Comparison study of scalable and cost-effective interconnection networks for HPC

Pablo Fuentes, Enrique Vallejo, Carmen Martínez, **Marina García**, Ramón Beivide

Department of Electronics and Computers, University of Cantabria

This work compares scalability, cost and performance of two network topologies proposed for large-radix routers: Concentrated Torus and Dragonflies. The comparison concludes that for middle-range networks both Concentrated Tori and Dragonflies are cost-efficient options, while for large networks Dragonflies dominate because of their cost and scalability.

Dragonfly topology

A *Dragonfly network* is hierarchically divided into 2 levels: the first level is composed of multiple groups of routers, with routers connected all-to-all within a group; and the second level connects the different groups in an all-to-all topology. The network is fully described with the following parameters:

p	Compute nodes attached to every router
a	Routers per group
h	Global links per router

In a well-balanced Dragonfly, with its performance not limited by the bisection bandwidth, the relation $a = 2p = 2h$ holds. Such network connects up to $\frac{k^3}{16}$ nodes.

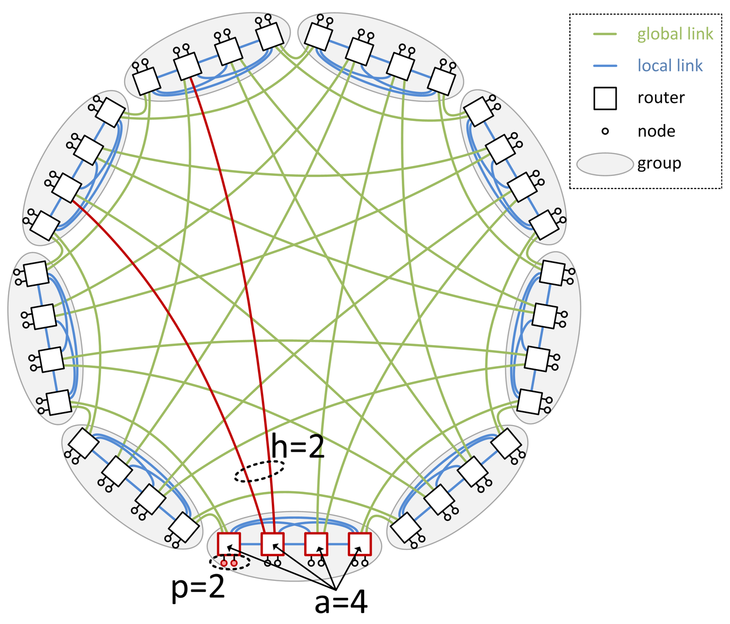

Figure 1: Dragonfly network with $h = 2$.

Concentrated Torus topology

We study symmetric Concentrated Tori with D dimensions. The routers on each dimension are connected as a ring, using t parallel links between each pair of consecutive routers. The network is fully described with the following parameters:

p	Compute nodes attached to every router
r	Routers per dimension
D	Number of dimensions
t	*Trunking factor*, phsyical wires per link
k	Router degree

The bisection bandwidth imposes $\frac{pr}{22} \le 2t \Rightarrow pr \le 8t$. We assume $p \geq t$ to avoid link overdimensioning with respect to the amount of computing nodes. The maximum number of nodes with these restrictions is

$$N = 8^D t \text{ and } t = p = \left\lfloor \frac{k}{2D+1} \right\rfloor$$

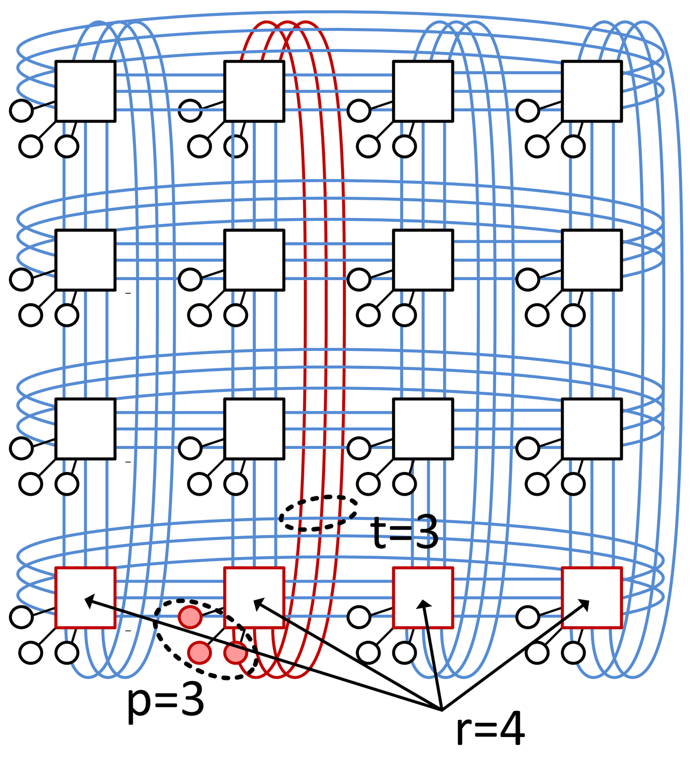

Figure 2: Concentrated Torus with $r = 4$.

Network scalability

Figure 3 presents the maximum network size with increasing router degree up to 64 ports, using both balanced Dragonflies and Concentrated Tori with 1 to 4 dimensions D. The maximum network size has been calculated based on the bisection bandwidth limits of each topology, as discussed above. We restrict to $D < 5$ for layout restrictions. Figure 3 shows that 4D Concentrated Tori can reach a higher number of computing nodes when the router degree is below 27 ports.

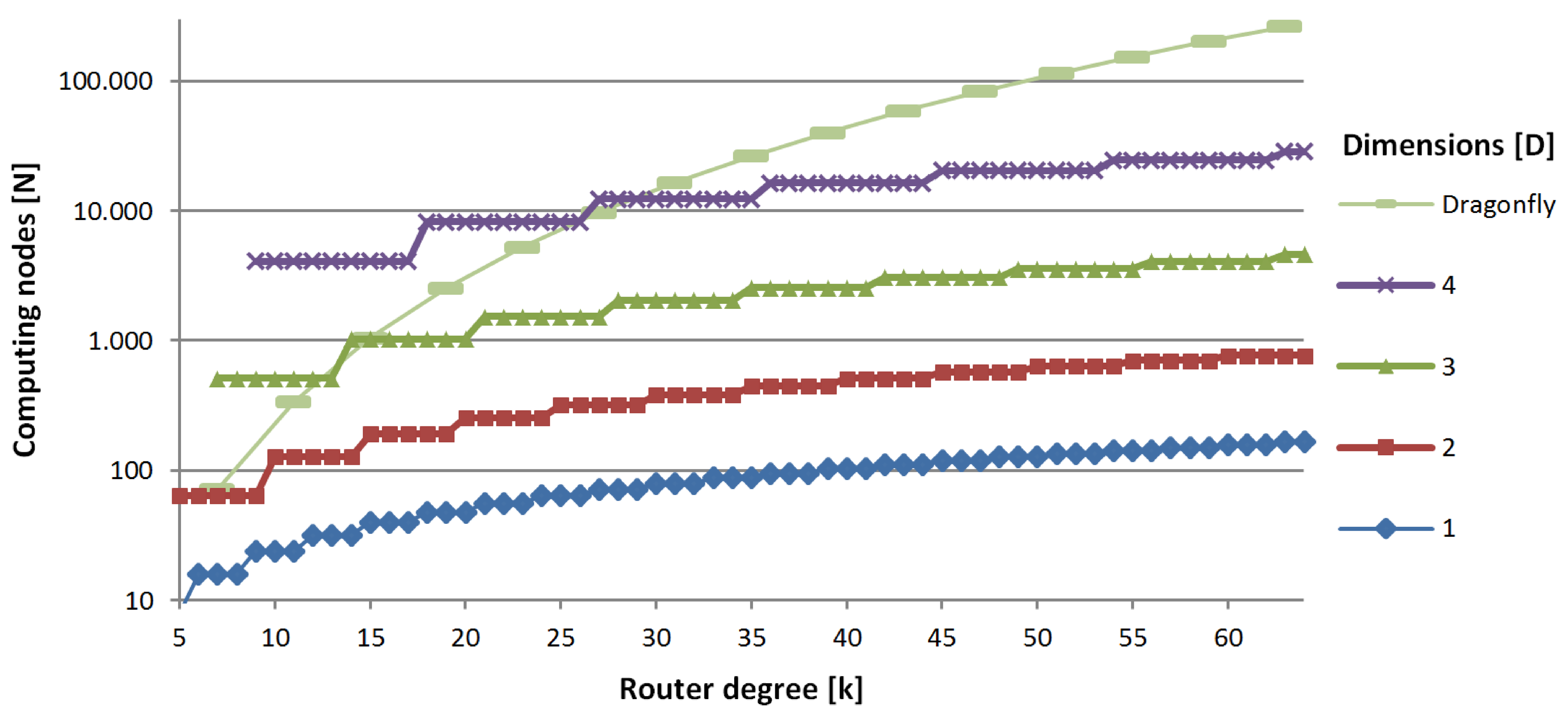

Figure 3: Maximum scalability for a router size

Cost comparison

Network cost is calculated from the number of routers and wires using the prices from [1]. Each router is 390$, and link cost varies with its length and technology. All links in Concentrated Tori are electrical wires, and their length is calculated from the distribution of nodes on racks, using a folded torus layout. The Dragonfly cost is calculated similarly to [2]. Unlike tori, Dragonflies have long optical wires ($\propto$ 220$ per wire). Figure 4 shows the cost per node for a network built with 64-port routers. 3D or 4D Concentrated Torus are cheaper than Dragonflies under 20.000 computation nodes.

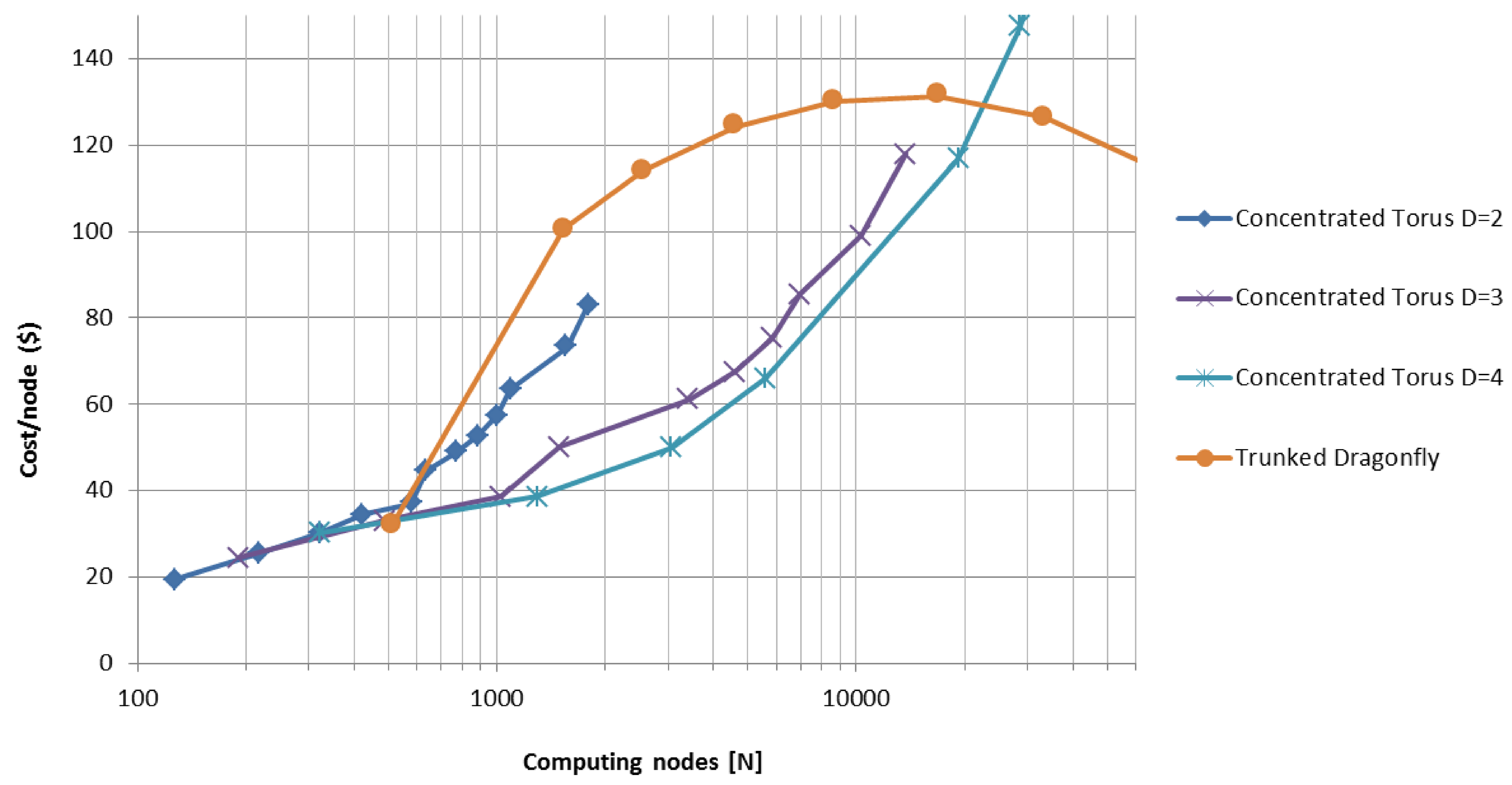

Figure 4: Cost comparison using router size 64.

Network performance

Figures 5 and 6 show network performance for a Concentrated Torus and a Dragonfly of similar size, using routers of degree $k = 27$ and random uniform traffic. Simulated link latencies are proportional to the average wire length in each case: 10/100 cycles in local/global Dragonfly links and 36 cycles in 4D Concentrated Torus (considering a folded layout). The figures depict maximum throughput per node and average latency per packet as the network load increases. The throughput is limited due to network contention, which is lower in the Concentrated Torus since it employs adaptive routing and it has several physical links per hop. By contrast, the Concentrated Torus has a longer average distance, what causes a higher network latency.

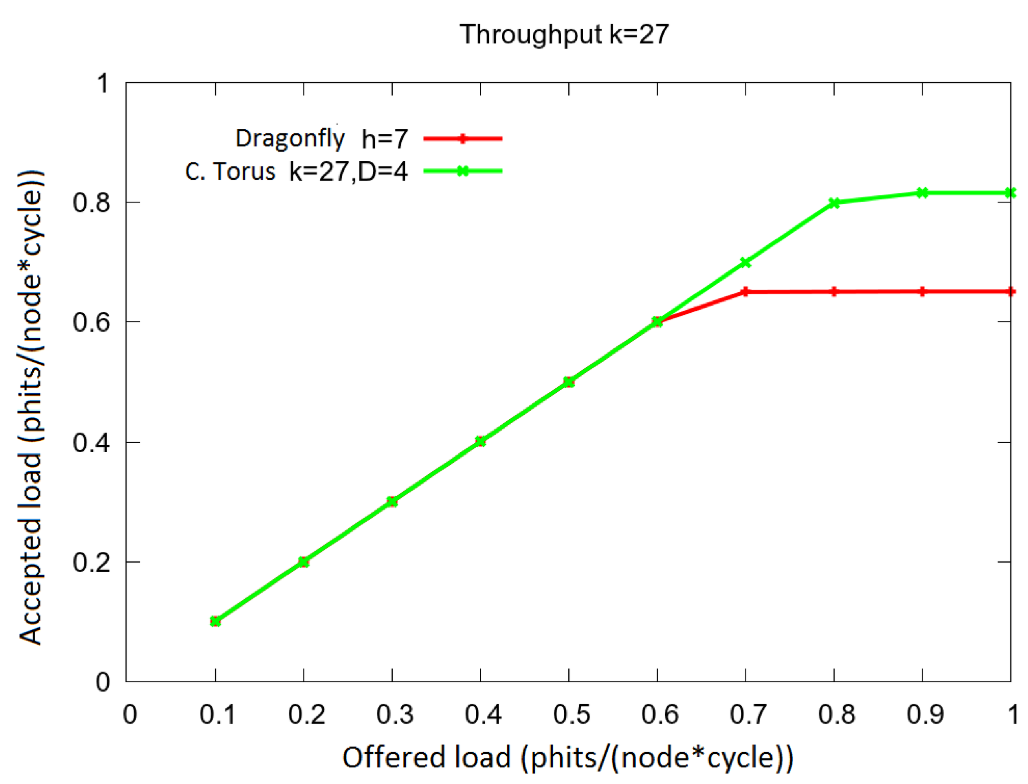

Figure 5: Throughput comparison using router size 27.

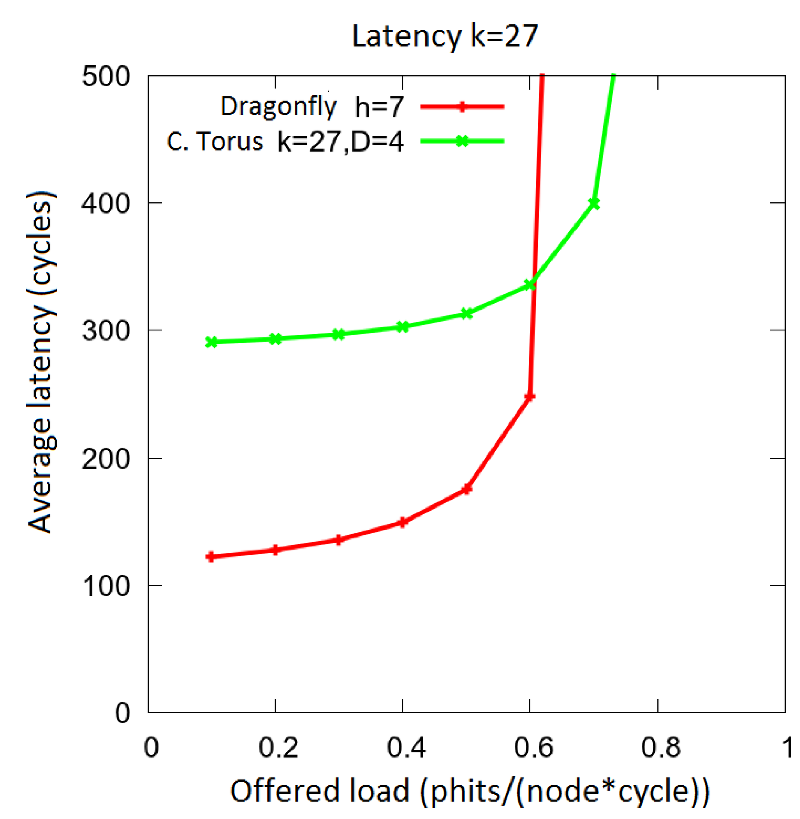

Figure 6: Latency comparison using router size 27.

Acknowledgements

This work has been supported by the Spanish Ministry of Science under contracts TIN2010-21291-C02-02 and CONSOLIDER Project CSD 2007-00050, and by the European HiPEAC Network of Excellence.

References

[1] J. Kim, W. Dally, and D. Abts, "Flattened butterfly: a cost-efficient topology for high-radix networks", in *ISCA '07*.

[2] J. Kim, W. Dally, S. Scott and D. Abts, "Technology-Driven, Highly-Scalable Dragonfly Topology", in *ISCA '08*.

Conclusions

A Concentrated Torus has better scalability and lower costs than the corresponding Dragonfly network for mid-sized networks with less than 10000 compute nodes. Evaluations using synthetic traffic show that a Concentrated Torus can provide higher throughput than a Dragonfly network, though latency is higher.