

The Achronix Integrated 2D NoC Enables High- Bandwidth Designs

WP028

2022.09.01

Copyright and Trademarks

Copyright © 2022 Achronix Semiconductor Corporation. All rights reserved. Achronix, Speedster and VectorPath are registered trademarks, and Speedcore and Speedchip are trademarks of Achronix Semiconductor Corporation. All other trademarks are the property of their prospective owners. All specifications subject to change without notice. WP028 V1.1

Notice of Disclaimer

The information given in this document is believed to be accurate and reliable. However, Achronix Semiconductor Corporation does not give any representations or warranties as to the completeness or accuracy of such information and shall have no liability for the use of the information contained herein. Achronix Semiconductor Corporation reserves the right to make changes to this document and the information contained herein at any time and without notice. All Achronix trademarks, registered trademarks, disclaimers and patents are listed at <http://www.achronix.com/legal>.

Achronix[®]
Data Acceleration

Achronix Semiconductor Corporation

2903 Bunker Hill Lane
Santa Clara, CA 95054
USA

Website: www.achronix.com
E-mail : info@achronix.com

Abstract

Devices aimed at addressing modern algorithm acceleration workloads must be able to efficiently move high-bandwidth data streams between high-speed interfaces and throughout the device. Achronix Speedster®7t FPGAs can process these high-bandwidth data streams via an integrated new and highly innovative two-dimensional network on chip (2D NoC). How does the Achronix innovative 2D NoC stack up against designing a flexible 2D NoC using traditional methods in an FPGA fabric? This white paper discusses the two methods of implementing a 2D NoC and presents an example design to show how the Achronix 2D NoC improves performance, reduces area, and reduces design time when compared to a soft 2D NoC implementation.

Introduction

Achronix has completely re-architected on-chip communications for the Speedster7t family to accommodate high-bandwidth data streams by integrating an innovative 2D NoC. On the periphery of the FPGA, this 2D NoC connects to all of the high-speed interfaces: multiple 400G Ethernet, PCIe Gen5, GDDR6 and DDR4/5 ports. In the interior of the FPGA (programmable fabric) are a series of high-speed row and column conduits that distribute network traffic horizontally and vertically respectively through the FPGA programmable fabric. In addition to these rows and columns, there are initiator and target NoC access points (NAPs) at the location where each row and column of the NoC cross. These NAPs serve as either a source or destination between the NoC and the resources located in the programmable fabric.

To compare the built-in 2D NoC with one created using traditional methods in the programmable fabric, several soft NoC designs were reviewed. Finally the soft 2D NoC from Milan Polytechnic (<https://github.com/agalimberti/NoCRouter>, 2017) was chosen based on peer reviews and ease of portability to an FPGA fabric. This soft NoC implements a wormhole lookahead predictive switching in a unidirectional mesh. When implemented it requires multiple memories at each mesh node to store and forward flits (flow control units).

To quantify the differences between the on-chip 2D NoC and the soft implementation using fabric resources, first a design was created that instantiated 19 instances of an AlexNet 2D convolution, then three main metrics were compared between the completed 2D NoC designs: resources needed, performance of the design, and design time (time to create design as well as time to compile the design in the tools). In all three cases, the integrated Achronix 2D NoC significantly outperforms a soft implementation.

2D NoC Reduces Resources Used

To compare the two different 2D NoC designs, both 2D NoCs were combined with an existing 2D convolution (conv2d) design. The conv2d design performs AlexNet 2D convolution on an input image. This conv2d design requires either one or two AXI-4 connections: one to read from the memory and one to write to the memory, or a shared AXI-4 performing both read and write. In order to achieve the best integration with the soft NoC, a single shared AXI-4 interface was chosen, with an instance of the conv2d module located at each mesh node. The soft NoC is then enabling data ingress and egress from the GDDR6 memory interface — in the soft NoC the memory interface is connected to the 20th mesh node; in the built-in NoC, this connection is already present. Within the overall design, node-to-node communication exists from the GDDR6 to each of the conv2d nodes, but the conv2d nodes do not communicate between each other.

Design Details with Achronix 2D NoC

The design has 19 instances of the conv2d module, each accessing GDDR6 memory. The 20th instance is free because the GDDR6 interface is connected directly to the integrated 2D NoC. Thirty-eight of the eighty available NoC access points (NAPs) are used to connect to conv2d instances. Each conv2d instance uses 64 machine learning processors (MLPs), which covers two NAPs vertically. Because of this arrangement versus the built-in 2D NoC, a dual AXI-4 approach was taken for connecting the conv2d module. The table below lists the resources used in this design.

Table 1 – Resources Used with the Achronix 2D NoC

LUTs	BRAMs	MLP
10812	1178	1140

The use of the integrated Achronix 2D NoC produces an elegant, repeatable structure to the design placement, and consumes less than half of the resources in the device. Below is a floor plan of the resources used in the AC7t1500 device.

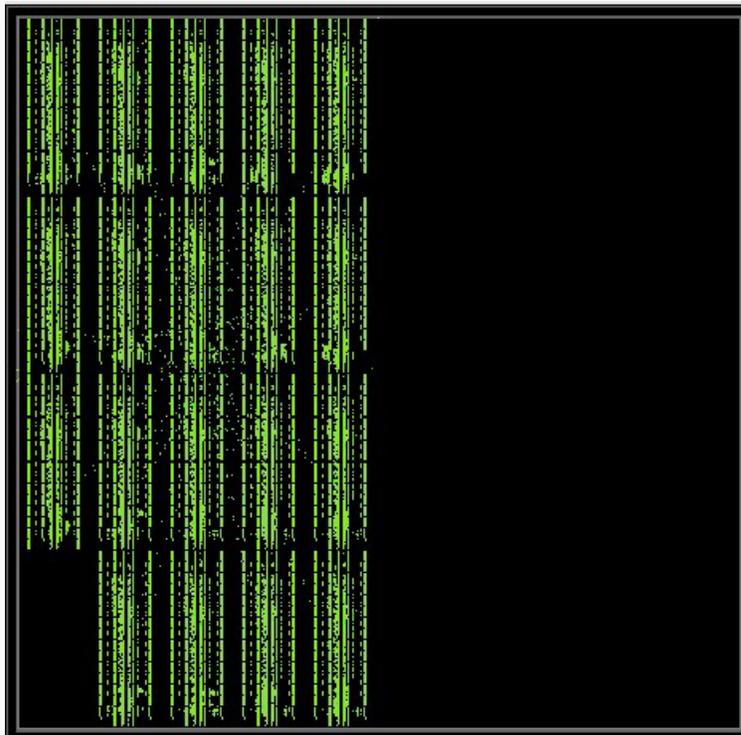


Figure 1 – Placement of Instances Using the Achronix 2D NoC in an AC7t1500

Design Details with Soft 2D NoC

The design is configured as a 5 × 4 mesh, with 19 instances of the conv2d module, each connected to a soft NoC node. The 20th mesh node is reserved for the GDDR6 interface. As a result, more logic is needed to manage the soft 2D NoC structure. This implementation also requires memory at each node in order to store and forward flits to the next node. The result is significantly higher usage of resources, as well as irregular placement across the

device. The table below lists the resources used; the figure following shows the floor plan of the resources used in the AC7t1500.

Table 2 – Resources Used with a Soft 2D NoC

LUTs	BRAMs	MLP
80365	1724	1140

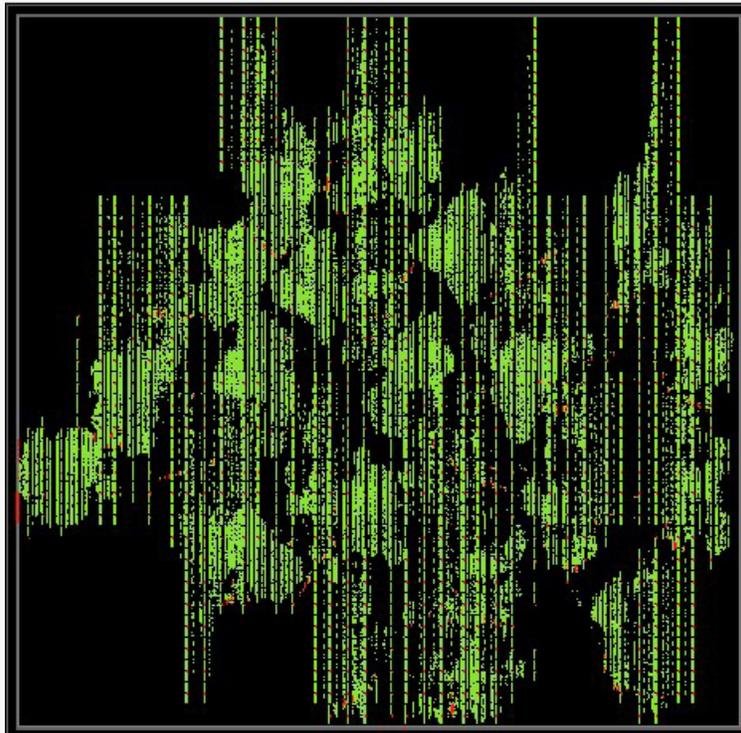


Figure 2 – Placement of Instances Using Soft 2D NoC

2D NoC Improves Performance

As stated before, by using the Achronix 2D NoC, the conv2d design produces a regular placement of resources, resulting in regular routing. The reduced logic reduces congestion as there is less logic to route. This solution achieves a maximum frequency of 565 MHz, with the critical path contained within the conv2d instance logic. As more conv2d nodes are added to the design, the maximum frequency does not reduce.

Figure 3 below shows the routing produced when using the Achronix 2D NoC.

Using the soft 2D NoC solution results in complex and irregular routing. Timing is also compromised as deep LUT logic is needed to select the appropriate paths in the soft 2D NoC.

Additionally, performance reduces as the mesh size increases. With a 2×3 mesh the design can achieve 94 MHz, while a 5×4 mesh can only achieve 82 MHz. The critical path is contained within the soft NoC mesh rather than in the conv2d logic. Timing on the soft 2D NoC could be improved further if more time were spent on optimizing the design for performance.

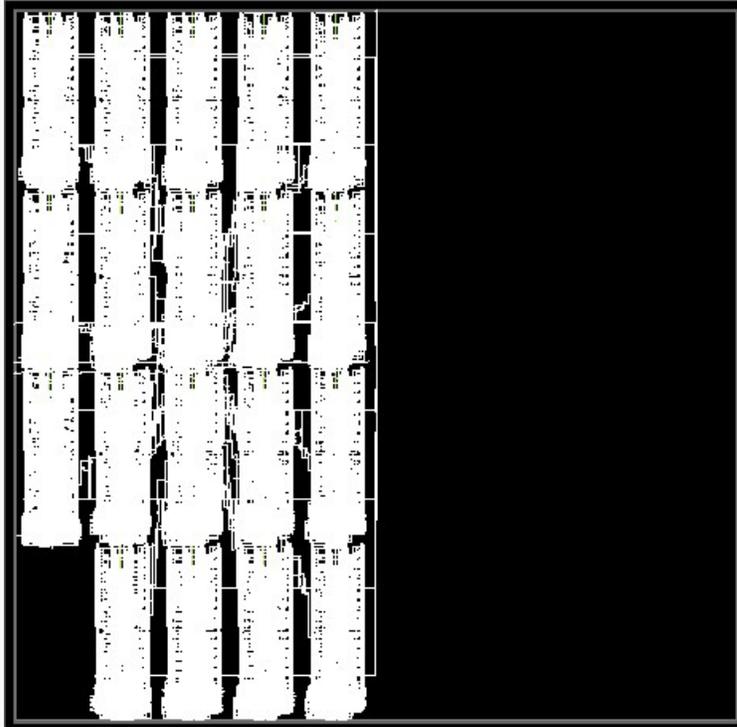


Figure 3 – Routing of the *cnv2d* Design Using the Achronix 2D NoC

Figure 4 displays the routing produced while using the soft 2D NoC design.



Figure 4 – Routing of *cnv2d* Design Using the Soft 2D NoC

2D NoC Improves Bandwidth

The Achronix 2D NoC operates at 2 GHz using a 256-bit bi-directional bus. Each conv2d instance connects to two NAPs, enabling a maximum bandwidth to/from the GDDR6 interface of 512 Gbps on one node. The block diagram below shows a close-up of the 2D NoC and one NAP connected to the local conv2d instance.

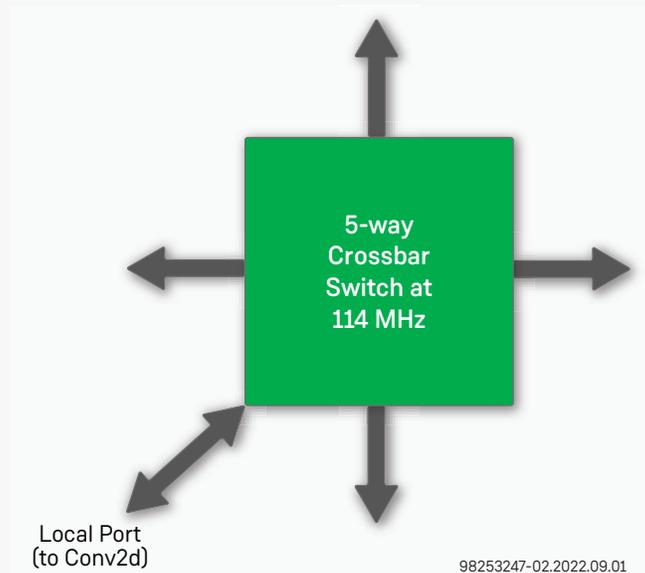


Figure 5 – Achronix 2D NoC and NAP

The soft 2D NoC uses a five-way crossbar switch. One port communicates with the local conv2d instance, while the other ports communicate with the next node in the mesh. This solution can achieve 82 MHz from node to node, producing a maximum bandwidth to/from the GDDR6 interface of 21 Gbps on one node. The block diagram below shows one crossbar switch in the soft 2D NoC mesh.

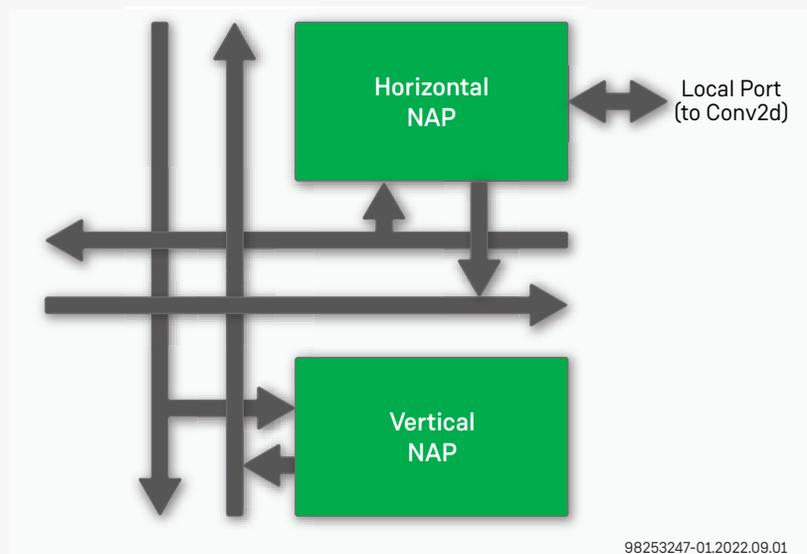


Figure 6 – Soft 2D NoC Crossbar Switch

2D NoC Reduces Design Time and Tool Run Time

The Achronix 2D NoC leverages the AXI-4 standard to communicate with the NAPs, which is an interface standard many FPGA designers are familiar with already. Additionally, the 2D NoC includes built-in features such as clock-domain crossing logic, transaction flow control, and decoding of addresses, that no longer need to be included in the user logic. The full-featured implementation of the Achronix 2D NoC eliminates a large amount of design work for the user, allowing designers to concentrate on the accelerator connecting to the 2D NoC.

Along with reduced design time, a design which utilizes the Achronix 2D NoC uses fewer resources than one which uses a soft 2D NoC. The result is less logic to place and route and results in faster run time through the tools. For example, the design using the Achronix 2D NoC takes less than half the time to place and route versus the implementation using the soft 2D NoC.

Conclusions

The integrated Speedster7t 2D NoC enables a fundamental shift in the FPGA design process. Achronix is the first FPGA company to integrate a 2D NoC which connects all of the system interfaces and the FPGA fabric. This new architecture makes Achronix FPGAs uniquely suitable for high-bandwidth applications while significantly improving designer productivity. Because the 2D NoC manages all of the networking functions between data accelerators designed into the FPGA fabric and the high-speed data interfaces, designers need only design their data accelerators and connect them to a NAP primitive. When compared to using a soft 2D NoC, designers can benefit from:

- Reduced logic utilization and increased overall FPGA performance
- Increased bandwidth
- Reduction in memory requirements
- Faster design time and less tool run time

Table 3 – Summary Comparison of Speedster7t 2D NoC versus Soft 2D NoC

Metric	Soft 2D NoC	Speedster7t 2D NoC	Improvement
Frequency	82 MHz	565 MHz	7×
Bandwidth	21 Gbps	512 Gbps	24×
LUTs	80365	10812	7×
BRAMs	1724	1178	1.5×
MLPs	1140	1140	1×
Run Time	120 minutes	50 minutes	2.4