Neuromorphic Array Communications Controller

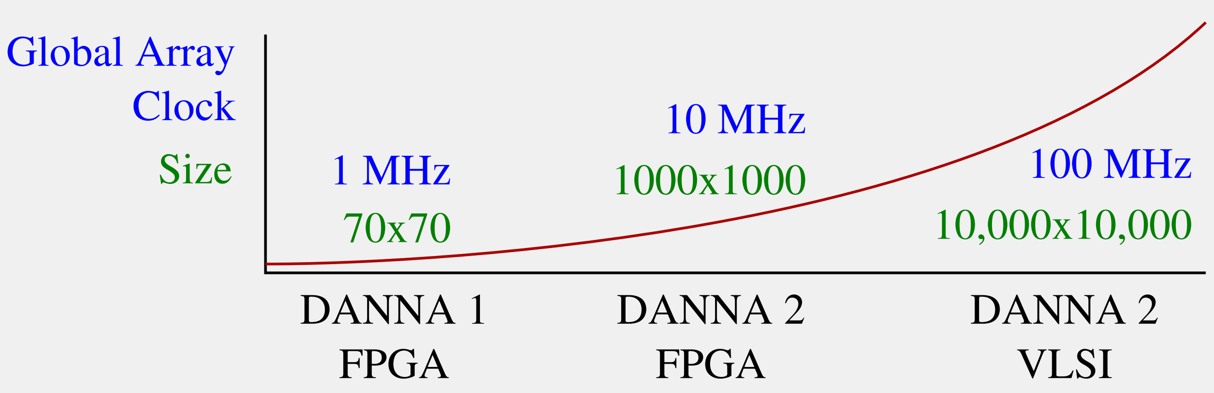
Aaron R. Young[1] and Mark E. Dean[1]

[1] University of Tennessee, Knoxville, Tennessee, U.S.A

TENN LAB NEUROMORPHIC ARCHITECTURES. LEARNING. APPLICATIONS.

Background and Motivation

- Neuromorphic hardware is increasing in both speed and size.

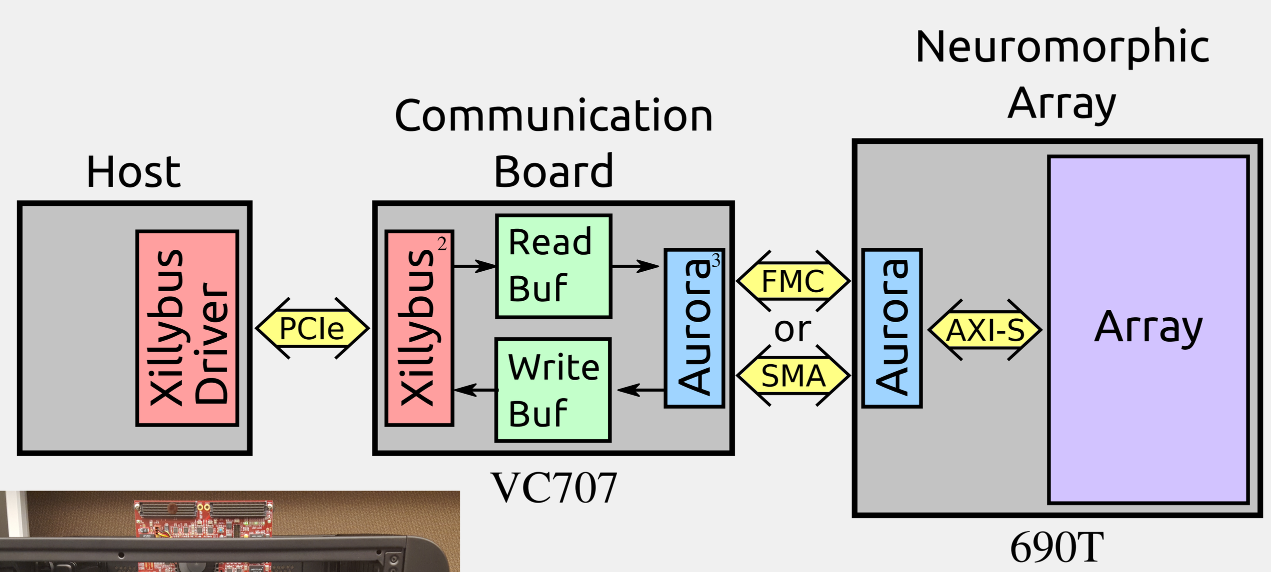


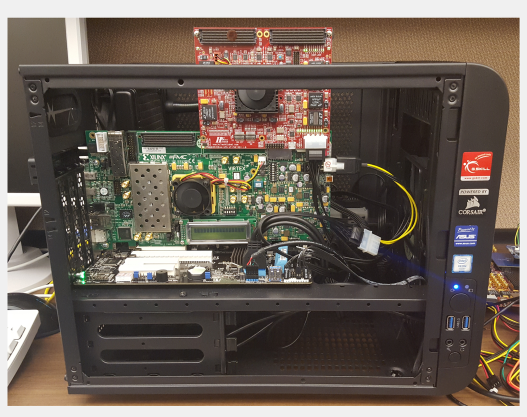
- Previous communication method with USB 3 Cypress EZ-USB FX3[1] was maxed out with DANNA 1.
- Communication patterns and requirements unique to spatial-temporal spiking data.

Communications Considerations

Monitoring	• Real-time monitoring for developing and debugging. • Provides valuable feedback to analyze the system. • Detect security or safety vulnerability.
Optimization	• Host can be used to drive real-time learning and optimization of the neuromorphic network via evolving networks at runtime.
Host to Array Communication	• Operational commands (Configuration, Control) • Real-time data (Input Spikes, Output Spikes)
Scale to External Interfaces	• Translate input/output between spiking and non-spiking to allow for connection with external devices.
Inter Sub-array Operation	• The sub-arrays need to be able to function together as a large array of elements, capable of running large neural networks.

Communication Testing

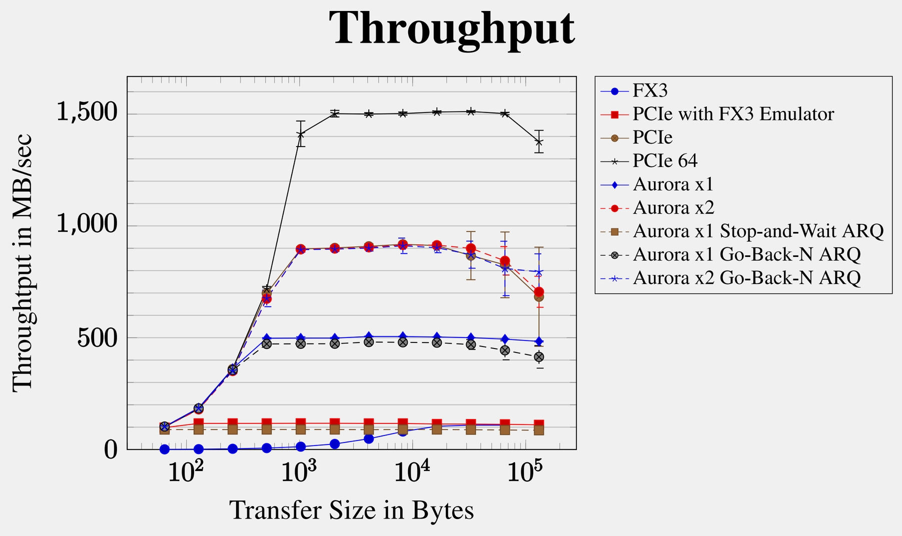




FPGA Boards used for test setup:

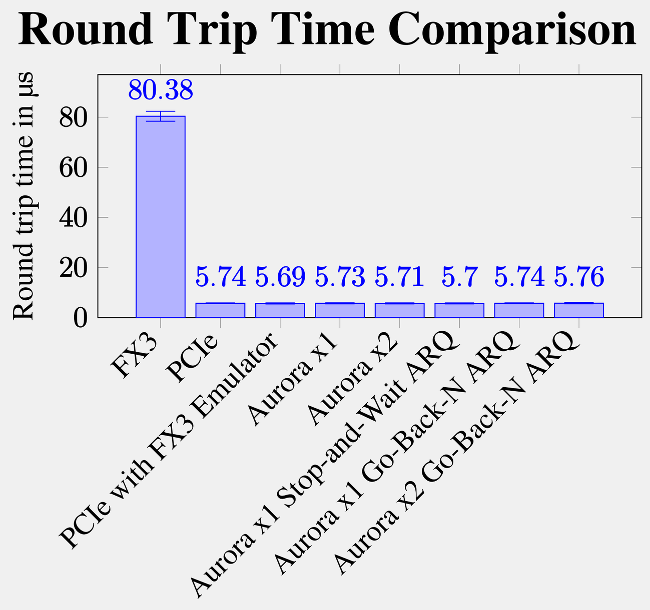
- VC707 with Virtex7 485T.
- HTG-777 with Virtex7 690T.

Throughput



Throughput experiment, varying the amount of data transmitted per function call.

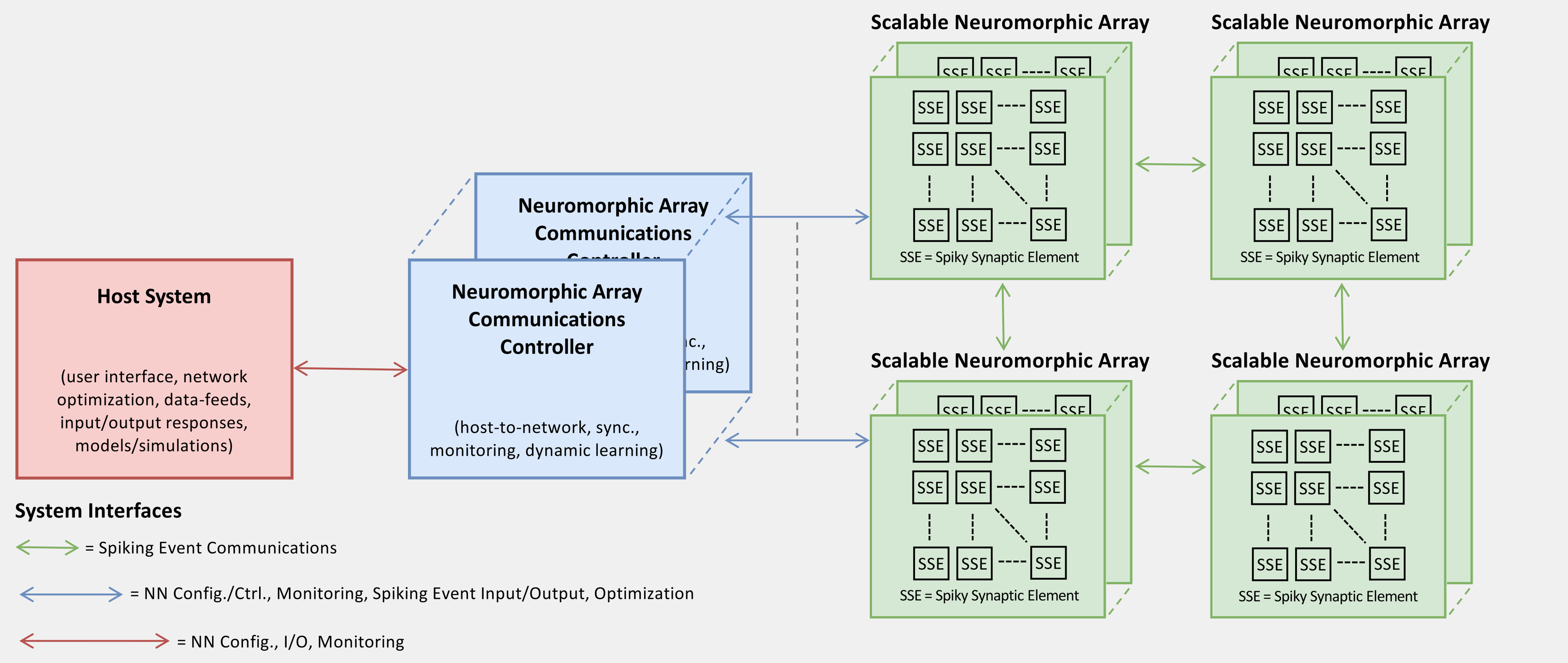
Round Trip Time Comparison

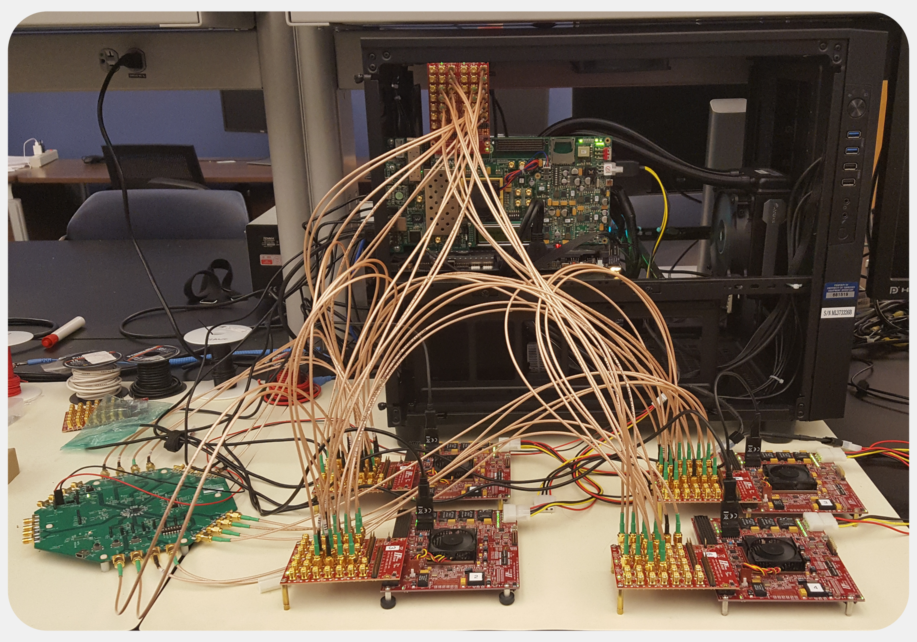


Conclusions

- Room to scale.
- Surpasses limitations of FX3.
- Maximum throughput occurs with large transfers.
- Hardware is able to evaluate arrays in constant time per cycle, whereas, the simulator grows linearly with the number of events.

Communications Board Design





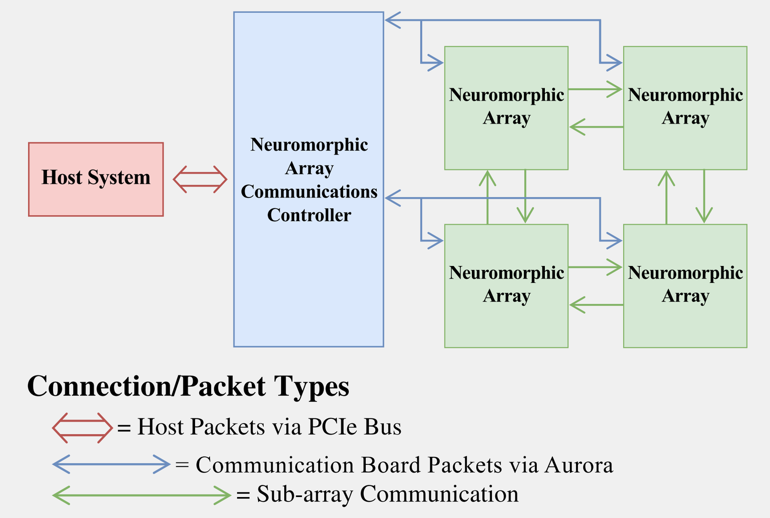
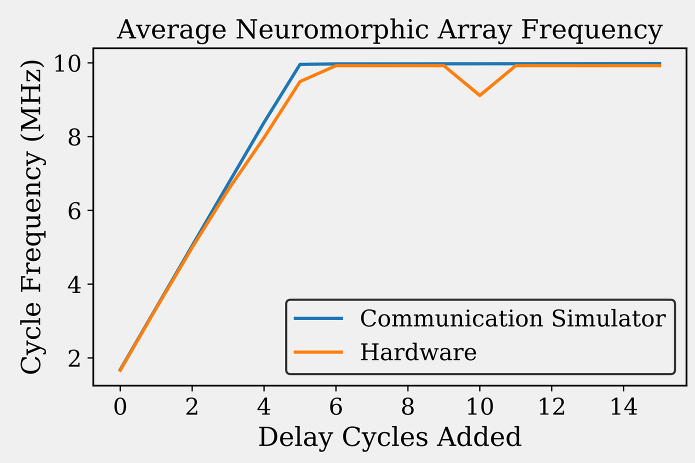
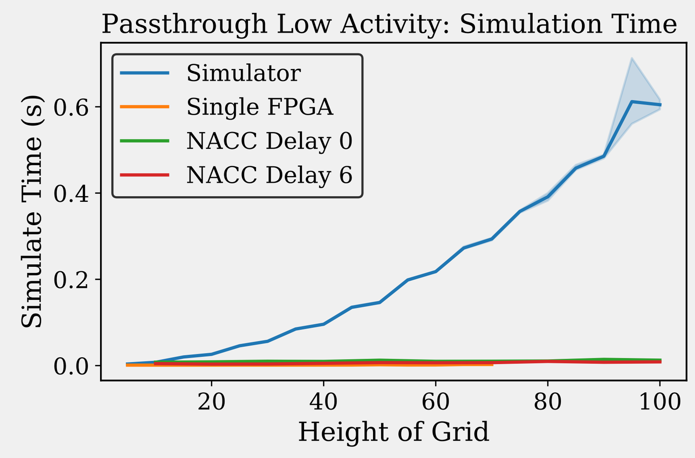


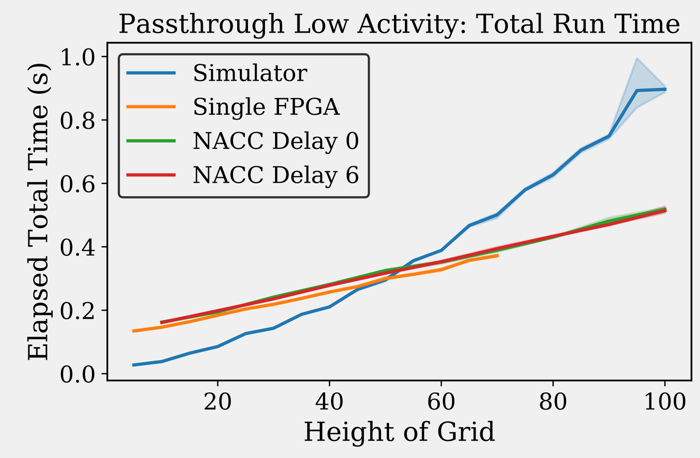
Diagram of Current NACC Setup (Pictured Left)

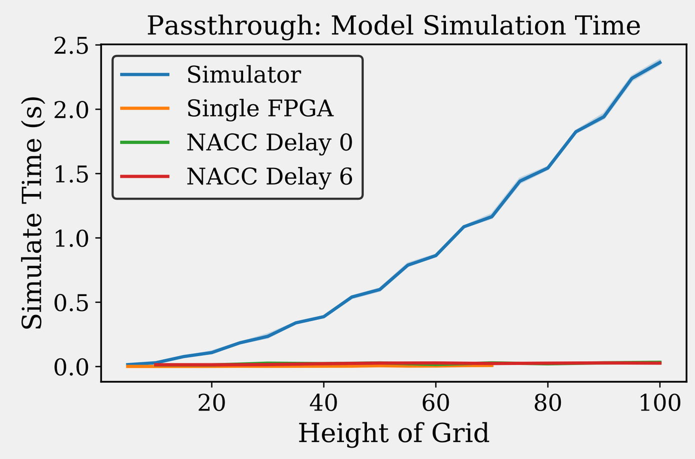
Results

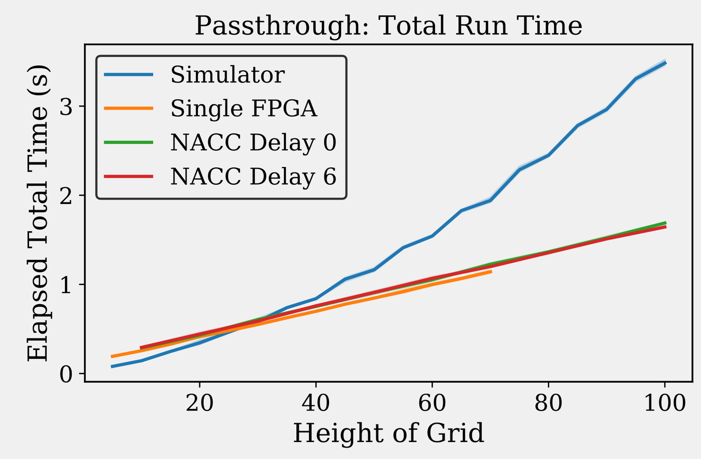


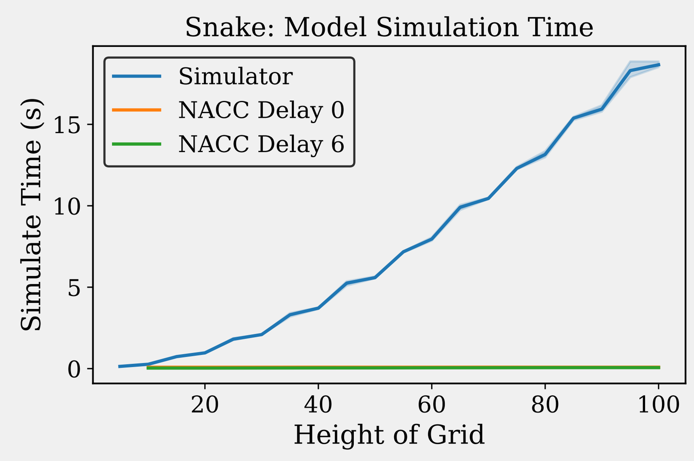
- The communication latency is slower than the neuromorphic arrays.
- Delay cycles are added to synapses crossing the board boundary to hide this latency.
- Hardware communication patterns match simulated communication patterns.
- The width of the grid is half of the height.
- The number of inputs/outputs is equal to the height.

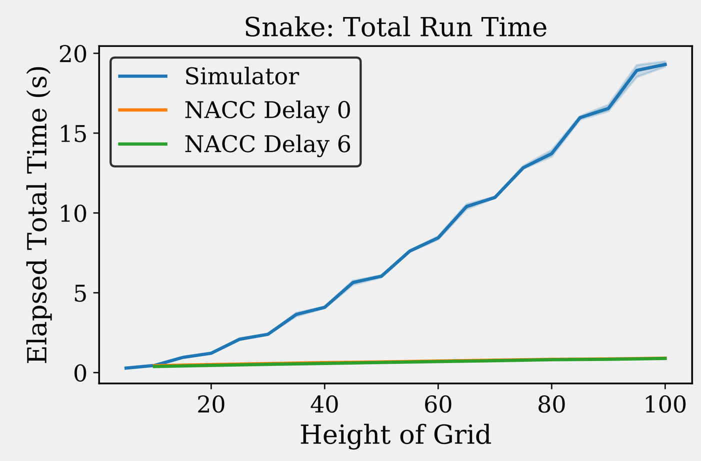












Internal Activity / Input Activity — Increases

- The hardware is faster when the network is simulated for more cycles, when the internal activity to input activity ratio is greater, and when the number of elements increases.

Acknowledgements

OAK RIDGE National Laboratory

THE UNIVERSITY OF TENNESSEE KNOXVILLE

Visit our website at neuromorphic.eecs.utk.edu

References

[1] http://www.cypress.com/products/ez-usb-fx3-superspeed-usb-30-peripheral-controller
[2] http://xillybus.com/
[3] https://www.xilinx.com/products/intellectual-property/aurora8b10b.html