DATE: October 8, 1981

TO: R & D Personnel

FROM: Suresh Jasrasaria, Rick Bahr and Humberto Rodriguez

SUBJECT: PERFORMANCE EVALUATION OF A BACK-END STORAGE NETWORK

REFERENCE: (See Bibliography)

KEYWORDS: System Performance, Simulation, Queuing Network Model

ABSTRACT

This document describes the modeling efforts done to measure the effects of a Back-end Storage Network (BSN) on system performance. A current Prime product, RINGNET, is selected as a BSN to replace the existing I/O bus in the PRIME series 50 computers. The resulting system is modeled using both analytic and simulation techniques. Input data for the models are obtained from the measurements of the existing system. The two main conclusions of the performance study are:

1. The effects on system performance due to the introduction of a BSN are minimal. Compared to the functionality gain obtained, they are almost negligible (less than 10 percent) for all practical cases of interest.

2. The predictions of the analytic model based on the Mean Value Analysis (MVA) technique are very close (errors less than 10 percent) to those of the simulation model.

1 INTRODUCTION

Due to the developments in communications technology, processor architecture, and peripheral device design, distributed processing has emerged as a particularly interesting concept in recent years. Business forecasters project that distributed processing will play an important role in the 1980s. Electronic funds transfer in Banking, robots in production line, and electronic mail in the office environment will revolutionize the automation of service and manufacturing industries of the future.

A Local Area Network (LAN) architecture provides us with a technology to implement distributed processing in a geographically limited environment. (The distance spanned by the network should be within a few kilometers. For a detailed study of LANs the reader is referred to [JAS80].) The key factor for a successful implementation of distributed processing is low-overhead resource sharing - a function provided by Back-end Storage Networks (BSNs). A BSN, Figure 1, is a logical subnetwork, not necessarily a physically separate hardware subnetwork, within a general purpose high performance LAN to provide shared storage services to a set of host computers [WAT80]. Here high performance denotes throughputs of the order of 100s of Mbits/sec, and packet delays of the order of 10ths of a second. Because of the low-overhead resource sharing requirement, high network performance is particularly important for BSN-based service of a LAN-based distributed processing system. The traditional networking advantages, when viewed in the context of a BSN, develop new interpretations:

1. Resource Sharing. Low-overhead sharing of peripheral devices among a user community is an obvious consequence of the adoption of a BSN. There is no concept of "owner" or "host" in the interconnection. While software strategies to exploit this are far from trivial, this is an opportunity for more effective use of what frequently dominates the cost of a computer facility. Load balancing is also more readily possible. Even such crude measures as a terminal user electing to run on a less busy host can show significant return. Moreover, special purpose processors, such as floating point or data base machines, can be easily shared by a number of hosts in the network, thus reducing the effective cost of providing such facilities.

2. Physical Distribution. From a hardware point of view as a controller attachment bus, a BSN offers considerable packaging flexibility to the controllers which may now reside with the devices they attach. In addition the BSN can supplant any ad hoc linkage scheme which interconnects a host and its I/O processors.

3. Reliability. BSNs provide multiple logical paths to a device in case of a host failure. They also provide a more uniform pool of storage resources in the event of a device failure.

Multiple BSNs, and hence multiple physical paths to a device, can even provide a higher degree of flexibility in this regard. The increased symmetry between host and controller responsibilities lends itself to simpler fault detection and recovery algorithms.

4. Modularity. Growth from small to large configurations is possible without such unexpected hurdles as I/O bus expanders/cages being encountered. The presence of a BSN holds the promise of "on-line" configuration change. A site may choose to incrementally expand processing ability (add a CPU) without massive I/O responsibility restructuring.

There are also some beneficial system implications in the adoption of a BSN which have no direct parallel in traditional networking applications. A BSN's emphasis on average, rather than peak bandwidth exchanges buffering cost for bus interface cost at both host and controller. A more uniform syntax for device interactions offers economy of mechanism in software. It also offers a higher degree of hardware isolation. With its high data to address ratio and implicit retry characteristics, the same syntax also offers savings in hardware which finds relatively easy to meet responsetime constraints and reduced multiplexing demands over typical current strategies.

2 DESCRIPTION OF THE SYSTEM

A current Prime product, RINGNET [PRI79], is selected to perform the functions of a BSN. For the purposes of this analysis, it is assumed that RINGNET will replace the current I/O bus from the disk controller to the host in the PRIME series 50 computers [PRI80]. RINGNET is a token based ring structured LAN [GOR79]. The data transfer rate on the ring is 1 Mbyte/sec. Packets can be of variable length. The bus access is arbitrated by a circulating token. Once a node is in possession of the token, it may choose to transmit. If it does so, the data circulates around the ring, encountering but a few bits of delay at intervening nodes, only to be removd by the originator. The receiver copies the message and provides an immediate low level acknowledgement in a reserved field which trails the packet. Once the packet has returned, or if a node does not wish to transmit, the token is relayed to the adjacent node.

The inclusion of a BSN additionally requires the imposition of store and forward buffers in the disk controller which can no longer stream data directly between the disk and the main store. It should be mentioned that only one such store and forward operation will be necessary, as the processor will be able to transfer to/from its main store directly.

The software disk access strategy is assumed to remain the same by this imposition. That is, reads and writes for record quantities (2048 bytes) will still be the only disk access modes with the data occupying a single packet. Ownership is assumed to

reside with one processor and functional responsibilities between controller and host are assumed unaltered.

3 PERFORMANCE EVALUATION OF THE SYSTEM

Analytic and simulation modeling techniques were used to evaluate the performance of the above system. Input data for the models were derived from the measurements of an existing system. The measurements were obtained with the help of the General Metering Tool (GMT) [ROD81], an event driven software monitor designed for internal use at PRIME to aid in performance analysis. The existing system consists of a CPU with a scaled instruction rate of 1.0, an 80 Mbyte disk, and a disk controller with a transfer rate of 1.25 Mbytes/sec. An I/O intensive workload with a multi-programming level (MPL) of six users was generated using a remote terminal emulator (RTE). Figure 2 represents the distributions of I/O service times. (Here the I/O service time includes all the three components, namely - seek, latency and transfer.) Distributions of the measurement data for CPU service times can be seen in Figure 3.

3.1 THE SIMULATION MODEL

The simulation views the running software as a collection of processes. The processes are characterized by "mean times" between file I/O and pagefault requests. All I/O activity is directed at disks and is comprised of single record (2048 byte) transfer at a time. The read-to-write request ratio of the benchmark has also been preserved in the model and some small amount of system overhead has been added for pre- and post-processing of an I/O request as well as process exchange.

The I/O subsystem consists of a few controllers (max=8), with a fixed number of disks per controller (max=4). File I/O requests are scattered uniformly among all disks while pagefault requests are directed to only one of them. An unlimited controller buffer availability is assumed. Disk rotational latencies are assumed to be uniformly distributed between 0 and 16 ms, the time to complete one revolution for a 3600 rpm disk. To mimic the observed seek time distribution (Figure 4) as closely as possible it is assumed in the model that half the I/O requests do not encounter any seek delay while the remainder of them face a uniform seek time distribution between 6 and 55 ms. (6 ms and 55 ms being the times to seek 1 and all 800 cylinders in the disk respectively.) The existing disk transfer bandwidth of 1.25 Mbyte/sec is used in the model.

The two I/O subsystem alternatives considered are the current I/O bus and RINGNET-based BSN. The simulation of the model advances time at a uniform rate, while gathering necessary statistics such as device (bus, CPU and disk) utilizations, mean transaction/response time, and mean device queue lengths.

3.2 THE ANALYTIC MODEL

A very simplistic approach is taken develop an analytic model of the system. A closed Queueing Network Model (QNM) of the system (Figure 5) consisting of service centers for CPU, disks, disk controllers and BSN is constructed and solved using Mean Value Analysis (MVA) technique [REI80]. (For a detailed study of QNMs and their solution techniques the reader is referred to [SAU80]. A special issue of ACM Computing Surveys on QNMs of Computer System Performance edited by G.S.Graham [GRA78] also gives a good overview of the area.) The mean values obtained from the measurement data are directly used for input to the model.

The I/O subsystem model is based on the fact that each disk controller can control up to 4 disks, i.e., 4 disks per controller can perform independent seeks at the same time. But, once the seek is complete only one such disk can transfer data at one time. Thus in terms of the model parameters, the disk service time is only the seek component of the I/O service time while the disk controller service time constitutes the remaining two parts of the I/O service time, namely rotational latency and data transfer time. The abstraction of these parameters from the mean I/O service time can be done by using the fact that:

1. the average packet length is 2 Kbytes,
2. the average latency is half cycle (3600 rpm disk), and
3. the disk controller transfer rate is 1.25 Mbytes/sec.

4 RESULTS

Both the simulation and the analytic models of the existing system have been validated against the results obtained from actual measurements. The same models with appropriate modifications have also been used to predict the performance of the extended current system with a BSN. The results obtained from both the models are surprisingly close to each other. (The difference in the predicted performance from both the models is within 10 percent.) For the sake of clarity and uniformity all the results presented here are taken from the solutions of the analytic model.

The formost measure of a BSN's performance is the degree to which it adequately provides a desired functionality or support for an application. If a BSN successfully supports a desired resource sharing scheme, for example, or allows a planned distribution of a computational workload over a collection of processors, then we may say the BSN is performing adequately [CHL80]. The main objective of this paper is to quantify the effect of the BSN in providing a resource sharing capability. This has been done by obtaining the system throughput and response time, and device utilizations and mean queue lengths for various load conditions, CPU speeds and I/O subsystem configurations.

The two I/O subsystem configurations considered in this study are:

(i) 1 disk and 1 disk controller.
(ii) 16 disks and 4 disk controllers.

The performance of these configurations, both with and without the BSN, is obtained under 3 different CPU speeds - 1, 3 and 5. (These are scaled CPU speeds where 1 represents the existing CPU speed.) In all these cases the load (MPL) is varied from 1 to 16, enough to capture the effects of saturation. The reason for selecting the above configurations and CPU speeds is that by studying their various combinations we can note the effects on system throughput and response time due to shifts in bottleneck from CPU to I/O subsystem and vice versa.

Figures 6 through 10 represent the results for a 1-disk and 1-controller I/O subsystem with a BSN. It is clear from these graphs that as we increase the CPU speed the bottleneck of the system shifts from CPU to the disks, thus making very little improvement (less than 10 percent) in system performance. The effect of introducing a BSN in this system can be concluded from the fact that the absolute BSN utilization (Figure 10) is less than 10 percent. To represent the effect of a BSN on the system more clearly Figures 11 and 12 represent the percentage difference in throughput and response time of the system with and without the BSN.

Figures 13 through 17 represent the results for a 16-disk and 4-controller I/O subsystem with a BSN. From these results it can be seen that the system is processor bound even with a CPU speed of 5 units. Moreover, in this configuration as we increase the speed of CPU from 1 to 5, there is a marked improvement in the system performance (Figures 13 and 14). The reason for such improvement is that there is a large number of disks in the I/O subsystem to share the load, as a result even if we increase the CPU speed by a factor of 5, the system still remains processor bound. (Disk utilization, Figure 16, always remains under 50 percent.) The effect on system performance due to a BSN can be seen in Figures 18 1nd 19. From these results it can be very easily concluded that compared to the functionality gain obtained the degradation in performance due to the BSN is within tolerance limits (less than 10 percent). For the sake of completeness Figures 20 and 21 represent the throughput-delay curves for the two I/O subsystem configurations considered in this study.

It is also interesting to note (Figure 22) that the 16-disk and 4-controller I/O subsystem configuration starts to feel the congestion, i.e., becomes I/O bound, only at a very high CPU speed (>10 units).

5 CONCLUSION

In this study we have shown two important results. Firstly, we have been able to show that an analytic model is as good a

tool as a simulation model for a high level performance prediction of a BSN-based system. Secondly, it is clear from the quantitive results presented that a BSN is capable of adequately providing a shared storage facility.

BIBLIOGRAPHY

[CHL80] I. Chlamtac, W.R. Franta, P.L. Patton, and B. Wells, "Performance Issues in Back-end Storage Networks", Computer February 1980, pp 18-31.

[GOR79] R.L. Gordon, W.W. Farr, and P. Levine, "Ringnet: A Packet Switched Local Network with Decentralized Control",Proc. of the 4th Conference on Local Computer Networks, Minneapolis, Minnesota, October 22-23 1979, pp 13-19.

[GRA73] G.S. Graham, Guest Editor, ACM Computing Surveys Vol. 10, No. 3, September 1978, a Special Issue on Queueing Network Models of Computer System Performance.

[JAS80] S.K. Jasrasaria, "Local Area Network Bus Access Protocols and Their Performance", M.Sc. Thesis, Department of Computer Science, University of Toronto, November 1980.

[PRI79] "The Primenet Guide IDR 3710", Prime Computer Inc., Technical Publications Department, 500 Old Connecticut Path, Framingham MA 01701 USA.

[PRI80] "System Architecture Reference Guide PDR 3060", Prime Computer Inc., Technical Publications Department, 500 Old Connecticut Path, Framingham MA 01701 USA.

[ROD81] H. Rodriguez Jr., "Design Specifications of General Metering Tool", Technical Report PE-TI-910, Prime Computer Inc., Technical Publications Department, 500 Old Connecticut Path, Framingham MA 01701 USA.

[SAU80] C.H. Sauer, and K.M. Chandy,"Computer System Performance Modeling: A Primer", Prentice Hall, 1980.

[WAT80] R.W. Watson, "Network Architecture Design for Back-end Storage Networks", Computer, February 1980, pp 32-48.

FIGURES

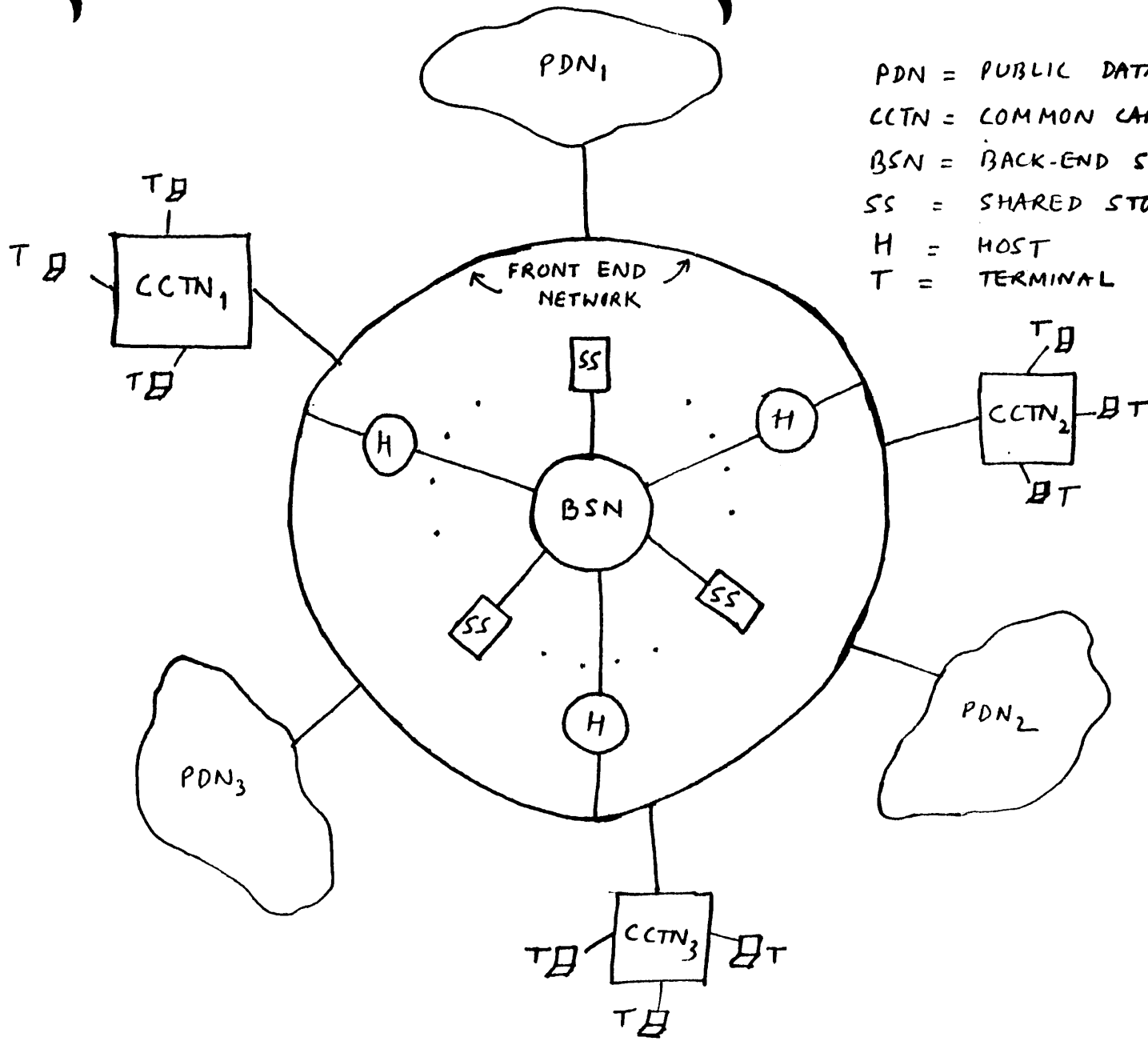

Figure 1 A BACK-END STORAGE NETWORK

:118 system running IO workload with 2048Kb memory.

Statistics for Actual I/O time - not counting queue time (msecs)

MEAN	STD DEV	MINIMUM	MAXIMUM	N
24.412	18.726	0.000	69.697	47874

Histogram for Actual I/O time - not counting queue time (msecs)

OBSRV FREQ	REL PCT	CUM PCT	UNITS
12125	25.3	25.3	0.0
3170	6.6	31.9	4.0
1434	3.0	34.9	8.0
2629	5.5	40.4	12.0
3526	7.4	47.8	16.0
2101	4.4	52.2	20.0
3233	6.8	58.9	24.0
1689	3.5	62.5	28.0
1884	3.9	66.4	32.0
4227	8.8	75.2	36.0
2114	4.4	79.7	40.0
2063	4.3	84.0	44.0
4021	8.4	92.4	48.0
1379	2.9	95.2	52.0
1024	2.1	97.4	56.0
1139	2.4	99.8	60.0
109	0.2	100.0	64.0
7	0.0	100.0	68.0
0	0.0	100.0	72.0
0	0.0	100.0	76.0
0	0.0	100.0	80.0
0	0.0	100.0	84.0
0	0.0	100.0	88.0
0	0.0	100.0	92.0
0	0.0	100.0	96.0
0	0.0	100.0	100.0
47874			

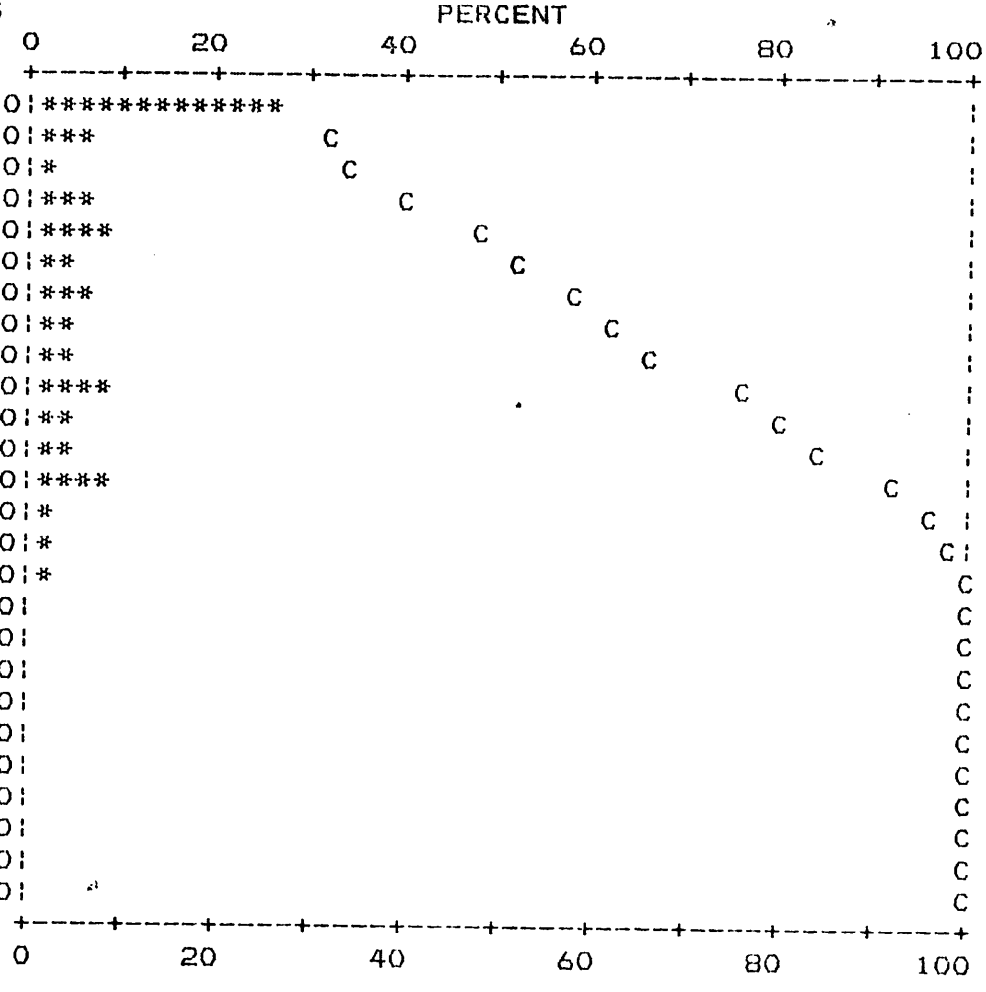

Figure 2: DISTRIBUTION OF I/O SERVICE TIMES

V 10 - **I/O** workload - 2Mb P550

Statistics for Virtual time (msecs) between I/O's

MEAN	STD DEV	MINIMUM	MAXIMUM	N
24.532	113.615	0.000	5966.848	47869

Histogram for Virtual time (msecs) between I/O's

OBSRV FREQ	REL PCT	CUM PCT	UNITS
14575	30.4	30.4	0.0
19412	40.6	71.0	2.0
3336	7.0	78.0	4.0
1210	2.5	80.5	6.0
600	1.3	81.8	8.0
190	0.4	82.1	10.0
177	0.4	82.5	12.0
471	1.0	83.5	14.0
560	1.2	84.7	16.0
693	1.4	86.1	18.0
347	0.7	86.9	20.0
204	0.4	87.3	22.0
146	0.3	87.6	24.0
98	0.2	87.8	26.0
41	0.1	87.9	28.0
17	0.0	87.9	30.0
20	0.0	88.0	32.0
70	0.0	95.4	190.0
42	0.1	95.5	192.0
33	0.1	95.6	194.0
30	0.1	95.7	196.0
44	0.1	95.8	198.0
2034	4.2	100.0	200.0 >
47869			

PERCENT

0 20 40 60 80 100

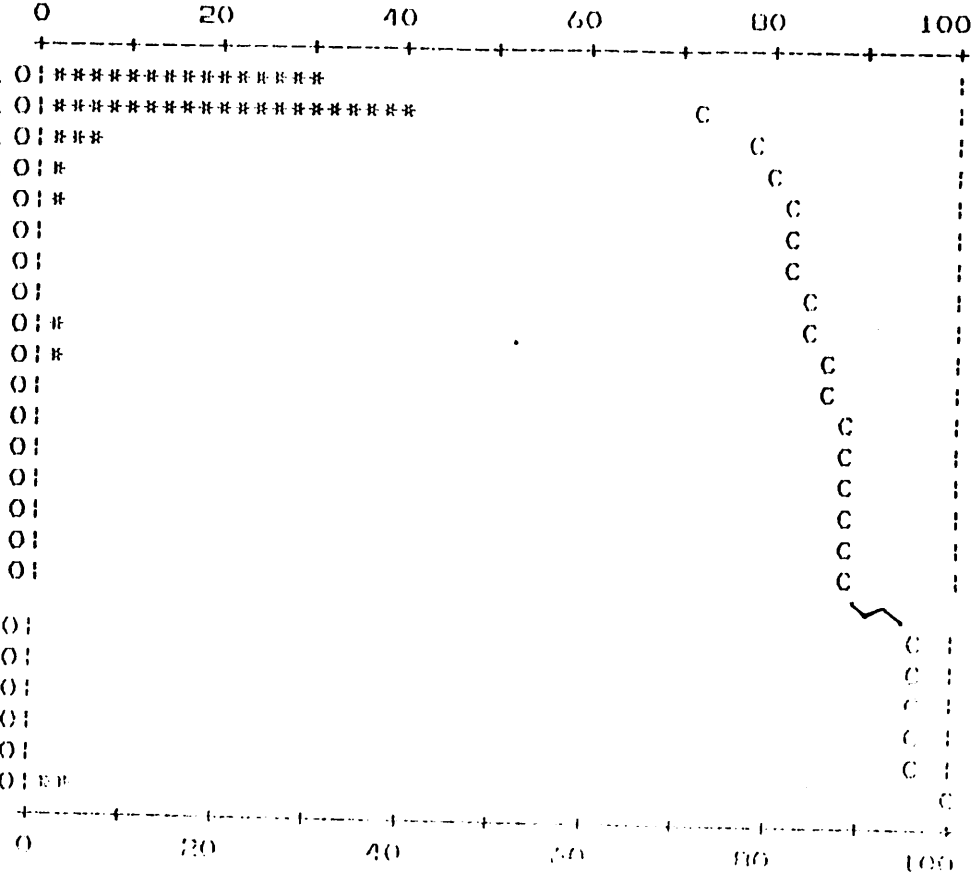

Figure 3: DISTRIBUTION OF CPU SERVICE TIMES

Statistics for Seek distances (# cylinders traversed)

MEAN	STD DEV	MINIMUM	MAXIMUM	N
156.424	212.785	0.000	737.000	47876

Histogram for Seek distances (# cylinders traversed)

BSRV FREQ	REL PCT	CUM PCT	UNITS	PERCENT 0 20 40 60 80 100
19131	40.0	40.0	0.0	**********
2656	5.5	45.5	1.0	*** C
176	0.4	45.9	2.0	C
109	0.2	46.1	3.0	C
107	0.2	46.3	4.0	C
95	0.2	46.5	5.0	C
137	0.3	46.8	6.0	C
91	0.2	47.0	7.0	C
87	0.2	47.2	8.0	C
85	0.2	47.4	9.0	C
98	0.2	47.6	10.0	C
102	0.2	47.8	11.0	C
111	0.2	48.0	12.0	C
144	0.3	48.3	13.0	C
101	0.2	48.5	14.0	C
58	0.1	48.6	15.0	C
64	0.1	48.8	16.0	C
92	0.2	49.0	17.0	C
113	0.2	49.2	18.0	C
21	0.0	69.1	195.0	C
12	0.0	69.2	196.0	C
8	0.0	69.2	197.0	C
19	0.0	69.3	198.0	C
49	0.1	69.4	199.0	C
19145	31.6	100.0	> 200.0	********** C
47876				0 20 40 60 80 100

Figure 4: STATISTICS FOR SEEK DISTANCES

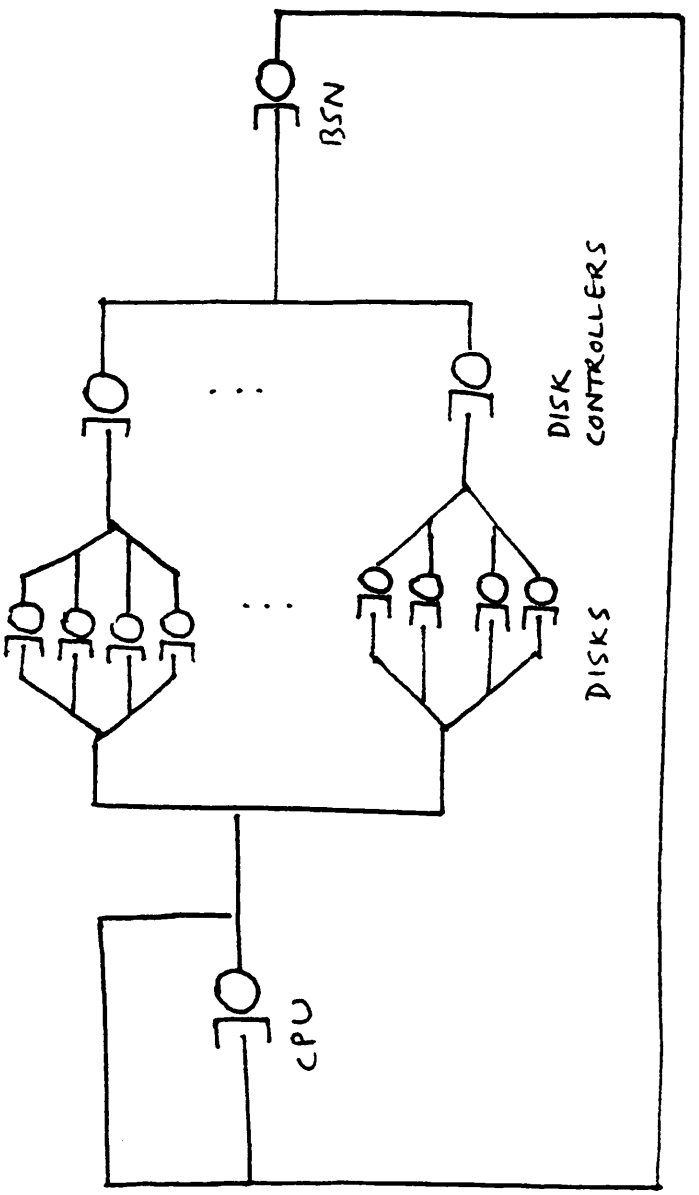

Figure 5 : THE ANALYTIC MODEL.

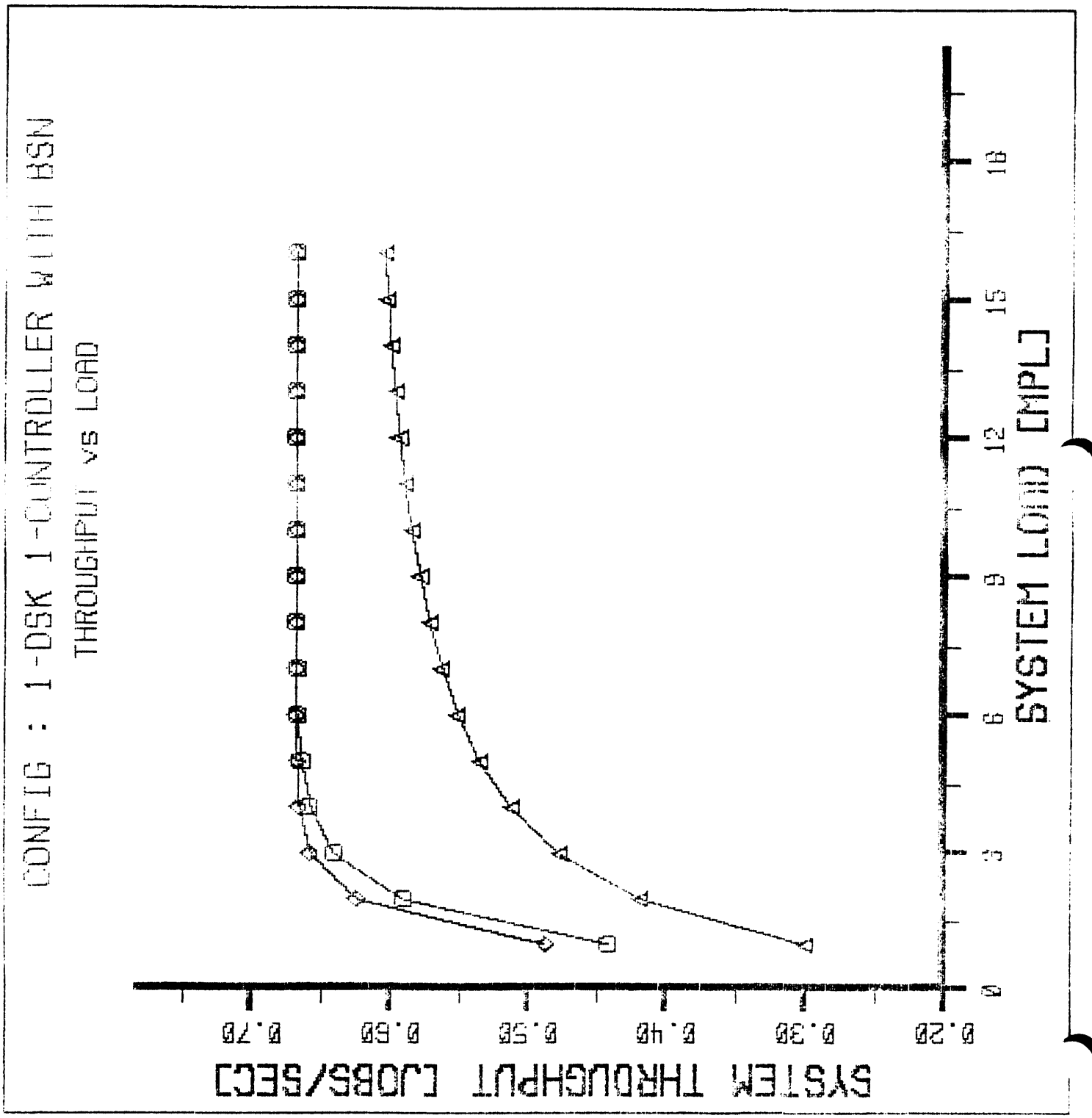
CONFIG : 1-DSK 1-CONTROLLER WITH BSN
THROUGHPUT vs LOAD
CURVE ID:
Δ CPU1
□ CPU3
◇ CPU5
SYSTEM THROUGHPUT [JOBS/SEC]
0.20
0.30
0.40
0.50
0.60
0.70
0
3
6
9
12
15
18
SYSTEM LOAD [MPL]

Figure 6

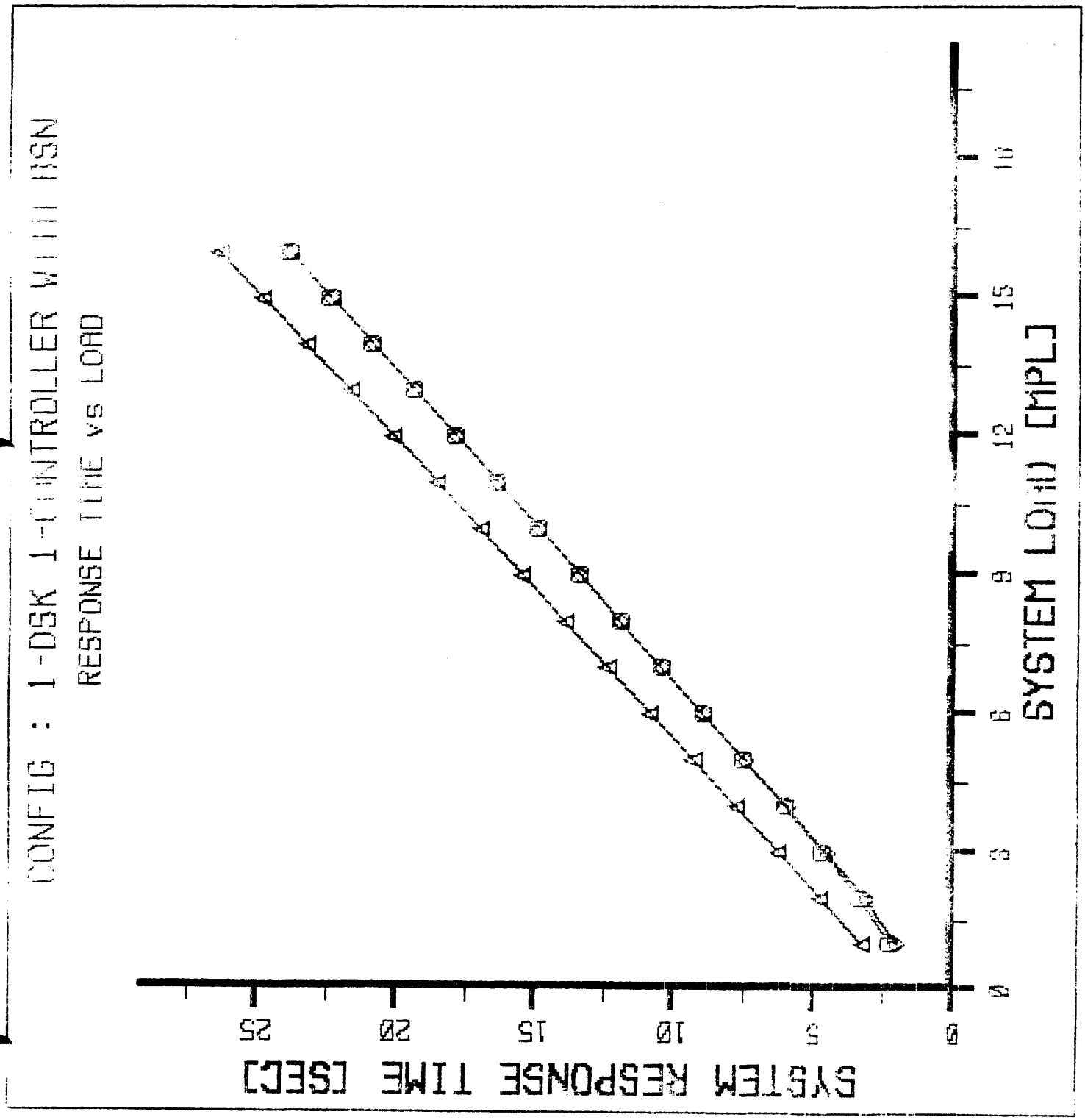
CONFIG : 1-DSK 1-CONTROLLER WITH BSN
RESPONSE TIME vs LOAD
CURVE ID:
△ CPU1
□ CPU3
◇ CPU5
SYSTEM RESPONSE TIME [SEC]
0
5
10
15
20
25
SYSTEM LOAD [MPL]
0
3
6
9
12
15
16

Figure 7

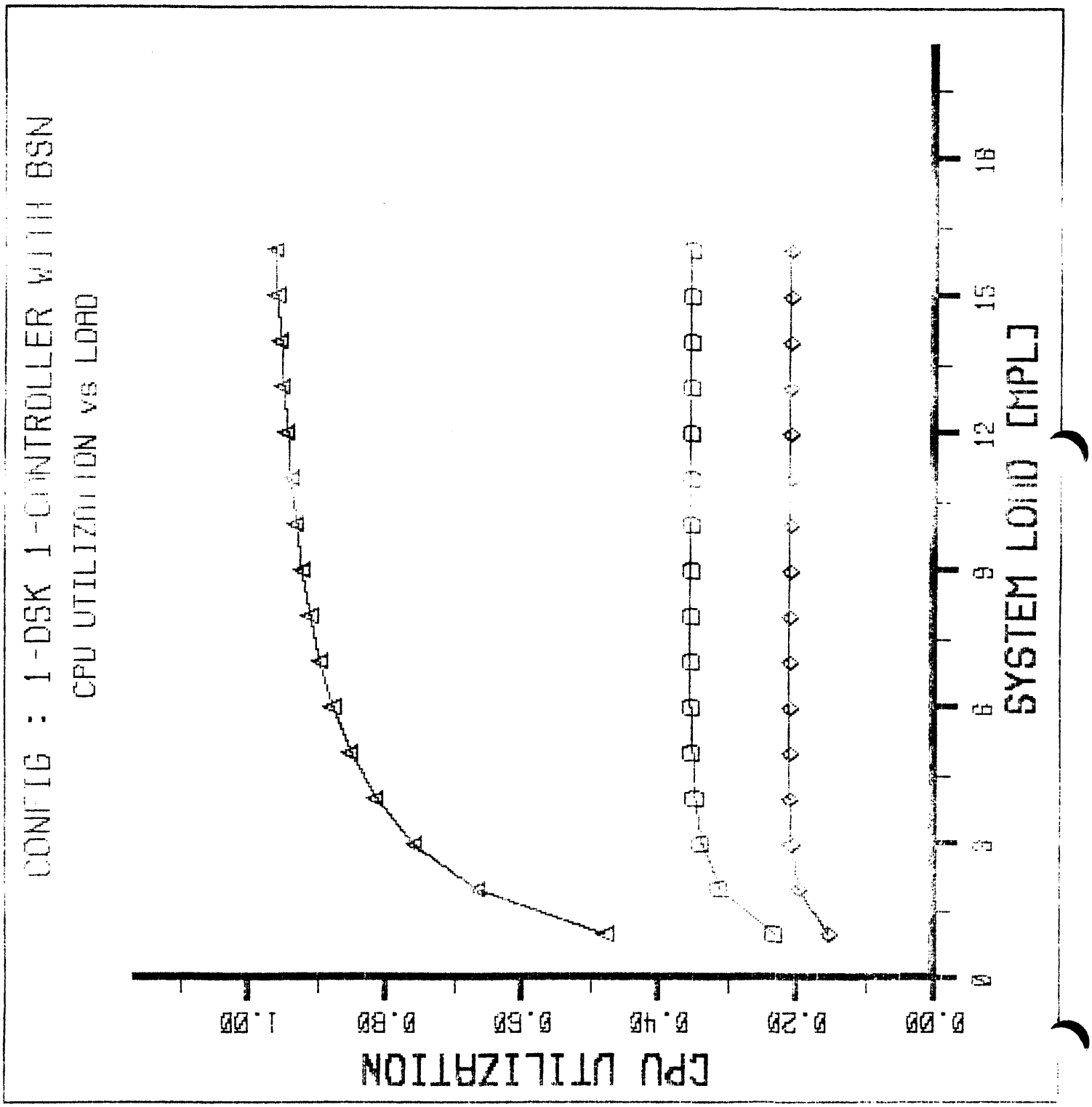
CONFIG : 1-DSK 1-CONTROLLER WITH BSN
CPU UTILIZATION vs LOAD
CURVE ID:
Δ CPU1
□ CPU3
◇ CPU5
CPU UTILIZATION
0.00
0.25
0.40
0.60
0.80
1.00
0
3
6
9
12
15
18
SYSTEM LOAD [MPL]

Figure 8

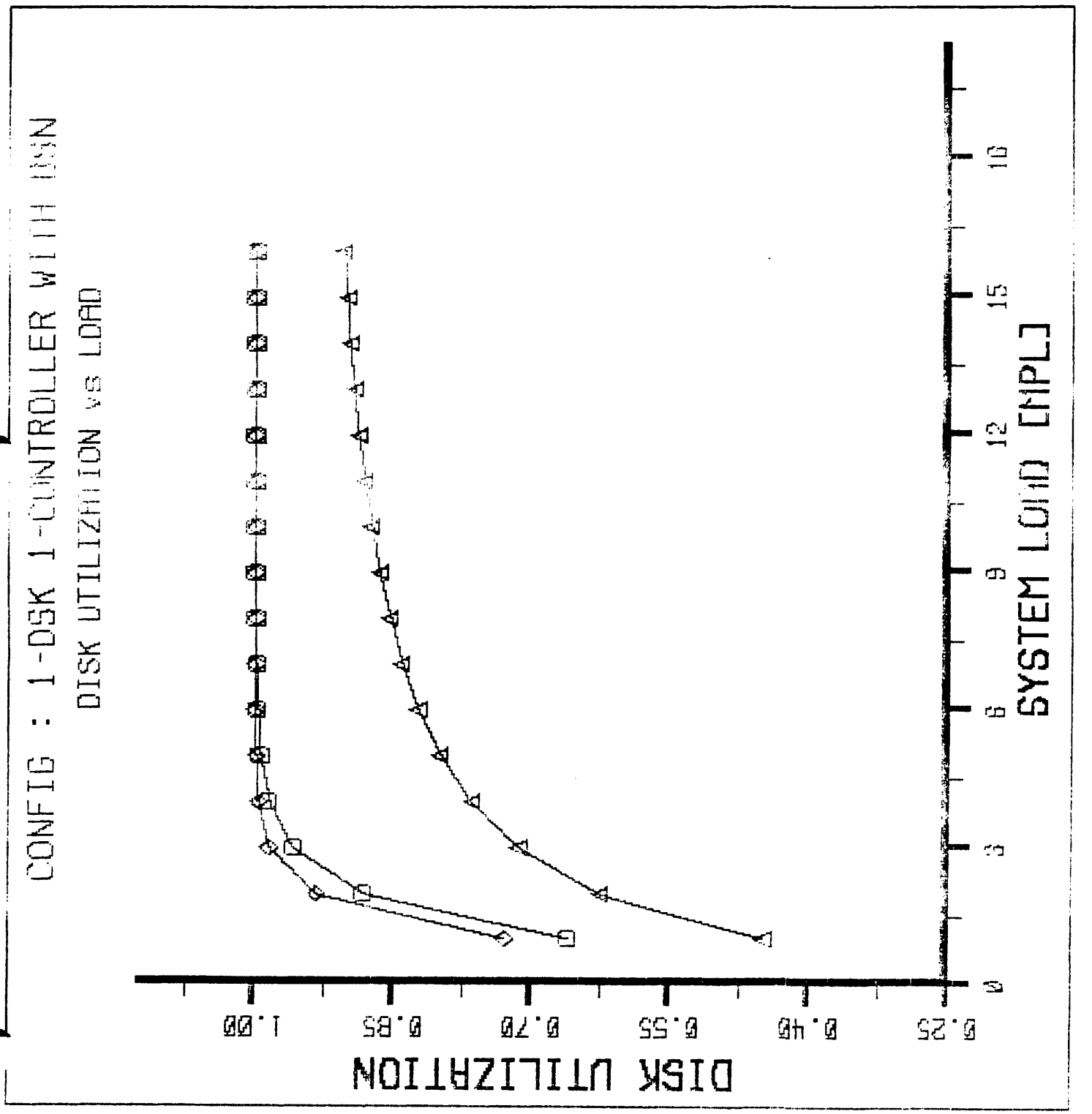
CONFIG : 1-DSK 1-CONTROLLER WITH BSN
DISK UTILIZATION vs LOAD
CURVE ID:
Δ CPU1
□ CPU3
◇ CPU5
DISK UTILIZATION
0.25
0.40
0.55
0.70
0.85
1.00
0
3
6
9
12
15
18
SYSTEM LOAD [NPL]

Figure 9

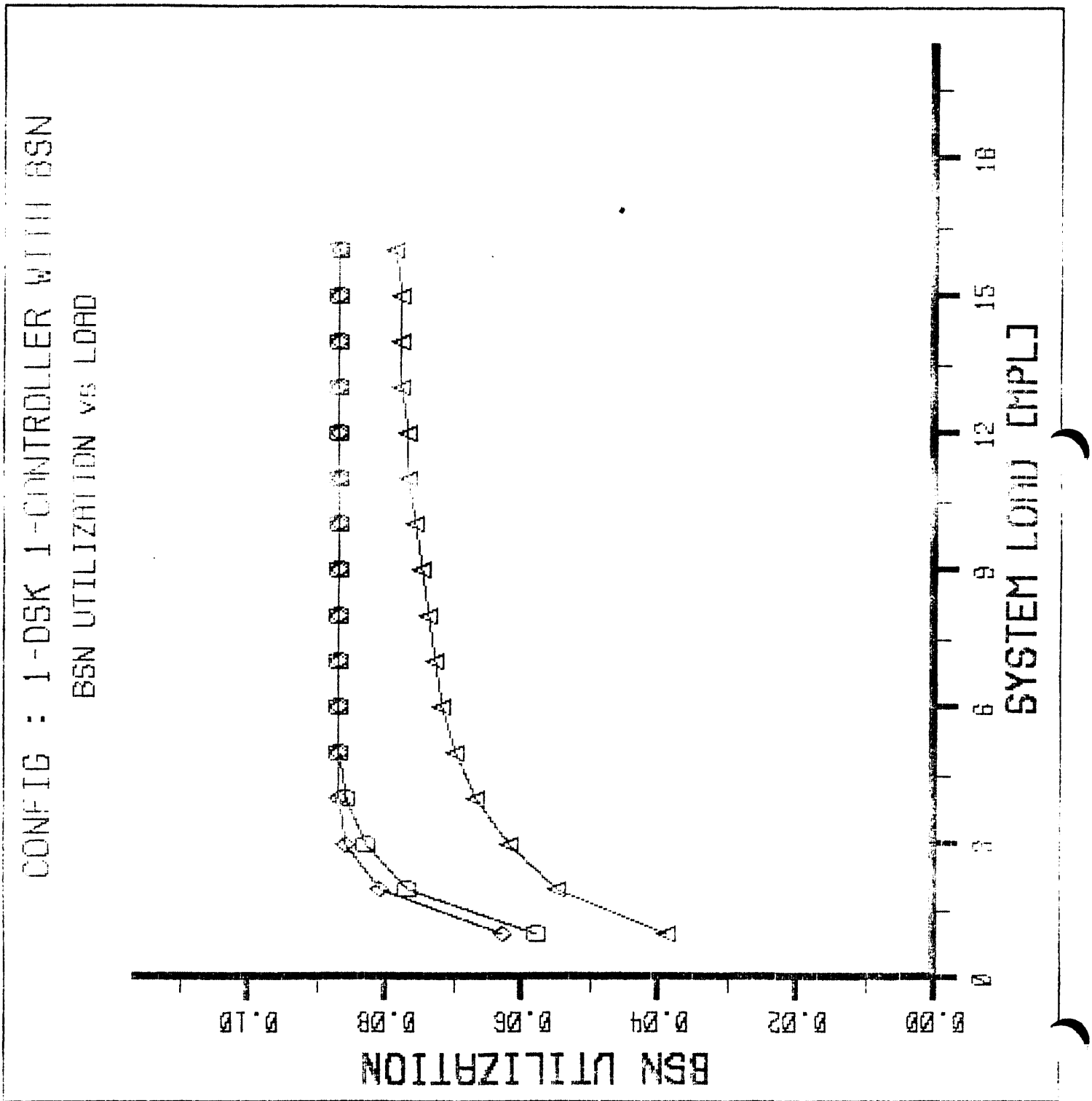
CONFIG : 1-DSK 1-CONTROLLER WITH BSN
BSN UTILIZATION vs LOAD
CURVE ID:
Δ CPU1
□ CPU3
◇ CPU5
BSN UTILIZATION
0.00
0.02
0.04
0.06
0.08
0.10
0
3
6
9
12
15
18
SYSTEM LOAD [MPL]

Figure 10

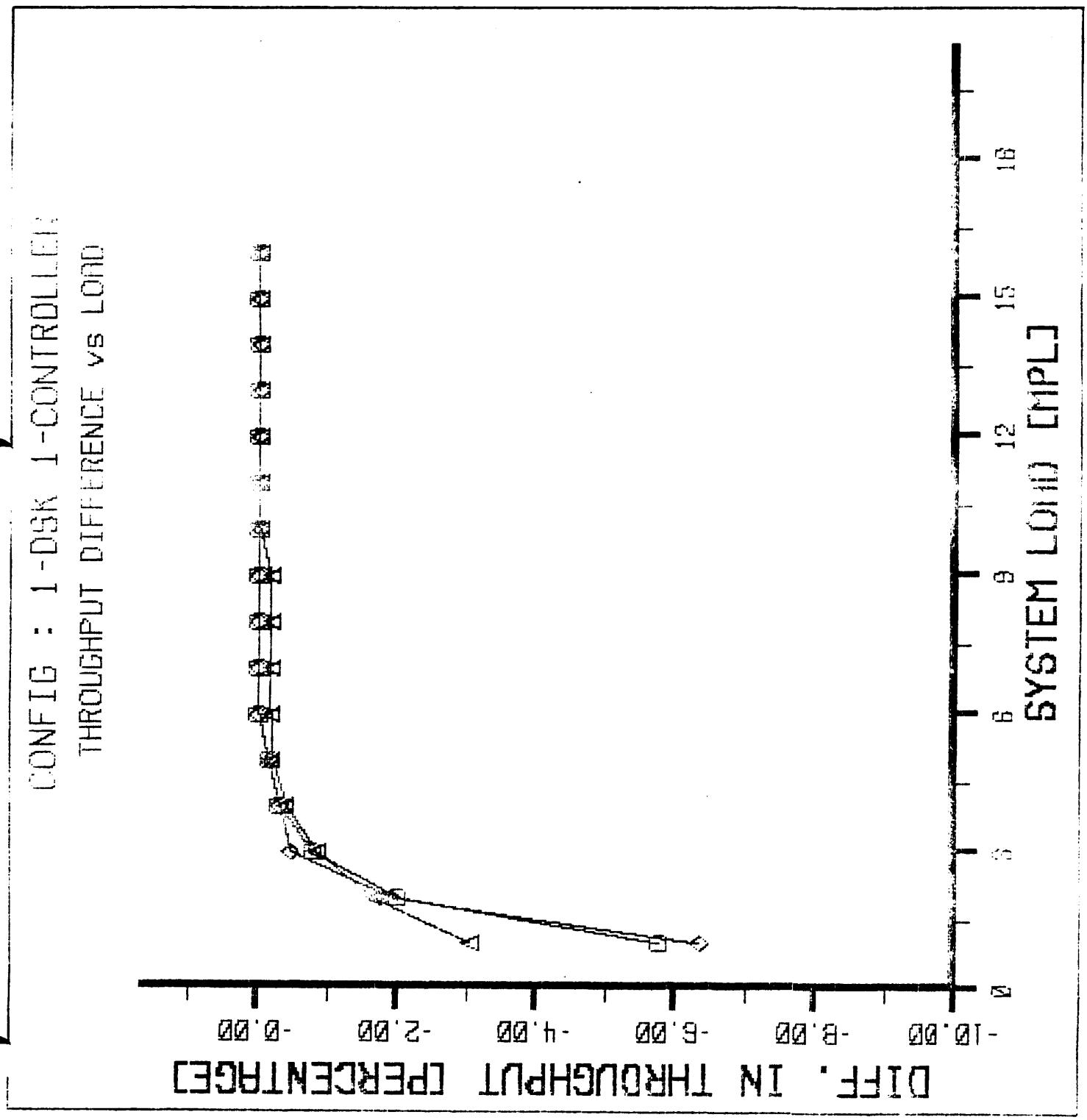
CONFIG : 1-DSK 1-CONTROLLER
THROUGHPUT DIFFERENCE vs LOAD
CURVE ID:
△ CPU1
□ CPU3
◇ CPU5
DIFF. IN THROUGHPUT [PERCENTAGE]
-10.00
-8.00
-6.00
-4.00
-2.00
-0.00
0
3
6
9
12
15
18
SYSTEM LOAD [MPL]

Figure 11

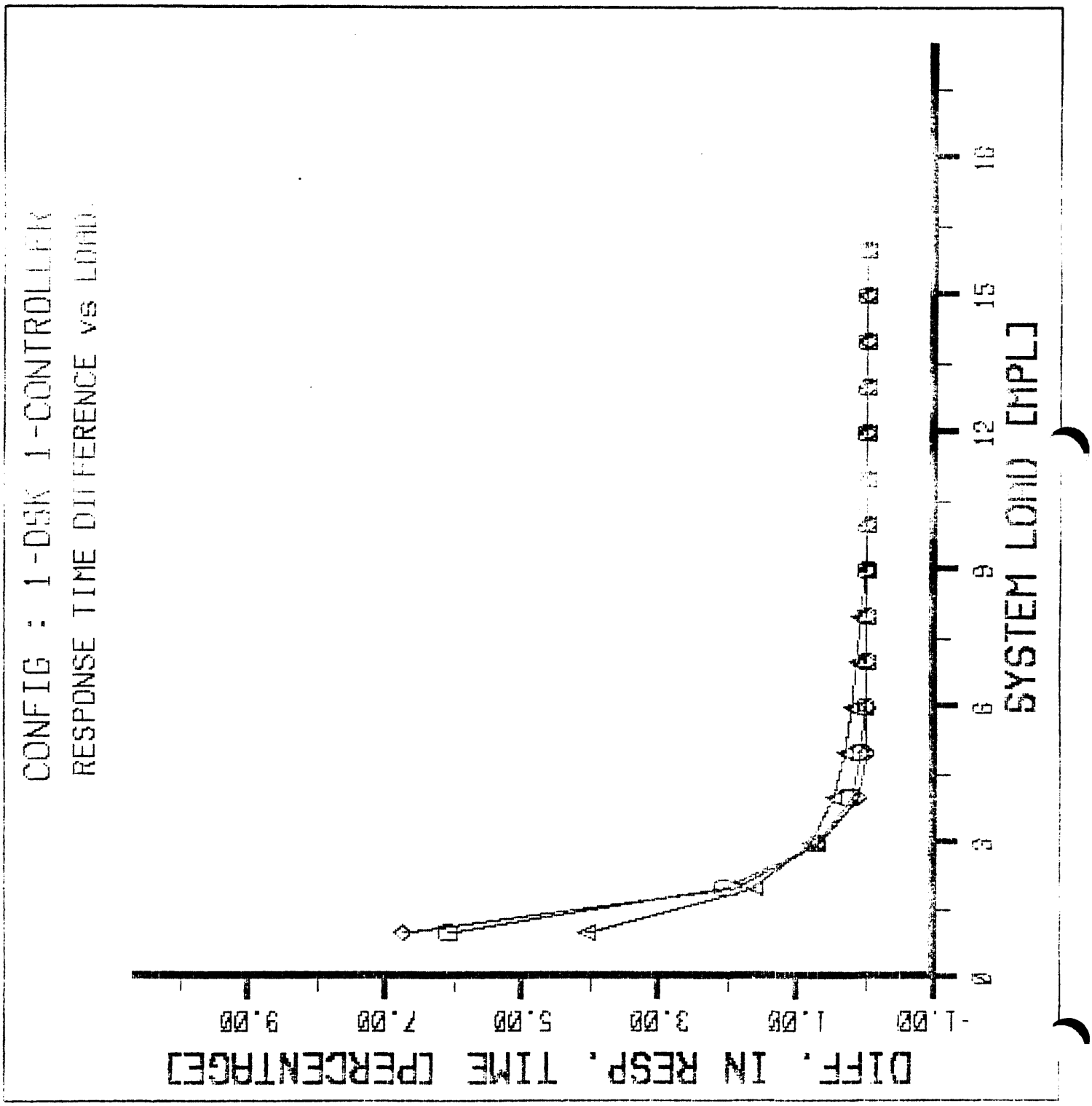
CONFIG : 1-DSK 1-CONTROLLER
RESPONSE TIME DIFFERENCE vs LOAD
CURVE ID:
△ CPU1
□ CPU3
◇ CPU5
DIFF. IN RESP. TIME [PERCENTAGE]
-1.00
1.00
3.00
5.00
7.00
9.00
0
3
6
9
12
15
18
SYSTEM LOAD [MPL]

Figure 12

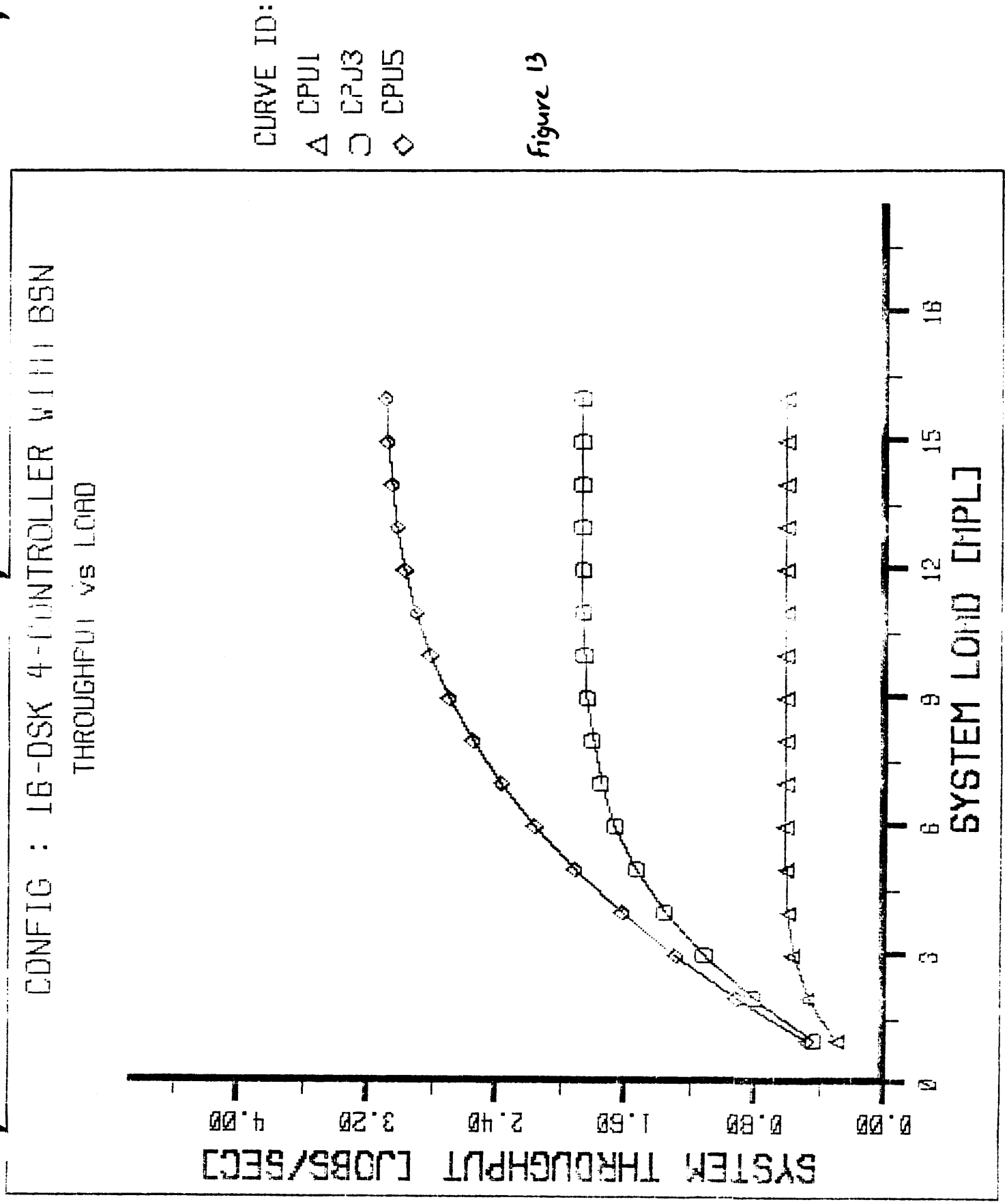
CONFIG : 16-DSK 4-CONTROLLER WITH BSN
THROUGHPUT vs LOAD
CURVE ID:
△ CPU1
□ CPU3
◇ CPU5
SYSTEM THROUGHPUT [JOBS/SEC]
0.00
0.80
1.60
2.40
3.20
4.00
SYSTEM LOAD [MPL]
0
3
6
9
12
15
18

Figure 13

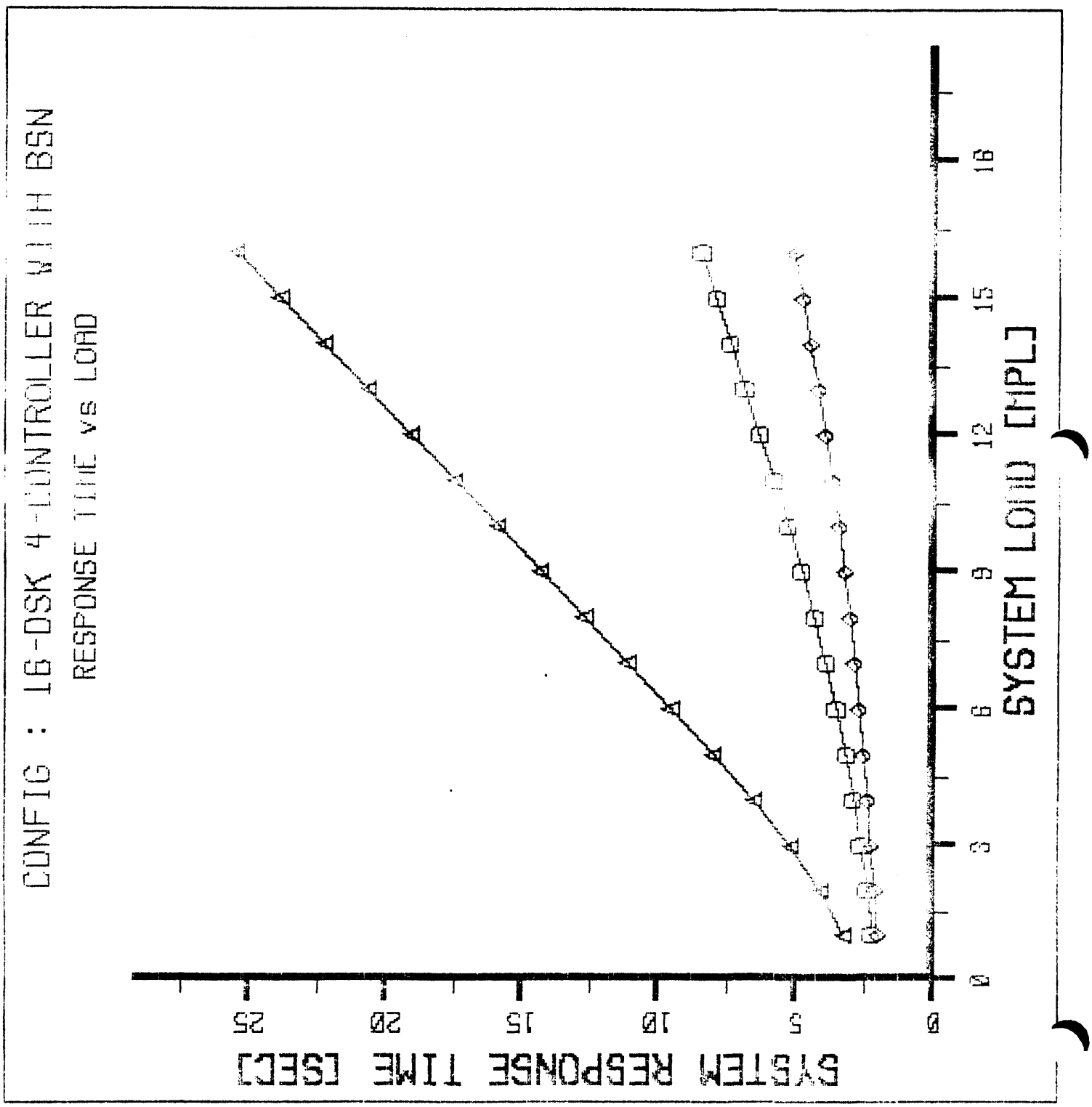
CONFIG : 16-DSK 4-CONTROLLER WITH BSN
RESPONSE TIME vs LOAD
CURVE ID:
△ CPU1
□ CPU3
◇ CPU5
SYSTEM RESPONSE TIME [SEC]
0
5
10
15
20
25
SYSTEM LOAD [MPL]
0
3
6
9
12
15
18

Figure 14

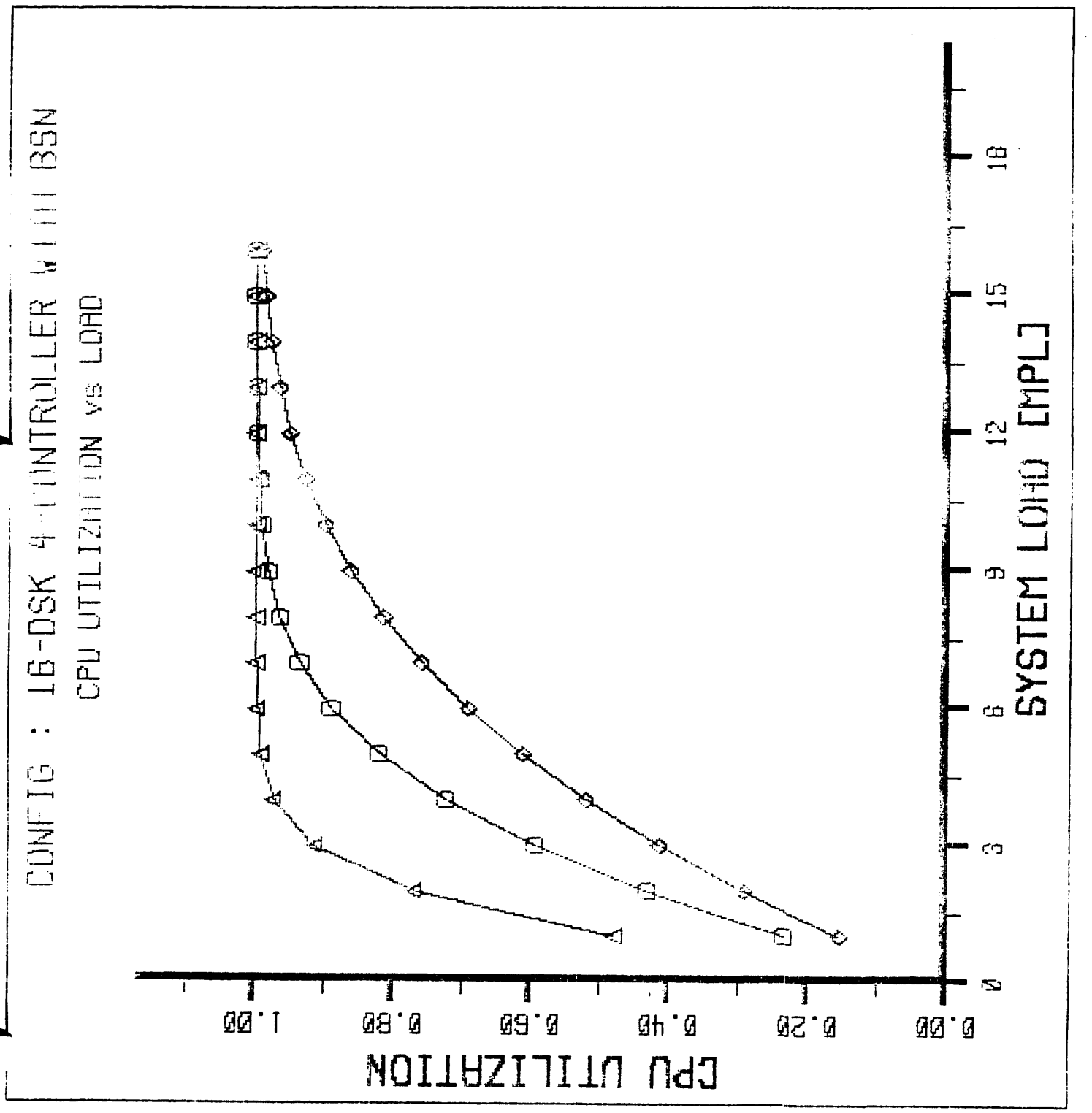
CONFIG : 16-DSK 4-CONTROLLER WITH BSN
CPU UTILIZATION vs LOAD
CURVE ID:
CPU1
CPU3
CPU5
CPU UTILIZATION
0.00
0.20
0.40
0.60
0.80
1.00
0
3
6
9
12
15
18
SYSTEM LOAD [MPL]

Figure 15

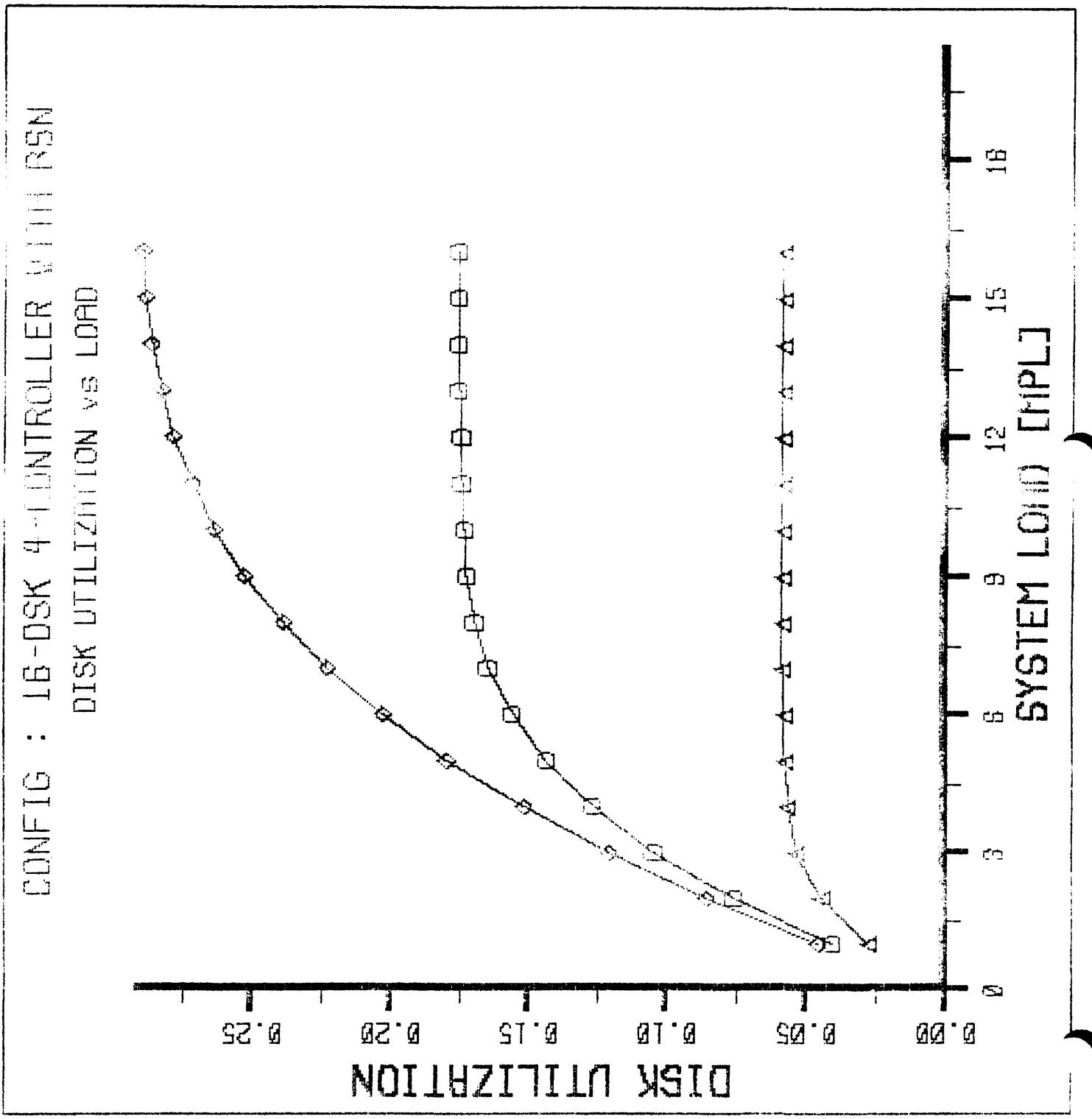
CONFIG : 16-DSK 4-CONTROLLER WITH RSN
DISK UTILIZATION vs LOAD
CURVE ID:
Δ CPU1
□ CPU3
◇ CPU5
DISK UTILIZATION
0.00
0.05
0.10
0.15
0.20
0.25
0
3
6
9
12
15
18
SYSTEM LOAD [TPL]

Figure 16

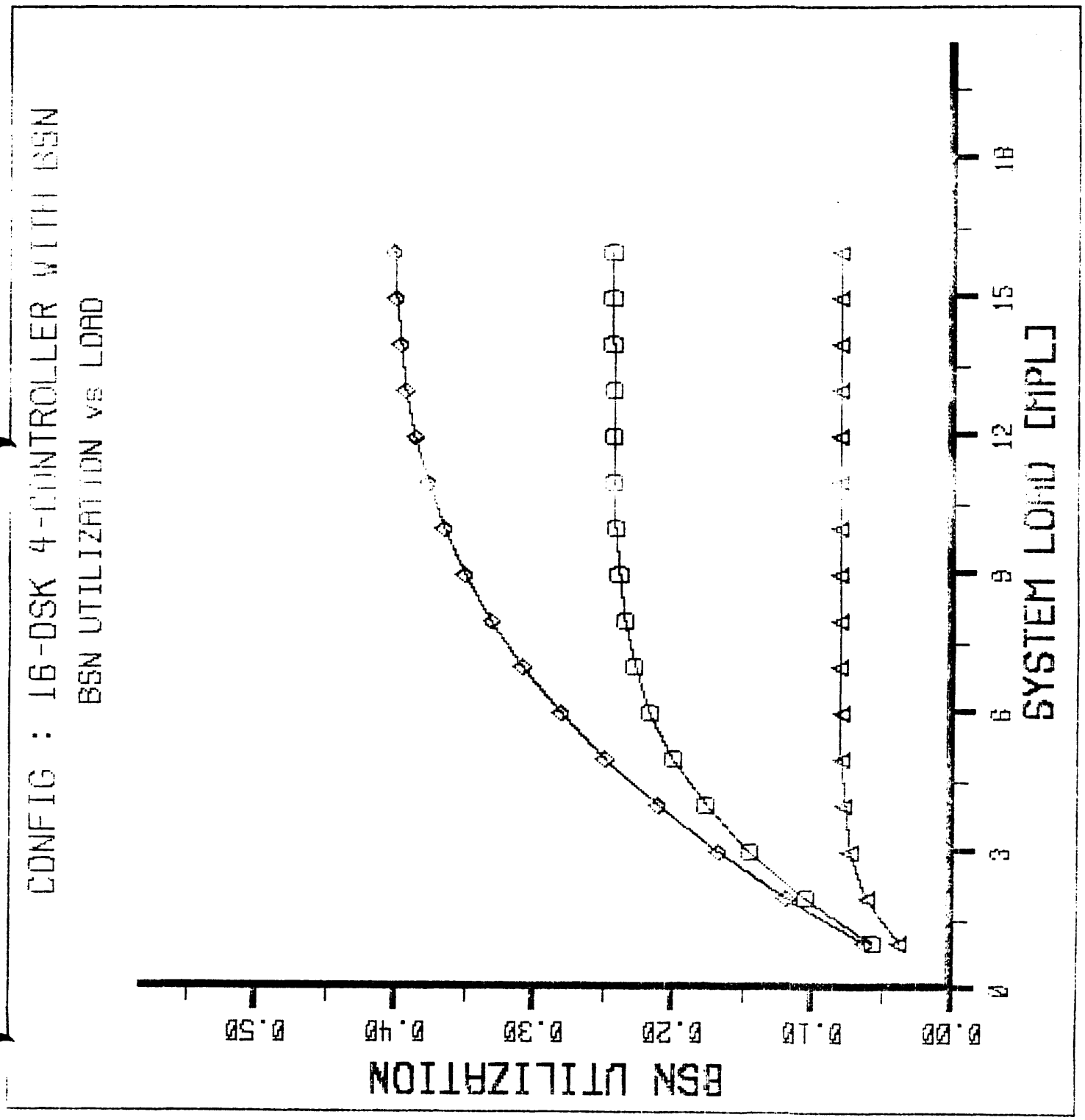
CONFIG : 16-DSK 4-CONTROLLER WITH BSN
BSN UTILIZATION vs LOAD
CURVE ID:
△ CPU1
□ CPU3
◇ CPU5
BSN UTILIZATION
0.00
0.10
0.20
0.30
0.40
0.50
0
3
6
9
12
15
18
SYSTEM LOAD [MPL]

Figure 17

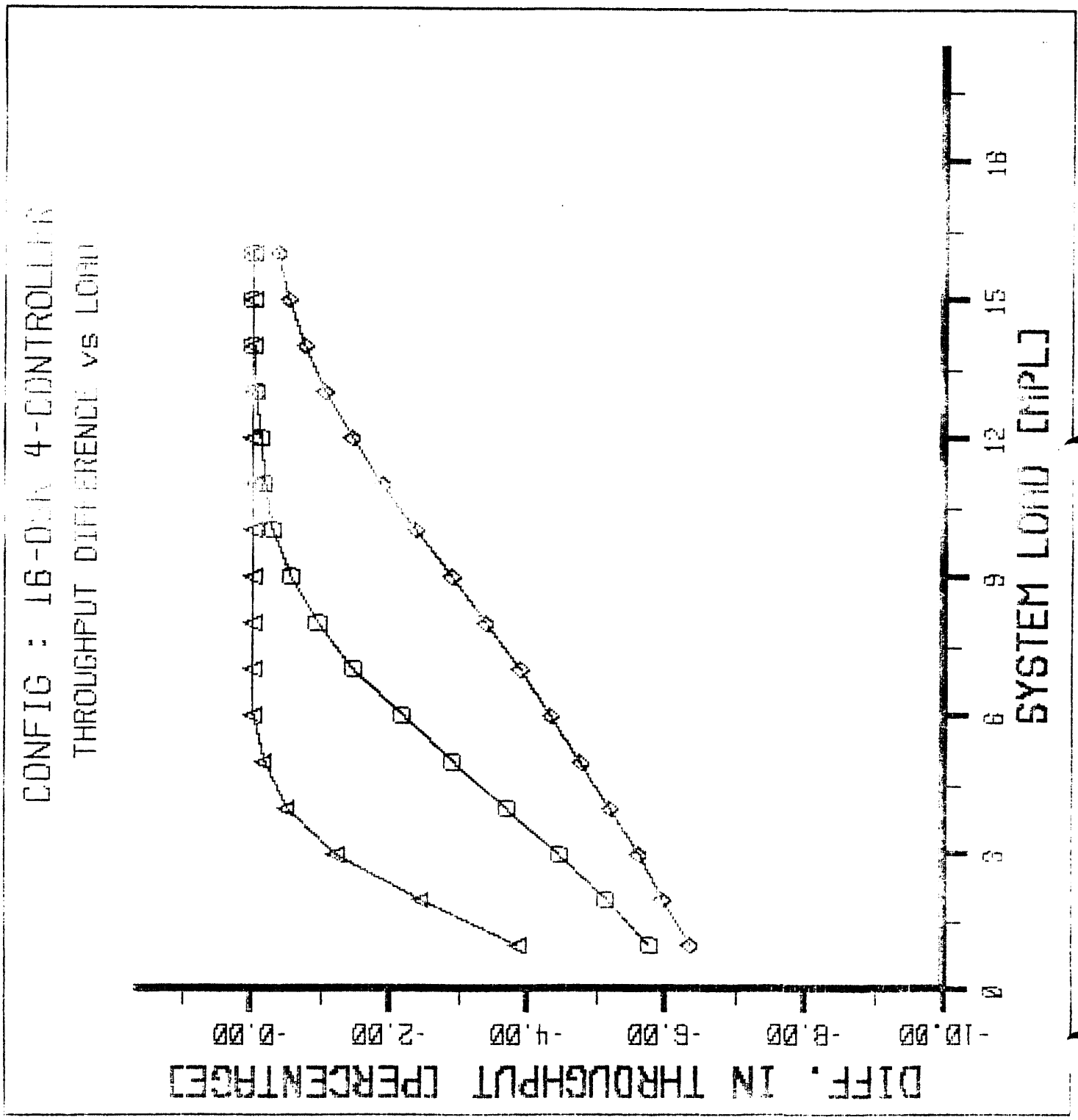
CONFIG : 16-DISK 4-CONTROLLER
THROUGHPUT DIFFERENCE vs LOAD
CURVE ID:
△ CPU1
□ CPU3
◇ CPU5
DIFF. IN THROUGHPUT (PERCENTAGE)
-10.00
-8.00
-6.00
-4.00
-2.00
-0.00
0
3
6
9
12
15
18
SYSTEM LOAD (TPS)

Figure 18

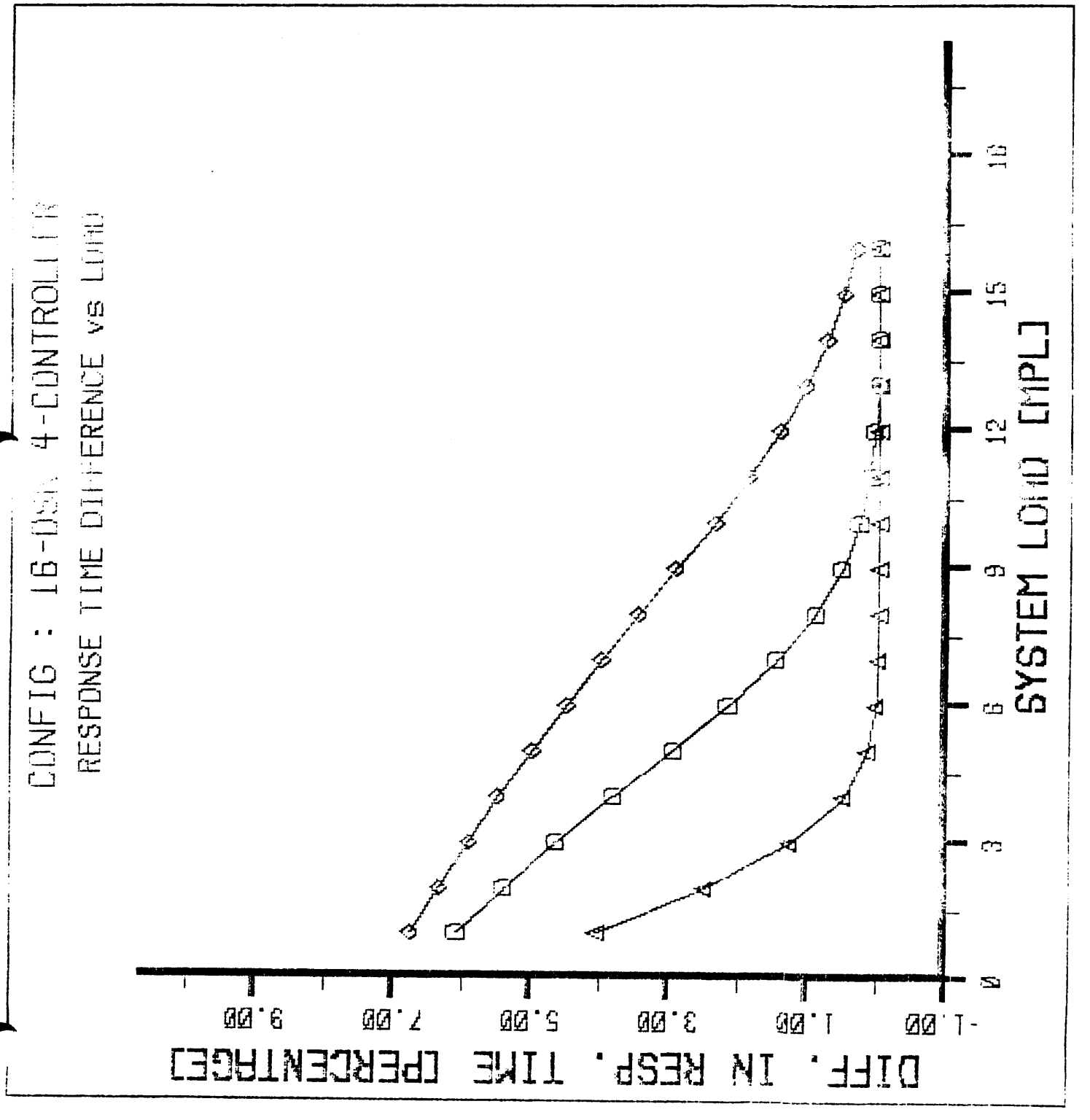
CONFIG : 16-DSK, 4-CONTROLLERS
RESPONSE TIME DIFFERENCE vs LOAD
CURVE ID:
CPU1
CPU3
CPU5
DIFF. IN RESP. TIME [PERCENTAGE]
-1.00
1.00
3.00
5.00
7.00
9.00
0
3
6
9
12
15
18
SYSTEM LOAD [MPL]

Figure 19

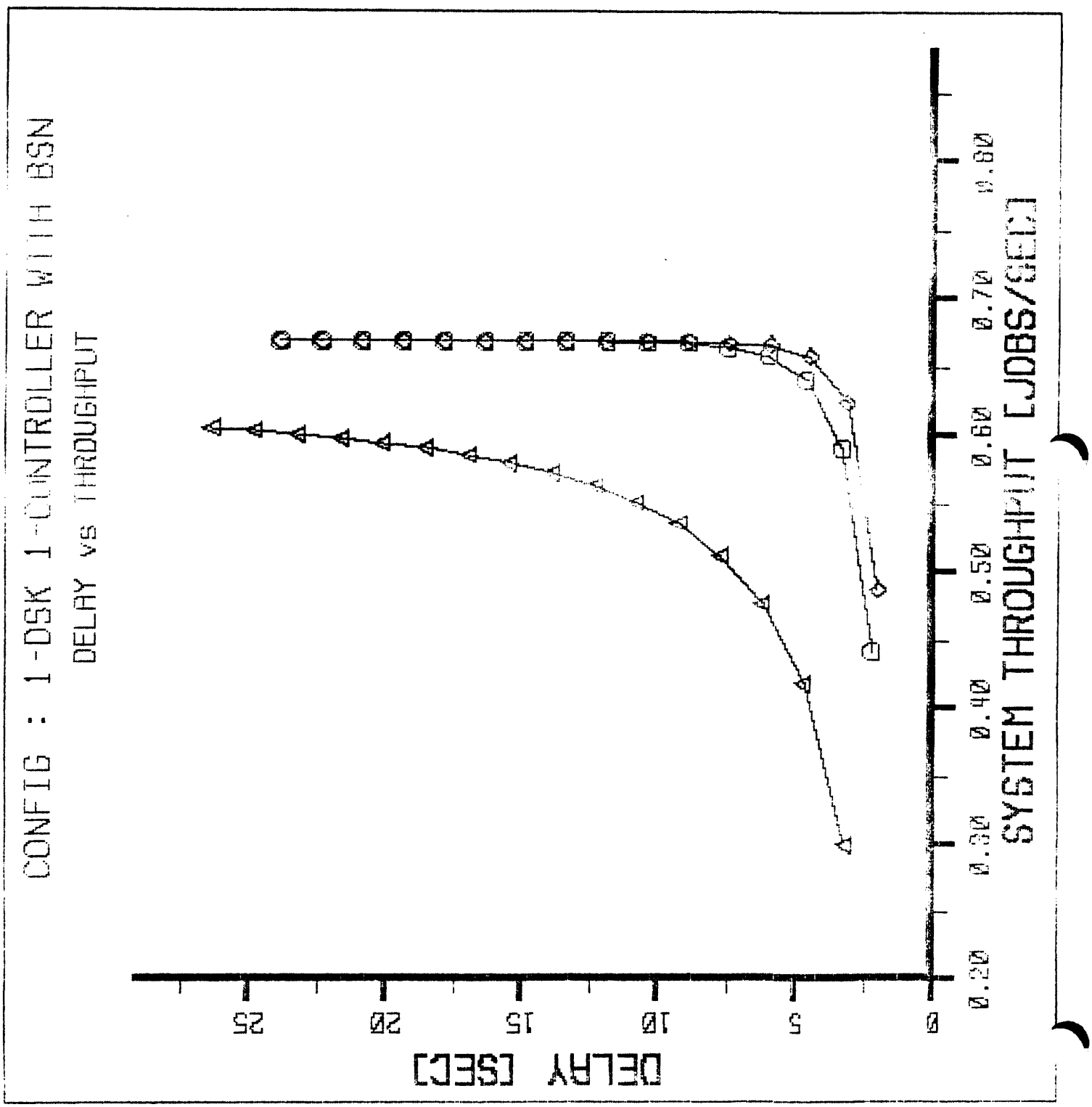
CONFIG : 1-DSK 1-CONTROLLER WITH BSN
DELAY vs THROUGHPUT
CURVE ID:
△ CPU1
○ CPU3
◇ CPU5
DELAY (SEC)
0
5
10
15
20
25
0.20
0.30
0.40
0.50
0.60
0.70
0.80
SYSTEM THROUGHPUT (JOBS/SEC)

Figure 20

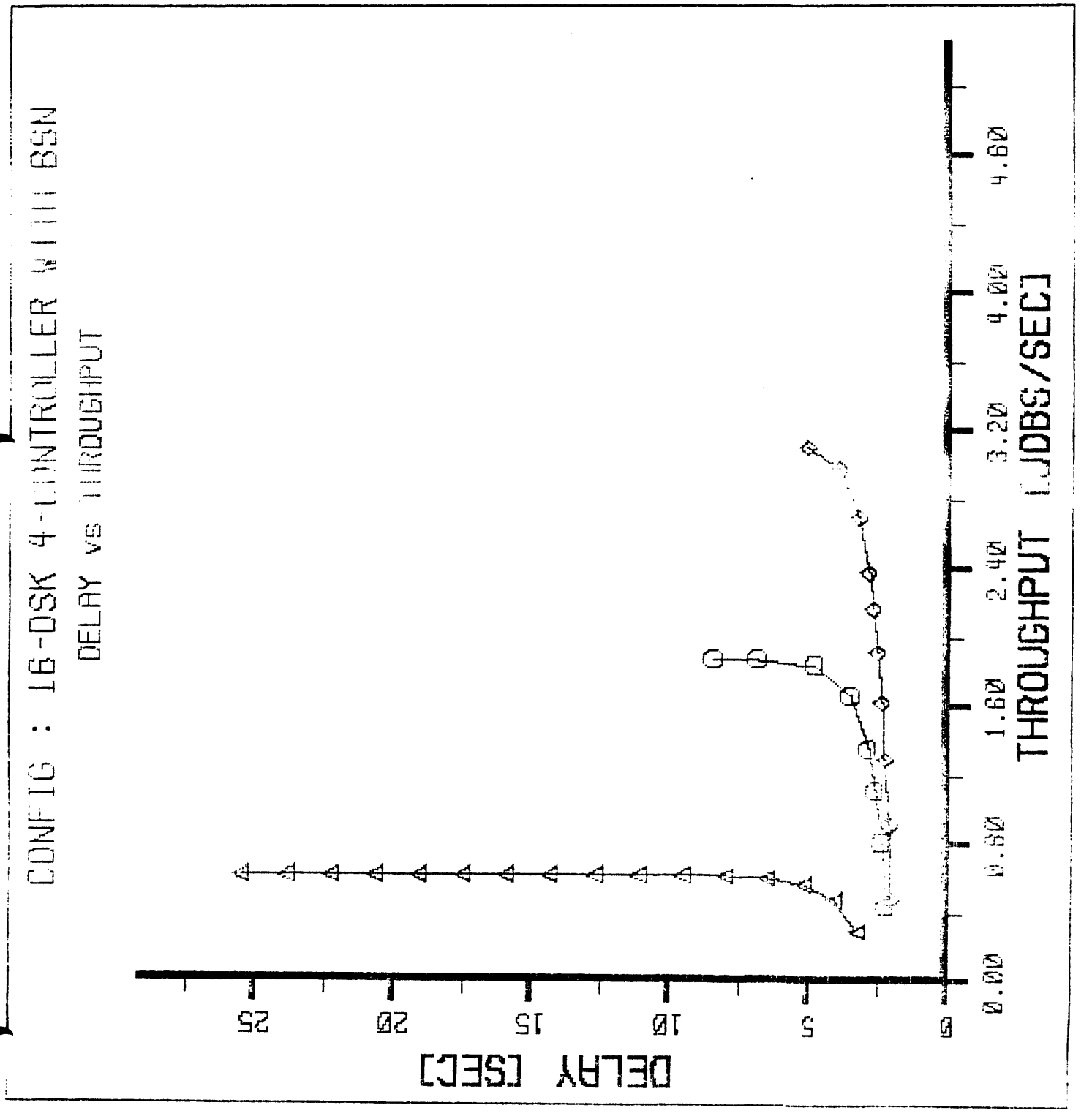
CONFIG : 16-DSK 4-CONTROLLER WITH BSN
DELAY VS THROUGHPUT
CURVE ID:
△ CPU1
□ CPU3
◇ CPU5
DELAY [SEC]
0
5
10
15
20
25
0.00
0.80
1.60
2.40
3.20
4.00
4.80
THROUGHPUT [JOBS/SEC]

Figure 21.

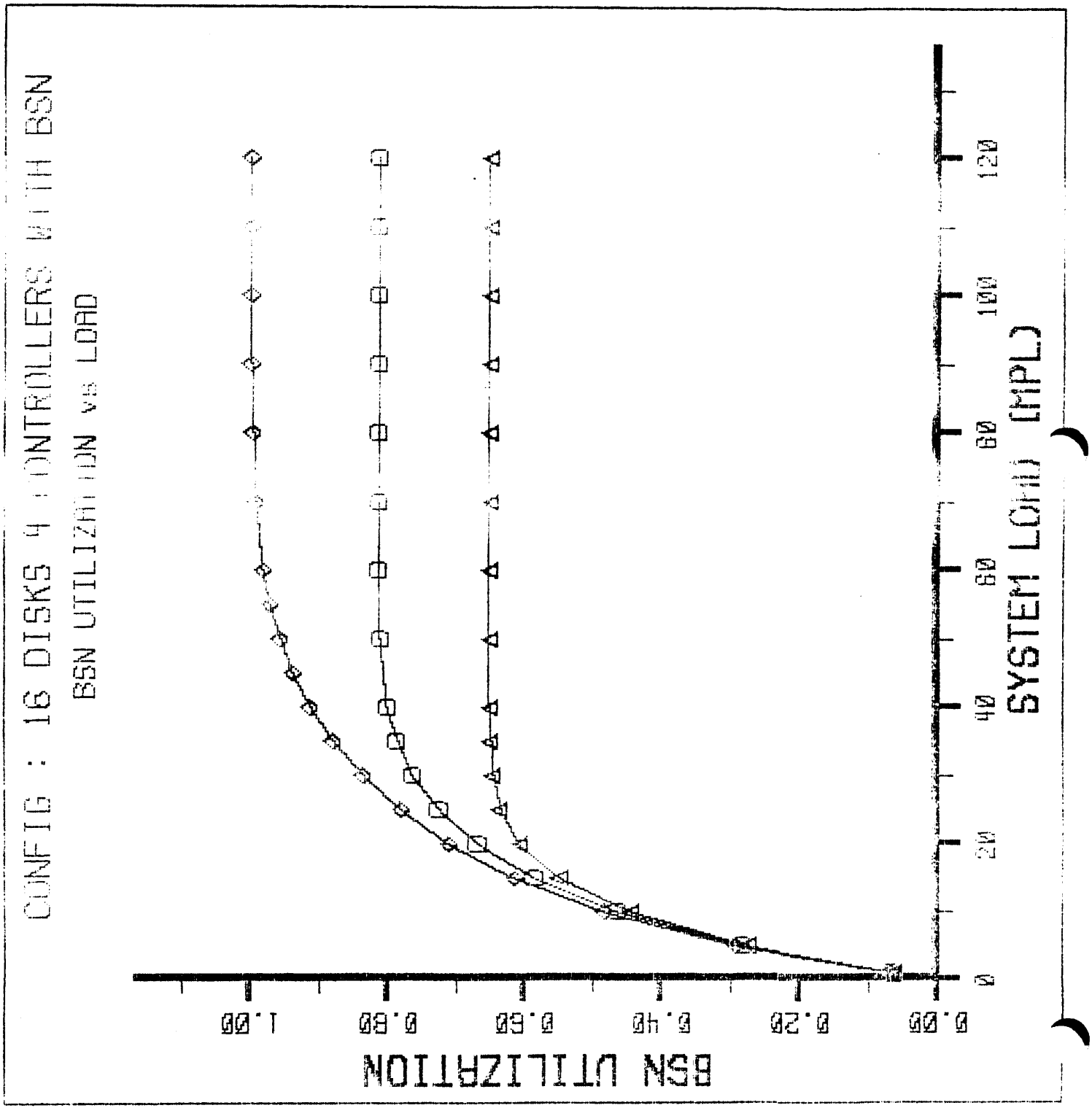
CONFIG : 16 DISKS 4 CONTROLLERS WITH BSN
BSN UTILIZATION vs LOAD
CURVE ID:
Δ CPU8
□ CPU10
◇ CPU15
BSN UTILIZATION
0.00
0.20
0.40
0.60
0.80
1.00
0
20
40
60
80
100
120
SYSTEM LOAD (MPL)

Figure 22