Performance Profile

IBM 3990-6 with 3390-3s

October 1995

Performance Associates, Inc.
72-687 Spyglass Lane
Palm Desert, CA 92260
(619) 346-0310

PAI/O Driver® is a trademark of Performance Associates, Inc.

MVS/SP ™, ***MVS/XA*** ™, ***MVS/ESA*** ™, ***MVS/DFP*** ™, ***ESA/390*** ™, ***DFHSM*** ™, ***DFSORT*** ™, ***Hiperspace*** ™, ***Hiperbatch*** ™, ***DFSMS*** ™, ***DFSMS/MVS*** ™, ***PR/SM*** ™, ***RACF*** ™, ***RAMAC*** ™, ***DFDSS*** ™, ***System/360*** ™, ***System/370*** ™, ***System/390*** ™, ***Parallel Sysplex*** ™, ***ESCON*** ™, ***VM/ESA*** ™, ***VSE/ESA*** ™, ***ES/3090*** ™, ***ES/9000*** ™, ***EMIF*** ™, and ***IBM*** ® are trademarks or registered trademarks of the IBM Corporation.

Revised April 7, 1996

Table of Contents

IBM 3990-6 with 3390-3s 1
1.1 Logical Structure 3
1.2 Experimental Results 7
1.2.1 Uniform Tests 9
1.2.2 Skewed Tests 13
1.2.3 Burn Through Tests 17
1.2.4 Maximum Stress Tests 22
1.2.5 Front End Bandwidth Tests 25
1.2.6 Record Level Cache Tests 28
1.2.7 Sequential Tests 30
1.3 Observations, and Comments, and Hypotheses 33
1.3.1 Observations 34
1.3.1.1 Ongoing Tuning Requirements 35
1.3.1.2 Aggregate Data Transfer Rates 36
1.3.1.3 Channel Data Transfer Rates 38
1.3.1.4 Focusing on Access Density 41
1.3.2 Comments 42
1.3.3 Hypotheses 44
1.4 Acquisition Strategies 45

IBM 3990-6 with 3390-3s

*Please note that the narrative of this performance profile assumes that you have read the acquisition methodology manual **MVS DASD Subsystems: Understanding, Evaluating, and Acquiring New Technologies**. Unless you have read this manual, it may be difficult for you to interpret the results presented in this document.*

The 3990-6 control unit configured with thirty-two or sixty-four 3390-3 DASD devices was IBM's highest performance storage subsystem offering during the early 1990s. While IBM has announced the withdrawal of the 3390 models 1, 2, and 3 along with all of the models of the 3990 prior to the 3990-6 effective April 26th, 1996, 3390-3s are both abundant and reasonably priced in the secondary market. These devices are particularly attractive for installations with short term storage requirements or for installations who do not require the environmental relief provided by IBM's RAMAC family or similar commodity device based offerings from other vendors.

This performance profile is the result of a subsystem test which was conducted in June of 1994, just prior to start of shipments of the initial RAMAC subsystem, with the maximum configuration for the subsystem. Hence, the results discussed in this DASD Subsystem Performance Profile may be considered as representative of the final delivered state of the subsystem. Should you have questions about this report, Performance Associates is pleased to consult via phone with its clients.

1.1 Logical Structure

The 3990-6 is a result of the continued evolution of the 3990 control unit family which was introduced in 1989. An overview of the logical structure of the subsystem configured with 3390-3 devices is provided in Figure 1.

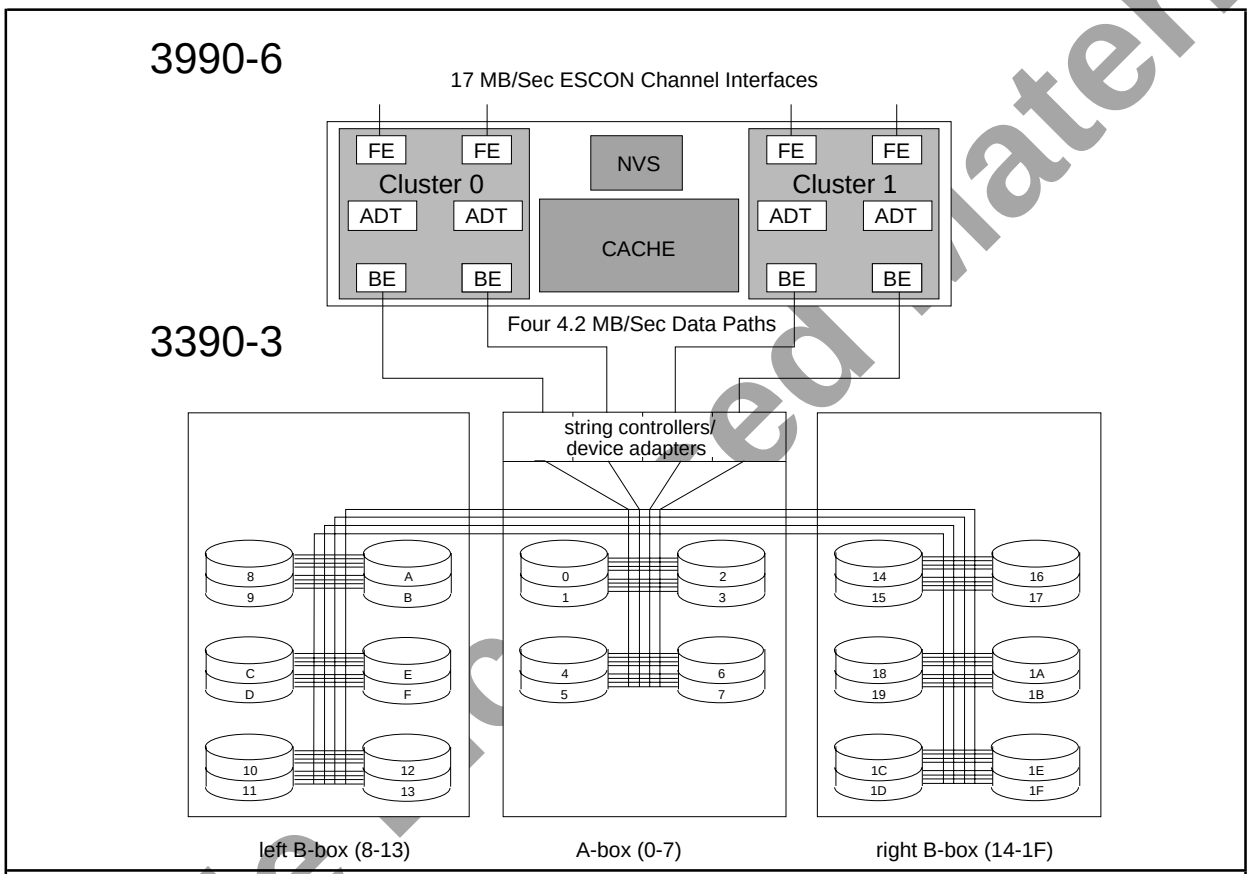

Figure 1. 3990-6 3390-3 Subsystem Logical Structure

As is the case with all of the members of the 3990 control unit family, the 3990-6 control unit is comprised of two storage clusters (i.e., Cluster 0 and 1) each of which supports a multipath storage director (MPSD). Each MPSD provides two storage paths for a total of

four storage paths through the subsystem. One of the key features of the 3990 architecture is that each device may be addressed by any of the four storage paths. IBM employs the term *device level select extended* (DLSE) to denote four path connectivity.

The 3990-6 provides four front end interfaces, each of which can support two ESCON or four parallel channel interfaces. The ESCON interfaces are rated at 17 MB/Sec. The control unit also provides four back end interfaces, each of which provides a 4.2 MB/Sec path to the storage devices. Depending on the characteristics of the I/O requests being processed, the front end and back end of each storage path may operate in conjunction (e.g., a read miss) or independently (e.g., the front end serves read hit while the back end stages or destages other data). While it is statistically unlikely, the control unit could support eight simultaneous data transfers. In any event, the maximum theoretical front end and back end bandwidths are 68 and 16.8 MB/Sec respectively.

The 3990-6 control unit also incorporates *non-volatile storage* (NVS) and cache resources for storing write and read data respectively. Please note that data waiting to be written is stored in both the NVS and cache to provide fault tolerance. That is, the failure of either of the cache resources does not result in the loss of data waiting to be written.

Below the control unit in the figure, a full string (i.e., 32 devices) of 3390-3s is depicted. These devices are packaged in either a front end interface box (i.e., an A8) or as a dedicated storage box (i.e., a B12). The configuration in the figure is comprised of an A8 plus two B12-s. The 3990-6 may support two full strings (i.e., 64 devices) of 3390-3.

From an architectural standpoint, there are several key questions which must be answered about the subsystem. Specifically, they are:

- what is the effective channel interface data transmission rate? Specifically, how much of the 17 MB/Sec data rate provided by ESCON channels can be exploited by the architecture?
- how effective is the subsystem's RLC I (record level cache) support for dealing with record level read and write miss workloads?
- what are the maximum aggregate data rates supported by the subsystem? That is, how much data can the subsystem actually transfer in read, write, and pure cache hit operations?

Each of these questions will be addressed in the analysis of the data collected during the study.

1.2 Experimental Results

PAI/O Driver engineering test series was employed to evaluate a 3990-6 3390-3 subsystem in June of 1994. The specific characteristics of the subsystem are provided in Table 1.

Table 1. 3990-6 3390-3 Subsystem Configuration

Control Unit Feature	Evaluated Configuration
Microcode Level	C87236
Cache Memory	1 GB
Nonvolatile Storage (NVS)	16 MB
Number of Channels	4
Channel Type	ESCON
Channel Data Rate	17 MB/Sec
Maximum Logical Paths	128
Device Type	3390-3s
Device Capacity	2.8 GB
Number of Physical Volumes	64
Subsystem Capacity (units of 10^9 bytes)	180
Aggregate Size of the Test Data Sets	40 GB

The reader may wish to note that the 3990-6 can support two full strings of 3390 model 1, 2, 3, or 9 devices. The results provided in this DASD subsystem performance profile are representative of the SSCH rates which can be sustained by model 1, 2, or 3 devices[1]. However, it would be invalid to apply these results to 3390-9 devices since the rotational

1 Of course, there would be significant differences in access density. Specifically, the access densities of 3390-1 devices would be three times larger than the 3390-3s since the 3390-1s provide only one third of the storage capacity. In the same manner, the access densities of the 3390-2s would be 1.5 as high as those for 3390-3s.

delay of the 3390-9 is three times higher than the other 3390 models. The physical implementation of the 3390-9 volume image, as provided by IBM[2], ***should be viewed as a bulk storage device rather than as a performance DASD subsystem.***

2 Other vendors provide other physical and logical implementation of the 3390-9 image whose performance profiles (in terms of SSCH rate) meet or exceed standard 3390-3 devices.

1.2.1 Uniform Tests

The uniform test series is intended to determine the write sensitivity of the subsystem as well as providing an estimate of the maximum *credible performance* for the subsystem. The uniform test series was conducted for the 3990-6 3390-3 subsystem using both 4K and half track (27998 bytes) block sizes. The results from the 4K study are shown in Figure 2.

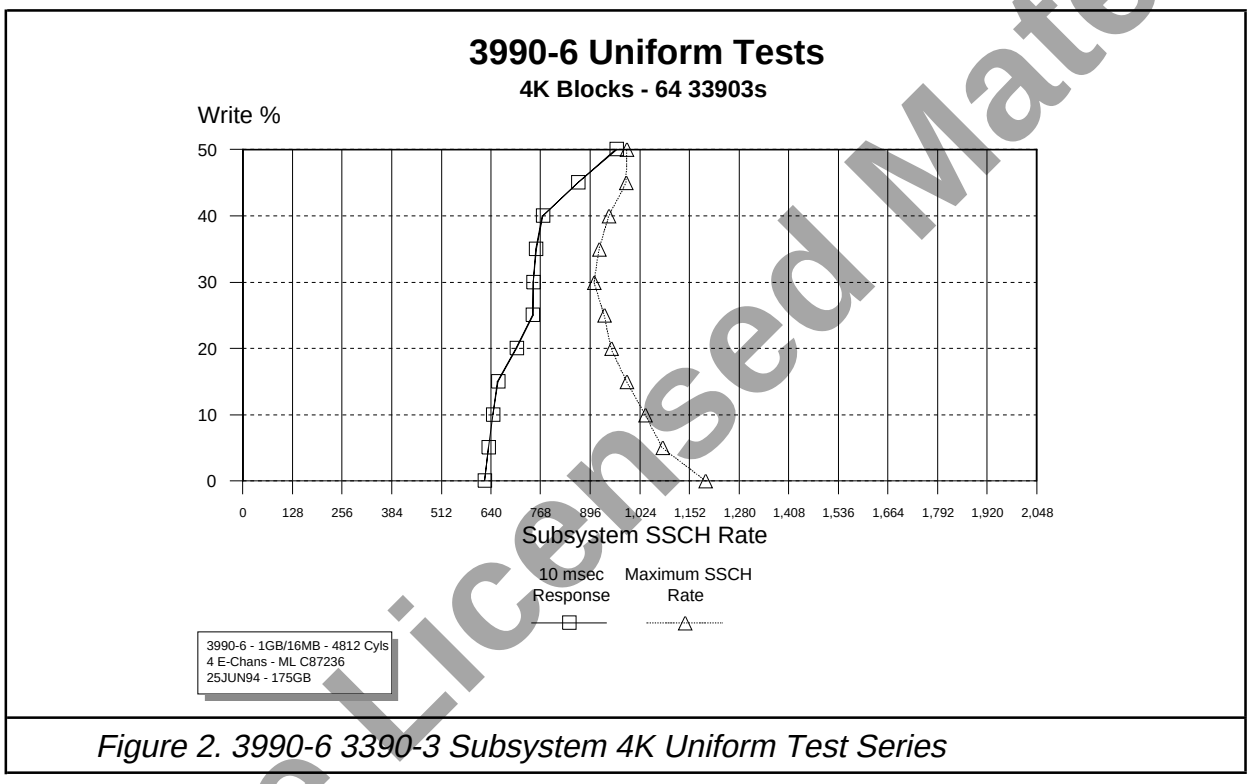

Figure 2. 3990-6 3390-3 Subsystem 4K Uniform Test Series

As can be seen in the figure, the 10 msec envelope for the 3990-6 3390-3 subsystem demonstrates proverse write sensitivity as depicted by the line with open squares. The subsystem offers from 640 to 960 SSCHs/Sec at write fractions varying from 0 to 50%. The maximum throughput line for the 4K uniform case demonstrates a slight adverse sensitivity to write fraction, depicted by the line with open triangles. After achieving a maximum rate of 1,190 SSCHs/Sec at a 0% write fraction, the maximum rate for the subsystem dropped to a minimum of 900 SSCHs/Sec at higher write fractions.

Figure 3 presents the 10 msec and maximum responses line for the 4K uniform test series results in access density format. Access density format was selected to avoid the invalid conclusions which are often drawn from SSCH rate oriented presentations. An ideal DASD subsystem would be linearly scaleable. That is, every time the capacity doubled, the service level constrained SSCH rate would double. For such an ideal subsystem, the access density would be the same, independent of the quantity of storage supported by the control unit.

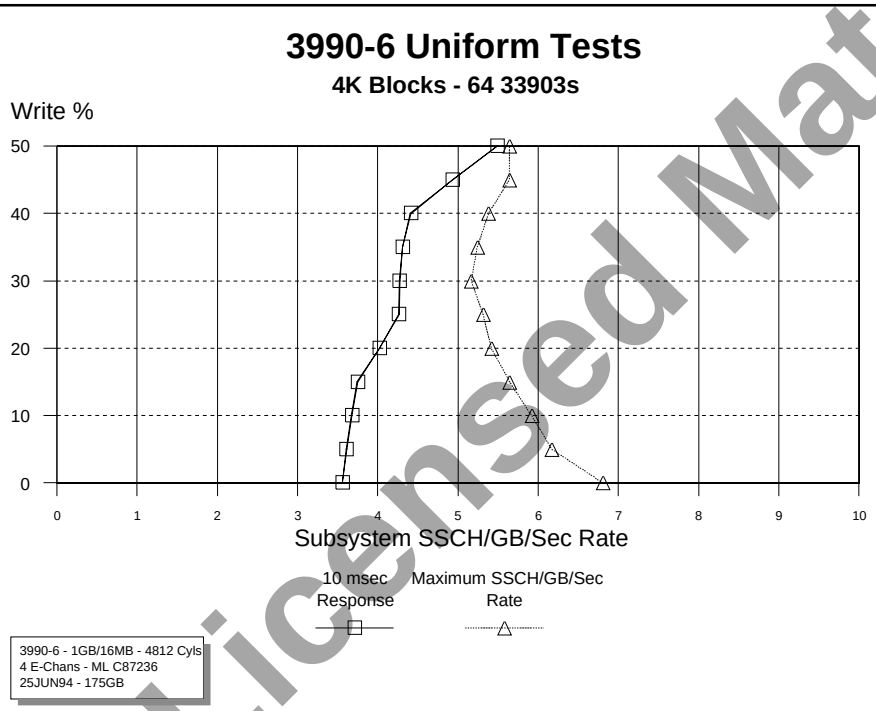

Figure 3. 3990-6 3390-3 Subsystem 4K Uniform Test Series, Access Density Format

The results for the 10 msec and maximum throughput values are depicted by the lines with open squares and triangles respectively. The 10 msec response line varies from 3.6 to 5.7 SSCHs/GB/Sec with a pronounced proverse write fraction sensitivity. In comparison, the maximum throughput line varies from 6.9 to 5.2 SSCHs/GB/Sec with an adverse write sensitivity.

Figure 4 provides the uniform results for the half track tests for the 64 volume configuration of the 3990-6 3390-3 subsystem. Please note that a 20 msec constant response time is employed for the half track tests. The 20 msec response line, depicted by the line with open squares, averages 640 SSCHs/Sec while the maximum throughput line ranges from 630 to 910 SSCHs/Sec with an adverse write sensitivity, depicted by the line with open triangles.

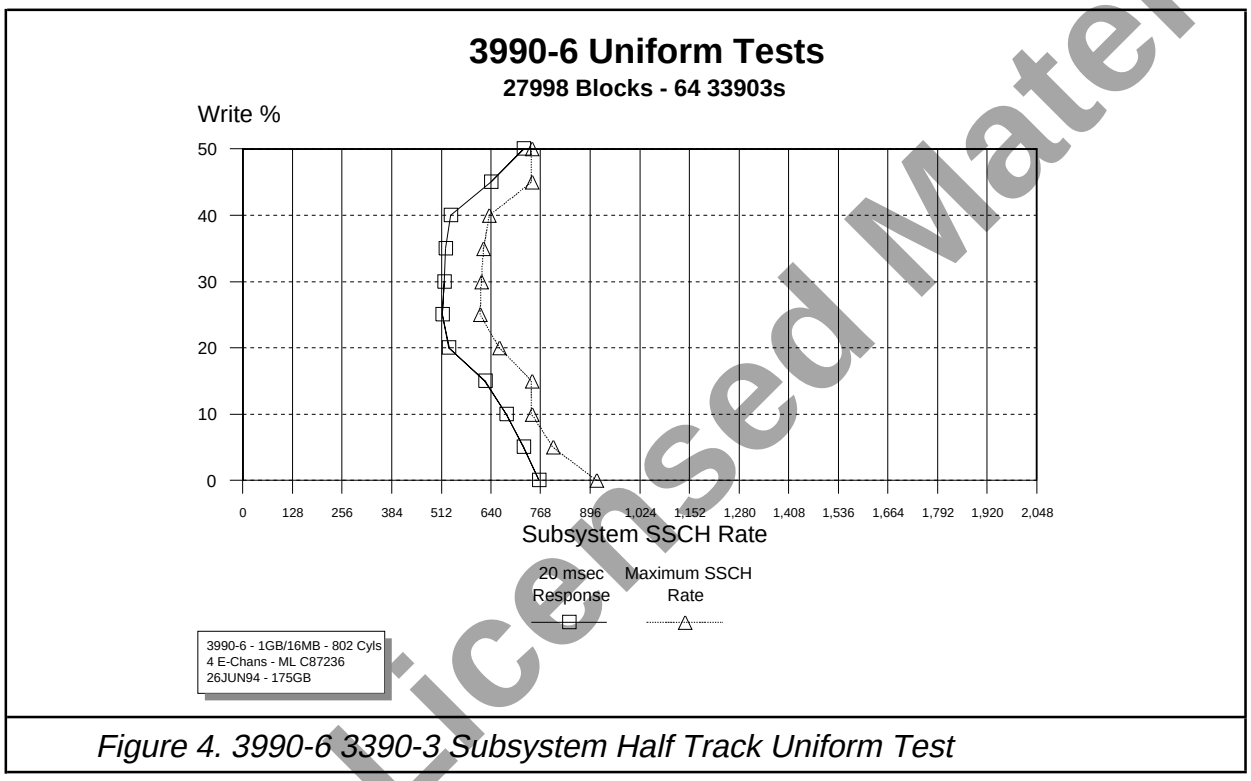

Figure 4. 3990-6 3390-3 Subsystem Half Track Uniform Test

Figure 5 presents half track results in access density format. The 20 msec response line, depicted by the line with open squares, averages 3.6 SSCHs/GB/Sec while the maximum throughput line ranges from 5.2 to 3.6 SSCHs/GB/Sec with an adverse write sensitivity, depicted by the line with open triangles.

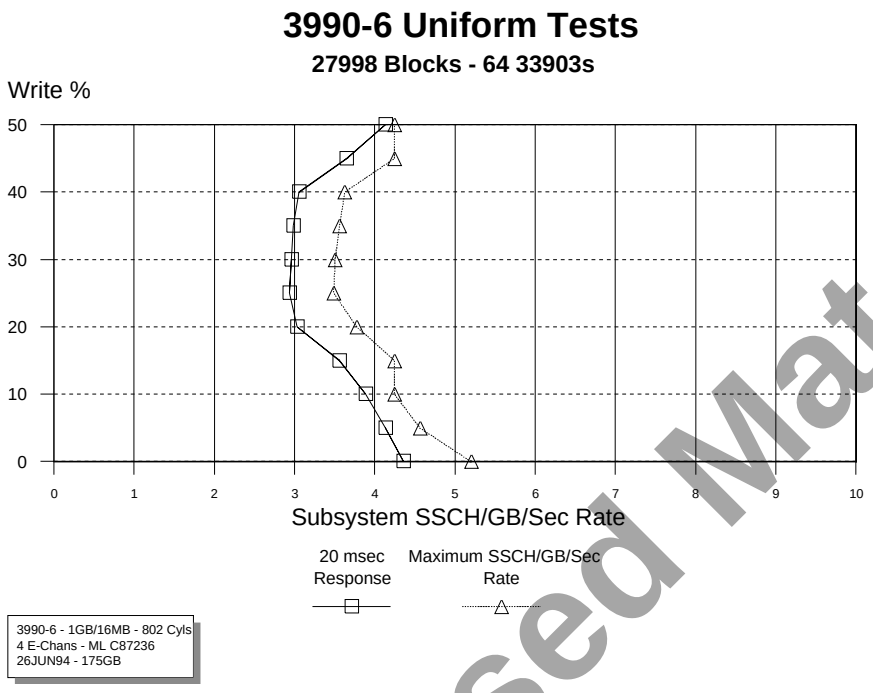

Figure 5. 3990-6 3390-3 Subsystem Half Track Uniform Test Series, Access Density Format

1.2.2 Skewed Tests

The skewed test series is intended to determine the performance and write sensitivity of the subsystem for a more realistic workload environment. The skewed test series was conducted for the 3990-6 3390-3 subsystem using both 4K and half track (27998 bytes) block sizes. The results from the 4K study are shown in Figure 6. It is important to remember that the I/O content (arrival rate, seek distances, and hit rates) of each skewed test point is identical to the corresponding uniform test point. Hence a comparison between the two studies will allow some inferences about the ongoing data set level tuning requirements for the subsystem to be made.

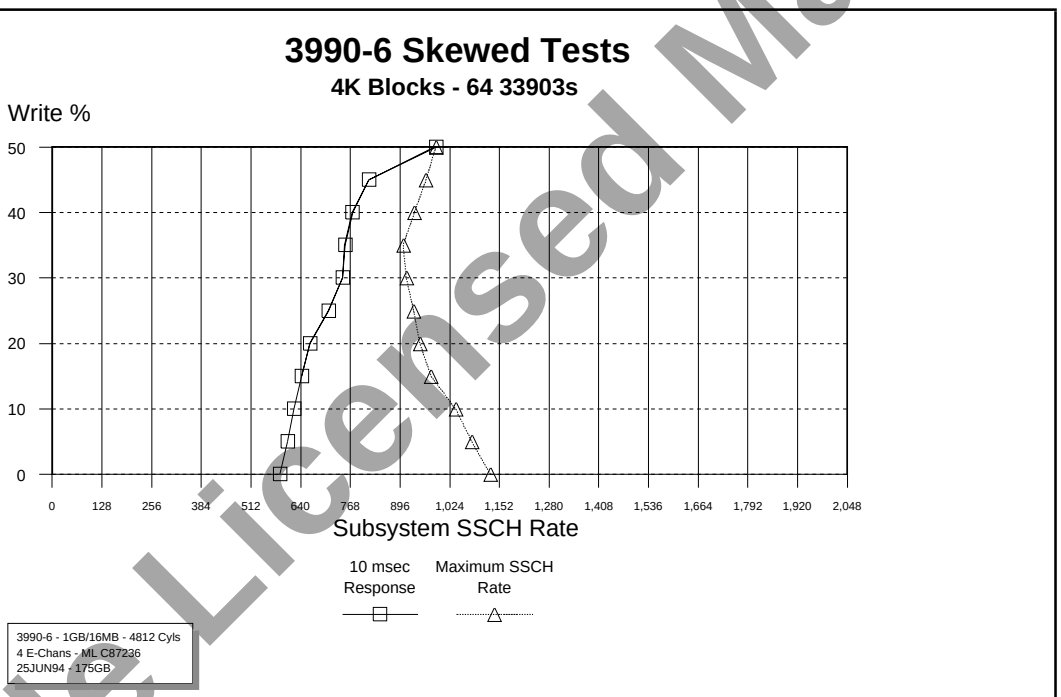

Figure 6. 3990-6 3390-3 Subsystem 4K Skewed Test Series

As can be seen in the figure, the 3990-6 3390-3 subsystem in this configuration demonstrates proverse write sensitivity, i.e., the 10 msec envelop increases with write fraction. Specifically, the subsystem offers 600 SSCHs/Sec at 0% write fraction and a maximum of 980 SSCHs/Sec at a 50% write fractions. This behavior is depicted by the

line with open squares in the figure. In comparison, the maximum throughput line for the 4K skewed case demonstrates an adverse sensitivity to write fraction, depicted by the line with open triangles which varies from 1140 to 900 SSCHs/Sec.

Figure 7 presents the data shown in the prior figure translated into access density format. The 10 msec line (lines with open squares) varies from 3.3 to 5.7 SSCHs/GB/Sec and the maximum throughput line (line with open triangles) varies from 6.4 to 5.15 SSCHs/GB/Sec.

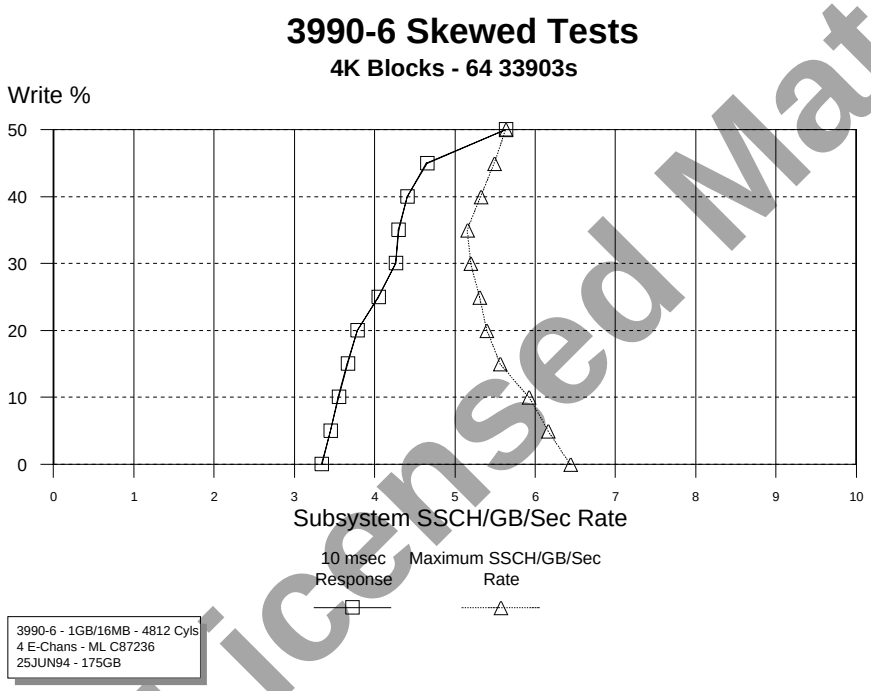

Figure 7. 3990-6 3390-3 Subsystem 4K Skewed Test Series, Access Density Format

The skewed results for a half track blocksize is shown in Figure 8. Both the 20 msec and maximum throughput lines demonstrate adverse write sensitivity. Please note that a 20 msec constant response time is employed for the half track tests. The 20 msec response line, depicted by the line with open squares, averages 650 SSCHs/Sec while the maximum throughput line ranges from 900 to 620 SSCHs/Sec, depicted by the line with open triangles.

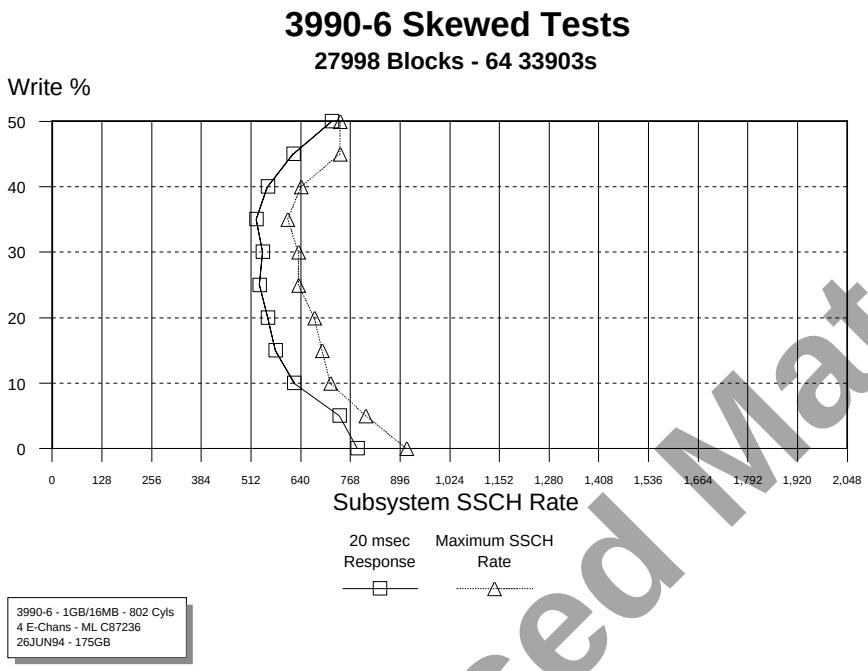

Figure 8. 3990-6 3390-3 Subsystem Half Track Skewed Test

The half track skewed results are presented in access density format in Figure 9. The 20 msec response line, depicted by the line with open squares, averages 3.8 SSCHs/GB/Sec while the maximum throughput line ranges from 5.2 to 3.5 SSCHs/GB/Sec, depicted by the line with open triangles.

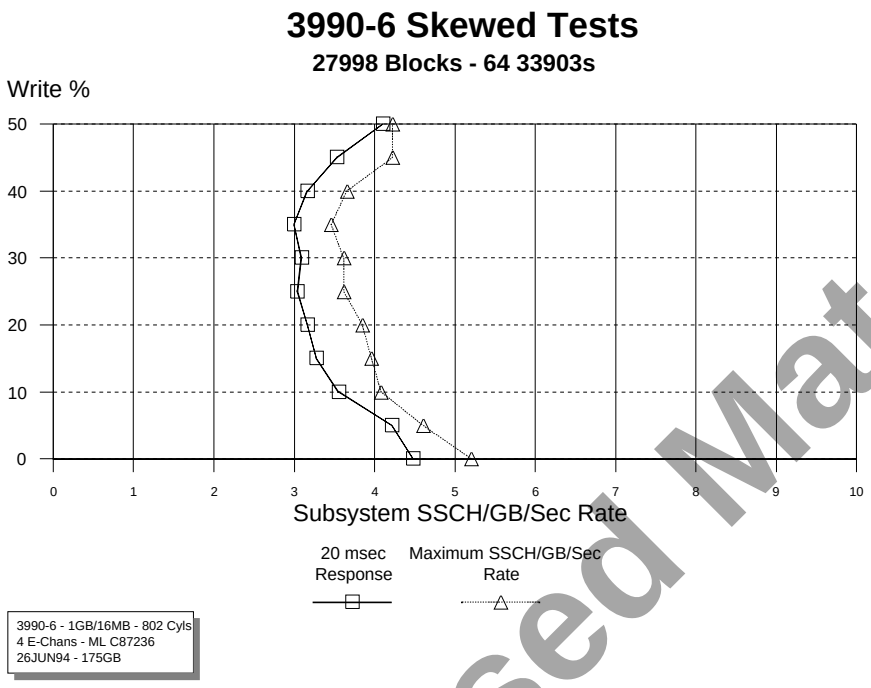

Figure 9.3990-6 3390-3 Subsystem Half Track Skewed Test Series, Access Density Format

One important observation can be made by comparing each of the four graphs in this section with their counterparts in Section 1.2.1, which discusses the uniform results. Specifically, the differences between the skewed and uniform envelopes range from 5 to 10%, with the maximum differences occurring for small block sizes with low write fractions. Hence, the 3990-6 3390-3 subsystem would tend to require some ongoing device level tuning if the subsystem were employed for DFSMS managed data sets.

1.2.3 Burn Through Tests

The burn through tests are intended to evaluate the impact of a single very busy device on the subsystem as a whole. In addition, the burn through tests also allow the subsystem to demonstrate the capability of the controller to boot-strap a high activity data set into cache. A total of four burn through tests were conducted. The first two were based on a 4K block size and evaluated the subsystem's response to random and very high locality stressed devices respectively.

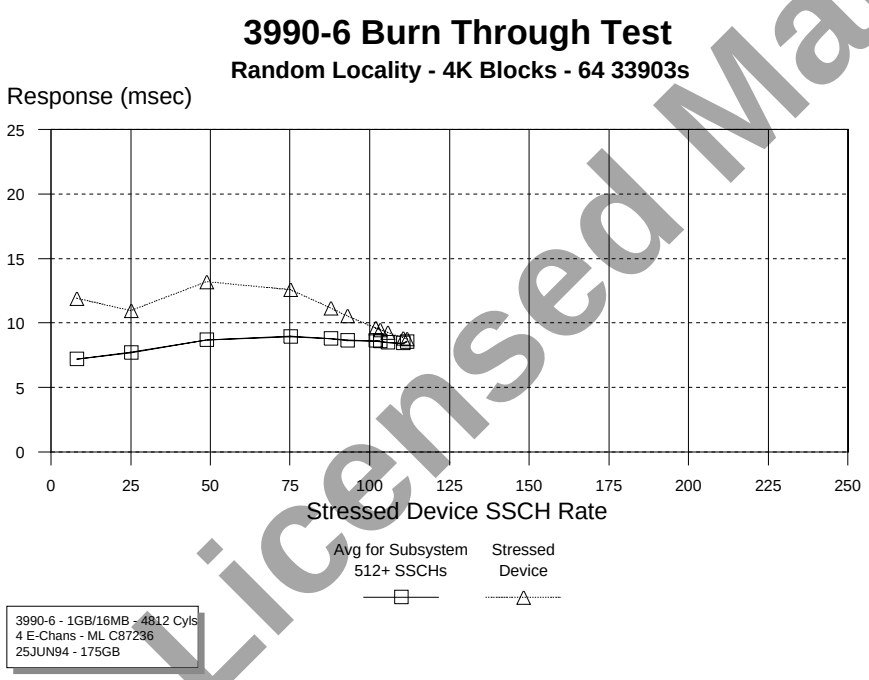

Figure 10. 3990-6 3390-3 Subsystem 4K Random Locality Burn Through Test

Figure 10 provides the results for a stressed random locality data set. There are two lines presented in the figure which are denoted with open squares and open triangles. The first is the average subsystem response time (open squares) and the second is the response time (open triangles) for the stressed device. It is important to note that both of these lines are plotted versus the arrival rate for the stressed device, even though the arrival rate for

the 63 background devices was held constant at 512 SSCH/Sec. The key result from this test is that, at worst, negligible degradation is observed in the average response time for the subsystem as a whole when the stressed device achieved its maximum SSCH rate, i.e., 114 SSCH/Sec. At this point, the stressed device utilization was approaching saturation since the arrival rate (114) times the response time (8.5 msec) is approximately 97%.

Figure 11 depicts the same results for a stressed very high locality data set. Rather than being greater than the average response time for the subsystem as a whole, the stressed data set demonstrates a lower average response time since it has a higher read hit percentage (83%) than the read hit percentage (70%) for the workload as a whole. *The critical characteristic demonstrated by both of these 4K tests is the complete insensitivity of the 3990-6 3390-3 subsystem's architecture to extreme skewing conditions.*

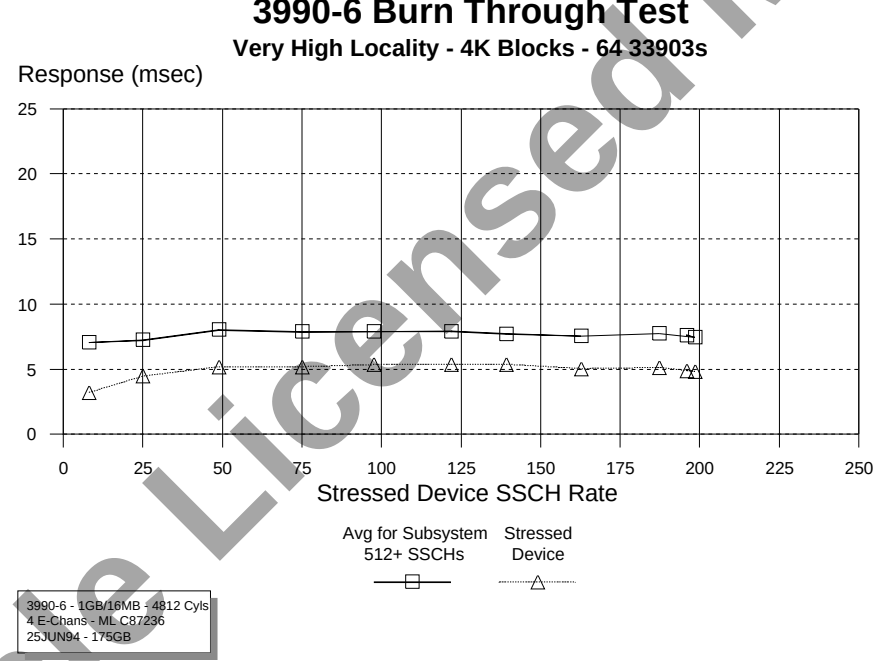

Figure 11. 3990-6 3390-3 Subsystem 4K Very High Locality Burn Through Test

The second set of burn through tests was conducted with a half track blocksize. The half track random locality results are shown in Figure 12 and the very high locality results are shown in Figure 13.

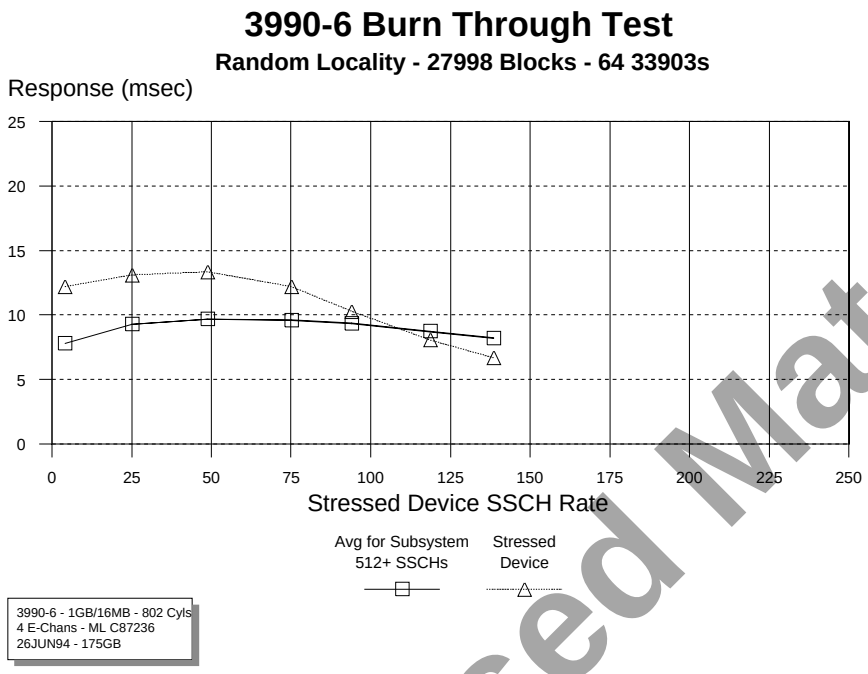

Figure 12. 3990-6 3390-3 Subsystem Half Track Random Locality Burn Through Test

As was the case with the 4K tests, the half track tests showed that the average response of the subsystem as a whole was unaffected by the activity of the stressed device. Moreover, the limit (i.e., highest stressed device SSCH rate) of both tests was defined by the saturation of the stressed device. *For example, the stressed device utilization for the final point of the random locality test was in excess of 75%, i.e., 143 SSCHs/Sec times 5.3 msec per SSCH.* These results completely corroborate the observations provided for the 4K tests.

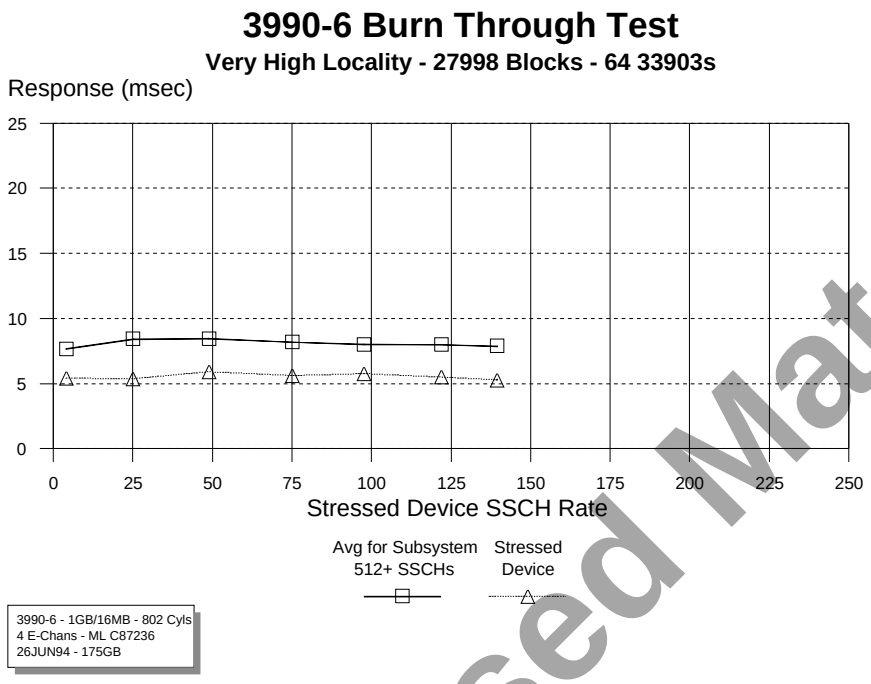

Figure 13. 3990-6 3390-3 Subsystem Half Track Very High Locality Burn Through Test

1.2.4 Maximum Stress Tests

The maximum stress test series is intended to determine the capability of the subsystem to process the type of intensive read/write update activity that typifies OLTP workloads. A total of four maximum stress tests were conducted. The first two are based on a 4K block size and evaluate the subsystem's response to random and very high locality read/update workloads. The second set of tests evaluated the same scenario using a half track blocksize.

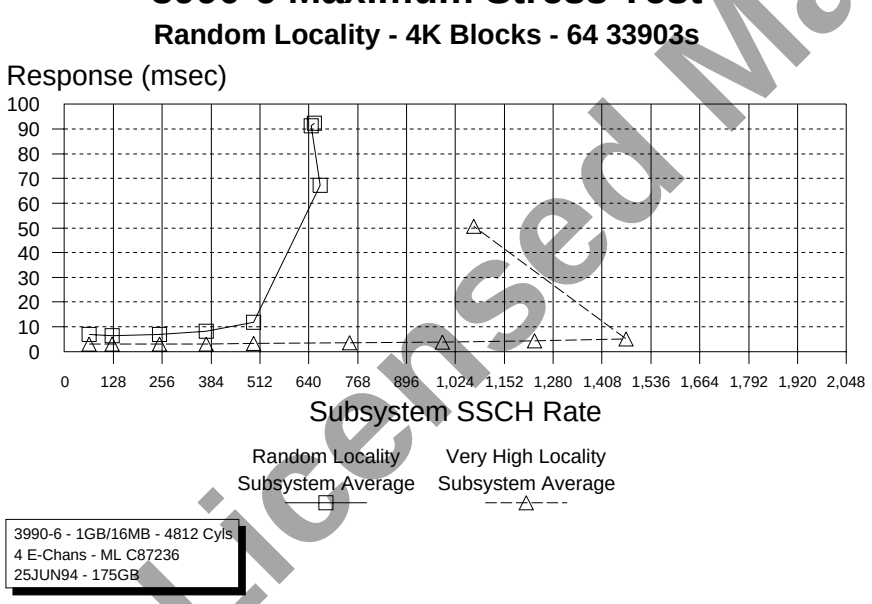

Figure 14. 3990-6 3390-3 Subsystem 4K Random and Very High Locality Maximum Stress Tests

Figure 14 provides an overview of the 4K maximum stress test for both random and very high locality data sets which are depicted by lines with open squares and lines with open triangles respectively. The maximum SSCH rate of the random locality test was approximately 660 SSCHs/Sec. It is interesting to note that the next point in the series

actually achieves a lower total SSCH rate with a higher response time. The reason for this behavior was writes which had to wait due to NVS saturation. The RMF Cache Reporter counts such waits as fast write delays.

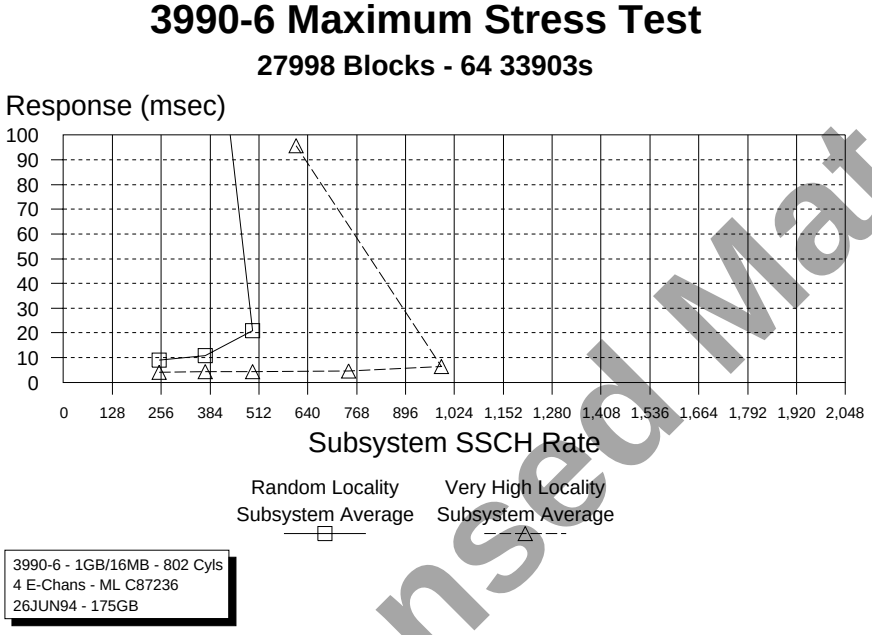

Figure 15. 3990-6 3390-3 Subsystem Half Track Random and Very High Locality Maximum Stress Tests

The second line in the figure provides the same results for the very high locality data sets, which experience an 88.7% hit ratio. The maximum SSCH rate achieved for the very high locality test series was 1490 SSCH/Sec. The significant performance limitation imposed by the subsystem's NVS implementation is demonstrated by the 40+% drop in the very high locality rate (i.e., 1490 to 1050) when saturation occurs. Please note the speed and bandwidth of the back end of the subsystem have very little impact in this test since approximately 90% of the reads and all of the writes can be served by just the front end and cache.

The same two tests were repeated using half track blocksizes. The results of these tests are shown in Figure 15. The results of the half track tests are similar in form to the 4K

results, with the higher read hit ratio data sets demonstrating higher throughputs and lower response times. Once again, the significant limitations imposed by the NVS design are evident in both the random and high locality results.

One interesting comparison may be made between the 4K and half track blocksizes. Specifically, it is the benefits which ESCON channels provide for large blocksizes. The response time for the half track blocks (at low SSCH rates before channel or device contention becomes a factor) are only about 30% greater than the 4K response times even though seven times as much data is being transferred. This is a result of the fact that the overheads associated with establishing an ESCON link are small compared to the substantial benefit of reduced data transfer time.

1.2.5 Front End Bandwidth Tests

The front end bandwidth test series is intended to evaluate the effective data transfer rate and ESCON protocol delays for the control unit. The front end bandwidth tests are conducted with 4K and half track blocksizes. Figure 16 provides the results for a 4K block size. Please note that the data presented in this figure is very likely to be similar to the ***"cache friendly results"*** often employed in a variety of vendor marketing presentations, where a specific definition of the term ***cache friendly*** is never provided. If you are fortunate enough to experience very high hit ratios (i.e., 95%+) in your own environment, then these results will probably be most representative of the actual performance you observe. The maximum demonstrated 4K front-end hit SSCH rate was 2600 SSCH/Sec as depicted by the line with open squares in the figure.

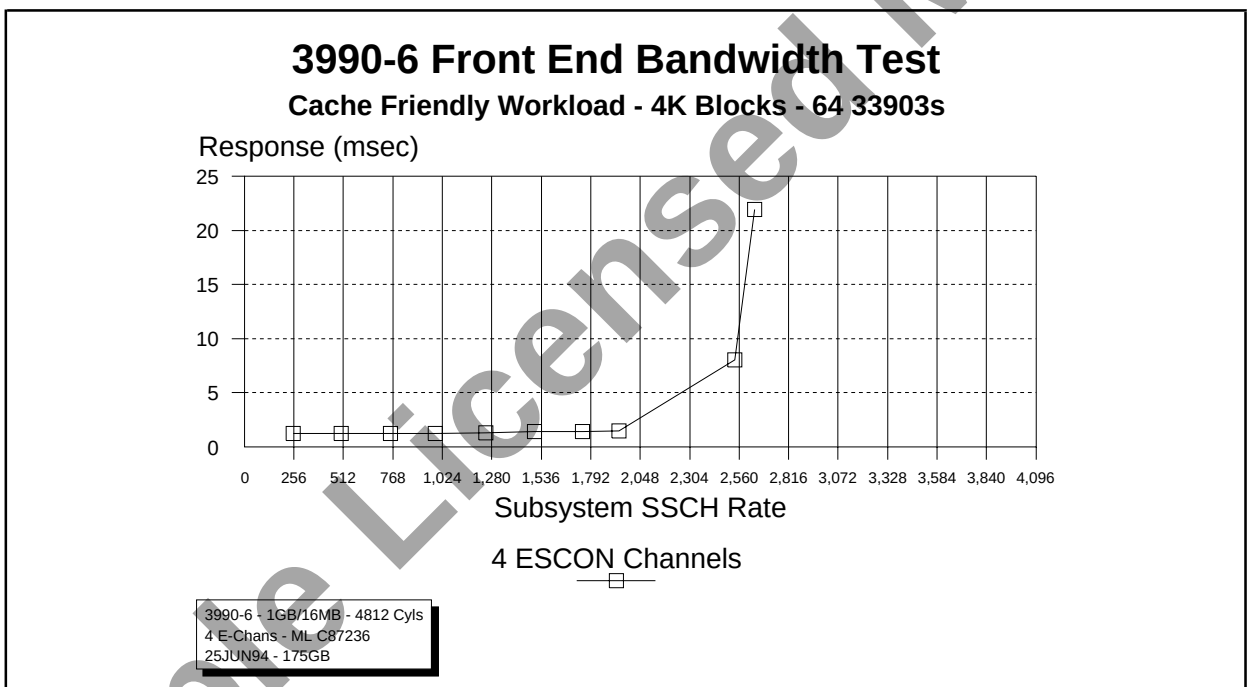

Figure 16. 3990-6 3390-3 Subsystem Front End 4K Bandwidth Test

While the 4K block size test is useful for determining the minimum protocol times for a controller, the half track block size tests provide an understanding of the maximum data transfer rate the control unit front end can support. Figure 17 provides the results for a half track block size. The maximum demonstrated half track front-end hit SSCH rate was 1415 SSCH/Sec as depicted by the line with open squares in the figure.

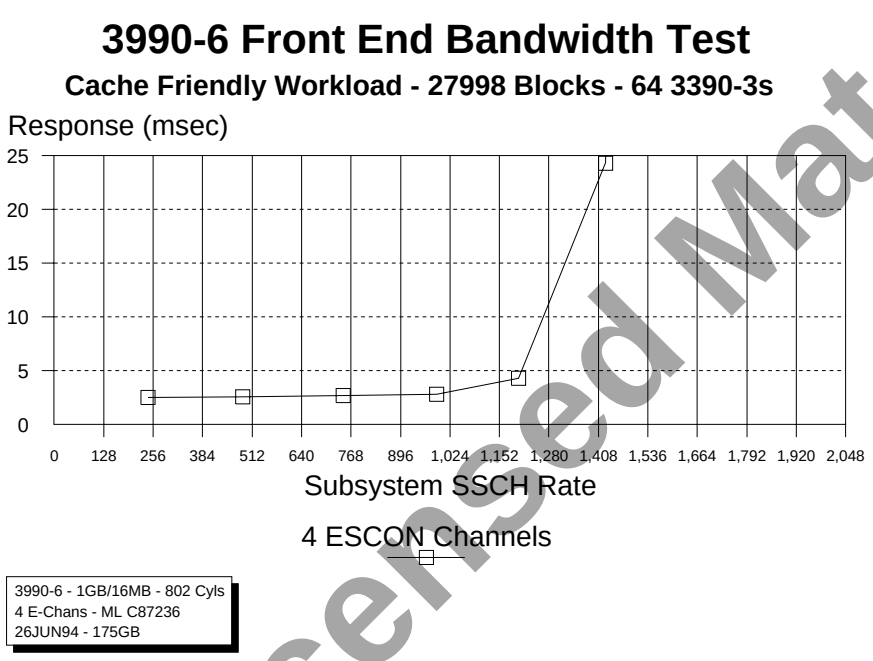

Figure 17. 3990-6 3390-3 Subsystem Front End Half Track Bandwidth Test

One other observation that can be drawn from the data is the benefit ESCON provides to large data transfers. Specifically, the minimum response time for a half track block (i.e., 2.6 msec) is less than twice that of a 4K block, even though almost seven times more data is being transferred.

Figure 18 presents the minimum response time achieved by the subsystem for 4k and half track blocksizes, 1.2 and 2.6 msec respectively.

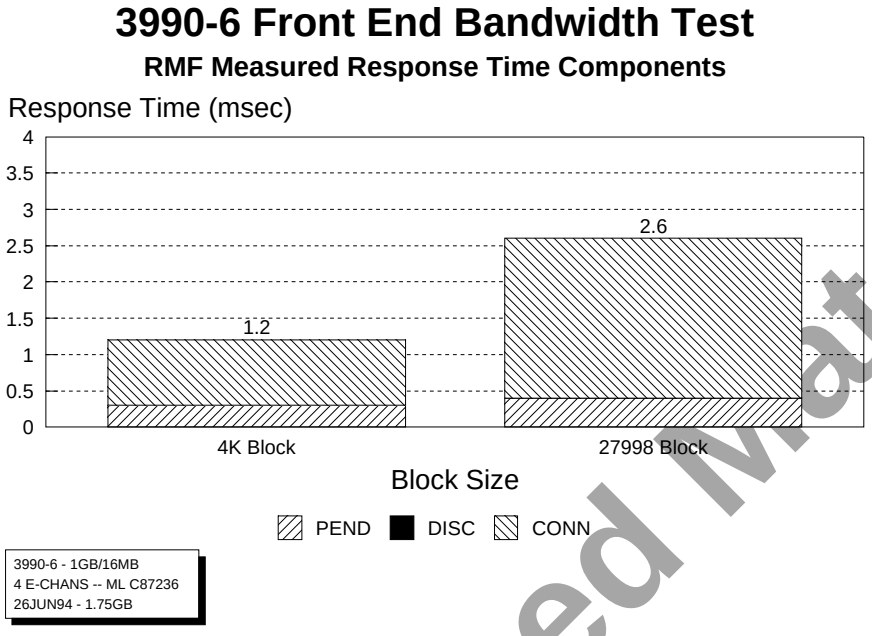

Figure 18. 3990-6 3390-3 Subsystem 4K and Half Track Front End Response Time Components

For the 4K transfer, the 1.4 msec minimum response time is comprised of PEND, DISC, and CONN time components of 0.3, 0.0, and 0.9 msec respectively. The corresponding values for the 2.6 msec half track response time were 0.4, 0.0, and 2.2 msec respectively.

1.2.6 Record Level Cache Tests

Perhaps the most critical feature of a control unit when it is applied to random read/write update OLTP workloads is record level caching. Specifically, record level caching avoids the overhead of staging the remainder of the track on a read-miss and allows all writes to be treated as write-hits even if the underlying track is not in the cache. IBM's initial approach to these problems was called RLC I. RLC I is a DFSMS implementation managed by DCME which addresses both record level staging as well as write hits for regular format data sets.

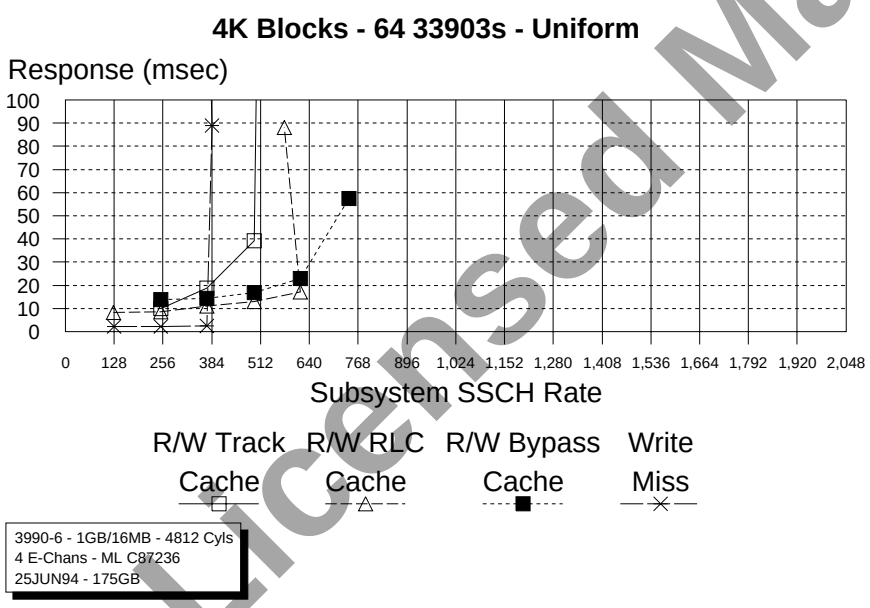

Figure 19. 3990-6 3390-3 Subsystem Record Level Cache Tests

Figure 19 summarizes the record level cache test series results. The results for R/W Track, R/W RLC Cache, R/W Bypass Cache, and 100% Write miss tests are depicted by the lines with open squares, triangles, solid squares, and crosses respectively.

The R/W Track results (line with open squares) demonstrates the performance of the subsystem for read-misses with one hit per stage when the control unit is in the normal *stage to end of track mode.* Specifically, the performance of the subsystem rapidly declines due to the saturation of the 3990-6's four back end data paths by the transfer of data which is never actually used. This bottleneck occurs at approximately 500 SSCH/sec.

The R/W RLC results (line with open triangles) demonstrate the benefits of staging only the desired record for low locality data sets, rather than the remainder of the track. Specifically, the subsystem response time is less than or equal to 15 msec until the control unit's NVS resource becomes saturated at 630 SSCH/Sec, as identified by RMF Cache Reporter statistics. The R/W Bypass Cache line shows the benefits of the traditional bypass cache strategy for poor cache candidates.

The final line in the figure (line with crosses) depicts the 100% Write Miss behavior of the subsystem. As can be seen in the figure, the response times recorded by the 100% write miss case are approximately equal to those record for the front end bandwidth test for 4K blocks until the NVS resource is saturated a 380 SSCH/sec.

1.2.7 Sequential Tests

Figure 20 provides the read and write SSCH rates and response time for the half track sequential tests for the 3990-6 3390-3 subsystem. The read cache, write cache, read bypass, and write bypass results are depicted by the lines with open squares, triangles, solid squares, and crosses respectively. The maximum SSCH rates for cached read and write cases were 620 and 256 SSCH/Sec respectively. When compared to the maximum theoretical limit of 630 reads per second for a 3990-3/6 class control unit,[3] it is clear that the subsystem has reached the 3990's theoretical limit for the back end data staging rate.

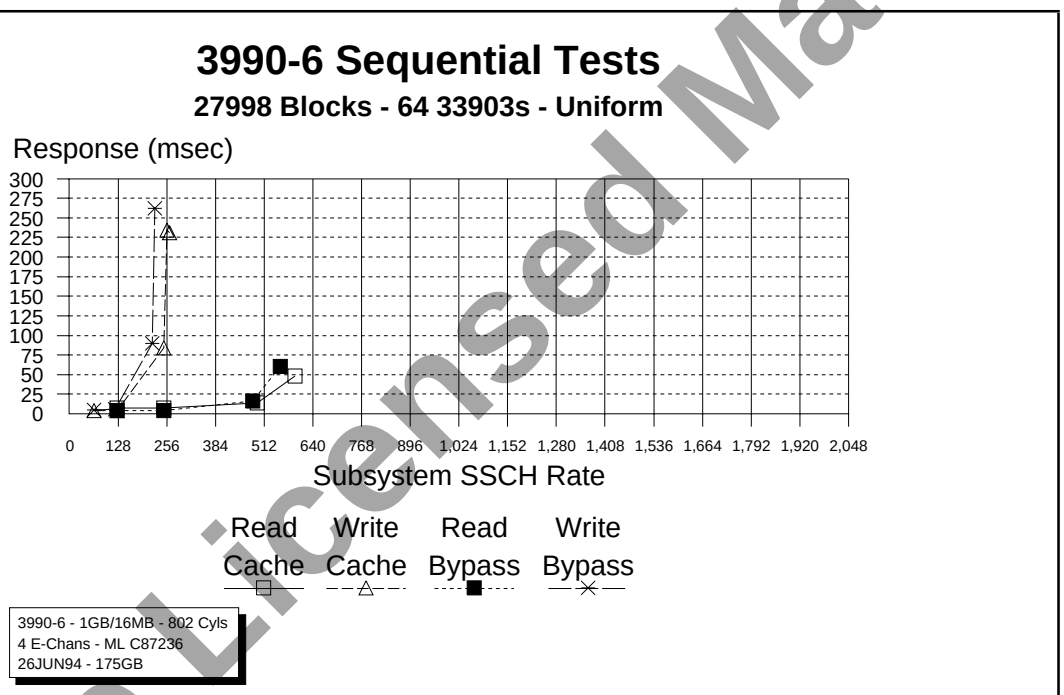

Figure 20. 3990-6 3390-3 Subsystem Sequential Tests

3 A 3990 class control unit includes four 4.2 MB/Sec back end data paths. Assuming 100% utilization for each path, such an architecture can transfer no more than 630 half track blocks a second.

Once again, NVS saturation limited the write performance of the subsystem. This observation is based on the number of fast write delays recorded by the RMF Cache Reporter.

1.3 Observations, and Comments, and Hypotheses

This section is intended to summarize the findings of the ***PAI/O Driver*** studies of the 3990-6 3390-3 subsystem. This section is divided into three primary areas. They are:

Observations: derived data points based on data elements collected in multiple test series. Observations are based on the analysis of measured subsystem behaviors,

Comments: provide a summary of the observations made in Section 1.2, and cross reference the supporting sections where the subsystem behaviors were identified,

Hypotheses: are the author's opinions regarding potential reasons for the observed subsystem behaviors and potential issues which will have to be addressed by architectural extensions in the future.

1.3.1 Observations

The following observations were made based on an overall analysis of the data rather than the test series level discussion which was provided in Section 1.2. Four principal observations were made based on an analysis of all of the evaluated configurations.

1.3.1.1 Ongoing Tuning Requirements

A comparison of the uniform and skewed envelopes for both the 4K and half track studies yields an important observation. Specifically, the difference between the two sets of envelopes indicates that the subsystem would require some ongoing volume level tuning to maintain performance in a DFSMS environment. Specifically, DFSMS's simplistic data set placement algorithms can result in volume level skewing which impacts (to some degree) the performance of a 3990-6 3390-3 subsystem.

1.3.1.2 Aggregate Data Transfer Rates

The aggregate data transfer rate values are derived from the half track front end and sequential test series. The aggregate data rates calculated based on these three tests define the limits of the subsystem's data transfer capabilities. The front-end rate provides insight into the maximum capability of subsystem for high cache-hit ratio workloads. The read and write sequential rates provide insight in to the subsystem's aggregate capability for dump and restore workloads respectively.

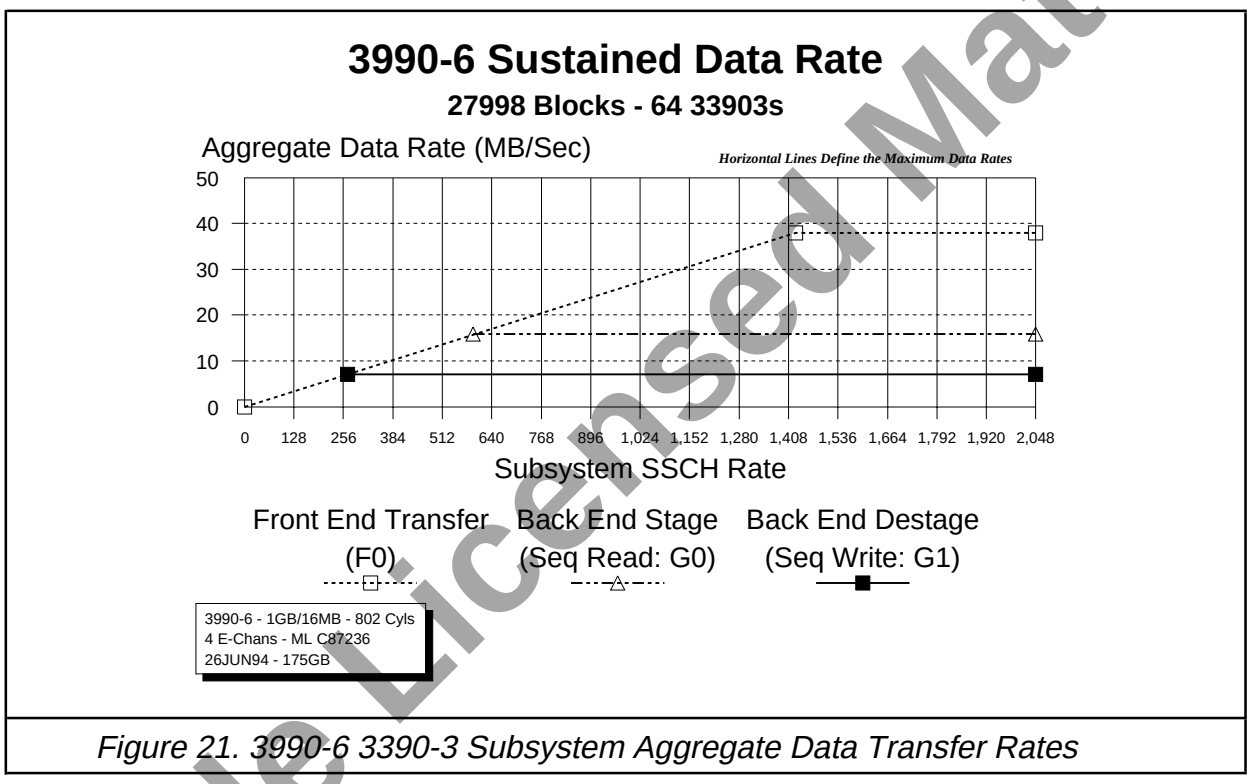

Figure 21. 3990-6 3390-3 Subsystem Aggregate Data Transfer Rates

Please note that higher aggregate read or write sequential data rates do not imply that the dump or restore time for an individual logical volume would be faster or slower than a subsystem with a different aggregate value. Rather, this is a measure of how quickly data can be moved from all of the volumes supported by the subsystem. The performance of

an individual volume is a function of the speed of the underlying devices. A higher aggregate sequential read data transfer rate is also a very important performance factor for data mining applications which can exploit query parallelism.

The front end, back end stage, and back end destage rates are depicted by open squares, open triangles, and solid squares respectively. As can be seen in Figure 21, the effective front end data rate of the subsystem is approximately 38 MB/Sec.

The back end stage rate of the subsystem is 16.7 MB/Sec (as depicted by the line with open triangles). This value is approximately 95% of the architectural maximum of the 3990 subsystem.

Once again, the back end destage rate is limited by the saturation of the subsystem's NVS resource. As can be seen in the figure (denoted by the line with solid squares), the maximum destage rate is approximately 8.2 MB/Sec.

1.3.1.3 Channel Data Transfer Rates

Perhaps one of the most important quantitative measures of an ESCON control unit is its ***effective***[4] channel data transfer rates. Simply stated, higher effective data rates translate into improved response times and decreased channel utilizations. Reducing channel utilization is a particularly critical concern for installations which employ EMIF to share physical ESCON channels between LPARs. The effective channel data transfer rate is calculated by dividing the aggregate data transfer rates presented in Section 1.3.1.2 by the number of channels multiplied by the average channel utilization.

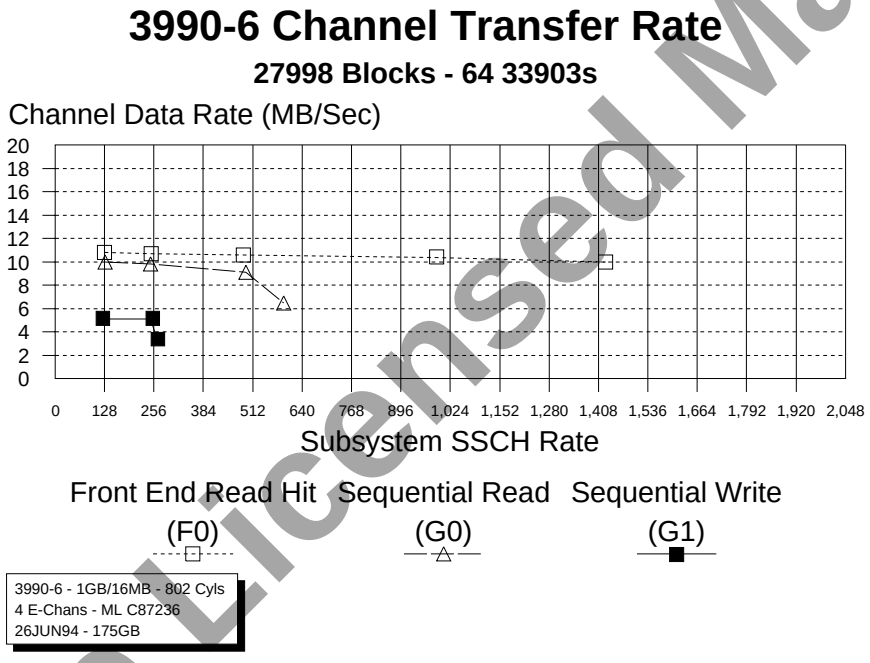

Figure 22. 3990-6 3390-3 Subsystem Channel Data Transfer Rates

4 The term effective is used to denote that only the number of bytes of user data transferred is employed to calculate this rate. Channel programs and ESCON protocol bytes are excluded from this calculation.

Based on the data collected during the front end bandwidth and sequential test series, the overall effective channel data rates for front end read hit, sequential read, and sequential write can be determined. These effective data rates[5] are shown in Figure 22. The front end, sequential read, and sequential write data rates are depicted by lines with open squares, open triangles, and solid squares respectively. Please note that the data rate calculations are based on the half track studies where the actual benefits of ESCON's higher data rates are realized. For 4K blocks, typical ESCON data transfer rates are in the range of 4 to 6 MB/Sec.

The front end data rate for the controller ranges between 11.3 and 10.0 MB/Sec per channel, decreasing with aggregate utilization. That is, as the subsystem transfers more data to the channels, it is reasonable to expect lower transfer rates as a result of higher channel and control unit microprocessor utilizations.

The read data rate ranges from 9.9 to 6.3 MB/Sec per channel, once again decreasing with aggregate arrival rate. The values are lower than those for the front end tests since physical data staging is required. *The significant decline in the data rate at high arrival rates is a result of physical 3390-3 device utilization.*

The final observation is for sequential writes, where data rates range from 5.6 to 3.8 MB/Sec per channel. The significant difference between the back end destage results and those for front end and back end stage cases is the effect of the overheads associated with destage and NVS saturation.

While one might surmise that increasing the size of the 3990-6's NVS would modify this behavior, it would in fact only delay it. Consider an example of a 5 gallon bucket with a drain which allows one-half gallon of water per minute to leave the container. If you add water to the bucket at a rate greater than one-half gallon per minute, then the bucket will overflow. This overflow is equivalent to the fast write delays for 3990 architecture. To continue our example, if you increased the size of the bucket, you would only delay the occurrence of the overflow, not prevent it. *The actual bottlenecks in the 3990 destage process are the speed of the microprocessors which execute the destage microcode and*

---

5 Please note that the actual data rate (including CCWs, ESCON overheads, and status information) would be somewhat higher than the values presented. The data rates shown in the figure are based only on actual user data transferred.

the bandwidth of the 3990's back end. While adding more NVS will allow the control unit to respond to higher momentary bursts of writes, it will not change the steady state behavior of the destage process.

Hence, when the ***term NVS saturation*** is employed in the remainder of Section 1.3, it should be interpreted as the whole destage process rather than the configured size of NVS for the controller. *In any event, these results were measured for a subsystem with the maximum configurable NVS size.*

1.3.1.4 Focusing on Access Density

It is important to realize that the majority of DASD subsystems which compete with the 3990-6 3390-3 subsystem offer substantially more storage capacity. Hence, a cursory review of SSCH rate based performance data for competing subsystems can easily lead you to an invalid conclusion. Simply stated, the maximum SSCH rate of a subsystem should not be your primary decision criteria during acquisition. Rather, the SSCH rates need to be normalized into access density format to provide an apples to apples comparison of the competing storage solutions. *As was emphasized in the acquisition methodology, we recommend that you select a DASD subsystem from all of the candidates whose access density meets or exceeds your requirements.*

When comparing a proposed 3990-6 3390-3 subsystem solution to other vendor offerings, it is likely that you may be considering a configuration of several 3990-6 3390-3 subsystems versus just one or two subsystems from another vendor to meet your storage requirements. Provided that all of the proposed solutions fit within your environmental and other configuration requirements, the final decision will likely based on cost and the ***n:1*** trade-off.

While the meaning of cost is clear, some explanation will be required for the ***n:1*** trade-off question. Specifically, in the event of a single subsystem failure, your potential loss is smaller for a DASD configuration comprised of several subsystems than that for a configuration with just one or two subsystems. Of course, it can be argued that if you lose just one critical subsystem (e.g., the one with the spool or system volume) the net effect is just the same.

On the other hand, it can be argued that fewer subsystems is better since it is difficult to ever achieve a perfect split (from a load balancing standpoint) of your DASD data sets. Hence, over the life of the subsystem substantially more data sets relocations might be required to maintain subsystem balance.

In any event, always focus your performance related decision criteria on access density in the event that you are comparing subsystems of different capacities.

1.3.2 Comments

Based on the engineering test series conducted for the 3990-6 3390-3 subsystem using ***PAI/O Driver***, the following are summary comments about the subsystem. The observations are presented in Tables 2 and 3 with references to the section of the analysis on which each observation is based.

Table 2. 3990-6 3390-3 Subsystem Comments and Observations

Observation	***Section***
The 3990-6 3390-3 subsystem offers 128 logical paths. This is more than adequate for the most complex EMIF and parallel sysplex configurations.	1.2
The 3990-6 3390-3 subsystem demonstrates a proverse write fraction sensitivity for small blocks.	1.2.1, Figure 2 1.2.2, Figure 6
The 3990-6 3390-3 subsystem demonstrates an adverse write fraction sensitivity (service constrained) for half track blocks.	1.2.1, Figure 3 1.2.2, Figure 7
The 3990-6 3390-3 subsystem would tend to require some ongoing device level tuning if the subsystem were employed for DFSMS managed data sets.	1.2.1 1.2.2 1.3.1.1
The 3990-6 3390-3 subsystem architecture is completely insensitive to the influence of a high activity device. That is, a subsystem can support a high activity file like a MIM data set, JES checkpoint data, or data base indices set without compromising the overall performance of the subsystem.	1.2.3
Depending on locality of reference, the 3990-6 3390-3 subsystem can provide a response time less than or equal to 10 msec at 360 to 1500 SSCHs/Sec for read/update OLTP like data sets with 4K blocksizes.	1.2.4, Figure 14
The minimum search, ESCON protocol, and transfer time for a 4K block is 1.2 msec.	1.2.5, Figure 18
The back end interfaces are effectively saturated for all but ***"cache friendly"*** reference patterns and are the fundamentally limiting resource of the subsystem.	1.3.1.2
The maximum effective ESCON data transfer rate measured for the subsystem was 11.3 MB/Sec.	1.3.1.3

Table 3. 3990-6 3390-3 Subsystem Comments and Observations

Observation	***Section***
RLC I provides an effective solution for read or read/write data with poor locality.	1.2.6
The 100% write miss support, i.e., quick write , works as advertised. However, the maximum supportable write miss rate is limited by NVS saturation.	1.2.6 1.3.1.3
The NVS scheme employed by the 3990 architecture imposes a strict limit on the subsystem's sustained write performance. This limitation imposes significant performance limitations on OLTP read/write update as well as sequential write access patterns. It also limits the maximum aggregate write data transfer rate as well as the effective channel rate for write operations.	1.2.4 1.2.6 1.3.1.2 1.3.1.3

1.3.3 Hypotheses

Since the 3990-6 3390-3 subsystem is a mature product which is well understood, no hypotheses are required.

1.4 Acquisition Strategies

This performance profile has provided a review of the 3990-6 3390-3 subsystem. As we have previously observed, the 3990 architecture has been refined for the past 16 years. Moreover, the performance characteristics of physical 3390-3 devices are well understood. While the 3990-6 3390-3 subsystem is no longer a product offering from IBM, they may easily be obtained in the secondary market.

Hence, we recommend that 3990-6 3390-3 subsystems be acquired as ***gap fillers*** for installation's that do not have environmental constraints. The critical factors which must be considered in an acquisition are:

- *the residual value will likely be zero at the end of the transaction,*
- depending on the age of the 3390-3s, you may experience a higher than normal HDA failure rate.
- long term, maintenance and environmental costs are likely to outweigh any initial savings that the subsystem provides.
- ***you do not intend to employ them for write intensive workloads with strict performance requirements.***

Based on this subsystem performance profile, ***Performance Associates would advise a client to only consider a used 3990-6 3390-3 subsystem as an interim (i.e., 12 to 18 months) solution to its storage subsystem requirements.*** While the performance levels are comparable with IBM's new offerings, environmental and maintenance expenses preclude the subsystem as a long term solution.