

# Enabling technologies for improved data management: Hardware

Kerstin van Dam-Kleese and Michael Hopewell  
*CLRC – Daresbury Laboratory, Daresbury,  
 Warrington, Cheshire, WA4 4AD, UK*  
*Tel.: +44 1925 603832; Fax: +44 1925 603634;*  
*E-mail: k.kleese@dl.ac.uk, mqhopewell@hotmail.com*

The most valuable assets in every scientific community are the expert work force and the research results/data produced. The last decade has seen new experimental and computational techniques developing at an ever-faster pace, encouraging the production of ever-larger quantities of data in ever-shorter time spans. Concurrently the traditional scientific working environment has changed beyond recognition. Today scientists can use a wide spectrum of experimental, computational and analytical facilities, often widely distributed over the UK and Europe. In this environment new challenges are posed for the Management of Data every day, but are we ready to tackle them? Do we know exactly what the challenges are? Is the right technology available and is it applied where necessary?

This part of enabling technologies investigates current hardware techniques and their functionalities and provides a comparison between various products.

## 1. What are the challenges?

For the majority of scientists the biggest challenges arise through the distribution of their work group and research facilities. The co-ordination of activities like: keeping track of research progress, recording and sampling results, discussing ongoing work, submitting and supervising research work (computational and experimental), exploring and preserving findings are major tasks for all groups, bearing in mind the increased distribution of personal and equipment. Relevant techniques are changing fast and therefore these tasks require time and expertise to be performed effectively and efficiently. With limited research funding they are usually not top of the project priority list and often only the most basic work is done. Unfortunately this has led to large amounts of lost or inaccessible research data. Currently the flood of new data is often overcome

by self-censorship, only the absolute minimum of data is kept leading sometimes to numerous repetitions of experiments and modelling calculations over the years.

At the top end challenges are posed by large-scale experiments, an example can be found in the Particle Physics community. The Large Hadron Collider (LHC) experiment at CERN will start its work in 2005 and annually produce 4 PetaByte of data ( $\sim 11$  TBytes or  $\sim 9$  million floppies *per day*) over a lifetime of 20 years. The data produced needs to be distributed to and analysed by a worldwide community. There are however other facilities which even today could produce similar amounts of data, e.g. modern area detectors in synchrotron or neutron sources with  $1000 \times 1000$  pixel arrays reading out results at millisecond or microsecond time scales. The reason that data is not produced at the afore mentioned rates at national facilities, is the lack of supporting data management infrastructure which has led to strong self-censorship in the community.

In computational science the biggest data producers in the UK are currently Quantum Chromodynamics (QCD), Cosmology and Climate Research, closely followed by Materials Science.

The data requirements of these communities vary significantly:

- Materials Science – small data files ( $\sim 20$  GB), but very many of them, with a limited life span
- QCD – large data sets (up to 4.5 TBytes) with a limited life span, e.g. 10 years
- Climate research – large data sets (up to 2 TBytes), unlimited life span, will still be valuable in 20 years time, if it can still be accessed

A widely distributed user community accesses all the above data. Studies have shown a clear relationship between available computing power and the size of the corresponding archives (Tera Flop systems will produce Terabyte data streams). Table 1 provides estimates from the German Climate Computing Centre (DKRZ) about their future CPU performance and storage requirements [1].

Table 1  
CPU performance/ storage archival size ratio

	2000	2001	2002	2003	2004	2005
CPU performance (Gflops)	10	150	3000	300	600	600
Data archival rate (TByte/year)	30	240	420	420	660	660
Total data Archival (TByte)	120	360	780	1200	1860	2520
Total archive capacity required (TByte) (TByte)	150	500	1000	1600	2400	3300

Based on these findings a number of data centres have calculated their data requirements for the next 5–6 years:

- DKRZ 2 PBytes by 2005 (currently 120 TB)
- UK Met. Office 1 PBytes by 2005 (currently ~ 50 TB)
- ECMWF 10 PBytes by 2006 (currently 35 TB pure private user data + ~ 150 TB public data archive)

Internationally the Bioinformatics community is also developing into a big data producer. Celera Genomics, USA, is analysing the whole human DNA looking for clues to causes of severe diseases. Collecting data from experimental facilities, it processes its sequences on the biggest available workstation cluster currently available (over 600 Compaq Alpha EV6 Processors).

In industry the oil sector and the chemical and pharmaceutical companies are probably presented with the biggest scientific data challenges. In the oil industry large volumes of seismic data, mapping data, bore hole information etc. need to be combined on a timely basis, here speed equals business advantage. In the chemical and pharmaceutical industry large amounts of data are produced by experiments and computational modelling, which need to be screened and searched for the relevant, desired properties. Again speed is a synonym for business advantage.

These are obvious examples, but data is becoming an issue in other fields too. Imagine for the future molecular dynamics simulations with 100 million particles. If one wanted to store the calculated trajectories over a reasonable simulation time, perhaps to search for rare dynamical events, then one could be dealing with data sets of tens of TBytes.

For the future it is CLRC's vision, shared by many, to combine experimental, computational, analytical and visualisation facilities in a seamless and transparent way for the scientists – a Grid. This will allow easy access to many national and international facilities without bureaucratic overheads and including sophisticated techniques for the transfer and storage of data.

*The increasing distribution of research resources has lead to rapidly growing co-ordination and integration*

*requirements for research groups. Within the next 3 – 5 years we will see exponentially growing data repositories, which will quickly reach the size of several PBytes. These data volumes need to be managed for, accessed by and distributed to an international scientific community. They will also need to be linked to other resources. The rapidly growing amount of data will inhibit the use of many traditional techniques and requires the development of new strategies to make the data available for computation and pre- and post-processing facilities, which are themselves, distributed. New initiatives like the 'Grid' offer great opportunities for the future, but also present big challenges for the management of data in this environment.*

## 2. Are the technical capabilities available to meet today's challenges?

Today's data storage hardware and software is capable to manage data holdings and provide and integrate, transparent hierarchical data storage systems (e.g. companies like StorageTek, FileTek, ADIC, Mass Storage) up to multi PetaByte data volumes – at a price. Catalogue systems have been available for a while to provide easy data access and enabling users to assess the content of particular data sets. Most of these systems are custom made. Generally available Network bandwidth is often unsatisfactory due to high usage and unsuitable for high transfer rates, but more and faster, dedicated connections can be hired (see also the new German 2.5 GBit network), although again at a price. For data exploration, numerous data mining tools are available for the commercial sector, though often custom made. A range of advanced visualization tools like Virtual Reality Caves or desk systems have also been developed their price very much depending on size. I/O support for codes with high requirements is available on hardware, system/software and application levels from a limited number of vendors (e.g. Fujitsu, IBM). Again the biggest challenges lies in the integration of the distributed and collaborative working environment that prevails today, tools are beginning to emerge, but their capabilities still need to be tested.

*In the following sections we will examine current hardware technology and its ability to support the requirements of the scientific community. We will be looking at the areas data production, data transfer and data storage. In a second part of this enabling technologies for improved data management report we will discuss software products, technologies and methodologies.*

## 2.1. Data production

Modern supercomputing technology enables us to produce increasingly realistic representations of the scientific processes. Many of today's modelling codes require large input data sets and produce significant amounts of output. The capabilities of components like processors, compilers and scientific libraries have improved dramatically over the past years. They are thereby promoting the faster production of more output in shorter time spans, but disk I/O and data archival and data retrieval mechanisms have not kept pace with this development, leaving a growing gap between the amount of input and output data the codes require and the capabilities of the components that are supposed to feed them.

*“Currently the speed at which data can be moved in and out of secondary and tertiary storage systems is an order of magnitude less than the rate at which data can be processed by the CPU. High Performance computers can operate at speeds exceeding a trillion operations per second, but I/O operations run closer to 10 million bytes per second on state of the art disk.” [2].*

In some applications program I/O has become a real limiting factor to code performance.

*“The data volumes generated by climate models can be very large and are a problem to deal with especially when the models generating this data have been optimised to run quickly and efficiently on HPC platforms.” [3].*

*“Even with multi-terabyte local disk sub-systems and multi-PetaByte archives, I/O can become a bottleneck in high performance computing.” [4].*

Different types of bottleneck can occur for different models. Current models often need large input data sets to be present at program start and during the program run. It is usually time consuming to retrieve these data sets from on-site mass storage facilities or from other data centres and it is not always possible to predict the exact time such retrieval will require (e.g. due to network traffic or heavy use of the facility). In addition

computing centres often charge for the amount of disk space that is used by the scientists, so intermediate storing of large data sets on local disc can be an expensive option. If local copies of the data sets are not available; they have to be retrieved at the start of the program run, expensive processors may be blocked for long periods, waiting for input. Some models require regular, time dependent input which is difficult to sustain on current architectures, similarly leading to a waste of resources.

The execution of coupled models, communicating via files, can be stalled, if the disk I/O capabilities can not keep pace with the model speed. If disk space is limited, output files need to be transferred to alternative storage during the program run, but if these processes do not match the speed of the model or are not available, the execution time can rise tremendously or the program run can even fail. Version control for data sets (gaining importance with the number of scientists involved in a project) and decisions on how and by whom the data will be used later on (affecting data formats) are important research project management tasks. Program I/O needs to be designed to support these decisions. We will be looking at hardware and system support for programs with high I/O requirements, in the future we hope to conduct a more detailed survey on primary storage techniques and fast disks.

### 2.1.1. Hardware I/O support

The main limiting factor in satisfying the high I/O requirements of modelling programs on HPC platforms is the underlying hardware.

*“SRAM/DRAM access times have only dropped by a factor of six within the last 15 years.” [5].*

In contrast to this the speed of high-end processors has increased by a factor of  $\sim 80$  in the same time span (based on the comparison of a Cray 1 and a NEC SX-5 processor). There is a growing gap between the processor capabilities and the actual ability of the disk sub-systems, fifteen years ago  $\sim 10$  instructions were necessary to ‘hide’ one memory access, today  $\sim 360$  are required. For codes with high I/O requirements such as climate modelling codes, this gap represents a serious performance bottleneck.

Today's systems are very unbalanced in their ability to produce I/O and to subsequently handle it. Leading hardware vendors like SGI, IBM and Fujitsu acknowledge the problem, but are not planning to focus significant research effort into this area. The main argument against it is that any major improvements would be too expensive or even physically impossible at the mo-

ment. Instead some of them e.g. Fujitsu are developing software solutions for this problem.

There are however two smaller companies, which have taken different approaches to deliver better-balanced systems: Tera Computing and SRC. Tera Computing is following the philosophy ‘*if you can’t beat it – hide it!*’ using cheap memory chips and hiding the long memory latency times with a thread based programming approach. SRC says that ‘*memory is too stupid*’ so they use cheap off the shelf processors in their systems and concentrate their main efforts on the development of fast memory and network solutions. Future benchmarking efforts have to show how well these approaches work. CLRC – Daresbury Laboratory has set up collaboration with Tera enabling it to test their 8-processor MTA system at the San Diego Super Computing Center. Currently an ocean model and an engineering code are being tested. Additionally Tera has recently bought Cray from SGI and it will be interesting to observe future developments in these product lines.

### 2.1.2. System I/O support

Most HPC systems today are quite unbalanced in their provision of processor speed and I/O supporting devices.

*“Performance bottlenecks: Memory bandwidth, long memory latency, poor cache behaviour and instruction level parallelism.” [6].*

As it can not be expected that a significantly better balanced system will become widely available in the next few years (see section Hardware I/O Support), it is necessary to look for other solutions to bridge the existing gap between processor speed and for example memory access times. In this section techniques and products will be discussed which help to improve the systems I/O performance, but do not require any changes to the applications themselves.

*Fast file systems.* Fujitsu promises to deliver the best-balanced I/O performance on the market with its new VPP500 range systems, which features a flexible and high performance file system (FPFS). The FPFS allows flexible I/O processing according to the I/O request length. Small requests will be transferred via the cache; large ones are transferred directly avoiding the system buffer. In addition parallel access is supported (e.g. multiple real volumes). The FPFS is coupled with the distributed parallel file system (DPFS) which divides I/O requests for externally stored data into multiple requests to PE’s connected to the external storage device (IO-PEs). Taking advantage of the large mem-

ory capacity of the VPP500 (up to 16 GB per processor) the memory resident file system (MRFS) improves access time to local data. Currently Fujitsu seems to be the only provider of MRFS.

IBM has developed the General Parallel File System (GPFS) as one of the key system software components of the ASCI Pacific Blue Supercomputer; it allows high-speed parallel file access. Originally invented for multimedia applications it has been extended to support the requirements of parallel applications. On ASCI Pacific Blue, GPFS enables parallel programs running on up to 488 nodes to read and write individual files at data rates of up to 1.2 GB/s to file systems of up to 15 TBytes. The product is available for RS/6000 SP systems of any size.

## 2.2. Data transfer

Once the data is produced it needs to be transferred into secondary and tertiary storage for further analysis or archival. Depending on where these facilities are various types of network infrastructure will be involved. In general switches, routers, hubs, ATM, FDDI, HiPPi, Ethernet, token ring, virtual LANs, wireless LANs and fibre channel products all fall into this category. However, we will concentrate in this section on high performance LAN technology, as this is the area where each organisation can really influence the network performance. We will however give a brief overview of current and future network capabilities and expected problems for the future.

### 2.2.1. High performance connections

There are various connection technologies available and each has their advantages and disadvantages. We will summarise the basic principles behind the current technology available on the market.

*SCSI.* SCSI (Small Computer Systems Interface) was a major advance in hardware interfaces. Prior to SCSI were interfaces that were not “intelligent”; that is, they were designed for specific peripherals. For example, there was the HDD interface for hard disks, a tape drive interface for tape drives and so on. With SCSI, a standard interface was defined for all hardware peripherals so that only a single adapter was required. Table 2 provides an overview of currently implemented SCSI technologies.

One of the critical advantages that helped SCSI over its rivals was that it had the ability to process multiple overlapped commands. This overlap I/O support feature, also known as multi-tasking, enables SCSI drives

Table 2  
Types of SCSI currently implemented

Data	Bus	Connection	Data rate (MB/s)	Comments
SCSI-2	8-bit (narrow)	50-pin (high density)	10	Used for peripherals like CD-R and Iomega Zip Drive
Wide-SCSI	16-bit (wide)	68-pin (high density)	20	Primarily used for HDD's
Fast SCSI	8-bit (narrow)	68-pin (high density)	10	Used for peripherals like CD-R and Iomega Zip Drive
Ultra SCSI	8-bit (narrow)	50-pin (high density)	20	Used for peripherals like CD-R and Iomega Zip Drive
SCSI-3	16-bit (wide)	68-pin (high density)	40	Primarily used for HDD's
Ultra2 SCSI	16-bit (wide)	68-pin (high density)	80	Used for HDD's
Ultra 160 SCSI		160		

to fully overlap their read/write operations with other drives in the system. This allows different SCSI drives to be processing commands simultaneously rather than serially.

The performance and throughput of SCSI standard have been enhanced over the past 10 years, but the underlying technologies and protocols have changed very little.

*Fibre channel.* FC is a full duplex, serial communications interface that has a maximum speed of 100 MB/s. Compared to SCSI; it uses much simpler cabling and connectors, which significantly reduces the complexity of the configurations. Because addressing is handled automatically, there is no need to set switches and jumpers. Furthermore, the architecture of FC means that there are no termination issues, unlike SCSI. There are three topologies that are implemented by FC and these can be configured from direct point-to-point connection to fabrics. A disk can also be connected using FC with one of the topologies called FC-AL (Fibre Channel Arbitrated Loop). FC can be viewed merely as a transport mechanism for the supported command set. This means that different protocols can be used over the same physical link to satisfy the needs of both channel and network users. Channel users usually adopt a SCSI command set over the FC link whereas network users run IP over a link. FC can transport a number of other channel and network protocols including ATM (Asynchronous Transfer Mode), TCP/IP, FDDI (Fibre Distributed Data Interface), HiPPi (High Performance Parallel Interface) and IPI.

- *Fibre Channel Arbitrated Loop (FC-AL).* Data is passed through a token scheme and can only pass one way, therefore if the loop is broken then all transmissions cease. A higher level of redundancy can be provided if a second loop is implemented.

- *Non-arbitration.* To avoid the need for arbitration, every peripheral must be on a dedicated link with an adjacent device to which it talks half the time. This means that there is no permission required, since one device “talks” the other “listens”. If multiple devices are combined by dedicated links, then a daisy chain results where multiple conversations can happen concurrently.
- *Serial Storage Architecture (SSA).* Serial storage architecture is non-arbitrated and it uses the dedicated links between peripherals to form a daisy chain. It is a high performance serial interface designed to connect disk drives, CD-ROMs, tape drives printer's etc to other peripherals such as personal computers, storage sub-systems and servers. It is a new industry standard that provides high-speed communication, but at a lower cost. It provides a series of point-to-point serial links, each of which can be up to 25 metres with low cost cable. Each node (peripheral) can have 2 connections known as ports. Links can be set up to form strings with up to 128 nodes or with loops with up to 127 nodes and no single point of failure.

*High Performance Parallel Interface (HiPPi).* This is an ANSI standard and is used to physically connect devices at short distances and at high speeds. Serial HiPPi is the fibre-optic version of HiPPi. HiPPi employs pairs of point-to-point simplex links that ensure full bandwidth to each station. The emerging Gigabyte System Network, previously also called HiPPi-6400 or SuperHiPPi, is expected to deliver 200 MB/s sustained. However GSN/TCP will not be able to support transfer rates up to 500 MB/s with only one connection. The new protocol stack Scheduled Transfer/ST might offer the needed capabilities, but will require support from the operating system and applications.

Table 3  
Comparison between connection types

Features	FC	SSA	Ultra 160 SCSI
Node-node Distance	100m	20m	20m
Max. Optical Distance	10,000m	2,450m	NA
Max. Copper Distance	100m	10m	6m
Current Speed	100MB/s	20MB/s	160MB/s
Future Speed	200MB/s	80MB/s	??MB/s
Max. Devices	126 (loop), 16million (sw)	128 (loop) 128 (sw)	15
Peripherals supported	All	RAID only	Limited types
Cost compared to SCSI	Higher but decreasing	Higher but decreasing	–
Serial Connectivity	Yes	Yes	No
Protocols supported	Universal	SSA	SCSI
ANSI standards	Yes	Draft stage	Yes
Dual ported operation	Yes	Yes	No

Note: sw = switches.

*Summary.* Table 3 provides a comprehensive overview between commonly used connection technologies. Currently the Ultra 160 SCSI has the fastest transfer rate, but Fibre Channel will catch up with its new release later this year. However even now the SCSI performance is not all it seems, with command overheads and maintenance of compatibility with older devices experts say that only a maximum performance of 140 MB/s is achievable, depending very much on the application mix. In the past SCSI had a price benefit in the storage area network (SAN) market, but this is expected to vanish over the coming year. Aside from performance and prize issues Fibre Channel offers other advantages, as it removes many physical constraints of SCSI such as distance and addressing.

#### 2.2.2. Network technology

The Internet connections between big cities or countries are determining the upper limit of the achievable network capacity. Real bottlenecks are connections with the US as there is only a very limited amount of bandwidth available. European data centres that require bulk transfers from the US (e.g. 20 GB SLAC data for CLRC) are struggling. At a recent EU workshop on eScience this problem was reiterated and in a general consent experts asked for a Terabit connection between Europe and the US, the subsequently released 4th Call of the IST programme reflects this demand and offers opportunities for developers of such technology. Currently experts see the limitations for such a connection not so much in the network itself as in the routers that need to feed them, and innovative solutions will be required to achieve such a connection in the foreseeable future. Throughout Europe the available bandwidth varies greatly, Germany has just received a 2.5 Gbit/s network connecting all major cities, whereas the UK's

academic network Janet has a capacity of 155 Mbit/s. However further developments are planned and the UK should reach 2.5 Gbit/s by January 2001 and 20 Gbit/s are planned for the end of 2002. Other countries like France, Spain or Portugal are currently trailing behind. Another obstacle that frequently becomes apparent is that there are too many single points of failure, problems with switches or routers can result in network failures for large regions. The metropolitan area networks are of varying capacity, but need and will receive an upgrade in the near future. The slowest link in the chain is often the on-site local area network; universities have over the past years often neglected to upgrade their facilities and are now struggling to cope with the traffic.

For the future it is expected that millions of small, intelligent devices (personal digital assistants [PDA], mobile phones) equipped with spontaneous networking capabilities will enter the market to access information and services available "on the net". First web sites already offer special services for mobile phone users.

#### 2.3. Data storage

In this section we present a review of secondary and tertiary data storage hardware which is currently available, including some older technology which is still widely used. Michael Hopewell, who worked as a placement student at Daresbury Laboratory in summer 1999 and 2000, has carried out a survey of RAID systems. For tertiary storage hardware we will look at three different areas: removable storage media (e.g. tapes), tape drives and archival systems (e.g. robots/silos) as those are the most important ones for a high performance storage system. Finally we will look at relevant standards for data storage systems.

### 2.3.1. Secondary storage

RAID (Redundant Array of Inexpensive (or independent Disks) technology is a collection of drives, which act as a single storage system. RAID technology was developed to let you add capacity to your server without degrading performance. Data files are broken into segments and distributed throughout the array so drive heads can access data segments simultaneously. The other advantage is that RAID provides better techniques to protect data against hardware failure by simultaneously duplicating complete data set or through parity design from which data may be reconstructed. RAID deploys all available network technologies described in the previous section, but which one is the most suitable is determined by the user's requirements.

Due to the development of larger RAID capacities, there are many vendors supplying small to medium capacities, providing a greater variety of systems to the user. Currently, there are not many vendors who supply RAID products that are in the terabyte region, but the explosive growth in LAN's and WAN's will mean that more of the existing vendors will supply RAID products with ever higher capacities. In the near future, it would not be surprising to see RAID systems reaching into the PBytes region. However, this high scale is dependent on the size of the HDD's (Hard Disk Drives) and very soon we will start seeing 76+ GB by IBM. The cost for Terabyte RAID products are unsurprisingly high and it is critical that the user studies the market to make the decisions – choosing a RAID solution that is compatible with existing systems, will meet all the requirements and at the lowest cost possible.

Due to the high-end RAID products being aimed at the more critical-mission applications, there are not many organisations incorporating terabyte RAID products. A few products have already been benchmarked by a few organisations and independent employees of computer related firms. For example, Tomas Pospisek (works for SPIN, an ISP) maintains the following web page which has many links to benchmarks to RAID and RAID-related devices.

See: <http://www.spin.ch/SPIN/tpo/homepage/bench/index.html>

Research is also carried out into various aspects of RAID systems; E.g. University of Versailles has a project to investigate measures to increase the performance of RAID systems and to avoid data loss [7].

For the future there are some interesting discussions on the way. Given the cost of some RAID controllers

and the continually dropping cost of drives, some vendors begin to question is there is a need for hardware based RAID systems at all. Bearing in mind all the intelligence that has already been implemented in SAN's, it is open to discussion if an intelligent switch could be used to control multiple streams to multiple disks.

In the meantime the industry continues its search for new levels of RAID capabilities to open up new opportunities. Table 4 provides a comprehensive overview between currently available RAID systems.

### 2.3.2. Tertiary storage

In this section we will present survey results on removable storage media, tape drives and archival systems.

#### 2.3.2.1. Removable storage media

There are various types of removable storage media available, using different technology and offering different storage capabilities. In general we can group them in four different categories: round tapes, cassettes, cartridges and optical disks. The difference between cartridges (e.g. IBM 3480, DLT) and cassettes (e.g. 4 mm, 8 mm, 19 mm, VHS) is, that cartridges have only one reel whereas cassettes have two, so that they may not need to be rewound to dismount them (time benefit). In addition we look at three different tape formats: longitudinal, serpentine and helical. Information on media lifetimes (longevity) can be found under <http://www.nml.org/MediaStability>.

There is still a large amount of data stored on older tape formats like the IBM 3420 9-track round tapes with a capacity of several hundred Mbytes or the IBM 3480 18-track 200 Mbytes cartridge which was introduced 1984. None of these technologies are used for the day to day business of providing storage capacity for HPC users. Table 5 shows the most commonly used storage media, whereas Table 6 shows recent introductions to the market which are gaining in importance. Table 7 at last shows a range of media technologies which due to come out soon or which are still under development.

Initially the release date for the Sony DTF-2 tape was 1999, but the delay was worse while as the technology has been improved significantly in the meantime, originally a native capacity of only 100 GB was planned. The tape offers a very price competitive option and further developments are planned with DFT-3 and 4 holding 400 and 800 GB native capacity respectively.

The Advanced Intelligent Tape (AIT) technology provides a Memory in Cassette (MIC) option. The MIC system consists of a 128 KBit for AIT2 memory

Table 5  
Most commonly used storage media

Name	IBM 3490E	IBM 3590	SD-3	DLT 7000	Exabyte Mammoth	9840	Sony AIT1
Form factor	1/2" cartridge	1/2" cartridge	19 mm cartridge	1/2" cartridge	8 mm cartridge	19 mm cartridge	8 mm cartridge
Capacity	0.8 GB native	10 GB native	10/25/50 GB native	35 GB native	20 GB native	20 GB native	25 GB native
Recording format	serpentine	serpentine	helical	serpentine	helical	helical	helical
Comments	Intr. 1992	Intr. 1994	StorageTek	Quantum		StorageTek	

chip built into the data cartridge, which holds the tape's system log and other useful information. Unlike other conventional tape types which have to be rewound to the start of the tape to read the system log and locate the desired file, the MIC system provides that information immediately. Advancing directly to the correct position on the tape can then access the file. Thus the AIT cartridge has largely reduced access times for read and write operations and meantime between access.

Super DLT from Quantum was first announced in summer 1998 with an expected shipping date of summer 1999, but at the time of publication it was not available yet.

The 'Keele ultra high density memory system' was developed by Prof. Ted Williams, formerly Keele University. Currently in its patenting phase this magneto optical system claims to be able to store 9.1 TBytes on a credit card sized device, with an expected price of 50\$ for one unit. As it also promises fast access rates this device could revolutionise the storage market.

Important for the choice of tape is not only the anticipated total amount of stored data, but also average file size and access time requirements. The common experience for smaller file sizes (3GB and less) is that the time to locate a file becomes just too long on the big tapes (e.g. D2 – 165 GB native), whereas these tapes are very efficient for really big data files (e.g. image files or animation sequences).

### 2.3.2.2. Tape drives

As important as the right choice of medium is the right choice of tape drive. The tape drive will determine the speed of data transfer from and to the medium, positioning time and the amount of data that can be stored (data compression). Important attributes for the choice of a tape drive are:

Specific capacity / data rate / position speed / rewind speed / load time / unload time / error rate / number of passes for medium / shelf time / MTBF for drive / costs.

Though not all of the new drives are suitable for High Performance Mass Storage Systems, this general development has also helped the HPSS market and today there is a much wider choice of tape and tape

drive available then 5 years ago. Currently tape drives for HPSS applications are produced by Exabyte, Sony, StorageTek, Quantum and IBM. They vary not only in price, but differ in the factors mentioned above and also on reliability and warranty issues.

*Tape technology.* The tape technology type supported by the tape drive not only determines which tapes can be read, but also which size of media can be read (this can be useful for compatible media from other vendors), which recording format is supported and how many tracks can be written at the same time (if applicable). Commonly used are media with form factors of 1/2", 8 mm (from the video industry) and 19 mm (from the data instrumentation industry). There are three recording formats longitudinal (barely used any longer), serpentine and helical (defined above). If the recording format is longitudinal or serpentine it is important to know how many tracks can be written onto the medium in total (capacity) and how many of them can be written in parallel (speed).

*Capacity.* The tape drive capacity is measured by the maximum amount of data that can be recorded on a single tape cartridge or cassette. Tape drive manufacturers increase the capacity of their drives by increasing the track density (number of tracks) on a given section of tape or by increasing the physical length of the tape in the cartridge or cassette. Built-in data compression is also used to increase the capacity of the tape drives. Therefore every vendor has to identify whether his numbers were achieved by 'native' or 'compressed' data recording. The data compression methods used are called Improved Data Recording Capability (IDRC) and (2) variants of the general Lempel-Ziv (LZ) compression algorithm LZ, LZ1, DLZ (Digital Lempel-Ziv) or ALDC (Adaptive Loss less Data Compression).

*Data transfer rates.* The data transfer rate defines the number of megabytes per second, which can be read from or written to the media. Again a native and a compressed capacity are usually given. The developers of the new LTO technology promise a transfer rate of up to 1TB per hour (~ 280 MB/s).

*Search speed.* The search speed, also known as Mean Time to Data Access (MTDA) is determined by adding the load time plus the average file access time.



Table 6  
Upcoming storage media technologies, recently introduced or entering the High-End market

Name	Sony AIT2	DLT8000	IBM Ultrium
Form factor	8 mm cartridge	1/2" cartridge	1/2" cartridge
Capacity	50 GB native	40 GB native	100 GB native
Recording format	helical	serpentine	serpentine
Comments		Quantum	LTO standard compliant

Table 7  
Storage Media Technologies under development

Name	DTF-2	Sony AIT3	Super DLT	???	KUHDMS*
Form factor	cartridge	8 mm cartridge	1/2" cartridge	1/2" one reel cartridge	Credit card sized
Capacity	200 GB native	100 GB native	500 GB native	60 GB native	9.1 TB
Recording format		helical	serpentine		optical
Comments	Expected May 2000	Expected 2001	Quantum, was expected mid 1999	StorageTek Expected 2H00	Still in patenting phase

\*KUHDMS – Keele ultra high density memory system.

Table 8  
Currently available tape drive technology

Name	Media size	Recording format	Number of tracks	Capacity	Data transfer rate	MTDA
				Native / Compressed	Native / compressed	
STK 9490E	1/2"	Serpentine	36	0.8 GB (IDRC)	6 MB/s	45 s
STK Redwood	1/2"	Helical		50 GB / (IDRC)	11 MB/s	200 s
STK 9840	1/2"	Helical	20 GB / 80 GB	(LZ1)	5 MB/s   20 MB/s	
Quantum DLT 7000	1/2"	Serpentine	208	35 GB / 70 GB (DLZ)	5 MB/s   10 MB/s	120 s
Quantum DLT 8000	1/2"			40 GB / 80 GB (DLZ)	6 MB/s   12 MB/s	
IBM Magstar (3590)	1/2"	Serpentine	128 (16 at a time)	10 GB / 30 GB (LZI)	9 MB/s   27 MB/s	60 s
Exabyte Mammoth	8 mm	Helical		20 GB / 40 GB (LZ)	3MB/s   6 MB/s	100 s
DST Quad	19 mm	Helical	50GB / 100GB	20 MB/s   40 MB/s	43 s	
Density	4.3 × 6.8 –		150GB / 300GB			
Tape drive	8.1 × 14.4 in.		330GB / 660GB			
Sony AIT1	8 mm 3.5 in	Helical	25 GB / 50 GB	3MB/s   6 MB/s	37.5 s	
Sony AIT2	8 mm 3.5 in	Helical	50 GB / 100 GB	6MB/s   12 MB/s		
SonyGY-8240 for DTF			200 GB / 518 GB	24MB/s   40 MB/s	(ALDC)	
IBM 3580 Ultrium Tape Drive		Ultrium		100 GB / 200 GB	15MB/s   30 MB/s	

Note that not all tape drives can be combined with all archival systems, so the choice might be limited.

Table 8 provides an overview of currently available tape drive technologies. Both IBM and StorageTek are to announce new tape drive technologies in ~ September 2000. IBM's 3590E is announced to have a capacity of 20 GB/40 GB (latter compressed) and a transfer rate of 14 MB/s native and 42 MB/s compressed. It will be possible buy it with SCSI connection or via adapter with FCAL. For 2001 they are planning to develop smaller sized drives (3590F) to allow installing more drives per library. Additionally they are planning to increase their transfer rate to 20–25 MB/s native and 30–50 MB/s compressed. There is a clear roadmap for higher transfer rates and capacity as well as a change

in media technology past this stage, but no expected release dates are known to date.

StorageTek plans the release a number of upgrades to its highly successful 9840 drive technology over the coming 18 month. Starting with the 9840A at the end of this year the transfer rate will gradually increase to 20 MB/s native by 2001. However we can also expect the release of a new high capacity tape drive in September with higher transfer rates and significantly increase capacity.

Although tape drives are an integral part of SAN's they relied until recently on bridges or routers to connect to Fibre Channel networks, however now both

Exabyte and StorageTek have announced native interfaces.

### 2.3.2.3. Archival systems

Today tape robot/silo systems are the most commonly used hardware for larger data archives, making human interference more or less obsolete and granting quick access to removable media 24 hours a day. For the high-end market (into the PetaByte range) there are only two major traditional providers: StorageTek and ADIC. Both companies allow mixing different types of medium / tape drives within one silo enabling the user to choose a media combination best suitable for his applications. Overall it can be said that the ADIC systems offer more flexibility in the choice of tape drives than the StorageTek solutions, but StorageTek can provide larger storage capacity as more units can be connected together. A new trend promoted by Mass Storage is the connection of large number of smaller tape libraries to form PB archives using AIT tape technology.

*StorageTek.* StorageTek is currently the market leader and most big computing or data centres will have StorageTek technology. They have been a provider for many years and although their tape drive technology has changed the silos have remained more or less the same for 10 years or more. Their current product is:

- Powderhorn 9310 (which used to be the ACS 4400). The silo can hold up to 6000 1/2" type tapes. It is possible to connect together up to 256 of these silos. Currently it is possible to connect StorageTek's own tape drives to the system, which can handle 3480, 3490 (9490), Redwood/D3 (SD-3) and the new 9840 tapes. It is also supposed to be possible to install IBM Magstar drives, but the common experience is that this can be very difficult and that it is better to stay with StorageTek's own product range.

Table 9 shows the maximal storage capacities, which can be reached with StorageTek solutions. In recent years StorageTek has been involved in many large-scale storage projects. It is also involved in the Accelerated Strategic Computing Initiative (ASCI) project related HPSS initiative. StorageTek hardware is used for example to provide storage capacity for the ASCI pacific blue system (IBM).

*ADIC.* The Advanced Digital Information Corporation only recently bought EMASS Inc., which some year's back bought Grau Automationsysteme. ADIC has the smaller market share of the two competitors, but

Table 9  
StorageTek archival capacities for different storage media

Tape Type	Capacity per Silo	Maximum Capacity of full system
9490	0.0048 PByte	1.2 Pbyte
Redwood	0.3 PByte	77 Pbyte
9840	0.48 PByte	123 PByte

Table 10  
Mass Storage Solutions archival capacity

Tape Type	Capacity per Library	Maximum Capacity of full system
AIT2 / Qualstar	18 TB	6.5 Pbyte
DTF-2 / Sony	9.8 TB	11.2 Pbyte

offers a good product range. Their main advantage is that they support tape drives from almost every vendor, giving the owner more flexibility in his media choice. They also support the mixing of different media types in one archival system. This not only safeguards investments, but also allows addressing of a large variety of storage requirements very effectively.

Currently available products in the high-end market are:

- AML/E which can store from 500 to 10000 cartridges (depending on the media size). Up to 50 tape drives can be attached to this system. Supported media types are 3480/3490, 3590, DLT, VHS, 8 mm, 5 1/4" Optical Disk, D-2S, D-2M and AIT.
- AML/2 which holds 5000 to 47000 cartridges (depending on media size). Up to 256 tape drives can be attached to this system. The AML/2 consists of ADIC Quadro Tower modules, each of these modules can hold up to 6000 tapes (depending on media size) and 8 of these towers can be connected together. Supported media types are 3480, 3590, DLT, VHS, 8mm, 5 1/4" Optical Disk, D-2S, D-2M and AIT. Based on AIT-2 tapes this technology can give you a peak of 7 PBytes of native data.

Currently ADIC are developing a new library type which will allow to overcome the limit of only 8 connectable units and offer faster tape access. The ECMWF is using an ADIC library with AIT2 tapes for their backup system.

*Mass Storage.* Mass Storage is offering two possible solutions in the PBytes range: Clusters of Qualstar AIT tape libraries and Clusters of Sony DMS-8400 modules with DTF2 tape drives.

The Qualstar AIT libraries come in different sizes holding between 10 and 360 AIT cartridges, that allows to store up to 18 TB native per library (AIT2). With the

support of the ‘Virtual Library Support’ it is possible to connect and manage up to 360 libraries. Current customers are Microsoft Hotmail (90+ libraries) Australian Olympics (240 libraries) and KODAK (650 TB). Due to the small media size and the possibility to stack the libraries, this is a very space saving alternative.

The new Sony Peta Side DMS solution with DTF-2 drives and tapes is the most cost effective one on the market, the libraries come in sizes of 0.6, 4.8, 6.2, 7.5 and 9.8 TB and can be clustered.

Table 10 shows the archival capacities, which can be reached with both the AIT2 and the DTF-2 storage solution. Furthermore both solutions are cheaper than their big competitors ( $\sim$  factor 3 on list price). For fast expanding data centres this is an ideal solution as new storage capacity can be brought in when required without major capital investments.

*IBM.* Approximately 15 years ago IBM decided that the high end storage market is no longer profitable enough and left the field to StorageTek, which have been the market leaders ever since. However in recent years IBM has attempted to re-enter the market, so far with limited success. Although offering a high quality product range for the medium range market, their products do not have the capacity yet to compete with StorageTek, ADIC or Mass Storage. Their newest library product, the 3584 UltraScalable Tape Library, holds up to 72 3580 Ultrium tape drives and offers a maximum capacity of 497 TB if the maximum of 6 library frames is connected. For the future they will very much depend on their tape drive and tape developments to gain the capacity that their competitors already offer, and although there are interesting plans for the foreseeable future no fixed release schedule have been announced.

### 2.3.3. Standards

We can divide the available standards into three categories: specific products, interfaces and general reference models.

#### 2.3.3.1. Standards for specific products and interfaces

A wide variety of old and new standards for various components of data management and storage systems can be found at the sites of the standardising organisations. The most important organisations are:

- WSSN – World Standards Services Network – [//www.wssn.net/WSSN/](http://www.wssn.net/WSSN/)
- ISO – Work group on Information Technology – [//www.iso.ch](http://www.iso.ch)
- IEEE – the organisation for electrical engineering including IT – [//computer.org](http://computer.org)

- SNIA – Storage Networking Industry Association (member list like who is who in storage) – [//www.snia.org](http://www.snia.org)
- LTO – High Performance Tape Standards (IBM, SONY, HP, EXABYTE etc) – [//www.Lto-technology.com](http://www.Lto-technology.com)
- OII – Standards and Specifications collected by the European Union – [//158.169.50.95:10080/oii/en/oii-home.html](http://158.169.50.95:10080/oii/en/oii-home.html)

A few new standards worth mentioning are:

- *The Linear Tape Open (LTO).* Technology was developed by HP, IBM and Seagate and is available for licensing to other tape producers. It aims to provide a quasi standard for tape technology to ensure compatibility between products. There are two main technologies Accellis for fast access and Ultrium for high capacity.
- *ISO/IEC 8802-11: 1999.* Telecommunication and information exchange between systems, local and metropolitan area networks incl. Wireless LANs.
- *ISO/IEC 16382: 2000.* Data interchange on 12.7 mm 208 track magnetic tape cartridges – DLT-6 format.
- *ISO/IEC 17913: 2000.* 12.7 mm 128 track magnetic tape cartridge for information interchange – parallel serpentine format
- *IEEE.* announced the 1st storage system standard, approved in June 2000. It incorporates so far 4 separate standards for the *IEEE Media Management System (MMS):* architecture, data model, core media management protocol and protocols for drive and library management. More standards are expected to follow soon.

#### 2.3.3.2 General reference models

There are generally two standards, which are of interest for the provision of data storage. One is the IEEE Reference Model for Open Storage Systems Interconnections (OSSSI), previously known as the IEEE Mass Storage Reference Model, and the ISO standard for a Reference Model for Open Archival Information Systems, which is still under development. Whereas the IEEE standard is mainly concerned with architecture, interface and terminology specifications and standards, the ISO standard focuses more on necessary operational issues and interactions between different parts of data archives. In this respect they can be seen as complementary.

*IEEE’s OSSSI.* As it started out as the IEEE Mass Storage Reference Model in the 80’s, this standard is very

much focused on technical details of mass storage systems, from storage media over drive technology to data management software. In the last few years more general, organisational functions have been added and the description of connections and interactions with other storage systems has gained in importance. Nowadays the Reference Model for Open Storage Systems Interconnection (OSSSI Model) provides a framework for the co-ordination of standards development for storage system interconnection and a common perspective for existing standards. The descriptions used are independent of existing technologies and applications and are therefore flexible enough to accommodate advanced technologies and the expansion of user demands.

*“It is not the intent of the OSSSI Model to serve as an implementation specification, to be a basis for appraising the conformance of actual implementations, or to define precisely the standards for services and protocols of the interconnection architecture. Rather, the OSSSI Model provides a conceptual and functional framework that allows teams of experts to work productively and independently on the development of standards for OSSSI.”*

IEEE 1994

The OSSSI Model is concerned with open storage systems and the environment that a storage system is required to work in. The model defines issues like:

- *Storage Objects* – e.g. Device, Cartridge, Physical Volume, Store
- *Storage Services* – e.g. Mover, Physical Volume Repository, Physical Volume Library, Virtual Storage Service
- *Storage System Environment* – e.g. Storage System Management Framework, Security Framework, Communication Services, Location services, Name Services

Many standards have been developed over the years within the framework of this reference model and its predecessor. The standards developed in this area are published under IEEE P1244 and can be found on the IEEE Web site or in their publications.

*ISO’s OAIS.* This new ISO standard has not yet been published and all following comments are based on a draft version of the standard from September 1998. CLRC’s Space Science Department is actively involved in the development of this standard and for up to date information David Giaretta should be contacted (d.giaretta@rl.ac.uk).

The Open Archival Information System (OAIS) aims to provide a framework for the operation of long term archives which serve a well specified community. Issues like data submission, data storage and data dissemination are discussed. Every function is seen in its entirety not only describing technical details but also human interventions and roles. For the purpose of this standard it has been decided that the information that is maintained needs long-term preservation, even if the OAIS itself will not exist through the whole time span.

*“The reference model addresses a full range of archival information preservation functions including ingest, archival storage, data management, access and dissemination. It also addresses the migration of digital information to new media and forms, the data models used to represent the information, the role of software in information preservation, and the exchange of digital information among archives. It identifies both internal and external interfaces to archive functions, and it identifies a number of high level services at these interfaces.”*

CCSDS, 1998

The OAIS concept is not only concerned about defining functionality within the archival system itself, but also about its interaction with other archives as well as with data producers, data consumers and archive facility managers.

### 3. Conclusions

In many scientific domains experimental devices and simulation programs generate large volumes of data. This data is highly valuable and should be made available to collaborating research groups and a wider scientific community at some point during its live cycle. The volumes of data can reach hundreds of TBytes or PBytes and therefore it is impractical if not impossible to store all of it on fast disk systems. Rather a sophisticated hierarchical storage management system (HSM) is required to deliver the necessary capacity for the lowest possible price. The HSM includes next to software products a range of hardware components, such as disk, RAID, tape and libraries as well as technology to support the high I/O performance of various devices, in particular high end computing systems. Network components also play an important role. We have discussed a wide range of these products and techniques in our

report and in Section 2.3.3 information can be found on relevant standardisation bodies and standards itself.

As demonstrated in the report a wide range of possible products is available in this increasingly fast changing market. The obvious question is what would be the ideal configuration? The answer is, there is no such thing as an ideal solution which will work well or even satisfactory for everyone. A detailed analysis of existing hardware and software products, applications, data volumes, data access volume and pattern, number of users, life span of archived data, typical file sizes etc. will have to be carried to make any valid suggestions. You also need to bear in mind that quite a number of products from different vendors do not work particularly well together (although sometimes stated differently by the vendors).

The German Climate Computing Centre (DKRZ) has established a range of 'rules of thumb' for its Centre to express the requirements for a climate modelling archive [8], which might also be useful for others:

- Basic data intensity expressed, as annual long term archival is 1Kbyte/1Flop (sustained) CPU performance.
- Long term storage capacity, expressed as a factor of 1.5 between model generation (and storage) and long term archival. Approximately 2/3 of the produced data is archived.
- Data access profile is characterised by a factor of 3 of the model data access versus model data generation. 3/4 of mass storage activity is reading, 1/4 writing.

Experience showed that the disk cache size should be at least 5–6% of the annual archive rate (the ECMWF has 4%).

A further consideration for the design of a suitable data management environment should be security (if you can afford it), avoiding single points of failure and choose easily exchangeable units (e.g. hot swap of disks, power supplies or drives). If you have data that is unique, irreplaceable or of long term value you might also want to consider a back up archive on a different site or building (e.g. ECMWF).

Although the hardware is the backbone of a good data management system it is not sufficient on its own to fulfil the requirements posed by today's large data volumes, intelligent software techniques and packages are required to complement the solution. In the second part of this report we will focus on software solutions for the management and exploration of the data.

Intelligent data management is an important issue, but fast changing technology makes it difficult to fol-

low all the relevant information. We also found that the information available is widely scattered and time consuming to collect. In an attempt to provide easier and up-to-date access to relevant hardware and software techniques, relevant projects and publications we will launch a 'Virtual Data Technology Centre' at the end of 2000. Please refer to <http://www.dl.ac.uk/TCSC/datamanagement/VDTC/index.html> for more information.

#### 4. Acronyms

CLRC – Central Laboratory of the Research Councils in the UK

LHC – Large Hadron Collider experiment at CERN

QCD – Quantum Chromodynamics

DKRZ - German Climate Computing Centre in Hamburg

ECMWF – European Centre for Medium Range Weather Forecast in Reading UK

HPC – High Performance Computing

LAN – Local Area Network

WAN – Wide Area Network

SCSI – Small Computer Systems Interface

FC – Fibre Channel

PDA – Personal Digital Assistant

HDD – Hard Disk Drives

RAID – Redundant Array of Inexpensive/Independent Disks

#### Acknowledgements

We would like to thank all vendors who kindly collaborated with us and provided us with essential information about current and future products.

#### References

- [1] H. Fichtel, Problem Areas and Potential Solutions for High Performance Data Management Systems, Data Management 2000, Conference proceedings of the 1, in: *International Workshop on Advanced Data Storage/Management Techniques for High Performance Computing*, K. Kleese and R. Allan, eds, CLRC – Daresbury Laboratory, UK, May 2000.
- [2] National Computer Science Alliance, Web Pages, 1998
- [3] A. O'Neill and L. Steenman-Clark, Modelling Climate Variability on HPC Platforms, in: *High Performance Computing*, R.J. Allan, M.F. Guest, A.D. Simpson, D.S. Henty and D.A. Nicole, eds, Plenum Publishing Company Ltd., London, 1998.
- [4] J. Jenness, Web Pages, Lawrence Livermore National Laboratory, ASCI project, 1998.

- [5] G. Astfalk, Hewlett Packard, Workshop presentation at the eight ECMWF workshop on the Use of Parallel Processors in Meteorology, Reading 1998.
- [6] D. Snelling, Fujitsu European Centre for Information Technology (FECIT), Workshop presentation at the eight ECMWF workshop on the Use of Parallel Processors in Meteorology, Reading 1998.
- [7] S. Zertal, Stripe Restriction mechanism to Avoid the Degraded Mode on RAID5, Data Management 2000, Conference proceedings of the 1, in: *International Workshop on Advanced Data Storage/Management Techniques for High Performance Computing*, K. Kleese and R. Allan, eds, CLRC – Daresbury Laboratory, UK, May 2000.
- [8] H. Fichtel and M. Lautenschlager, High Performance Storage Management Issues, Technical Watch Report of DIO and DSM Authors: M. Lautenschlager and S. Musgrave, Colchester, UK, 1999.
- [9] BCC Technology, Tape Drive Technology Comparison of: Tandberg Data MLR, Exabyte Mammoth, Sony AIT, Magstar MP (3570) and Magstar (3590). White Paper, BCC Technologies, Irvine California, USA, 1998.
- [10] M.I. Beare, The Southampton – East Anglia (SEA) Model: A General Purpose Parallel Ocean Circulation Model, in: *High Performance Computing*, R.J. Allan, M.F. Guest, A.D. Simpson, D.S. Henty and D.A. Nicole, eds, Plenum Publishing Company Ltd., London, 1998.
- [11] H. Biggar, Fibre Channel comes to tape, *InfoStor* 4(4) April 2000.
- [12] P. Churchill, Centre for Earth Observation (CEO), in: *The Globe on the Net*, (Issue 25), 1995.
- [13] D. Dixon, ECMWF weathers enormous growth with help from IBM, ADSTAR Distributed Storage Manager customer experience, WWW, 1998.
- [14] Environmental Change Network(ECN), Baseline data and information from long-term monitoring sites in the UK, <http://mwnta.nmw.ac.uk/eddemo/>, 1997.
- [15] J. Faundeen, *Digital Archives for EO: An American Perspective*, US Geological Survey, EROS Data Centre, 1998.
- [16] J. Haystead, Raid cost/performance trade-offs, *InfoStor* 4(7), July 2000.
- [17] J. Gray, Evolution of Data Management, Computer, IEEE, October 1996, pp. 38–46.
- [18] S.R. Hall, The STAR File: A New Format for Electronic Data Transfer and Archiving, *J. Chem. Inf. Comput. Sci.* 31 (1991), 326–333.
- [19] High Performance Strategy Committee, Discussion Document on the Strategy for High-End Computing for the UK Research Councils, EPSRC, September 1999.
- [20] K. Kleese, Data Management in Climate Research, *ERCIM News* 34, July 1998.
- [21] K. Kleese, Requirements for a Data Management Infrastructure to support UK High-End Computing, Technical Report DL-TR-99-004, Central Laboratory of the Research Councils, Daresbury, UK, November 1999.
- [22] A. Linn, The World Data Centre System, <http://www.ngdc.noaa.gov/wdc/wdcmain.html#wdc>.
- [23] Looking Forward, The NERC Strategy for Science, <http://www.nerc.ac.uk/science/strategy/>, May 1998.
- [24] D.J. Webb, Running a global ocean Model on the T3D, Proceedings of the Simulation of Fluid Flow on Parallel Machines workshop, Imperial College, London, June 1998.







**Hindawi**

Submit your manuscripts at  
<http://www.hindawi.com>

