

# The Memory Channel PAR Justification

Bob Davis  
Summit Computer Systems, Inc.

## Name – Memory Channel Standard

### Scope

Develop a flexible, scalable, secure data interface to transfer data to and from, and between, addressable storage. This protocol will be technology independent, to be supported by current communications links, and will remove size and distance limitations on data transfer, with redundancy and coherency methods.

### Purpose

To develop a memory data transport protocol capable of supporting data growth and data security requirements in the changing microprocessor environment. Memory is stored data in many forms and locations. The Memory Channel protocol shall be independent of link technology. This protocol shall be capable of transparent, secure, access to large, local and remote memory systems, employing coherent and redundant storage methods.

### Why specify a new Memory Channel?

The Memory Channel addresses the location and transfer of data to and from a data storage location. This new Memory Channel does not want, or need, specific information on the structure of physical storage mechanism. This channel does not address data structures within the target storage environment. Only a common understanding of a memory unit address is needed. This will expanding, significantly, the size and possible location of usable memory.

### *Non-Goals of the Memory Channel*

No Specification of new physical technologies. No Redefinition of communications protocols. The following assumptions are also made: MTU is a problem of the carrier. Segmentation and Reassembly is a function of each carrier

### Goals for the new Memory Channel

#### *Memory is Memory is Memory*

Data storage takes many forms from CPU registers to Archive Tape in a vault. The difference between the various forms of storage is, physical media and blocking factors, the access time and bandwidth and permanency of the data.

#### 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 – xx ns
- Level 4 – “Main Memory” – up to 1 TB – xx-xxx ns
- Level 5 – Virtual Memory – Paged Memory XTB – x us
- 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 - Geographical distribution – Business continuance
- Level 9 – “ColdStor” – Petabyte increments – 30sec - RAID, dump targets, large data bases, redundant
- Level 10 – “Tape Storage” Archive – Min, Hours, Days - Dumps, transport – questionable recovery

All different levels of memory store the same information with different access times. The byte retrieved or written is the same.

#### *Support Large Memory*

Current memory channels segments have had upper limits of 4GB based on the availability of 32 bit addressing. The Memory Channel should be planned with a minimum of 10 year design life expectancy. If the history of the ISA and PCI architectures are any kind of a predictor, it will probably still be viable in 15 years.

Consider normal trends, with history, as indicators of future developments. Interconnect speeds still double every 18 months. Typical and Max Memory sizes increase, requiring a additional bit of address every year to 18 months (12 bits in 1974 + 30 years = 42 bits now). The Sun SunFire 15K is offering 512 GB of RAM per domain and currently up to 18 domains now (44 bits) and IBM is just behind at 256 GB of RAM. This would lead to 48-60 bits of memory address space per user requiring the capability of 64 bit addressing.

Alternate forms of storage will become available that will be significantly denser and less expensive over the life of this specification. While the current technology allows for devices in the 200 million devices per die, predictions of this growth by 2014 expand this by at least 2 orders of magnitude. These new devices will both need, support, and provide these larger memory spaces.

Processing elements will continue to grow, based on the industry long term technology roadmap. These roadmaps also show technology advancing to 35nm IC technology by 2009, with smaller than 25nm technology within the planning horizon of this Memory Channel. With the smaller geometries, the greater probability of geometry induced faults and the need for better error detection in all circuit elements. Planning/allowing for future versions of the specification will help grow to places we have not thought about yet.

### ***Memory Technology Independent***

Memory technology in all it forms is progressing rapidly and independent of the processor technology and instruction set technology.

Many of the current processors are designed to support a given memory technology such as DDRx. The complexities of these memorys are designed into the processor which also makes the processor slave to that technology. This was much more important prior to the local caching of data when all accesses went to processor external memory. As shown in the hierarchy of memory above the access of die is at level 3 or 4 and moving to higher levels to support the increasing execution speeds of these same processors.

It is now time to divorce the actual memory technology from the memory request operation.

### ***Message Based – not bit optimized***

Most accesses to storage are now block oriented with a number of bytes(octets), quadlets, or octlets transferred for each memory access, such as a cache line. This is effectively a message being transfer to the storage from the requesting element for action on a block of data. The Memory Channel recognizes this reality and builds the protocol to support a general purpose storage access and transfer message.

I/O operations have also become memory operations. The difference in delays and the small number of I/O address has long since transitioned to the only difference between I/O data transfers and Memory data transfers is the resolved address.

### ***Common Protocol Layer***

The heart of this work is the building of a common protocol that has sufficient capabilities to be used for all memory transfer operations and have the longevity to survive technology of 10 or more years. This is simply a protocol for the handling of the general purpose storage transactions.

### ***Physical Layer Independent***

Transportation of the Memory Channel protocol is independent of the protocol. Data link development is an ongoing effort with extensive work in every possible area of development. The Memory Channel will gladly follow and make use of this work and be conveyed over it.

I anticipate that multiple different physical layer technologies will be used in each access that extends more than several millimeters from the processor element. Clearly inter-chassis data transfer requires different technology than 20 cm access methods.

### ***Memory Distance Independent***

The demand for memory speed is always present and mediated only by the speed of light and the technology of the devices. Caching has been the most prevalent answer to this problem. This has also reduced the need, with in this context, for the same access time from all memory elements. Please see again the hierarchy of memory above. With time as a relative and controllable quantity, the location of the memory can extended to whatever length needed to support the use of that memory. Very large memories can not be physically located next to the processing element due to size constraints. The system designer will now have the ability to tradeoff response time, cost of memory, size of memory, protection of memory as independent variables.

### ***Source and Destination Addressing***

Extending the memory domain with global storage requires that the memory channel message has both a destination and source identifiers. This is required for authentication of the source of the message, the return path for the return data, if any, and for security of the transfer.

This Memory Channel Protocol specifies the destination address as a combination of destination node and memory offset within that node. Node addressing is derived from a globally unique identifier. This is a 64 bit identifier in the format of, either an EUI (Extended Unique Identifier) or a WWN (World Wide Name). An EUI consist of the IEEE Registration Authority Committee assigned 24 bit OUI (Organizational Unique Identifier) plus the organization guaranteed 40 bit unique number. The WWN consists of 3 fields within the 64 bits consisting of a 4 bit NAA (Naming Authority) mostly fixed at the value 5h which is the IEEE RAC, the 24 bits of the IEEE assigned OUI and 36 bits of organization guaranteed unique number. This forms a 128 bit address that is globally unique.

### ***Security – Authentication, Validation, Encryption***

Any channel used to carry customer data must support Authentication of the user(s), validation of the operation, and hiding of the data with encryption. Authentication identifies the source of the request and the access privileges of the requestor for the transaction proposed. Validation assures the request is complete as received with all information correct. Encryption hides the nature of the transfer to all, assumed, snooping eyes.

### ***Effective RDMA and SRDMA***

This Memory Channel supports Remote Direct Memory Access and the Secure Remote Direct Memory Access methods for direct access to the large memory spaces. Remote access is defined as access where the identification of the source is required. In the Memory Channel this is anytime other than when the shortest address form is used in a closed system.

### ***Support Redundant Memory Operation***

Data reliability now requires redundant storage of all forms. This will necessarily include RAM, DRAM, DISK and any other form of storage for data envisioned. Redundant storage will need to cover both: 1) The reliability, or unreliability, of the storage mechanism and 2) Disaster Recovery of the data from outside the identified threat zone. Mechanism covered by the method will include RAID, RAIIM, and one or more remote asynchronous, and possibly synchronous, backups and mirrors of the live data.

Each of these mechanisms will have a maintenance plan associated with it. The mechanism and maintenance plan will be dictated by the criticality of the particular data set and will vary from one data set to another and NOT part of this standard. This redundant operation is orthogonal to the coherency requirement but can probably used a similar mechanism. While the coherency mechanism is tracking active copies of the live data base this portion of the Directory based, linked list approach to identifying the existing copies, will maintain the coherency of the stored data in the distributed domain.

### ***Support Cache Coherent Operations***

Almost all systems today use one form of caching or another to maintain acceptable performance. With all the caching taking place, a mechanism is require to maintain the coherency of these caches to prevent stale data at the user site. The best current method for maintaining coherence over a distributed caching system is the SCI protocols, IEEE Std 1596-1992 for a Directory Based Cache Coherency schema. Use of these protocols will require some modification to embrace the nature of the communications protocols and the size of the memory landscape. SCI directory based cache coherency system was designed around 64K nodes with a 48 bit offset into each node and deals with cache lines of data. This system will expand that to  $2^{64}$  potential, although sparsely populated, nodes with a 64 bit offset into node address space and will deal with much larger scalable objects and files over a much larger set of domains. This extension work is part of the project and standard.

### ***Scalable Design***

Every aspect of this Memory Channel is aimed at a fully scalable system and data design.

### ***Multiple Implementation levels***

While there is one protocol level for the Memory Channel, various features of that channel can be selectively disabled where they are not needed.

The level of implementation is discernable at the beginning of the header for each transaction. If the destination of a message is not able to comply with the expected performance, a negative acknowledgement is returned with the error code.

### ***Plug and Play***

With a common interface, all memory devices should be able to work together. And example is USB flash memory system.

### ***Cost Effective***

Supporting slower and larger memory systems can save money in a system. Likewise the enhanced memory system with faster response over a wider channel will improve the performance where needed. These features can be traded off for optimum performance.

### **Create Vendor Opportunities aka USB Memory**

With a common Memory Channel interface, vendors can create memory of various performance and cost parameters for general use or at a targeted market segment.

### **Support Vendor Differentiation**

The common interface allow vendors to differentiate there products to improve performance, reliability or cost. Examples are additional levels of caching or different vendor created caching algorithms to improve performance.

### **Separately define Encapsulations**

This document discusses the protocol for the message and does not define the interface with the transport mechanism. Among the different structures that can carry the Memory Channel protocol are: Infiniband, PCI Express, USB, Firewire, Ethernet, TCP/IP and the other variant of IP traffic defined through the IETF.

### **Beneficiaries of the New Memory Channel**

The benefits of the new Memory Channel are distributed throughout the supply channel from processor and memory vendors to the end user base.

### **Processor Vendors**

Processor designer and suppliers will have a common target implementation of memory for several generations of processor designs. All constraints of specific knowledge of the operation of the memory are removed and transferred to the vendor of the memory. Operation with respect to decisions respecting the anticipated delay are already being made by the processor and the execution path modified while a memory access is being performed.

Requirements for the refreshing of dynamic RAM and other memory specific function, such as pre-charge cycles become the function of the memory greatly simplifying the design of the processor or memory controller.

### **Memory Vendors**

Memory vendors gain by being able to optimize the overall memory product rather than just meeting the "DIMM" standard. Innovation by the memory designer can improve the performance of a block of memory using statistical techniques similar to those used in the cache on the processor. This is a value add for the vendor allowing for specialization into the different areas in which memory is utilized in the system.

The physical constraints of the packaging are removed and allow the optimization of form factors to meet design parameters. This Memory Channel interface may be added to new memory die or to a local controller similar to the currently being designed AMB device to supply FBDIMM interface. This provides a far more general operation on the response to a memory request than the AMB design, as currently envisioned although in may be slower in the larger forms.

The larger vision of the interface being implemented at chassis level with multiple technology memory devices providing a uniform face to the requesting process is also valid.

### **System Builders**

The benefit to system builders is a greater variety of memory products available for solving the system objectives. A generalized memory channel interface removes specific design constraints with respect to system layout and the ability to upgrade the memory with changing product needs.

Individual memory system design can optimize the design of the memory subsystem signal integrity needs. The tracking of a specific memory cell requirements are removed to the maker of those memory cells.

Different speed memory subsystems will work together and perform per there individual design. Adding faster or slower memory may affect overall systems performance but will always work!

Faster time to market is realized with a general purpose memory interface that is insensitive to the type of memory involved.

Memory constraints of the application are removed as the number of DIMM slots is no long a limit on the overall size of memory.

### **Memory Module Vendors**

Memory Module Vendors can compete for business with more than the best cost for a given configuration. This standard opens the path for memory innovation to meet a set of specification defined by the application and the user. Combinations of memory technologies in one unit should provide a far better solution to specific design goals and computational situations.

Design innovation can provide opportunities for specific design wins that better meet the system builders and end users goals. The care and feeding of this particular module is the property of the module. RAM with higher refresh requirements can be used on modules with that specific module spending more refresh cycles to take care of that memory. This adjustment is transparent to any other memory in the system.

### ***End Customer, Users***

The end user wins by having the ability to upgrade the memory to meet the changing needs. Plug and Play is automatically enabled with each memory sub system taking care of itself. The EUI makes each module and its memory space unique. Compatibility with existing memory in the system is not a concern when adding new memory, added memory takes care of itself.

Memory modules can be designed to meet special needs, such as graphic, or portable application, while maintain transparency with the other memory systems.

### **Design Objectives**

#### ***Generic Memory Interconnect***

This protocol will support all types of memory and is independent of technology, speed, distance, and physical implementation.

#### ***Remove any memory size and location constraints***

Scalability in memory capacity, size and physical location are a requirement of the design.

#### ***Extensible Design – Room for Options and Future Developments***

Specific room is added to the design to allow modification in the future with features not conceived in the current implementation.

#### ***Memory Channel Protocol separated from Link Technology***

A specific design objective is to support link technology as it develops over the next generation. The design must not be sensitive to the link implementation employed. A working assumption is that three or more different links may be involved in each path from requestor to target memory.

#### ***Does not change at each new Processor***

Memory Channel design is independent of the processor type and is constant over several generations of processor. This is the primary economic benefit to the processor and memory vendors.

#### ***Plug'n'Play***

When a new memory module is discovered by the processor, its availability can be accessed. Discovery methods may be beyond the scope of this project. The EUI makes each memory module unique and prevents miss-addressing of the information.

#### ***Capable of RAIMM Operation (RAID with and without rotation)***

The Memory Channel support Redundant Arrays of Inexpensive Memory Modules to maintain the availability of a specific memory in the event of failure of one component of the array. This is similar to the methods applied to disks in a RAID environment.

#### ***Extensive Data Protection in addition to link level protection***

The header and the data packet are validated separately from the validation of the link layer.

#### ***Data security required***

All data must be capable of being secured as it travels over the links. This security requires authentication, validation, and encryption of the message being transferred.

#### ***Smart Memory Module***

All memory modules are required to provide for their own maintenance.

#### ***Memory Caching Write and Read***

As memory caching will occur, the cache coherency protocol is needed to assure the management of the data.

Approval – This project, with the above Scope and Purpose was approved by the Sponsor, the MSC, on 12 April 2004.