A Low Latency Scalable 3D NoC Using BFT Topology with Table Based Uniform Routing

Avik Bose†, Prasun Ghosal‡, Saraju P. Mohanty‡
† Indian Institute of Engineering Science and Technology, Shibpur, Howrah 711103, WB, India
‡ University of North Texas, Denton, TX 76203, USA
Email: avikbose1145@gmail.com, prasun@ieee.org, saraju.mohanty@unt.edu

PROBLEM DESCRIPTION

Due to the close proximity of layers in 3D NoC, the signals travelling in the vertical (inter-layer) direction is much faster than the horizontal (intralayer) in their 2D counterpart.

An inter-layer connection requires the addition of two more links (up and down) to each router that leads to an increase in complexity as well as the blocking probability inside the routers.

Being a multi-hop communication fabric, the traditional NoC routers cannot be placed on the vertical path in a NoC as the multi-hop delay and the router delay would overshadow the ultra-fast propagation time.

Thus, it is desirable to have simple hop communication among the layers because of the short distance between them.

Also, the number of vertical pillars should be kept low to reduce the manufacturing cost of a 3D NoC.

It introduces a new problem for the IP blocks with close proximity to the pillar nodes on a layer giving a maximum advantage in case of inter-layer communication thus these that are at relatively distant position from the pillar nodes.

Objective

To propose a novel 3D NoC topology with proper routing method that can address all the above issues.

DIFFERENT ROUTING SCOPES

Layer Scope: A 3D NoC chip consists of several layers. Every layer is made up of four BFTs (denoted by L). In every BFT there are four regions. Each region is made up of two IP blocks connected to it.

Tree Scope: Each region is made up of four localities, where each locality comprises of one local router (denoted by R) and four IP blocks connected to it.

Regional Scope: The circular DTDMA pillar node is shown in the center of the floor plan picture of a single BFT. It is connected to a special bus architectures requires only one additional link (in the place of two) on NoC router.

Regional Scope: The circular DTDMA pillar node is shown in the center of the floor plan picture of a single BFT. It is connected to a special bus architectures requires only one additional link (in the place of two) on NoC router.

Local Scope: The circular DTDMA pillar node is shown in the center of the floor plan picture of a single BFT. It is connected to a special bus architectures requires only one additional link (in the place of two) on NoC router.

ROUTING ALGORITHMS FOR DIFFERENT SCOPES

Connection Structure in a typical 3D Mesh topology.

PROPOSED SOLUTION

To propose a novel 3D NoC topology with proper routing method that can address all the above issues.

DISTRIBUTION ALGORITHM FOR DIFFERENT SCOPES

DISTRIBUTION ALGORITHM FOR DIFFERENT SCOPES

DISTRIBUTION ALGORITHM FOR DIFFERENT SCOPES

A Tree Routing Table (Node Number, Link Number) is used to route the packets in the network. The table is updated periodically and the routing decision is based on the Table.

Routing Table

Layer Number

Tree Number

Region Number

Locality Number

Node Number

Routing Algorithm

Layer Scope

Tree Scope

Regional Scope

Local Scope

EXPERIMENTAL RESULTS

Simulation Results for Torus and BFT

Simulation Results for Flattened Butterfly and BFT

Simulation Results for Butterfly and BFT

Simulation Results for Mesh and BFT

SUMMARY OF COMPARATIVE IMPROVEMENTS (%) FOR DIFFERENT PERFORMANCE METRICS

CONCLUSION AND FUTURE WORK

Proposed BFT topology can withstand heavy workload while still maintaining low latency, and the acceptance rate also increases with increasing injection rate. This is because of the uniform and load balancing connectivity of BFT where we have more than one path between a pair of source and destination but with same hop count. On the other hand, all other topological designs have failed to balance the load and sometimes crash.

If routers can be designed in such a way that they can have the capability to balance load and control congestion efficiently, then with this design we can achieve a highly efficient NoC system for interactive applications with throughput capability.

Future work may be in the direction of investigating the thermal effects and optimizing it accordingly with a suitable core placement strategy, investigating and improving performance using real time application mapping and so on.

REFERENCES


