

## Modern Interconnects – Myrinet

- 190 of top 500 clusters have Myrinet interconnects

## Origins

- based on MPP architecture – multicomputer parallel network
  - packet-based
  - regular topology – i.e., can figure out address mapping
  - cut-through routing
  - flow control on every link – asynchronous signals
  - low error rate assumption
  
- Caltech Mosaic multicomputer

## Myrinet Design

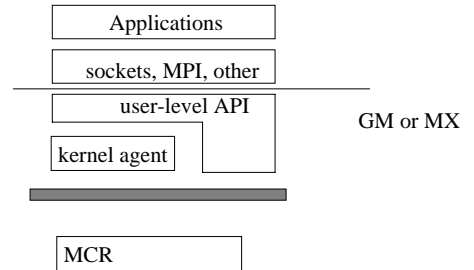
- point-to-point, host NIs and switches
- flow control w/ stop and go signals
  - stop and go at threshold values, not at min/max to allow for some slack
- it's own frame format
  - header, payload (arbitrary length!), CRC at the end, on entire packet, interpacket gap...
- blocking-cut-through routing

## Myrinet NICs

- LaNai chips
  - SRAM, bus (originally proprietary E-Bus, LaNai-X with PCI-X), DMA engine, packet engine
  - DMA engine can compute checksum on data transferred on bus
  - programmable processor – Myrinet Control Program (in memory write protected by LaNai processor).

## Host interface

- command and acknowledgment queues
- scatter/gather DMA
- 1 vs. zero-copy data transfers
  - user space to NIC memory, checksum by DMA engine
    - packet in multiple dma transfers...
- interrupt enable/disable on rcv.
- automatically select address mappers in network
- IP multicast support



- third generation 2x2Gbps
- fourth 2x10Gbps, 10GbE compliant
- LaNai-X – multiple interconnects supported in firmware (Myrinet, GbE, Infiniband).
  - PCI-X, plus proprietary SAN interface...