IBM

Technical Report

Page/Swap Configurations

Thomas Beretvas

PAGE/SWAP CONFIGURATIONS

November 1982

Thomas Beretvas

IBM IS&TG
Poughkeepsie, NY 12602

This paper describes relatively easy ways to evaluate proposed paging/swapping configurations, to configure systems, and to assess the performance impact of such configurations. The methodology is applied to MVS systems but is applicable in part to VM systems.

The paper presents simple modeling formulae, which can be used to compare paging device capabilities in terms of throughput (pages/second) and response times. A system configuration methodology is provided satisfying paging rate and page space demands.

First Edition (November 1982)

The information contained in this document has not been submitted to any formal IBM test and is distributed without any warranty either expressed or implied. The use of this information is a customer responsibility and depends on the customer's ability to evaluate and integrate it into their operational environment.

References in this publication to IBM products, programs, or services do not imply that IBM intends to make these available in all countries in which IBM operates.

Publications are not stocked at the address given below. Requests for copies of IBM publications should be made to your IBM representative or to the branch office serving your locality.

Address comments concerning the contents of this publication to IBM Corporation, Marketing Publications, Dept. 824, 1133 Westchester Avenue, White Plains, New York 10604. IBM may use or distribute any of the information you supply in any way it believes appropriate without incurring any obligation to you.

PREFACE

```
-> Performance data apply to specific <-
->   configurations and workloads.    <-
-> This presentation contains        <-
->   mostly modeling data              <-

-> Care should be exercised when      <-
->extrapolating to other environments.<-

Presentation of the data does not
constitute a warranty that any other
installation will obtain comparable
    or better performance.

I gratefully acknowledge the
     contributions of
C. Dowling, G. King and M. Goldfeder
```

Figure 1. FOIL 1--TITLE: PAGE/SWAP CONFIGURATIONS

Figure 1 contains the title foil of the presentation.

This paper is being made available to IBM and customer personnel. It has not been subject to any formal review. The information herein has not been submitted to any formal IBM test and is presented on an "AS IS" BASIS WITHOUT ANY WARRANTY EITHER EXPRESS OR IMPLIED. The use of this information or the implementation of any of these techniques is a customer responsibility and depends on the customer's ability to evaluate and integrate them into the customer's operational environment. While each item may have been reviewed by IBM for accuracy in a specific situation, there is no guarantee that the same or similar results will be obtained elsewhere. Customers attempting to adapt these techniques to their own environments do so at their own risk.

The performance data contained in this document is modeling data; the results which may be obtained in real operating environments may therefore vary significantly. Users of this document should verify the applicable data for their specific environment.

ACKNOWLEDGEMENTS

I want to thank Mr. C. Dowling, Dr. M. Goldfeder, and Mr. G. King for their contributions to the technical material in the paper. In particular, APL models and the graphs were produced by Mr. Dowling.

I want to thank Ms. R. Bibolet and Messrs. J. Colosimo, J. Cord, G. Dodson, S. Friesenborg, S.Goldstein, G. Neaga, R. Siederer, S. Samson, Y. Singh and R. Wasserman for reviewing the paper, and making suggestions and comments.

TABLE OF CONTENTS

LIST OF ILLUSTRATIONS

1.0 INTRODUCTION

- Provide simple formulae for evaluation of paging/swapping configurations. Applicable to MVS (and VM)
- Use formulae to obtain device capabilities.
  - -Throughput : pages/sec
  - -Response time (per I/O request)
- Provide evaluation methodology satisfying paging rate & space demands
- Illustrate methodology by examples.
- Develop easy-to-use design graphs.
- Provide tuning hints.

Figure 2. FOIL 2--OBJECTIVES

Figure 2 contains the foil with the objectives of the presentation.

The objective of this paper is to provide relatively easy ways for the evaluation of proposed paging/swapping configurations, to configure systems, and to assess the performance impact of such configurations. The paper presents simple modeling formulae, which can be used to compare paging device capabilities in terms of throughput (pages/second) and response times. A system configuration methodology is provided satisfying paging rate and page space demands. The methodology is illustrated by examples. Design graphs are also given to facilitate design.

The paper presents a modeling methodology, which is probably too detailed for most applications. The modeling methodology is used to obtain tables of performance capacity for both swapping and paging devices. Design graphs are also plotted using the modeling method. In essence, three levels of guidelines are provided:

1. Swap and demand page device use limit tables. The use of these tables is very simple, and does not require detailed understanding of the modeling method. The use of these tables provides recommended limits for paging rates sustainable by the devices. While the tables were derived for dedicated paging channels, their use is applicable for dedicated paging devices on non-dedicated paging channels.

2. Design graphs. These graphs were generated for dedicated paging channels. They provide guidance for paging response times obtainable at given paging rates. They can be used to design a TSO system with a predetermined response time, and to compare the performance of various paging devices.

3. Detailed modeling methodology. Examples are given for dedicated and nondedicated paging channels. Detailed performance information is obtained in the form of device, channel and control unit utilization and response times for given paging rates. The use of this (detailed) methodology is somewhat more complex, even though simplified formulae are provided where appropriate. The methodology presented can be used to evaluate the performance of unusual paging configurations, such as mixed paging/data base configurations.

		3380	3375	3350	3880-11	2305
SWAP	PAGES/ACT	120	90	60	240	72
MAX.	ACT/ PATH	2	2	2	1	2
DEMAND	PAGES/ACT	15-20	15	15	60-80	20
MAX.	ACT/ PATH	4+	4+	4+	2	4

Figure 3. RECOMMENDED PAGING DEVICE USE LIMITS

For the casual user Figure 3 summarizes the paging device use information explained in detail in the text. It lists the paging loads sustainable by one actuator, (in this context the 3880-11 is considered an actuator).

Thus, for example one 3380 actuator can be loaded to a limit of 120 swap pages/second, and a 3880-11 to 240 pages per second. The maximum number of swap actuators per path is 2 (1 for the 3880-11).

A 3880-11 can be loaded to a limit of 60-80 demand pages/second. One 3380 actuator can be loaded to 15-20 demand pages per second. The maximum number of actuators on a dedicated demand paging path is 4. 2 demand paging 3880-11s can be placed per path.

2.0 PAGING CONFIGURATION POSSIBILITIES

->Paging configuration possibilities.

- Device characteristics.
- Paging space size and rate calculations.
- Concepts of I/O activity and formulae.
- Examples.
- Swap and page device limits.
- Configuration examples.

Figure 4. FOIL 3--OVERVIEW

Figure 4 introduces a new topic in the foil presentation by a foil of the type shown. The next subject is indicated as "Paging configuration possibilities".

In the rest of this paper these guidepost foils are omitted, causing an apparent discontinuity in the foil numbers.

Before presenting the quantitative performance methodology, a qualitative overview of configuration alternatives and their characteristics is provided.

2.1 PAGING TYPES

PAGING TYPE	PAGES/ IO REQ.	PAGE DATA SET TYPE
PAGE-IN (FAULT)	1	LOCAL
PAGE-OUT (STEAL)	1-2...2	LOCAL
VIO	1-10..4	LOCAL
SWAP TRIM	1-8...5	LOCAL
SWAPIN/OUT	1-12..12	SWAP
SWAPIN/OUT	1-?...?	LOCAL
PLPA	1 ..1	PLPA
COMMON	1-2 ..1	COMMON

Key Decision: whether to separate "demand" paging from "swapping"

Figure 5. FOIL 4--PAGING TYPES. (MVS)

The various paging types were discussed in Reference 1. Figure 5 gives a review of these paging types. In MVS, "local" paging devices can handle all paging traffic except for PLPA and COMMON pages. In this paper all discussion is oriented to paging that can go to LOCAL paging devices, since most paging problems involve swap and demand paging that could go to local page data sets. Thus, COMMON and PLPA paging are not discussed further.

In MVS/SP 1.3 all page faults are separate I/O requests; no other I/O is associated with the page fault, i.e., upon completion of the (single) page read the system (the user) is notified via a PCI operation.

When the system steals pages from an address space, one or two pages are taken at one time and, if they had been changed, are written as a single I/O request, usually, but not necessarily into consecutive locations, i.e., most likely single seek, probably on same track. For the purposes of calculation it is reasonable to assume that two (consecutive) pages are written.

Swap pages are written to and read from swap data sets (if such are defined) as swap sets of up to 12 pages per I/O request, always with a single arm motion (seek) and into contiguous slots. For the purposes of calculation it is reasonable to assume that 12 pages are read/written.

If swap data sets are not defined, contiguous slot allocation places pages to be swapped into contiguous slots (if possible) on local page data sets, but the number of contiguously allocated slots is dynamically adjusted depending on the ASM[1] burst size calculation.

ASM attempts to achieve a burst size of 80 ms, i.e., attempts to create a chain long enough to achieve a response time of 80 ms for the I/O request. The burst size in terms of pages depends on the device used, it is reasonable to assume that on 3350s it is 8-10 slots long, and on 3380s the length could be 18-30 slots long.

At swap-out time, unreferenced but changed pages are "trimmed" from the address space and are written out contiguously to local page data sets. Pages which are both unreferenced and unchanged are unaffected.

VIO windows are contiguously allocated on local page data sets. VIO writes are blocked according to the window size. The VIO window size is determined by the "simulated" device type, which has no relationship to the real device. For simulated 3350 VIO devices it is reasonable to assume a VIO window size of four pages. For simulated 3380 VIO devices it is reasonable to assume a VIO window size of ten pages. The use of simulated 3380 VIO devices results in the use of larger block sizes, which is advisable for large VIO data sets.

The response time of the paging subsystem (and consequently system interactive response time) is directly determined by <u>page-in and swap-in</u> response times; page-out and swap-out response times do not directly impact system response times. Also, the characteristics of a page-in operation are very different from the characteristics of a swap operation. Swap operations keep the channel busy with long data transfers while demand page operations leave the channel relatively free, even though the paging device can be fairly busy.

Consequently a key decision to be made by the system administrator is whether to mix swap and demand paging load on the same devices, and even if on separate devices, whether to mix them on the same channels, or to separate them. Much of the subsequent discussion centers in quantitatively assessing the response time impact of this and related decisions. One thing should be fairly clear however, even without detailed discussion: heavy swap paging that keeps the devices and the channels busy causes an elongation of the demand paging response time on the same device/channel.

---

[1] Auxiliary Storage Manager, the MVS component handling auxiliary storage.

2.2 PAGING CATEGORIES

```
1. SWAPPING
 -36+ pages transferred in/out at once.

 -Good swap-in resp. time if 3+ I/O
  requests on separate, parallel paths.
  Important: fast data transfer rate.

 -Swap space size requirements: low
  --->  high access density.
----------------------------------------

2. DEMAND PAGING.
 -Resolution of single page fault.

 -Best page-in response time
  from device with fast access time.
  3880-11

  Space size requirements: high
   ---> very low access density.
   (compared to swap space)
```

Figure 6. FOIL 5--PAGING CATEGORIES

Figure 6 contains a summary of paging categories.

With extended swap, (with the IUP 5796-PNW, or with MVS/SP 1.3 and later releases) a TSO swap group normally contains at least 36 pages. Thus, if swap data sets had been defined, at least 3 swap sets are swapped in or out. All swap I/O (in or out) requests for an address space are started concurrently, thus good swap-in response time can be obtained when the number of separate, independent, parallel swap paths is at least as large as the number of swap I/O requests (swap sets if swap data sets exist). A path means the combination of a channel, control unit and head of string. Parallel operations can take place if the paths used are all free.

For good swap response time devices with fast data transfer rates are desirable, such as 3380s or 3880-11s.

Swap data set size requirements are very small. Detailed calculations are given later. High paging rates and low space occupancy means high access density.[2]

Good demand paging performance can be obtained by devices with fast access times and adequate capacity, such as the 3880-11.

The speed of the data transfer is not as important, since for most devices the time spent in transferring a single page is fairly small in comparison with the access and queue times.

Most of the paging I/O activity occurs on TSO systems because of physical swapping. When swapping is separated from demand paging on such systems, the (demand) page data sets have large page space requirements, but low rate of accesses, i.e., low access density when compared with swap data sets.

The characteristics of swap and demand page data sets can thus be contrasted: swap data sets are characterized by small space requirements and very high paging loads, just the opposite of demand paging data sets. This can be summarized by saying that swap data sets have very high access densities, and demand page data sets have low access densities.

---

[2] Access density is defined as paging rate divided by the active (used, i.e., not allocated but unused) paging space.

2.3 SYSTEM ORIENTATION

```
•TSO RESPONSE TIME. ---> PRODUCTIVITY
 -TSO favored over batch.
 -Trivial TSO favored over nontrivial.
 -First: increase main storage to the
  limit. It is the best paging device.
 -Next: improve paging subsystem
  Swap and demand paging

  Actions:
 -> Maximum parallelism for swap.
 -> Separate swap from demand paging.

->>> Channel capacity permitting:<<<-
  Enhance swap to reduce demand paging
 -> Storage isolation for maximum swap.
 -> Reduce logical swap by reducing
    system think time (LSCTMTE)

•Batch throughput.
 ->Use SRM control to favor batch.
   (E.g. dispatching  priority)
 ->The rules above should be followed.
```

Figure 7. FOIL 6--SYSTEM ORIENTATION

Figure 7 discusses the orientation of a system.

It is possible to design the system to provide fast interactive (TSO) response times or make it more throughput (batch) oriented. Designing for response time is very important, since it results in better people productivity (Reference 2), which reduces people cost.

One approach to this kind of system orientation is favoring TSO over batch and trivial TSO transactions over nontrivial TSO transactions. There are system tuning parameters available to accomplish this goal.

The first action to consider is to increase main storage. When main storage is increased, paging (of all kinds) is reduced. The effect is to decrease:

1. CPU cycles associated with paging.

2. The required number of paging devices.

3. The paging load carried by channels.

When the main storage used is at its limit, then it is worthwhile considering an enhancement of the overall paging subsystem, both swapping and demand paging.

1. Swapping can be enhanced by providing maximum parallelism. This means that the swap paging load is distributed to preferably three or more, not necessarily dedicated paths (channels, control units, heads of string). As swapping takes place faster, the address space takes up main storage for a shorter period of time, and this in turn reduces demand paging. Main storage occupancy is measured in terms of frame-seconds, and the shorter the period the address space spends in waiting for swap-in, (or page faults), the less real storage is occupied by that address space. As an example, an address space with 36 frames and a response time of 0.5 seconds occupies 36 x 0.5 = 18 frame-seconds. The same transaction with a one second response time occupies twice as much. The smaller this figure is, the smaller impact it represents to the system.

2. If swap paging is separated from demand paging, at least in terms of devices, superior demand paging response time can be obtained, since

demand paging is not queued behind long swap chains. Separation of demand and swap paging often means, however, that more paths carry paging load than without separation. It also means, primarily in the case of small to medium systems, a requirement for more page/swap devices.

3. A better paging subsystem can be obtained by using more advanced paging devices, i.e., faster devices for swapping, and faster access devices for demand paging.

A better performing paging subsystem can be provided by enhancing and increasing swapping and, simultaneously, reducing demand paging. This approach is workable if and only if adequate path (channel, control unit and device) capacity is available to handle the additional swap load. The methods listed below require the additional swap capacity, and they are not applicable if only a limited number of saturated swap paths are available in the system.

4. By introducing storage isolation (using the PWSS parameter in the IPS, e.g. PWSS=(36,*)) for trivial TSO transactions, the demand page load is decreased but the swap load is increased. As an example, an internal IBM system, TSO with an average swap group size of 37 pages, was storage isolated to 48 frames resulting in an average swap group size of 45 pages. The trivial response time before storage isolation was one second.

   As a result of using storage isolation the overall main storage occupancy of an address space is reduced. The short term main storage occupancy may increase, the overall main storage occupancy (the frame-second product) decreases due to the shorter time spent in main storage waiting for page faults.

   In the example above, the swap group size increase of eight frames could result in saving four page faults per transaction. With a paging response time of 50 milliseconds (37 x 0.05 x 4) = 7.3 frame-seconds are saved, and the response time is reduced to 0.8 seconds. (In fact, the figure is somewhat larger since the size of the address space is increased by the page faults). Assume that because of the swap group size four additional and subsequently unreferenced frames are swapped in, and the transaction response time is 0.8 second. The cost is 4 x .8 = 3.2 frame-seconds.

   Thus, a gain of (7.3 - 3.2 = 4.1) frame-seconds is achieved. (The 50 % rereference ratio has been measured on real systems - as per communication from Gary King.)

5. If the maximum "system think time", LSCTMTE, defined in the IEAOPTXX member of SYS1.PARMLIB is reduced to a small number, say 1-5 seconds from the default of 30 seconds, only those users are logically swapped who interact with the system very fast. As a result, the main storage occupancy due to logically swapped users is reduced, the main storage available for page buffer, i.e., the number of frames, containing unreferenced pages with steadily increasing UIC values is increased and demand paging is reduced.

   As an example, with a maximum think time of one second a logical swap percentage of 40 % has been observed on an internal IBM TSO/batch production system, i.e., even a system think time of only one second can produce significant logical swapping.

   In most cases the presence of at least the minimum logical swapping permitted by a one second system think time is very desirable, since without it channel loads may become very high. Thus, the minimum value of LSCTMTE should be set to at least 1 second.

If TSO responsiveness is enhanced, batch tends to suffer. This occurs because TSO is seldom in page wait status anymore, and the system tends to dispatch the higher priority TSO address spaces. If higher batch throughput is desired, then the SRM controls in the IPS have to be used to allocate system resources between the two kinds of load. In other words, the enhancement given to TSO paging via the described mechanisms should be done even when batch is to be favored, but SRM controls are to be used for favoring batch. For example certain kinds of batch jobs can be dispatched at a higher priority than long running TSO transactions.

2.4 THE EFFECT OF SEPARATE SWAP DATA SETS

```
Advantages:
 -Swap I/O separated from demand paging
 -Low seeks <-- small (swap) data sets.
 -Higher channel utilization possible
  on dedicated swap channels.

Disadvantage:
 -On <medium systems loss of granularity
  (may be more actuators required.)
 -Slightly more tuning effort required.

Recommend:
 -use on combined DB/DC-TSO systems.
  so (TSO) swapping does not interfere
  with (DB/DC) demand paging.
 -use on large (>=3033UP) systems with
  many (>50) TSO users, especially
  with different paging device types.

Corollary:
 -Separate demand page data sets.
 -Good demand page response time.
```

Figure 8. FOIL 7--SEPARATE SWAP DATA SETS

Figure 8 illustrates the effect of separate swap data sets.

The existence of separate swap data sets means that the heavy swap load is concentrated on the swap devices. The "local" page data sets therefore carry only demand paging load, providing better demand paging response times.

Swap data sets are small, therefore seeking on swap data sets is kept to a minimum, thereby improving response times.

If the swap data sets reside on dedicated swap channels, higher swap channel utilizations become possible. Experience shows that a maximum channel (control unit) utilization of 50 % is achievable with dedicated swap channels.

On small systems the swap paging rate may be small, and so is the demand paging rate. The use of separate swap data sets may require the presence of more paging actuators than the use of "all local" page data sets. This is called the granularity effect.

The use of swap data sets also necessitates a slightly more intensive tuning exercise than the use of all local page data sets.

The use of swap data sets is strongly recommended whenever the system has a DB/DC subsystem (e.g. IMS or CICS) in addition to the TSO users, so that swap paging does not interfere with demand paging from the on line data base system. Separate swap data sets should also be used on large systems, i.e., on 3033 UPs and above, with many i.e., more than 50 TSO users, especially when different paging device types are used, for example 3380 swap and 3880-11 demand paging devices.

The separation of swap data sets implies demand paging local page data sets, which means that the demand paging on these data sets is not queued behind long swap paging chains. The result is significantly reduced demand paging response times.

2.5 THE USE OF ALL "LOCAL" PAGE DATA SETS

```
Advantage:
 -Better granularity,
  potentially fewer actuators.
 -Less tuning effort required.


Disadvantage:
 -Paging queued behind swapping.


Recommend:
 -Use on <large dedicated TSO systems.
 -Use on non-TSO systems.
 -When the number of paging actuators
  is less than 6-8.
 -Single paging device type used.
```

Figure 9. FOIL 8--ALL "LOCAL" PAGE DATA SETS

Figure 9 illustrates the effect of "all local" data sets.

The use of all local page data sets has the opposite effect from the use of separate demand page and swap data sets. It provides better granularity, and requires less tuning effort. The use of these data sets is recommended on small to medium size TSO systems, and on systems without TSO, i.e., no swapping. In order to explain the granularity effect, one might say that for optimum swap parallelism at least three separated swap data sets are required. If the total number of paging actuators used for both demand and swap paging is less than six, then in most cases best performance can be obtained by having an all local configuration. If the number of actuators exceeds eight, then the pool becomes large enough, so that separation of paging from swapping becomes (usually) worthwhile.

2.6 GUIDANCE FOR PAGING CONFIGURATION MIX

```
HEAVY TSO OR BATCH SWAPPING.
|
+-------> NO - NO SWAP DATA SETS
|
YES
|
DB/DC --> YES - CONSIDER USE OF SWAP DATA SETS.
|
NO
|
ONLY ONE PAGING DEVICE TYPE (E.G. 3350)
|
+-------> NO  - CONSIDER USE OF SWAP DATA SETS.
|
YES
|
PAGING DEVICE TYPE IS 3880-11
|
+-------> YES - CONSIDER USE OF SWAP DATA SETS.
|
+-------> NO  - CONSIDER USE OF ALL LOCALS.
```

Figure 10. FOIL 9--DETERMINING PAGING CONFIGURATION MIX

Figure 10 contains a guide in flowchart form, as to whether swap data sets should be used or not. For example, if there is heavy swapping and a DB/DC environment which is sensitive to demand paging, swap data set use should be considered.

If only one paging device type is used (unless it is the 3880-11) local data sets should be considered even in a TSO environment.

In non-TSO environments swap data sets should not be used.

2.7 PAGING DATA SET REQUIREMENTS

```
•Swap (or "all local" page) data sets
 in a heavy swap environment.
 -Characterized by high paging rates.
-> Devices with high data transfer rates
   (3880-11, 3380, 3375, 2305, 3350)
   Maximum parallelism:
   (Swap group size / 12 = # of paths.)
   At least 3 actuators, distributed on
   at least 3 (non-dedicated) paths.
   Or at least 2 dedicated paths.
   Asymmetric connection of strings.

•Demand paging data sets.
  MVS: Low load (because of swap ds's).
  Low access time is desirable.
  Recommend: 3880-11, 3380, 3375, 3350
  Use 3880-11s if avail., DASD otherwise.
  Not 2305s (low access density)
   unless enough (capacity) available.
  VM:  High load (no swap data sets).
       Use devices with low access times.
  Recommend: 3880-11, 2305, 3380.
```

Figure 11. FOIL 10--PAGE DATA SET REQUIREMENTS IN A SWAP ENVIRONMENT

Figure 11 describes the requirements for paging data sets when used in a swap environment.

When swap (or "all local" page) data sets are used in a TSO environment for swapping, they have some common characteristics:

1. The page data sets carrying swap traffic usually carry high paging loads, and more particularly long paging chains, which in turn makes the use of devices with fast data transfer rates desirable.

   Thus, the IBM devices with the highest data transfer rates available should be selected. The devices ordered by data transfer speed are: 3880-11, 3380, 3375, 2305, 3350.

2. Maximum parallelism is often desired for the best swap response time. In order to gain maximum parallelism, at least three or more independent paths should contain the data sets used for swapping. This goal can be accomplished by using 3880-11s or 3380s on multiple separate paths. Often better granularity can be achieved by using multiple 3380 actuators on (nondedicated) paths. Alternatively, if the installation wishes to isolate the swap paths, then at least two dedicated paths could be used. The paper evaluates numerically the effects of both nondedicated and dedicated paths.

3. Since ASM distributes the paging load nearly uniformly to most paging devices, string switching for dedicated paging devices is not desirable from a performance viewpoint, (Reference 3), even though it may be desirable from an availability point of view. Thus, in an MP envi-

ronment dedicated paging strings could be attached asymmetrically for best device performance.

This means attaching them to only one side of the MP. The result is better performance. If the single remaining path fails, MVS requests the system operator to vary another path on. The impact of path contention is greater on demand paging strings than on swap strings.

When the swap paging load is separated from the demand paging load on MVS, the demand paging data sets carrying that load have different requirements from the swap devices carrying swap load:

1. The load on demand paging devices in terms of pages/second, is significantly lower than the swap paging load. This happens in TSO, because the swap paging load represents most of the paging.

   Also demand paging (with one or two pages per I/O request) causes higher device utilization per page transferred, but for reasonable response times the device utilization has to be kept low.

2. For best performance, demand page-in response time, low access times are required. The device with the best access time is the 3880-11, thus whenever excellent demand paging response time is desired, the 3880-11 is the device of choice. If 3880-11 is not available, then other DASD (3380, 3375, 3350) can be used.

3. The 2305-2 drums provide good access times, but, because their capacity is limited, their use for demand paging is not advisable in an MVS/SP 1.3 environment. The drums could be used if enough such devices are available to contain the total required demand page space. The 3880-11 with its automatic migration and faster access times is the device of choice.

4. The situation on VM/370 is slightly different, because the only form of paging is demand paging, which may become very heavy. The best demand paging device is the 3880-11, and its use is recommended. VM/370 support of 3880-11s becomes available in March 1983. Until that time 2305-2s can be used for VM demand paging. In the mean time, installations with 2305s and both MVS and VM systems may wish to shift the 2305s to their VM systems. If neither 3880-11s nor 2305s are available then 3380s provide acceptable demand paging support.

```
Dedicated devices. (Recommended)
  Advantages:
   -Improved seeking.
    (No seeking on 3880-11 for hits)
   -Less interference with non-page I/O.
   -Better response times.

  Disadvantage:
   -Only small fraction of device used.

  Tradeoff:
   -Space on dedicated device is
    "wasted", but  performance gained.

   -Space can be used for data "seldom"
    used during interactive periods
    E.g. HSM archives,
         Sysgen libraries,
         Night Batch Databases.
```

Figure 12. FOIL 11--DEDICATING PAGING DEVICES

Figure 12 summarizes the effects of dedicating paging devices.

The question of dedicated paging devices often generates heated controversy.

Installations say "You mean to tell me that I have a 3380 actuator with an available storage space of about 630 MB, and you are using only 20 MB for paging and wasting the rest?" The distinction between optimal use of space and optimal use of an actuator from a performance viewpoint must be clearly established.

Dedicated paging devices provide significant advantages:

1. Reduced seeking results because ASM remembers the last cylinder used for paging purposes, but is unaware of other uses. Also, if only a paging data set of limited size is used, seeking is reduced because of the limited span the arm has to travel. There is no seeking for writes or read hits on the 3880-11.

2. On a dedicated paging device, paging cannot, at least as far as the device is concerned, interfere with non-paging I/O and vice versa. Therefore better paging and nonpaging response times are obtained.

Of course the use of only a relatively small portion of the device may be considered a disadvantage.

In order to overcome this difficulty, an installation may judiciously attempt to use the space on the paging devices for nonpaging purposes. Often, heavy interactive use of the system is restricted to the day time, thus heavy paging is also limited to the day time.

Consequently, the space can be used at night as archives for HSM, SYSGEN libraries,temporary space, or data bases for batch jobs run only at night. The risk is response time deterioration on an exception basis, when the data sets that were to be used at night are accessed during the day for some reason.

An installation may make a conscious tradeoff decision that performance (response time) is less important than the use of DASD space. If this is the case, then it is possible to distribute local page data sets across many actuators, and use the unallocated space for data bases or temporary space, despite the performance implications.

```
•Dedicated swap and page channels.
  Advantages:
   -Higher swap channel utilization
    possible: up to 50 %.
   -No interference with non-page I/O
    from long swap chains.
   -No interference with page I/O
    by e.g. VTOC search.
  Disadvantages:
   -If not enough swap paths, then
    Loss of parallelism, worse resp. time
  Recommend: Dedicated demand paging channels,
   avoids interference with demand paging.

 Distributed swap and paging devices
  Recommend: For swap devices (& for VM)
  Advantage: More parallelism, good response time.
  Disadvantages:
   -Interference by non-paging I/O.
   -Lower channel utilization (30-40 % maximum)
```

Figure 13. FOIL 12--DEDICATING PAGING PATHS

Figure 13 discusses the effect of dedicating paging paths.

Dedicating paging paths yields similar, but perhaps less significant beneficial results to dedicating paging devices.

This configuration concept avoids paging I/O interference with data base activity and vice versa. Dedication of a swap path in particular permits higher channel utilization (up to 50%), than a non-dedicated channel (30-40%). On the other hand, maximum parallelism for swapping requires multiple paths. In many practical cases the multiple paths can be made available only if they are not dedicated. Consequently, it is recommended that demand paging paths be dedicated if possible, but swap paths do not have to be.

In other words, distribute dedicated swap devices (and VM paging devices) across many paths.

Having done so, the old channel utilization rule of thumb (30-40% maximum channel utilization) should be adhered to.

2.8 SUMMARY OF RECOMMENDATIONS

- For best TSO response time:

  -increase main storage to the maximum.

  -separate swapping from demand paging

  -improve the paging subsystem

  -use multiple parallel 3880-11 or 3380 paths for swapping.

  -use 3880-11 for demand paging.

  PATH CAPACITY PERMITTING:

  -use storage isolation for reduced demand paging.

  -reduce logical swap think time for increased physical swapping.

- For best batch throughput:

  -do as outlined above.

  -use SRM controls.

Figure 14. FOIL 13--SUMMARY OF RECOMMENDATIONS

Figure 14 summarizes the recommendations. It is suggested that the best possible TSO response times i.e., sub-second response time desired by interactive users (Reference 2) should be the target of the installation, and this can be accomplished by:

1. Increasing main storage to the maximum to minimize paging.
2. Separation of swapping from demand paging by use of swap data sets.
3. Improvement of the paging subsystem by using fast devices.
4. Use of multiple parallel 3880-11 or 3380 paths for swapping.
5. Use of 3880-11s for demand paging.
6. If path capacity permits then the system can be altered so that more swapping and less demand paging occurs:
   a. Maximizing swapping and minimizing demand paging by the use of storage isolation.
   b. Maximizing swapping and minimizing demand paging by the reduction of logical swapping (LSCTMTE).

3.0 DEVICE CHARACTERISTICS

DEVICE TYPE	3350	2305	3375	3380
DATA XFER RATES MB/S	1.2	1.5	1.8	3.0
TIME OF ROTATION MS	16.7	10	20	16.7
AVERAGE LATENCY MS	8.3	5	10	8.3
MAXIMUM SEEK MS	50	0	38	30
AVERAGE SEEK MS	25	0	19	16
MINIMUM SEEK MS	10	0	4	3
NO OF PGS PER TRACK	4+	3+	8	10
NO OF TRACKS PER CYL	30	8	12	15
NO OF SLOTS PER CYL	120	24	96	150
NO OF CYLINDERS	555	96	959	885
CAPACITY MB/ACT	319	11.2	410	630
NO PG SLOTS PER ACT *	65K	2470	65K	65K
NO SWAP SETS/ACT	5.5K	190	7672	10K

* MVS supports a maximum of 65 k slots

Figure 15. FOIL 15--CHARACTERISTICS OF PAGING DEVICES

For convenience Figure 15 summarizes the physical characteristics of IBM paging DASD. Observe that 3380s have the highest data rate (3MB/sec) which makes them excellent swap devices. 3375s also have a high data transfer rate, so they can be also used for swapping. Also, the minimum seeks on 3375s and 3380s are so low that seeking does not heavily interfere with their use as swap devices.

In contrast, the time for one revolution is quite high for all of the moving head DASD. Realizing that the duration of initial latency is half the revolution, which represents a significant portion of the access time, it becomes clear why DASD does not provide the best demand paging response time.

Observe that the 2305-2 device has only limited space capacity, which makes its use inappropriate as a demand paging device on large MVS systems where large page space capacities are required.

The number of slots and swap sets per actuator listed in the foil is the maximum definable for an actuator. The maximum comes from either the actual physical limitations of the device (e.g., 2470 slots or 190 swap sets for the 2305), or from ASM, which limits a local page data set size to 65K slots.

3.1 THE 3880-11 BUFFERED PAGING DEVICE

```
     ---------------------
    | Channel             |
     ---+------------+----
        |            |
        |            +--Other C.Us
 -------|-------
|       |       |   Control Unit
|    ---+-------|------------
|   | Cache     |            |
|    ---+-------|-----+------
|       |       |     |
|       |       |     +--Other string
|       |       +-------------
|       |                     |
|       +------------         |
|       |            |        |
|   +---+---+    +---+---+    |
|   | 3350  |    | 3350  |    |
|   +-------+    +-------+    |
|      Back store          |
 ----------------------------
```

Figure 16. FOIL 16--THE 3880-11

Figure 16 shows the schematic of a 3880-11 paging device with two 3350 back store actuators. The cache memory (8.3 MB) supports only the paging activity, in conjunction with the two 3350 actuators. The other half of the 3880 Control unit is independent, supporting 3330 and 3350 DASD.

Pages that cannot be accomodated in the cache are migrated with an LRU algorithm to the back store. Only pages, that are marked changed, have to be written out onto the back store.

The dynamic migration capability of the device enables it to handle large page spaces.

The 3880-11 has multiple exposures, so that it can accept multiple page requests from the system at the same time.

3.2 3880-11 DEVICE CHARACTERISTICS

```
CHANNEL DATA XFER RATE MB/SEC        3

CACHE DATA XFER TIME PER PAGE     MS 1.3
CACHE LATENCY TIME PER PAGE       MS 0.7
CACHE BUSY TIME PER PAGE          MS 2

ONE DATA TRANSFER AT ONE TIME.
DASD ACCESS IS ADDITIONAL TO CACHE
 TIME, IF A MISS OCCURS.
MULTIPLE EXPOSURES TO PREVENT HITS
 FROM WAITING FOR MISSES.
TWO DEVICES ATTACHABLE TO A CHANNEL.
LESS SENSITIVITY TO RPS MISSES.

HIT RATIO = HR = ratio of pages found in
 cache to total paging rate to 3880-11.

 HR= .9 ->   Miss ratio = M = 1 - HR = .1
 I.E. ABOUT 10% OF TIME DASD IS ACCESSED.
```

Figure 17. FOIL 17--3880-11 DEVICE CHARACTERISTICS

Figure 17 discusses the device characteristics of the 3880-11.

The 3880-11 buffered paging device was designed with the objective of being able to produce excellent paging response times on large systems.

It has a 3 MB/sec data transfer rate on the channel, making it into a good swap device.

In this paper, the 3880-11 page transfer time is considered to be 2 ms, (2.4 ms transfer time was announced). This is obtained by the data transfer time of about 1.3 ms on the channel, and cache latency time of about 0.7 ms. This low access time makes the 3880-11 an excellent paging device.

The device can do one data transfer activity at one time; either from the back store to the cache or from the cache to the channel.

The multiple exposure capability is provided, so that when a page is not found in the cache (miss), requiring accesses to the back store, it can still proceed to honor another request, which may not require back store access.

Hit ratio is defined as the proportion of the pages accessed that were found in the cache; alternatively it can be stated as the number of pages requested that were found in the cache vs the total number requested. Miss ratio is obtained by subtracting the hit ratio from 1. A miss ratio of 0.1 means that 10% of the time the referenced slot is not found in the cache, it is missed, and a DASD reference may result (immediately or later). As a first approximation, a 10% miss ratio implies that 10% of the time the back store DASD is referenced.

Whenever ASM requests a (nonswap) page write, it does so because the page is changed.

Consequently all page write requests to the 3880-11 involve a changed page. If the original copy of the changed page is in the cache already, it is called a write hit, otherwise it is a write miss. It is assumed, that in case of write misses there is always an empty slot in the cache, i.e., cache management can always provide an empty slot by freeing up unchanged slots or by appropriate migration or destaging of changed pages, and keeping a pool of available slots. No distinction is made between write hits and misses in terms of modeling; it is assumed that the cost of finding an empty slot is negligible i.e., both write hits and misses take 2 ms. A write hit requires a change to a page still in the cache, this means that multiple writes to the cache may change the same slot in the cache repeatedly. If the cache LRU algorithm decides to cause the migration of a changed slot, then a write to the back store results. Back store write requests do not directly impact response times.

The more interesting phenomenon is that of the read requests, since they are the only ones directly contributing to response times. A read hit can be directly serviced with a service time of 2 ms, but a read miss requires a read first from the back store into the cache, followed by a read from the cache to the channel.

In most of the calculations performed subsequently it is assumed that the number of reads and writes to the 3880-11 is about the same. This is certainly the case for swap devices. It is further assumed for simplicity, that read hit ratio i.e., read hit rate/read request rate, is the same as the overall hit ratio.

With these assumptions, writes occur to the DASD slightly more frequently than reads.

Each read miss requires an immediate DASD access and a write miss may require a deferred write to the back store. Note that many read or write hits may occur to the slot first impacted by a write miss, before it is eventually aged out of the cache. In addition, if the system changed a previously read page, i.e., a read miss was followed by a write hit, this may also require a deferred write to the back store. Therefore a DASD access is caused by:

1. A read miss. (Immediate)

2. A write miss. (Deferred, other write hits may still follow).

3. A write hit following a read miss. (Deferred, other write hits may still follow).

If the miss ratio is M, and the paging rate to the device is P, made up of half reads and half writes, then with our assumptions, the read miss rate is (P/2) x M and empirically (communication from Dr. Goldfeder) the destaging rate with these assumptions is (P x M / (M+1)). Consequently the DASD ratio, i.e., the number of paging rate to the back store divided by the paging rate to the 3880-11 is:

DASD ratio = (M / 2) + (M / (M+1))

The following table can be built:

Miss Ratio M	Read Miss Ratio M / 2	Destage Rate M/(M+1)	DASD Ratio	DASD Ratio / Miss Ratio
.05	.025	.047	.07	1.5
.1	.05	.09	.14	1.4
.2	.1	.17	.27	1.35
.4	.2	.29	.49	1.2

The DASD Ratio-Miss Ratio quotient is in the 1.2-1.5 range for practical miss ratios.

The miss ratio therefore provides an approximation to the paging load addressed to the back store, but the actual load is somewhat higher. For an example, assuming a paging load of 200 pages per second to the 3880-11 and a miss ratio of 0.1, a DASD ratio of 0.14 is obtained, yielding a paging rate to the back store of 28 pages per second. As a first approximation, the miss ratio could be used directly, yielding a paging rate to the back store of 20 pages per second.

Caution: if a measurement of 3880-11 "cache effectiveness" is obtained, that may be defined as the paging rate to the 3880-11 without requiring back store I/O, divided by the total paging rate. The cache effectiveness so defined is the same as (1 - DASD ratio). Thus if a cache effectiveness of 86% is observed, then the DASD ratio is 14%, corresponding to about a hit ratio of 90% and a miss ratio of 10%. (Communication from Sandor Roehlich).

In the examples, the DASD ratio is used to calculate the back store page load.

4.0 PAGING SPACE SIZE AND RATE CALCULATIONS

Having done a qualitative analysis of the paging environment, quantitative analysis can begin. In this section paging space size and paging rate requirements are calculated.

4.1 PAGING SPACE SIZE REQUIREMENTS

```
•Swap Data Sets

 Swap Group Size, SGS = 36 slots
  (It can be set to 48, or even 60)

 Swap space size=
  Number of Users x Swap Group Size
    SDS = NU x SGS

  e.g., for 100 users
  SDS = 100 x 36 = 3600 slots =
      = 3600 x 4K = 15 MB

•Demand Page Data Sets

 Paging space size=
  Number of users x 120 pages

  e.g., for 100 users
  SDS = 100 x 120 = 12000 slots =
      = 12000 x 4K = 48 MB = 50 MB
```

Figure 18. FOIL 19--PAGING SPACE SIZE REQUIREMENTS

Figure 18 shows the calculation of swap and demand page space sizes.

4.1.1 SWAP SPACE SIZES

TSO installations can determine the swap group size from their RMF report. Usually the swap-out size is greater than the swap-in size, since swap-out size includes swap trim. If there is a significant difference between the two numbers, that may be an indication that storage isolation was not used, or its size was inadequate. In any case, if swap configuration and specifically channel capacity permits , it makes sense to round the swap-out size upward to an integral multiple of 12 e.g., from 37 to 48, and storage isolate the trivial TSO performance period at that level. Having done that, swap-out size approximates swap-in size, or the swap group size, with few demand page-ins for trivial TSO transactions. An increase in swap group size requires a corresponding increase in swap configuration capability however.

In this paper it is assumed that a swap group size of 36 is used, for both swap-ins and swap-outs. Once this assumption is made, the swap space size can be calculated if the number of active users is known. Thus, with 100 active users, the swap space size can be defined as (100 x 36) = 3600 slots, or about 15 MB. The actual swap space allocated should be about twice as much, to allow for momentary overloads, batch swapping, etc.

With increasing and more sophisticated use of full screen terminals, there is a tendency towards higher swap group sizes, thus it is also reasonable to assume a swap group size of 48 slots, requiring 20 MB of swap space for 100 TSO users.

4.1.2 DEMAND PAGE SPACE SIZES

As a rule of thumb (on the basis of experience) a demand paging space (swap space on separate swap data sets) space of about 120 pages per user is required, i.e., for 100 users the size of the space needed is about 50 MB. The actual page space allocated should be about two to three times as much to allow for momentary overloads, VIO, etc. Also, without spare space available contiguous slot allocation is unable to perform its function of contiguous allocation. On the other hand allocation of more than <u>three</u> times the space required may result in excessive seeking.

Observe that demand page space required is much larger in size than the swap space.

4.2 SWAP PAGING RATES

```
                             NO OF ACTIVE USERS
TRANSACTION RATE, TR= ------------------
                             AVERAGE THINK TIME

MAXIMUM SWAP PAGING RATE =
2 x SWAP GROUP SIZE X TRANS. RATE

SWAP PAGING RATE =
2 X SWAP GROUP SIZE X TR X (1-LOG.SWAP%)

E. G. NUMBER of USERS = 100
      SWAP GROUP SIZE =  36 slots
      THINK TIME      =  25 seconds
      LOGICAL SWAP    =  17%

 SWAP PAGING RATE=
  2 x 100 x 36 x .83 / 25 = 240 pgs/sec
```

Figure 19. FOIL 20--SWAP PAGING RATES

Figure 19 illustrates the calculation of swap paging rates.

An easy way of estimating the TSO transaction rate is to use Little's theorem (Reference 1) and divide the number of users by the average think time. This is an approximation, since response time should also be in the denominator. Thus, 100 users with an average think time of 25 seconds yields a transaction rate of 4 transactions per second. Please note that the average think time of such high value includes people who have coffee, talk on the telephone, and are noninteractive for a variety of reasons, but are logged on. There is also evidence that the reduction of response time leads to a reduction in the value of think times, i.e., user productivity goes up. This effect is not taken into account in what follows, and must be evaluated for the individual installation.

Given the transaction rate, it is possible to estimate the maximum possible swap paging rate, by multiplying the transaction rate by the swap group size and by a factor of 2, to allow for both swap-in and swap-out. Thus, for a transaction rate of 4 transactions per second and a swap group size of 36 the swap paging rate becomes (4 x 36 x 2) = 288 pages/sec.

In most practical cases, however, logical swapping occurs and this reduces the physical swapping rate. In this paper a logical swap rate of 17% is assumed, somewhat arbitrarily. A different logical swap ratio would impact the calculated results, but would not change the methodology. This ratio was selected for ease of calculation. Then the actual swap paging rate reduces to (288 x (1-0.17)) = 240 pages/sec.

Each installation can perform these calculations using data from the "heaviest" period of the day.

4.3 DEMAND PAGING RATES

```
->Difficult to calculate.<-

But: minimize "demand" paging rate by

- Using storage isolation
  forces swapin = swap-out rate.
  little swap-out trim.
  Minimizes TSO demand paging.
  (3-5 page faults per transaction)

- Minimize overinitiation

- Use only paging rate control in SRM
  to limit MPL. (RCCPTRT)
  -> Consistency

- Limit: 30-80 pages/sec on 3033 UP
```

Figure 20. FOIL 21--DEMAND PAGING RATES

Figure 20 illustrates the calculation of demand paging rates.

Demand paging rate manifests itself in its clearest form in those cases where the swap paging is directed to swap data sets, and there is minimal use of VIO. The remaining paging load on the local page data sets can then be viewed as demand paging. It is very difficult to estimate this paging load, but it is possible to measure it and control it. If storage isolation is used, then swap trim is minimized. In a system without serious main storage constraints, i.e., where some logical swapping occurs, or where the UIC (unreferenced interval count) has a value of about 20, a TSO trivial transaction can encounter up to 3-5 page faults after storage isolation (as observed on some systems). Consequently the demand paging load is at least (2 x transaction rate x 3) for such systems, assuming 3 demand pages read and written per transaction. In addition, nontrivial TSO transactions and batch can also cause significant demand paging.

By limiting overinitiation and by using the SRM page fault control mechanism (RCCPTRT), the multiprogramming level in the system, and consequently the demand paging rate can be controlled. The advantage of doing this is consistency of response time. (Reference 4).

As an example, on a 16 MB 3033 TSO-batch system (100 users) the "demand paging" rate was controlled in the 30-60 pages/second range. This range (maximum) could be applicable to most systems with CPU power equal to or slower than a 3033 UP with storage limitations. If main storage is increased (for example to 24 MB), the demand paging rate can be significantly reduced. As a first approximation, it can be assumed that for CPUs faster than a 3033 UP, the demand paging rate can be linearly scaled upwards with the instruction execution rate of the CPU, assuming that the number of interactive users is also linearly increased.

5.0 BASIC CONCEPTS OF I/O ACTIVITY AND ASSOCIATED FORMULAE

In this section a basic explanation of the I/O activity is given, together with its modeling representation. The modeling used is an analytic model using a fairly standard M/D/1 formulation. It is not necessary to understand the modeling methodology in detail to use the formulae, thus the following detailed discussion can be skipped, and the reader can continue with the examples directly.

"M" means exponential arrival of requests to the devices.
"D" means fixed service time at the servers (devices).
"1" means single queue assumed.

Other sources have used M/M/1 formulation (exponential distribution of service time), but system measurements indicate to this writer that a somewhat better match with measured results can be obtained for at least 3380s and 3880-11s with an M/D/1 formulation. The basic difference in calculations between M/M/1 and M/D/1 models is that an M/D/1 model essentially provides a "wait in queue time" half as long as the M/M/1 model. Furthermore, it is also true, that when swapping and demand paging are separated, the service times on the devices do not vary exponentially, and a fixed service time assumption makes more sense than the assumption of exponentially distributed service times. The service time variations introduced by seeking is also significantly reduced for these modern devices.

The question of exponential arrivals has to be addressed. It is likely that demand paging arrivals are randomly distributed, i.e., the assumption of exponential arrivals is a reasonable approximation. It is also clear that the arrival of swap sets is not randomly distributed, since each swap group contains multiple swap sets, which arrive in a clustered manner. The arrivals of <u>swap groups</u> can be considered randomly distributed however. Consequently, the assumption of exponential distribution of swap set arrivals is really valid only if there are enough independent swap devices and paths, so that each swap device is affected only by one swap set from each swap group. If this is not the case, (for example, there are two swap devices on the same path), the assumption would be incorrect. It is the author's belief that the calculations assuming random arrivals yield reasonable results.

Consider for example a single swap device on a path, which is confronted by the clustered arrival of two swap sets. The two swap sets can be represented as a larger entity, a pseudo swap set consisting of two swap sets (24 pages), and the arrival of these entities is distributed exponentially once more. In this case the calculations could be validly performed on pseudo swap sets of 24 pages. The formulae used would result, not surprisingly, in response times about twice as large as for a single swap set. Thus, an approximation could be made of calculating single swap set response times, and doubling it for the evaluation of the response time associated with the pseudo swap set. The section "Swap Measurements" discusses this problem in more detail.

Two swap devices on the same path represents the worst case from a modeling viewpoint, since RPS misses occur, and the problem of clustered arrival is distributed between the two devices. The clustered arrival problem is almost identical to the case of single device. The calculation of the RPS miss (with clustered arrivals) presents a special problem. The intention of this paper is to obtain a reasonable evaluation of device capabilities, and no special effort is made to handle this special case differently from the general case. The approach represents an approximation.

In this paper, for simplicity the arrival of swap sets is considered as being exponentially distributed, so that the swap set response times are valid only if there are a sufficient number of independent devices. The overall swap response times are correct however because of the exponential arrivals of the swap groups. A later section of the paper illustrates the consequences regarding the measured data.

It should be emphasized that the numbers calculated by models are modeling numbers, not measured numbers, and while an attempt was made to "validate" them by comparing them to measured numbers, modeling at best represents an approximation of reality. Reality does not necessarily follow modeling assumptions. Furthermore, modeling makes simplifying assumptions, which often make the models suspect. Consequently, utilization and throughput figures presented may match reality more closely than the response times calculated. On the other hand, the modeling results represent relative evaluation of the devices presented, even if the absolute values prove to be somewhat inaccurate.

5.1 COMPONENTS OF I/O ACTIVITY

Conceptually:

Queue, Q - due to device busy

Command delay, CD - SIOF,CC¬=0, path busy

Seek

Seek reconnect, SRO (CCW fetch),
due to path busy (not on 3880s)

Latency = half revolution

RPS miss and reconnect

Search (start 3-5 sectors before)

Data Transfer

Figure 21. FOIL 23--COMPONENTS OF I/O ACTIVITY

Figure 21 defines the following components of I/O activity:

1. Device queue.

   A device queue is created when the system tries to access a device and the device is found to be busy. In MVS, the device queue explicitly exists for swap devices: for each swap device two (MVS/SP 1.3.0) or three (MVS/SP 1.3.1) requests can be initiated because of the number of IORBs available; any further requests are queued in ASM. Nonetheless, in most cases, only one request can be in progress at one time (exceptions are 2305s and 3880-11s). Paging requests for a local paging device are dynamically appended to the device channel program, but this only says that the queue explicitly occurs at the device. The device queue delay, Q, is calculated from the following (M/D/1) formula:

   Q = (DST x DUT) / (2 x (1-DUT)), where

   DST = Device Service Time
   DUT = Device Utilization

   If an M/M/1 model were used, the factor "2" would not appear in the denominator.

2. Command Delay.

   The path (channel or control unit) may be busy servicing another request (from another device on the same path, if any), so that the start of the I/O request which has to be passed through the path is delayed. This is manifested in the system for a swap device by a condition code of non-zero to the SIO. This delay does not occur for local paging devices as frequently, because of the dynamic appending of channel requests, but for consistency it was calculated for all DASD. The average delay depends on the channel busy state caused by other devices, and the average service time for other devices. On average, the request sees half the duration of another channel service request. The command delay is calculated from the following formula:

CD = Average Channel Service Time x RCB / 2, where
RCB = Relative Channel Busy (See below).

3. Seeks.

```
2.1 + 0.9 x (Number of cylinders) ** .5 --- 3380

2.9 + 1.1 x (Number of cylinders) ** .5 --- 3375

7.8 + 1.8 x (Number of cylinders) ** .5 --- 3350
```

Figure 22. DASD SEEK FORMULAE

Figure 22 defines approximate formulae for seek times. Thus, for a 40 cylinder data set size on a 3380 one obtains a maximum seek time of:

2.1 + 0.9 x (40) ** .5 = 7.8 ms

In the calculations an average seek size of one cylinder is assumed for swap data sets. Seeks on local page data sets are larger, but depend on the data set size.

4. Seek Reconnect.

Seek Reconnect occurs on 3830 control units, not on 3880s. It means that after the initial DASD seek has been completed, another channel command has to be fetched through the channel (usually the set sector command). On 3880s the second command is prefetched, so this delay does not occur. Since on most paging devices only 50% or less of the seeks are non-zero seeks, the delay occurs only about 50% of the time. When it occurs, it does so, because the channel is busy, so its formula is similar to the command delay, except it occurs only 1/2 the time.

SRO = CD / 2 (Not on 3880s)

5. Latency.

On average, in order to arrive at a page requested, a DASD will have to complete half a rotation, called latency; this keeps the device busy.

Latency = half a revolution

6. Search.

The set sector command has to be issued 3-5 sectors prior to the desired sector. This causes a channel time which is longer than could be calculated from the data transfer time by itself. In the calculations presented below this is approximated by increasing the single page data transfer time slightly (as shown below).

7. Data Transfer.

Page transfer times are calculated and rounded:

```
3380: 4096 / 3   MB / sec =  1.3 ms --> 1.5 ms
3375: 4096 / 1.8 MB / sec =  2.3 ms --> 2.6 ms
3350: 4096 / 1.2 MB / sec =  3.3 ms --> 3.6 ms
2305: 4096 / 1.5 MB / sec =  2.7 ms --> 3.0 ms
```

swap set transfer times are calculated by counting the number of pages per track and the tracks per swap set, and multiplying the product by the rotation time. This accounts for inter-record gaps and partial track usage.

```
3350: 3   revolutions = 3   x 16.7 ms = 50 ms
2305: 4   revolutions = 4   x 10   ms = 40 ms
3375: 1.5 revolutions = 1.5 x 20.1 ms = 30 ms
3380: 1.2 revolutions = 1.2 x 16.7 ms = 20 ms
```

5.2 CONCURRENCY FORMULAE

```
Multiple concurrent swap I/Os

-Calculate single path response time
-Adjust for number of swap sets/paths
-Maximum is NO of swap sets x resp.time

SWAP RESPONSE TIME =                RT
SWAP SET RESPONSE TIME            = SRT
NUMBER OF SWAP SETS IN GROUP      = STS
NUMBER OF SWAP SETS PER CHN       = SPC
CONTROL UNIT UTILIZATION          = CUT
NUMBER OF CHANNELS                = NCH

RT = SRT X (SPC+(STS x CUT ** (NCH-1)))
RT = Lesser of (RT, STS x SRT)


NCH=3, SRT=50, STS=4, CUT=.3, SPC=1.33
-> RT = 50 x (1.33+(4x(.3) ** 2)) =  85

--> EXPONENTIAL DECREASE WITH CHANNELS
```

Figure 23. FOIL 24--CONCURRENCY FORMULAE

Figure 23 quantifies the effect of concurrency, i.e., the effect of transmitting multiple swap sets concurrently on several channels. An empirical formula, based on a simulation model, was developed.

First, the swap set response time is evaluated by the methods yet to be demonstrated. The number of swap sets in the swap group and the number of channels used for swapping clearly influences the overall swap response time. Thus, if only one device is used for three swap sets, then the swap response time is three times the swap set response time, this represents the worst case. On the other hand, if three devices on three separate paths are available, and the paths are free, then the swap response time is the same as the swap set response time, and this is the best case. The formula used must yield correct results for these boundary conditions. In general, the paths are not always free, and while some overlap is possible, complete overlap is unlikely, in part because of the lack of synchronism between channels, devices and swap set sizes (partial swap sets do occur). The more the paths are loaded, the less likely the overlap becomes.

Simulation indicates that the response time elongation can be calculated by taking the larger of the path (or the device) busy fraction to the power (number of channels - 1), multiplied by the number of swap sets and added to the natural elongation. The natural elongation occurs when the number of swap sets per channel is greater than 1.

Assume:
- -swap set response time = 50 ms
- -Four swap sets
- -Maximum channel utilization / C.U. utilization / device utilization = 30%

NUMBER OF CHANELS (NCH)	SWAP SETS PER CHANNEL (SPC)	SWAP RESPONSE TIME (RT)
1	4	200
2	2	160
3	1.33	85
4	1	55

Figure 24. FOIL 25--PARALLELISM AND SWAP RESPONSE TIME

If the calculated swap set response time were 50 ms and this were to be obtained on 1, 2, 3, and 4 channels for a swap group with four swap sets, with the greatest of channel, control unit and device utilization kept at 30%, the numerical results shown in Figure 24 could be established showing the effects of parallelism on response time. The most significant benefits are obtained when the number of paths increases from 2 to 3. As the number of paths increase, little additional benefit is gained from concurrency, even though benefits may accrue in a real situation because of lessened path utilization. In the calculation above a fixed utilization of 30% was assumed.

5.3 CONCEPT OF RELATIVE CHANNEL BUSY

Consider a swap channel with 2 devices.
Let channel busy = 50%
Relative channel busy = RCB = 1/3

When device D1 tries to reconnect, it
finds channel busy 25% of time for D2,
finds channel free 50% of time

$$RCB(D1) = \frac{CH.BSY(\neg D1)}{CH.BSY(\neg D1)+CH.FREE} = \frac{25}{25+50} = \frac{1}{3}$$

CHN FREE	D1 BUSY	D2 BUSY
50%	25%	25%

$$RPS = 16.7 \times \frac{1/3}{1-(1/3)} = 8.4 \text{ ms}$$

Figure 25. FOIL 26--CONCEPT OF RELATIVE CHANNEL BUSY

Figure 25 introduces the important concept of Relative Channel Busy (Reference 5). RPS misses or command delays occur as the device does not find the channel (the control unit, the path) free because other devices keep it busy. This means that the channel busy factor customarily used in the RPS miss and command delay calculations should be the relative channel busy (RCB).

The concept can be best understood by considering an example:

A channel is 50% busy, because each of two devices keeps it busy 25% of the time. Thus, the channel is free 50% of the time, and a device (D1) sees a channel busy of 25% (due to the other device, D2) and 50% channel free time. The relative channel busy then as seen by device D1 is the channel busy time due to other devices (D2) divided by the the total time when device D1 is not busy itself. The relative channel busy then is

25 / (25+50) = 1/3

5.4 BASIC CONCEPTS (REVIEW)

```
M/D/1 Queue -    Q  = (DST x DUT) / (2 x (1-DUT))

Relative Channel Busy, RCB: as seen by device

Command delay:  CD = Channel service time x RCB /2

Seek (One cylinder for swap data sets)

Seek reconnect: SRO = CD/2 (not on 3880)

Latency:        LAT = ROT / 2

RPS reconnect:  RPS = (ROT * RCB) / (1-RCB)
Search: a few sectors

Transfer data: (3380 examples)
page: (4096 / 3x10**6) + = 1.5 ms

swap: 10 pages/track = 16.7 ms
      but swap set = 12 pages =
      16.7 + (2/10) x 16.7 = 20 ms
```

Figure 26. FOIL 27--BASIC CONCEPTS (REVIEW)

Figure 26 summarizes the formulae used in the calculations.

6.0 EXAMPLES

In order to demonstrate the use of these techniques, a few examples are presented. The examples encompass both swap and demand page data set evaluation.

6.1 SWAP EXAMPLES

In all the swap examples that follow, it is assumed that there are 12 pages per swap set, and each swap set is transmitted by one SIO. (This assumption will be incorrect when partial swap sets exist, or when the number of swap actuators is smaller than the number of swap sets in a swap group and the system is heavily loaded).

6.1.1 2 3380 SWAP ACTUATORS ON CHANNEL

```
12  PAGES PER SIO
SIOS  PER SECOND PER CHANNEL:  20
CONTROL UNIT OVERHEAD              2.6 MS
DATA TRANSFER TIME                20   MS

CU. UTIL.:     20 x 22.6=452 ms/s = 45%
REL.CH.UT.: RCB = 22.5/(22.5+55) = 29%

COMMAND DELAY, CD = 22.6x.29 / 2= 3.3 MS
OVERHEAD TIME                      2.6 MS
SEEK TIME                          3   MS

RPS MISS = 16.7 x .29 / (1-.71) = 6.8 MS
LATENCY TIME                       8.3 MS
DATA TRANSFER                     20   MS
DEVICE SERVICE TIME               40.7 MS
DEVICE UTILIZATION = 20x40.7/2 = 40.7 %

Q (M/D/1)=
 =40.7 x.407 / (2 x (1-.407))  = 14   MS

SWAP SET READ
RESPONSE TIME= 40.7 +14  + 3.3 = 58   MS
```

Figure 27. FOIL 29--2 3380 SWAP ACTUATORS ON CHANNEL

Figure 27 provides an evaluation of two 3380 swap actuators. It is assumed that two 3380 actuators are used per channel (or control unit, used interchangably in the example), both for swapping. It is assumed that the SIO rate is 20 SIOs per second. The control unit overhead (protocol) is 2.6 ms and the data transfer time for a swap set is 20 ms, for a total of 22.6 ms. The control unit utilization is obtained by multiplying the SIO rate with the control unit utilization per swap set. A control unit utilization of 45% (just below the magic 50%) is obtained.

One device sees 22.5% channel busy and 55% channel free. Thus, the relative channel busy is obtained as 22.5 / (22.5+55) = 29%. A one cylinder seek of 3 ms is assumed and the service time of a device is calculated by adding the overhead time, seek time, RPS miss, latency, and the time for the data transfer. The service time obtained (40.7 ms) multiplied by the SIO rate associated with one device (10) yields the device utilization, (40.7%).

The device queue is calculated by taking the device service time, multiplying it by the device utilization, dividing it by 2 times (1 - device utilization); and a queue wait time of 14.3 ms is obtained.

Response time is then calculated by adding the device service time, queue wait time and command delay.

The command delay is insignificant compared with the the queue wait and service times. The service time itself consists of about half data transfer, half associated overhead. Data transfer represents one third of the response time.

6.1.2 3880-11 SWAP DEVICE WITH 2 3350S

```
12 PAGES PER SIO. CACHE BUSY: 2 MS / PG
HIT RATIO= .95, MISS RATIO = .05
 DASD RATIO = .07

SIOS PER SECOND = SPS =        20
TOTAL PAGING RATE = 20 x 12 = 240 PGS/SEC
SIO'S TO 3350S: = SPS x DR = 20x.07=1.4
3350 PGNG PER ACT. 1.4 x 12 /2 = 8.4 PG/S
CACHE BUSY - CHANNEL= (12 x 2) = 24   MS
CACHE BUSY - 3350=.07 x (50+1) = 3.6  MS
CACHE BUSY TIME/SIO: 24 + 3.6  = 27.6 MS

CONTROL UNIT UTIL. 20 x 27.6   = 55   %
WEIGHTED SERVICE TIME, DST:
 CACHE BUSY TIME / 1.047       = 26.4 MS
QUEUE WAIT, Q = (DST x DU) / 2 x (1-DU)
 = 26.4 x .55 / (2x(1 - .55 )) = 16.1 MS
3350 RESPONSE TIME- (ABOUT)    = 91   MS

SWAP SET READ RESPONSE TIME:
 24 + (.05 x 91) + 16.1        = 44.6 MS
```

Figure 28. 3880-11 SWAP DEVICE

Figure 28 presents an example of 3880-11 swapping. A swap load of 240 pages/second is used going to a 3880-11 swap device with two actuators. A hit ratio of 95% is assumed, which is achievable, as measured on an internal IBM production system, when the active swap data set size does not exceed 16 MB. In other words, coverage does not fall below 50%. Coverage is defined as the size of the cache, 8 MB - divided by the supported page space, 16 MB in this case.

In calculating the SIO rate to DASD, it was assumed that swap sets of 12 pages maintain their unit character. In fact units of eight are used, making DASD utilization slightly worse. At these high hit ratios the effect of the back store on performance is negligible. The SIO rate to the 3880-11 is multiplied by the DASD ratio of .07, this gives us a DASD rate of (20 x 0.07) = 1.4 SIOs per second or 16.8 pages per second, or 8.4 pages per second per 3350 actuator. Note that the pages transferred are consecutively located on the DASD.

The calculation of the cache busy per SIO is obtained by adding the cache/channel transfer busy time of 24 ms for 12 pages to the proportionate cache/3350 busy time. The 3350 busy time is 50 ms data transfer per swap set and 1 ms for protocol for a total of 51 ms. Since the DASD ratio is .07, the cache/3350 busy time per SIO on the average is (0.07 x 51) = 3.6 ms. Total cache busy time of 27.6 ms yields a 55.2% device (or control unit) utilization which appoaches the limit by our ground rules.

In order to calculate the queue time the average service time must be obtained. Three different activities have to be considered:

1. Channel/cache/backstore transfers: .05 I/Os per SIO
2. Channel/cache transfers: .95 I/Os per SIO
3. Cache/backstore transfers: .047 I/Os per SIO
4. Altogether 1.047 I/O activities per SIO.

The average service time (which is the length of an I/O activity by the cache) is obtained by dividing the cache busy per SIO by the number of I/O activities, i.e., 1.047. Notice that the average cache busy time could have been taken, (just as easily, 27.6 ms instead of 26.4 ms) without making too big an error.

The queue wait is calculated, assuming the cache is a single server. This approach ignores the effect of the multiple exposures, but in the case of swap sets, the dominating factor is the data transfer time and the queues resulting from heavy data transfer activities through the channel/cache path. Multiple exposure helps in the case of a read hit following a read miss. With the high hit ratios observed few misses occur.

The device (swap set) read response time is obtained by adding the service time required for cache/channel transfers (24 ms) to the queue wait calculated, and then adding the portion contributed by read misses. The read miss portion is calculated by establishing the 3350 response time (91 ms, calculated below) for a read miss and multiplying it by the read miss ratio (0.05).

```
2 3350s ON A SWAP 3880-11

 95% HIT RATIO. MISS RATIO = .05
 TOTAL PAGING RATE =          240 PGS/SEC
 DR = M/2 + M/(M+1)
 DR(READ)           = .05 / 2       =.025
 DR(WRITE)          = .05 / 1.05    =.045
 DASD RATIO:        = .025 +0.045   =.07
 PGS TO 3350s: 240 x .07      = 16.8 PGS/S
 SIOS TO 3350S: = SPS x DR = 20x.07=1.4

 CACHE  BUSY (CHANNEL) 240 x 2     = 48  %
 CACHE-3350 PG: (16.8 / 12) x (50+1)= 7.1%

 RCB(3350) =(.48 +.036) / (1-.036) = 53  %

 OVERHEAD TIME      (3350-CACHE)        1  MS
 SEEK TIME                             10  MS
 RPS MISS(P)=16.7 x .53  / (1-.53 )= 18.8MS
 LATENCY TIME                      =  8.3MS
 DATA TRANSFER FOR SWAP SET         = 50 MS

 SWAP SET RD TIME=1+10+18.8+8.3+50  =88.1MS
 3350 DEV UT  = (1.4 / 2) x  88.1  =  6  %
 3350 Q    = 88.1 x .06 / (2 x .94 )=  2.8MS
 3350 SWAP SET RD RT  = 88.1 + 2.8  = 90.9MS
```

Figure 29. 3350 SWAP SET RESPONSE TIME ON A 3880-11

Figure 29 addresses the modeling of the back store device. The back store device (the 3350) is to be handled as an additional queueing problem, similar to the one solved in the preceding example for 3380s. In this case, the load to the 3350s is known, and the cache is represented as a control unit with known utilization. The "relative channel busy" can be calculated by observing that the busy time seen by a back store device is the sum of cache/channel busy time (48%), plus the cache busy time due to the other actuator (3.6%). An average seek of one cylinder (10 ms) can be assumed, and the RPS miss can be calculated, using the RCB. This is followed by the calculation of the service time for the 3350, which consists of 51 ms data transfer and protocol, 10 ms seek, 8.3 ms latency and the RPS miss of 18.8

ms for a total of 88.1 ms. The device utilization is the service time (88.1) times the SIO rate per actuator (1.4 / 2). The device utilization is about 6%. The device queue is calculated as about 3 ms. The 3350 response time can be calculated as the sum of the service time and the queue wait, yielding a 3350 response time of about 91 ms, which was used in the 3880-11 response time calculation before. This response time however contributes little to the average 3880-11 read response times because of the high hit ratio.

The calculations demonstrate that at these high hit ratios the back store (and its blocking factor) have very little impact on the overall performance.

```
12 PAGES PER SIO. CACHE BUSY: 2 MS / PG
HIT RATIO= .95, MISS RATIO = .05
SIOS PER SECOND = SPS =        20
TOTAL PAGING RATE = 20 x 12 = 240 PGS/SEC

SIOS TO 3350S: = SPS x M = 20 x .05 =1
CACHE BUSY - CHANNEL= (12 x 2) = 24   MS
CACHE BUSY - 3350=.05 x (50+1) =  2.6 MS
CACHE BUSY TIME:      24 + 2.6  = 26.6 MS
CONTROL UNIT UTIL. 20 x 26.6    = 53.2 %

QUEUE WAIT, Q = (DST x DU) / 2 x (1-DU)
 = 24 x .532 /(2x(1 - .532))    = 13.6 MS

3350 RESPONSE TIME (ASSUMED)    = 84   MS
READ MISS TIME = 84 x 0.05      =  4.2 MS

3880-11 SWAP SET READ RESPONSE TIME:
 24 + 4.2 + 13.6                = 41.8 MS
```

Figure 30. FOIL 30--3880-11 SWAP (APPROXIMATE METHOD)

Figure 30 gives an approximate method for evaluating the performance of a 3880-11 swap device.

Assume, that the miss ratio equals the DASD ratio. Calculate cache busy time per SIO as the sum of (Miss ratio x 51) and 24. Calculate the cache busy as the product of the SIO rate to the device and the calculated device busy time per SIO. Calculate the queue for the cache, and calculate the response time by adding the hit time (24 ms), the calculated queue time (13.6 ms), and the read miss time (4.2 ms). The read miss time is represented here as the approximate value of the 3350 read response time, 84 ms, times the read miss ratio, .05, i.e., (84 x 0.05 = 4.2). The 84 ms value was chosen somewhat arbitrarily and is used repeatedly in the paper.

6.1.3 DISTRIBUTED SWAP DATA SETS

For a final example of swap, assume that 3 swap actuators are distributed to three nondedicated channels handling the total swap load of 240 pages per second. In addition, the remaining three data base actuators on each one of the channels handle 46 single 4096 byte long data base I/O activity per channel. The relative channel busy calculation is different for the swap actuator and the three data base actuators. The swap actuator sees only the data base actuators while the data base actuators see two data base actuators and one swap actuator. Assume that the data base load is evenly distributed among the data base actuators.

```
1 3380 SWAP ACTUATOR PER CHANNEL
3 DATA BASE ACTUATORS PER CHANNEL
3 CHANNELS

DATA BASE SIOS PER SEC PER channel = 46
DATA TRANSFER: 1.5 + 2.6           =  4.1 MS
CHANNEL UTIL. DUE TO DATA:46 x 4.1= 18.9%

SWAP SIOS PER SEC PER CHANNEL      =  6.7
CHN.UTIL. DUE TO SWAP:(6.7 x 22.6)= 15.1%

TOTAL CHN UTIL: 18.9 + 15.1        = 34  %

RCB(SWAP) = 18.9 / (66   + 18.9)   = 22  %
RCB(DATA) = 27.7 / (66   + 27.7)   = 30  %

SERV.TIME (SEEN BY DATA) =
( ( 6.7 x 22.6) + (2/3 x 46 x 4.1 ))
/ (6.7 + (2/3 x 46 ))              =  7.4 MS
COMMAND DELAY, CD(DATA)= 7.4x.3 /2=  1.1 MS
COMMAND DELAY, CD(SWAP)= 4.1x.22/2=  0.4 MS
OVERHEAD TIME                      =  2.6 MS

RCB(SWAP) = 22% ; RCB(DATA)       = 30  %
SEEK(SWAP)=  3 MS; SEEK(DATA)      =  6   MS
RPS(SWAP) =4.7 MS; RPS(DATA)       =  7.2 MS

LATENCY   =8.3 MS;                 =  8.3 MS
XFER(SWAP)  20 MS; XFER(DATA)      =  1.5 MS
S.T.(SWAP)  38.6 ; S.T. (DATA)     = 25.6 MS

DEVICE UTIL.(SWAP) = 6.7x 38.6     = 25.9%
DEVICE UTIL.(DATA) =15.3x25.6      = 39.1%

QUEUE (SWAP) = 38.6x.26/(2x.74)    =  6.8 MS
QUEUE (DATA) = 25.6x.39/(2x.61)    =  8.2 MS

SRT (SWAP) = .4 + 6.8 + 38.6       = 45.8 MS
R.T.(DATA) =1.1 + 8.2 + 26.8       = 36.1 MS
SWAP RESPONSE TIME = RT (CONCURRENCY)
RT = SRT X (SPC+(STS x CUT ** (NCH-1)))
61.7=45.8x ( 1+ ( 3  x.34  ** ( 3 -1)))
SWAP RESPONSE TIME                 = 61.7 MS
```

Figure 31. DISTRIBUTED SWAP DATA SETS

```
1 3380 SWAP, 3 DB ACT'S PER CHAN. 3 CHANS

DATA, SWAP SIOS PER SEC PER CHAN =  46;6.7
DATA TRANSFER: 1.5 + 2.6         =  4.1 MS
CH.UT.(SW)=15.1% ; CH.UT.(DATA)  = 18.9%
TOTAL CHN UTIL: 18.9 + 15.1      = 34  %

CD(SWAP)    = 0.4 MS; CD(DATA)    =  1.1 MS
OVERHEAD    = 2.6 MS;             =  2.6 MS
RCB(SWAP)   =22  % ; RCB(DATA)   = 30  %
SEEK(SWAP) = 3   MS; SEEK(DATA)  =  6   MS
RPS(SWAP)   = 4.7 MS; RPS(DATA)   =  7.2 MS
LATENCY     = 8.3 MS;             =  8.3 MS
XFER(SWAP) =20   MS; XFER(DATA)  =  1.5 MS
S.T.(SWAP) =38.6 MS; S.T. (DATA) = 25.6 MS
DEV.UT.(SW)=25.9% ; DEV.UT(DATA)= 39.1%
QUEUE(SWAP)= 6.8 MS; QUEUE (DATA)=  8.2 MS

SRT (SWAP) =45.8 MS; R.T.(DATA)  = 36.1 MS

SWAP RESPONSE TIME =
 45.8 x ( 1+ ( 3 x.34 ** ( 3 -1)))=61.7 MS
```

Figure 32. FOIL 31--DISTRIBUTED SWAP DATA SETS

An example of performance modeling for distributed swap data sets was constructed to yield a channel utilization of about 35%. Figure 31 on page 32 presents the detailed calculation and Figure 32 presents in foil format the summary of calculations.

The concurrency formula was used to evaluate the swap response time. The data response time obtained is excellent in spite of the mixed (nondedicated) channel and the heavy swapload (36.1 ms).

Precisely the same methodology can be used for the mixture of three demand paging device and a swap device on the same channel. In fact, this example does represent the (unlikely) case when the demand paging load is 46 pages per second per channel in addition to the swap load.

6.2 DEMAND PAGING EXAMPLES

Examples of the 3880-11 and the 3380 demand paging configurations are evaluated.

6.2.1 3880-11 DEMAND PAGER WITH 2 3350S

```
1 PAGE PER READ SIO. 2 PAGES PER WRITES
HR=.6; M=.4; DR=.49;CACHE BUSY= 2 MS/PG
TOTAL PAGING RATE =               60 PGS/SEC
DR (READ)           = .4  / 2       =  .2
DR WRITE(PAGE)      = .4  / 1.4     =  .29
3350 PG RD SIOS PER SEC: 60 x .2  =12
3350 PG WR SIOS PER SEC: 60x.29/2 = 8.7

CACHE  BUSY (CHAN.) =60 x 2=120MS =12  %
CACHE-3350 PG: (12 x4.6)+(8.7x8.2)=12.7%
CONTROL UNIT UTILIZATION=12 + 12.7=24.7%
WEIGHTED SERVICE TIME, DST:
 247 / (30 + (1.29 x 15))           = 5  MS
QUEUE WAIT, Q = (DST x DU) /(2x(1-DU))
 = 5 x .247 /(2x(1 - .247))         =  .8MS
3350 RESPONSE TIME-                 =33  MS

DEMAND PAGE READ RESPONSE TIME:
 2 + (.4 x 33) + 0.8                =16  MS
```

Figure 33. 3880-11 DEMAND PAGER

Figure 33 presents the modeling evaluation of a 3880-11 demand pager.

At an internal IBM site a 3880-11 supported 60 pages/sec with about a 60% hit ratio and 20% coverage, i.e., 8 MB of cache supporting 40 MB of active (not allocated) page space. Similarly, a hit ratio of 80% was observed with 40% coverage. (i.e., two 3880-11s with 16 MB of cache). Both these observations occured in a TSO-batch environment with up to 100 TSO users being active.

The techniques used for the evaluation of the 3880-11 demand paging are similar to the ones used for its evaluation as a swap device. The figure above shows the pertinent calculations for the 3880-11. The control unit utilization is calculated as the sum of utilization due to cache to channel and cache to back store transfers. A read from the 3350 to the cache takes (3.6 + 1) = 4.6 ms, while a write of 2 pages takes (7.2 + 1)= 8.2 ms. Note, that the (total) cache utilization is kept below 35%. The weighted service time is determined by considering the three different activities occuring: reads (including misses), writes and destage activities. The weighted service time is then used to determine the queue wait time, which for this low utilization is found to be negligible. The average read time is the sum of cache to channel read transfer time, the queue time and the 3350 read response time multiplied by the (read) miss ratio.

The average response time is shown as the sum of the cache-channel transfer time (2 ms), the queueing dalay (.8 ms), and the share of the read miss time (13.2 ms), for an average read response time of 16 ms.

```
2 3350S ON A 3880-11

60% HIT RATIO. MISS RATIO = .4
TOTAL PAGING RATE =            60 PGS/SEC
DR = M/2 + M/(M+1)
DR READ (PAGE)     = .4   / 2      =.2
DR WRITE(PAGE)     = .4   / 1.4    =.29
DR (PAGE)          = .2   + .29    =.49
PGS TO 3350s: 60 x .49        = 29.4 PGS/S

CACHE BUSY: 1 PAGE READ PER SIO=(3.6+1) MS
            2 PAGE WRITES/SIO  =(7.2+1) MS

3350 PG RD SIOS PER SEC: 60 x .2     =12
3350 PG WR SIOS PER SEC: 60 x .29 /2= 8.7
CACHE  BUSY (CHANNEL) 60 x 2         =12 %
CACHE-3350 PG: (12 x4.6)+(8.7x8.2)  =12.7%

RCB(3350) =(.120+.063)/(1-.063)    = 19.5%

OVERHEAD TIME      (3350-CACHE)        1  MS
SEEK TIME                             10  MS
RPS MISS(P)=16.7 x .195 / (1-.195)=  4.1MS
LATENCY TIME                      =  8.3MS

PAGE READ TIME= 1+10+ 4.1+8.3+3.6 = 27  MS
PAGE WRT  TIME= 1+10+ 4.1+8.3+7.2 = 30.6MS
3350 UT =(6 x 27) + (4.35 x 30.6) = 29.5%
SERVICE TIME  = 295 / (6  + 4.35) = 28.5MS
3350 Q   = 28.5 x.295 / (2 x .705)=  6  MS
3350 PAGE READ R.T. = 27   + 6    = 33  MS
```

Figure 34. 3350 PAGING RESPONSE TIME ON A 3880-11

Figure 34 calculates the 3350 paging response time on the 3880-11. The interesting part of this is the calculation of the relative channel busy. From the viewpoint of the 3350, the cache is considered busy when it transfers to the channel (12%), and when it is busy transferring to the other 3350, (6.7%). In the calculations distinction is made between the single page reads and double page writes. The page read response time is 33 ms with this load, (14.7 page/second per actuator). Notice that the cache utilization is low enough not to cause significant RPS misses, and the 3350 utilization is kept low enough so as not to cause high queueing delays on the device itself.

If the calculated 3350 utilization exceeded 35%, it is worth while considering using <u>four 3350 actuators</u> instead of two, so as to reduce the queueing delay caused by device busy. The use of more actuators is a possibility for a demand paging environment, but is unnecessary for the swapping case.

It is possible to quickly estimate the effects of using four 3350 actuators instead of two. The only major change in the calculation from Figure 34 would be the number of SIOs going to an actuator, which would be half of the two actuator case. Consequently the 3350 device utilization is reduced from about 29.5% to about 15%. The 3350 Q-wait time is then reduced from 6 ms to 2.5 ms, reducing the 3350 response time from about 33 ms to 30 ms and the 3880-11 response time from about 16 ms to about 15 ms. Thus, the improvement is not very significant, since the utilization of the 3350 actuators was only 30 % in the two actuator case.

If the paging load were increased from 60 pages per second to 80 pages per second to the 3880-11 the situation would be different. Control unit utilization then increases to about 33 %, RPS miss to 8 ms, and 3350 utilization (for two actuators) increases to 45 %. The 3350 queue wait time increases to about 13 ms for a 3350 response time of 46 ms yielding a 3880-11 response time of 22 ms. With four actuators 3350 utilization falls to about 23 %, for a 3350 response time of about 37 ms yielding a 3880-11 response time of 18 ms. Thus, with a paging load of 80 pages per second to the 3880-11, maintaining the hit ratio of 0.6, the use of four actuators improves response time significantly.

The response time calculations illustrate the main characteristics of the demand paging environment. The load is low, and queue time is almost negligible. If the hit ratio were 100%, the demand paging response time would be roughly 2 ms. As the hit ratio is dropped by supporting a large page space, the key component of the average response time is the portion contributed by reads from the back store. This observation leads to a very simple calculation as a first approximation.

```
2 3350S ON A 3880-11
HIT RATIO = .6 ;  MISS RATIO       = .4
TOTAL PAGING RATE               = 60 PGS/SEC

CACHE BUSY TIME PER PAGE FOR HITS = 2  MS
3350 RESPONSE TIME - (ASSUMED)    =35  MS
MISS CONTRIBUTION = .4 x 35       =14  MS

CACHE  BUSY (CHANNEL)=60 x 2=120MS=12  %
CACHE-3350 PG: (60 x .4 x 4.6)    =11  %
CONTROL UNIT UTILIZATION=12 + 11  =23  %

DEMAND PAGE READ RESPONSE TIME:
 2 + (.4 x 35)                    =16  MS

• READ HIT RATIOS USUALLY HIGHER
  THAN OVERALL HIT RATIOS.

• TSO INTERACTIVE READS ARE USUALLY HITS
  (THE MISSES ARE FOR BATCH.)
```

Figure 35. FOIL 32--3880-11 DEMAND PAGING (APPROXIMATE METHOD)

Figure 35 contains a simple, approximate method for evaluation of 3880-11 demand paging response time. Assume a hit ratio of 60%, and the demand paging response time is (2 + (Miss Ratio x 35)).

The average response time is shown as the sum of the cache-channel transfer time (2 ms), and the share of the read miss time (14 ms), for an average read response time of 16 ms.

There are several reasons why the average demand paging read response observed on an actual system is better than the time calculated by the methodology presented here.

1. Read hit ratios are often higher than write hit ratios; thus, by making the assumption that the read hit ratio is the same as the average hit ratio, the results calculated are conservative. (A write miss causes the use of an available slot, it does not cause activity to the DASD. Write miss ratios are separately reported however). Thus, if an overall hit ratio of 60% is assumed, it is not unreasonable to assume a read hit ratio of 65-70%.

2. The most important, and not easily measurable, effect is that demand page reads for interactive TSO users tend to be serviced by read hits, while read misses would be encountered mostly by long running TSO transactions and batch jobs. The reason for this phenomenon is the nature of the LRU algorithm of the cache. Pages belonging to interactive transactions would tend to stay in the cache, while old pages belonging to batch jobs would have to be retrieved by misses.

3. An additional effect is the existence of multiple exposures, which is ignored in the (averaging) calculations used here. The multiple exposure feature allows a read hit to proceed while a miss is in the process of being handled. Multiple exposures allow these hits to proceed unimpeded; and are therefore clearly important, but their existence does not improve the response time for read misses, (it makes them

slightly worse by potentially causing an additional RPS miss). Thus, TSO trivial transactions serviced by hits, obtain an even better response time, while misses, for batch are slightly delayed. In terms of the average response times, (in contrast to the specific "hit" response times), the impact is not great, since the average queueing delay, as calculated above, is very small.

The actual demand page read response time seen by trivial TSO transactions is much better, since they tend to be serviced by hits rather than misses. Thus, the 3880-11 is an excellent demand paging device with the ability to handle large page spaces (up to 5 times the size of the cache); and should be the device of choice in response-oriented environments.

As a final cautionary observation, the calculations indicate that, for excellent demand paging response times, the cache and the back store utilizations should be kept below 35%.

6.2.2 FOUR 3380 DEMAND PAGE ACTUATORS ON A CHANNEL

```
1  PAGE  PER SIO

SIOS  PER SECOND PER CHANNEL:  60
CONTROL UNIT OVERHEAD              2.6 MS
DATA TRANSFER TIME                 1.5 MS
CHAN. UTIL.: 60 x 4.1=246 ms/s = 24.6 %
CHAN. UTIL./ACTUATOR =  .246 / 4 = .061
REL.CHAN UTIL.: RCB = 3 x .061 /(1-.061)
                    = 0.195 =    20   %

COMMAND DELAY, CD = 3.6 x.20 /2=  0.4 MS
OVERHEAD TIME                     2.6 MS
SEEK TIME                         6.3 MS

RPS MISS = 16.7 x .2 / (1-.2 )    4.2 MS
LATENCY TIME                      8.3 MS
DATA TRANSFER                     1.5 MS
DEVICE SERVICE TIME              22.9 MS

DEVICE UTILIZATION = 15 x 22.9 = 34   %

Q = (22.9 x .34)/(2 x (1-.34)) =  5.9 MS
RESPONSE TIME: 22.9 + 5.9+ .4  = 29.2 MS
```

Figure 36. FOIL 33--3380 DEMAND PAGING

Figure 36 shows an evaluation of 3380 demand paging. If 3880-11s are not available to handle the demand paging load, an alternative is to use available DASD, such as 3380s. The most important consideration in the use of such DASD is to define the demand page data sets small enough, so that the seeks encountered are small. In the example above a 60 pages/second load was assumed, which is reasonable to support about 100 TSO users. The required demand paging data set size is about 50 MB. If all four actuators are grossly overallocated to about 25 MB each, then the span of a data set allocated is about 40 cylinders. The average seek distance is then about 13 cylinders for reads, i.e., not very high. The user of the modeling methodology must make his own assumptions about seeks. (One can assume that 50% of the writes require no seeks).

In the example above, a pessimistic seek value of 6.3 ms was assumed. The advantage of the use of 3380s for demand paging is the ability to define a rather large data set size, such as 40 MB, on each actuator without incurring a heavy seek penalty.

The calculation shows that the most significant portion of the demand page read response time is, on average, almost unchangeable: seeks (maybe about 4-6 ms), latency of 8.3 ms, an overhead of 2.6 ms, a data transfer of 1.5 ms and some RPS miss are always encountered. Queue wait is not very significant, thus a good approximate calculation would state that the demand paging response time obtainable from 3380 actuators ranges between 20 and 30 ms, when there are sufficient actuators to handle the load.

The above calculation assumed that all demand paging goes to 3380s in single page bursts. It is known of course that page writes occur with about two pages per SIO, consequently the device utilization is lower (less seeks and RPS misses) and the device therefore can handle a somewhat higher paging load with heavy write content.

The design graphs and device limit tables make use of the fact that writes tend to be blocked, and the device utilization is therefore somewhat lowered.

7.0 SWAP AND DEMAND PAGE DEVICE LIMITS

The calculations illustrated in the previous chapter can be performed for a variety of paging devices, and paging loads. An APL model was written to do this, and curves were plotted from the model output. In addition, the calculations can be used to establish device use limits, i.e., conditions under which the device utilizations and response times are reasonable and the queues occuring are not excessive. Somewhat arbitrarily these limits were defined in such a way that swap device and path utilizations were kept in the neighborhood of 50%, while demand paging path and device utilizations were kept below 35%. Experience indicates that these are reasonable guidelines.

Device use limit tables can be constructed with these guidelines. In the tables the configurations listed are those that can be attached to a control unit on a dedicated path for the purpose illustrated. Thus, for example there are two swap actuators listed for any of the devices per path. On the other hand, four disk actuators per path are listed for demand paging. The per path configurations listed for different device types are definitely not equivalent.

In a later section of the paper, design graphs of the device modeled response times with varying paging loads are presented. There may be small differences between the modeling results plotted and the results calculated in this presentation, since the hand calculation used slightly more approximations. The design graphs can be used to obtain the tables below.

7.1 SUMMARY OF SWAP DEVICE USE LIMITS

- Massive data transfers
- Important: parallelism (multiple paths)
- Important: fast transfer rates
- Access time is of least importance
- Two or more 3880-11s or 3-4 3380 actuators distributed.
- Dedicated swap channels can be loaded to 50% utilization.

	3380	3375	3350	3880-11	2305
Actuators	2	2	2	(2)	2
HIT RATIO				.95	
DS SIZE MB	20	20	8	16	20
SIO RATE	20	15	10	20	12
PG RATE	240	180	120	240	144
PGS/ACT	120	90	60		72
RESP.TIME	58	100	130	45	70

Figure 37. FOIL 35--SUMMARY OF SWAP DEVICE USE LIMITS

Figure 37 contains a summary of swap device use limits. It shows that a pair of 3350 actuators can be loaded to a maximum of 120 swap pages/second and a pair of 2305 drums to 144 swap pages per second. The response time obtained on the 3350s is worse than on the 2305s however. 3375s are able to support a higher swap paging load than 2305s.

The 3 MB data transfer rates on the 3880-11 and 3380 allows almost twice the paging load than on 2305 and with better response times. The 3880-11 provides the best swap set response time among the configurations considered.

7.2 SUMMARY OF DEMAND PAGE DEVICE LIMITS

- Low data transfer rates
- Most important is access time
- Of almost equal importance is capacity.
- Device of choice is 3880-11.
- Good performance: 4 3380 actuators
- Dedicated paging control unit utilization: 35%

	3380	3375	3350	3880-11	2305
Actuators	4	4	4	(2)	2
HIT RATIO				.6	
DS SIZE MB	50	50	50	40	20
SIO RATE	60-80	60	60		40
PG RATE	60-80	60	60	60-80	40
PGS/ACT	15-20	15	15		20
RESP.TIME	26	41	52	16	11

Figure 38. FOIL 36--SUMMARY OF DEMAND PAGE DEVICE USE LIMITS

Figure 38 contains a summary of demand page device use limits. In calculating the values in the above table it was assumed that page-outs are grouped two consecutive pages per SIO, while each page-in is done one page per SIO. The average block size is then 1.33 pages per SIO, when writes and reads are equal. The effect of swap trim on the local page datasets is considered negligible in this calculation.

The table shows that the 3880-11 is the best demand paging device. It provides the best response times among the paging devices for the same paging rate of 60 pages per control unit.

The table also illustrates the limitations of the 2305 device, which, because of its limited storage capacity is unable to carry a heavy demand paging load.

3380 disks can carry the same demand paging load with four actuators per path, as a 3880-11, but the response time is considerably higher.

Recall, that the read response time is better on the 3880-11 than the average response time; this fact is not factored into the values in the table.

7.3 ADDITIONAL TUNING CONSIDERATIONS

A number of tuning points should be observed in configuring a system:

1. In defining swap page space sizes estimate the swap group size required, say the average size is 36 slots. Notice that an average of 36 means the existence of 35 and 37 (for example), which translates into usage of 3 and 4 swap sets respectively. Thus, over-allocate swap space size by a factor of two, in order to avoid overflowing.

2. In defining demand page space sizes estimate the size required. then for safety define two-three times as much space. This is necessary to allow for VIO data sets, momentary overruns, etc. Do not overallocate to avoid possibility of excessive seeks.

3. When installing a 3880-11 as the primary demand paging device, the size of the backstore data sets should follow a similar rule, to provide the desired "coverage". In order to obtain good hit ratios and demand paging response times it is a wise precaution to gradually increase the size of the back store, but not to over-allocate initially. An initial starting point of 3:1-4:1 coverage is suggested, i.e., allocation of 24-32 MB on the back store. This size can be increased subsequently. Initially, overflow paging devices (3350s or 3380s) could be retained, and subsequently removed when satisfactory coverage and paging response time is observed.

7.4 CONFIGURATION CHECKLIST

A configuration checklist is given below. These rules should be viewed as recommendations only.

1. In configuring the number of actuators carrying swap paging load, it is wise to assure that the number of swap paths (channels, control units, heads of strings) used is at least two and if possible, it equals the number obtained by dividing the swap group size by 12.
2. Fastest devices should be used for swapping. If possible, only one type of swap device should be used.
3. The utilization of a dedicated swap path (channel, control unit, head of string) does not exceed 50%.
4. The utilization of a dedicated swap device does not exceed 50%.
5. The utilization of a dedicated demand paging path (channel, control unit, head of string) does not exceed 35%.
6. The utilization of a dedicated demand paging device does not exceed 35%.
7. 2305 drums are better used for swapping than demand paging.
8. In a mixed swap/demand paging environment observe the demand paging (utilization) rules.
9. Overallocate demand and swap space by a factor of two to three.

8.0 CONFIGURATION EXAMPLES

In this section the methodology previously presented is used to evaluate page/swap configuration alternatives.

8.1 SYSTEM AND WORKLOAD DESCRIPTION

```
•key for paging config.s is page rate.
 Size is of concern for 2305 and 3880-11.

•Assume 3033 UP, 100 TSO users
 -Think time                     = 25 secs
 -Transaction rate  = 100 / 25   =  4 tr/s
 -Logical swap percentage        = 17 %

 3033 UP, page and swap separation
 -Swap group size = 36
  3 page faults / trivial transaction
 -Swap data set size = 100 x 36  = 15 MB

 -Page data set size = 100 x 120 = 50 MB

 -Swap paging rate:
         4 x 36 x 2 x (1-.17)    = 240 pg/s

 -Non-swap paging rate           =  60 pg/s
```

Figure 39. FOIL 38--SYSTEM AND WORKLOAD DESCRIPTION

Figure 39 contains the description of the system and workload considered. In configuring a page/swap subsystem, both paging rate and space sizes must be considered for demand paging and swapping. In the example presented, logical swapping limits the total swap load to 240 pages per second and a demand paging rate of 60 pages per second is attained by using RCCPTRT. There are three page faults per trivial transaction.

8.2 CONFIGURING FOR SWAPPING

- Space = 15 MB, paging rate = 240, at least 2 paths, better: 3+ paths (36/12 = 3)
- 3350: Capability = 60 pages per second / actuator
  240/60 = 4 actuators required = 2 + paths
  Allocate 15 x 2 / 4 = 8 MB / actuator
- 2305: Capability = 72 pages per second / actuator
  240/ 72 = 4 actuators required. = 2 + paths
  Allocate 15 x 2 / 4 = 7.5 MB / actuator
- 3375: Capability = 90 pages per second / actuator
  240/90 = 3 actuators required = 2 + paths
  Allocate 15 x 2 / 3 = 10 MB / actuator
- 3380: Capability= 120 pages per second / actuator
  240/120 = 2 actuators required = 1 + paths
  Allocate 15 x 2 / 2 = 15 MB / actuator
- 3880-11 Capability= 240 pages per second / device
  240/240 = 1 device required, 1 path , 2 actuators
  Allocation required: 15 MB, cache 50% of space
  1 device sufficient.
  Allocate 15 x 2 / 2 = 15 MB / actuator

Figure 40. FOIL 39--CONFIGURING FOR SWAPPING

Figure 40 shows the alternatives when configuring for swapping. Using the page throughput capbilities presented earlier in table form, it is possible to configure the swap configuration desired. Four 3350 or 2305 actuators are required for the swap load, or three 3375 actuators, or two 3380 actuators, or one 3880-11 device. No problems are presented from a space capacity view point either: The four 2305 actuators yield usable space of 40 MB, when only 15 MB is required. Also the coverage obtained on the 3880-11 is just below the maximum 50%, in order to attain a 95% hit ratio.

For maximum parallelism, i.e., good swap-in response time, it would be desirable to have at least three parallel swap paths.

8.3 CONFIGURING FOR DEMAND PAGING

- Space = 50 MB, paging rate = 60 pages per second
- 3350-3375: Capability= 15 pages per second/ actuator
  60/15 = 4 actuators required. 1+ paths
  Allocate 50 x 2 / 4 = 25 MB / actuator
- 2305: Size= 10 MB/actuator
  Allocation required: 50 MB
  50/10 = 5 actuators required + 2 overflow 3350s
- 3380: Capability= 15-20 pages per second / actuator
  60/15 = 4 actuators required. 1+ path
  Allocate 50 x 2 / 4 = 25 MB / actuator
- 3880-11: Cache size = 20% of page space
  Allocation required: 50 MB
  .2 x 50 = 10 MB --> 1+ device.
  With 1 device: slightly lower HR,
  and slightly higher response time
  Capability: 60-80 pages per second/device
  60 / 60 = 1 device.

Figure 41. FOIL 40--CONFIGURING FOR DEMAND PAGING

Figure 41 shows the alternatives when configuring for demand paging. The method is similar; the previously presented table is used. 60 pages/second requires four 3350, 3375 or 3380 actuators on a dedicated path. It is advisable to over-allocate by a factor of two (to three) the DASD to about 100 - 150 MB.

2305s present a space problem. The required 50 MB page space means that at least five actuators would be needed with some overflow capability to 3350s. The 2305 drums are not very effective in a demand paging environment, each drum would carry a load of only 12 pages/second, less than the 3350 actuators would, if configured as recommended.

The 3880-11 presents a slight coverage problem; the 20% coverage requirement would allow only 40 MB of page space, instead of 50 MB. Remembering that the rules presented are guidelines only, it is possible to satisfy the 50 MB page space demands with the 8 MB cache with slightly longer response time than indicated by the guideline. The response time is degraded, because the larger coverage requirement can cause a lower hit ratio. In terms of paging load, the required 60 pages per second can easily be handled by the 3880-11. The 3880-11 gives the best demand paging response time.

8.4 "MIXED" CONFIGURATIONS

Clearly the separation of paging and swapping is advantageous, and provides good performance, especially on large systems, where granularity permits it. This solution is not possible for all installations however. The installation may not have large enough paging loads, may want to use only dedicated channels, (not distributing swap devices). Another reason may be that the installation decides to use 3880-11s for paging. Multiple 3880-11s provide excellent swap performance, and it is possible that there is spare capacity on these devices to provide for the demand paging load as well.

Is the methodology presented applicable for mixed environments? The M/D/1 assumptions imply fixed service times, assume no service time distributions. Mixed configurations depart from these assumptions. Nonetheless, the author thinks it is worthwhile applying the modeling methodology to assess the potential impact of the mixing strategy.

In the mixed cases presented below the same workload is assumed, i.e., 240 swap pages/second and 60 demand pages/second and 3 page faults per trivial transaction, but it is assumed that the installation wishes to have dedicated paging channels. There are two choices available, to have all local paging devices or separate demand paging and swap devices on the same channel. At first, 3380 mixed configurations are examined.

8.4.1 3380 PAGING AND SWAPPING DATA SETS ON SAME CHANNEL

```
2x2 3380 SWAP & PAGE ACTS (2 PER CHAN.)
12 ;1; 2  PAGES PER SIO(SWAP;PAGE RD;WR)

SWAP;PAGE SIOS PER SEC PER CHN =10;15;7.5
CONTROL UNIT OVERHEAD            MS    2.6
SWAP; PAGE RD; WR; DATA XFER   20; 1.5; 3
CU.UT.: (10x22.6)+(15x4.1)+(7.5x5.6)= 33%
RCB(S)=10.4/77.4=.13;RCB(P)=22.6/89.6=.25
ST(SEEN BY SW) =4.6 ;ST(SEEN BY PG) =22.6

CD(SW)=4.6x.13/2=0.3;CD(P)=22.6x.25/2=2.8
SEEK TIME    SWAP: 3   MS; PAGE:   6.3MS
RPS MISS =   SWAP: 2.5 MS; PAGE:   5.6MS
LATENCY TIME SWAP: 8.3 MS; PAGE:   8.3MS
SERVICE TIME SWAP:36.4 MS; PG R:  24.3MS
SERVICE TIME               ; PG W:  25.8MS
DEV UT (SWAP): 36.4 x 10 =         36.4%
DEV Q (SWAP): 36.4x.364/(2x.636) =10.4MS
DEV UT(PAGE):(15x24.3)+(7.5x25.8)=55.8%
SERVICE TIME (PAGE): 558 / 22.5  =24.8MS
DEV Q (PAGE): 24.8x.558/(2x.442) =15.7MS

PAGE READ TIME=24.3 +15.7 +2.8  = 42.8MS
SWAP SET R.T.  =36.4 +10.4 +0.3  = 47.1MS
SWAP RT=47.1x(1.5+3x.364**(2-1))=122.1MS
SUBSYSTEM R.T.=122.1+(3 x 42.8) =250.5MS
```

Figure 42. 3380 PAGE AND SWAP DEVICES ON 2 CHANNELS

Figure 42 shows the effect of using four 3380 actuators on two dedicated paths, an AA4 box, for paging. It assumes that demand paging occurs as a single page read per I/O request, but two consecutive stolen pages are written per I/O request. All swapping occurs with 12 pages per I/O request. There is a demand paging and a swap device on the same path,

[3] In the notation "S" stands for a swap device, L for a local paging device, and (SL, SL) means two paths with one swap and local paging device on each of the paths.

i.e., the configuration is of the form: (SL, SL).[3] Channel (control unit) busy is calculated by separately summing the swap, page read and page write loads. For example, page writes have an overhead of 2.6 ms plus two page transfers (2 x 1.5 ms) for a total of 5.6 ms per request, and there are 7.5 such requests per second. In calculating RCB, the swap device sees only the paging device and vice versa. In the command delay calculation the same principle is used. Seek distance on a swap device is smaller than on the demand paging device. The calculation proceeds essentially as before.

The demand paging device presents a granularity problem. It sustains 15 page reads and 7.5 x 2 page writes, for a total of 30 pages/second. Not surprisingly, the device utilization calculated is high, 55.8%, and the page read response time is high. The subsystem response time is 250.5 ms, half of it from demand paging. The demand paging response time in this configuration is somewhat high, this configuration may present performance problems.

The granularity problem may be solved by using "all local" paging devices. Instead of having demand paging devices with very high device utilizations and swap devices with relatively low utilizations, distribute both kinds of paging among all the devices. The next example shows the effects.

8.4.2 3380 ALL LOCAL PAGING

```
2x2 3380 LOCAL ACTUATORS (2 PER CHAN.)
12;1;2 PAGES PER SIO (SWAP;PAGE RD; WR)

SWAP;PAGE SIOS PER SEC PER CHN =10;15;7.5
CONTROL UNIT OVERHEAD             2.6  MS
SWAP/PAGE DATA TRANSFER TIME 20;1.5;3  MS
CU.UT.: (10x22.6)+(15x4.1)+(7.5x5.6)= 33%
AVG CHN SERV. TIME  330 / 32.5 =  10.2 MS
REL.CHAN UTIL.: RCB = .165/(.67+.165)=20%

COMMAND DELAY, CD= 10 x .20/2  =  1.0 MS
SEEK TIME                         6.3 MS
RPS MISS = 16.7 x .20 / (1-.20)=  4.2 MS
LATENCY TIME                   =  8.3 MS
PAGE READ/WRITE TIME      22.9; 24.4   MS
SWAP SERVICE TIME              = 41.4 MS
DEV.UT=5x41.4 +7.5x22.9 +3.75x24.4= 47%
DEV S.T.  = 470 / (5+7.5+3.75) = 28.9 MS
DEV QUEUE = 28.9x.47 / (2 x.53)= 12.8 MS

PAGE READ TIME= 22.9 + 12.8 +1 = 36.7 MS
SWAP SET R.T.  = 41.4 + 12.8 +1 = 55.2 MS
SWAP RT.= 55.2x(1.5+3x.47 **(2-1))=160.6
SUBSYSTEM R.T.= 160.6 +(3x36.7)=260.7 MS
```

Figure 43. 3380 LOCAL DEV'S ON 2 CHANS, BL=12

Figure 43 shows the effect of using four 3380 actuators on two dedicated paths, an "AA4" box, and all four devices are local paging devices, i.e., the configuration is of the form: (LL, LL). It assumes that demand paging occurs as a single page read per I/O request, but two consecutive stolen pages are written per I/O request. All swapping occurs with 12 pages per I/O request. The calculations are simple, and they yield a much better page read time and a worse swap set response time, than in the previous example. Overall, the paging subsytem response time appears worse, than in the previous (SL, SL) case. This is not the whole story however.

ASM establishes a nominal burst size of 80 ms. This configuration results only in a swap burst of about 55 ms. It is probable therefore that ASM blocks more swap pages in an I/O request. The next example assumes a blocking factor of 18 pages per swap I/O request.

```
2x2 3380 LOCAL ACTUATORS (2 PER CHAN.)
18;1;2 PAGES PER SIO (SWAP;PAGE RD; WR)

SWAP;PAGE SIOS PER SEC PER CH =6.7;15;7.5
CONTROL UNIT OVERHEAD           2.6    MS
SWAP/PAGE DATA TRANSFER TIME  30;1.5;3 MS
CU.UT.:(6.7x32.6)+(15x4.1)+(7.5x5.6)= 32%
AVG CHN SERV. TIME  320 / 29.2 = 11    MS
REL.CHAN UTIL.: RCB = .16/(.68+.16)=19%

COMMAND DELAY, CD=  9 x .19/2 =    0.9 MS
SEEK TIME                          6.3 MS
RPS MISS = 16.7 x .19 / (1-.19)=   3.9 MS
LATENCY TIME                       8.3 MS
PAGE READ/WRITE TIME        22.6; 24.1 MS
SWAP SERVICE TIME                 51.1 MS
DEV.UT=3.4x51.1+7.5x22.6+3.75x24.1= 43.4%
DEV.ST=   434 / (3.4 +7.5 +3.75)= 29.6 MS
DEV QUEUE =(29.6)x.434/(2x.566) = 11.3 MS

PAGE READ TIME= 22.6 + 11.3+.9  = 34.8 MS
SWAP SET R. T. = 51.1 + 11.3+.9 = 63.3 MS
SWAP RT.=63.3x(1 +2x.434**(2-1))=118   MS
SUBSYSTEM R.T. = 118  + 3 x 34.8=222.4 MS
```

Figure 44. 3380 LOCAL DEV.S ON 2 CHANS, BL=18

Figure 44 shows the effect of using four 3380 actuators on two dedicated paths, an AA4 box, and all four devices are local paging devices, i.e., the configuration is of the form: (LL, LL). This example demonstrates the beneficial impact of blocking with contiguous slot allocation. It shows that the swap set response time for 18 pages is only marginally higher than that for 12 pages in the previous case, but device utilization and demand paging response times are much lower. Overall, the subsytem response time is significantly better, but the demand paging response time obtained is still somewhat high (about 35 ms). This configuration (LL,LL) therefore is clearly superior to the other configuration considered (SL,SL), since read time is better, even though swap set response time is worse. It provides very good overall results, i.e., if the system is not very sensitive to demand paging because of the presence of a DB/DC subsystem.

8.4.3 CONFIGURATION WITH 3 (MIXED) 3880-11S

The system could be configured with two 3880-11s. From the configuration guidelines we concluded that at least one 3880-11 is required for swapping and slightly more than one 3880-11 could be used for demand paging.

One 3880-11 swap device provides no parallelism. Swap response time is 3 times the swap set response time. Furthermore, one swap device and one paging device do not readily provide resilience to momentary overloads. If the active swap or paging spaces increase, the hit ratios go down, and response time deteriorates. The problem is that of granularity. A solution might be to use three 3880-11s and use them for both paging and swapping. This solution provides potentially maximum parallelism (three way in this instance) definitely improving swap response time. The question that remains is, what happens to demand paging response time.

Measurements show that with two (mixed) 3880-11s better performance can be attained by defining both swap and local data sets on the back store 3350 actuators than by the use of only local page datasets. Therefore the 3880-11 back store devices are configured in the form: (LS), i.e., one local and one swap device per controller.

Another, and better, alternative is to configure a small swap data set and a larger local page data set for each back store actuator. The order of definition should be a swap data set first followed by local page data set, so that seeks are minimized. Measurements indicate that this controller configuration 2 x (SL, SL) is more resilient to higher paging loads, than the configuration discussed above; and it is expected to be so because of granularity reasons. For simplicity the simpler controller configuration (LS) is used in the evaluations. Three 3880-11s are used in the configuration evaluated.

```
2 3350S ON EACH OF 3 3880-11s.

LET HIT RATIO (OVERALL)= .83
LET "RAW" HR(PAGE)=0.7; HR(SWAP)=0.96
SWAP/PAGE RD/WR SIO RATES= 20 / 30 / 15
SWAP/DEMAND PAGING RATES =  80/20 =4:1
Thus: ((4 X .96Y) + (1 X .7Y)) / 5 = .83
    4.54Y/5 = .83---> Y=.91
HR(SW) = .91 x .96 = .87, M(SW)=.13
HR(PG) = .91 x .7  = .64, M(PG)=.36
DR = M/2 + M/(M+1)
DR READ (PAGE)     = .36 / 2      =.18
DR WRITE(PAGE)     = .36 / 1.36   =.26
DR (PAGE)          = .18  + .26   =.44
DR READ(SWAP)  = (.13/2)          =.07
DR WRITE(SWAP) = (.13/1.13)       =.12
DR (SWAP)      = .07  + .12       =.19

 DEMAND PG:1 PG RD OR 2 PG WR / SIO
 SWAP   PAGING: 12 PAGES      / SIO
PGS TO 3350s:(20x.44)+(80x.19)= 24 PGS/S

3350 PG RD SIOS PER SEC: 20 x .18    = 3.6
3350 PG WR SIOS PER SEC: 20 x .26 /2= 2.6
3350 SWAP  SIOS PER SEC: 80 x .19/12= 1.27
3350 PG RD;WR;SW XFER TIME   4.6;8.2;51 MS

CACHE/CHN BSY:  100 x 2    = 200 MS =20  %
CACHE-3350 PG: (3.6x4.6)+(2.6x8.2)  = 3.8%
CACHE-3350 SW:  1.27x 51            = 6.5%
CACHE-TOTAL  :                      =30.4%

AVERAGE SERVICE TIME =
304 / (10 + (5 x 1.26 ) + (6.7 x 1.12))
= 304 / 23.8  =  12.8 MS

QUEUE TIME, Q = (DST x DU) / (2X(1-DU))
 = 12.8 x .304 /(2x (1 - .304))=  2.8 MS
```

Figure 45. THE USE OF 3880-11 IN MIXED MODE (1)

```
RCB(3350 PAGE) =(.201+.065)/(1-.038)= 28%
RCB(3350 SWAP) =(.201+.038)/(1-.065)= 26%

OVERHEAD TIME      (3350)            1  MS
SEEK TIME (3350 PAGE)               10  MS
SEEK TIME (3350 SWAP)               10  MS
RPS MISS(S)=16.7 x .26 / (1-.26)=   5.9MS
RPS MISS(P)=16.7 x .28 / (1-.28)=   6.5MS
LATENCY TIME                     =  8.3MS

3350 PG RD ST = 1+10+ 6.5+8.3+3.6= 29.4MS
3350 PG WR ST = 1+10+ 6.5+8.3+7.2= 33  MS
DEV UT (3350 PG)=(3.6x29.4)+(2.6x33)=19.2%
3350 PG ST    = 192 / (3.6 +2.6) = 31  MS
3350 PG Q= 31 x .192 / (2 x .808)=  3.7MS
3350 PAGE READ R.T. = 29.4  + 3.7= 33.1MS

3350 SWAP ST  = 1+10+ 5.9+8.3+50 = 75.2MS
DEV UT (3350 SWAP)=  1.3 x 75.2  =  9.8%
3350 SWAP Q=75.2 x.098/(2 x .902)=  4.1MS
3350 SWAP SET R.T. =  75.2 + 4.1  = 79.3MS

3880-11 PAGE RT.= 2+(.36x41.2)+ 2.8 =19.6
3880-11 SWAP RT.=24+(.13x79.3)+ 2.8 =37.1
SWAP RT.= 37.1x(1 +3x(.304**(3-1))=47.4MS
TOTAL PAGING SUBSYSTEM READ RESPONSE TIME=
SWAP RESPONSE TIME + (3 x PAGE RESP.TIME)
 =  47.4 + (19.6 x 3) =        106.2 MS
```

Figure 46. THE USE OF 3880-11 IN MIXED MODE (2)

Figure 45 on page 48 and Figure 46 contain the evaluation of configuration of 3 3880-11s in mixed mode. In the calculations the first question to be answered is the hit ratio. Two mixed 3880-11s have been measured with 75% hit ratio. Also the demand paging hit ratio of a single 3880-11 was 60%, and that of two 3880-11s about 80%. The hit ratio with 2 swap paging 3880-11s was established at 98%. When swap paging and demand paging are mixed, therefore the hit ratio must be between 60 and 98% and is larger than 75%, (since three 3880-11s are used).

Assume that we have 1.5 (!) 3880-11 cache dedicated to demand paging. A raw hit ratio of about 70% may be obtained. Similarly a raw swap hit ratio of 96% can be assumed for another 1.5 3880-11. When the 1.5 demand paging and 1.5 swap paging cache is combined into 3 3880-11s for both swapping and demand paging, the overall hit ratio for 3 3880-11s may be assumed to be 83%.

The next question is that, given 83% overall hit ratio, what would be the separate demand paging and swap paging hit ratios. The not unreasonable assumption was made that the ratio of nominal (raw) hit ratios remains the same, and the overall hit ratio is obtained by weighting the individual hit ratios with the respective paging rates. With these assumptions a demand paging hit ratio of 64% and a swap paging hit ratio of 87% were calculated.

An overall cache (control unit) utilization of about 30% is calculated. The queue wait on the cache is about 3 ms. The response time for the 3350 back store devices is calculated, yielding about 33 ms for demand page read (miss) and about 80 ms for swap set read (miss). The 3880-11 demand page read response time is about 20 ms, while the swap set read response time is about 37 ms. The overall paging subsystem response time at 106 ms is very low indeed, and the demand page read response time of 20 ms is also very low. This configuration provides excellent all-round results.

```
2 3350S ON A MIXED 3880-11.

HIT RATIO (OVERALL)= .83
SWAP          PAGING RATE    = 80 pgs/sec
DEMAND        PAGING RATES   = 20 pgs/sec
HR(PAGE) = 0.64; HR(SWAP) = 0.87
SWAP SIO RATE                = 6.7 SIO/sec
PAGE READ SIO RATE           =10   SIO/sec
PAGE WRITE SIO RATE          = 5   SIO/sec

CACHE/CHN BSY:(100 x 2)               = 20  %
CACHE-3350    : 100 x .17 x 4.6       =  7.2%
CACHE-TOTAL   :                       = 27.2%

AVERAGE SERVICE TIME =
272 / (10 + 5 + 6.7 )                 = 12.8 MS

QUEUE TIME, Q = (DST x DU) / (2X(1-DU))
 = 12.8 x .272 /(2x (1 - .272))       =  2.4 MS

3880-11 PAGE READ RESPONSE TIME
 =  2 + (.36 x 35) + 2.4              = 17   MS

3880-11 SWAP SET READ RESPONSE TIME
 = (12 x 2) + (.13 x 84) + 2.4        = 37.3 MS
```

Figure 47. FOIL 41--MIXED 3880-11 (APPROXIMATE METHOD)

The mixed 3880-11 calculations first presented are very cumbersome. A simple approximation to the detailed methodology is provided in Figure 47. An overall hit ratio of .83 is assumed. The demand paging and swap hit ratios of 0.64 and 0.87 can be calculated as before, or assumed. Cache utilization can be calculated by using the overall miss ratio and assuming 4.6 ms cache busy time (3.6 ms data transfer and 1 ms protocol time) per missed page. Average service time and hence the queue wait is calculated, by dividing cache busy with the overall SIO rate. For the response time calculations 3350 demand page read response time of 35 ms and a swap set response time of 84 ms are used.

The detailed results show that mixing the swap and demand paging load on the three 3880-11s is a good solution. It provides good parallelism, and the demand paging response time does not become high, because the cache utilization is kept low enough (30%), so that 3350 RPS misses due to cache busy time are kept low. The availability of three separate paths and cache storage also provides excellent resilience to momentary overloads.

More concretely, the swap response time obtained is better than in any other configuration considered. The demand paging response time is slightly worse than the dedicated 3880-11 demand paging case, but still better than obtainable from other devices. The overall paging subsystem response time, i.e., swap response time plus three demand paging read response times, is excellent.

Recall that the read hit ratio in fact may be higher than the write hit ratio, and that trivial TSO transactions would tend to see read hits instead of read misses. Thus, the trivial TSO demand page read response times should be better than the average read response times calculated.

8.4.4 CONFIGURATIONS WITH MIXED 3350S

It is worth briefly discussing a mixed 3350 environment. In this case four swap actuators would be required on two (or more) paths and also four demand paging actuators on a separate path to support the system. It is possible to distribute the 3350 swap actuators to four (nondedicated) paths to obtain maximum parallelism. This solution provides the best response time. If the installation wished to isolate paging and swapping from other activity, then it needs at least three paths. It is better to spread the four swap actuators to three separate paths (granularity problem) and the four demand paging actuators to the same three paging paths, than to concentrate them on their separate swap and demand paging paths. The following configuration is feasible: (SLLL,SS,SL); i.e., place three local page data sets on the same path as a swap device, keep one dedicated swap path, and place a swap and a local page data set on the third path. This configuration would result in unbalanced paths, since path (SLLL) could be used much more than path (SL). If long strings are available, the following configurations would work much better: (SSLL,SSLL,SSLL) or (SLLL,SLLL,SLLL).

On the other hand, once the decision is made for dedicated paging paths, the following configuration could also work: (LLLL, LL, LL). The following configuration would have difficulties: (LL, LL, LL), since there are not enough actuators. Measurements indicate, that in these cases the all local configurations tend to work better with higher loads, than the mixed (page/swap) configurations, if the number of actuators is limited.

CONFIGURATION NO	TYPE	BURST SIZE PGS	PAGE RESPONSE MS	SWAP TIME	OVERALL	PAGE DEV UT %
1	3x(LLLL)	8	62	226	412	35
2	3x(LLLL)	10	62	194	380	35
3	3x(SSLL)		68	154	358	37
4	3x(SLLL)		61	180	363	24

LOAD: 80 SWAP PAGES PER CHANNEL.
20 DEMAND PAGES PER CHANNEL.
3 DEMAND PAGES PER TRANSACTION.
SWAP GROUP SIZE = 36.
3 PARALLEL PATHS.

Figure 48. MODELLED MIXED 3350 CONFIGURATIONS

Figure 48 shows the results of modeling a variety of mixed 3350 configurations. With three paths and a swap group of 36 pages the burst size cannot be kept to 80 ms if the number of pages exceeds eight. The table shows the tradeoffs. There is very little to choose between the three configurations (2, 3, and 4). If ASM does not succeed in blocking at ten pages per SIO, (Conf. 2) only at eight pages per SIO, (Conf. 1) then the all local configuration (Conf. 1) yields definitely poorer performance than the ones with swap data sets. Configuration 4 appears to provide the most resilience to demand paging overloads, and would appear to be the configuration of choice to the point where swap devices are overloaded.

The overall response time is calculated as the sum of the swap response time and (three times the page-in response time).

8.5 TSO RESPONSE TIME ANALYSIS

```
Trivial Response time objective: 500 ms
Trivial TSO transaction components.
-CPU cost: about 250k instructions.
-I/O cost: 2 I/Os per transaction.
-Swap-in time
-3 page faults per tr. (Stor. isol.)
CPU service time = .25 / 5 = .05 sec
CPU utilization due to TSO trivial
4 tr/sec x .05 sec/tr = .2  = 20%
Assume 10% higher pty tasks CPU

Then:
CPU response time
  .05 / ((1-.2) ) = 60 ms, about 70 ms

I/O response time: 50 ms per I/O
Total I/O response time: 2 x 50 = 100 ms

Paging response time: ?
Swap   response time: ?
```

Figure 49. FOIL 42--TSO RESPONSE TIME ANALYSIS

Figure 49 shows an analysis of (trivial) TSO response times. In the system discussed, the installation is aiming at a trivial response time of 500 ms, or better. It determines that each trivial TSO transaction consumes 50 ms in CPU time, which, together with waiting for the CPU, costs about 70 ms in response time. Also 2 I/O operations per transaction are required on average, resulting in an I/O response time of about 100 ms. The sum of CPU and I/O response times is 170 ms, leaving 330 ms for paging and swapping. Assume that there are three demand page faults per trivial transaction.

8.6 COMPARISON OF SUBSYSTEMS

DEVICE	ACTS	PATHS	RESPONSE TIME SWAP	PER PAGE	OVERALL
1.S3380	2	1	174		
P3380	4	1		26	392
2.L3380	4	2	118	35	373
3.S3880-11	2	1	135		
P3880-11	2	1		16	330
4.S3380(dist)	3	3	62		
P3380	4	1		26	299
5.S3380(dist)	3	3	62		
P3880-11	2	1		16	*269*
6.M3880-11	6	3	47	20	*269*
7.S3350	4	2	390		
P3350	4	1		52	650

Figure 50. FOIL 43--COMPARISON OF SUBSYSTEMS

Figure 50 contains seven alternative configurations and the trivial TSO transaction response calculated for them.

The first configuration uses two 3380 actuators on a dedicated swap channel (not advocated), and four 3380 demand page actuators on another dedicated demand paging channel. The swap response time obtained is 3 x 58 = 174 ms for a physical swap; taking logical swap into account, it is 0.83 x 174 = 144 ms, while the demand page read response time is 26 ms. The overall (transaction) response time is (170 + 144 + 3 x 26) = 392 ms.

The configuration can be improved by introducing two 3380 local paging paths with four actuators. This (as shown in detailed calculations) provides a modicum of parallelism.

Further improvement is achieved if the paging configuration used is two 3880-11s, one used for swapping and the other one for demand paging. The response time is 330 ms. The use of 3880-11s improves the response times. There is no parallelism however for swapping, and swap response time represents a significant portion of the overall response time.

The effects of parallelism are demonstrated when with three distributed 3380 swap actuators the swap response time is cut to 62 ms, which cuts the overall response time to 299 ms, even though only 3380 demand paging is used. In this configuration the problem is the demand paging response time, which is relatively high.

The next improvement is to use a 3880-11 for demand paging, but retaining the 3380 swap actuators. This results in a further improvement and a response time of 269 ms. This solution clearly demonstrates the advantage of using 3880-11s in the demand paging environment. This configuration should be an excellent choice.

Another alternative is to use three 3880-11s in mixed mode, i.e., one swap and one page data set defined as back store devices for one control unit. Maximum parallelism is achieved for swapping, yielding excellent swap response time, with a slight degradation in demand paging response time compared to the 3880-11 being dedicated to demand paging. The overall response time is excellent. The actual response time should be better since the trivial TSO transactions see mostly read hits, not read misses. Observe, that the three 3880-11s provide for <u>all</u> the swap and demand paging in this configuration.

As a last alternative, an all 3350 solution is also evaluated. This solution is adequate from the viewpoint of throughput, but does not provide the required trivial response time.

9.0 INTERPRETATION OF MEASUREMENTS

It is sometimes difficult to interpret measurement numbers obtained from RMF. It is even more difficult to compare the values used in a configuration design with the values measured. This section attempts to address the problem of interpreting the measurements to some small degree.

9.1 SWAP MEASUREMENTS

```
•RMF reports swap set service time
 in swap data set report,
 this does not include all queueing.
 ASM queueing included in RCVSWPTM.

•swap sets do not arrive exponentially
 but lumped. Consider 4 swap sets.

 --- --- ---    ----    ---    ---
|   |   |   |  |IORB|  |   |  |   |
|ST |ST | Q |--+----+--+DEV+---+CU |
|   |   |   |  |IORB|  |   |  |   |
 --- --- ---    ----    ---    ---

 --- --- ---    ----    ---    ---
|   |   |   |  |IORB|  |   |  |   |
|ST |ST | Q |--+----+--+DEV+---+CU |
|   |   |   |  |IORB|  |   |  |   |
 --- --- ---    ----    ---    ---

RT = ((Q+ST) + (Q+2xST)) / 2 = Q + ST x 1.5
```

Figure 51. FOIL 44--SWAP MEASUREMENTS (1)

Figure 51 discusses the interpretation of swap measurements by RMF. RMF reports swap service time in the swap data set report. This includes some queueing, because there are two (MVS/SP 1.3.0) or three (MVS/SP 1.3.1) IORBs per data set, i.e., that many requests can be started by ASM to a given data set. Any further requests are queued in ASM itself. Thus, RCVSWPTM is a better measure of swap set response time than the one reported in the swap data set report.

It is very easy to assume exponential arrival of swap sets for modeling purposes. It is known however, that all the swap sets belonging to a swap group tend to arrive to the swap data sets almost simultaneously. Thus, the situation illustrated on the foil occurs. In this case four swap sets exist in the swap group which are distributed evenly to the two devices, almost instantaneously. If there are not enough IORBs, i.e., actuators, then it is possible that multiple requests queue on one IORB, causing an elongation of the swap set response time. In any case, the requests spend time waiting (Q), then are serviced with a service time (ST), one after another. The response time for the first swap set is (Q + ST), for the second, (Q +(2 x ST)). The average therefore is (Q + (ST x 1.5)) instead of ((Q / 2) + ST). Any attempt to compare measured swap set response time with modeled numbers requires a calculation similar to the one shown above taking the number of swap sets in the swap group into account.

•RMF does not (directly) report
swap response time.

•RCVSWPTM usually > swap set response
time.
(Reported swap data set service time)

•IF RCVSWPTM >> swap set response
time, then serious queueing in ASM.

Figure 52. FOIL 45--SWAP MEASUREMENT (2)

Figure 52 continues the discussion of swap measurements by RMF. It is possible to evaluate swap response times from the RMF workload report as shown in Reference 1. The swap delay <u>for trivial TSO</u> is obtained by

$$\frac{\text{Absorption Rate-Service Rate}}{\text{Absorption Rate}} \times \text{Response Time} = \text{Swap Delay}$$

In order to obtain physical swap-in time, the additional calculation shown below must be performed, to adjust for logically swapped transactions.

Swap-in time = swap delay / (1 - (logical swap ratio)

Logical swap ratio = logical swap-out ratio
x logical swap-in ratio

Note that whenever RCVSWPTM is much higher than swap set response time, (which is reported in the swap data set report as service time) the swap set configuration should be adjusted.

9.2 PAGING MEASUREMENTS

•Local page data sets have
- page faults (1 per "SIO")
- page writes (1.7+ per "SIO")
- page trims (1-10 per "SIO")

•Local page data sets have
I/O request counts, but
What is blocking factor???
Must calculate it by assumptions.

• Page read response time not
reported, only page transfer time.
Read response is probably worse
than page transfer time.

Figure 53. FOIL 46--PAGING MEASUREMENT

Figure 53 discusses the interpretation of paging measurements by RMF. Local paging devices service a mixture of requests; page faults of one page per I/O request, page writes of two pages per I/O request, swap trims and possibly swaps of almost any size up to the burst. Whenever measurement results are analyzed, this mixture must be kept in mind.

The page data set report contains I/O request counts for each data set, but appended requests do not show up as separate I/O requests. The average blocking factor for local page data sets cannot readily be calculated from RMF, except by making assumptions of the kind given in this paper.

There is a measured number, the average page transfer time, which is meaningful. This page transfer time combines all the various paging requests arriving at the device. Thus, on DASD, the demand page read response time, which is the important value, must be assumed to be worse than the page transfer time. The reason for this is that a single page request encounters essentially the same overhead, (control unit overhead, seek, RPS miss, latency) as a multi-page request, but the average page transfer time is significantly improved by multipage requests, e.g. swapping.

For the 3880-11, page transfer times is an average of read hits, read misses and write hits. Since reads are a mixture of read hits and read misses only, the read hit time will also be slightly worse, than the page transfer time. On the other hand, trivial TSO sees mostly read hits, so it sees better than average page read transfer times.

Measured page transfer time numbers give a general indication of the relative goodness of the configuration, but not necessarily absolute demand page-in times.

10.0 DESIGN GRAPHS

The results of the modeling methodology have been plotted on six attached design graphs; these can be used to configure a system without the need for the elaborate calculations demonstrated.

1. Swap set response time on two actuators each of 3350, 3375, 2305 and 3380 on a dedicated path using 12 pages per SIO and M/D/1 formulation.

2. Swap set response time on two actuators of 3380 on a dedicated path, and swap set response time on a 3880-11 with hit ratios of .9 and .95 on a dedicated path using 12 pages per SIO and M/D/1 formulation.

3. Swap response times on four actuators of 3380 and 2305s on two dedicated paths, and swap set response time on two 3880-11s with hit ratios of .9 and .95 on two dedicated paths using 12 pages per SIO and M/D/1 formulation, with a swap group size of 36.

4. Swap response times on six actuators of 3380 and 2305s on three dedicated paths, and swap response time on 3 3880-11s with hit ratios of .9 and .95 on three dedicated paths using 12 pages per SIO and M/D/1 formulation, with a swap group size of 36.

5. Demand paging response time on four actuators each of 3350, 3375 and 3380 and 2 actuators of 2305 on a dedicated path using 1.33 pages per SIO and M/D/1 formulation. Single page reads and double page writes yield an average of 1.33 pages per SIO, if the number of reads and writes is equal.

6. Demand paging response time on four actuators of 3380 on a dedicated path and a 3880-11 with hit ratios of 0.6 and 0.8 on a dedicated path using 1.33 pages per SIO and M/D/1 formulation. (The combination of single demand page read and two page writes for equal number of pages read and written yields an average of 1.33 pages per SIO. It is assumed that swap trim is negligible). The response time plotted is that of the read responses, but read hit ratio is assumed to be the same as the overall hit ratio.

10.1 USING THE DESIGN GRAPHS

As an example of using the design graphs, the previous example is redesigned. A system is to be configured with a swap paging rate of 240 pages/second and a demand paging rate of 60 pages/second. Total page space requirements are 50 MB for demand paging and 15 MB for swap paging. Each trivial TSO transaction encounters three page faults.

Graph 2 yields swap set response times of

1. 60 ms for two 3380 swap actuators.

2. 46 ms for a 3880-11 with a hit ratio of 0.95, coverage is (8 MB / 15 MB = 53%), so the 95% hit ratio is achievable.)

3. 60 ms for a 3880-11 with a hit ratio of 0.9.

Since the swap group size was assumed to be 36 slots, a physical swap response time of three times the above value is obtained, i.e., 180, 138 and 180 ms respectively.

From graph 6 demand paging response times can be obtained.

1. 26 ms for four 3380 demand paging actuators.

2. 16 ms for a 3880-11 with a hit ratio of 0.6, coverage is (8 MB / 50 MB = 16%), so the 60% hit ratio should be close.)

3. 8 ms for a 3880-11 with a hit ratio of 0.8. This hit ratio is almost achievable with two 3880-11s, i.e., with a coverage of (16MB / 50 MB = 32%).

If it is desired to improve swap response times, it is possible to dedicate two control units to swapping. Graph 2 provides swap response times directly.

1. 155 ms for four 2305 swap actuators on two paths.
2. 93 ms for four 3380 swap actuators on two paths.
3. 75 ms for two 3880-11s with a hit ratio of 0.95, coverage is (16 MB / 15 MB = 100%), so the 95% hit ratio is underestimating actual performance.

11.0 SUMMARY OF METHODOLOGY

•Determine swap and demand paging rates

-construct paging configurations using device limit tables

-for response time calculations use design graphs (for dedicated channels)

-for mixed configurations use detailed modeling calculations

Figure 54. FOIL 47--SUMMARY OF METHODOLOGY

Figure 54 contains a summary of the modeling methodology.

Configuration methodology was presented consisting of three separate alternatives. In all cases the first step is to determine the paging load desired.

1. The swap and demand page device limit tables can be used to determine the number of actuators required, if demand and swap paging are separated to distinct devices.

2. If the response times are also desired, then the design graphs can be used to evaluate swap set and demand paging response times for any dedicated paging/swapping channel (control unit). The design graphs can be used directly for swap response time evaluation in dedicated swap channel configurations.

   If an estimate of overall TSO response time is also desired, then CPU and I/O response times have to be determined, e.g., from RMF, also the number of pagefaults per transaction has to be estimated, e.g. from SMF or RMF.

3. For the evaluation of mixed configuration it is best to use the full modeling methodology.

12.0 CONCLUSIONS

- For best performance

  -use multiple 3380 or 3880-11 swap paths

  -use 3880-11 demand paging.

- It is possible to determine

  page/swap configurations

  by simple calculations.

Figure 55. FOIL 48--CONCLUSIONS

Figure 55 contains the conclusions of this paper. In conclusion the 3880-11 provides the best demand paging response times.

For the best swap reponse, multiple swap paths are required. Thus, multiple 3380 actuators distributed on non-dedicated paths could be used. Alternatively, multiple 3880-11s can be used.

The paper demonstrated that simple calculations can be used for the sizing of adequate paging configurations.

13.0 REFERENCES

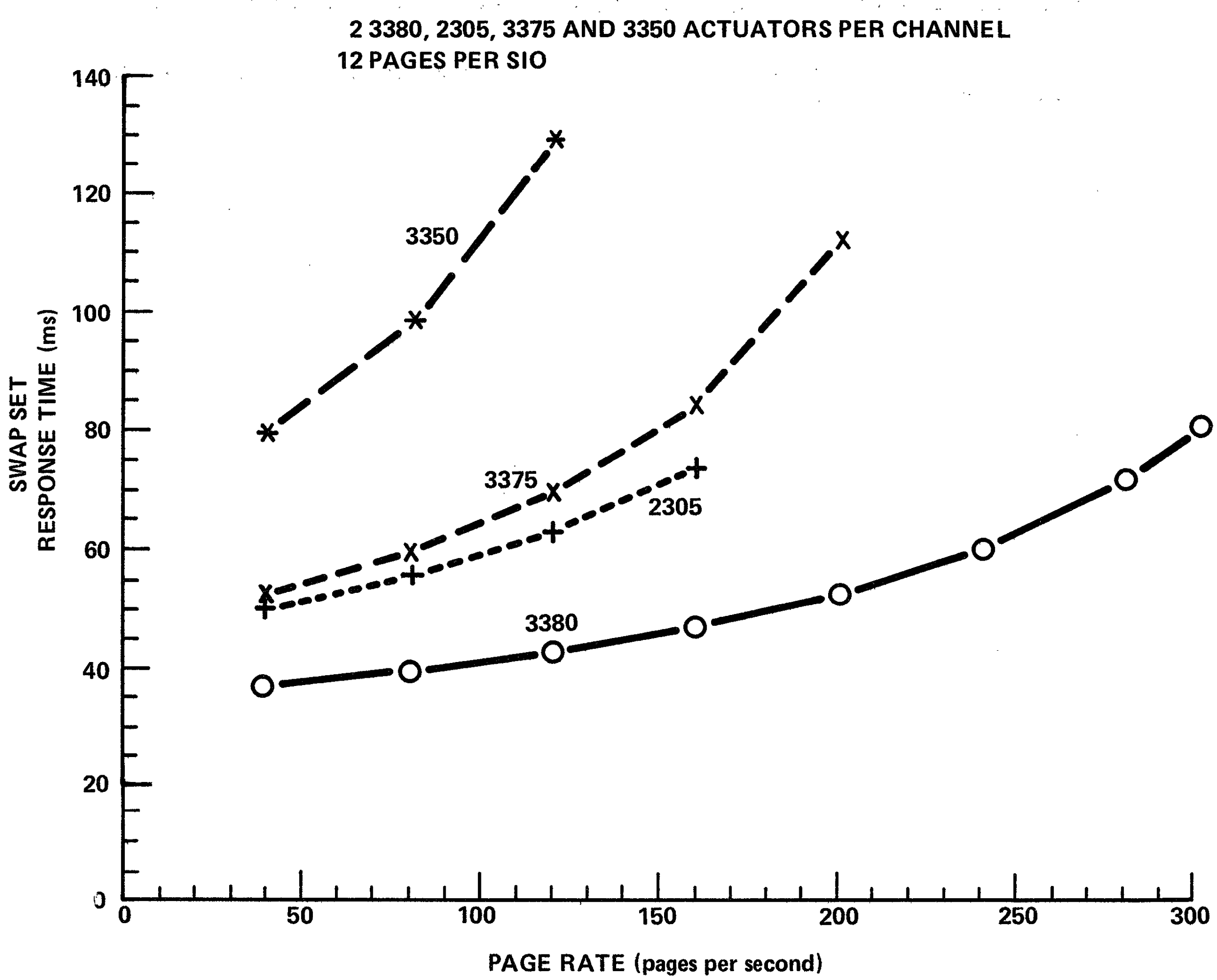
1. T. Beretvas: "Paging in MVS/SP with Extended Swap," IBM Washington System Center Technical Bulletin, GG22-9206-1, October 1981

2. A. J. Thadhani: "Interactive User Productivity," IBM System Journal, 20, No 4. 407-423 (1981).

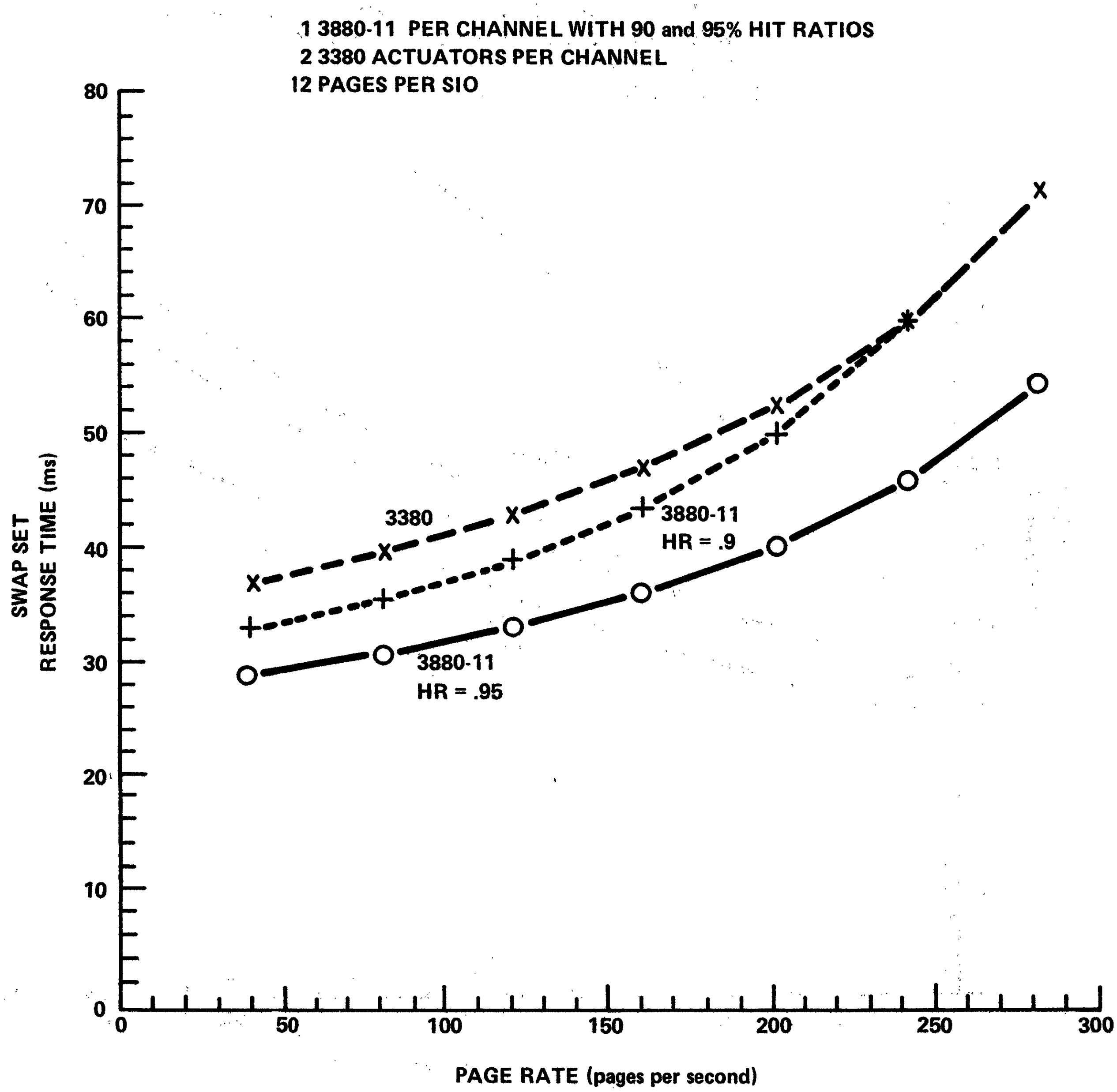
3. S. Friesenborg: "DASD Path and Device Contention Considerations," IBM Washington System Center Technical Bulletin, GG22-9217-0, March 1981.

4. S. Friesenborg: "MVS Paging Performance Considerations," IBM Washington System Center Technical Bulletin, GG22-9264-0, November 1981.

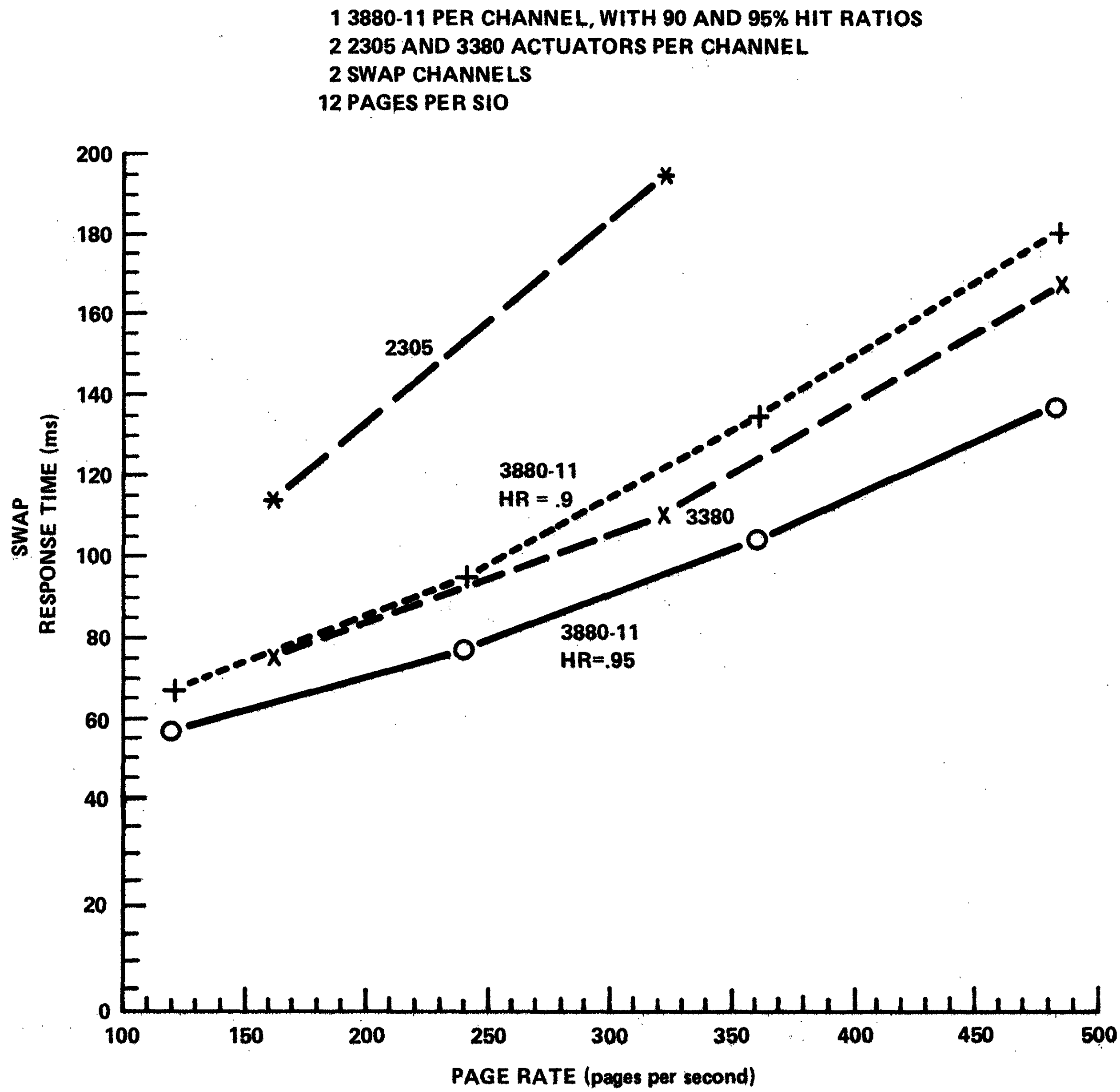
5. D. Hunter: "Modeling Real DASD configurations," IBM Yorktown Research Bulletin, RC8606, December 1980.



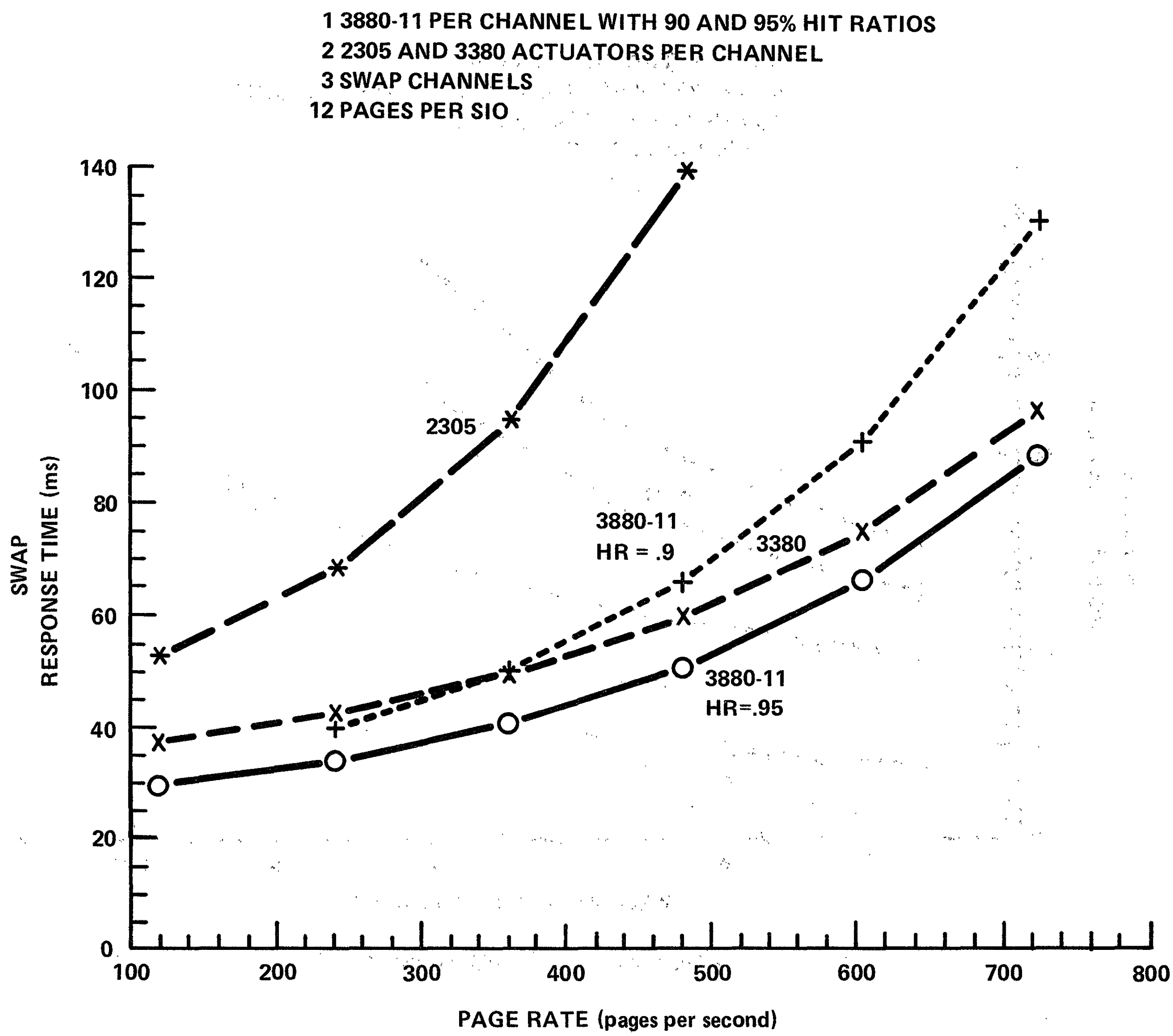
Design Graph No. 1 (for Swapping)



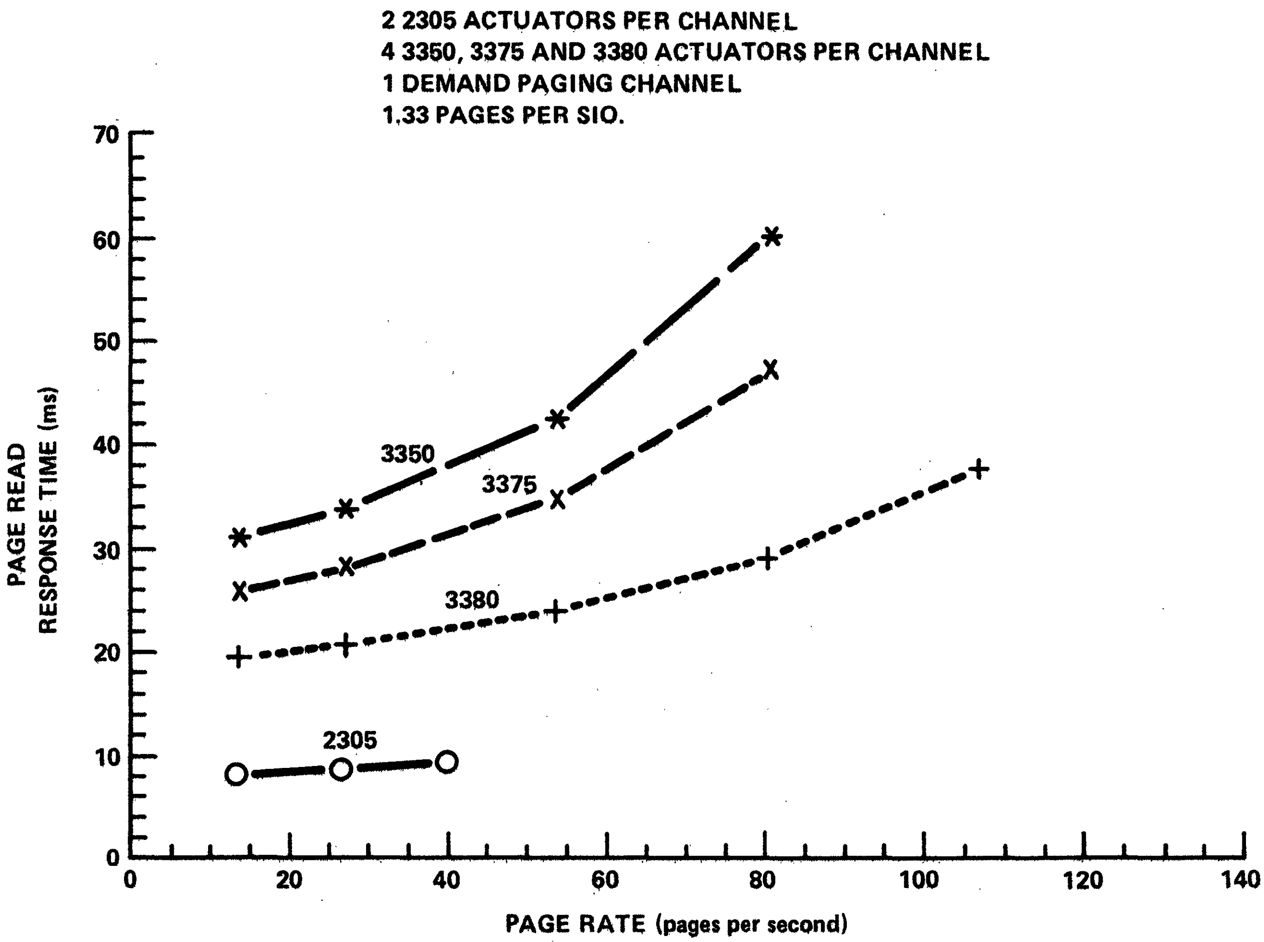
Design Graph No. 2 (for Swapping)



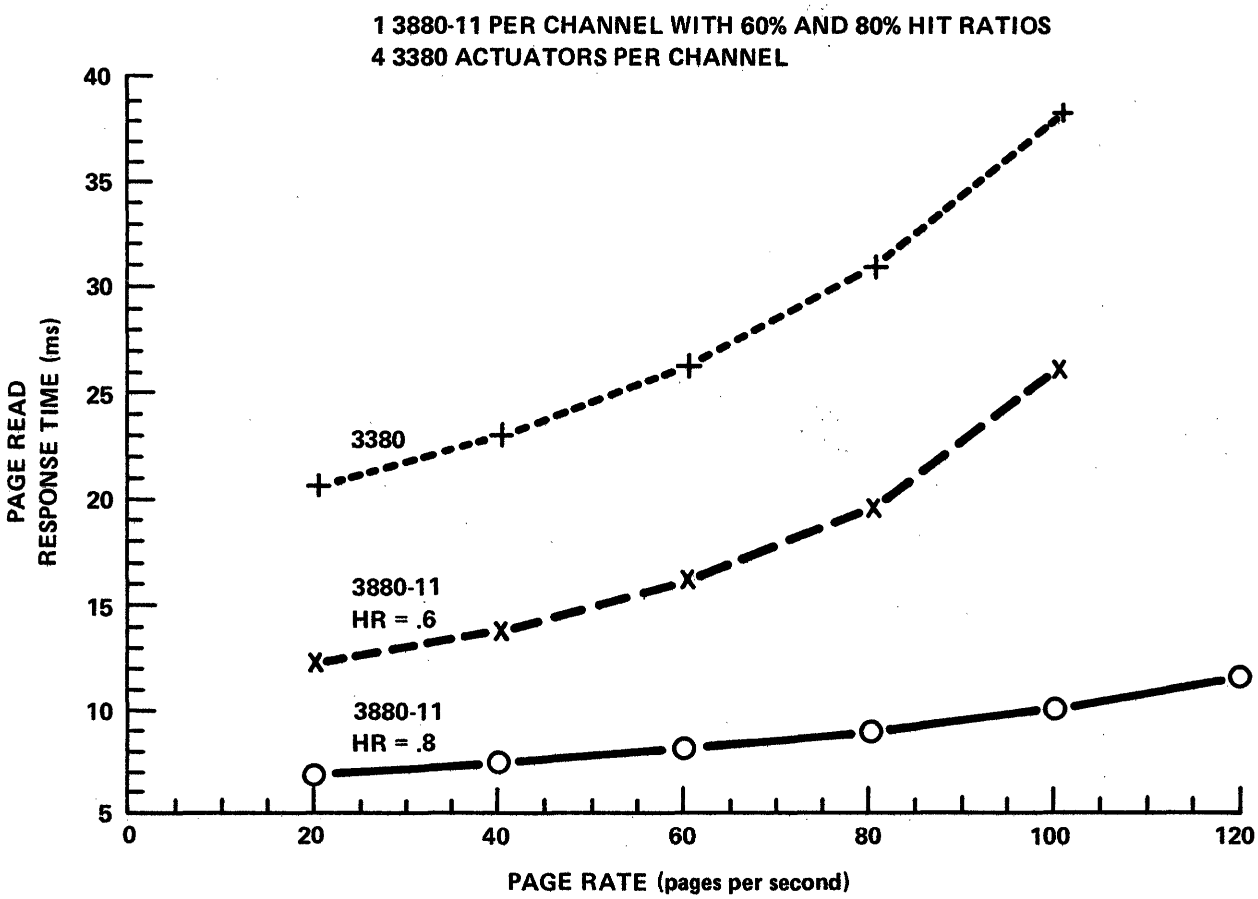
Design Graph No. 3 (for Swapping)



Design Graph No. 4 (for Swapping)



Design Graph No. 5 (for Demand Paging)



Design Graph No. 6 (for Demand Paging)

PAGE - SWAP CONFIGURATIONS.

-> PERFORMANCE DATA APPLY TO SPECIFIC<-
-> CONFIGURATIONS AND WORKLOADS. <-
-> THIS PRESENTATION CONTAINS <-
-> MOSTLY MODELING DATA <-

-> CARE SHOULD BE EXERCISED WHEN <-
->EXTRAPOLATING TO OTHER ENVIRONMENTS<-

PRESENTATION OF THE DATA DOES NOT CONSTITUTE A WARRANTY THAT ANY OTHER INSTALLATION WILL OBTAIN COMPARABLE OR BETTER PERFORMANCE.

I GRATEFULLY ACKNOWLEDGE THE CONTRIBUTIONS OF
C. DOWLING, G. KING AND M. GOLDFEDER

OBJECTIVES.

- PROVIDE SIMPLE FORMULAE FOR EVALUATION OF PAGING/SWAPPING CONFIGURATIONS. APPLICABLE TO MVS (AND VM)

- USE FORMULAE TO OBTAIN DEVICE CAPABILITIES.
  - -THROUGHPUT : PAGES/SEC
  - -RESPONSE TIME (PER I/O REQUEST)

- PROVIDE EVALUATION METHODOLOGY SATISFYING PAGING RATE & SPACE DEMANDS

- ILLUSTRATE METHODOLOGY BY EXAMPLES.

- DEVELOP EASY-TO-USE DESIGN GRAPHS.

- PROVIDE TUNING HINTS.

OVERVIEW.

->PAGING CONFIGURATION POSSIBILITIES.

- DEVICE CHARACTERISTICS.

- PAGING SPACE SIZE AND RATE CALCULATIONS.

- CONCEPTS OF I/O ACTIVITY AND FORMULAE

- EXAMPLES.

- SWAP AND PAGE DEVICE LIMITS.

- CONFIGURATION EXAMPLES.

PAGING TYPES. (MVS)

PAGING TYPE	PAGES/ IO REQ.	PAGE DATASET TYPE
PAGE-IN (FAULT)	1	LOCAL
PAGE-OUT (STEAL)	1-2....2	LOCAL
VIO	1-10...4	LOCAL
SWAP TRIM	1-8....5	LOCAL
SWAPIN/OUT	1-12..12	SWAP
SWAPIN/OUT	1-?...?	LOCAL
PLPA	1 ..1	PLPA
COMMON	1-2 ..1	COMMON

KEY DECISION: WHETHER TO SEPARATE "DEMAND" PAGING FROM "SWAPPING"

PAGING CATEGORIES.

1. SWAPPING
-36+ PAGES TRANSFERRED IN/OUT AT ONCE.

-GOOD SWAP-IN RESPONSE TIME IF 3+ I/O REQUESTS ON SEPARATE, PARALLEL PATHS. IMPORTANT: FAST DATA TRANSFER RATE. 3880-11, 3380

-SWAP SPACE SIZE REQUIREMENTS: LOW ---> HIGH ACCESS DENSITY.

---

2. DEMAND PAGING.
-RESOLUTION OF SINGLE PAGE FAULT.

-BEST PAGE-IN RESPONSE TIME FROM DEVICE WITH FAST ACCESS TIME. 3880-11

-SPACE SIZE REQUIREMENTS: HIGH ---> VERY LOW ACCESS DENSITY. (COMPARED TO SWAP SPACE)

SYSTEM ORIENTATION.

- TSO RESPONSE TIME. ---> PRODUCTIVITY
  - -TSO FAVORED OVER BATCH.
  - -TRIVIAL TSO FAVORED OVER NONTRIVIAL
  - -FIRST: INCREASE MAIN STORAGE TO THE LIMIT. IT IS THE BEST PAGING DEVICE
  - -NEXT: IMPROVE PAGING SUBSYSTEM BETTER SWAP AND DEMAND PAGING

  ACTIONS:
  - -> MAXIMUM PARALLELISM FOR SWAP.
  - -> SEPARATE SWAP FROM DEMAND PAGING.
  - -> FASTER SWAP AND DEMAND PAGING DEV'S.

->>>CHANNEL CAPACITY PERMITTING:<<<-

ENHANCE SWAP = REDUCE DEMAND PAGING
- -> STORAGE ISOLATION FOR MORE SWAP
- -> REDUCE LOGICAL SWAP BY REDUCING SYSTEM THINK TIME (LSCTMTE)

- BATCH THROUGHPUT.
  - ->USE SRM CONTROL TO FAVOR BATCH. (E.G. DISPATCHING PRIORITY)
  - ->THE RULES ABOVE SHOULD BE FOLLOWED.

SEPARATE SWAP DATASETS.

ADVANTAGES:

- -SWAP I/O SEPARATED FROM DEMAND PAGING.
- -LOW SEEKS <-- SMALL (SWAP) DATASETS.
- -HIGHER SWAP CHANNEL UTIL. POSSIBLE

DISADVANTAGE:

- -<MEDIUM SYSTEMS: LOSS OF GRANULARITY (MAY BE MORE ACTUATORS REQUIRED.)
- -SLIGHTLY MORE TUNING EFFORT REQUIRED.

RECOMMEND:

- -USE ON COMBINED DB/DC-TSO SYSTEMS, SO (TSO) SWAPPING DOES NOT INTERFERE WITH (DB/DC) DEMAND PAGING.
- -USE ON LARGE (>=3033UP) SYSTEMS WITH MANY (>50) TSO USERS, ESPECIALLY WITH DIFFERENT PAGING DEVICE TYPES.

COROLLARY:

- -SEPARATE DEMAND PAGE DATASETS
- -GOOD DEMAND PAGE RESPONSE TIME.

ALL "LOCAL" PAGE DATASETS. (MVS)

ADVANTAGE:

-BETTER GRANULARITY,
POTENTIALLY FEWER ACTUATORS.

-LESS TUNING EFFORT REQUIRED.

DISADVANTAGE:

-PAGING QUEUED BEHIND SWAPPING.

RECOMMEND:

-USE ON <LARGE DEDICATED TSO SYSTEMS.

-USE ON NON-TSO SYSTEMS.

-WHEN THE NUMBER OF PAGING ACTUATORS IS LESS THAN 6-8.

-SINGLE PAGING DEVICE TYPE USED.

DETERMINING PAGING CONFIGURATION MIX.

```
HEAVY TSO OR BATCH SWAPPING.
|
+-------> NO - NO SWAP DATASETS
|
YES
|
DB/DC --> YES - CONSIDER SWAP DS'S.
|
NO
|
ONLY ONE PAGING DEVICE TYPE (E.G. 3350)
|
+-------> NO  - CONSIDER SWAP DS'S.
|
YES
|
PAGING DEVICE TYPE IS 3880-11
|
+-------> YES - CONSIDER SWAP DS'S.
|
+-------> NO  - CONSIDER ALL LOCALS.
```

PAGE DATASET REQUIREMENTS.

- SWAP (OR "ALL LOCAL" PAGE DATASETS)
  IN A HEAVY SWAP ENVIRONMENT.
  -CHARACTERIZED BY HIGH PAGING RATES.

-> DEVICES WITH HIGH DATA XFER RATES
   (3880-11, 3380, 3375, 2305, 3350)
   MAXIMUM PARALLELISM:
   (SWAP GROUP SIZE / 12 = # OF PATHS.)
   AT LEAST 3 ACTUATORS, DISTRIBUTED ON
   AT LEAST 3 (NON-DEDICATED) PATHS.
   OR AT LEAST 2 DEDICATED PATHS.
   ASYMMETRIC CONNECTION OF STRINGS.

- DEMAND PAGING DATASETS.
  MVS: LOW LOAD (BECAUSE OF SWAP DS'S)
  LOW ACCESS TIME IS DESIRABLE.
  RECOMMEND: 3880-11, 3380, 3375, 3350
  USE 3880-11 IF AVAIL., DASD OTHERWISE
  NOT 2305'S (LOW ACCESS DENSITY)
   UNLESS ENOUGH (CAPACITY) AVAILABLE.
  VM: HIGH LOAD (NO SWAP DATASETS)
      USE DEVICES WITH LOW ACCESS TIMES
  RECOMMEND: 3880-11, 2305, 3380.

DEDICATING PAGING DEVICES.

DEDICATED DEVICES. (RECOMMENDED)

ADVANTAGES:

-IMPROVED SEEKING.
(NO SEEKING ON 3880-11 FOR HITS).

-LESS INTERFERENCE WITH NON-PAGE I/O.

-BETTER RESPONSE TIMES.

DISADVANTAGE:

-ONLY SMALL FRACTION OF DEVICE USED.

TRADEOFF:

-SPACE ON DEDICATED DEVICE IS "WASTED", BUT PERFORMANCE GAINED.

-SPACE CAN BE USED FOR DATA "SELDOM" USED DURING INTERACTIVE PERIODS
E.G. HSM ARCHIVES,
SYSGEN LIBRARIES,
NIGHT BATCH DATABASES.

DEDICATING PAGING PATHS.

- DEDICATED SWAP AND PAGE CHANNELS.
  ADVANTAGES:
  - -HIGHER SWAP CHANNEL UTILIZATION POSSIBLE: UP TO 50 %.
  - -NO INTERFERENCE WITH NON-PAGE I/O FROM LONG SWAP CHAINS.
  - -NO INTERFERENCE WITH PAGE I/O BY E.G. VTOC SEARCH.

  DISADVANTAGES:
  - -IF NOT ENOUGH SWAP PATHS, THEN LOSS OF PARALLELISM, WORSE RESP. TIME

  RECOMMEND: DEDICATED DMD PAGING PATHS,
  AVOIDS INTERFERENCE WITH DMD PGING.

- DISTRIBUTED SWAP AND PAGING DEVICES
  RECOMMEND: FOR SWAP DEVICES (& FOR VM)
  ADVANTAGE: MORE PARALLELISM, GOOD R.T.
  DISADVANTAGES:
  - -INTERFERENCE BY NON-PAGING I/O.
  - -LOWER CHAN. UTIL. (30-40 % MAX.)

SUMMARY OF RECOMMENDATIONS.

- FOR BEST TSO RESPONSE TIME:
  - -INCREASE MAIN STORAGE TO MAXIMUM.
  - -SEPARATE SWAPPING FROM DEMAND PAGING
  - -IMPROVE THE PAGING SUBSYSTEM
  - -USE MULTIPLE PARALLEL 3880-11 OR 3380 PATHS FOR SWAPPING.
  - -USE 3880-11 FOR DEMAND PAGING.

PATH CAPACITY PERMITTING:

- -USE STORAGE ISOLATION FOR REDUCED DEMAND PAGING.
- -MINIMIZE LOGICAL SWAP THINK TIME FOR INCREASED PHYSICAL SWAPPING.

- FOR BEST BATCH THROUGHPUT:
  - -DO, AS OUTLINED ABOVE.
  - -USE SRM CONTROLS.

OVERVIEW.

- CONFIGURATION POSSIBILITIES.

->DEVICE CHARACTERISTICS.

- DATASET AND PAGING RATE CALCULATIONS.

- CONCEPTS OF I/O ACTIVITY AND FORMULAE

- EXAMPLES.

- SWAP AND PAGE DEVICE LIMITS.

- CONFIGURATION EXAMPLES.

DEVICE CHARACTERISTICS.

DEVICE TYPE		3350	2305	3375	3380
DATA XFER RATES	MB/S	1.2	1.5	1.8	3.0
TIME OF ROTATION	MS	16.7	10	20	16.7
AVERAGE LATENCY	MS	8.3	5	10	8.3
MAXIMUM SEEK	MS	50	0	38	30
AVERAGE SEEK	MS	25	0	19	16
MINIMUM SEEK	MS	10	0	4	3
NO OF PGS PER TRACK		4+	3+	8	10
NO OF TRACKS PER CYL		30	8	12	15
NO OF SLOTS PER CYL		120	24	96	150
NO OF CYLINDERS		555	96	959	885
CAPACITY	MB/ACT	319	11.2	410	630
NO PG SLTS PER ACT *		65K	2470	65K	65K
NO SWAPSETS / ACT		5.5K	190	7672	10K

* MVS SUPPORTS A MAXIMUM OF 65 K SLOTS

THE 3880-11 BUFFERED PAGING DEVICE.

```
    +---------------------+
    | CHANNEL             |
    +---+------------+----+
        |            |
        |            +--OTHER C.U'S
+-------|-------+
|       |       |   CONTROL UNIT
|   +---+-------|------------+
|   | CACHE     |            |
|   +---+-------|-----+------+
|       |       |     |
|       |       |     +--OTHER STRING
|       |       +-------------+
|       |                     |
|       +------------+        |
|       |            |        |
|   +---+---+    +---+---+    |
|   | 3350  |    | 3350  |    |
|   +-------+    +-------+    |
|       BACK STORE            |
+-----------------------------+
```

3880-11 DEVICE CHARACTERISTICS

CHANNEL DATA XFER RATE MB/SEC		3
CACHE DATA XFER TIME PER PAGE	MS	1.3
CACHE LATENCY TIME PER PAGE	MS	0.7
CACHE BUSY TIME PER PAGE	MS	2

ONE DATA TRANSFER AT ONE TIME.
DASD ACCESS IS ADDITIONAL TO CACHE TIME, IF A MISS OCCURS.
MULTIPLE EXPOSURES TO PREVENT HITS FROM WAITING FOR MISSES.
TWO DEVICES ATTACHABLE TO A CHANNEL.
LESS SENSITIVITY TO RPS MISSES.

HIT RATIO =HR= RATIO OF PAGES FOUND IN CACHE TO TOTAL PAGING RATE TO 3880-11.

HR= .9 -> MISS RATIO = M = 1 - HR = .1
I.E. ABOUT 10% OF TIME DASD IS ACCESSED

OVERVIEW.

- PAGING CONFIGURATION POSSIBILITIES.
- DEVICE CHARACTERISTICS.
- ->PAGING SPACE SIZE AND RATE CALCULATIONS.
- CONCEPTS OF I/O ACTIVITY AND FORMULAE
- EXAMPLES.
- SWAP AND PAGE DEVICE LIMITS.
- CONFIGURATION EXAMPLES.

PAGING SPACE SIZE REQUIREMENTS.

- SWAP DATASETS

SWAP GROUP SIZE, SGS = 36 SLOTS
(IT CAN BE SET TO 48 OR EVEN 60)

SWAP SPACE SIZE=
NUMBER OF USERS X SWAP GROUP SIZE
SDS = NU X SGS

E. G. FOR 100 USERS
SDS = 100 X 36 = 3600 SLOTS =
= 3600 X 4K = 15 MB

- DEMAND PAGE DATASETS

PAGING SPACE SIZE=
NUMBER OF USERS X 120 PAGES

E. G. FOR 100 USERS
SDS = 100 X 120 = 12000 SLOTS =
= 12000 X 4K = 48 MB = 50 MB

SWAP PAGING RATES.

$$\text{TRANSACTION RATE, TR} = \frac{\text{NO OF ACTIVE USERS}}{\text{AVERAGE THINK TIME}}$$

MAXIMUM SWAP PAGING RATE =
2 X SWAP GROUP SIZE X TRANS. RATE

SWAP PAGING RATE =
2 X SWP GROUP SIZE X TR X (1-LOG.SWP %)

E. G. NUMBER OF USERS = 100
SWAP GROUP SIZE = 36 SLOTS
THINK TIME = 25 SECONDS
LOGICAL SWAP = 17 %

SWAP PAGING RATE=
2 X 100 X 36 X .83 / 25 = 240 PGS/SEC

DEMAND PAGING RATES.

->DIFFICULT TO CALCULATE.<-
BUT: POSSIBLE TO CONTROL

MINIMIZE "DEMAND" PAGING RATE BY

- USING STORAGE ISOLATION
  FORCES SWAPIN = SWAPOUT RATE.
  LITTLE SWAPOUT TRIM.
  MINIMIZES TSO DEMAND PAGING.
  (3-5 PAGEFAULTS PER TRANSACTION)

- MINIMIZE OVERINITIATION

- USE ONLY PAGING RATE CONTROL IN SRM
  TO LIMIT MPL. (RCCPTRT)
  -> CONSISTENCY

- LIMIT: 30-60 PAGES/SEC ON 3033 UP

OVERVIEW.

- PAGING CONFIGURATION POSSIBILITIES.
- DEVICE CHARACTERISTICS.
- PAGING SPACE SIZE AND RATE CALCULATIONS.
- >CONCEPTS OF I/O ACTIVITY AND FORMULAE
- EXAMPLES.
- SWAP AND PAGE DEVICE LIMITS.
- CONFIGURATION EXAMPLES.

COMPONENTS OF I/O ACTIVITY.

CONCEPTUALLY:

QUEUE, Q - DUE TO DEVICE BUSY

COMMAND DELAY,CD -SIOF,CC¬=0, PATH BUSY

SEEK

SEEK RECONNECT, SRO (CCW FETCH),
DUE TO PATH BUSY (NOT ON 3880'S)

LATENCY = HALF REVOLUTION

RPS MISS AND RECONNECT

SEARCH (START 3-5 SECTORS BEFORE)

DATA TRANSFER

CONCURRENCY FORMULAE.

MULTIPLE CONCURRENT SWAP I/O'S

-CALCULATE SINGLE PATH RESPONSE TIME

-ADJUST FOR NUMBER OF SWAPSETS/PATHS

-MAXIMUM IS NO OF SWAPSETS X RESP.TIME

SWAP RESPONSE TIME	= RT
SWAPSET RESPONSE TIME	= SRT
NO SWAPSETS IN GROUP	= STS
NO SWAPSETS PER PATH CHANNEL	= SPC
CONTROL UNIT (PATH) UTILIZATION	= CUT
NUMBER OF PATHS (CHANNELS)	= NCH

RT = SRT X (SPC+(STS X CUT ** (NCH-1)))

RT = LESSER OF (RT, STS X SRT)

NCH=3, SRT=50, STS=4, CUT=.3, SPC=1.33

-> RT = 50 X (1.33+(4X(.3) ** 2)) = 85

--> EXPONENTIAL DECREASE WITH PATHS

PARALLELISM AND SWAP RESPONSE TIME.

ASSUME:

- SWAPSET RESPONSE TIME = 50 MS
- FOUR SWAPSETS
- MAXIMUM
  - +-CHANNEL UTILIZATION
  - -I C.U. UTILIZATION = 30 %
  - +-DEVICE UTILIZATION

NO OF PATHS (NCH)	SWAPSETS PER PATH (SPC)	SWAP RESPONSE TIME (RT)
1	4	200
2	2	160
3	1.33	85
4	1	55

CONCEPT OF RELATIVE CHANNEL BUSY.

CONSIDER A SWAP CHANNEL WITH 2 DEVICES
LET CHANNEL BUSY = 50 %
RELATIVE CHANNEL BUSY = RCB = 1/3

WHEN DEVICE D1 TRIES TO RECONNECT, IT
FINDS CHANNEL BUSY 25 % OF TIME FOR D2
FINDS CHANNEL FREE 50 % OF TIME

$$RCB(D1) = \frac{CH.BSY(\neg D1)}{CH.BSY(\neg D1)+CH.FREE} = \frac{25}{25+50} = \frac{1}{3}$$

CHN FREE	D1 BUSY	D2 BUSY
50 %	25 %	25 %

$$RPS = 16.7 \times \frac{1/3}{1-(1/3)} = 8.4 \text{ MS}$$

BASIC CONCEPTS. (REVIEW.)

M/D/1 QUEUE - (DST X DUT)/(2 X (1-DUT))

RELATIVE CHANNEL BUSY: SEEN BY DEVICE

COMMAND DELAY: CD = CHAN SERV.X RCB /2

SEEK (ONE CYLINDER FOR SWAP DATASETS)

SEEK RECONNECT:SRO = CD/2 (NOT ON 3880)

LATENCY: LAT = ROT / 2

RPS RECONNECT: RPS =(ROT *RCB)/(1-RCB)

SEARCH: A FEW SECTORS

TRANSFER DATA: (3380 EXAMPLES)
PAGE: (4096 / 3X10**6) + = 1.5 MS

SWAP: 10 PAGES/TRACK = 16.7 MS
BUT SWAPSET = 12 PAGES =
16.7 + (2/10) X 16.7 = 20 MS

OVERVIEW.

- PAGING CONFIGURATION POSSIBILITIES.
- DEVICE CHARACTERISTICS.
- PAGING SPACE SIZE AND RATE CALCULATIONS.
- CONCEPTS OF I/O ACTIVITY AND FORMULAE

->EXAMPLES.

- SWAP AND PAGE DEVICE LIMITS.
- CONFIGURATION EXAMPLES.

2 3380 SWAP ACTUATORS ON CHANNEL

12 PAGES PER SIO
SIOS PER SECOND PER CHANNEL: 20
CONTROL UNIT OVERHEAD 2.6 MS
DATA TRANSFER TIME 20 MS

CU. UTIL.: 20 X 22.6=452 MS/S = 45 %
REL.CH.UT.: RCB = 22.5/(22.5+55) = 29 %

COMMAND DELAY, CD = 22.6X.29 / 2=3.3 MS
OVERHEAD TIME 2.6 MS
SEEK TIME 3 MS

RPS MISS = 16.7 X .29 / (1-.71) =6.8 MS
LATENCY TIME 8.3 MS
DATA TRANSFER 20 MS
DEVICE SERVICE TIME 40.7 MS
DEVICE UTILIZATION = 20X40.7/ 2=40.7 %

Q (M/D/1)
=40.7 X .407 / (2 X (1-.407)) =14 MS
SWAPSET READ
RESPONSE TIME= 40.7 +14 + 3.3 =58 MS

THIS CALCULATION IS AN APPROXIMATION TO THE DETAILED METHOD

3880-11 SWAP DEVICE WITH 2 3350'S.

12 PAGES PER SIO. CACHE BUSY: 2 MS / PG
HIT RATIO= .95, MISS RATIO = .05
SIOS PER SECOND = SPS = 20
TOTAL PAGING RATE = 20 X 12 = 240 PGS/S

SIO'S TO 3350'S: = SPS X M = 20 X .05=1
CACHE BUSY - CHANNEL= (12 X 2) = 24 MS
CACHE BUSY - 3350=.05 X (50+1) = 2.6MS
CACHE BUSY TIME: 24 + 2.6 = 26.6MS
CONTROL UNIT UTIL. 20 X 26.6 = 53.2 %

QUEUE WAIT, Q = (DST X DU) / 2 X (1-DU)
= 24 X .532 /(2X(1 - .532)) = 13.6MS

3350 RESPONSE TIME (ASSUMED) = 84 MS
READ MISS TIME = 84 X 0.05 = 4.2MS

3880-11 SWAPSET READ RESPONSE TIME:
24 + 4.2 + 13.6 = 41.8MS

THIS CALCULATION IS AN APPROXIMATION TO THE DETAILED METHOD

DISTRIBUTED SWAP DATASETS.

1 3380 SWAP, 3 DB ACT'S PER CHN. 3 CHNS

```
DATA,SWAP SIOS PER SEC PER CHAN =46;6.7
DATA TRANSFER: 1.5 + 2.6        = 4.1MS
CH.UT.(SW)=15.1% ; CH.UT.(DATA) =18.9 %
TOTAL CHN UTIL: 18.9 + 15.1     =34   %

CD(SWAP)    = 0.4MS; CD(DATA)    = 1.1MS
OVERHEAD    = 2.6MS;             = 2.6MS
RCB(SWAP)   =22   % ; RCB(DATA)  =30   %
SEEK(SWAP)  = 3   MS; SEEK(DATA) = 6   MS
RPS(SWAP)   = 4.7MS; RPS(DATA)   = 7.2MS
LATENCY     = 8.3MS;             = 8.3MS
XFER(SWAP)  =20   MS; XFER(DATA) = 1.5MS
S.T.(SWAP)  =38.6MS; S.T. (DATA) =25.6MS
DEV.UT.(SW)=25.9% ; DEV.UT(DATA)=39.1 %
QUEUE(SWAP)= 6.8  ; QUEUE (DATA)= 8.2MS

SRT (SWAP) =45.8MS; R.T.(DATA)  =36.1MS

SWAP RESPONSE TIME =
 45.8 X (1+( 3 X.34 ** ( 3 -1)))=61.7MS
```

3880-11 DEMAND PAGING

2 3350'S ON A 3880-11
HIT RATIO = .6 ; MISS RATIO = .4
TOTAL PAGING RATE = 60 PGS/S

CACHE BUSY TIME PER PAGE FOR HITS = 2MS
3350 RESPONSE TIME - (ASSUMED) =35MS
MISS CONTRIBUTION = .4 X 35 =14MS

CACHE BUSY (CHANNEL)=60 X 2=120MS=12 %
CACHE-3350 PG: (60 X .4 X 4.6) =11 %
CONTROL UNIT UTILIZATION=12 + 11 =23 %

DEMAND PAGE READ RESPONSE TIME:
2 + (.4 X 35) =16MS

- READ HIT RATIOS USUALLY HIGHER THAN OVERALL HIT RATIOS.

- TSO INTERACTIVE READS ARE USUALLY HITS (THE MISSES ARE FOR BATCH.)

THIS CALCULATION IS AN APPROXIMATION TO THE DETAILED METHOD

4 3380 DEMAND PAGE ACTUATORS ON CHAN.

---

1 PAGE PER SIO

SIOS PER SECOND PER CHANNEL: 60
CONTROL UNIT OVERHEAD = 2.6MS
DATA TRANSFER TIME = 1.5MS
CHAN. UTIL.: 60 X 4.1=246 MS/S =24.6 %
CHAN. UTIL./ACTUATOR = .246 / 4 = .061
REL.CHAN UTIL.: RCB = 3 X.061 /(1-.061)
= 0.195 =20 %

COMMAND DELAY, CD = 3.6 X.20 /2 = 0.4MS
OVERHEAD TIME = 2.6MS
SEEK TIME = 6.3MS

RPS MISS = 16.7 X .2 / (1-.2) = 4.2MS
LATENCY TIME = 8.3MS
DATA TRANSFER = 1.5MS
DEVICE SERVICE TIME =22.9MS

DEVICE UTILIZATION = 15 X 22.9 =34 %

Q=(22.9 X .34) / (2 X (1-.34)) = 5.9MS
RESPONSE TIME: 22.9 + 5.9+ .4 =29.2MS

OVERVIEW.

- PAGING CONFIGURATION POSSIBILITIES.
- DEVICE CHARACTERISTICS.
- PAGING SPACE SIZE AND RATE CALCULATIONS.
- CONCEPTS OF I/O ACTIVITY AND FORMULAE
- EXAMPLES.

->SWAP AND PAGE DEVICE LIMITS.

- CONFIGURATION EXAMPLES.

SUMMARY OF SWAP DEVICE USE LIMITS.

- MASSIVE DATA TRANSFERS
- IMPORTANT:PARALLELISM (MULTIPLE PATHS)
- IMPORTANT: FAST TRANSFER RATES
- ACCESS TIME IS OF LEAST IMPORTANCE
- TWO OR MORE 3880-11'S
  OR 3-4 3380 ACTUATORS DISTRIBUTED.
- DEDICATED SWAP CHANNELS CAN BE
  LOADED TO 50 % UTILIZATION.

	3380	3375	3350	3880-11	2305
ACTUATORS	2	2	2	(2)	2
HIT RATIO				.95	
DS SIZE MB	20	20	8	16	20
SIO RATE	20	15	10	20	12
PG RATE	240	180	120	240	144
PGS/ACT	120	90	60		72
RESP.TIME	58	100	130	45	70

SUMMARY OF DEMAND PAGE DEVICE LIMITS.

- LOW DATA TRANSFER RATES.
- MOST IMPORTANT IS ACCESS TIME.
- OF ALMOST EQUAL IMPORTANCE IS CAPACITY
- DEVICE OF CHOICE IS 3880-11.
- GOOD PERFORMANCE: 4 3380 ACTUATORS
- DEDICATED PAGING C. U. UTIL: 35%

	3380	3375	3350	3800-11	2305
ACTUATORS	4	4	4	(2)	2
HIT RATIO				.6	
DS SIZE MB	50	50	50	40	20
SIO RATE	60-80	60	60		40
PG RATE	60-80	60	60	60-80	40
PGS/ACT	15-20	15	15		20
RESP.TIME	26	41	52	16	11

OVERVIEW.

- PAGING CONFIGURATION POSSIBILITIES.
- DEVICE CHARACTERISTICS.
- PAGING SPACE SIZE AND RATE CALCULATIONS.
- CONCEPTS OF I/O ACTIVITY AND FORMULAE
- EXAMPLES.
- SWAP AND PAGE DEVICE LIMITS.

->CONFIGURATION EXAMPLES.

SYSTEM AND WORKLOAD DESCRIPTION.

- KEY FOR PAGING CONFIG'S IS PAGE RATE
  PAGING SPACE SIZE IS OF CONCERN
  FOR 2305 AND 3880-11

- ASSUME 3033 UP, 100 TSO USERS
  - THINK TIME = 25 SECS
  - TRANSACTION RATE = 100 / 25 = 4 TR/S
  - LOGICAL SWAP PERCENTAGE = 17 %

  - PAGE AND SWAP SEPARATION
  - SWAP GROUP SIZE = 36
    3 PAGEFAULTS / TRIVIAL TRANSACTION

  - SWAP SPACE SIZE = 100 X 36 = 15 MB

  - PAGE SPACE SIZE = 100 X 120 = 50 MB

  - SWAP PAGING RATE:
    4 X 36 X 2 X (1-.17) =240 PG/S

  - NON-SWAP PAGING RATE = 60 PG/S

CONFIGURING FOR SWAPPING.

- SPACE = 15 MB, PG RATE=240; AT LEAST 2 PATHS, BETTER: TO 3+ PATHS

- 3350: CAPABILITY= 60 PGS PER SEC/ ACT
  240/60 = 4 ACT'S REQ'D = 2 + PATHS
  ALLOCATE 15 X 2 / 4 = 8 MB / ACTUATOR

- 2305: CAPABILITY= 72 PGS PER SEC/ ACT
  240/ 72 = 4 ACT'S REQ'D. = 2 + PATHS
  ALLOCATE 15 X 2 / 4 =7.5MB / ACTUATOR

- 3375: CAPABILITY= 90 PGS PER SEC/ ACT
  240/90 = 3 ACT'S REQ'D = 2 + PATHS
  ALLOCATE 15X2/3 = 10 MB / ACTUATOR

- 3380: CAPABILITY=120 PGS PER SEC/ ACT
  240/120 = 2 ACT'S REQ'D. = 1 + PATHS
  ALLOCATE 15X2/2 = 15 MB / ACTUATOR

- 3880-11: CAPABILITY=240 PGS / DEVICE
  240/240 = 1 DEV.REQ'D, 1+ PATH, 2 ACTS
  ALLOC.: 15 MB, CACHE = 50 % OF SPACE
  1 DEV. REQ'D. ALLOC: 15X2/2 = 15MB/ACT

CONFIGURING FOR DEMAND PAGING.

- SPACE = 15 MB, PG RATE=60 PGS/SEC
- 3350-3375: CAPABILITY: 15 PGS/ACT
  60/15 = 4 ACT'S REQ'D. 1 PATH
  ALLOCATE 50 X 2 / 4 = 25 MB / ACTUATOR
- 2305: SIZE= 10 MB/ACTUATOR
  ALLOCATION REQUIRED: 50 MB
  50/10 = 5 ACT'S REQ'D + 2 OVFLO 3350'S
- 3380: CAPABILITY=15-20 PGS PER SEC/ACT
  60/15 = 4 ACT'S REQ'D. 1 PATH
  ALLOCATE 50 X 2 / 4 = 25 MB / ACTUATOR
- 3880-11: CACHE SIZE=20% OF PAGE SPACE
  ALLOCATION REQUIRED: 50 MB
  .2 X 50 = 10 MB --> 1+ DEVICE.
  WITH 1 DEVICE: SLIGHTLY LOWER HR,
  AND SLIGHTLY HIGHER RESPONSE TIME
  CAPABILITY: 60-80 PGS PER SEC/ DEVICE
  60 / 60 = 1 DEVICE.

MIXED 3880-11

2 3350 BACK STORE ACTUATORS.
HIT RATIO (OVERALL)= .83

SWAP PAGING RATE = 80 PGS/SEC
DEMAND PAGING RATE = 20 PGS/SEC
SWAP SIO RATE = 6.7 PER SEC
PAGE READ/WRITE SIO RATES= 10 / 5
HR(PAGE)=0.64; HR(SWAP)=0.87

CACHE/CHN BSY: 100 X 2 =20 %
CACHE-3350 : 100 X .17 X 4.6 = 7.2%
CACHE-TOTAL : =27.2%

AVERAGE SERVICE TIME =
272 / (10 + 5 + 6.7) = 12.8 MS

QUEUE TIME, Q = (DST X DU) / (2X(1-DU))
= 12.8 X .272 /(2X (1 - .272))= 2.4MS

3880-11 PAGE READ RESPONSE TIME
= 2 + (.36 X 35) + 2.4 =17 MS

3880-11 SWAPSET READ RESPONSE TIME
=(12X2) + (.13 X 84) + 2.4 =37.3MS

THIS CALCULATION IS AN APPROXIMATION TO THE DETAILED METHOD

TSO RESPONSE TIME ANALYSIS.

TRIVIAL RESPONSE TIME OBJECTIVE: 500 MS
TRIVIAL TSO TRANSACTION COMPONENTS.
-CPU COST: ABOUT 250K INSTRUCTIONS.
-I/O COST: 2 I/O'S PER TRANSACTION.
-SWAP-IN TIME
-3 PAGE FAULTS PER TR. (STOR. ISOL.)

CPU SERVICE TIME = .25 / 5 = .05 SEC
CPU UTILIZATION DUE TO TSO TRIVIAL
4 TR/SEC X .05 SEC/TR = .2 = 20 %
ASSUME 10 % HIGHER PTY TASKS CPU

THEN:
CPU RESPONSE TIME
.05 / ((1-.2)) = 60 MS, ABOUT 70 MS

I/O RESPONSE TIME: 50 MS PER I/O
TOTAL I/O RESPONSE TIME: 2 X 50= 100 MS

PAGING RESPONSE TIME: ?
SWAP RESPONSE TIME: ?

COMPARISON OF SUBSYSTEMS.

DEVICE	ACTS	PATHS	RESPONSE TIME SWAP	PER PAGE	TOTAL
1. S3380	2	1	174		
P3380	4	1		26	392
2. L3380	4	2	118	35	373
3. S3880-11	2	1	135		
P3880-11	2	1		16	330
4. S3380(DIST)	3	3	62		
P3380	4	1		26	299
5. S3380(DIST)	3	3	62		
P3880-11	2	1		16	*269*
6. M3880-11	6	3	47	20	*269*
7. S3350	4	2	390		
P3350	4	1		52	650

SWAP MEASUREMENTS (1)

- RMF REPORTS SWAPSET SERVICE TIME IN SWAP DATASET REPORT, THIS DOES NOT INCLUDE ALL QUEUEING. ASM QUEUEING INCLUDED IN RCVSWPTM.

- SWAPSETS DO NOT ARRIVE EXPONENTIALLY BUT LUMPED. CONSIDER 4 SWAPSETS.

```
+---+---+---+   +----+   +---+    +---+
|   |   |   |   |IORB|   |   |    |   |
|ST |ST | Q |--+----+--+DEV+---+CU  |
|   |   |   |   |IORB|   |   |    |   |
+---+---+---+   +----+   +---+    +---+

+---+---+---+   +----+   +---+    +---+
|   |   |   |   |IORB|   |   |    |   |
|ST |ST | Q |--+----+--+DEV+---+CU  |
|   |   |   |   |IORB|   |   |    |   |
+---+---+---+   +----+   +---+    +---+
```

RT= ((Q+ST)+(Q+2XST))/2 = Q + ST X 1.5

SWAP MEASUREMENT (2)

- RMF DOES NOT (DIRECTLY) REPORT SWAP RESPONSE TIME.

- RCVSWPTM USUALLY > SWAPSET RESPONSE TIME.
  (REPORTED SWAP DATASET SERVICE TIME)

- IF RCVSWPTM >> SWAPSET RESPONSE TIME, THEN SERIOUS QUEUEING IN ASM.

PAGING MEASUREMENTS

- LOCAL PAGE DATASETS HAVE
  - PAGE FAULTS (1 PER "SIO")
  - PAGE WRITES (1.7+ PER "SIO")
  - PAGE TRIMS (1-8 PER "SIO")

- LOCAL PAGE DATASETS HAVE
  I/O REQUEST COUNTS, BUT
  WHAT IS BLOCKING FACTOR???
  MUST CALCULATE IT BY ASSUMPTIONS.

- PAGE READ RESPONSE TIME NOT
  REPORTED, ONLY PAGE TRANSFER TIME.
  READ RESPONSE IS PROBABLY WORSE
  THAN PAGE TRANSFER TIME.

SUMMARY OF METHODOLOGY

- DETERMINE SWAP AND DEMAND PAGING RATES.

  - CONSTRUCT PAGING CONFIGURATIONS USING DEVICE LIMIT TABLES

  - FOR RESPONSE TIME CALCULATIONS USE DESIGN GRAPHS (FOR DEDICATED CHANNELS)

  - FOR MIXED CONFIGURATIONS USE DETAILED MODELING METHODOLOGY

CONCLUSIONS

- FOR BEST PERFORMANCE
  - USE MULTIPLE 3380 OR 3880-11 SWAP PATHS.
  - USE 3880-11 DEMAND PAGING.
- IT IS POSSIBLE TO DETERMINE PAGE/SWAP CONFIGURATIONS BY SIMPLE CALCULATIONS.

G320-0053-0

Page/Swap Configurations Technical Report Printed in U.S.A. G320-0053-0