Memory Interconnect Directions

DIMM Interface Replacement
JEDEC JC-42 Project

Bob Davis
bob@scsi.com
Summit Computer Systems, Inc.
8 Sept 2003

Agenda

- Why am I here?
- Why Design a New Memory Interface?
- What's Changing
- Memory Hierarchy
  - Regularizing the Stored Data Interface
- Memory Vendor Design Freedom
- An Possible "Ideal" Interface
- Memory Connection Capabilities
- Possible Memory Interface Solutions
- Who Benefits from this approach?
- Issues

Why a New Memory Interface?

- Support the Changing Memory Needs
- Extend the Addressing Capability
- Improve Reliability (RAS) – Support Redundancy
- Allow design freedom for Memory Vendors
- Recognize Memory Hierarchy and use
- Regularize storage interconnect
- Improve Signal Integrity and usage models
- Reduce Pin Count – "It's the PIN's Stupid – the PINS".
- Improve Pin (ball) Utilization – Use better technology
- Vastly Increase Scalability, Availability, Performance

What In Memory is Changing

- More On-Die processor Cache
  - Level Three Cache use and size is growing
- Total Memory Size Requirements Growing
- Memory is Memory is Memory
- Remote Memory – Make RDMA "Natural"
- Reliability of Memory
- Memory Technology is changing Getting smaller - nano
- Memory Speed is not tracking processor speed
  - Physical delays
- Layout constraints dominating designs
  - Signal Integrity
  - EMI
- Need Plug-n-Play
- Memory operations are for Cache Lines and Blocks not Bytes or bits

Memory Hierarchy

- Level 0 – Processor Registers – access – xxx ps
- Level 1 – First level cache – I and D Cache – xxx ps
- Level 2 – Second level cache – unified 1- 8 MB – x ns
- Level 3 – Third level cache – General – IBM etc – x ns
- Level 4 – “Main Memory” – up to 1 TB – xx ns
- Level 5 – Virtual Memory – Paged Memory XTB – x ms
- Level 6 – Fast Disk – 100GB – access x ms
- Level 7 – Active Storage – 2x growth from 8TB – xx ms
  - RAID and “Synchronous” backup - Failover
- Level 8 – Near Line Storage – from 96TB – xxx ms
  - RAID and Geographical distribution – Business continuance
- Level 9 – “ColdStore” – Petabyte increments – 30s
  - RAID, dump targets, large data bases, redundant
- Level 10 – “Tape Storage” Archive – Min, Hours, Days
  - Dumps, transport – questionable recovery

Ideal Memory Interface

- Generic Memory Interconnect
- Extensible Design – Room for Options and Future Developments
- Byte Lane design - Support 1 to 64 Byte Lanes
- Does not change at each new Processor
- Plug'n'Play
- Self Timed From Interface – No distributed clocks
- AC Coupled - 8B10B encoded for DC Balance
- Longer Interconnects - Extendable
- Possible switch or router Enabled
- Capable of RAIMM Operation (RAID without rotation)
- Transmission Path CRC32 Protected
- Smart Memory Module
  - Memory Caching Write and Read
  - Module Maintenance performed by Module
- Configuration and control via in-band messages

Needed Connection Capabilities

- Addressing large memory
- Variable Bandwidth Thin to Wide 1-n “wires”
- Follow technology – 3Gb-> 6Gb->10Gb-> + per second per wire pair
- Independent of actual Memory Technology
- Be independent of distance
- Scale 64 + bits of address
- Source and Destination identification
- Secure – Authentication and Encrypt
- Support redundancy
- Support MEMORY Coherency
- RDMA? Capability
- Scalable design
- Multiple implementation levels
- Higher Pin utilization
- Cost Effective
- Vendors make money
- Vendor differentiation
- Message optimized – not bit optimized

Memory Vendor Design Freedom

- Define the interface tightly for commonality
- Plug-n-Play for user
- Implementation details to vendor
  - Tune for various price point – interface width
  - Tune for various reliability needs
  - Make available various sizes 2^10 – 2^64
  - Make available different power level
  - Make available different speed requirements
  - Change Technologies at will
  - Add Caching for better performance in interface
  - Add redundancy features – RAIMM techniques
  - Supply multiple technologies in same interface
  - Optimize yield
  - Differentiate and compete other than by cost
  - Meet Additional needs of systems vendors

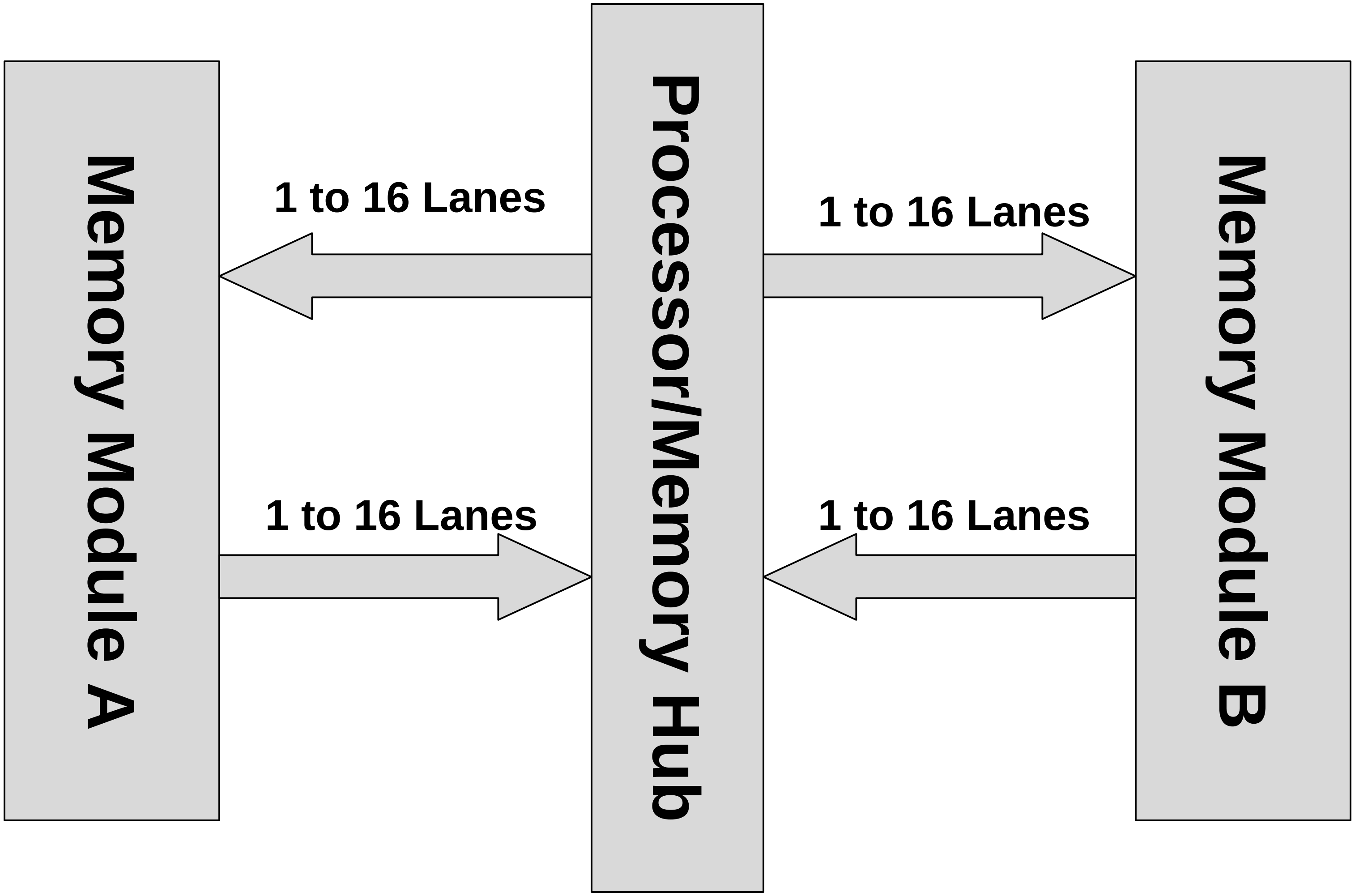
Byte Lane Design Advantages

- Scalable in Width and Depth
- Capable of Redundant Memory Techniques
  – Redundant Array of Inexpensive Memory Modules
  – Support really paranoid customers
- Hot Replacement Possible
- Discovery and Configuration – In band
- Self Timed - Clock embedded in data
- Support orderly command structure
- Distance Limitations removed – except latency
- Reduce cost
- Well known SerDes designs
  – Ongoing development effort
- Possible switcher and router configurations

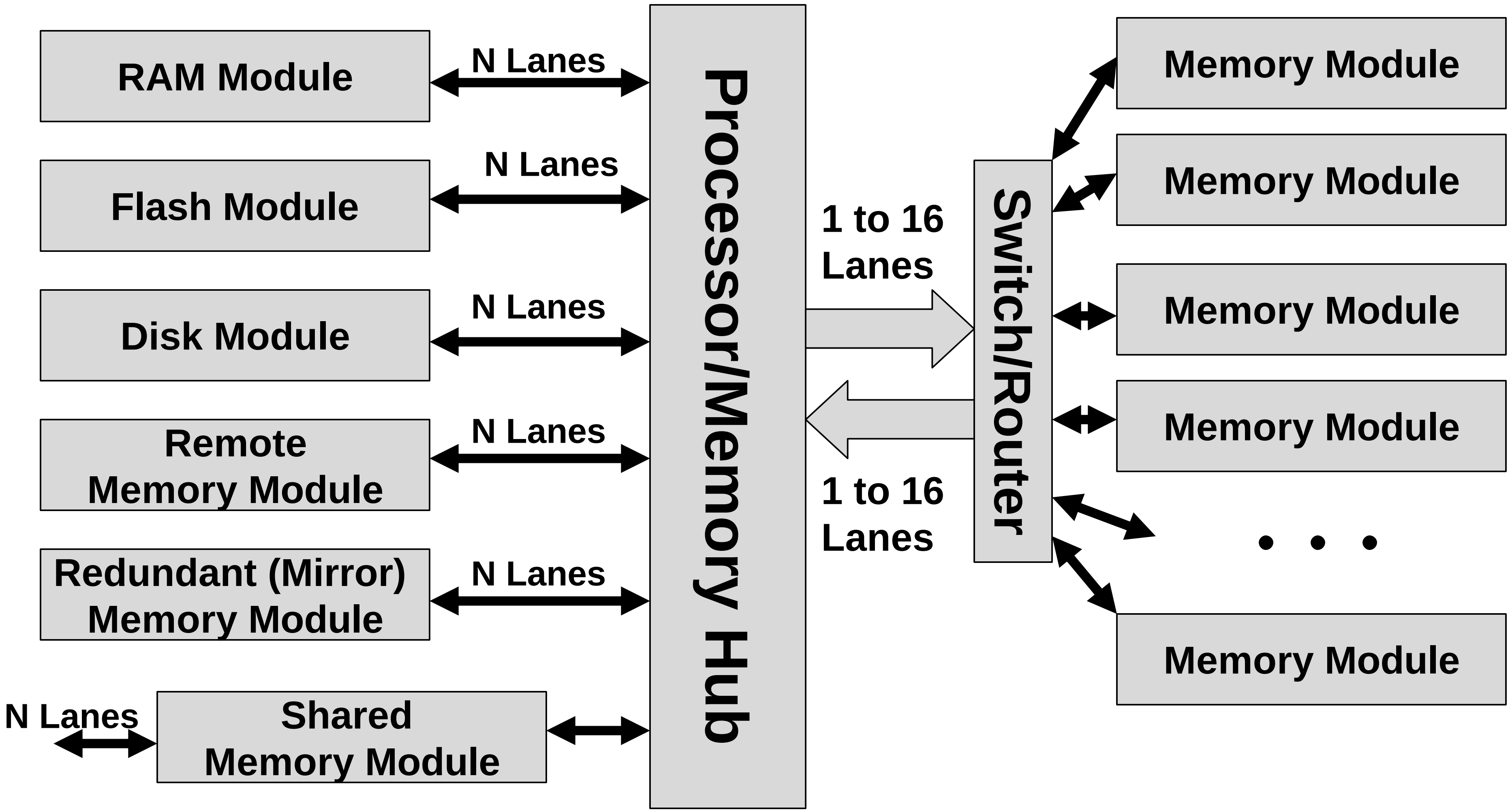
Transfer Peak Bandwidth

	1 Lane 4 Wires	4 Lanes 16 Wires	16 Lanes 64 Wires	32 Lanes 128 Wires	64 Lanes 256 Wires
2,5 GBaud 2.0 Gb/s	**250 MB/S**	**1 GB/s**	**4.0 GB/s**	**8.0 GB/s**	**16.0 GB/s**
3.125 GBaud 2.5 Gb/s	**312.5 MB/s**	**1.25 GB/s**	**5.0 GB/s**	**10 GB/s**	**20 GB/s**
6.25 GBaud 5.0 Gb/s	**625 MB/s**	**2.5 GB/s**	**10 GB/s**	**20 GB/s**	**40 GB/s**
9.375 GBaud 7.5 Gb/s	**937.5 MB/s**	**3.75 GB/s**	**15 GB/s**	**30 GB/s**	**60 GB/s**
12.5 Gbaud 10 Gb/s	**1.25 GB/s**	**5.0 GB/s**	**20 GB/s**	**40 GB/s**	**80 GB/s**

Simple Structure



More Complex Structure



Possible Solution Technology

- Extend Current RAS/CAS
  - Poor Pin utilization
  - Poor flexibility
- PCI Ex AS protocol
- Enet/IP based Protocol - RDMA
- InfiniBand
- SCI – Scalable Coherent Interface
- New Message Based Interface
  - Byte Lane optimized

Packet Protocol

- Expandable - Extensible
- Multi-Generation Capable
- Not Specific to one type of Memory
- Write-only Packets
- Reliable
- Routable
- Best ROI
- Utilize Current Packet Technology

PCI Express AS Base

- Advantages
  - Routing switches being built
  - Is somewhat scalable
  - Could define new Packet protocol similar to definition I made for larger systems.
  - EUI identification built in.
- Disadvantages
  - Limitations of AS in large systems
  - Addressing poor beyond 64bits
  - Lacks configuration in peer to peer modes
  - Source routing –

Beneficiaries of New Interface

- Processor Vendors
  - No needed knowledge of Memory Module Architecture
- Memory Vendors
  - More Optimized Memory Design – Less physical driving contraints
  - Add interface to chip for smallest systems
- System Builders
  - More Options for Memory
  - Better Signal Integrity design
  - Faster Time To Market
  - Capable of UPGrade after sale
  - Expanded System Architecture to meet specific design goals
  - Component Compatibility over time for longer product life
- Memory Module Vendors
  - More design freedom in sub-unit design
  - Opportunity to add capabilities for differentiation
  - Memory is Memory is Memory –different technologies
- Users
  - Plug-n-Play
  - Capable of upgrading system memory and performance
  - Ability to support special applications needs.

Next Steps

- Gather the Experts
- Define Long Range Architecture
- Choose the Technology trend
- Choose the Interface Definition
- Plan for expansion with Options
- Put in Place by 2005/6
- Follow and Expand on SCI…

Conclusions – Final Plea

- Break the tradition of the IBM PC
- Don’t bit optimize as first principle
- Make a Long Lived Generic Memory Interface
- Make and Interface that can be implemented over a wide variety of applications
- Expand the scope to larger Memory objects
- Shift to a message style packet based Interface
- Move care of the Memory to the Memory
- Use Optional Fields to build in very dumb to very capable module capabilities
- Remove Distance limitations
- Allow authentication of module and Redundancy