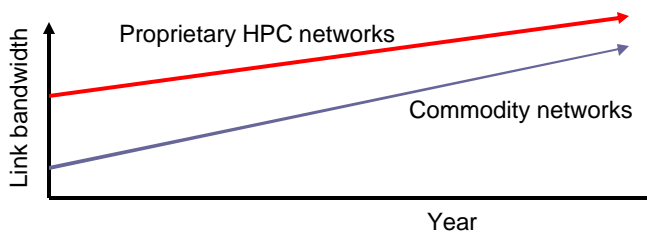


Future trends in high-performance interconnects for parallel computers

Craig Stunkel
IBM T.J. Watson Research
Center

Commodity networks

- An increasing portion of customers will be satisfied with commodity networks
- Need more performance than that?
 - You may need to help fund development!



Network interfaces

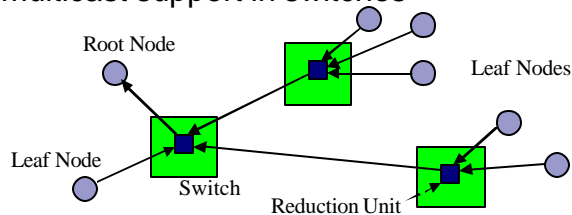
- Will they become more intelligent?
 - Not necessarily
 - Already programmable in most cases
 - “Killer” micros are hard to compete with
 - But more common functions will be offloaded
 - Most high-performance network interfaces are leveraging very similar techniques

Network interfaces

- Will they move on-chip?
 - Likely, and gradually
 - Reduces latency of I/O bus crossing
 - Closer to memory nest
 - Closer to the (multiple) processors
- More network-friendly processors and memory nests will be crucial to continued performance gains

Collective communication

- Efficient collectives are increasingly important
 - Some app performance is collective-bound
 - Integrated network support is a big advantage
 - For indirect networks, reduction (combining) and multicast support in switches

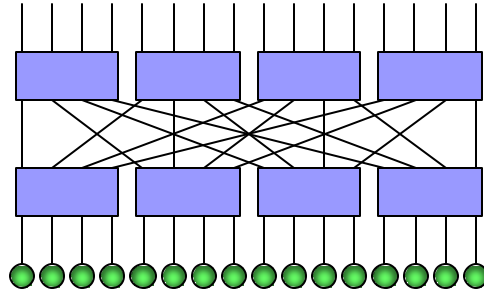


Collective communication

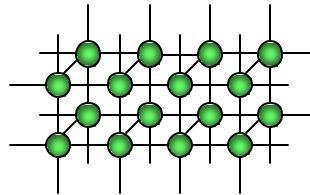
- Separate collective network?
 - Preferable to integrate with unicast network
- Not always possible to integrate
 - For example, if augmenting a commodity network

Topologies

- Only two of any significance today:
 - Fat-trees and similar indirect networks
 - Most prevalent today



- k-ary n-cubes (mesh, tori, hypercubes)
 - Used in several tightly-packaged MPPs/clusters



Hot Interconnects 2004

7

Topologies

- Mesh/tori cost scales linearly
 - Good for constrained bisection
 - Good for local communication
- Fat-tree bandwidth scales linearly
 - In practice, bisection is rarely constrained
 - Especially for clusters
 - Good for random or uniform-distributed traffic
 - Good for difficult permutations
- On-chip networks
 - Meshes best because of the constrained bisection

Hot Interconnects 2004

8

The effect of optics

- Optical links make distance “free”
 - Aiding topologies which scale bandwidth
- Biggest cost of many future networks will be in the optoelectronic transceivers
- Fast-switching optical switches will remain more costly than their electronic counterparts
- Slow-switching optical switches could be cheaper
 - Can they be useful for computer systems?

Switch radix (number of ports)

- Switches are getting larger
 - Translates to less hops
 - Higher performance, given same buffering per port
- However, it is difficult to grow both port bandwidth and radix
 - Requires multiple-chip solutions
 - Minimum packet size suffers (grows)



Switch radix (number of ports)

- One solution: *don't* try to achieve both high radix and high port bandwidth
 - Just go for high radix ...
 - ... and have multiple switch planes/networks
 - Packets can arrive out of order
 - Commodity switches may be more easily leveraged