DESIGN OF SPACE STATION ERA MASS STORAGE SYSTEMS

Ward P. Homer, James R. Chesney
Data Systems Technology Division, Code 520
Mission Operations and Data Systems Directorate
NASA, Goddard Space Flight Center
Greenbelt, Maryland 20771

ABSTRACT

Mass storage systems used in the Space Station era must be able to store, retrieve, process, and distribute data, in near real time, at rates up to 300 Mbps and with fast access storage greater than one terabit. To meet NASA's required speed, capacity, flexibility, and reliability at a reasonable cost, these systems will need to employ modem techniques of data presorting, disk stripping, and parallel redundancy. These systems will, also, have to include special, NASA specific subsystems which pre-process telemetry data in order to reduce direct CPU loading and overall system complexity.

Commercial storage systems and components available today do provide the fundamental elements required for the development of such a high performance mass storage system. In particular, new commercial parallel drive array systems and parallel drive controllers provide an opportunity to develop and prototype architectures which are suitable and cost effective for NASA's applications. In addition, by utilizing experienced already gained in the use and application of VLSI technology, various required NASA specific functions can be integrated with these commercial storage components to develop an intelligent mass storage system prototype. This paper describes the architecture, components, and technical approach for such a mass storage system prototype.

1. INTRODUCTION

Space Station era ground data processing will require an enormous increase in NASA's present capability to process, store and distribute spacecraft data. The extent of this problem is exemplified by NASA's projected average daily spacecraft telemetry volume of 2400 gigabytes expected by mid 1990's, as compared to the 1985 volume of 1 gigabyte per day (1). To accommodate and process this volume of data, new system architectures and storage subsystems are needed which can support this growth by accommodating projected improvements in storage component technology. Various studies have verified the potential mass storage problems awaiting the space communications network in the Space Station era and have suggested actions needed to arrive at approaches to these problems. Beyond requirement changes, these approaches depend on early experience with the advanced storage technology and components. This includes the development of new architectures and custom subsystems suited to the high performance telemetry data systems of the Space Station era. The Data Systems Technology Division at Goddard Space Flight Center (GSFC) has the responsibility to explore, develop, and validate technologies applicable to NASA's specific data processing and communications problems. As part of this effort, the Data Systems Section has undertaken the investigation, design, development, and test of a prototype Space Station era Mass Storage System.

While initial efforts indicate that a fast access read/write disk buffer (300 Mbps, 1,000 Gbit store) suitable for working storage and rate conversion applications (e.g. Level Zero Processing) is possible with current technology and at a reasonable cost, the same situation does not seem clear with respect to very large store magnetic or optical tape systems for data capture applications. A selection of tape storage systems seem sufficiently mature to satisfy line outage protection in the 1992-1998 time frame but they are generally quite expensive. Automated tape libraries and jukeboxes seem best suited to data protection and deferred delivery needs but they are also expensive and are either new or soon to be released commercial products. In any case, these systems will require fast access staging buffers to efficiently interface to the other processing levels. As a result, the Mass Storage System described in this paper is intended to prototype a potential architecture utilizing commercial parallel drive arrays and custom controllers and interfaces, primarily, to meet NASA's projected 1992 ground data processing requirements for a fast access working storage buffer. Tape storage is included in the prototype in order to gain more experience with high performance VCR digital tape technology and to insure compatibility with this technology in the basic prototype architecture.

The architecture selected was tailored to provide both a high probability of meeting storage and buffering requirements and for direct development with commercially available components. Other key criteria includes cost, reliability, use of commercial standards for interfaces/data structures, size, and facility requirements. In order to minimize prototype development time and to optimize compatibility with on-going VLSI based telemetry data systems projects within the Data Systems Technology Division at GSFC, the prototype design of this system is based on the VLSI generic telemetry data systems developed within this division. As a result, direct testing and validation of this system with such projects as the Customer Data Operation Service (CDOS) testbed and the First Element Launch FEL) / Data Interface Facility (DIF) prototype is possible (2). Various NASA specific functions such as those required for the Level Zero Processing (LZP) function (e.g. source packet ID sorting, data reversal, overlap deletion etc.) were included in the design and development of the prototype Mass Storage System's architecture.

Additionally, this system will take advantage of such special projects as the Packet Processor card development at GSFC.

2. PROTOTYPE MASS STORAGE SYSTEM

2.1 SYSTEM REQUIREMENTS

The prototype mass storage system is being designed around a single generic architecture that can be configured to satisfy a variety of NASA data storage requirements. The primary data storage needs call for such functions as line outage recording, rate buffering, data archiving, and level zero type processing. The types of data storage functions that are needed for tomorrow's systems will vary; however, a number of the primary data storage needs can be satisfied by a system that can be configured to provide three basic types of storage:

1) Block storage, where data is stored in fixed size blocks and accessed by a unique block ID number.

2) File storage, where data is stored into variable size files and accessed by file name.

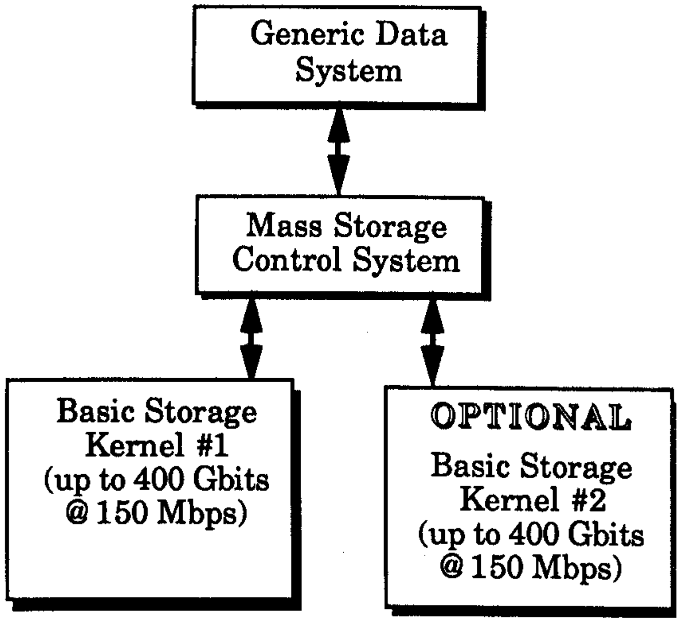
3) Data driven storage, where data is stored and processed according to its type and accessed by its sequence within this type.

The prototype mass storage system hardware will be configurable to support all three modes of operation at up to 300 Mbps.

2.2 GENERAL ARCHITECTURE

The system architecture consists of two distinct components: a configurable high level *mass storage control system* and *basic storage kernel* (see FIGURE 1 - HIGH LEVEL BLOCK DIAGRAM OF THE MASS STORAGE SYSTEM). There will be only one mass storage control system, while, there could be multiple basic storage kernels.

The mass storage control system provides an interface between a generic telemetry data system element (e.g. FEL, DIF, or packet processor) and one or more basic storage kernels, but it will contain no commercial mass storage devices itself. It will act as the master data flow director and control the data flow in and out of the basic storage kernels. It win be used to direct the activity of the basic storage kernels and to perform overall database management functions as required by the system. Additionally, if more than one basic storage kernel is used, it will be responsible for stripping the data across them.



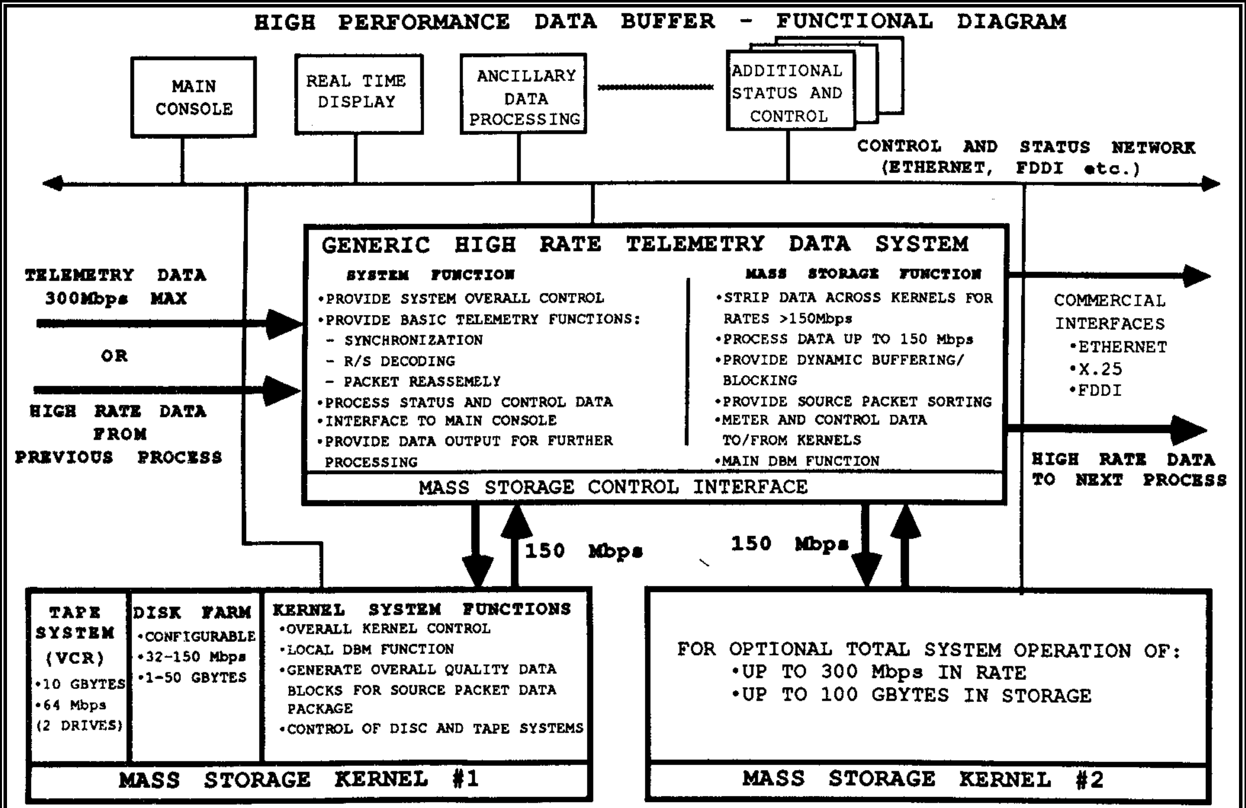
HIGH LEVEL BLOCK DIAGRAM OF THE MASS STORAGE SYSTEM

FIGURE 1

The basic storage kernel acts as a "data sponge" and contains all of the system's commercial storage devices. It will consist of a configurable arrangement of multiple high performance commercial parallel drive arrays and multiple VCR tape units for data storage. Also, it will include high rate input and output subsystems and local intelligence to perform kernel level data base management functions as required. It will directly control the data flow in and out of the storage devices and provide overall quality blocks for this data.

Functionally, the basic storage kernel may be viewed as a tightly coupled collection of high performance commercial parallel drive arrays, very large data storage VCR tape units for short term storage, and specialized single board computer data processors. These units are all bound together by a single centralized data bus (VME bus) that controls multiple storage device buses (SCSI bus).

The control, status, and processing functions required for the system operation will be optimally distributed across system elements and may be implemented in hardware, software, or some combination of both (see FIGURE 2 - TOP LEVEL FUNCTIONAL DIAGRAM). Exact distribution and implementation will be defined during the design concept stage of development which will include compatibility and integration with on-going VLSI based data capture systems work at GSFC. Some of these functions include: data preprocessing (e.g. source data sorting and time ordering); management and



TOP LEVEL FUNCTIONAL DIAGRAM

FIGURE 2

distribution of high rate streams over multiple disk controllers; data capture, backup, and recovery; generation of user data and quality packages; rate translation; and overall system control and status.

2.3 SYSTEM COMPONENTS

2.3.1 THE MASS STORAGE CONTROL SYSTEM

The mass storage control system will be configured to meet the particular system requirements. For lower cost, lower rate telemetry data systems (under 150 Mbps), it will be a single compatible plug-in card that directly interfaces with the generic telemetry data system. This configuration controls only one basic storage kernel. For high performance telemetry data systems (up to 300 Mbps), it will be a separate card cage with a custom interface to the generic telemetry data system. This configuration controls two basic storage kernels in order to achieve the necessary data rate and capacity.

The mass storage control system will control the data flow in and out of the basic storage kernels. It will provide data blocking and quality information accumulation on those blocks. It will provide rate smoothing to ensure a near constant 150 Mbps data movement in and out of the kernel.

For lower rate systems (single kernel systems, under 150 Mbps) that need only simple block storage or file access storage, the mass storage control system can be fairly simple. It will be used primarily as an interface between the generic telemetry data system and the basic storage kernel. The basic storage kernel will perform most of the data base management functions via high level commands from the mass storage control system.

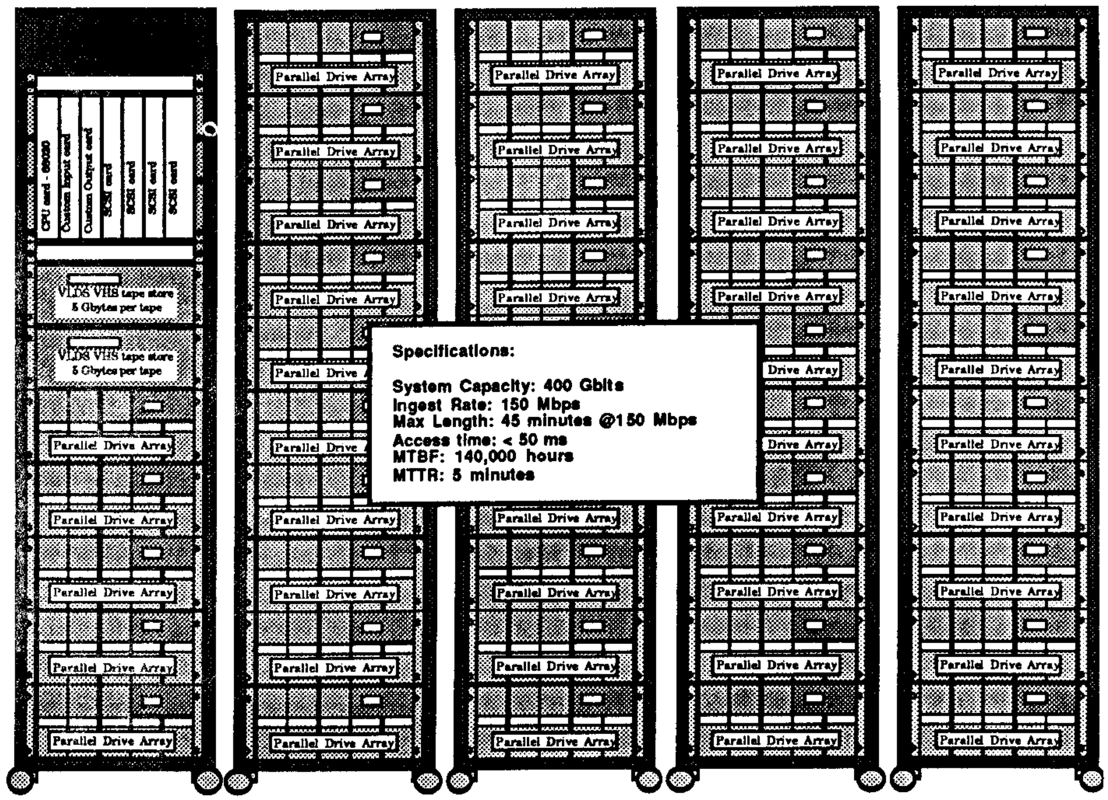
For systems that need data driven storage (e.g. level zero processing systems), the mass storage control system will be much more complex. It will include specialized VLSI gate arrays to perform such NASA specific functions as data sorting and blocking. It will be functionally divided into two areas: a *data storage control subsystem*, to control the data flow from the generic telemetry data system to the kernels; and a *data output control subsystem*, to control the data flow from the kernels to the generic telemetry data system and/or local area network and common carrier interfaces.

The *data storage control subsystem* will be further divided into three areas: an input section, for handling the incoming data from the generic telemetry data system; a sorting section, for sorting the data by unique ID into blocks; and an output section, for multiplexing and transmission of the data blocks to the basic storage kernels.

The *data output control subsystem* also contains the input and output sections, but will not contain the sorting section. Also, the output section will control data flow not only to the generic telemetry data system, but also, to external local area network and common carrier interfaces.

2.3.2 THE BASIC STORAGE KERNEL

The basic storage kernels form the core of the mass storage system. They contain all of the system's commercial storage devices. A basic storage kernel can be configured with interfaces for simple block storage, file storage, and data driven storage modes. Physically, the basic storage kernel will consist of a single VME bus based card cage, multiple parallel drive arrays, and VCR tape units. A single kernel can provide from 9.6 to 400 Gbits of fast access storage. Fully expanded, at 400 Gbits and 2 VCR tape units, the kernel could be housed in 5 six foot racks (see FIGURE 3 - FULLY ECPANDED BASIC STORAGE KERNEL).



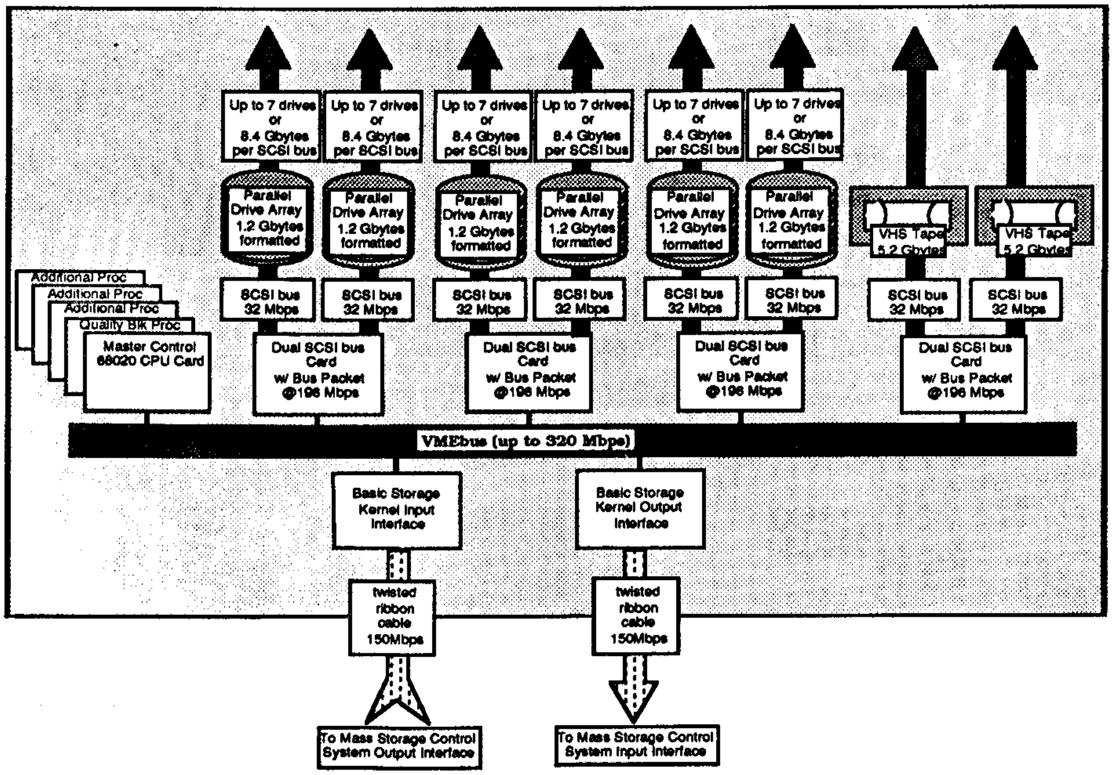
FULLY EXPANDED BASIC STORAGE KERNEL

FIGURE 3

In order to maintain the high data throughput, the kernel will perform pipelined data movement. Separate input and output subsystems will insure that these operations are performed autonomously. On data input, the data received from the mass storage system controller will move across the VME bus backplane and out to one of 6 SCSI bus interfaces onto the appropriate parallel drive array (see FIGURE 4 - BASIC STORAGE KERNEL CONTROL SYSTEM).

In a block storage and file access modes, the kernel will strip the incoming data across six parallel drive arrays simultaneously. The kernel will respond to block oriented type of accesses (e.g. a request for a unique data block by ID) by computing the necessary logical to physical block parameters so that the external interface need only be concerned with the unique block ID number. The kernel will also respond to file access oriented commands (e.g. open/close, create/delete, read/write, reposition, etc.) by consulting its directory of files and disk usage.

The data driven mode of operation will be streamlined for level zero type processing functions. It is assumed that there will be prior fundamental knowledge of the various data rates for each type of data (i.e. source packet ID) and how many types of data will be present in the telemetry stream. The data will not be automatically stripped across all six



BASIC STORAGE KERNEL CONTROL SYSTEM

FIGURE 4

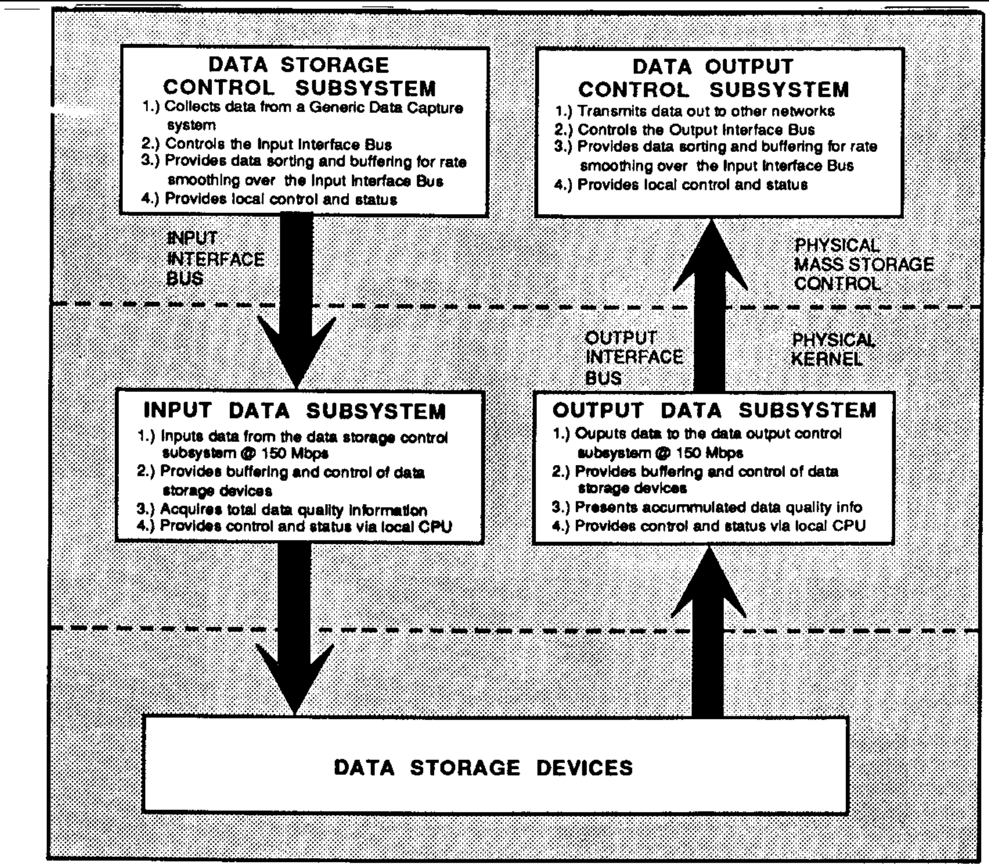
disk drives as in block and file access modes; although it will be stored in a presorted format by data type. The system will be preset to autonomously store data by type and perform stripping only if necessary to maintain the rate for a given type of data. The mass storage system controller will presort the data by type into blocks and pass these blocks along to the kernel for storage.

2.4. TECHNICAL APPROACH

2.4.1 DATA COMMUNICATION INTERFACES

MASS STORAGE CONTROL SYSTEM INTERFACE

The interface between the mass storage control system and the basic storage kernel will be accomplished with 4 custom subsystems (see FIGURE 5 - CONCEPTUAL MODEL OF THE MASS STORAGE SYSTEM). Physically located in the mass storage control system are the *data storage control subsystem* and the *data output control subsystem*; and in the basic storage kernel are the *input data subsystem* and the *output data subsystem.* These four subsystems will be connected by two interfaces: the input interface bus and the output



CONCEPTUAL MODEL OF THE MASS STORAGE SYSTEM

FIGURE 5

interface bus. Each interface will consist of dual 16 bit synchronous data channels and associated control bits. All data will include a 6 bit ECC code that will be used to minimize transmission errors. The use of separate intelligent input and output interface buses will make simultaneous input and output operations possible.

THE DISK FARM INTERFACE

The basic storage kernel will use the industry standard VME bus as the back bone for this interface. This bus is capable of moving data at up to 320 Mbps, however there are no memories available at a reasonable cost capable of accessing data at this speed. Using standard high speed static ram memories, data movement at greater than 160 Mbps sustained has been measured, which is quite adequate for this system.

All data storage devices used for the working storage (“disk farm”) will be controlled via the industry standard SCSI bus. The interface between the SCSI bus and the VME bus will be accomplished with a commercial high performance dual SCSI bus controller. There will be 8 SCSI buses tied to the kernel’s VME bus back bone. Six of the SCSI buses will be used exclusively for disk farm control. The other 2 SCSI buses will control the 2 VCR tape drives. Each SCSI bus is capable of supporting 7 parallel drive arrays. The full scale basic storage kernel will therefore have 42 parallel drive arrays. Each SCSI bus has a bandwidth of 32 Mbps sustained. In order to achieve the desired 150 Mbps bandwidth, these SCSI buses will have to be used in parallel. Data will transfer into the basic storage kernel at 150 Mbps, be split 6 ways and stored on 6 drives simultaneously. Data will be stored at a rate of 192 Mbps. This ensures that the storage rate is always faster than the input/output rate of 150 Mbps.

2.4.2 DATA STORAGE DEVICES

REQUIREMENTS

The data storage devices used for the disk farm are critical to the overall performance and reliability of the system as a whole. The storage devices used must be able to ingest data at 32 Mbps and access that data in less than 50 ms. In addition, this system will have a great number of disk platters, it is inevitable that some of the platters will fail during real time operations. Therefore, the system must be able to lose 1 or more platters without the loss of existing data or the loss of new incoming data. For this reason, high mean time between failure (MTBF) and parallel redundancy are critical.

PARALLEL DRIVE ARRAYS

Parallel drive arrays and array controllers that meet these requirements are available today. One such parallel drive array features five synchronized disk drives in parallel, 4 drives for data and 1 drive for parity. The outstanding feature of this drive is its reliability. Due to the addition of the fifth parity drive, the loss of any single drive in the array will cause no loss of data either during on-line real time operations or off-line. In fact, the loss of any single drive in the array will not affect data storage operations in any manner what so ever. If the failed drive is replaced within 72 hours, then the MTBF is estimated at 1,400,000 hours. It is capable of storing 32 Mbps sustained and 40 Mbps burst and holds 9.6 Gbits of data after formatting. Other manufacturers have announced even higher performance systems which will utilize this parallel array architecture.

SHORT TERM DATA STORAGE FUNCTIONS

Short term data storage functions will be performed by commercially available VCR tape drives. Each VCR drive can ingest data at 32 Mbps sustained and will hold 41.6 Gbits. The basic system will include 2 such drives each on its own SCSI bus allowing short term data storage functions to run at 64 Mbps. The basic storage kernel may be expanded to include up to 14 of these drives (7 per SCSI bus). An alternative basic storage kernel setup, would use SCSI compatible optical disk drives in place of or in addition to the VCR tape drives. These needs would be determined by the end user.

2.4.3 SPECIALIZED DATA PROCESSORS

Each basic storage kernel will have one or more specialized data processors (single board CPUs). These processors will have access to all data stored on the system. They can be used for data sorting, data quality information, data statistics gathering, and many other functions that might need to be performed on the stored data. These processors will perform their tasks either in parallel with the data input/output operations to extend the power of the input/output subsystems, or they may be used in an off line mode to perform more time consuming tasks on the data. These processors could be selected from a variety of commercial single board CPU units and may include custom built sorting engines, local high speed data storage, or other custom functions in the form of add-on mezzanine boards.

3. CONCLUSION

The magnitude of the mass storage problem facing NASA system engineers and planners is truly immense when compared to past requirements and needs. NASA's requirements in the Space Station era, to store, process, and distribute data described in terms of gigabits and hundreds of megabits per second, will challenge even the most optimistic predictions for advancement in the field of commercial storage technology. The solution to these problems will come through the integration of the latest state-of-the-art storage systems with modern systems design techniques utilizing application specific hardware and distributed processing. In order to accurately predict and optimize system performance and cost, early prototype experience with such advanced system architectures is essential.

At the Goddard Space Flight Center, a mass storage prototype system is now under development. This prototype will utilize the latest commercial parallel drive array systems and this center's experience in the development of NASA application specific VLSI systems, to provide a high performance, random access, working storage buffer. This prototype is intended as a "proof of concept" system for NASA's future high performance, configurable, modular, and low cost telemetry mass storage systems.

REFERENCES

1. "Mass Storage Systems for Data Transport in the Early Space Station Era 1992-1998", NASA Technical Memorandum 87826. Carper, Dalton, Healey, Kempster, Martin, McCaleb, Sobieski, and Sos, NASA, Goddard Space Flight Center, Greenbelt, MD, July 1987.

2. “DIF Cost and Technology Evaluation Study’, NASCOM Task S224. Ford Aerospace and Communications Corporation, College Park, MD, December 1987.

NOMENCLATURE

CDOS	Customer Data Operation Service
CPU	Central Processing Unit
ECC	Error Correcting Code
FDDI	Fiber Optic Distributed Data Interface
FEL	First Element Launch
GSFC	Goddard Space Flight Center
LZP	Level Zero Processor
Mbps	Megabits per second
MTBF	Mean Time Between Failure
MTTR	Mean Time To Repair
NASA	National Aeronautics and Space Administration
SCSI	Small Computer System Interface
VCR	Video Cassette Recorder
VLSI	Very Large Scale Integration
VME	Versabus Module Eurocard Format