect for neuro-glia networks. NoCs were originally introduced as a solution to the interconnecting problem arising from the increasing number of cores on System on Chip (SoC) and Multi-Processor System on Chip (MPSoC) technology. The increased numbers of cores also enhanced the complexity of the wiring structure as there is a direct correlation with the number of individual cores and wires needed to connect these cores on

a SoC. NoCs use an interconnect architecture where routers and packets of data are used to transmit information between cores [14]–[16]. The many advantages of NoCs for SoC implementation include scalability, reduced overhead, low power consumption and reduced complexity. NoCs are appropriate for the SNN interconnect due to the many processing cores of a SoC. The communication between these cores is similar to the neuron/synapse connections and the communication between neurons: A SNN consists of processor cores (neurons), communication channels (synapses) and a topology structure (complex neuron interconnect) [24].

Fig. 3 illustrates the HNoC hierarchy [1], which uses 3 levels of routers in a hierarchical topology to connect neurons or node facilities. Level 1 (the neuron facility) consists of 10 neurons connected via a single node router in a star (direct) topology. Level 2 is the tile facility which connects 10 node routers via a tile router. Level 3 is the cluster facility, each cluster consists of four tile facilities. One cluster facility connects 400 neurons, and communicates between neurons using up and downstream communications between the internal tile and node facilities. The three layers of routing facilities provides HNoC with the architecture to connect many neural facilities. HNoC allows connections between cluster facilities thereby allowing a higher number of neurons to be connected by connecting neural tiles in North (N), East (E), South (S) and West (W) directions; i.e. there is no need for irregular wiring layouts which leads to complex and inefficient routing structures in hardware. Using routers in a hierarchical manner allows information to be passed from neuron to neuron using a bi-directional routing algorithm, with an up and downstream connection via routers. Therefore, neurons in one cluster may communicate to neurons in a separate cluster in a short number of hops between routers. Neurons communicate through transmission of spike events via synapses in the neural network while astrocytes communicate with each other via IP_3 in the astrocyte network. As the network scales this becomes an enormous and vast number of processing elements to be connected efficiently. Previous work has successfully implemented astrocyte communication between astrocyte cells using NoC [25], i.e. high-level communication using a novel ‘astrocyte-NoC’. However, it does not account for communicating between neurons and astrocytes at a low-level which is necessary for self-repair.

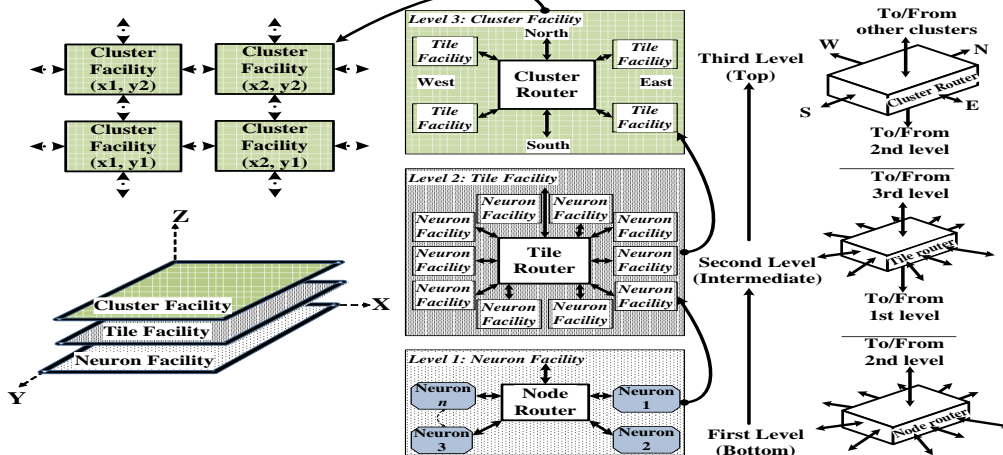


Fig. 3. The internal architecture of HNoC [1]

III. LOW-LEVEL NEURO-GLIA INTERCONNECT

This section presents the novel low-level NoC hardware communication architecture, which extends the HNoC paradigm to support scalable Neuro-glia network hardware. While the HNoC currently provides scalable communication for the SNN, the key focus of this paper is to facilitating low-level communication between the network of astrocyte cells and the SNN using a ring-topology (astrocyte-NoC). Implementing a neuro-glia network in hardware requires the low-level communication to not affect normal SNN activity while also allowing interactions between large numbers of spiking neurons via synaptic connections. In effect, the aim is to facilitate communication within and between each astrocyte and the SNN network, while forming a single unified neuro-glia network. Neurons communicate in a temporal manner while astrocytes communicate continuously, and at a much slower rate than that of neurons. Fig. 4 provides an overview of the interactions and communication exchange between neurons and astrocytes (inter-) and astrocyte and astrocyte (intra-) communication exchange.

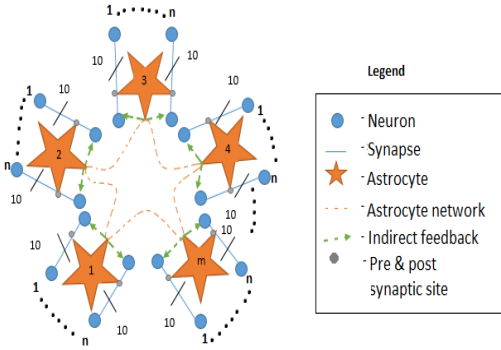


Fig. 4. Neuro-glia network overview

Astrocyte data is numerical in value, with more data (traffic) is exchanged at less demanding throughput rates than spike events. Neuron data is digital spike events. Fig. 5 identifies the key signals communicated in a neuro-glia network, with the astrocyte cell/cores and neurons depicted and illustrates the key parameters in the astrocyte repair process [7]. Glutamate is released from the presynaptic dendrite, 2AG as a result is released by the postsynaptic dendrite and triggers the oscillation of Ca^{2+} in the astrocyte. As a result, the Ca^{2+} and IP3 is used for the communication between astrocytes as well as driving the e-Sp and DSE signal which controls the probability of the synapse.

A. Communicating e-SP Data within Astrocyte-NoC

The HNoC and astrocyte-NoC operate in parallel and exchange data as required. The HNoC is based on a hierarchical structure consisting of a Node Facility, Tile Facility and Cluster Facility. At the lowest level, ten neurons are directly connected to a node router using a point to point or star topology. This provides each neuron with a direct connection to the node router where each spike is picked up by the node router. This data is then assembled and configured into packet form [1].

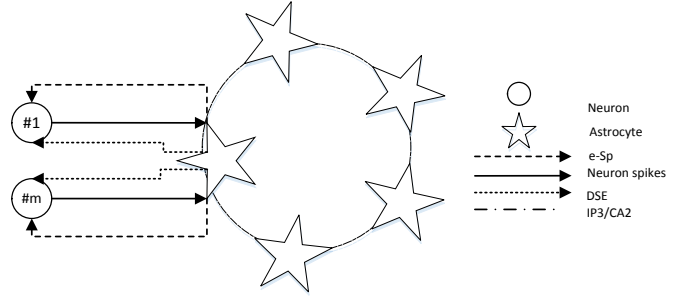


Fig. 5. Key signal communication in a neuro-glia network

The astrocyte captures data non-intrusively from the HNoC via an additional output port within the node router which clones the HNoC data packet and sends it to the astrocyte core. HNoC continues to communicate spike events via its interconnect, thereby permitting the astrocyte-NoC for inter-astrocyte and low-level astrocyte-neuron communication; i.e. does not impact on traffic data load in the HNoC. The astrocyte model consists of two 2-AG generators and an astrocyte; spikes from neurons produce DSE and 2-AG signals where the latter produces the e-Sp signal, as seen in Fig. 6.

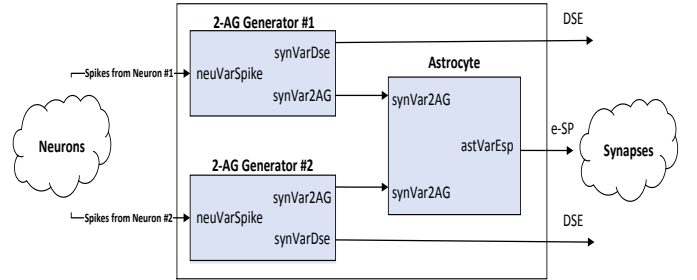


Fig. 6. 2-AG generator communicating to the Astrocyte producing e-SP

Figure 7 illustrates the HNoC-astrocyte-NoC interface. The astrocyte-NoC is therefore indirectly connected to ten neurons via a node router. The e-Sp signal is a global signal and is identical for every synapse and is therefore the most challenging to address due to its one-to-all connectivity requirement. At the lowest level of HNoC the node router is directly connected in a point-to-point fashion with the neurons (shown as #1 to #10) via ns synapses, (where ns is the max number of synapses per neuron). The node router connects to the upper tile and cluster facilities of the HNoC hierarchy. The synapses of the node interfaced with the astrocyte core using an ‘e-SP comms’ module. This module consists of an ‘e-SP Tx’ transmitter block and several ‘e-SP Rx’ receiver blocks (one connected to each neuron) which communicate via a serial link. The ‘e-SP Tx’ within the astrocyte NoC performs parallel to serial packetisation and the ‘e-SP Rx’ performs serial to parallel de-packetisation and storage (Fig. 8). The e-SP packet is serially propagated through each e-SP Rx module via the ring topology enabling all 10 neurons to receive and store the e-SP data value.

This, in essence, shows the two separate networks (HNoC and the astrocyte-NoC network) interacting at the lowest level possible and although they communicate at this low level, they work independently reducing the number of connections and information exchanges between both networks.

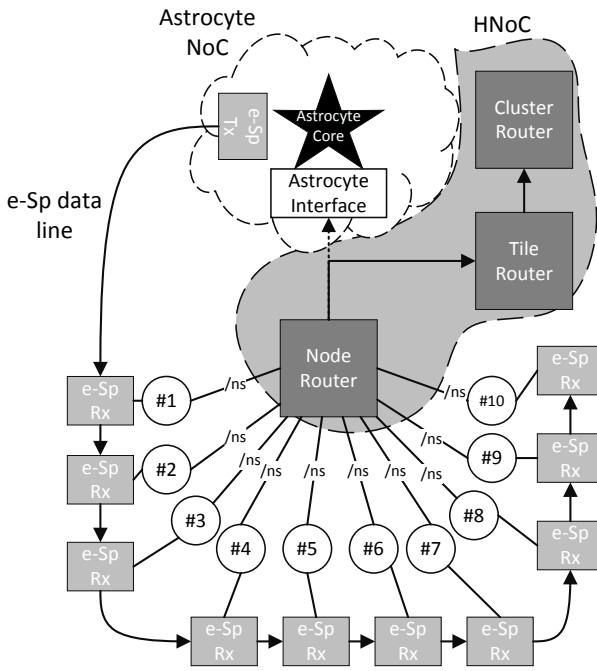


Fig. 7. The node facility of HNoC, the node router is responsible for receiving the spikes and depending on the address, it packetizes the information and either forwards it up a level to the tile facility or back down to the neurons within its facility, therefore it is the most efficient and least complex level to implement the astrocyte communication.

The e-SP packet is serialized in the ‘e-SP Tx’ block and then forwarded through a single wire to the ‘e-SP Rx’ block. Therefore, the e-SP signal is communicated in a serial fashion on a single line thereby reducing the level of wires required. Using one start bit the e-SP indicates the start of the information to be communicated and the ‘e-SP Tx’ block forwards the packet (in a series of bits) through the flip flop and into a shift register which stores the global e-SP value. A ring topology is used to address two key low-level issues within the neuro-glia implementation; (1) reduce the number of physical wires per node facility and (2), exploit the slower communication speeds of the biological e-SP signal (i.e. spike events are typically 2-3 orders of magnitude faster in exchange rates). The ring topology has previously shown benefits in area-speed trade-offs for area SNN hardware [26]. Therefore, the e-SP communication is based on a ring or daisy chain topology for scalability. It exploits the one-to-many global communication

between the neurons and shown in Fig.7 to traverse in a ring fashion around all 10 neurons in a node tile of the HNoC.

B. NoC Data Format and Ring Protocol

In Fig. 8, the ‘e-SP comms Tx’ module outputs a 64 bit packet which is 64 bits [13]. The 64 bit precision is a result of the double point precision based on the current astrocyte core [13]. this is significant in bit-size, in terms of communicating the value globally, it becomes area inefficient to communicate directly to each neuron (640 direct lines). Due to the astrocyte communicating at a slower rate in biology compared to neurons, the propagation speed of the e-SP signal is not demanding; this slow communication time may be exploited by using a serial link and ring topology.

Fundamentally, the e-SP packet size depends on the bit-resolution of the astrocyte model. However, the worst case scenario of 64 bits indicates that there are 64 FIFOs (First in First out). The ‘e-SP comms’ module which is made up of one ‘e-SP Tx’ logic block used to interface the astrocyte core back to the neurons within HNoC. The e-SP data is converted into a series of bits using a parallel to serial conversion. This series of bits is then forwarded to the ‘e-SP Rx’ block, within which is a D flip-flop which simply acts as a buffer. The output is fed into a JK flip flop and a shift register. The serial data is passed through to a shift register (i.e. FIFO) consisting of a series of registers (each neuron from #1 to #m has FIFOs from bit 1 to bit N_R) the size of shift register depends on the size of the packet. The JK flip-flop within the ‘e-SP Rx’ is utilised to automatically stop the shift register from receiving data bits when the FIFO has received all of its data. The size of the packet has a direct impact on the size of the shift register. Fig. 9 and 10 show HNoC and the e-SP packet layout, respectively. The payload within the ‘e-SP comms’ packet size is indicated as 64 bits, this increases the number of wires as it correlates with the increasing packet size and thus the buffer size increases. Using a ring topology minimizes the overhead induced by the increasing buffer size. The e-SP packet is decoded and sent serially, therefore there is one wire from the astrocyte communicated back to the synapses. The serial message will continue in this same manner through all the neurons and associated synapses. Although the tradeoff is speed of communication, there is no traffic congestion caused by the ring approach due to the information eventually making its way to associated synapses at each of the neurons.

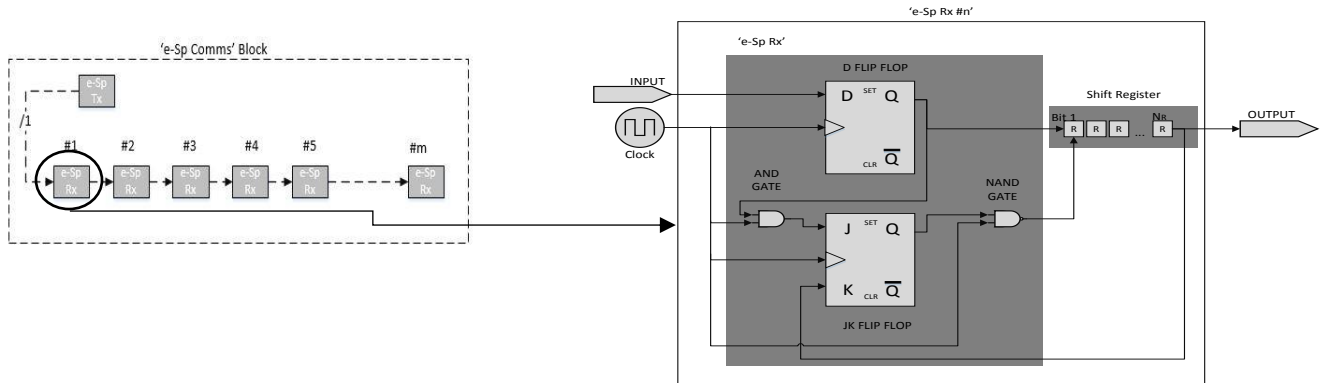


Fig. 8. ‘e-SP comms’ Block using a series of low level flip flops, the packet is broken down into the relative number of bits and sent using the ‘e-SP Tx’ module. This module communicates the e-SP to every ‘e-SP Rx’ connected in the ring topology. The ‘e-SP Rx’ requires a start bit which will begin the transmission of the bits through the buffer, it also requires an end bit which will indicate the end of data to be communicated and the ‘e-SP Tx’ will automatically stop sending data.

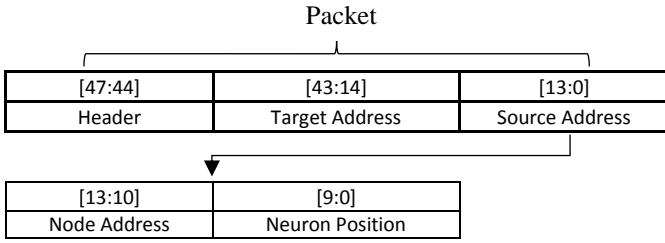


Fig. 9. This is a typical HNoC packet, it is 48 bits consisting of a Header, Target Address and Source Address. This contains all the information for traversing data in HNoC.

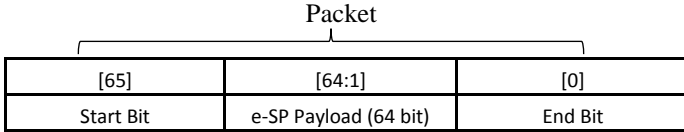


Fig. 10. A typical e-SP packet is 66 bits in size. Due to a direct connection from the astrocyte to all the neurons in a single node facility there is no need for an address as the data across all synapses is identical. There is a start bit and end bit, required to switch the transmission on or off.

IV. RESULTS

This section outlines the test bench and provides performance analysis of the ring-topology for the ‘e-SP comms’ block. The area overhead of the e-SP communication is compared with the HNoC and an astrocyte core to demonstrate its compactness and it is assessed regarding area utilization for various packet sizes.

A. Testbench and setup

The HNoC neuron facility, astrocyte cell and astrocyte NoC ring have been captured in VHDL and synthesised for a Xilinx Virtex-7 XC7VX485T-2FFG1761C FPGA evaluation board using Xilinx Vivado 2016.2. The NoC ring topology with astrocyte to neuron communication has been validated on the FPGA Xilinx ARTY 35T evaluation board. The astrocyte accepted packets in the form of 8 bits (a spike would be represented as binary “00000001”) and produced a number of signals including the global e-SP signal, this signal was then interfaced with a low level logic block known as the ‘e-SP comms’ block and as the information was in the format of a 64 bit packet, the packet was then decoded into a series of bits, starting with a ‘1’ bit and ending with a ‘1’ bit (this was to autonomously start and end the information communicated through the ring topology). The ‘e-SP comms’ block was then interfaced with the neurons using a single wire and a ring topology and communicated the series of bits through to a FIFO shift register, the e-SP signal was therefore communicated serially to all neurons within a single neuron facility.

B. Analysis of the ‘e-SP comms’ Block

The area overhead incurred by the ‘e-SP comms’ block, defined in Table 1, correlates with the size of the e-SP packet data i.e. the size of the packet generated by the astrocyte, which at this point in time is a fixed value. In terms of scalability the number of look up tables (LUTs) and slice registers were used to determine how the ‘e-SP comms’ block scaled in comparison to both the area consumed by the HNoC Neuron facility and the astrocyte core. Fig. 11 shows the difference in the number of physical wires and latency incurred when using a ring topology

against using a star topology. An increased number of wires is required using a direct topology and results in higher latency with a much higher communication speed. However, because of the inherent slow speed of the astrocyte communication in biology the trade-off for speed is to have a more efficient communication in terms of less latency and lower area overhead. Due to a low number of wires the ring topology is area efficient with no latency as the network traffic traverses around the ring without congestion. In terms of both LUTs and slice registers (area utilization) the ‘e-SP comms’ block is very small compared to the HNoC neuron facility with which it communicates. Fig. 12 illustrates the LUT resource usage of the ‘e-SP comms’ as the network scales, with reference to the HNoC neuron facility. The ‘e-SP comms’ block scales more linearly than the HNoC element.

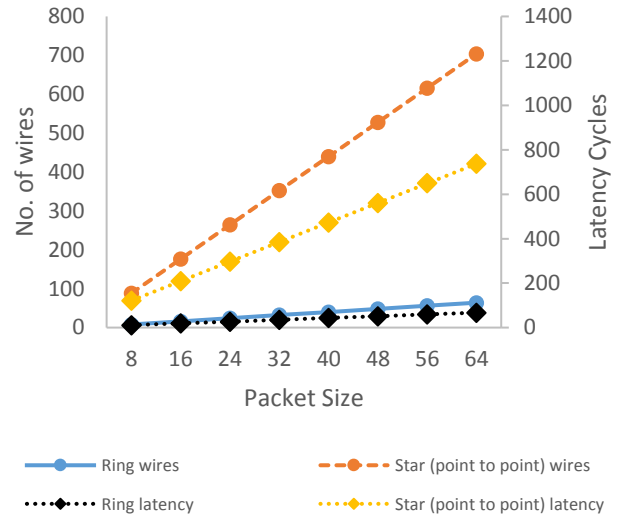


Fig. 11. ‘e-SP comms’ block scalability

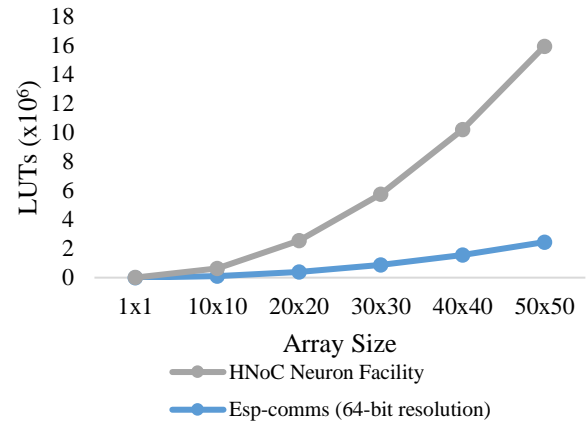


Fig. 12. ‘e-SP comms’ block vs HNoC neuron facility

Table I. compares area overhead with the astrocyte and HNoC neuron facility the area utilization is shown in terms of percentages due to the size of the astrocyte core. Therefore referencing the astrocyte as the maximum in terms of area overhead, the HNoC neuron facility is around 3.34% in terms of LUTs and 4.51% in terms of slice registers. The e-SP a mere 0.61% in terms of LUTs and 1.2% in slice registers at a 1:1 ratio.

TABLE I. ‘e-SP COMMS’ BLOCK EVALUATION

Component	LUTs	(%)	Slice Registers	(%)
Astrocyte	16,182	-	16,305	-
HNoC Neuron-Facility	540	3.34	735	4.51
Esp-comms	98	0.61	198	1.21

This indicates that the ‘e-SP comms’ block incurs very little area as a result of successfully communicating e-SP back to synapses. The second constraint which affects the area overhead is the number of bits in the ‘e-SP comms’ signal, i.e. the packet size. Due to the worst case scenario of 64 bits being used, it is important to realize that a more efficient packet size would result in the area overhead decreasing; therefore this indicates the impact a 64 bit packet has when the neuro-glia network array size is increased. Fig. 13 shows the number of LUTs (area utilization) in the ‘e-SP comms’ block, when applied across different packet sizes (e.g. between 8 to 64 bit) it shows the area overhead incurred when this is also scaled up (between 10x10 and 50x50) this shows the LUT (area) has a gradual increase and can be reduced by minimizing the e-SP packet size; i.e. optimize the astrocyte core.

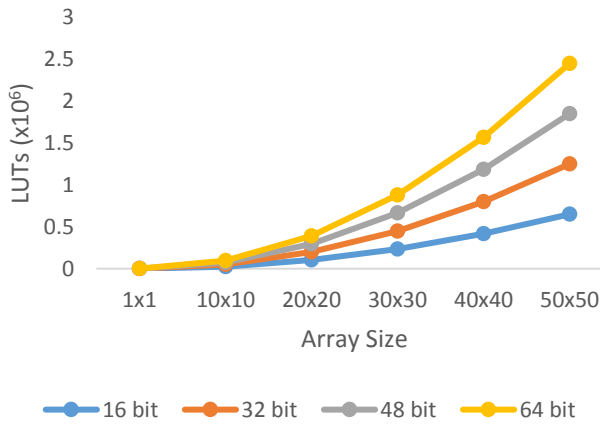


Fig. 13. ‘e-SP comms’ block affected by packet size

The results show that the area overhead incurred by adding the ‘e-SP comms’ is minimal. Ultimately, provision of astrocyte communication with minimal area overhead enables neuro-glia network sizes which can be used to realize future self-repair systems.

V. CONCLUSION

Implementing a neuro-glia network requires an interconnect with a low area overhead, due to the vast size of both individual networks, the number of processing elements (neurons and astrocytes) and communication signals, the interfacing of the two completely different networks is a complex challenge. Previous work on high level astrocyte to astrocyte communications and this low level astrocyte to neuron/synapse communication indicates NoCs provide a scalable solution to the interconnect challenge. The use of the ring topology in the NoC provides a good trade-off between reducing area/wire

overheads and relaxing the communication speed of data provided by the astrocyte to synapses/neurons.

This novel NoC interconnection scheme for communicating e-SP enables a significant number of astrocytes to communicate with neurons within a minimal area constraint. This enables self-repair emulation with a distributed and fine grained nature without a central controller and is based on the biological and computational models of previous works, thereby allowing a SNN to operate in parallel with an astrocyte network. This low level interconnect in addition to the high level astrocyte communications already developed provide a platform for developing a neuro-glia interconnect for future inspired computing paradigms regarding self-repair strategies in hardware. Future work with astrocyte-neuron networks aims to reduce latency and area using a more efficient communication protocol; also explore reduction of the packet size which influences the NoC area. NoC communication in this manner is a promising interconnect solution with minimal area overhead providing the communication which is capable of facilitating self-repair at a fine grained and distributed level.

REFERENCES

- [1] S. Carrillo, J. Harkin, L. J. McDaid, F. Morgan, S. Pande, S. Cawley, and B. McGinley, “Scalable hierarchical network-on-chip architecture for spiking neural network hardware implementations,” *IEEE Trans. Parallel Distrib. Syst.*, vol. 24, no. 12, pp. 2451–2461, 2013.
- [2] K. Reick, P. N. Sanda, S. Swaney, J. W. Kellington, M. Mack, M. Floyd, and D. Henderson, “Fault-tolerant design of the IBM Power6 microprocessor,” *IEEE Micro*, vol. 28, no. 2, pp. 30–38, 2008.
- [3] K. Kyriakoulakos and D. Pnevmatikatos, “A novel SRAM-based FPGA architecture for efficient TMR fault tolerance support,” *FPL 09 19th Int. Conf. F. Program. Log. Appl.*, pp. 193–198, 2009.
- [4] J. Liu, J. Harkin, Y. Li, and L. Maguire, “Online fault detection for Networks-on-Chip interconnect,” *Proc. 2014 NASA/ESA Conf. Adapt. Hardw. Syst. AHS 2014*, pp. 31–38, 2014.
- [5] J. J. Wade, L. J. McDaid, J. Harkin, V. Crunelli, J. A. S. Kelso, and V. Beiu, “Exploring retrograde signaling via astrocytes as a mechanism for self repair,” *Proc. Int. Jt. Conf. Neural Networks*, pp. 3149–3155, Jul. 2011.
- [6] J. Wade, L. McDaid, J. Harkin, V. Crunelli, and S. Kelso, “Self-repair in a bidirectionally coupled astrocyte-neuron (AN) system based on retrograde signaling,” *Front. Comput. Neurosci.*, vol. 6, no. September, p. 76, Jan. 2012.
- [7] M. Naeem, L. J. McDaid, J. Harkin, J. J. Wade, and J. Marsland, “On the Role of Astroglial Syncytia in Self-Repairing Spiking Neural Networks,” *IEEE Trans. Neural Networks Learn. Syst.*, vol. 26, no. 10, pp. 2370–2380, 2015.
- [8] Y. Irizarry-Valle and A. C. Parker, “Astrocyte on neuronal phase synchrony in CMOS,” *Proc. - IEEE Int. Symp. Circuits Syst.*, pp. 261–264, 2014.
- [9] Y. Irizarry-Valle and A. C. Parker, “An astrocyte neuromorphic circuit that influences neuronal phase synchrony,” *IEEE Trans. Biomed. Circuits Syst.*, vol. 9, no. 2, pp. 175–187, 2015.
- [10] B. A. Abed, A. Ismail, and N. A. Aziz, “Real time astrocyte in spiking neural network,” *SAI Intell. Syst. Conf. 2015*, pp. 565–570, 2015.
- [11] H. Soleimani, M. Bavandpour, A. Ahmadi, and D. Abbott, “Digital implementation of a biological astrocyte model and its application,” *IEEE Trans. Neural Networks Learn. Syst.*, vol. 26, no. 1, pp. 127–139, Jan. 2015.
- [12] M. Hayati, M. Nouri, S. Haghiri, and D. Abbott, “A Digital Realization of Astrocyte and Neural Glial Interactions,” *IEEE Trans. Biomed. Circuits Syst.*, vol. 10, no. 2, pp. 518–529, 2016.

- [13] J. Liu, J. Harkin, L. Maguire, L. McDaid, J. Wade, and M. McElholm, "Self-Repairing Hardware with Astrocyte-Neuron Networks," *IEEE Int. Symp. Circuits Syst.*, pp. 1–4, 2016.
- [14] L. Benini and G. De Micheli, "Networks on chips: A new SoC paradigm," *Computer (Long Beach, Calif.)*, vol. 35, no. 1, pp. 70–78, 2002.
- [15] W. J. Dally and B. Towles, "Design Automation Conference, 2001. Proceedings," *Des. Autom. Conf. 2001. Proc.*, pp. 684–689, 2001.
- [16] A. Hemani, A. Jantsch, S. Kumar, A. Postula, J. Oberg, M. Millberg, and D. Lindqvist, "Network on chip: An architecture for billion transistor era," *Proceeding IEEE NorChip Conf.*, pp. 166–173, 2000.
- [17] S. Carrillo, J. Harkin, L. McDaid, S. Pande, S. Cawley, and F. Morgan, "Adaptive routing strategies for large scale spiking neural network hardware implementations," *Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics)*, vol. 6791 LNCS, no. PART 1, pp. 77–84, 2011.
- [18] B. V. Benjamin, P. Gao, E. McQuinn, S. Choudhary, A. R. Chandrasekaran, J. M. Bussat, R. Alvarez-Icaza, J. V. Arthur, P. A. Merolla, and K. Boahen, "Neurogrid: A mixed-analog-digital multichip system for large-scale neural simulations," *Proc. IEEE*, vol. 102, no. 5, pp. 699–716, May 2014.
- [19] E. Painkras, L. A. Plana, J. Garside, S. Temple, F. Galluppi, C. Patterson, D. R. Lester, A. D. Brown, and S. B. Furber, "SpiNNaker: A 1-W 18-core system-on-chip for massively-parallel neural network simulation," *IEEE J. Solid-State Circuits*, vol. 48, no. 8, pp. 1943–1953, 2013.
- [20] J. Schemmel, J. Fieres, and K. Meier, "Wafer-scale integration of analog neural networks," *Proc. Int. Jt. Conf. Neural Networks*, pp. 431–438, Jun. 2008.
- [21] S. Murali, T. Theocharides, N. Vijaykrishnan, M. J. Irwin, L. Benini, and G. De Micheli, "Analysis of error recovery schemes for networks on chips," *IEEE Des. Test Comput.*, vol. 22, no. 5, pp. 434–442, 2005.
- [22] S. Mitra, W. J. Huang, N. R. Saxena, S. Y. Yu, and E. J. McCluskey, "Reconfigurable architecture for autonomous self-repair," *IEEE Des. Test Comput.*, vol. 21, no. 3, pp. 228–240, 2004.
- [23] J. J. Wade, L. J. McDaid, J. Harkin, V. Crunelli, and J. A. S. Kelso, "Bidirectional coupling between astrocytes and neurons mediates learning and dynamic coordination in the brain: A multiple modeling approach," *PLoS One*, vol. 6, no. 12, p. e29445, Jan. 2011.
- [24] R. Emery, A. Yakovlev, and G. Chester, "Connection-centric network for spiking neural networks," *Proc. - 2009 3rd ACM/IEEE Int. Symp. Networks-on-Chip, NoCS 2009*, pp. 144–152, 2009.
- [25] J. Liu, J. Harkin, L. McDaid, and G. Martin, "Hierarchical Networks-on-Chip Interconnect for Astrocyte-Neuron Network Hardware," *International Conference on Artificial Neural Networks* Sept, 2016 In Press.
- [26] S. Pande, F. Morgan, G. Smit, T. Bruintjes, J. Rutgers, B. McGinley, S. Cawley, J. Harkin, and L. McDaid, "Fixed latency on-chip interconnect for hardware spiking neural network architectures," *Parallel Comput.*, vol. 39, no. 9, pp. 357–371, 2013.