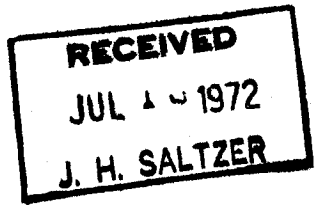
DATE: JULY 10, 1972




TO: C.T. CLINGEN
F.J. CORBATO'
R.C. DALEY
J.W. GINTELL
R.F. MABEE
N.I. MORRIS
J.H. SALTZER ←
R.B. SNYDER
D.R. VINOGRAD
S.H. WEBBER

FROM: L. SCHEFFLER

SUBJECT: ON-LINE DISK STORAGE FOR FOLLOW-ON MULTICS

This is a draft of a document on the long-range strategy for bringing up on-line disk storage for the Multics follow-on. It is intended to generate questions, comments and gripes. Talk to me about them at Honeywell, X229, or send mail to Scheffler, Multics.

Any gaps in terminology can probably be filled in by skimming through "General Description of DSS-181 and DSS-190 Subsystems."

It has not been decided whether this document is an MSB, MCB, MOSN, or what. Any suggestions are welcome.

L. Scheffler

L. Scheffler

att.
/ms

CONTENTS


| | Page |
|---|---|
| Introduction | 1 |
| Initial Multics 645 FO Operation | 2 |
| First Interim DSU-181 DIM | 3 |
| Second Interim DSU-181 DIM | 6 |
| Final DSU-181 Configuration | 8 |
| Final DSU-181 DIM | 10 |
| IOM_manager reprogramming | 12 |
| DSU-190 Configuration | 13 |
| DSU-190 DIM | 15 |
| Optimum Sector Size Considerations | 17 |

## INTRODUCTION

It is expected that on-line disk storage for the Multics follow-on installation at MIT will eventually be provided entirely by DSU-190 disk units. (The DSU-190's are a new line of disk units currently under development by Honeywell. They are similar in operation, size, and performance to the IBM 330 disk units.) The path by which this goal is reached, while at all times guaranteeing on-line disk storage of sufficient size and speed for the Multics follow-on development effort and initial service period is rather complex. This document outlines the several stages involved, equipment necessary, features to be added, and expected performance improvements in each stage, and alternative development paths should equipment/programming/debugging schedules slip. Specific performance figures mentioned here are abstracted from Multics Performance Log (MPL) # 67, "Crude Analysis of Multics Follow-On Disk Subsystem Performance". Assumptions about follow-on mean time between disk page faults (mtbdpf) are made on the basis of MPL #56, "Crude Estimates of Multics Performance with 645 Follow-On Hardware".

INITIAL MULTICS 645 FO Operation

DSU-270 fixed head disk units are currently being used for the debugging of the follow-on hardware in Phoenix. When the first follow-on CPU and memory arrive, sufficient DSU-270 disk storage will be available to bring the initial system up. Therefore, although the first DSS-181 and DSS-190 equipment is expected to arrive at about the same time as the first CPU, operational DSS-181 and DSS-190 subsystems are not immediate necessities.

FIRST INTERIM DSU-181 DIM

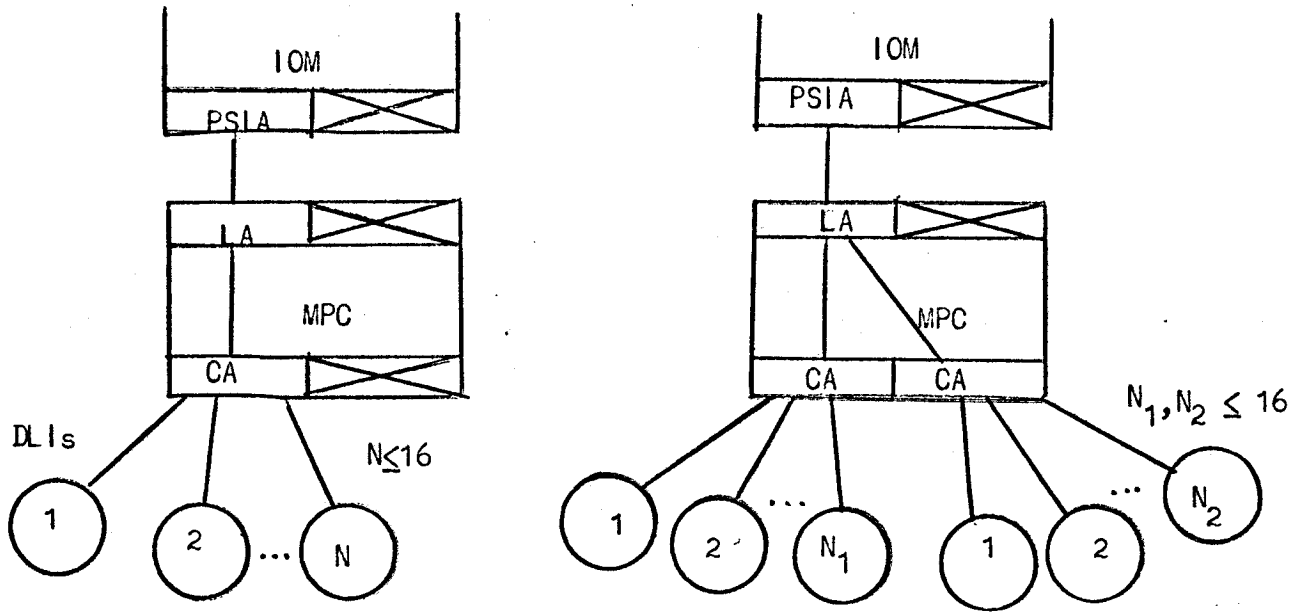
As the DSU-181 disk units, associated MPC controller and microprogramming, and channel and adapter hardware, will be tried and tested products by the time they arrive (the first MPC was only recently delivered to a customer), and the DSS-190 equipment is freshly off the drawing board, it is most reasonable to first bring up a DSS-181 subsystem. The only major stumbling block here is producing a working DSU-181 DIM for Multics. The current plan is to iterate through two or three DSU-181 DIMs, each adding additional features and improving performance, so as to make the DSU-181 disks available as early as possible for follow-on development use.

The DSU-181 disk units and MPC controller can be operated at reasonable efficiency in a fashion very similar to the way the soon-to-be-installed overlapping-seeks DSU-170 DIM operates the DSU-170 disk units. Without attempting to take advantage of the many additional capabilities of the DSS-181 over the DSS-170, the DSU-170 DIM needs to be modified only slightly (error messages and device identifier) to operate the DSU-181 units using only a single IOM channel. Thus, only a short period of relatively high priority work will provide the follow-on development 1-CPU configuration with more than enough disk storage (up to 16 DSU-181 disks, or 71,000 pages with a single controller adapter), with expected performance

comparable to that of the DSU-170 DIM:

| | |
|---|---|
| mean disk page fetvh time | 85 ms |
| 90% disk page fetch time (90% of requests see less page fetch time than this) (This assumes a mtbdpf of approximately 75 ms for a 1-CPU system). | 170 ms |
| maximum average number of page faults per second | 19.4 page faults/sec |

This DIM will be capable of handling either of these two configurations:




PSIA = Peripheral Subsystem Interface Adapter

LA = Link Adapter

MPC = Microprogrammed Peripheral Controller

CA = Controller Adapter

DLI = Device Level Interface

(These are figures 2.1.4.1 and 2.1.4.2 from EPS-1 43A23985, DSS-181 and DSS-190 Removable Media Disk Storage Sub-systems.)

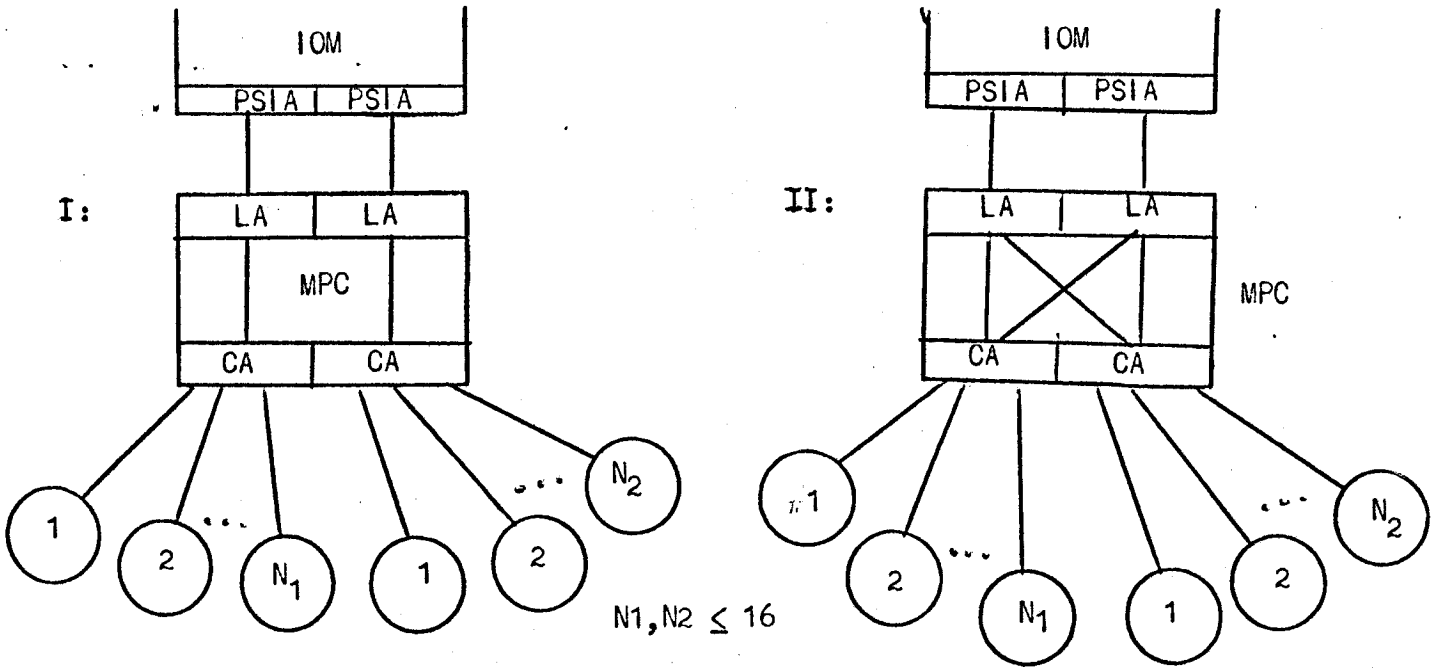
As 9 DSU-181 units are expected to arrive with the first CPU, this DIM will probably start off on the 1-CA configuration.

## SECOND INTERIM DSU-181 DIM

Depending on equipment arrivals and the progress system debugging, there are two possible paths from this point. If hardware (a second PSIA, LA, and CA) and microprogramming for two-channel operation of the DSC-181 controller have arrived, and a second 645FO CPU has arrived and there is a need for non-disk-channel-limited 2-CPU operation (very doubtful), a second interim DSU-181 DIM, capable of operating two IOM channels, can be installed. On the other hand, if two-channel operation is not immediately necessary, and if optimum-disk-sector-size and limited crossbar microprogramming have been completed for the DSC-181 MPC controller, then work can begin on debugging the final DSU-181 DIM. (See next section.)

This second interim DIM would also not attempt to take advantage of new capabilities offered by the DSS-181 controller or disk units, but would merely implement the DSU-170 DIM overlapped-seeks strategy for two channels.

Depending on the status of MPC microprogramming, this DIM will be able to operate either one-channel configuration, and one of the two following configurations:




(These are figures 2.1.4.4 and 2.1.4.5, respectively.)

Configuration II is preferable because it may be operated, at reduced performance, with only one PSIA-LA channel operational.

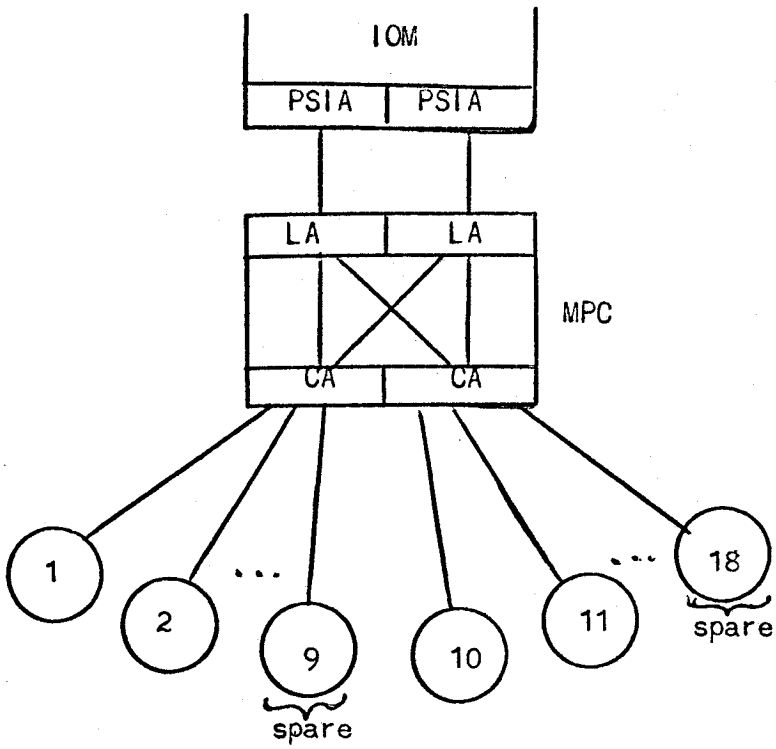
For two-channel operation, both configurations are expected to perform similarly:

| | |
|---|---|
| mean page fetch time | 83 ms |
| 90% page fetch time | 170 ms |
| maximum average frequency of page faults | 41.6 page faults/sec |

(This assumes a mtbdpf of approximately 41 ms for a 2-CPU system.)

FINAL DSU-181 CONFIGURATION

The objectives of low-cost, large capacity (at least 80-90K pages), and fast access time combine to make the following configuration the most attractive for the MIT Multics Installation:




This can supply up to 133,320 pages under present (64-word sector) formatting, or 151,500 pages under sector sizes of 128 or 256 words, with two spare drives. It affords one or two-channel access to all disks. Reliability studies predict a mean time between failure for the MPC controller of between 3000 and 4000 hours, with a mean-time-to-fix of 1.5 hours. Only failures in the basic MPC or one of the CA's need crash the system; failure in a PSIA, LA, or disk unit can be circumvented, at reduced performance, until the system is shutdown gracefully.

We expect to receive a total of 18 DSU-181 units, or 80K pages of storage capacity (90K pages with 128 or 256 word sectors).

FINAL DSU-181 DIM

Although this configuration appears optimum for the MIT Multics, it is not clear that it will be optimum or even very good for other installations. It appears at this stage that, with a minimum of extra code, both final DSS-181 and DSU-190 DIMs can be made able to handle most of the possible configurations at reasonable efficiency. This will have the advantages of flexibility of operation during equipment maintenance and experimentation, and easy exportability to other installations.

The final DSU-181 DIM will take advantage of most of the capabilities of the DSU-181 disk units and DSC-181 controller. Briefly, the read (or write) of a single page involves two sequences of operations. The pre-seek sequence (which takes less than a millisecond to initiate and several tens of milliseconds to complete) causes the read/write heads on the appropriate disk unit to begin seeking for the cylinder where the desired page resides. When the heads finally arrive at the cylinder, the system is informed by means of a Special Interrupt. The data-transfer sequence causes the heads to begin reading (or writing) when the first word of the desired page comes under the heads, and to stop reading when a specified number of words have been read.

The general strategy is quite simply stated. Arriving requests are kept in two queues - a high priority, and a low priority queue. All requests in the high priority queue are completed before any data transfer for requests in the low priority queue are begun. Normally, read requests (page faults) go into the high priority queue, and write requests into the low priority queue. Whenever a path (i.e., a PSIA-LA-CA) becomes free, all possible pre-seek sequences are initiated on disk units accessible via all free paths. When no more pre-seeks can be initiated and no seeks previously initiated are completed yet, the DIM returns, to wait for a Special Interrupt signalling the completion of a seek. When at least one seek has completed and at least one PSIA-LA-CA path is open, a search is made of the queues (high priority queue first), and a data transfer sequence is initiated for the first request with an open path and its seek complete.

This strategy preserves, as much as possible, the order of arriving disk read requests in the order of request completions. And since the strategy depends only on the concept of a "path" to a unit consisting of a PSIA, LA, CA, and DLI, the several possible configuration categories may be accomodated simply by a suitable means of defining the allowable paths between each disk unit and the system.

More detailed information on the operation of this DSU-181 DIM will be available in the initial design document, to be out soon.

The expected performance of the final DSU-181 subsystem is:

| | |
|---|---|
| mean page fetch time | 78 ms |
| 90% page fetch time | 182 ms |
| max.avg. frequency of page faults | 41.6 page faults/sec |

(This assumes a mtbdpt to the DSU-181's of approximately 41 ms for a 2-CPU system.)

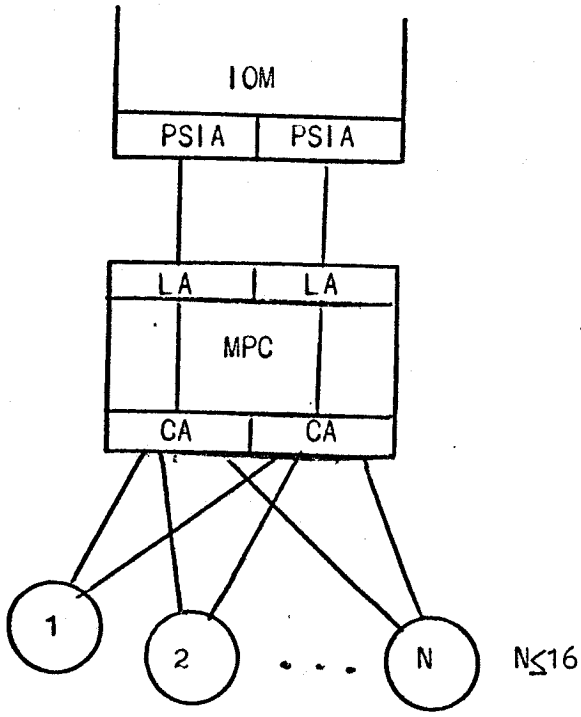
In addition, the DIM will contain facilities for the eventual addition of on-line disk T&D (test and diagnostic) programs to allow troublesome devices to be located, packs changed, and drive troubles located and repaired, all with Multics running.

IOM_MANAGER REPROGRAMMING

To enable the use of Special Interrupt status storage to convey unit identifiers at the completion of pre-seeks, code will have to be added to the iom_manager to manage a status queue of at least 32 double-words of pending status (one pair per possible drive).

DSU-190 CONFIGURATION

The eventual on-line secondary storage requirement for the Multics follow-on, as predicted in MPL #56 "Crude Estimates of Multics Performance with 645 Follow-On Hardware", is in the neighborhood of 300 million words. Given Multics optimum sector-size formatting on the DSU-190 disks, each disk can hold more than 19000 Multics pages. A single MPC can accomodate up to 16 such disk units (15 active plus 1 spare) in a two-channel fully crossbarred (two cables to each disk unit) configuration, as shown below




1X2*16 Configuration

Dual Port Option on Disk Units

This gives a storage capacity of up to 291 million words (285,000 pages), with one drive spare. This would appear to satisfy the needs of some 2500 registered users, each with 100 pages of quota. The reliability of this configuration is quite high, since a failure in any one module (other than the basic MPC itself) can do no worse than block access to one disk (until the pack is moved to another drive), or reduce performance.

Eight DSU-190 units are expected to be available when the follow-on starts its service, or 133,000 pages of storage with 256 word sectors, in addition to DSU-181 disks.

DSU-190 DIM

Although the first DSU-190 disk units and controller are scheduled to arrive very soon after the arrival of the first CPU, there is no need to have them immediately operational. The already-operational DSU-270's that arrive with the first CPU, and the DSU-181's that will become operational soon after, will be sufficient for the follow-on development effort. The added capacity of the DSU-190's will not be needed until the follow-on is ready for service sometime in Spring 1973. Therefore, there is no need for a layered approach of interim DIMs for the DSU-190.

The DSU-190 DIM will be quite different from DSU-170, 270, and 181 DIMS, primarily because of many features previously implemented in the DIMs are now implemented in the MPC microprogramming. These features are described in "General Description of DSS-181 and DSS-190 Subsystems".

Briefly, the DIM is expected to operate as follows: Arriving requests to read or write single pages are stored in two queues, a high-priority (read) queue, and a low-priority (write) queue. All requests in the high-priority queue are completed before any in the low-priority queue are begun.

Each physical channel (PSIA/LA) has four logical channels associated with it. This means that up to four complete DCW lists for four different disk units (seek IDCW, seek data IOTD, read/write

IDCW, and data transfer IOTD), may be accepted by the MPC over each physical channel, for parallel execution by the MPC. The MPC will issue seeks, and when more than one unit has seek complete, the MPC interrogates the disk units' rotational position sensing (RPS) to determine which read/write head will encounter its data soonest, and performs that data transfer first.

The DIM merely needs to keep track of which units are busy, and which logical channels are open on which physical channels. It is relatively simpler than previous high-performance DIMS, as it no longer needs special interrupts and uses only one channel connect per request.

The expected performance of the 1X2*N configuration under this DIM with four logical channels per physical channel is:

| | |
|---|---|
| mean page fetch time | 47 ms |
| 90% page fetch time | 122 ms |
| max avg. frequency of page faults | 88.5 page faults/sec |

(This assumes a mtbdpf to the DSU-190's of about 41 ms for a 2-CPU System.)

## OPTIMUM SECTOR SIZE CONSIDERATIONS

Because of the difficulty of changing hard-wired capabilities in the present DSU-170 controller, Multics has been forced to be satisfied with using 64-words/sector formatting on the DSU-170's. Quick calculations show that this causes some 10-15% waste of storage space in inter-sector gaps and format information.

The DSC-181 and DSC-190 controller, being microprogrammed, are not so restrictive. As the table on the next page shows, changing the microprogramming to allow 256-word sector sizes can increase DSU-181 pack capacity by about 13.6%, and DSU-190 pack capacity by a whopping 30.7%, or about 4500 pages/pack.

With these arguments, it is perhaps not clear, for the MIT Multics whose DSU-181 units will disappear when the DSU-190's become operational, that optimum sector size microprogramming for DSU-181's is immediately necessary. However, if other Multics installations will be using DSU-181's, it may be worthwhile.

However, when considering the storage requirements of the follow-on MIT Multics, the lack of optimum sector size microprogramming means having to buy a second MPC and channel and associated hardware to meet targeted 2-CPU performance and storage capacity requirements.

## SECTOR SIZE POSSIBILITIES

### DSU-181

| 36-Bit Words/Sector | 8-Bit Bytes/Sector | Sectors Per 1024-Word Page | Sectors Per Track | Pages Per Cylinder | Pages Per Pack |
|---|---|---|---|---|---|
| 64 | 288 | 16 | 18 | 22 | 4444 |
| 128 | 576 | 8 | 10 | 25 | 5050 |
| 256 | 1152 | 4 | 5 | 25 | 5050 |
| 512 | 2304 | 2 | 2 | 20 | 4000 |
| 1024 | 4608 | 1 | 1 | 20 | 4000 |
| 320 | 1440 | 4 | 4 | 20 | 4000 |
| 1600 | 7200 | 1 | 1 | 20 | 4000 |
| 376 | 1666 | 4 | 4 | 26 | 5252 |
| 176 | 792 | 8 | 8 | 26 | 5252 |

### DSU 190

| | | | | | |
|---|---|---|---|---|---|
| 64 | 288 | 16 | 31 | 36 | 14724 |
| 128 | 576 | 8 | 18 | 42 | 17178 |
| 256 | 1152 | 4 | 10 | 47 | 19233 |
| 512 | 2304 | 2 | 5 | 47 | 19223 |
| 1024 | 4608 | 1 | 2 | 38 | 15542 |
| 320 | 1440 | 4 | 8 | 38 | 15542 |
| 2880 | 12960 | 1 | 1 | 19 | 8871 |

## DSU-190 CAPACITIES

| | |
|---|---|
| Target Capacity | $\sim$ 300 million words |
| 1X2*16 Configuration w/64 word sectors (1 spare drive) | 226.2 million words |
| 1X2*16 Configuration w/256 word sectors (1 spare drive) | 295.4 million words |

These figures are based on the assumption that DSU-190 pack formatting follows the rules of IBM2314 A1 pack formatting within about 5%. Reliable information on this is being sought.