Technical Report Ref. TR-OASIS-NP-102015, ASL, 04/2015

OASIS-NP

Abderazek Ben Abdallah
Adaptive Systems Laboratory

Abstract

The biological brain implements massively parallel computations using a complex architecture that is different from current Von Neumann machine. The brain is a low-power, fault-tolerant and high-performance machine! It consumes only about 20W and brain circuits continue to operate as the organism needs even when the circuit (neuron, neuroglia, etc.) is perturbed or died. Our goal in this project is to research and develop in hardware an adaptive low-power and reliable neuro-inspired manycore SoC with on-chip learning and cognitive capabilities targeted for pattern recognition and complex cognitive tasks. Our other goal is to investigate and develop a low-power and low-cost platform for running large-scale simulations of biological brains in real-time targeted for neuroscience applications. Currently, we are investigating the following problems: the communication network for neuro-inspired chips, reconfigurability and adaptability methods, fault-tolerance, and learning circuits. In addition to these two target applications, lessons learned from this project will also be used to optimize power & performance of the conventional architectures.

Introduction

Current neural processing methods are based on so-called spiking neural networks (SNNs), which differ from conventional artificial NN models. In conventional SNNs information is transmitted using spikes or pulses (see Fig. 1). Brain-inspired models such as SNNs offer the opportunity to design a robust architecture for performing computation with low power. In particular, SNNs offer the potential to mimic the ability of the brain to tolerate faults and repair itself [1]. The brain often replaces neurons by reconnecting to newly generated neurons via synaptic junctions. This adaptability allows the brain to overcome faults.

Software simulations of traditional SNN networks face the problem of scalability in that biological computing systems are inherently parallel in their architecture whereas conventional systems are based on sequential processing architectures.

Hardware SNNs have the advantage of computational speed over software simulations and can take full advantage of their inherent parallelism. In particular, hardware SNNs can respond to the demands of real-time and fault-tolerant applications. However, the current problem of interneuron connectivity is prohibiting the implementation of these SNNs in hardware. This connectivity challenge more challenging when neuro-glia networks are considered.

One promising approach to this challenge is the use of Network-on-Chip (NoC) to support the fast exchange of information within an SNN. NoC uses packet to exchange information between different processing elements within a system.

Our goal is to develop dedicated mixed signals programmable Neural-Network architecture which will support the implementation of large-scale SNN applications.

Current FPGAs do not provide sufficient dedicated programmable hardware resources (i.e. synaptic junctions, inline training support) which needed for efficient SNN implementations. Also, FPGAs routing structures cannot accommodate the high levels of interconnection found in SNNs.

The proposed OSIS-NP architecture avoids these problems by making use of small-scale low-powered analog spiking neuron cells as the central PE. Combining small size low power neuron cells with the OASIS-NoC routing capability will lead to OASIS-NP system that can significantly increase the processing power of SNNs towards the biological scale. However, the use of the OASIS-NoC as an interconnection fabric for large scale neuro-glia networks demands a more challenging trade-off between power consumption, throughput and network scalability, and therefore an improved solution is required. Scaling is one of the key issues associated with all complex electronic systems because interconnect consumes large areas of chip real-estate and subsequently causes longer critical path delays [1].

To solve the large scale implementation issues of reconfigurability and programmability, lightweight synapse weight storage and interneuron connectivity, a new efficient architecture is required.

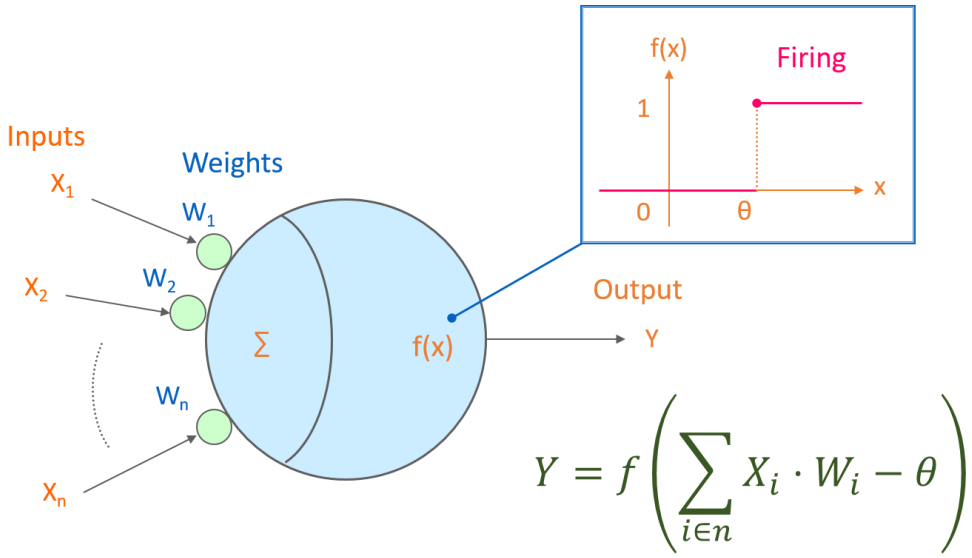
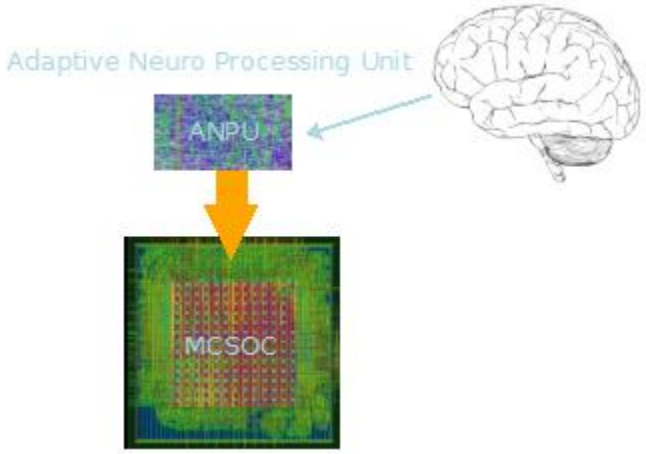


Figure 1. Neuron Model (McCulloch-Pitts Model of Neuron)

To demonstrate the OASIS-NoC implementation of an SNN, an FPGA hardware implementation will be developed and interfaced to a drone or a robot. This project is based on an existing scalable OASIS-NoC routers developed at ASL [2-21] and explored the following issues in the development of a new NoC strategy for SNN:
(1) An efficient topology and configurations which enable reconfiguration of different NoC networks in hardware;
(2) A mixture of different basic NoC topologies (OASIS-2D, OASIS-3D, etc.) will be investigated;
(3) Performance evaluation and comparisons against existing systems.



OASIS-NP High-Level View

Hardware SNN Architectures

A major challenge in developing SNNs in hardware is the development of neuro-inspired platforms which can support scalable low-power realizations and reprogrammable interconnect. In particular, the problem of interneuron connectivity is the major problem that prohibits the implementation of such SNNs.

Current attempts to realize SNNs using multiprocessors in hardware have included limited numbers of neurons due to connectivity issues [22-27].
Several full-custom neuromorphic architecture devices have been proposed [28-30] which aim to address the inefficiencies of FPGAs by including optimized synaptic cells and Address Event Representation (AER) routing [33]. However, these systems are not reconfigurable and cannot scale due to limited connectivity provided by AER between neurons.

References

1. Jim Harkin, Fearghal Morgan, Liam McDaid, Steve Hall, Brian McGinley, and Seamus Cawley, “A Reconfigurable and Biologically Inspired Paradigm for Computation Using Network-On-Chip and Spiking Neural Networks, “International Journal of Reconfigurable Computing, vol. 2009, Article ID 908740, 13 pages, 2009. doi:10.1155/2009/908740
2. Khanh N. Dang, Michael Meyer, Yuichi Okuyama, Abderazek Ben Abdallah, Xuan-Tu Tran, “A Soft-Error Resilient 3D Network-on-Chip Router”, Proc. of IEEE 7th International Conference on Awareness Science and Technology (iCAST 2015), pp. 84 – 90, Sep. 22-24, 2015.[Slides.pdf]
3. Abderazek Ben Abdallah, Mitsuhiro Nakamura, Akram Ben Ahmed, Michael Meyer, Yuichi Okuyama, “Fault-tolerant Router for Highly-reliable Many-core 3D-NoC Systems”, Proc. of the 3rd International Scientific Conference on Engineering and Applied Sciences (ISCEAS 2015), July 29-31, 2015, Okinawa, Japan.
4. Akram Ben Ahmed, High Throughput Architecture and Routing Algorithms Towards the Design of Reliable Many-Core Network-on-Chip Systems, Doctoral Thesis Preliminary Presentation, October 20, 2014.
5. A. Ben Ahmed, M. Meyer, Y. Okuyama, and A. Ben Abdallah, Adaptive Error- and Traffic Aware Router Architecture for 3D Network-on-Chip Systems, IEEE Proceedings of the 8th International Symposium on Embedded Multicore/Many-core SoCs (MCSoC-14), pp. 197-2014, Sept. 2014.
6. A. Ben Ahmed, A. Ben Abdallah, OASIS 3D-Router Hardware Physical Design, Technical Report, Adaptive Systems Laboratory, Division of Computer Engineering, School of Computer Science and Engineering, University of Aizu, July 8, 2014.

7. Akram Ben Ahmed, A. Ben Abdallah, Graceful Deadlock-Free Fault-Tolerant Routing Algorithm for 3D Network-on-Chip Architectures, Journal of Parallel and Distributed Computing, 2014. [DOI]
8. Akram Ben Ahmed, Achraf Ben Ahmed, A. Ben Abdallah, Deadlock Recovery Support for Fault-tolerant Routing Algorithms in 3D-NoC Architectures, IEEE Proceedings of the 7th International Symposium on Embedded Multicore/Many-core SoCs (MCSoC-13), pp., 2013. [DOI]
9. Akram Ben Ahmed, A. Ben Abdallah, Architecture and Design of High-throughput, Low-latency and Fault-Tolerant Routing Algorithm for 3D-Network-on-Chip, The Jnl. of Supercomputing, December 2013, Volume 66, Issue 3, pp 1507-1532. [DOI]
10. Akram Ben Ahmed, T. Ouchi, S. Miura, A. Ben Abdallah, ”Run-Time Monitoring Mechanism for Efficient Design of Application-specific NoC Architectures in Multi/Manycore Era,” ”’ IEEE Proc. of the 6th International Workshop on Engineering Parallel and Multicore Systems (ePaMuS2013′), July 2013.”’ [DOI]
11. Akram Ben Ahmed, T. Ouchi, S. Miura, A. Ben Abdallah, Run-Time Monitoring Mechanism for Efficient Design of Application-specific NoC Architectures in Multi/Manycore Era, Proc. IEEE 6th International Workshop on Engineering Parallel and Multicore Systems (ePaMuS2013′), July 2013.
12. Akram Ben Ahmed, A. Ben Abdallah, ”Low-overhead Routing Algorithm for 3D Network-on-Chip”, ”’IEEE Proc. of the The Third International Conference on Networking and Computing (ICNC’12), pp. 23-32, 2012.”’ [DOI]
13. Akram Ben Ahmed, A. Ben Abdallah, ”LA-XYZ: Low Latency, High Throughput Look-Ahead Routing Algorithm for 3D Network-on-Chip (3D-NoC) Architecture”, ”’IEEE Proceedings of the 6th International Symposium on Embedded Multicore SoCs (MCSoC-12), pp. 167-174, 2012. [DOI]
14. Akram Ben Ahmed, A. Ben Abdallah, ”ONoC-SPL Customized Network-on-Chip (NoC) Architecture and Prototyping for Data-intensive Computation Applications,” ”’IEEE Proceedings of The 4th International Conference on Awareness Science and Technology, pp. 257-262, 2012. DOI
15. A. Ben Ahmed, A. Ben Abdallah, K. Kuroda, Architecture and Design of Efficient 3D Network-on-Chip (3D NoC) for Custom Multicore SoC, IEEE Proc. of the 5th International Conference on Broadband, Wireless Computing, Communication and Applications (BWCCA-2010), pp.67-73, Nov. 2010. (best paper award) (Slides), (Papers)
16. K. Mori, A. Esch, A. Ben Abdallah, K. Kuroda, Advanced Design Issues for OASIS Network-on-Chip Architecture, IEEE Proc. of the 5th International Conference on Broadband, Wireless Computing, Communication and Applications (BWCCA-2010),pp.74-79, Nov. 2010. (Slides); (Paper)
17. Shohei Miura, Abderazek Ben Abdallah, Kenichi Kuroda, 設計空間探索と MCSoC の生成に適している parameterizable NoC (PNoC)のハードウェア設計と事前評価, 第 34 回パルテノン研究会, pp.105-108. Aug. 2009. SLIDE
18. Shohei Miura, Abderazek Ben Abdallah, Kenichi Kuroda, PNoC: Design and Preliminary Evaluation of a Parameterizable NoC for MCSoC Generation and Design Space Exploration, The 19th Intelligent System Symposium (FAN 2009), pp.314-317, Sep.2009. (Slides)
19. Kenichi Mori, Abderazek Ben Abdallah, Kenichi Kuroda, Design and Evaluation of a Complexity-Effective Network-on-Chip Architecture on FPGA, The 19th Intelligent System Symposium (FAN 2009), pp.318-321, Sep. 2009. (Slides)
20. A. Ben Abdallah, T. Yoshinaga and M. Sowa, “Mathematical Model for Multiobjective Synthesis of NoC Architectures,” IEEE Proc. of the 36th International Conference on Parallel Processing, Sept. 4-8, 2007. (PDF)
21. A. Ben Abdallah, Masahiro Sowa, “Basic Network-on-Chip Interconnection for Future Gigascale MCSoCs Applications: Communication and Computation Orthogonalization,” JASSST2006, Dec. 4-9th, 2006. (PDF)

22. L. P. Maguire, T. M. McGinnity, B. Glackin, A. Ghani, A. Belatreche, and J. Harkin, “Challenges for large-scale implementations of spiking neural networks on FPGAs,” *Neurocomputing*, vol. 71, no. 1–3, pp. 13–29, 2007.
23. B. Glackin, et al., “Novel approach for the implementation of large-scale spiking neural networks on FPGAs,” in *Proceedings of the Artificial Neural Network Conference*, pp. 552–563, 2005.
24. E. Ros, E. M. Ortigosa, R. Agis, R. Carrillo, and M. Arnold, “Real-time computing platform for spiking neurons (RTspike),” *IEEE Transactions on Neural Networks*, vol. 17, no. 4, pp. 1050–1063, 2006.
25. M. M. Khan, D. R. Lester, L. A. Plana, et al., “SpiNNaker: mapping neural networks onto a massively-parallel chip multiprocessor,” in *Proceedings of the International Joint Conference on Neural Networks (IJCNN ’08)*, pp. 2849–2856, Hong Kong, June 2008.
26. A. D. Rast, S. Yang, M. Khan, and S. B. Furber, “Virtual synaptic interconnect using an asynchronous network-on-chip,” in *Proceedings of the International Joint Conference on Neural Networks (IJCNN ’08)*, pp. 2727–2734, Hong Kong, June 2008.
27. B. Glackin, J. Harkin, T. M. McGinnity, and L. P. Maguire, “A hardware accelerated simulation environment for spiking neural networks,” in *Proceedings of the Applied Reconfigurable Computing Workshop*, pp. 336–341, 2009.
28. R. J. Vogelstein, U. Mallik, J. T. Vogelstein, and G. Cauwenberghs, “Dynamically reconfigurable silicon array of spiking neurons with conductance-based synapses,” *IEEE Transactions on Neural Networks*, vol. 18, no. 1, pp. 253–265, 2007
29. P. A. Merolla, J. V. Arthur, B. E. Shi, and K. A. Boahen, “Expandable networks for neuromorphic chips,” *IEEE Transactions on Circuits and Systems I*, vol. 54, no. 2, pp. 301–311, 2007.
32. J. Schemmel, et al., “Mixed-mode analog NN using current steering synapses,” *Analog Integrated Circuits & Signal Processing*, vol. 38, 2004.
33. J. Harkin and M. McElholm, “Novel Interconnect strategy for large scale implementations of NNs,” *IEEE Soft Computing in Industrial Applications*, July 2007.