Comparison of Physical and Virtual Express Topologies for Future Many-core On-Chip Networks

Chia-Hsin Owen Chen\(^1\), Niket Agarwal\(^1\), Tushar Krishna\(^1\), Kyung-Hoae Koo\(^2\), Li-Shiuan Peh\(^3\), and Krishna C. Saraswat\(^2\)

\(^1\)Dept. of Electrical Engineering, Princeton University, Princeton, NJ 08544
\(^2\)Electrical Engineering Department, Center of Integrated Systems, Stanford University, Stanford, CA 94305
\(^3\)\{chiac, niketa, tkrishna, peh\}@princeton.edu, \(^2\}\{koo1028, saraswat\}@stanford.edu,

Abstract

With the advent of many-core chips, the number of cores present on-chip are increasing rapidly. In these chips, the on-chip network that connects these cores needs to scale efficiently. The topology of on-chip networks is one of the important design choices that affects how these networks scale. Most current on-chip networks use 2-D mesh topologies to connect various on-chip nodes. 2-D meshes do not scale to large node counts due to their large diameter and energy inefficiency. To tackle the scalability problem of 2-D meshes, various express topologies have been proposed. These proposals fall within two categories: physical express topologies and virtual express topologies. The physical express topologies employ long-hop wires to enable bypassing of intermediate hops for non-local communication. This leads to lower latency and energy consumption. There are even more savings if the long hops use recently proposed link designs like capacitively driven low-swing interconnects (CDLSI). However, the addition of long-hop channels leads to additional router ports, larger crossbars and extra physical channels. The virtual express topologies employ opportunistic flow control techniques to enable bypassing of router pipelines, except the crossbar traversal stage. This leads to savings of router pipeline cycles as well as lower energy consumption. The virtual bypasses are not as aggressive as physical bypasses since router crossbars and local links still have to be traversed. So, there is a clear trade-off between the two bypassing techniques. However, there has not been an extensive comparative study of these two approaches in the past.

In this work, we present a detailed characterization of how CDLSI links can be used to model long-hop links in physical express topologies. We also compare physical express topologies to that of virtual express topologies using both synthetic network traffic as well as full-system evaluations. We show that CDLSI-based physical express topologies are attractive for designs that operate at relatively lower network traffic, in which case such topologies exhibit high performance at lower power. For networks that operate at higher injection rates, virtual express topologies are a better fit to the high throughput requirements of such systems.

1 Introduction

With Moore’s law providing more and more transistors on chip, architects have embraced many-core architectures to deal with increasing design complexity and power consumption of conventional single-processor chips. There are already fabricated designs \([13, 26]\) with 12s of cores on a single chip. Going forward, we could have 100s or even 1000s of cores on a single die. These many-core architectures employ many simpler cores and interconnect the cores using a scalable on chip network fabric. To meet the bandwidth demands of these cores, designers use packet-switched on chip interconnection networks and have moved beyond conventional shared buses.

Selecting an appropriate topology is one of the most critical decisions in the design of on chip networks; it impacts the zero load latency and sustainable bandwidth and also influences the power consumption of the network. Most existing on chip networks \([10, 13, 26]\) utilize a 2-D mesh topology, as meshes have lower design complexity and map well to the 2-D chip substrate. However, they are energy inefficient because the high network diameter leads to extra router hops to reach the destination, and router energy is high relative to link energy \([13]\). This poses serious scalability concerns for such topologies as node count increases in the future. To tackle the scalability issues of 2-D mesh, various express topologies has been proposed in the literature. One set of proposals \([8, 11, 17]\) employ long physical links between non-local routers to reduce the the ef-
fective network diameter. We will refer to these topologies as physical express topologies. Another set of proposals [20, 21, 23] employ various techniques to opportunistically bypass router pipelines at intermediate network hops and thus save on network latency and power. We will refer to these techniques as virtual express topologies.

Both physical and virtual express topologies have their advantages and disadvantages. Physical express topologies reduce the network diameter, thereby saving the latency and power due to bypassing of intermediate router hops. To accomplish the physical bypass paths, however, extra router ports, larger crossbars and extra physical channels are required. Extra router ports and larger crossbars lead to higher router area and power. While on chip wires are relatively abundant, use of large number of dedicated point-to-point links lead to a large area footprint and low channel utilization. Virtual express topologies have the advantage that they do not use extra long physical links. However, the bypass of routers in such designs is opportunistic and not guaranteed. The virtual router bypass also differs from physical bypass in that the flits still have to go through the router crossbar in order to reach the output link, while in a physical bypass, all intermediate routers are bypassed completely.

The long links in the physical express topologies can also leverage some of the recent innovative low-swing interconnect proposals. Specifically, capacitively driven low-swing interconnects (CDLSI) [12] have the potential to save significant energy while also providing modest latency reductions. CDLSI links can be used as single-cycle multi-hop express links and that provides an additional latency saving since long hops can be covered in a single cycle and the power consumption would also be less, owing to the low-swing nature of these links. Although there are clear trade-offs between various physical and virtual express topologies, to the best of our knowledge, there has been no work comparing these various topologies alternatives for many-core network on chips (NoCs). In this paper, we compare a particular physical express topology (express cubes [8]) to an express virtual topology proposal (express virtual channels (EVCs) [21]) for large node count on chip networks. We present energy, area, latency and throughput results for 256 node configurations with synthetic network traffic. We also do full-system evaluations for 64 node systems with SPLASH-2 benchmarks to see the network impact of various network topology proposals on real traffic.

The main contributions of this work are as follows.

- We present a detailed characterization of how CDLSI links could be leveraged for long hop links to be used in physical express topologies.
- We model the energy, area, delay and throughput of physical and virtual express topologies.

CDLSI was proposed by Ho. et al. in [12]. It has been shown to have excellent energy savings without the degradation of performance. CDLSI achieves these properties by driving the wires capacitively. The coupling capacitor not only eliminates the need for a second power supply, but also extends the wire bandwidth by introducing pre-emphasizing effects and significantly reducing the driver energy consumption. Figure 1 shows a simplified repeated CDLSI wire model. The differential and twisted wires enable the signals to be sent at a low voltage swing, and eliminate the coupling noise. But this comes at the cost of 2X wiring area.

We applied a delay-power optimization, which is similar to the approach in [15], to both conventional wires and CDLSI wires. Detailed analysis can be found in [18]. For conventional wires, more than half of the total power is consumed by the wire. On the other hand, for CDLSI wires, transceivers consume most of the power. As a result, for a given delay penalty, the impact on power saving in CDLSI wires is greater than in conventional wires. Table 1 shows the latency and energy consumption of a CDLSI wire as a function of different wire lengths and bandwidth with 192
nm wire width at 32 nm technology node. The wire width is chosen to be four times the minimum wire width defined in ITRS [1] to meet the bandwidth requirements in our experiments. The table also shows the numbers for conventional copper wires. The energy consumption of a CDLSI wire is approximately 10x less than that of a conventional wire with the same wire length and bandwidth below 3 Gbps. It should be noted that CDLSI wires remain ultra-low energy consuming even as the length of wire increases. This property enables the use of CDLSI wires as express links for multi-hop communication. For example, the delay for a 4 mm long CDLSI wire is 368 ps, that is, it is within two cycles for a 4 GHz clock frequency.

### Table 1. Delay(D) and energy(E) consumption of CDLSI wire and conventional wire for different lengths and bandwidths, where wire width is 192 nm

<table>
<thead>
<tr>
<th>wire length</th>
<th>1mm</th>
<th>2mm</th>
<th>3mm</th>
<th>4mm</th>
</tr>
</thead>
<tbody>
<tr>
<td>CDLSI 1Gbps</td>
<td>108</td>
<td>22.4</td>
<td>215</td>
<td>27</td>
</tr>
<tr>
<td>CDLSI 2Gbps</td>
<td>108</td>
<td>22.4</td>
<td>215</td>
<td>27</td>
</tr>
<tr>
<td>CDLSI 3Gbps</td>
<td>108</td>
<td>22.4</td>
<td>194</td>
<td>30.4</td>
</tr>
<tr>
<td>CDLSI 4Gbps</td>
<td>101</td>
<td>27.3</td>
<td>175</td>
<td>35.8</td>
</tr>
<tr>
<td>Conventional 1Gbps – 10Gbps</td>
<td>108</td>
<td>133</td>
<td>215</td>
<td>265</td>
</tr>
</tbody>
</table>

3 Express Topologies

An ideal interconnection network between processing cores in CMPs would be point-to-point, where the latency of the messages is equal to just the link latency, and there are no additional penalties due to contention and corresponding arbitration. However, a point-to-point network is not scalable, requiring \( N \) – 1 links per node in a \( N \) node network, which will blow up the area footprint. This has led to researchers proposing packet-switched networks with intermediate routers/switches to multiplex the available bandwidth. A state-of-the-art network router [9] (as shown in Figure 2(a)) typically has a 5-stage router pipeline, with four stages in the router, and one in the link, as shown in Figure 2(b). The stages are:

- **BW (Buffer Write):** The incoming flit gets buffered. In parallel, it performs the Route Computation (RC) for determining the output port to go out from.
- **VA (VC Allocation):** The flits at all input ports which wish to depart from a certain output port, arbitrate for virtual channels (VCs) at this output port.
- **SA (Switch Allocation):** The flits at all input ports, which have been assigned output VCs, now arbitrate for using the crossbar that will connect this input port to the required output port.
- **ST (Switch Traversal):** The flits that win the switch traverse the crossbar.
- **LT (Link Traversal):** The flits coming out of the switch, traverse the link till the next router.

A packet can go through VA and SA multiple times depending on the contention, until it succeeds in obtaining a VC and switch. Thus each flit entering a router takes multiple cycles before it can move on to the next hop.

This solution of having hop-by-hop traversal through multiple intermediate routers is simple and effective. However, spending multiple arbitration cycles in intermediate routers adds unnecessary delay to packets traveling longer distances. Moreover, the buffering and arbitrations at the routers add to the power consumed. Thus, there have been many proposals that advocate for packets traveling longer distances to **bypass** intermediate routers. This bypassing may be enabled by either

1. adding physical channels that skip intermediate routers, or
2. by having means to bypass arbitration and buffering at the intermediate routers.

Express cubes [8], Flattened Butterfly [17], MECS [11] and [25] follow approach 1, while Express Virtual Channels [21], Token Flow Control [20], Prediction Router [23] are techniques that fall under approach 2. We describe Express Cubes and Express Virtual channels here briefly.

3.1 Express Cubes

Express cubes [8] was a proposal for the multi-chassis shared-memory multi-processor (SMP) domain, that argued for physical express lanes between far away nodes. This would allow non-local messages to traverse these express links and avoid getting delayed at intermediate nodes, and
3.2 Express Virtual Channels

Express virtual channels (EVCs) [21] is a flow-control and router microarchitecture design to reduce the router energy/delay overhead by providing virtual express lanes in the network which can be used by packets to bypass the pipelines of intermediate routers along a dimension. The flits traveling on these lanes move straight to switch traversal (ST) as soon as they are received at a router, thereby reducing the latency, and dynamic power (as buffer reads/writes and VC/Switch arbitrations are skipped). This technique also leads to a reduction in the total number of buffers required in the network to support the same network throughput, which in turn reduces the leakage power.

In traditional routers, during the VA stage, the buffered packets arbitrate for VCs at the next router, where they would get buffered again. The basic idea behind express virtual channels is that packets arbitrate for VCs at routers that are \( k \) hops away, and only get buffered at that router, bypassing all the intermediate routers. Each router can act as either a source/sink node, which allows flits to get buffered and move through the normal router pipeline, or as a bypass node which gives priority to flits on EVCs to pass straight through the switch, without having to be buffered. The EVC bypass pipeline thus consists of only the ST and LT stages. This bypassing is enabled by sending a lookahead flit one cycle in advance which sets up the switch at the next router so that the incoming flit in the next cycle can traverse the crossbar straight away. Thus the VCs at all router ports are partitioned into NVCs (Normal Virtual Channels), which carry flits one hop; and \( k \)-hop EVCs (\( k \) can take all values between 2 and some \( l_{\text{max}} \)), which can carry flits \( k \) hops at a time, bypassing the pipelines of the \( k - 1 \) routers in between, in that dimension (XY routing is assumed). In the source/sink routers, the head flits arbitrate for the \( k \)-hop EVCs, or NVCs depending on their route (i.e. the number of hops remaining in that dimension), and the availability of each type of VCs. The body and tail flits follow on the same VC and release it when the tail flit leaves. Once a \( k \)-hop EVC is obtained, the intermediate \( k - 1 \) nodes can be bypassed. All packets try to use a combination of EVCs to best match their route, such that they are able to bypass most nodes in a dimension.
Figure 4. Comparison of path taken by Baseline, EVC and Express Cube for a flit going from router 0 to router 5

Token Flow Control [20] also enables bypassing, but it supports adaptive routing and allows for bypassing of entire routes from source to destination. Prediction Router [23] predicts the output port and sets up the switch so that the incoming flit can bypass the arbitrations.

The basic difference between the baseline, EVCs and Express Cubes is highlighted in Figure 4. A flit going from router 0 to router 5 would go hop by hop in the baseline design, through the 5-stage router pipeline at every intermediate router. In EVCs, assuming the maximum EVC length is three, the flit would use a 3-hop EVC to go from router 0 to router 3, and a 2-hop EVC to go from router 3 to router 5. It will go through the 2-stage bypass pipeline at the intermediate routers. In the express cubes design, assuming 4-hop express links, the flit would go from router 0 to router 4 on the express link, completely bypassing the intermediate routers, and then proceed to router 5.

4 Evaluation

In this section, we present a detailed evaluation of EVCs and express cubes. We compare both topologies with a baseline 2-D mesh topology. Next we will present the detailed evaluation methodology followed by results.

4.1 Simulation Infrastructure and Configuration

4.1.1 Simulation Infrastructure

To compare the different configurations, we used a cycle-level network-only simulator, GARNET [3]. GARNET models a detailed state-of-the-art 5-stage router pipeline as described in Section 3. We used ORION 2.0 [14], an architecture level network energy and area model, to evaluate the power consumption and area footprint of routers and links. More details about the energy and area modeling in ORION 2.0 are provided later in Section 4.2. All our experiments were done at 32 nm technology node. Evaluation with synthetic traffic uses packets which uniformly consist of either eight byte control packets or 72 byte data packets. We choose three message classes out of which two always carry control packets and one is dedicated to data packets. This was done to closely match modern cache coherence protocol requirements. Table 2 lists various network parameters that are common to all the configurations we evaluated for synthetic traffic. We next describe the various topology configurations we evaluated.

<table>
<thead>
<tr>
<th>Technology</th>
<th>32 nm</th>
</tr>
</thead>
<tbody>
<tr>
<td>$V_{dd}$</td>
<td>1.0 V</td>
</tr>
<tr>
<td>$V_{threshold}$</td>
<td>0.7 V</td>
</tr>
<tr>
<td>Frequency</td>
<td>4 GHz</td>
</tr>
<tr>
<td>Topology</td>
<td>16-ary 2-mesh (The physical express topologies have additional express links)</td>
</tr>
<tr>
<td>Routing</td>
<td>Dimension ordered X-Y routing</td>
</tr>
<tr>
<td>Synthetic Traffic</td>
<td>Uniform Random, Bit Complement, Tornado</td>
</tr>
</tbody>
</table>

Table 2. Simulation parameters

4.1.2 Configuration

<table>
<thead>
<tr>
<th>Baseline</th>
<th>72-byte links, 8 VCs/message class, 24 72-byte buffers per port</th>
</tr>
</thead>
<tbody>
<tr>
<td>Baseline-24B (per physical network)</td>
<td>24-byte links, 8 VCs/message class, 24 24-byte buffers per port</td>
</tr>
<tr>
<td>EVC-network</td>
<td>72-byte links, 8 VCs/message class, 24 72-byte buffers per port</td>
</tr>
<tr>
<td>EVC-network-24B (per physical network)</td>
<td>24-byte links, 8 VCs/message class, 24 24-byte buffers per port</td>
</tr>
<tr>
<td>Express-normal</td>
<td>24-byte links, 8 VCs/message class, 48 24-byte buffers per port</td>
</tr>
<tr>
<td>Express-CDLSI</td>
<td>12-byte links, 8 VCs/message class 96 12-byte buffers per port</td>
</tr>
</tbody>
</table>

Table 3. Topology parameters.
for the baseline and EVC networks was a 16x16 mesh, and is shown in Figure 5(a). For the express cubes topology, we used 4-hop express links, as shown in Figure 5(b). Express links in the x and y dimensions start at every alternate router. This topology design was chosen to allow many symmetrical express links in the 16x16 network, without increasing router ports significantly, since the power consumed by the crossbar increases quadratically with the increase in number of ports. This topology has 7 ports per router, as compared to the 5-port routers in the baseline. We kept the bisection metal area budget and input buffer area budget constant across all the configurations to do a fair comparison. We assumed 12-byte links in these CDLSI express paths, which means that the control packets comprise of a single flit, while the data packets are composed of six flits. We will refer to this topology as Express-CDLSI. We also evaluated the same express topology using conventional Cu express links between the 4-hop routers. As the CDLSI links take twice the metal area as compared to conventional links (due to the differential nature of the wires), the Cu express links were assumed to be 24 bytes wide, which implies three flits in each data packet. We will refer to this topology as Express-normal. We assumed 1mm links between routers, and 4mm express links. It takes 2 cycles to traverse these express paths in both Express-CDLSI and Express-normal, as discussed in Section 2. The physical express topologies have three times more channels across the bisection cut as the baseline, and again equalizing the metal area gave the baseline 72-byte wide links. These fat links meant that the baseline, and the EVC design, have one-flit data packets. We will refer to these as Baseline and EVC-network. The EVC-network has up to 3-hop EVCs. Having higher hop EVCs gives diminishing returns due to starvation of local flits at routers. Moreover, longer EVCs require more VCs and buffers as explained in [21] which we wanted to keep fixed across the different configurations as explained later. Though the 72-byte links in the Baseline and EVC-network designs equalize the area of the links with those in the express networks, they are not a fair equalization in terms of performance (or bandwidth). This is because the 1-flit packets beat the purpose of packet-switched networks (where multiple flits from different packets interleave for a better link utilization and to push throughput). Intuitively this can be understood by looking at the way the bisection bandwidth is being divided among different kinds of links in one row in the Baseline/EVC-network and the Express-CDLSI/normal networks. In the former, there is only one fat link that can carry the one flit packet at a time, while the latter has three sets of links which would be carrying three flits from three packets at a time. Hence a better configuration for dividing the 72-byte links would be to have multiple physical networks with thinner links. We thus evaluated two other network configurations for baseline and EVC, where we use thinner 24-byte links, similar to the express networks, and have three such physical networks to mimic a 72-byte channel network. The throughput, power and area of a single physical network are thus multiplied by three. We call these configurations Baseline-24B and EVC-network-24B. Note that the results we report for these configurations are already multiplied by three.

We modeled private buffers per VC in our design, except for the EVC-network (which requires shared buffers). We chose the number of buffers such that the total SRAM buffer area at routers taken up by each configuration is almost the same. For example, the Baseline network has 5 ports at every router which leads to 5 ports × 3 message class per port × 8 VCs per message class × 1 buffer per VC × 72 B per buffer = 8640 B of buffers. The EVC-network was

Figure 5. Network topology

(a) 16x16 Baseline and EVC-network. All links are 72 bytes.

(b) 16x16 Express Cubes network. All links are 24 bytes for Express-normal topology and 12 bytes for Express-CDLSI topology. Note that the normal 1-hop links are not shown.
also given the same number of total buffers per port. The Express-normal configuration has 7 ports per router which amounts to 7 ports × 3 message class per port × 8 VCs per message class × 2 buffers per VC × 24 B per buffer = 8064 B of buffer space. The Express-CDLSI also has 7 ports per router amounting to 7 ports × 3 message class per port × 8 VCs per message class × 4 buffers per VC × 12 B per buffer = 8064 B. The Baseline-24B and EVC-network-24B on the other hand, were modeled similar to the Express-CDLSI network, with 3 message classes per port, 8 VCs per message class and 1 buffer per VC to study the potential throughput achieved by these two networks. Table 3 presents various router parameters for all the configurations.

4.2 Area and Power Modeling

We used ORION 2.0 for modeling the area and power of the on-chip network. ORION 2.0 is an architecture level network energy and area model. For the 65nm technology node, parameters are obtained directly from TSMC process libraries, while parameters for 32nm are scaled using MASTAR [2] and ITRS [1]. Its models for major router building blocks include FIFO buffers, crossbar switches, arbiters, and links between routers. It also models the power due to clocking of router blocks. Dynamic power consumption in CMOS circuits is formulated as $P = \frac{1}{2}\alpha CV_{dd}^2 f_{clk}$, where $f_{clk}$ is the clock frequency, $\alpha$ is the switching activity, $C$ is the switching capacitance, and $V_{dd}$ is the supply voltage. ORION 2.0 estimates switching capacitance of register-based FIFO buffers, clocking due to routers, and physical links with detailed parameterized equations. The leakage analysis in ORION 2.0 follows the same methodology proposed in [7] with addition of gate leakage.

In our experiments, we use the power and area model in ORION 2.0 only for the routers. For links, we use the numbers in Table 1 for power consumption. Link area is formulated as $A_{normal} = (2w_{flit} - 1)lw$ and $A_{cdlsi} = (4w_{flit} - 1)lw$, where $w_{flit}$ is the flit width, $l$ is the link length, and $w$ is both the width of the link and the spacing between two bit-lines. To meet our bandwidth requirement, we chose $w$ to be four times the minimum width defined in ITRS for 32nm technology, i.e. $w$ is equal to 192nm in our experiments. The spacing between the links is assumed to be same as the width in order to maximize the bisectional bandwidth and also minimize the area without violating the design rule.

4.3 Results

Figure 6 shows the relative performance of various topologies for synthetic traffic.
4.3.1 Performance

We observe that the Baseline has the highest latency at low loads. The EVC-network has the second highest latency. The Express-normal and Express-CDLSI networks have consistently the lowest latency at low injection rates. These results are as expected. The high hop count of the non-physical express networks lead to extra router and link hops resulting in higher low-load latencies. The EVC-network bypasses some of the intermediate router pipelines and has a lower average network latency than the Baseline at low loads. The express networks utilize express long-hop links to “completely” bypass intermediate routers and links and thus have the lowest average network latency at low injection rates. The latency savings of the express network is the highest in bit complement traffic, followed by uniform random traffic and the least in tornado traffic. This is because in the bit complement traffic, the source and destination are farthest apart and the express links are best utilized in such a traffic pattern. Tornado traffic has packets with the source and destination in the same dimension. This leaves relatively fewer opportunities for exploiting express links. Uniform random traffic is in between because the sources choose the destinations in a random fashion.

The throughput trends differ from the latency trends. The Baseline-24B and the EVC-network-24B have way better saturation throughput than all other topologies for all traffic patterns. This is because multiple physical networks best utilize the available link bandwidth and push the saturation throughput curve towards ideal. The EVC-network-24B outperforms the Baseline-24B due to router bypasses. The express networks have a higher saturation throughput than the Baseline and EVC-network for uniform random and tornado traffic. This is because the bypassing of intermediate routers in the express networks leads to lower contention and hence pushes the saturation throughput. However, for the bit-complement traffic, the express topologies saturate before the Baseline and EVC-network. As discussed earlier, the bit-complement traffic has source and destinations which are far apart and the packets try to use the express links to bypass routers. The routing function in the express network that we implemented was such that if a packet has to travel four hops away from a router in the same dimension and it has the choice of an express link, it will “always” use the express links. If a lot of packets contend for these express links, it leads to lower channel utilization in the 1-hop links. Thus, for the bit complement traffic, the express networks have a lower saturation throughput. The EVC-network improves on the throughput of the Baseline by bypassing router pipelines which results in lower contention at routers and better utilization of network links. However, the EVC-network is not able to match the saturation throughput of Baseline-24B and EVC-network-24B.

The latency and throughput trends indicate that physical express topologies are desirable for systems that want lowest network latency for low injection rates. On the other hand, if higher network throughput is desired then virtual express topologies are preferable. Although the performance numbers show latency-throughput characteristics of various topologies, they only tell half the story. Network area and power are also important design constraints that future on-chip networks would have to adhere to.

4.3.2 Area

Figure 7 shows the area footprint of various topologies normalized to that of the baseline. Since we equalized the bisection cut metal area for all the topologies, they have similar link area consumption. The physical express networks have a slightly lower link area than the Baseline and EVC networks. This is because at the bisection cut, there are three links on top of each other (two express and one 1-hop link) but near the edges of the chip, there are only two links (one express and one 1-hop). The buffer area and the area consumed by the allocators are also similar. The crossbar area of the Baseline and EVC-network is about five times that of the Express-normal network. The crossbar area depends quadratically on the number of ports and the flit width. The Express-normal network has seven ports, as compared to five in Baseline, and has flit width of 24 B, as compared to 72 B is the Baseline. So the crossbar area of the Express-normal should be approximately (49/25)/9 which is close to one-fifth. The Express-CDLSI has about one-fourth the crossbar area that of the Express-normal because the Express-normal has crossbar twice the flit width. The crossbar area of the Baseline-24B and EVC-network-24B is higher than the physical express networks but 3 times less than the Baseline and EVC-network. Overall, the area of Express-CDLSI turns out to be half that of the Baseline-24B and EVC-network-24B given the same bisection cut.
metal area. It should be noted though that the Baseline-24B and EVC-network-24B have 3 physical networks. Ideally, the number of physical networks could be reduced and flit width increased for these networks and that would lead to interesting experiments to see which network topology has the best throughput given the same area footprint. We leave this for future work.

4.3.3 Power

Figure 8 shows the relative power consumption of various topologies. We show dynamic power for all router components and total power for the links. This is because the ORION 2.0 version that we were using did not have up-to-date static power estimates for 32nm. We leave this as future work. All numbers are plotted just before the saturation point of each configuration. We see that the link power of the physical express networks are much lower than the rest. This is because for every flit that traverses an express link, there are corresponding more hops that are traversed in the other topologies. Express-CDLSI has even more power savings, less than one-tenth the power of Baseline, due to the low power nature of the CDLSI lines. The crossbar power is also lower for the physical express topologies because of relatively shorter flit sizes. Baseline and EVC-network have similar link and crossbar power. However, the Baseline-24B and EVC-network-24B consume much lesser total crossbar power. This is because of smaller flit widths in these topologies and the crossbar power is a quadratic function of the flit width. The power consumed by input buffers is also lower for physical express topologies as compared to the Baseline and Baseline-24B. This is because the bypassing of physical routers in the physical express topologies leads to lower buffer usage and hence power consumption. The EVC-network and EVC-network-24B also have much lower buffer power consumption as compared to the Baseline networks. This is because the EVC-network allows bypassing of routers virtually leading to lesser buffer usage. The VC allocator power of the physical express networks is lower than that of Baseline and EVC-network due to bypassing, which leads to lower contention at routers for VC’s. The switch allocator power of all the express networks is lower because of lesser contention at routers as a result of bypassing. The physical express topologies have a lower clock power because they have a smaller router area, due to a narrower flit size, which leads to the clock tree not needing to spread over many circuit nodes. Overall the total router power consumption of the physical express topologies is lower than the virtual express topologies and all the express topologies consume lower power than the Baseline networks. As mentioned earlier, the Baseline-24B and EVC-network-24B have 3 physical networks and more experimentation should be ideally done to find out which topology has the best throughput characteristics, given the same on-chip network power budget. We intend to do that in the future.

4.4 Full-system Experiments

To understand the impact of various express topologies with real workload traffic, we evaluated the baseline mesh network, the express cube topology and express virtual channels in the GEMS [22] + GARNET framework. For the express cube topology we only evaluated Express-CDLSI since that has the same performance as Express-Normal but at a lower area footprint and also consumes less power. We performed extensive full-system experiments with the system level parameters remaining the same, while varying the on-chip network topologies. Our simulation setup and results are presented next.

Target system: We simulated a 64-core CMP with a shared L2 cache distributed in a 8x8 tiled mesh. Each core consisted of a two-issue in-order SPARC processor with 64 KB L1 I&D caches. Each tile also included a 1 MB L2 bank. DRAM is attached to the CMP via eight memory controllers placed along the edges of the chip. The DRAM access latency was modeled as 275 cycles. The router parameters were the same as those described in table 3. We again tried to normalize the bisection bandwidth of all topologies and hence gave the Baseline and EVC-network 72-byte links. Evaluating the three 24B separate physical networks configuration with real full-system traffic would have required instantiating three physical networks and changing the network interface to inject messages from different message classes into separate physical networks. We leave this experimentation as future work. In this section, we evaluate the Baseline, EVC-network, Express-normal and Express-CDLSI topologies. The router parameters were the same as those described in table 3.
Protocols: We evaluated a MOESI based directory protocol, which is part of the GEMS release. The protocol is a directory protocol in which the direction state is statically distributed along with the L2 banks and all L1 misses first go to a “home node” L2 bank, much similar to state-of-the-art directory protocols.

Workloads: We ran SPLASH-2 applications [27] on the above-mentioned configuration. We ran the parallel portion of each workload to completion for each configuration. All benchmarks were warmed up and check-pointed to avoid cold-start effects. To address the variability in parallel workloads, we simulate each design point multiple times with small, pseudo-random perturbations of request latencies to cause alternative paths being taken in each run [4]. We average the results of the runs.

Evaluation results: Figure 9 shows the comparison of the average network latencies of all the topologies, normalized to that of the baseline network. We see that the EVC-network has the best overall network performance. They outperform the Baseline and Express-CDLSI by 22% and 5% respectively. These numbers are consistent across all the benchmarks we studied. This, is because the benchmarks we ran, did not stress the on-chip network and mostly ran at low-loads. As we had seen earlier in Figure 6, the express networks have lower average network latency at lower loads as compared to the baseline network. The reason why EVC-network does better than Express-CDLSI for 8x8 networks is because the EVC-network allows bypassing of 1-hop, 2-hop and 3-hop flits but the Express-CDLSI network allows only 4-hop bypasses. For smaller diameter networks like 8x8, the long-hop express links are not as useful. We wish to also study the overall full-system impact of these network topologies. That is left as future work.

5 Related Work

5.1 Research on low-swing interconnects

A number of papers highlight the potential of low-swing links. It has been shown that a capacitive feed-forward method of global interconnect [12, 24] achieves nearly single-cycle delay for long RC wires, but with voltage-mode signaling. [5] employed not only capacitive pre-emphasis, but also RFI de-emphasis in order to improve the bandwidth. [16] use a current mode low-swing technique. They employed a FFE (feed forward equalization) technique at the transmitter and used TIA (Transimpedence amplifier to boost the small current) in order to enable better bandwidth. [19] uses capacitively driven low-swing links as global wires to broadcast control information for more efficient management of network resources. Our work however uses these CDLSI interconnects on the data path itself, enabling low power transmission of data.

5.2 Research on express topologies

Various express topologies have been proposed in the past to address the scalability concerns of the 2-D mesh topology. One solution proposed by researchers is a concentrated mesh [6]. In the concentrated mesh network, each network interface is shared among multiple terminals via a crossbar switch. This leads to a reduction in the total number of nodes in the network. By reducing the number of nodes, diameter of the network and the overall area footprint, concentrated meshes scale better than 2-D meshes. However, they still face similar problems as that of 2-D meshes when moving to larger nodes.

Various physical express topologies have been proposed to tackle the long diameters of 2-D meshes. We have used one such proposal, the express cubes topology [8] in this work. Express cubes employ long physical links to connect far away routers, leading to a reduction in the network diameter. These physical express links are used to bypass intermediate routers and links, thereby saving latency and power in the network. These bypass paths, however, require additional router ports, larger crossbars and extra physical channels. In our work, we have used the express cubes idea from the SMP domain, and applied it to the CMP domain by using high-speed and low-power interconnects to form these express paths. Another recently proposed physical express topology is the flattened butterfly topology [17]. It is a richly connected butterfly network mapped onto a two-dimensional substrate using a two-level hierarchy. The flattened butterfly topology reduces the network diameter to two, thus minimizing the impact of multiple router hops. However, in the flattened butterfly topology, the total number of links required in each dimension grows quadratically
with the number of nodes in the dimension. Multi-drop express channels (MECS) [11] uses a one-to-many communication model enabling a high degree of connectivity in a bandwidth efficient manner. MECS solves the problem of too many links of flattened butterfly.

Virtual express topologies try to mimic the performance of dedicated links by bypassing buffering and arbitrations at the routers. In our work, we compared our physical express topologies with one such virtual bypassing scheme, Express Virtual Channels [21]. Other virtual bypassing proposals include Token Flow Control [20] and Prediction Router [23] which both allow turns as well. However, EVCs seem a reasonable virtual alternative to express cubes (which don’t turn in our design) and hence we used it as one of our exploratory configurations.

6 Conclusion

In this paper, we have evaluated the effectiveness of physical and virtual express topologies, pointing out the latency benefits of physical express topologies vs. the throughput benefits of virtual express topologies. Clearly, hybrid express topologies that leverage both physical express links and bypass intermediate routers virtually can trade off the pros and cons of these two types of techniques. We seek to further investigate ways to leverage the best of both physical and virtual express topologies, and drive towards the energy-delay-throughput of an ideal interconnect fabric.

References


