EE382C Project Proposal

P. Nicolas Kokkalis, Njuguna Njoroge, Ernesto Staroswiecki, Stergios Stergiou

1 SUMMARY
The present report outlines preliminary design space exploration performed along with a tentative project schedule.

2 TOPOLOGY EXPLORATION

Even though the allowed circuit real-estate is 20mm x 20mm, as a first approximation to the design, routers are restricted to be no larger than the size of a tile, 1mm x 1mm. This imposes a constraint of ~1000 wires per dimension on the router and subsequently on the physical channels per dimension x channel width product.

The traffic pattern on the NoC is bimodal, with Write Request / Read Reply packets of ~600 bits and Read Request / Write Reply packets of ~20 – ~60 bits. Since there is no speedup on the NoC to motivate the adoption of smaller length flits, it would be preferable to utilize flits of ~600 bits size. However, transmitting smaller size requests would result in decreased resource utilization. Therefore the interleaving of two NoCs handling different length flits with identical topology is decided upon.

Since such a design decision would restrict the number of wide unidirectional channels per dimension to one, it is very likely that round robin arbitration on a flit basis on the single wide channel is to be performed. Since there exists residual space for more than one thin channel, more than one thin bidirectional channel per dimension is considered.

2.1 Point—to—Point Topology

In order to be able to evaluate the performance benefits of considered topologies, a point—to—point topology will be developed and integrated into booksim. Benchmarking on this topology will provide a lower bound on the performance of any real—world topology and will help streamline research / development efforts. For example, if the performance of the already present topologies is close to the ideal point—to—point, efforts will be directed to area optimization. Realistic design restrictions will be integrated, such as actual link distance to be traversed per tile tuple for an optimum tile layout.
2.2 Considered Topologies

A number of topology related publications has been collected from the vast depository on the subject. Their trade-offs regarding required chip real-estate are being analyzed. Since these topologies are variations on cube and fat trees, it is expected that analysis will be possible with manageable effort under boostim. Full exploration on them will be performed after the traffic pattern from the specific application to be mapped onto the NoC is provided.

2.2.1 Honeycomb and Diamond Networks

Honeycomb and diamond networks have been proposed as alternatives to mesh and torus architectures for parallel processing (Fig. 1). When wraparound links are included in honeycomb and diamond networks, the resulting structures can be viewed as having been derived via a systematic pruning scheme applied to the links of 2D and 3D tori, respectively. The removal of links, which is performed along a diagonal pruning direction, preserves the network's node-symmetry and diameter, while reducing its implementation complexity and VLSI layout area. It has been shown [1] that honeycomb and diamond networks are special subgraphs of complete 2D and 3D tori, respectively, with important implications for their physical layouts and routing schemes (Fig. 2). Because pruning reduces the node degree without increasing the network diameter, the pruned networks have an advantage when the degree-diameter product is used as a figure of merit. Additionally, if the reduced node degree is used as an opportunity to increase the link bandwidths to equalize the costs of pruned and unpruned networks, a gain in communication performance may result.

2.2.2 Folded Fat H-Tree

Folded Fat H-Tree [2] is a novel on-chip network topology that includes both a fat tree and a torus structure (Fig. 3.) For on-chip implementation, folding layout is performed on wraparound links as on torus. Both topologies, classic and Fat H-Trees, are being considered, with classic H-Tree being likely more viable for the specific design. A layout methodology for fat trees is presented in [3].

3 Routing

Since routing policy is highly topology—dependent, no conclusion has been drawn as to whether deterministic or oblivious is preferable. This decision will be made after initial experimental results on implemented topologies are collected. Adaptive routing does not seem to be a viable approach for NoCs mostly due to the excess buffering requirements.
per hop that are required. Since deadlock issues need additionally be addressed, it is likely that deterministic routing will

4 Flow Control

The flow control policy is again related to the topology selection and the specifications of the router. However, since total area dedicated to buffering appears to be a dominating factor on the design of the specified NoC, effort is going to be devoted to minimizing it, provided that this translates to actual minimization of the total chip area. In this respect, the adoption of Credit-based flow control along with distributing buffers onto the channels [9][10] is very likely.
5 Router Microarchitecture

Redesigning the router requires a considerable effort. While this is naturally of a lower priority in this project, we would like to explore the impact on performance of integrating lighter, single clock cycle routers onto booksim. A reasonable way to do this within the time budget of this project would be to explore the tradeoff between virtual channel count and redundant switch inputs.

REFERENCES


