GEM TN-93-286

GEM Detector Computing Study: Study of Combined On-Line (Level 3) and Off-Line Facility

K. McFarlane and L. Cormell
SSC Laboratory

September 18, 1992

Abstract:

This report describes the results of a study carried out by E-Systems, Inc. that used discrete-event system simulation to analyze various configurations of an architecture for the GEM detector on-line and off-line computing facility. Primary emphasis was given to the computing resources required to perform Level 3, Pass 1 and Pass 2 processing. Also included was an analysis of the mass storage requirements to store all Level 3 and Pass 1 output and to retrieve all Pass 2 data as a post-processing activity. The results were that the chosen architecture was scalable and behaved in a predictable way. Input data rates of order 3 G Bytes/s along with a total computational capacity of a million SSCUPs and a storage capacity of a few PB were used.

Preface

The following document is the result of a study of a combined Level 3 and off-line facility. The basic assumption was that events from an event builder would be processed at up to 3 GB/s input rate.

The facility would simultaneously be doing a re-calibration repeat of PASS1, called PASS2 in this report. PASS1, simulation and other off-line tasks, and the data flow through the network, CPU's and storage systems are simulated in detail.

A variety of issues are studied: different assumed Level 3 processing times, fluctuations in event rate, optimization of the storage system, and scalability of the architecture.

The report demonstrates the feasibility of the architecture and shows that it scales easily to allow optimization of performance and cost, though work remains to be done. The study was carried out be E-Systems of Garland, Texas with requirements set by K. McFarlane and Laird Cormell, among others.

SUPERCONDUCTING SUPER COLLIDER LABORATORY

GEM Detector Computing Study

FINAL REPORT

ABSTRACT

This report describes the results of a study that used discrete-event system simulation to analyze various configurations of an architecture for the GEM detector on-line and off-line computing facility. Primary emphasis was given to the computing resources required to perform Level 3, Pass 1 and Pass 2 processing. Also included was an analysis of the mass storage requirements to store all Level 3 and Pass 1 output and to retrieve all Pass 2 data as a post-processing activity.

E-SYSTEMS

GARLAND DIVISION

E-SYSTEMS INC. • P.O. BOX 660023 • DALLAS, TEXAS 75266-0023

DOCUMENT NUMBER 416-34065

18 SEPTEMBER 1992

E-SYSTEMS INC., GARLAND DIVISION • P.O. BOX 660023 • DALLAS, TEXAS 75266-0023

Contents

Figures

Tables

This page intentionally left blank

1.0 Introduction

On May 4, 1992 E-Systems began a five month study contract to analyze and quantify various computing and data storage configurations corresponding to the on-line and off-line processing requirements of the GEM detector. This report describes the results of this study and is divided into 5 major sections.

Section 1.0, the Introduction, describes the organization of this report and the objectives of the overall study. Section 2.0, System Overview, briefly describes the GEM computing requirements and system operation. Sections 3.0 and 4.0 describe the simulation model design and simulation results respectively. The last section, 5.0, provides a summary of the study results. For readers who desire an abbreviated review of this report, sections 1.0, 2.0 and 5.0 are recommended.

The proposed GEM computing facility provides all hardware and software resources required to perform Level 3 processing (also referred to as On-Line) and Pass 1 and Pass 2 processing (also referred to as Off-Line). Projected GEM detector trigger rates range from 300 to 3,000 one megabyte (MB) events per second. In addition, the computing facility will support various analysis and simulation activities. SSC personnel had selected an architecture of loosely-coupled high-end workstations connected by high speed networks and accessible to a petabyte size mass storage system.

At the end of the requirements phase of this study contract, five questions were identified:

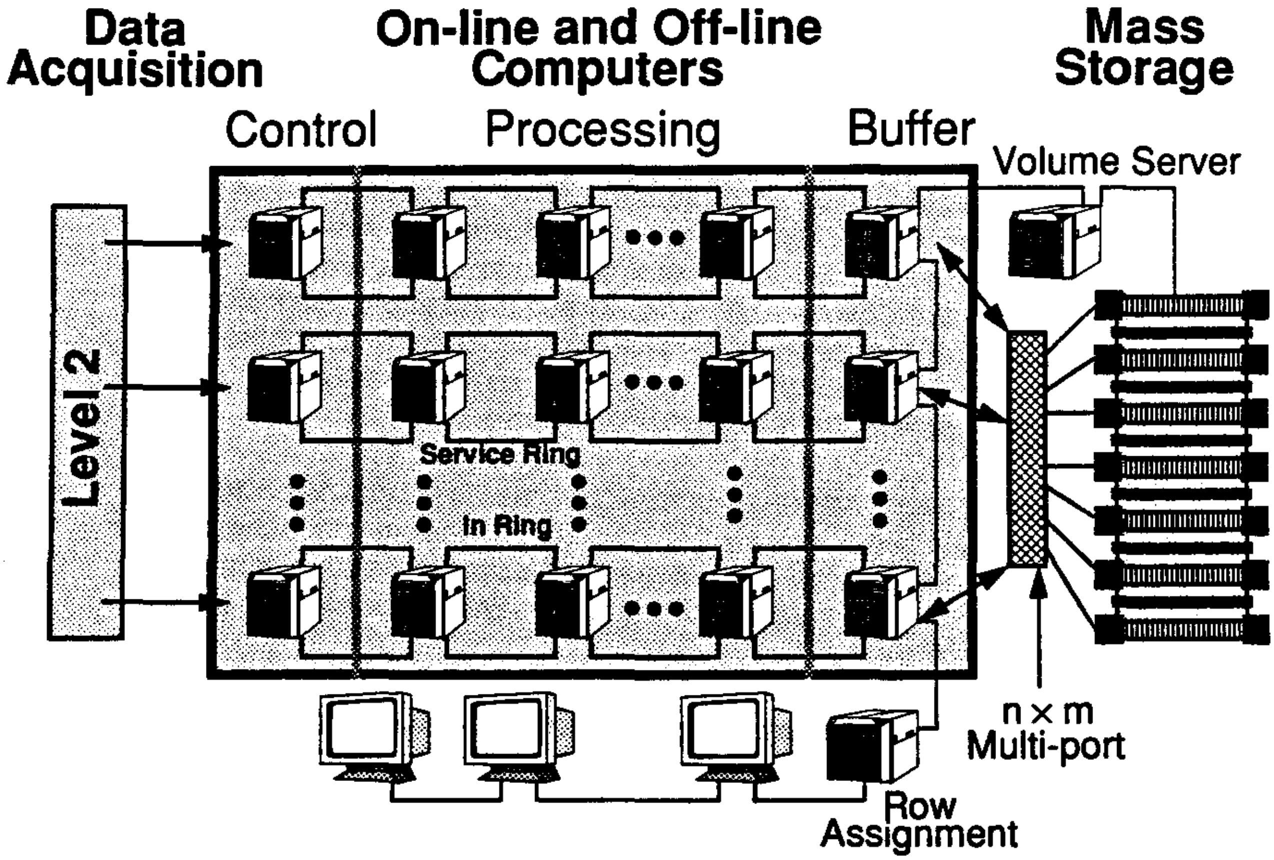
- Can Pass 1 be performed on the same CPU and immediately after Level 3?
- How many workstations will be required?
- How should the workstations be connected? For example, many short rows or a few long rows?
- What type of network throughput will be required?
- Assuming the use of an aisle-based mass storage system, how many robots and tape drives would be required?

In order to answer these questions, E-Systems developed a discrete-event system simulation model of the GEM computing facility. The model is highly parameterized and can be configured to represent a wide variety of specific hardware configurations. From a functional point-of-view, the model simulates all time delays associated with Level 3, Pass 1 and Pass 2 processing.

2.0 System Overview

The system architecture investigated in this study consists of a large collection (at least 1,000) of high-end workstations connected by high-speed networks and a mass storage system containing up to 15 petabytes of data. This architecture is designed to handle both On-line and Off-line processing without having two distinct sets of computers. Figure 2.0-1 shows a conceptual drawing of this system
.

Figure 2.0-1 System Design



The computing ranch consists of rows of multi-CPU UNIX™ workstations connected by high-speed networks. Each CPU row is made up of at least three processors, each having a specific task. The "Control Processor" is designed to temporarily buffer incoming events until they can be assigned and transferred to one of the "Computing Processors".

The "Computing Processors" are designed to perform all Level 3, Pass 1, and Pass 2 processing. Each Computing Processor will be assigned one event per CPU. As soon as a Computing Processor completes a processing task, the output data is transferred to the "Mass Storage System (MSS) Buffer Processor" and another event is assigned to the workstation. Each Computing Processor is assumed to have enough main memory so that each CPU within the workstation can be concurrently processing a different event without going to disk. Pass 1 processing is always performed immediately after and with the same CPU as Level 3 processing.

After each processing step the output data is transferred to the MSS Buffer Processor. The MSS Buffer Processor is responsible for two tasks. One task is to accumulate processed events into large files which are then stored to tape in the mass storage system. The other task is to receive files from the mass storage system and then break each file into events for Pass 2 processing (Off-Line).

The workstations on each CPU row have two high speed network connections. The "In Ring" is a uni-directional network and carries data from the Control Processor to a Computing Processor. The bi-directional "Service Ring" network carries processed data from a Computing Processor to its respective MSS Buffer Processor. The Service Ring also carries data that has been retrieved from mass storage to its respective Control Processor.

The computing ranch is very scalable. Its dimension could vary from a single CPU row with many Computing Processors to multiple CPU rows with only a few or even one Computing Processor. One objective of the simulation study was to find the most appropriate dimension.

Three important software-oriented assumptions were made while performing this study. First, Level 2 is expected to assign each event to a specific Control Processor (or CPU row) in a cyclic manner. Second, the Control Processor can differentiate events for On-Line or Off-Line processing with events for On-Line Processing given priority during allocation of CPUs. And third, during On-Line processing, the Computing Processors have software which will create a process to store Level 3 processed events while continuing Pass 1 processing.

The architecture of the mass storage system (MSS) consists of aisles containing media with high-speed tape drives at the ends of each aisle. Robotics are used to move media from a shelf to a tape drive and back. The MSS holds up to 15 petabytes of data. A volume server computer controls the allocation of robots and tape drives. This volume server computer is connected to each of the MSS Buffer Processors.

An nxm Multiport is used as the connection between tape drives in the mass storage system and the MSS Buffer Processors. This port allows data on any row to flow to any tape drive and similarly from any tape drive to any MSS Buffer Processor.

Two possibilities exist for the database containing location information of a file on a media. First, the database can be distributed across the MSS Buffer Processors such that each of these processors can access every file in the MSS. Or second, each MSS Buffer Processor contains only a subset of the entire database for which a meta-database computer is required to keep track of files associated with a specific CPU row. In both cases a processor is needed for assigning Off-Line tasks to a specific CPU row for the purpose of load balancing.

This page intentionally left blank

3.0 Simulation Model Description

By designing a simulation model of the system described in Section 2.0, several objectives can be achieved. First, by modelling the system in detail, problems which might otherwise be overlooked can be encountered thus leading to improved decisions in design. Second, "back of the envelope" calculations can be validated while greater insight to the system is being observed. A picture of system performance and the nuances contained within becomes much clearer. Furthermore, by studying the system in steady-state, the feasibility of the architecture can be determined.

The following sections describe in detail the model parameters, elements, and methodology used to simulate the system.

3.1 Model Overview

The system for which the simulation model was built was described in Section 2.0. Certain assumptions were made for modelling ease. Such assumptions include a distributed database across the MSS Buffer Processors and groups of tape drives being restricted to a specific robot. Further assumptions were made with regards to parameter values. Since this system will not be built for another four or five years, some parameter values reflect extrapolated performance of future products. Examples are commercially available networks with 100 megabytes per second (MB/s) effective throughputs and workstations with multiple CPUs.

Sections 3.2 and 3.3 give a detailed representation of the parameters and the simulation model. These details reflect important considerations of the system's operation and should be given careful attention. Graphical representations of the simulation model are included in the Appendix.

3.2 Parameters

The parameters used as input to the simulation model can be divided into four basic types:

- Run-time
- System Configuration
- Mass Storage Configuration
- Derived Input

Run-time parameters are listed in Table 3.2-I and are associated with the length of simulated run time, frequency of reports, and an identifier for reports. The parameter *warmup_time* is the length of time to let the system run before collecting statistics. The default value of 120 seconds was chosen so that the first file requests for Pass 2 data would be completely retrieved to the MSS Buffer Processor before statistic collection began. Otherwise some statistics would have been biased downward due to system inactivity. Statistics are collected on a batch basis. By having several contiguous batches, non-steady state performance is more easily identified. The number of batches is set by *number_of_batches* with the time length of each batch being *batch_time*. Statistics for each batch were appended to a report file. Snapshot files were also created which contained system state information every *snapshot_interval* seconds. Each file created during a simulation run had a prefix of *run_name*.

Table 3.2-I Run-Time Parameters

Description	Name	Default Value
Warm-Up Time (seconds)	*warmup_time*	120
Batch Time (seconds)	*batch_time*	900
Snapshot Interval (seconds)	*snapshot_interval*	5
Number of Batches	*number_of_batches*	4
Run Identifier	*run_name*	test

The next type of parameters are system related parameters. These are parameters associated with the configuration of the processors including connectivity and the On-line and Off-line processing requirements. The parameters are listed in Table 3.2-II and are discussed further in Section 3.3.

Table 3.2-II System Parameters

Description	Name	Default Value
Number of CPU Box Rows	*cpu_box_rows*	32
Number of CPU Boxes per Row	*boxes_per_row*	5
Number of CPUs per Box	*cpus_per_box*	8
Control CPUs Memory Buffer Size (MB)	*control_buffer*	150
Level 3 Input (Events per Second)	*l3_arrival_rate*	3,000
Changed Level 3 Input (Events per Second)	*new_l3_arrival_rate*	3,000
Level 3 Arrival Rate Change Time (seconds)	*rate_change_time*	7200
Level 3 Mean Event Size (MB)	*l3_event_size*	1.00
Standard Deviation of Level 3 Event Size	*l3_std_dev*	0.15
Control Processing Time (seconds)	*control_processor_time*	0.001
MSS Buffer Processing Time (seconds)	*buffer_processor_time*	0.001
CPU Power (SSCUPs)	*cpu_power*	500
In Ring Effective Rate (MB/s)	*in_ring_rate*	100
Service Ring Effective Rate (MB/s)	*service_ring_rate*	100
Vertical Ring Switch	*vertical_ring_switch*	OFF
Vertical Ring Effective Rate (MB/s)	*vertical_ring_rate*	100
Number of Vertical Ring Trunks	*vertical_ring_trunks*	2
Packet Size on Networks (MB)	*network_packet_size*	1.00
Level 3 Processing (SSCUPs/MB)	*l3_processing*	67
Level 3 Output Ratio	*l3_output_ratio*	1.05
Level 3 Events of NO Interest	*l3_no_interest*	0.967
Store Level 3 Data Switch	*store_l3_switch*	YES
Pass 1 Processing (SSCUPs/MB)	*p1_processing*	2100
Pass 1 Output Ratio	*p1_output_ratio*	2.00
Pass 1 Events of NO Interest	*p1_no_interest*	0.50
Percent Control Memory Required for Pass 2 Data	*p2_control_mem_req*	0.67

The third type of input parameters are associated with the Mass Storage System. These parameters are listed in Table 3.2-III and are also discussed in *Section* 3.3.

Table 3.2-III MSS Parameters

Description	Name	Default Value
File Size Written to Tape (MB)	*archive_file_size*	500
File Size Retrieved from Tape (MB)	*retrieve_file_size*	500
Number of Robots	*number_of_robots*	12
Store Tape Drives per Robot	*s_drives_per_robot*	1
Retrieve Tape Drives per Robot	*r_drives_per_robot*	2
Robot Delay Time (seconds)	*robot_delay*	20
Load and Position Tape for Store (seconds)	*store_load_position*	9
Load and Position Tape for Retrieve (seconds)	*retrv_load_position*	20
Rewind and Eject Tape Time (seconds)	*rewind_eject*	13
Effective Tape Drive Transfer Rate (MB/s)	*transfer_rate*	30

The final type of parameters are derived from the input parameters. These parameters are listed in Table 3.2-IV. Note especially that all parameters associated with Pass 2 events are derived from Level 3 and Pass 1 parameters.

Table 3.2-IV Derived Parameters

Description	Name	DefaultValue
Total Tape Drives per Robot	*drives_per_robot*	3
Total CPUs per Row	*cpus_per_row*	40
Total CPUs	*total_cpus*	1280
Pass 2 Mean Event Size (MB)	*p2_event_size*	2.10
Standard Deviation of Pass 2 Event Size	*p2_std_dev*	0.315
Pass 2 Arrival Rate (Files per second)	*p2_arrival_rate*	0.21
Changed Pass 2 Arrival Rate	*new_p2_arrival_rate*	0.21
Pass 2 Processing (SSCUPs/MB)	*p2_processing*	1050
Memory Limit Allowing Pass 2 Processing (MB)	*p2_memory_cutoff*	50
Total Run Time (seconds)	*total_run_time*	3720

The derived parameters are calculated as shown in Table 3.2-IV:

Table 3.2-V Derived Parameter Calculations

Parameter Name	Calculation
drives_per_robot	= s_drives_per_robot + r_drives_per_robot
cpus_per_row	= boxes_per_row x cpus_per_box
total_cpus	= cpus_per_row x cpu_box_rows
p2_event_size	= l3_event_size x l3_output_ratio x p1_output_ratio
p2_std_dev	= l3_std_dev x l3_output_ratio x p1_output_ratio
p2_arrival_rate	= [l3_arrival_rate x (1 - l3_no_interest) x (1 - p1_no_interest) x p2_event_size] /retrieve_file_size
p2_std_dev	= l3_std_dev x l3_output_ratio x p1_output_ratio
p2_std_dev	= l3_std_dev x l3_output_ratio x p1_output_ratio
new_p2_arrival_rate	= [new_l3_arrival_rate x (1 - l3_no_interest) x (1 - p1_no_interest) x p2_event_size] / retrieve_file_size
p2_processing	$= \frac{p1\_processing}{p1\_output\_ratio}$
p2_memory_cutoff	= control_buffer x (1 – p2_control_mem_req)
total_run_time	= warmup_time + (batch_time x number_of_batches)

Each of the previous tables have a listed default value. Most of these values are based on GEM requirements and can be considered as constants for the basis of this analysis. The remaining default parameter values, which deal with actual configuration sizing like the number of CPU rows or number of robots, were determined from preliminary analysis which suggested that a particular configuration could meet the requirements of a 3,000 MB/s Level 3 arrival rate.

3.3 Detailed Design

The system modelled can be divided into four subsystems which are the networks, the control processors, the MSS buffer processors, and the mass storage system. These four subsystems and the event flow through the system are discussed in the following sections.

3.3.1 Level 3 and Pass 1 Processing

Level 3 events are generated with interarrival times following an exponential distribution with mean 1 / *l3_arrival_rate*. If *rate_change_time* is less than *total_run_time* then at *rate_change_time* the mean becomes (1 / *new_l3_arrival_rate*). The size of these events follow a normal distribution with mean *l3_event_size* and standard deviation *l3_std_dev.*

The Level 3 events are routed cyclically to a control processor on one of *cpu_box_rows* rows. Control processor software allocates a CPU on the associated row for the Level 3 processing. If a CPU is not available, then the event queues. If a CPU is available, the event flows to the CPU for processing via the In Ring network. The allocated CPU will not be assigned a new event until all Level 3 and Pass 1 processing has been completed. The processing time for Level 3 events is equal to (*event_size* * *l3_processing* / *cpu_power*). After this processing, events may be deleted with probability *l3_no_interest*. The *event_size* of events not deleted increases to *event_size***l3_output_ratio* and Pass 1 processing begins immediately on the same CPU. If the switch *store_l3_switch* is ON, then a copy of the Level 3 processed event is sent to the MSS Buffer Processor via the Service Network and then stored to tape shortly thereafter.

The Pass 1 processing time for an event is equal to (*event_size* * *p1_processing* / *cpu_power*). After Pass 1 processing, events may be deleted with probability *p1_no_interest*. The *event_size* of events not deleted increases to *event_size***p1_output_ratio* and are then passed to the MSS Buffer Processor via the Service Ring and then stored to tape shortly thereafter.

3.3.2 Pass 2 Processing

Requests to the mass storage system for files containing data for Pass 2 processing are generated with interarrival times following an exponential distribution with mean (1 / *p2_arrival_rate*). If *rate_change_time* is less than *total_run_time*, at *rate_change_time* the mean becomes (1 / *new_p2_arrival_rate*). The size of the file retrieved is *retrieve_file_size* MB and it is written to an MSS Buffer Processor on a CPU row which was assigned cyclically. At this point, events are split from the file and routed to be processed. The size of the Pass 2 events are normally distributed with mean *p2_event_size* and standard deviation *p2_std_dev*. The switch *vertical_ring_switch*, which was added due to a design change, specifies if events will be routed from the MSS Buffer Processor across CPU rows or if events will stay on the same CPU row for event processing.

The first design choice (*vertical_ring_switch* = YES) specified a number of network trunks (*vertical_ring_trunks*) between MSS Buffer Processors. Events were then assigned cyclically to a CPU row and sent to that row's MSS Buffer Processor via the vertical ring network. From each MSS Buffer Processor the events were sent down the service ring network to the corresponding Control Processor where they waited until a CPU on that row was available for allocation. Although this design processed Pass 2 events quickly, it also had two problems. First, a considerable cost was added for having the vertical ring networks, especially due to the high bandwidth needed. Second, this original design left the possibility of a Pass 2 event being deleted from the Control Buffer in the event of a memory buffer overflow.

The second design choice (*vertical_ring_switch* = NO) removed the existing problems with the first design. The vertical ring networks are removed and events stay on the CPU row associated with the MSS Buffer Processor to which the file was written. Furthermore, Pass 2 events flow one at a time to the corresponding Control Processor via a service ring network, but only as long as the Control Processor's memory is less than *p2_memory_cutoff* MB. Given appropriate parameter values, Pass 2 events will never be overflowed from the Control Processor's buffer.

Once a Pass 2 event is to the Control Processor, it waits for an available CPU. Upon allocation of a CPU, the event flows to the CPU via the In Ring network and incurs a processing delay of (*event_size* * *p2_processing* / *cpu_power*) seconds. The processed event is then passed to the MSS Buffer Processor via the Service Ring and then stored to tape shortly thereafter.

3.3.3 Networks

Three different networks are used in the simulation model. The In Ring, the Service Ring, and the Vertical Ring. The In Ring is used for moving data from the Control Processor to a CPU. The Service Ring is used for moving processed data from a CPU to the MSS Buffer Processor and for moving Pass 2 pre-processed data from the MSS Buffer Processor to the Control Processor. One In Ring and one Service Ring exist per CPU row. The Vertical Ring is used to move pre-processed Pass 2 data from one MSS Buffer Processor to another. The Vertical Ring also has a number of trunks which are chosen randomly to help maintain the high data rate.

Each network is considered to be a token ring with packets of size *network_packet_size* and is modelled as a single server round robin queue with time slice of *network_packet_size / network_rate* where *network_rate* is either *in_ring_rate*, *service_ring_rate*, or *vertical_ring_rate*.

Since the In Ring network has only one source point, namely the Control Processor, the controlling software is assumed to send an entire event to a CPU before sending the next event. Without this assumption, the network transfer time could become significantly increased. For any length queue, one packet would be sent at a time for each event in the queue. Furthermore, with this assumption the In Ring is not required to be a token ring network.

3.3.4 Control Processor

The Control Processor allocates CPUs for event processing. There is one Control Processor at the beginning of each CPU row. Each time an event passes through the Control Processor, a processing delay of *control_processor_time* is incurred for CPU allocation. If no CPU on the particular row is available, the event is added to the control queue. Events are released from the queue such that Level 3 has priority over Pass 2. When an event is released from the control queue, another *control_processor_time* is incurred.

For Pass 2 events, a signal is sent to the MSS Buffer Processor to release the next Pass 2 event if the Control Processor's memory is less than *p2_memory_cutoff* MB. If not, then a flag is set so that a Level 3 event will signal the MSS Buffer Processor to release the next Pass 2 event once the memory has dropped below *p2_memory_cutoff*. This was a modelling technique that should give similar results for an operating system sending a signal to the MSS Buffer Processor when memory drops below *p2_memory_cutoff*.

Incoming Level 3 events are lost when the control queue exceeds *control_buffer* MB of data with the addition of the new event.

3.3.5 Mass Storage System Buffer Processor

The MSS Buffer Processor acts as the interface between the CPUs and the Mass Storage System. There is one MSS Buffer Processor at the end of each CPU row and each has a direct connection to the mass storage nxm multiport. After each event is processed, it is passed to an MSS Buffer Processor where it is aggregated to one file based on event type (Level 3, Pass 1, or Pass 2). Each time an event arrives, a processing delay of *buffer_processor_time* seconds is incurred. Each aggregated file is sent to the mass storage system for storage to tape when the file reaches a size of *archive_file_size* MB. Once the file is on tape, the file is deleted from the MSS Buffer Processor.

The MSS Buffer Processor also receives files containing data for Pass 2 processing from the mass storage system. The files retrieved have a file size of *retrieve_file_size* MB. Upon retrieval of a file, a processing delay of *buffer_processor_time* is again incurred. The file is broken down into events. The events are then released one at a time to the Control Processor each time an appropriate signal is received. The event is removed from the MSS Buffer Processor as soon as the event has arrived to the Control Processor.

3.3.6 Mass Storage System

The Mass Storage System controls the robots and tape drives associated with storing a file to tape or retrieving a file from tape. Tape drives are assumed to be accessed by only one robot. Each tape drive is assumed to perform either data storage or data retrieval, but not both. This is a conservative approach since system software could dynamically allocate tape drives for either data storage or retrieval during peak periods of On-line or Off-line processing, thus better utilizing the tape drives. The model assumes a full up system in steady state and therefore the number of tape drives are set using the parameters *s_drives_per_robot* and *r_drives_per_robot* given *number_of_robots.*

Retrieves and stores are modelled quite differently. For retrieves, an aisle is randomly selected to simulate the random placement of a tape. A request is queued until a retrieve tape drive on that aisle can be allocated. Once the tape drive is allocated, the request is queued for a robot. When the robot is available, a robotic delay of *robot_delay* seconds is incurred to get the tape and move it to the tape drive. Then a tape load and position delay of *retrv_load_position* seconds is followed by a data transfer delay of *(retrieve_file_size / transfer_rate)* seconds. At this point, the file is now in the MSS Buffer Processor and the timeline continues outside the Mass Storage System. However, in the background the tape is rewound and ejected with a delay of *rewind_eject* seconds and then queued for a robot. When the robot is available, a robotic delay of *robot_delay* seconds is again incurred, this time to move the tape from the tape drive to its bin. Once the tape is back on the shelf, the retrieve tape drive is deallocated.

Store requests typically come at a higher rate, therefore stores are modelled to minimize all resource wait time and usage. Each store request is allocated a store tape drive in a cyclic fashion among store tape drives. Furthermore, robots are only requested when the tape in a store tape drive is full. Most of the time a store tape drive will be allocated with a 0 second wait time with the tape drive being deallocated directly after the data transfer (*archive_file_size / transfer_rate* seconds). The tape stays in the tape drive at its ending position. If the file to be stored will not fit on the tape in the allocated tape drive, then the tape is rewound and ejected with delay *rewind_eject* seconds and moved by robotics from the tape drive to a bin with a delay of *robot_delay* seconds plus any robot queue wait time. The robot then gets a blank tape and takes it to the tape drive for another *robot_delay* seconds. A load and position tape delay of *store_load_position* seconds is followed by a data transfer delay of *(archive_file_size / transfer_rate)* seconds. Now the store tape drive is deallocated and the store process is complete. Store requests assigned tape drives which are already in the process of getting a new tape will incur a considerable tape drive queue wait time.

So that model runs would begin in an environment close to steady state, certain model variables were initialized with values other than 0. All store tape drives were initialized having a tape loaded with the tape fill varying uniformly between 0 and *archive_file_size* across store tape drives. The Level 3 and Pass1 files being aggregated for store were each assigned an initial size varying uniformly from 0 and *archive_file_size* across CPU rows. Since files for Pass 2 processing are not processed until retrieved from the Mass Storage System and all events of a file are processed on the same row, the initial size of all Pass 2 aggregated files is 0.

4.0 Configuration Analysis

4.1 Methodology

The choice of simulation runs were based on three objectives. First, determining the needed number of resources. Second, determining if the configuration was scalable. And third, determining how system efficiency was effected by increasing arrival rates for a given configuration.

Some preliminary runs were made to determine what size configuration would serve as a good baseline. In the process, we learned that execution time of the model would become a limiting factor to the total number of runs that could be made. Simulating one hour of Level 3 events arriving at 3,000 MB per second took 15-16 hours to execute. Although a one hour simulation was not necessary to see if a configuration could handle Level 3 and Pass 1 processing, it was necessary to ensure that the Mass Storage System was keeping up with the load.

The simulation runs are grouped into three categories. The first group of runs were made to investigate the sizing of the Mass Storage System. Mass Storage related parameters were changed while the Level 3 arrival rate was held constant at 3,000 MB per second. The results of these runs are discussed in Sections 4.2.1 and 4.2.2.

The second category of runs were made to investigate Level 3 efficiency, scalability, and CPU and network requirements. Runs were made for four system configurations with the load varying for each configuration. Load refers to the Level 3 arrival rate which also defines the Pass 2 arrival rate. The results of these runs are discussed in Sections 4.2.3 through 4.2.7.

The final category is one run to see how the overall system reacts to changes in load. For a 40 CPU row configuration, the system started with Level 3 events arriving at 4,000 MB per second. After 45 simulated minutes, the arrival rate was dropped to 2,000 MB per second. The results of this run are addressed in Section 4.2.8.

4.2 Simulation Results

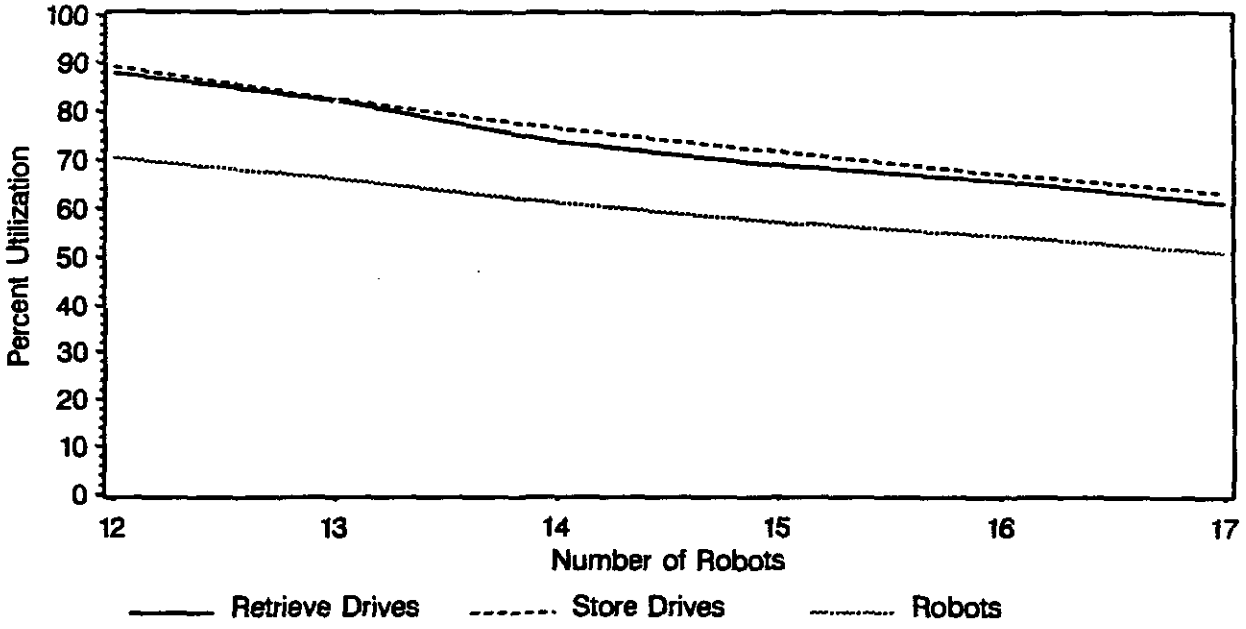
The following sections contain results of simulation runs which were mentioned in the previous section (4.1). The parameter values of all runs are the default values listed in Tables 3.2-I-IV unless otherwise specified.

4.2.1 Varying Number of Robots

Preliminary runs showed that 17 robots provided satisfactory performance with 3 tape drives per robot. However, due to cost constraints, the absolute minimum number of robots was of considerable importance. Simulation runs were made varying the number of robots from 17 down to 12. By nature of the model design, the number of tape drives was also decreased by 3 each time the number of robots was decreased.

As one would expect, the utilization of tape drives and robots increases as the number of robots decreases. This is easily seen in Figure 4.2.1-1. The tape drive utilization of 90% for 12 robots is too high when considering down time for maintenance on the tape drives. The robot utilization is acceptable.

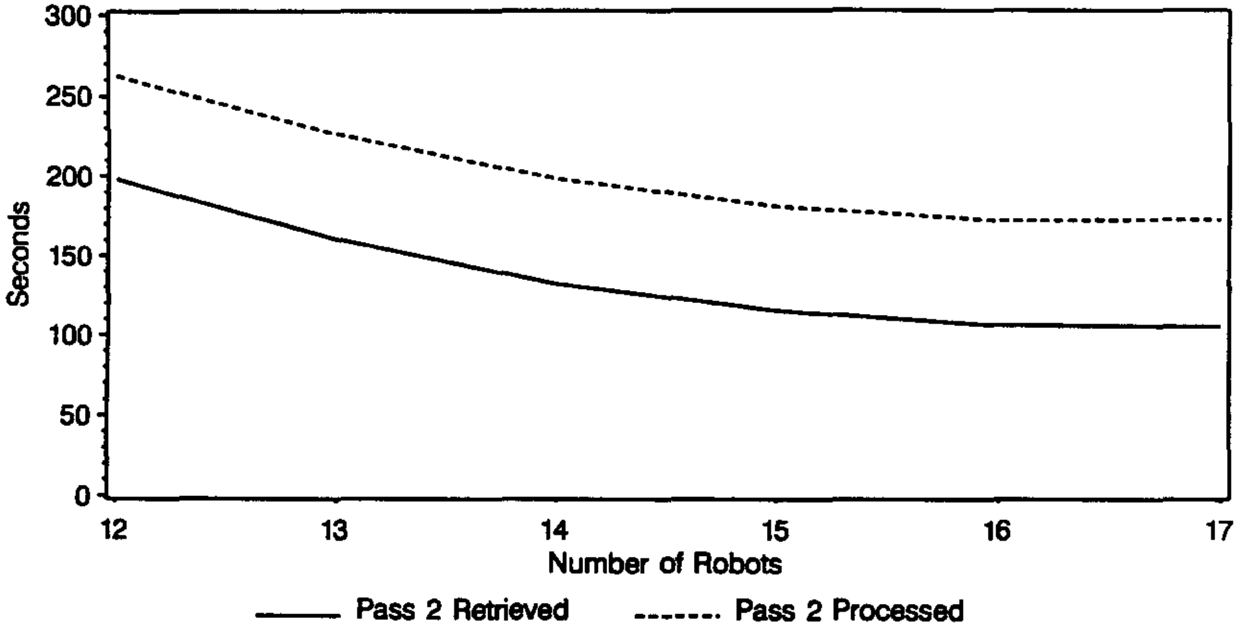
Figure 4.2.1-1 MSS Resource Utilization



NOTE: 3 Drives per Robot (1 for Stores, 2 for Retrieves)

In Figure 4.2.1-2, the delay times for the retrieval of a Pass 2 file are plotted. Beginning with 14 robots and down, the delay times are beginning to curve upward. Note, however, that once the file has been retrieved, the average time to process each event in the file (difference between the two lines) is constant.

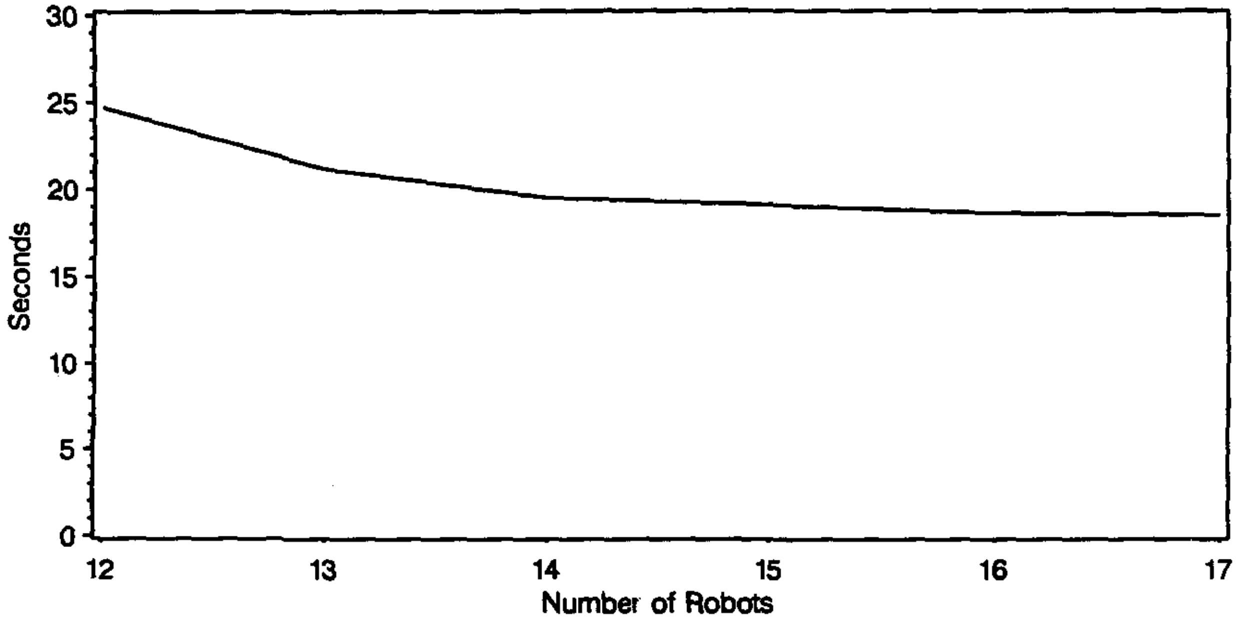
Figure 4.2.1-2 Pass 2 Completion Times



NOTE: 3 Drives per Robot (1 for Stores, 2 for Retrieves)

Figure 4.2.1-3 shows the average time to store a file to tape from the MSS Buffer Processor. These times are considerably *faster* than the retrieve times as only about 14 of 2200 store requests per hour require a robot. Similar to the retrieve tape drives, store tape drives are becoming limited with 13 and 12 robots as seen with the time delays bending upward.

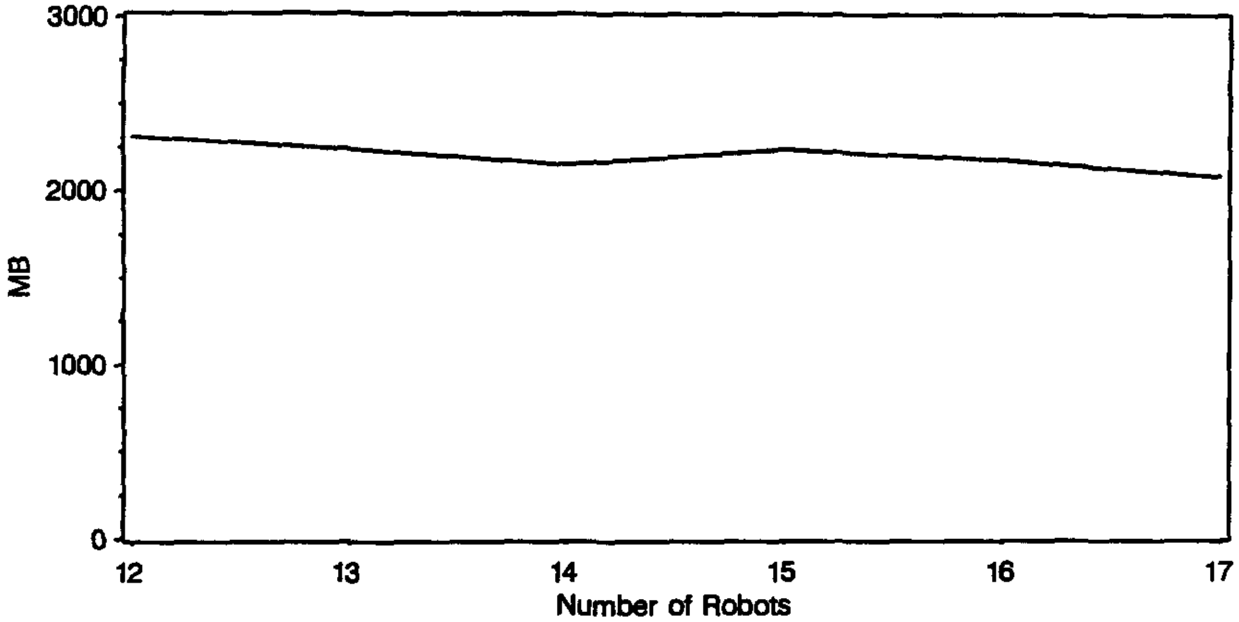
Figure 4.2.1-3 MSS Store Delay Times



NOTE: 3 Drives per Robot (1 for Stores, 2 for Retrieves)

Varying the number of robots and tape drives had very little impact on the performance of other aspects of the system, given the constant load. Only the Mass Storage System's performance was affected. Network and CPU utilization did not change. Figure 4.2.1-4 shows the maximum fill of an MSS Buffer Processor for each of the runs. Although there is some slight variation, it is not significant. The MSS Buffer fill is not expected to grow higher until there exists a severe shortage of store tape drives.

Figure 4.2.1-4 Maximum MSS Buffer Fill (per processor)



NOTE: 3 Drives per Robot (1 for Stores, 2 for Retrieves)

4.2.2 Adequate Retrieve Tape Drives

As noted in the previous section, retrieve tape drives are a scarce resource when using 12 robots. To gain insight on the number of tape drives needed for good system performance, additional runs with 3 retrieve tape drives per robot were made. Figures 4.2.2-1 and 4.2.2-2 reflect the differences in Mass Storage System related times for retrieves and stores respectively. These runs used 40 CPU rows as opposed to the 32 row configuration used in the previous section, but kept the same number of CPUs.

Figure 4.2.2-1 Retrieve Delay Times

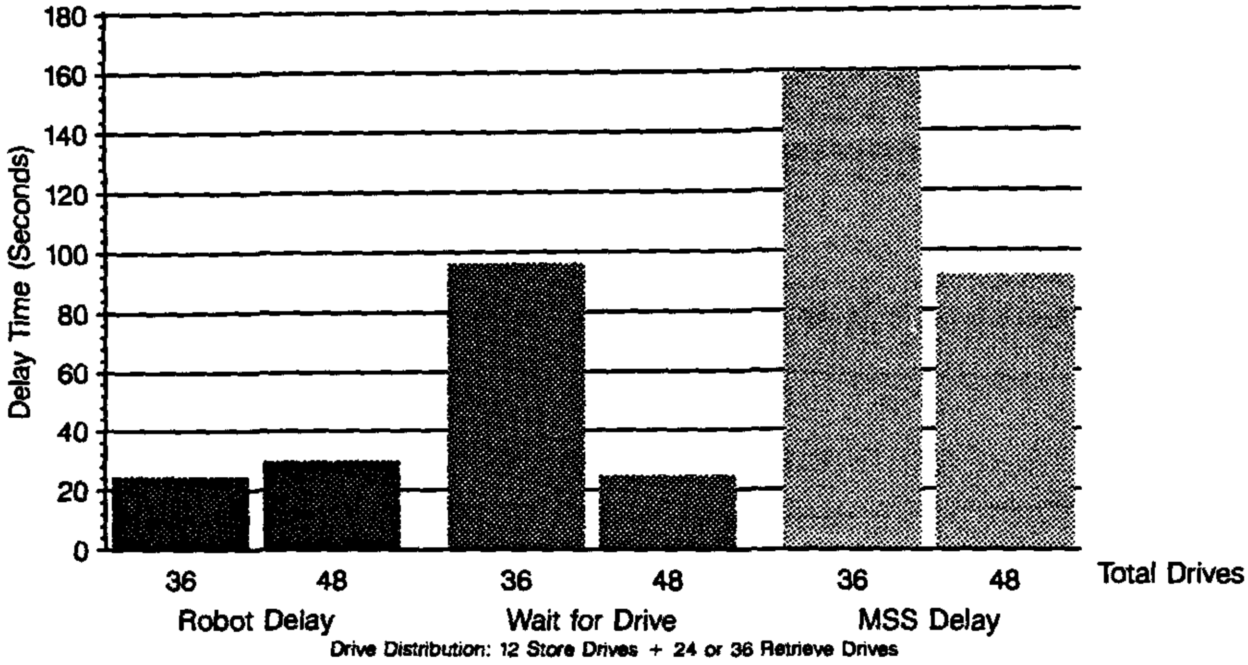
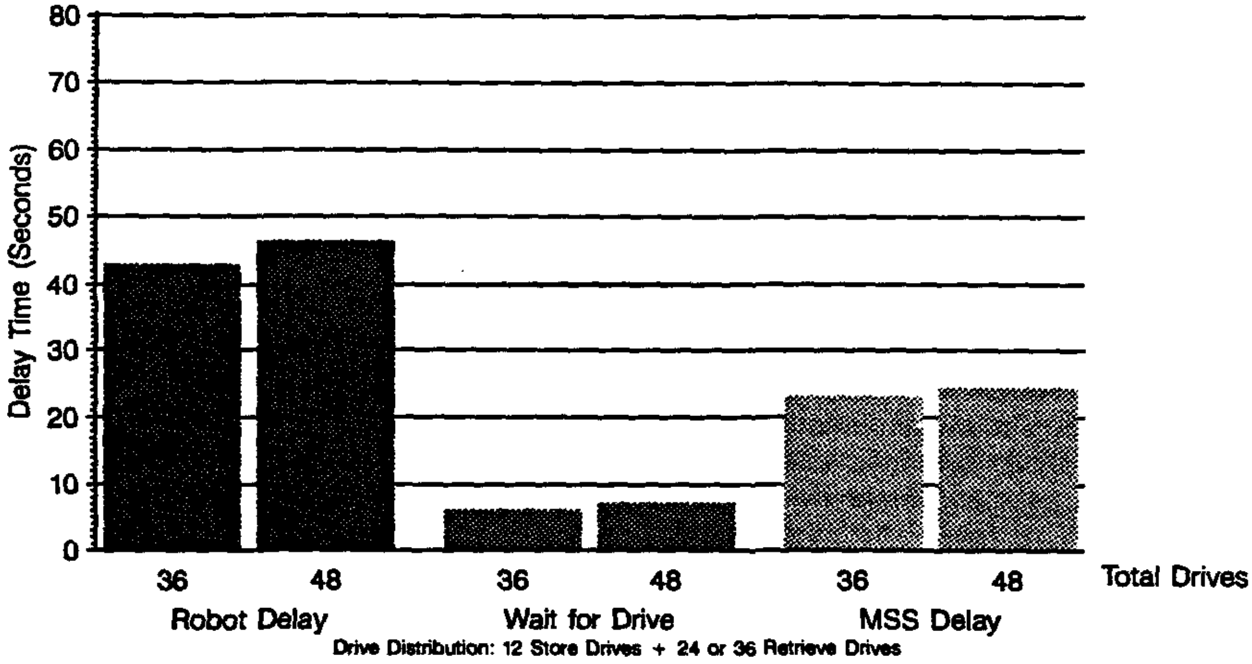


Figure 4.2.2-2 Store Delay Times



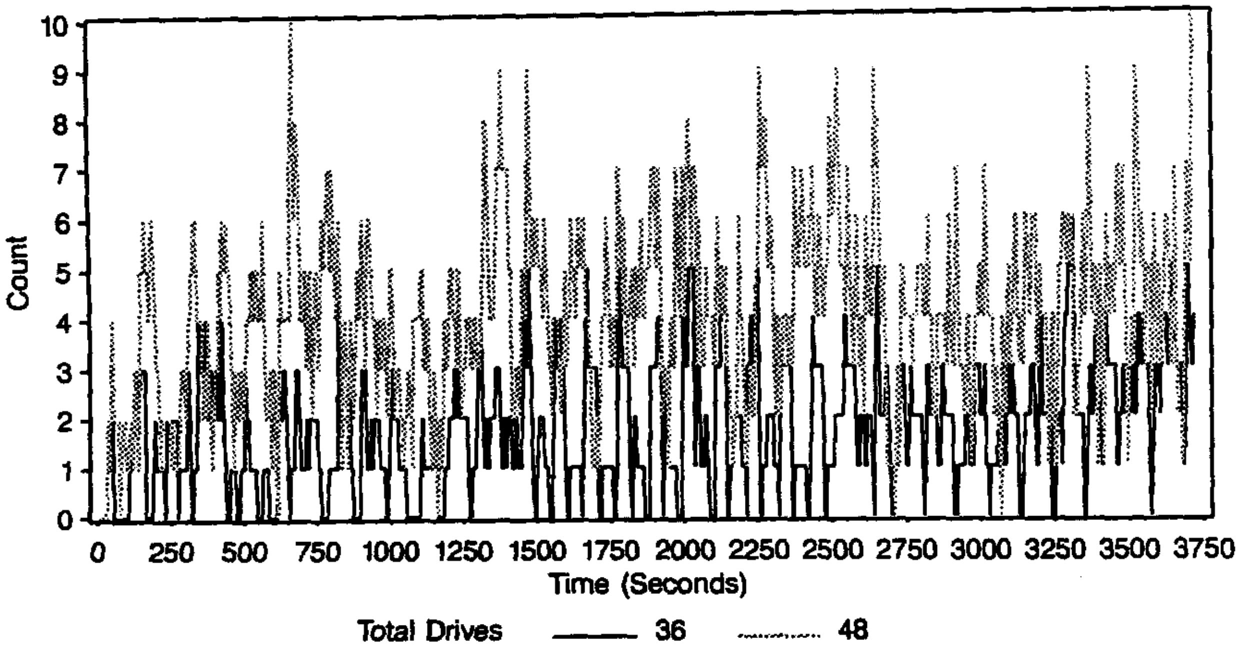
For retrieves, the tape drive queue wait time has been reduced considerably with the extra tape drive per robot. The total time through the Mass Storage System has dropped by almost 50%. Note however the robot delay has slightly increased. This is due to a higher utilization of robots. The robots are doing more work since more tape drives are available.

The addition of the retrieve tape drives has had a slightly adverse effect on store times. The robot time, tape drive queue time, and total MSS time have all slightly increased. This is all attributed to the higher robot utilization. On the few occasions when a robot is requested, the wait time is now higher and thus affects overall store times. Do not be concerned that the robot delay for stores is twice that of the delay for retrieves. Remember that the model design for stores requests robots only when a tape is full and incurs both a dismount and a mount, whereas retrieves only incur a mount.

To emphasize the need for more retrieve tape drives and the adequacy of 12 robots, Figures 4.2.2-3 and 4.2.2-4 on page 24 show a comparison of resource queues over time, first for robots and then for retrieve tape drives. The comparison of robot queues does infer a higher utilization of robots with 3 retrieve tape drives per robot since the queue length is higher. Although the queue is higher, it still appears to have some stability which implies that 12 robots should be satisfactory. However, there is a considerable difference between the retrieve tape drive queues. The system with 3 retrieve tape drives per robot is quite stable. However, the fewer tape drive system seems to be getting only further and further behind.

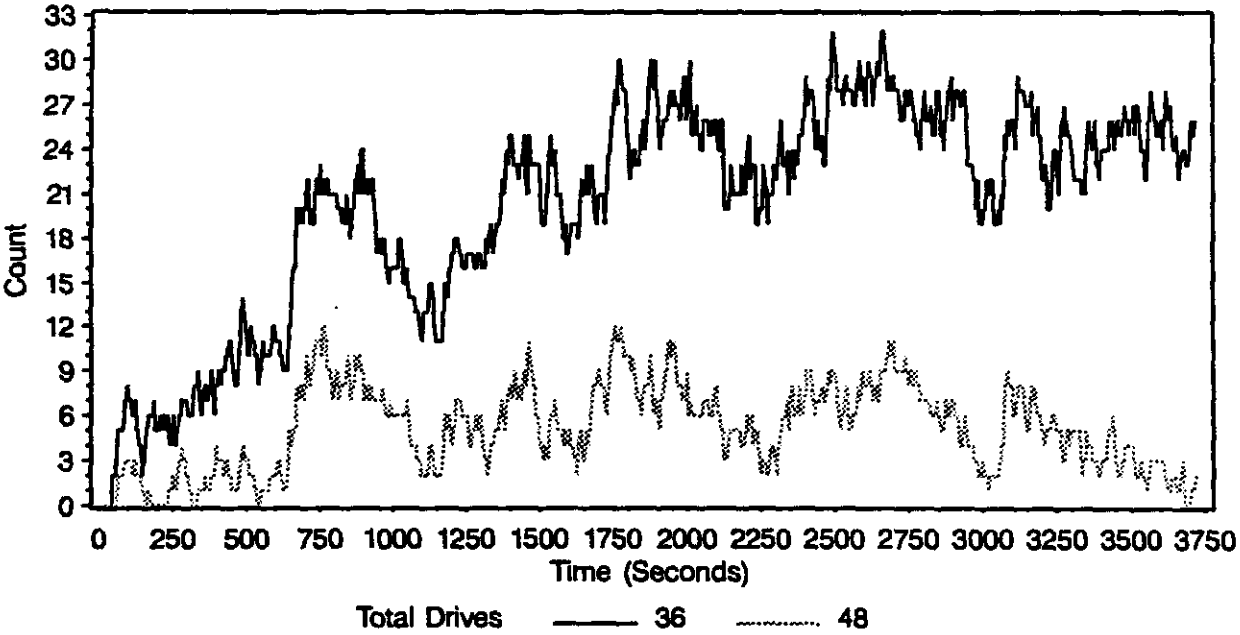
Model design did not allow for a different number of tape drives on each aisle. However, some insight to the number needed can be gained from snapshot data that was collected. Over time the maximum number of tape drives in use for the 36 retrieve tape drive system was 30. Therefore, 30 retrieve tape drives would be a suggested minimum number.

Figure 4.2.2-3 Comparison of Robot Queues (12 Robots)



Drive Distribution: 12 Store Drives + 24 or 36 Retrieve Drives

Figure 4.2.2-4 Comparison of Retrieve Tape Drive Queues (12 Robots)



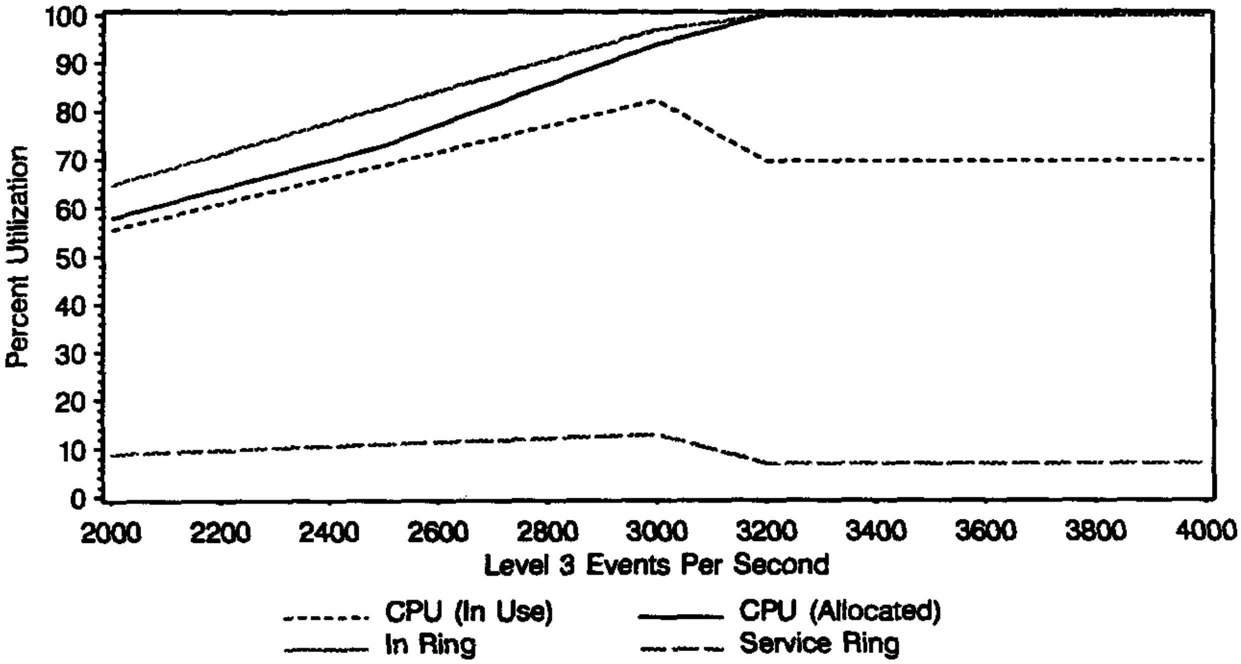
Drive Distribution: 12 Store Drives + 24 or 36 Retrieve Drives

4.2.3 32 CPU Row Configuration

The main purpose of the runs in this section was to investigate system performance as a function of load. Again, load refers to the Level 3 arrival rate which also defines the Pass 2 arrival rate. These runs were based on the default parameter values with loads varying from 2,000 MB/s to 4,000 MB/s.

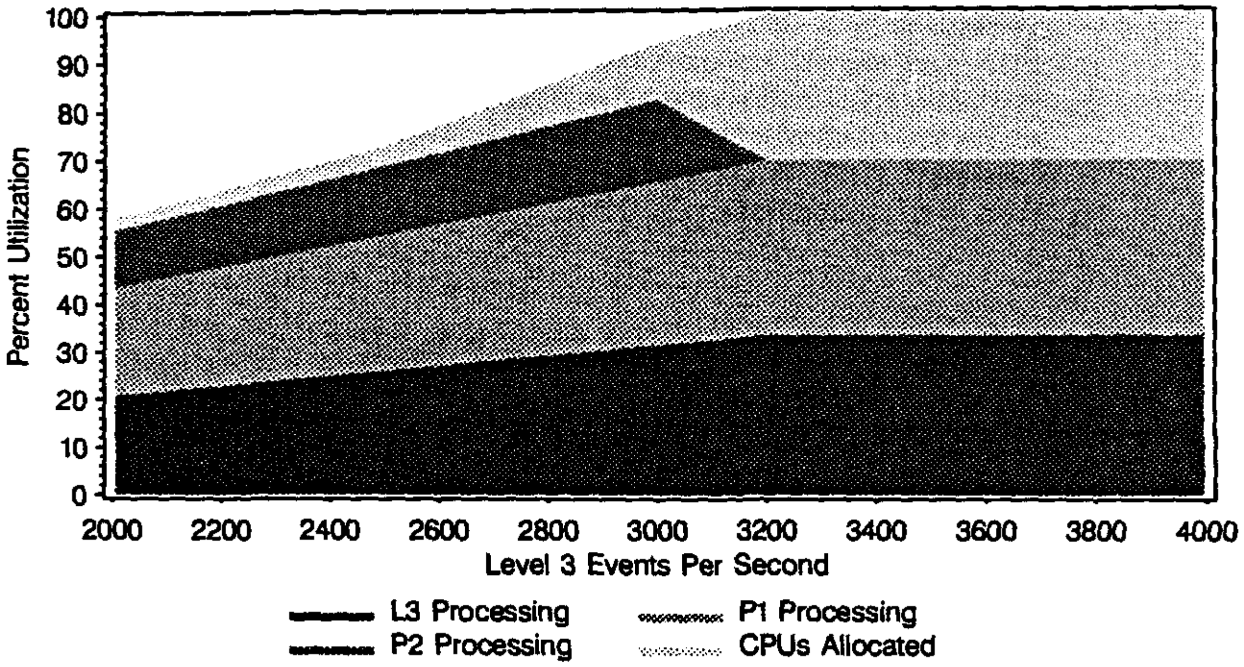
CPU and network utilization is reflected in Figure 4.2.3-1. Note that the percentage of CPUs in use is always smaller than the percentage of CPUs allocated. This is due to the time for an event to travel over the In Ring to a CPU after allocation. As the In Ring utilization increases, the effective rate is decreasing. This explains why the gap between the CPU In Use and CPU Allocated lines is expanding until the In Ring Utilization peaks at 100%.

Figure 4.2.3-1 CPU and Network Utilization



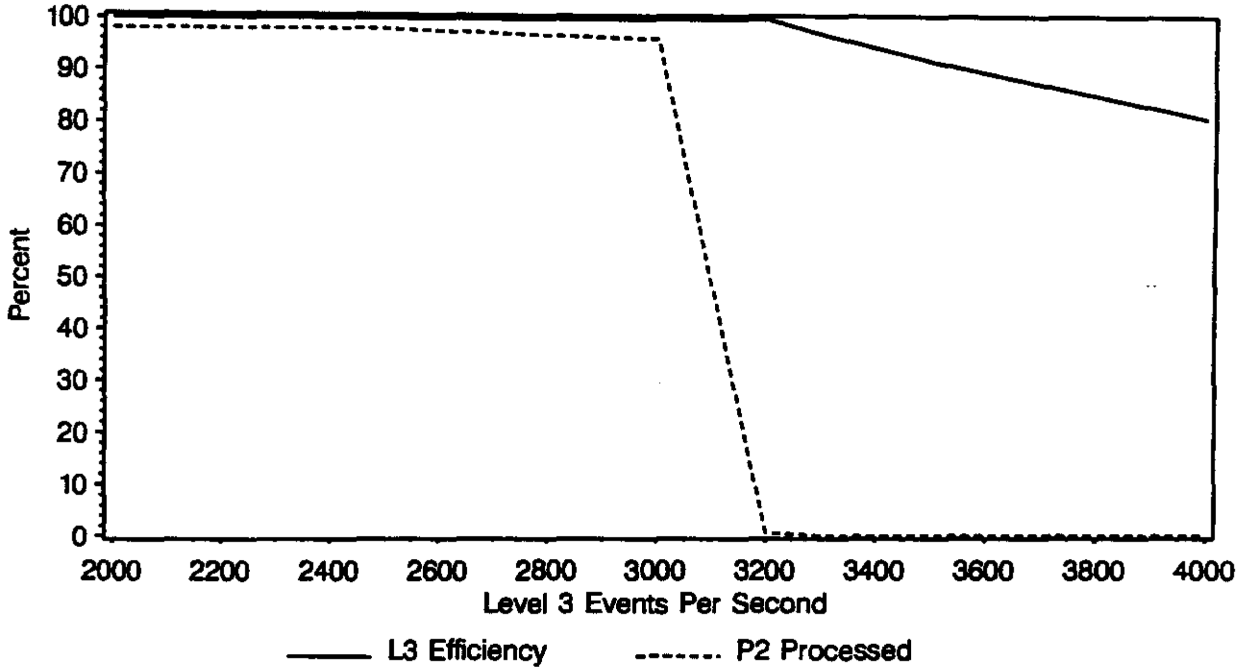
The Service Ring utilization increases only slightly up to a load of 3,000 MB/s and then suddenly decreases. The reason for this decrease can be seen in Figure 4.2.3-2 on page 26. Beginning with a load of 3,000 MB/s, Pass 2 Processing is being done less and less up to a load of 3200 MB per second where Pass 2 has been completely turned off. With Pass 2 turned off, the only traffic occurring on the Service Rings is between the CPUs and the MSS Buffer Processors. Once Pass 2 is turned off, the Service Ring utilization again continues increasing slightly with load.

Figure 4.2.3-2 CPU Utilization Subdivided by Event Processing



Note also from Figure 4.2.3-2 that almost 40% of the time, CPUs are waiting for events to get to them when loads are above 3200 MB/s. The amount of Level 3 and Pass 1 processing also levels off at this point. The reason for this levelling off can be seen in Figure 4.2.3-3 on page 27 where the Level 3 efficiency is no longer at 100% and dropping steadily. The percentage of Pass 2 data being processed given the load is also shown in this chart and again one can see that Pass 2 processing has been turned off with loads exceeding 3200 MB/s. Pass 2 processing begins to be quickly disabled when the load reaches 3,000 MB/s.

Figure 4.2.3-3 Level 3 Efficiency and Pass 2 Completion Rate



The tape drive and robot utilization increases as load increases as seen in Figure 4.2.3-4 on page 28. The point in which store tape drive utilization begins to fall off matches the same point in which Pass 2 processing becomes throttled.

The decreasing In Ring effective rate mentioned earlier can be seen in Figure 4.2.3-5 on page 28. With loads of 3200 MB/s and upward, the In Ring effective rate is constant. The Control Processor Buffers can no longer handle the incoming data rate and efficiency is therefore dropping. This leads to a constant input rate from the Control Processors to the CPUs. This same graph reflects the maximum amount of buffer fill in the Control and MSS Buffer Processors. Note the Control Processor's buffer will not exceed the parameter value for maximum Control Buffer size. The right vertical axis is associated with these lines. The fill increase in the MSS Buffer Processors is only slight until the amount of Pass 2 processing begins to be cut off. Then a sharp transition period takes place until Pass 2 is totally shut down. At this point the buffer fill again continues to increase at a slow rate relative to load as it did while Pass 2 was processing at 100%.

Figure 4.2.3-4 MSS Resource Utilization

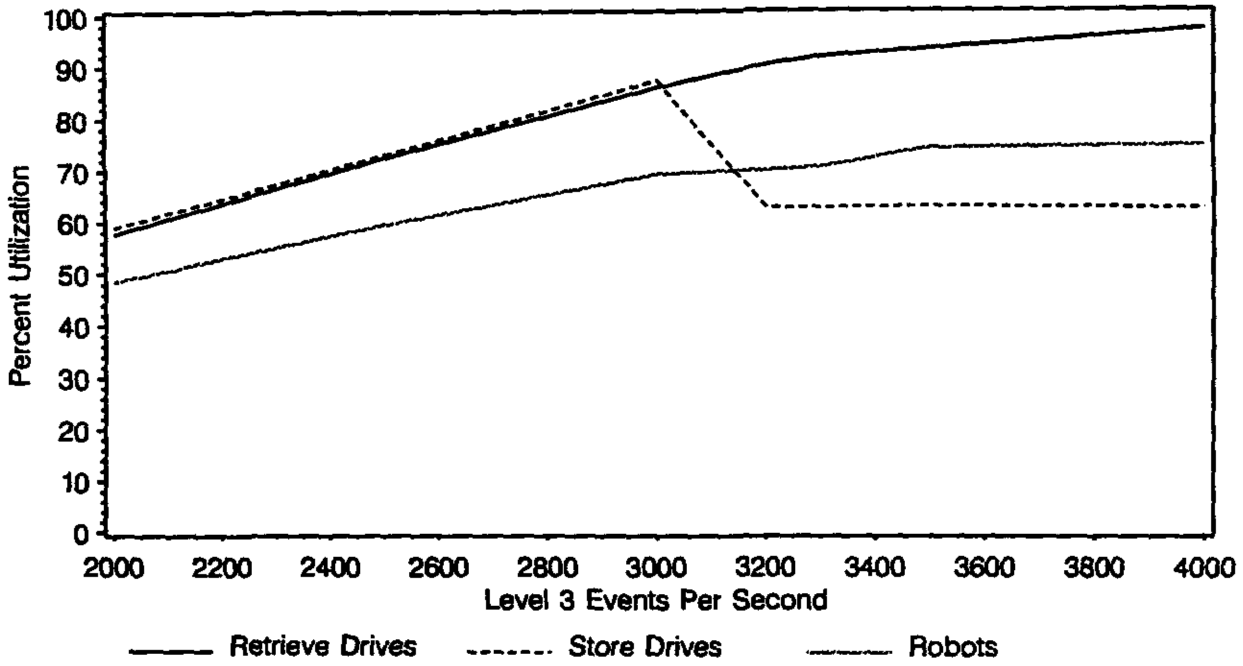
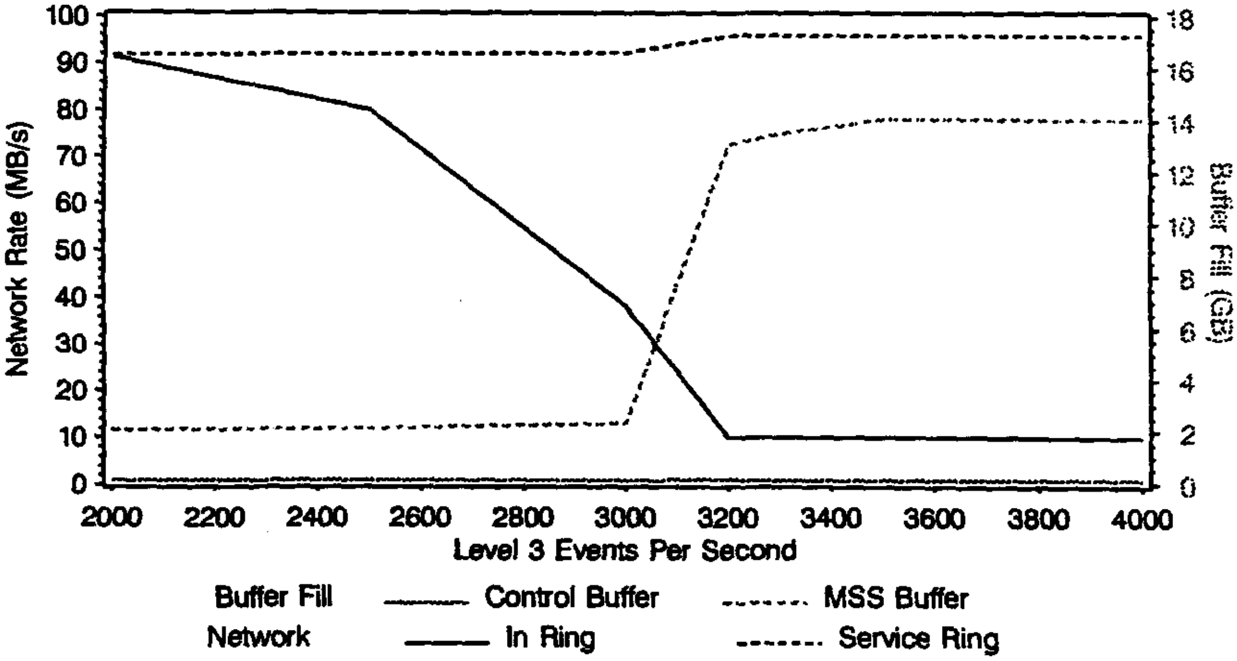


Figure 4.2.3-5 Network Rates and Buffer Fill

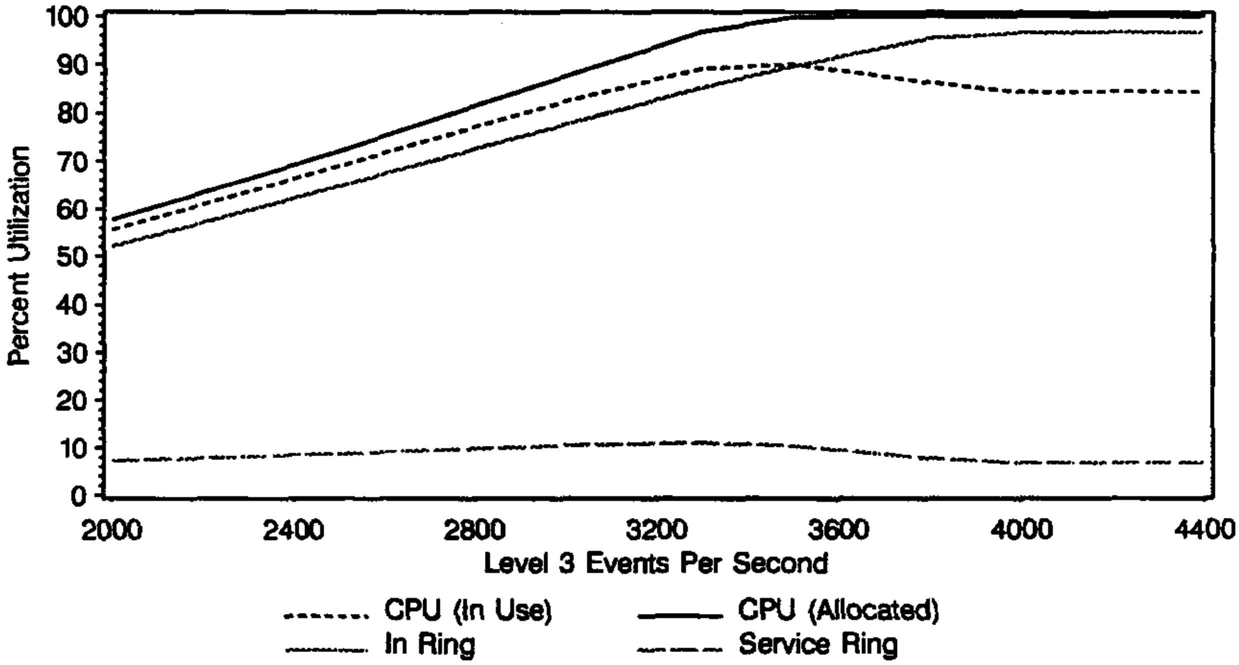


4.2.4 40 CPU Row Configuration

The system configuration for this set of runs consisted of 40 CPU rows with 32 CPUs per row giving a total of 1280 CPUs. Compared to the previous set of runs, 8 CPU rows were added while keeping the total number of CPUs constant.

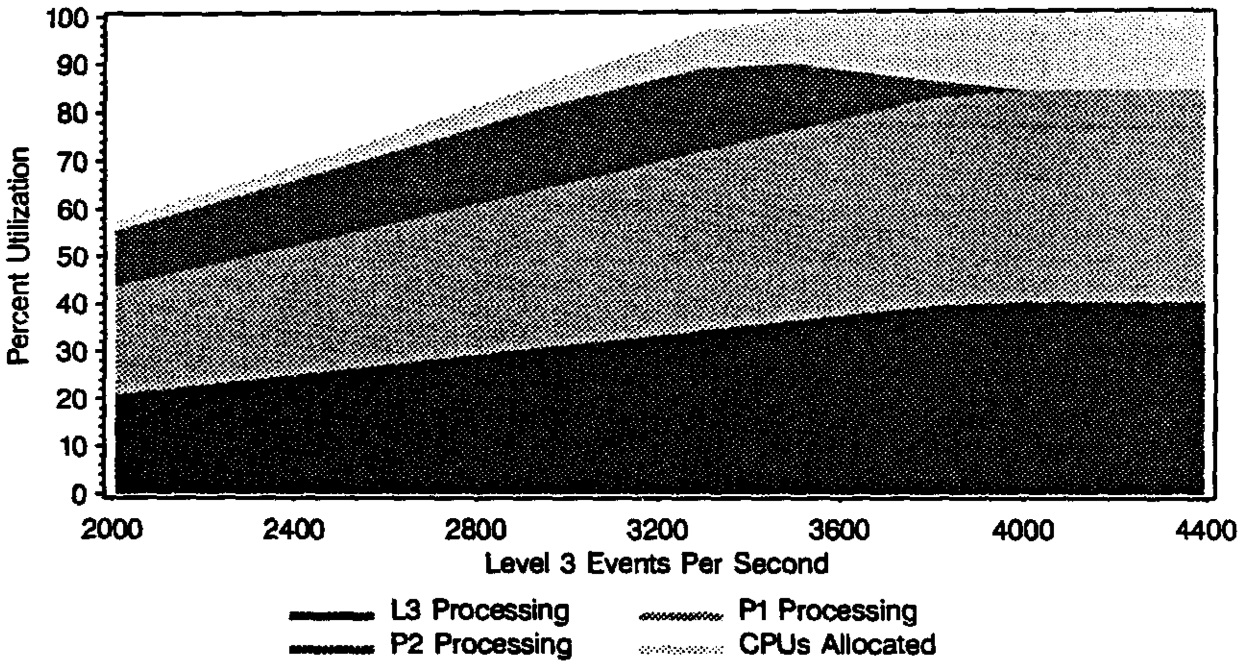
In general, the trends in utilization were very similar to those of the 32 row configuration. However, performance was considerably improved. Figure 4.2.4-1 shows the 40 CPU row configuration handling a much higher load. The number of CPUs has become the limiting factor, and not the In Ring.

Figure 4.2.4-1 CPU and Network Utilization



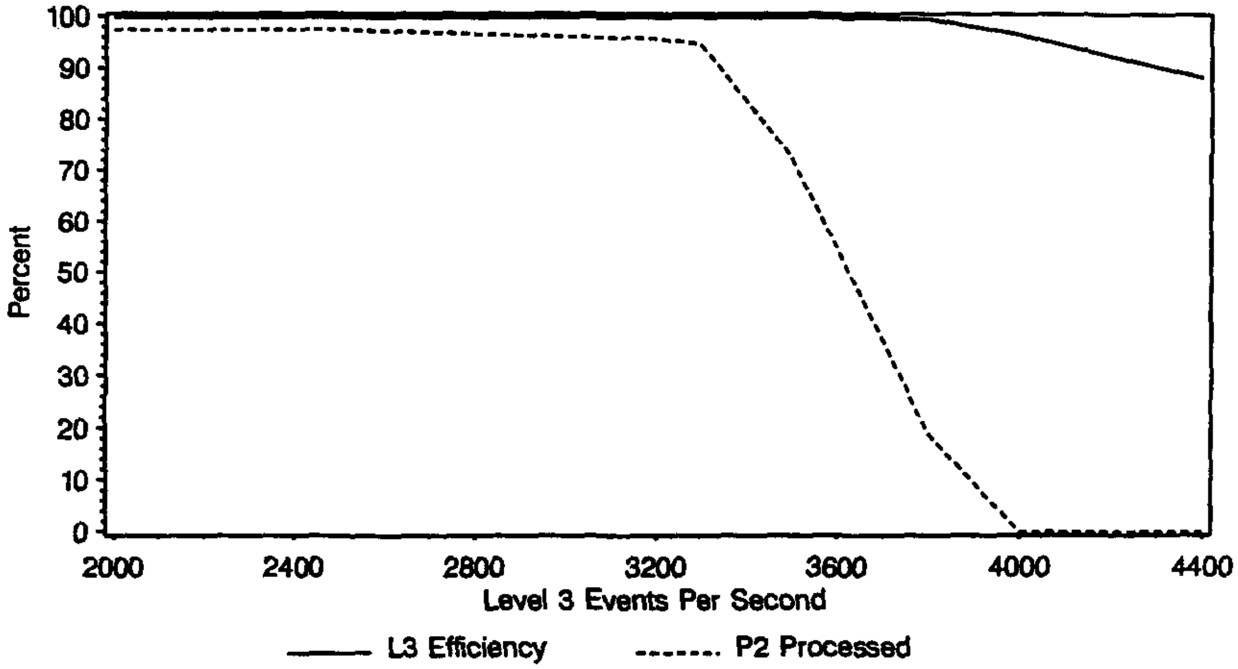
The Service Ring utilization now increases slightly up to a load of 3500 MB/s and then decreases until the load reaches 4,000 MB/s. By looking at Figure 4.2.4-2, Pass 2 processing is not stopped until the load of 4,000 MB/s is reached. It appears this configuration can handle a load of 3200 MB/s before Pass 2 begins to be throttled. Furthermore, the CPUs are much better utilized. With a load of 3200 MB/s, actual CPU usage peaks at about 90% and is at least 84% for higher loads. This is about 15% better than the 32 row configuration.

Figure 4.2.4-2 CPU Utilization Subdivided by Event Processing



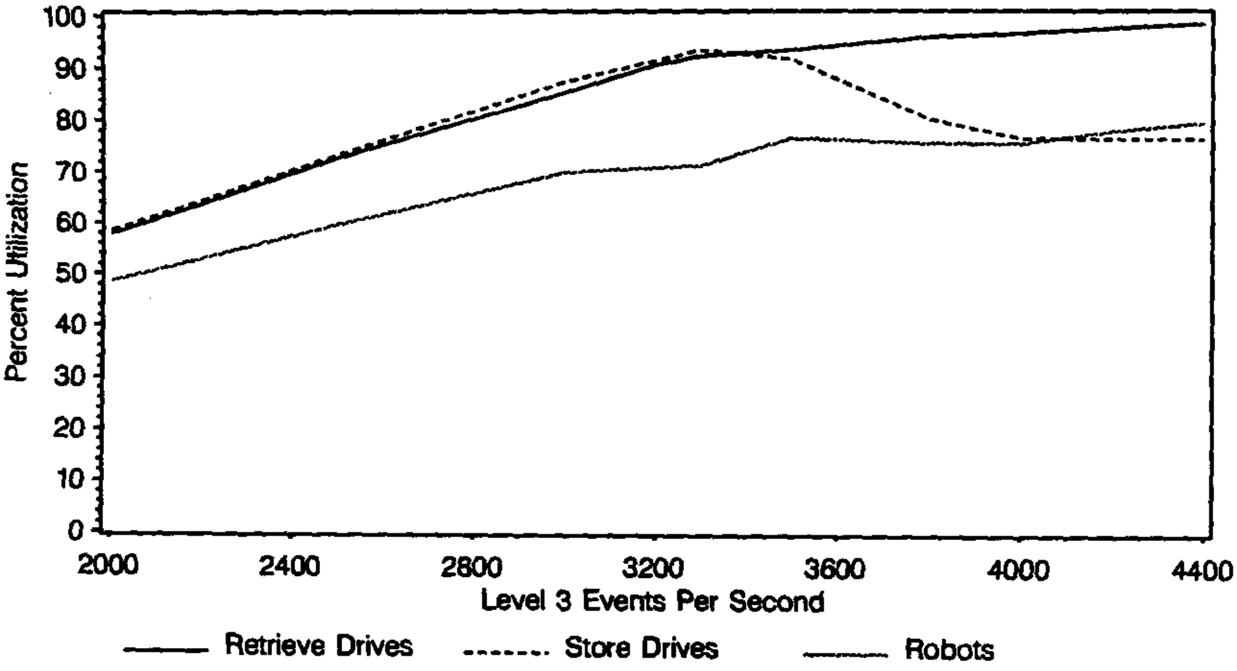
The percentage of Pass 2 data being processed given the load is shown in Figure 4.2.4-3 and Pass 2 processing is clearly ended with loads exceeding 4,000 MB/s. The load range in which Pass 2 Processing transitions to 0 has increased to 800 MB/s. With a load of 3800 MB/s, the Level 3 efficiency is beginning to drop. Note this is before Pass 2 is completely turned off.

Figure 4.2.4-3 Level 3 Efficiency and Pass 2 Completion Rate



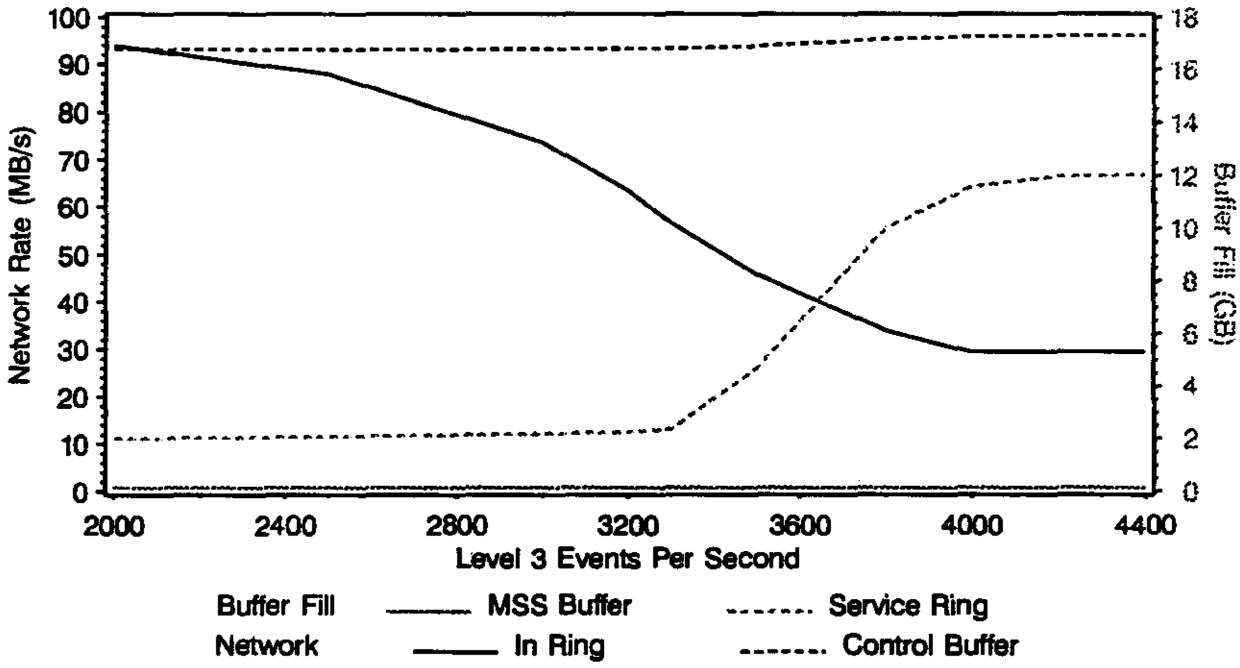
Similar to the 32 row configuration, the tape drive and robot utilization increases as load increases as seen in Figure 4.2.4-4. The point in which store tape drive utilization begins to fall off matches the same point in which Pass 2 processing becomes throttled.

Figure 4.2.4-4 MSS Resource Utilization



The In Ring effective rate, as seen in Figure 4.2.4-5, levels at about 30 MB/s for loads of 4,000 MB/s and upward. This effective rate is higher than the 32 row configuration's In Rings. This same graph reflects the maximum amount of buffer fill in a Control and MSS Buffer Processor. The right vertical axis is associated with these lines. As with the previous configuration, the fill increase in the MSS Buffer Processors is only slight until the amount of Pass 2 processing begins to be cut off. This time a smoother transition period takes place until Pass 2 is totally shut down. Note that 2 less gigabytes per MSS Buffer Processor were required (12 GB vs. 14 GB). This can be misleading since the total MSS Buffer fill is higher for the 40 row configuration (12x40=480 > 14x32=448). However, while both systems were at 100% efficiency and Pass 2 at full rate, only 2.2 GB were required.

Figure 4.2.4-5 Network Rates and Buffer Fill



4.2.5 24 CPU Row Configuration

The system configuration for this set of runs consisted of 24 CPU rows with 40 CPUs per row giving a total of 960 CPUs. This set of runs is essentially the 32 row configuration sized down to 24 rows.

When comparing the next five graphs with the respective ones in Section 4.2.3, one can see that the 32 row configuration and the 24 row configuration operate almost identically. The key difference is in the x-axis. Whereas the 32 row configuration could handle a load up to 3200 MB/s before completely stopping Pass 2 processing, the 24 row configuration can only handle 2400 MB/s. Another minor difference is that the load range in which Pass 2 processing transitions to 0 is smaller for the 24 row configuration.

Figure 4.2.5-1 CPU and Network Utilization

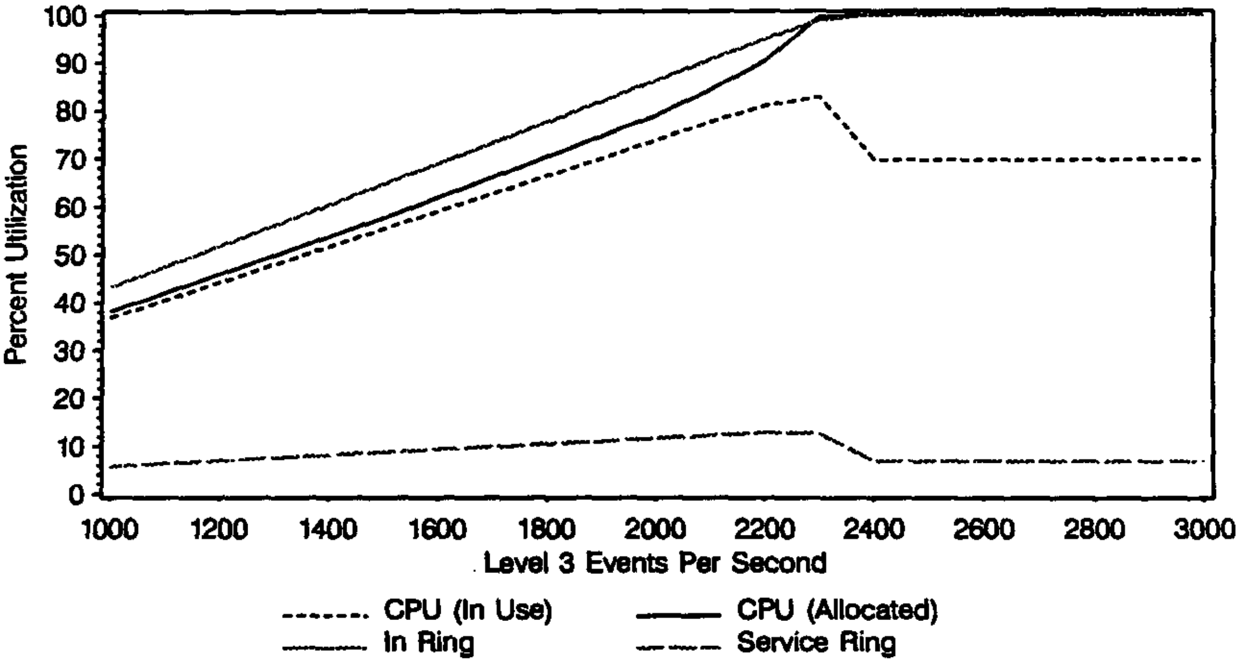


Figure 4.2.5-2 CPU Utilization Subdivided by Event Processing

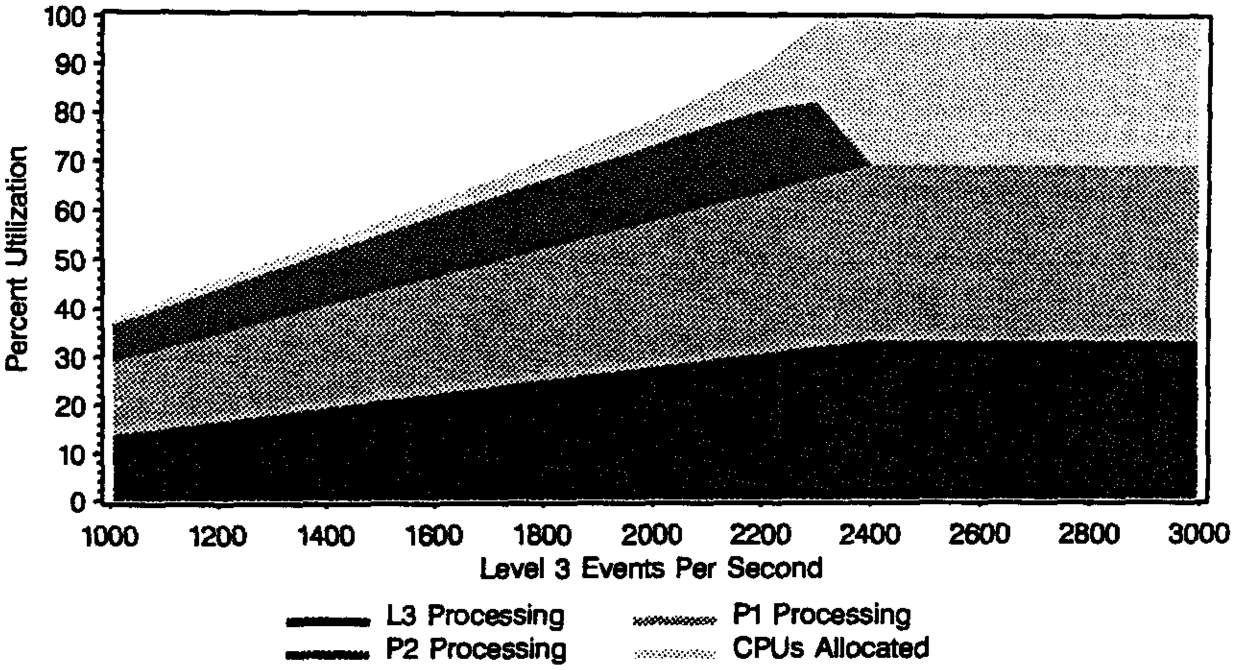


Figure 4.2.5-3 Level 3 Efficiency and Pass 2 Completion Rate

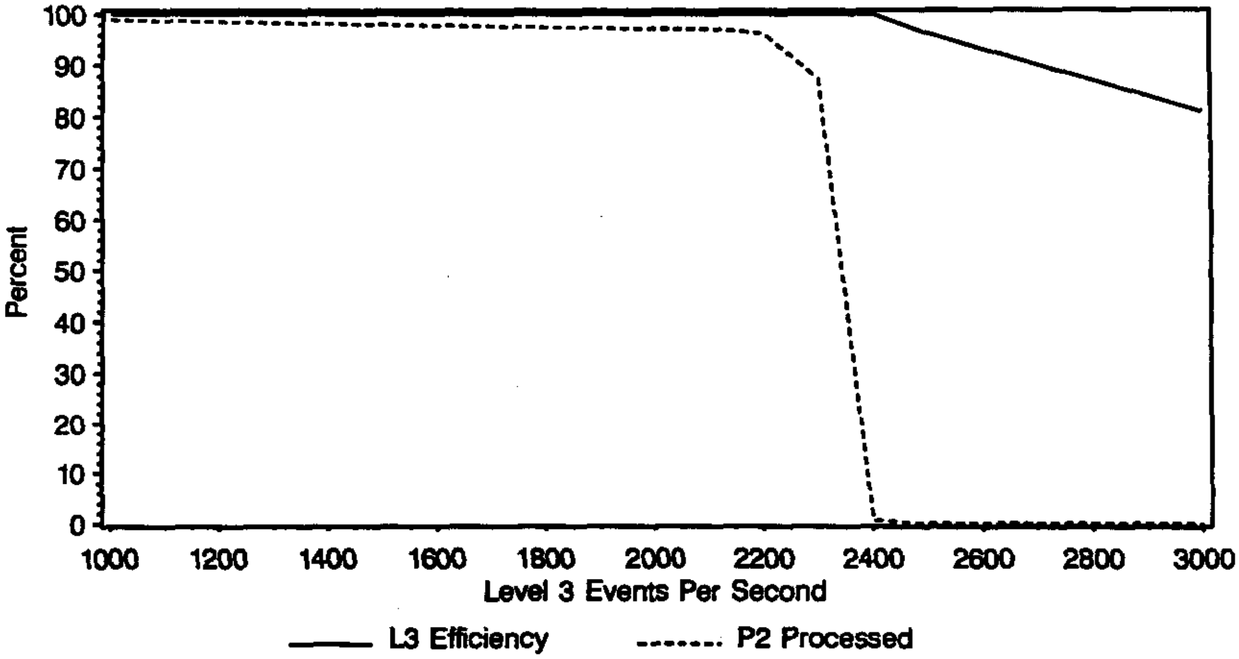


Figure 4.2.5-4 MSS Resource Utilization

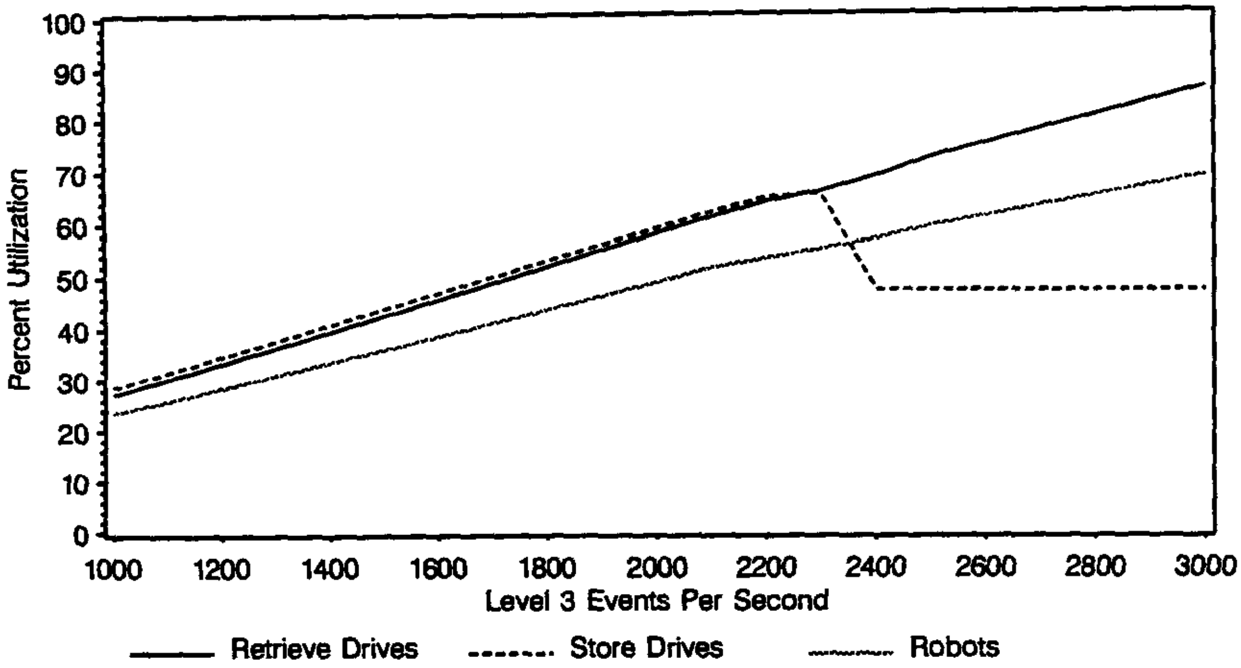
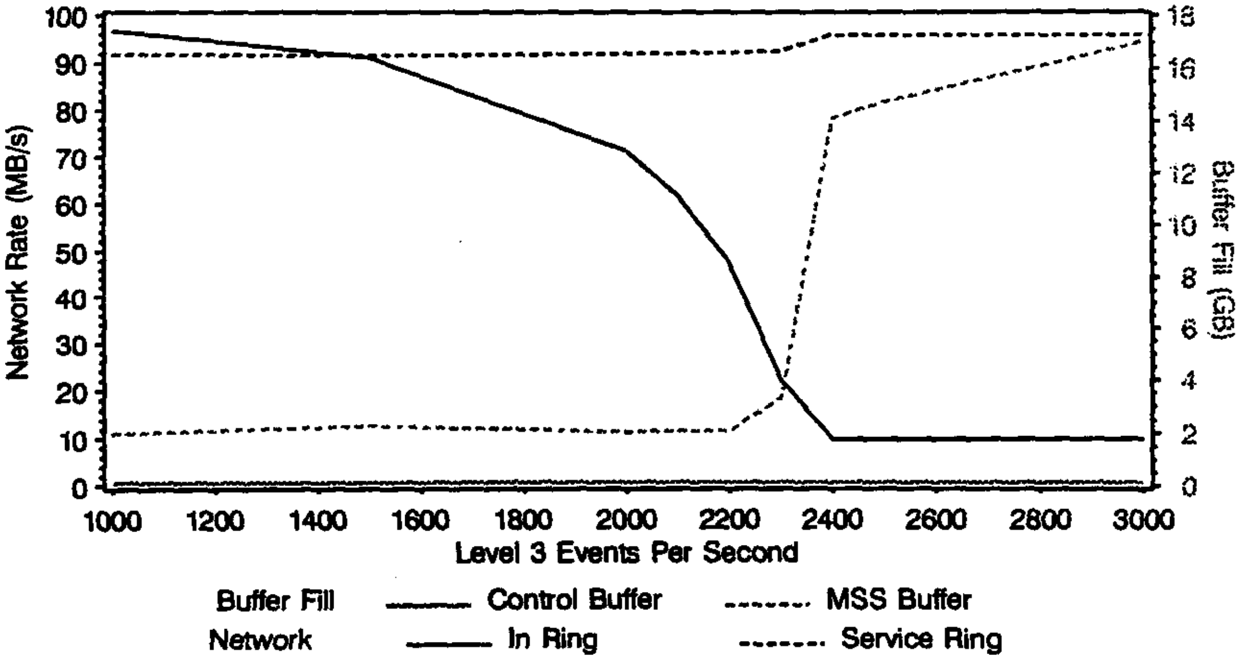


Figure 4.2.5-5 Network Rates and Buffer Fill



4.2.6 32 CPU Rows (Level 3 Processing Time Doubled)

The results of this section reflect what would happen if the Level 3 processing time was doubled. The 32 row configuration from Section 4.2.3 was used and the parameter *l3_processing* was changed from 67 SSCUPs to 134 SSCUPs.

Note in Figure 4.2.6-1 that CPU utilization (allocated) has all but reached 100% with a load of 2500 MB/s. This is 700 MB/s lower than when the Level 3 processing time was not doubled. CPUs are now the limiting resource rather than the In Ring networks.

Figure 4.2.6-1 CPU and Network Utilization

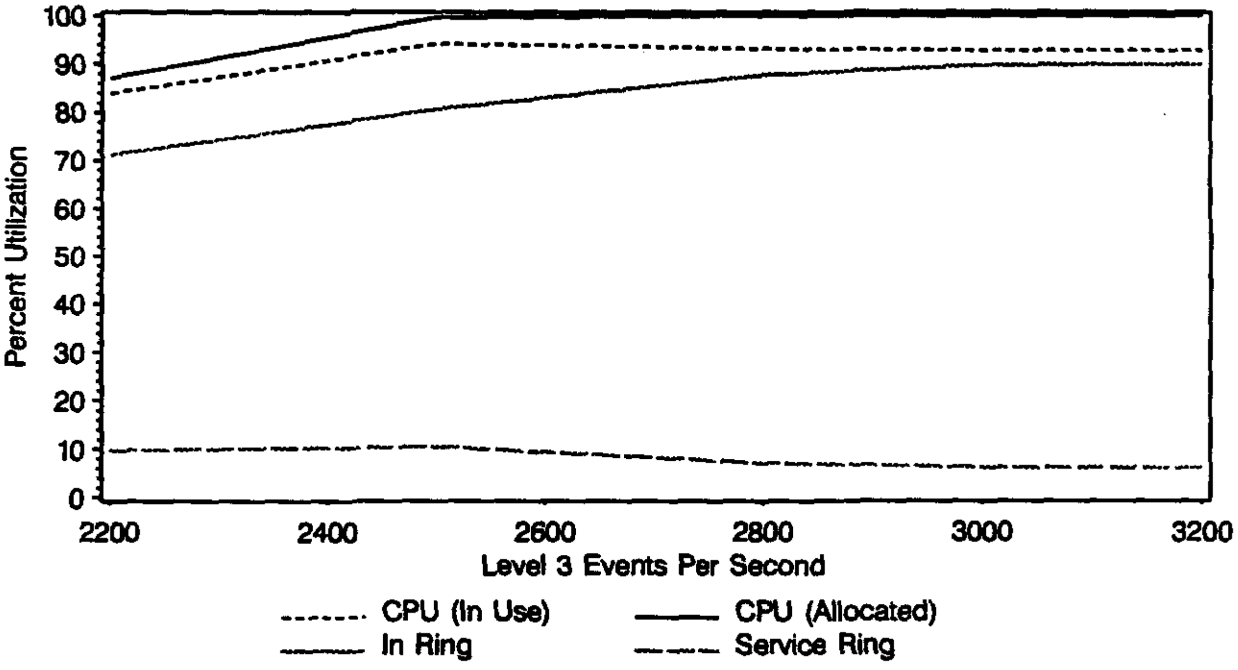


Figure 4.2.6-2 shows over 50% of the processing being done is now for Level 3. Unlike the previous sets of runs, Pass 2 processing is being throttled very slowly, beginning with a load of 2500 MB/s and not completely stopped until the load reaches 3,000 MB/s.

By looking at Figure 4.2.6-3, Pass 2 processing appears to be moving downward with even a load of 2200 MB/s. The Level 3 efficiency has dropped to 99.7% with a load of 2800 MB/s and continues to fall off from there.

Figure 4.2.6-2 CPU Utilization Subdivided by Event Processing

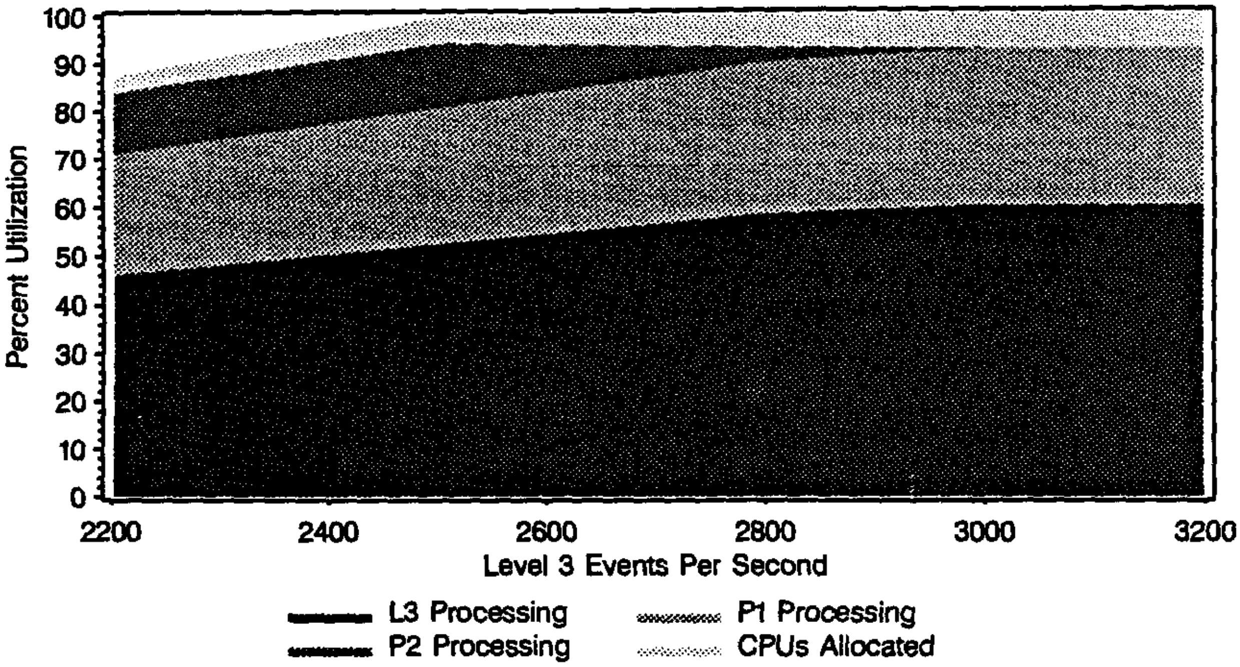
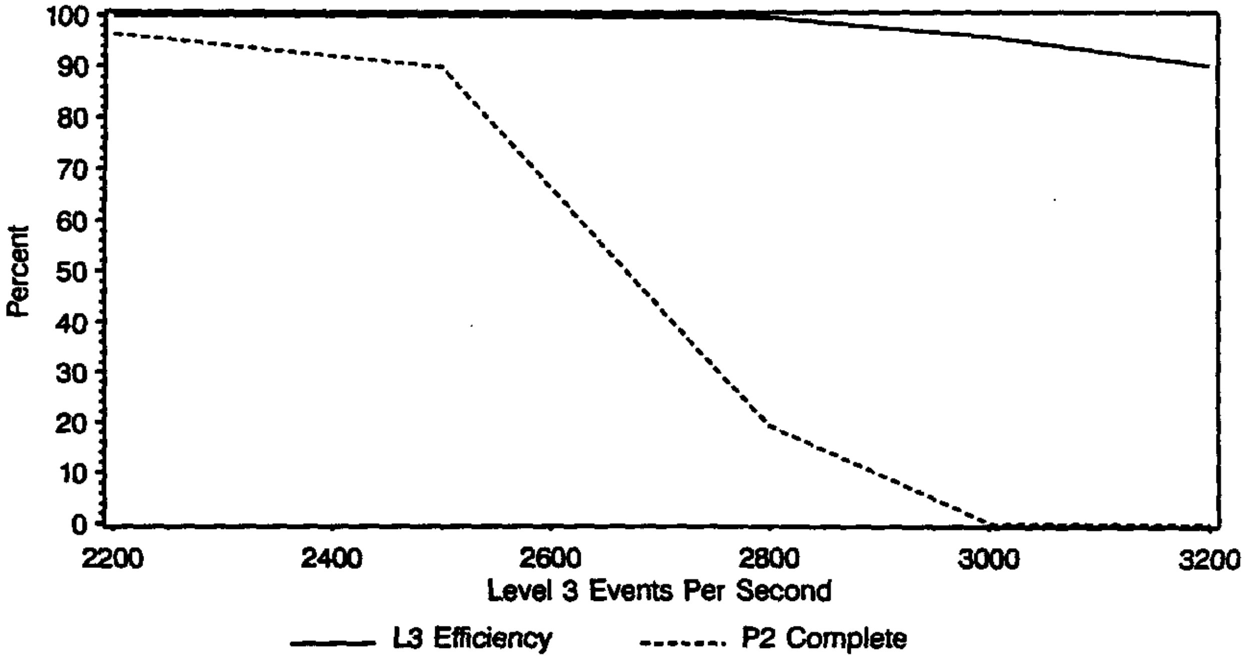
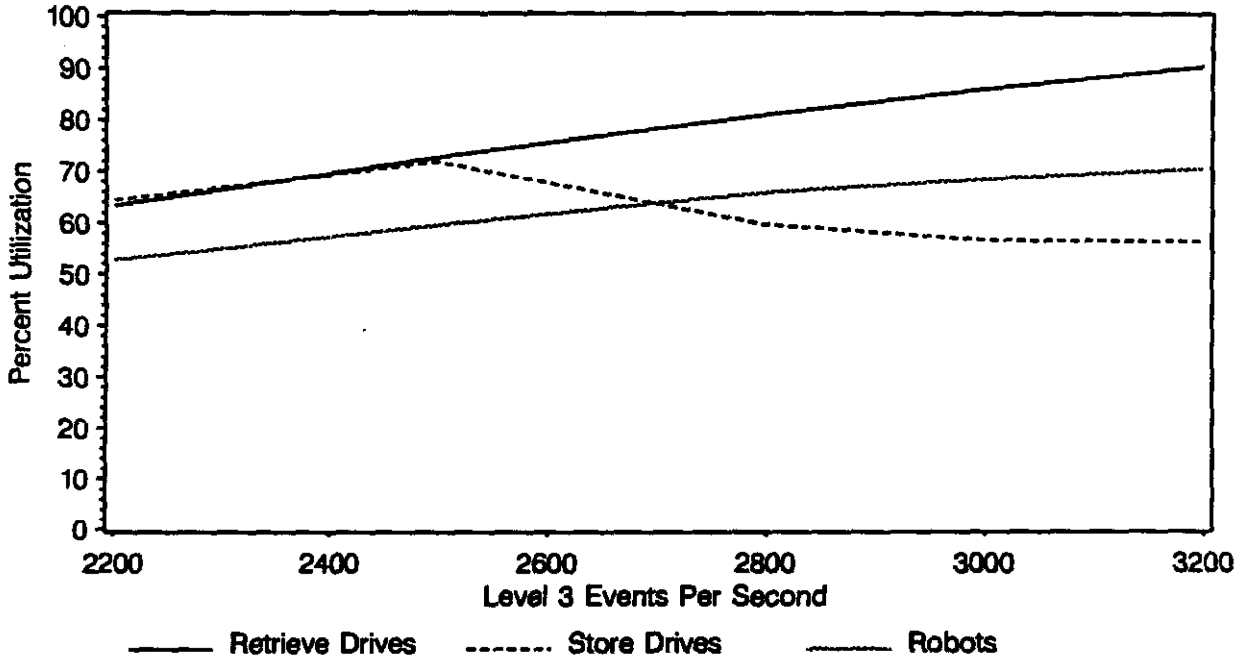


Figure 4.2.6-3 Level 3 Efficiency and Pass 2 Completion Rate



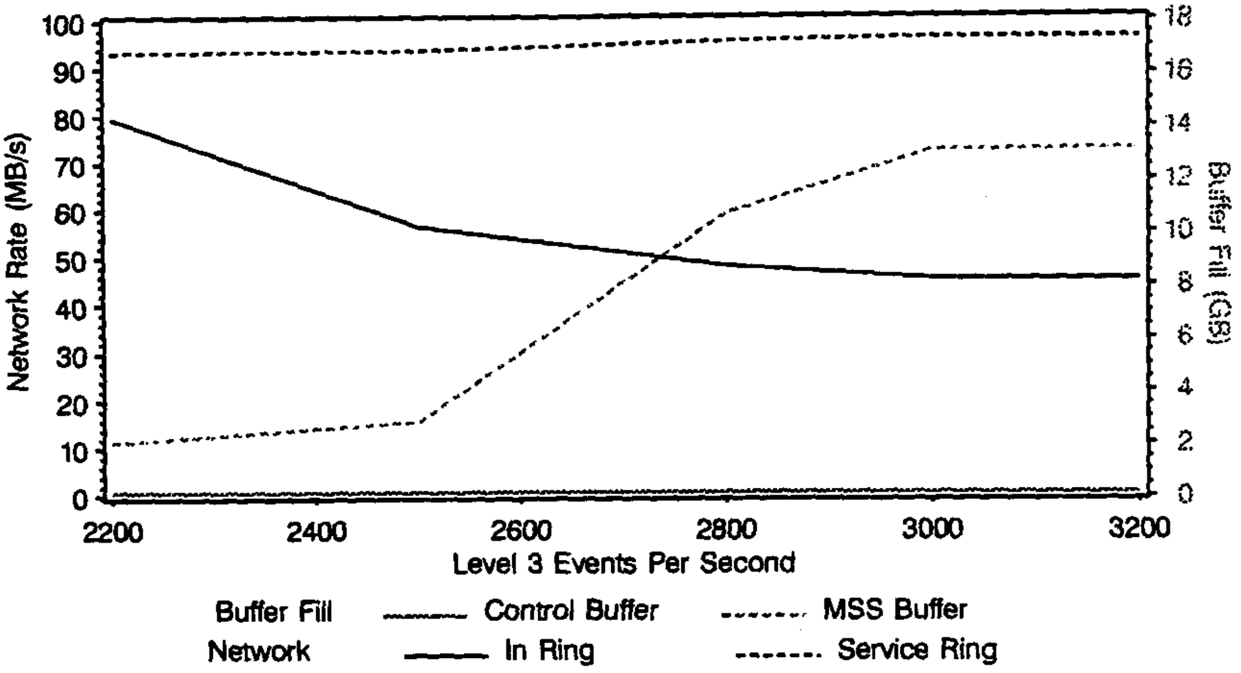
The utilization of retrieve tape drives and robots did not change much with the additional processing as seen in Figure 4.2.6-4 again compared to Section 4.2.3. The store tape drive utilization drops off much slower. As with the previous runs, store tape drive utilization decreases when Pass 2 processing decreases. Furthermore, the rate of the decrease is the same for both.

Figure 4.2.6-4 MSS Resource Utilization



In Figure 4.2.6-5 on page 40, the effective In Ring rate diminishes to just under 50 MB/s. Another sign that the In Ring is not a limited resource, but rather the CPUs. As with the decrease in store tape drive utilization, the increase in maximum MSS Buffer fill is very gradual. Again this is due to the gradual decrease in Pass 2 processing.

Figure 4.2.6-5 Network Rates and Buffer Fill



4.2.7 Processing Timelines

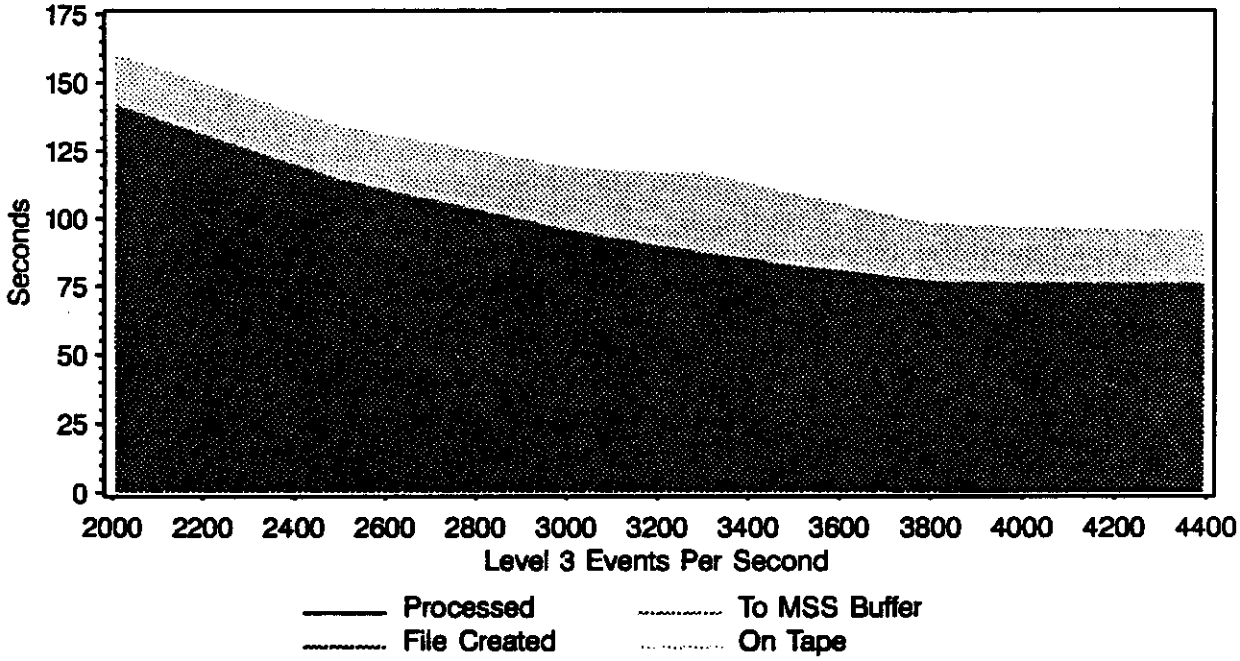
Of the configurations discussed in the previous few sections, the 40 CPU row *configuration* is probably the most desired. This section addresses the timelines associated with processing events with the 40 row configuration. While the actual times differ among configurations, the trends that are present are the same.

Figure 4.2.7-1 on page 41 shows the cumulative times for Level 3 events from processing completion to the MSS Buffer Processor to being in a fully aggregated file to being stored on tape. The processing time is under one second until the efficiency begins to drop. At that point the processing time becomes noticeable in the graph. The time to pass the processed event to the MSS Buffer Processor is essentially 0.01 seconds as the Service Ring has an effective rate close to 100 MB/s.

The major part of the timeline is the time the event stays in the MSS Buffer Processor until enough events have been aggregated to a file for storage. The average time for an event in this state was considered to be one-half the average time for a file to reach its desired size. This time will be shorter for configurations with fewer

CPU rows as rows can be cycled through faster. However, the time to be processed will increase as the In Ring rate degrades and the peak load will also be lower.

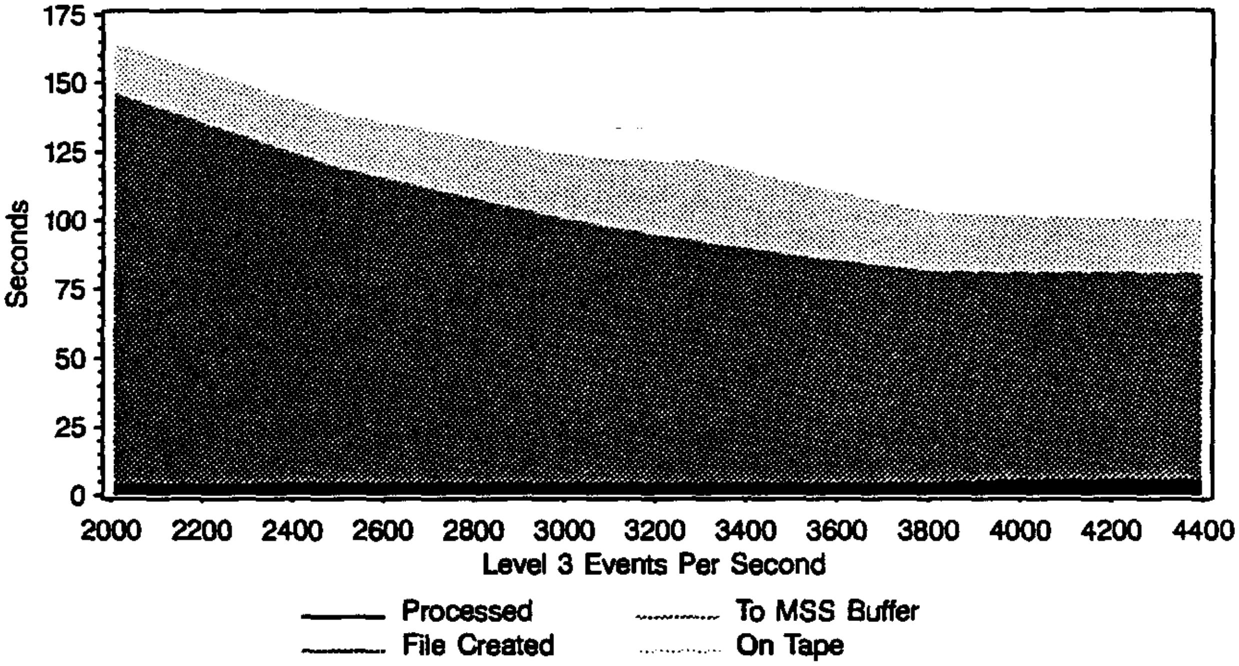
Figure 4.2.7-1 Level 3 Timeline



The final part of the Level 3 event's timeline is the store delay. Note this time increases around a load of 3200 MB/s to 3500 MB/s and then decreases. This is related to the higher store tape drive demand created by more Pass 2 processing. The decrease occurs as Pass 2 becomes throttled.

The timelines for Pass 1, seen in Figure 4.2.7-2, are almost identical to Figure 4.2.7-1 except that Pass 1 processing takes about 4 seconds.

Figure 4.2.7-2 Pass 1 Timeline

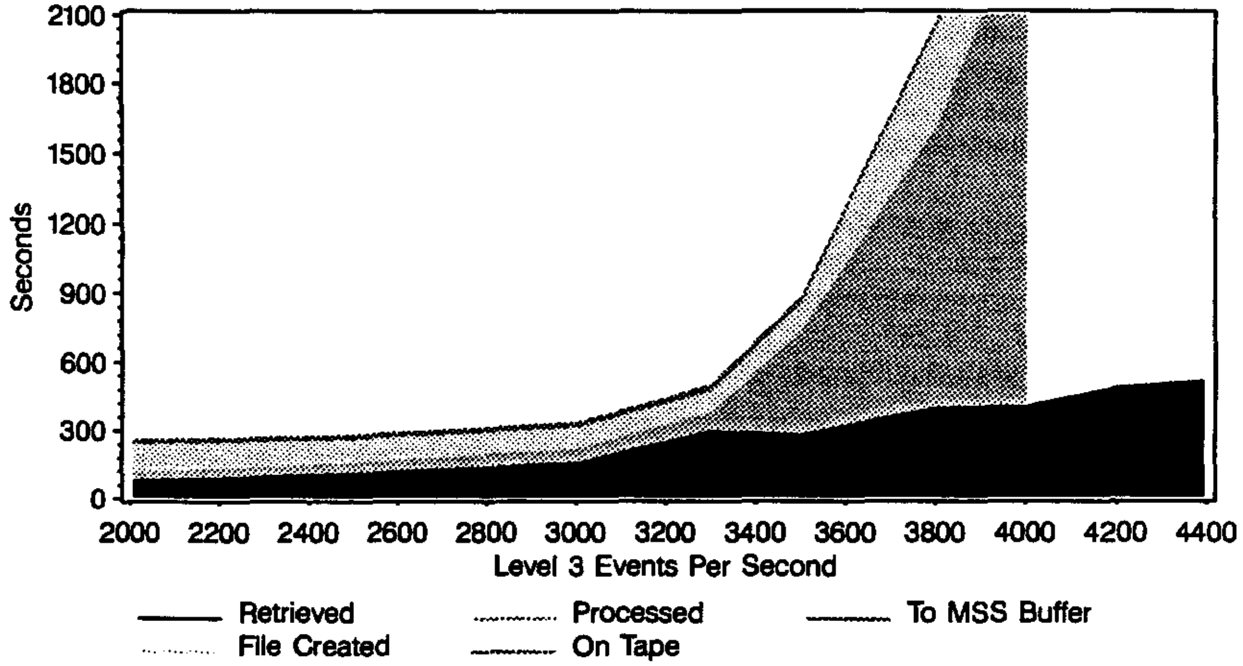


The effect of Pass 2 events having lower priority than Level 3 events becomes greatly noticeable in Figure 4.2.7-3. The Pass 2 event timelines approach infinity as Pass 2 processing becomes less and less frequent.

The Pass 2 timeline begins with a file being retrieved from the mass storage system. This time steadily increases as load increases. Do remember from Section 4.2.2 that this mass storage system's configuration cannot handle loads above 3,000 MB/s and reflect the mean time only after one hour. Once the file is retrieved and broken into events, the event's average time until being processed is about 60 seconds. As in the 3,000 MB/s load case, the time until processed ranges from 3 seconds for the first event in the file to 115 seconds for the last event in the file. Once Pass 2 processing becomes throttled, the mean time until an event is processed increases quite rapidly. When a Pass 2 event is processed it travels to the MSS Buffer in hundredths of a second. As Pass 2 becomes throttled the time to create an aggregated file of Pass 2 events also increases toward infinity. The time to store a tape remains in the 20 second range.

When Pass 2 processing is totally shut down, the only Pass 2 related activity is the retrieval of files from the mass storage system.

Figure 4.2.7-3 Pass 2 Timeline



4.2.8 Trigger Rate Transition from 4,000 to 2,000 MB/s

The purpose of this simulation run was to investigate how a system reacts to a change in load. This particular run used a system consisting of 40 CPU rows with 32 CPUs per row and 12 robots. The initial Level 3 input rate was 4,000 MB/s and was decreased to 2,000 MB/s after 45 minutes of simulated time. The system was intentionally started in a state which would overload it. Previous analysis has already shown that 12 robots and 40 CPU rows cannot handle 4,000 MB/s.

In Figure 4.2.8-1, the data being retrieved for Pass 2 processing is being accumulated in the MSS Buffer Processors up to the time Pass 2 processing started, which is when the arrival rate dropped to 2,000 MB/s. Similarly, the Control Processors' buffers are constantly full until the rate change.

Figure 4.2.8-1 Total MSS Buffer Fill Over Time

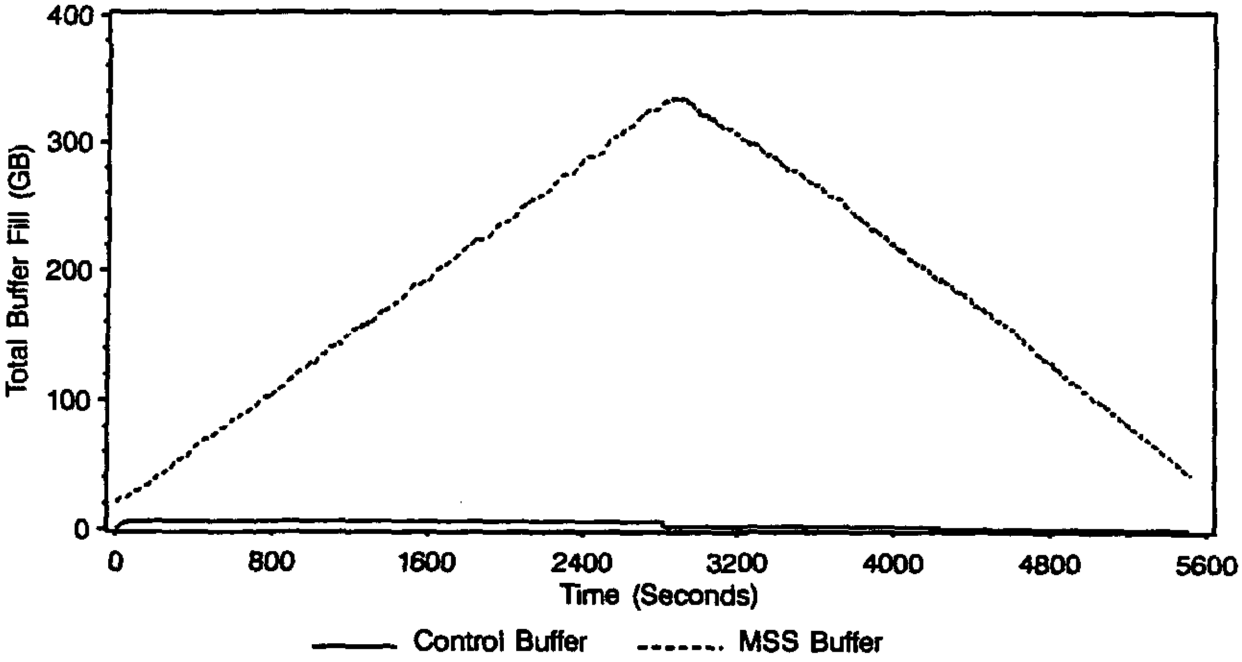
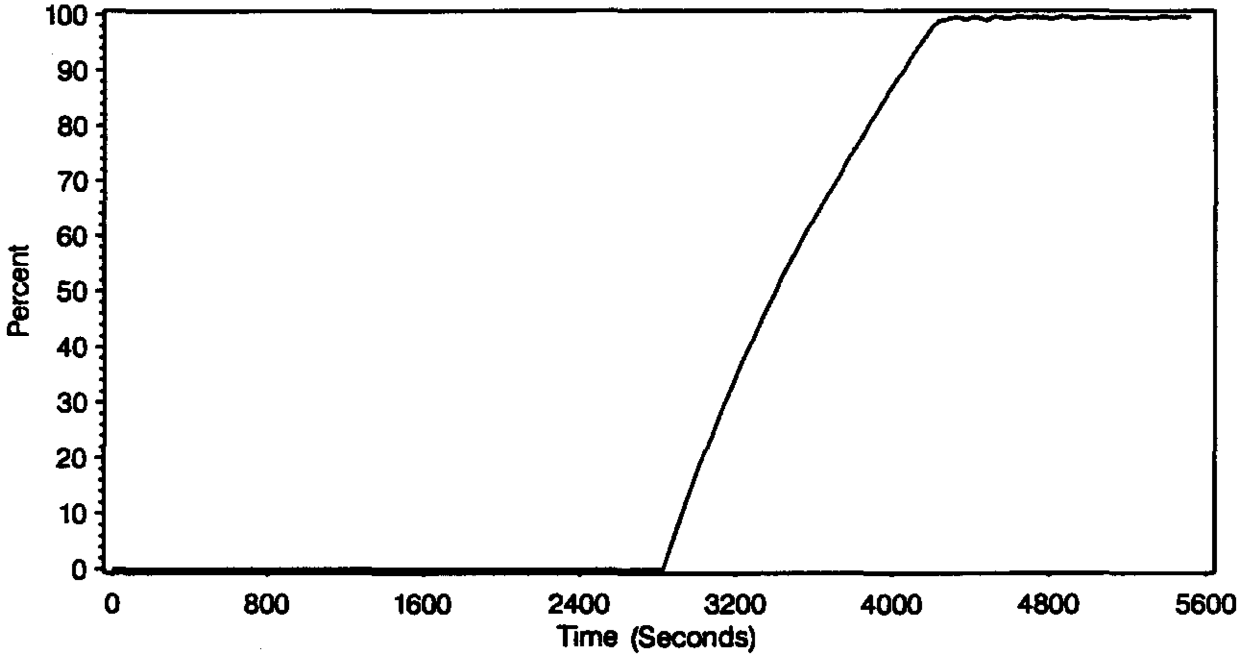


Figure 4.2.8-2 shows the percentage of Pass 2 data arrived to the MSS Buffer Processors that has been processed. Note that the processing part of the system takes just over 20 minutes to be back in a normal operations mode (i.e. 98% Pass 2 data retrieved has been processed). This is not obvious from Figure 4.2.8-1 in which a considerable amount of data remains in the MSS Buffer Processors.

Figure 4.2.8-2 Pass 2 Completion Rate Since Time 0



By looking at the retrieve tape drive utilization and the store tape drive utilization, Figures 4.2.8-3 and 4.2.8-4 respectively, the explanation becomes quite clear. Note the peak of the retrieve tape drive queue matches the peak of the MSS Buffer fill. However, the retrieve tape drive queue falls off faster than the buffer fill. The store tape drive queue grows considerably once Pass 2 processing begins. Therefore, a large portion of the data in the MSS Buffer Processors after time 4,000 is Pass 2 processed data. There is more data waiting to be stored rather than waiting to be processed.

Figure 4.2.8-3 Retrieve Tape Drive Utilization Over Time

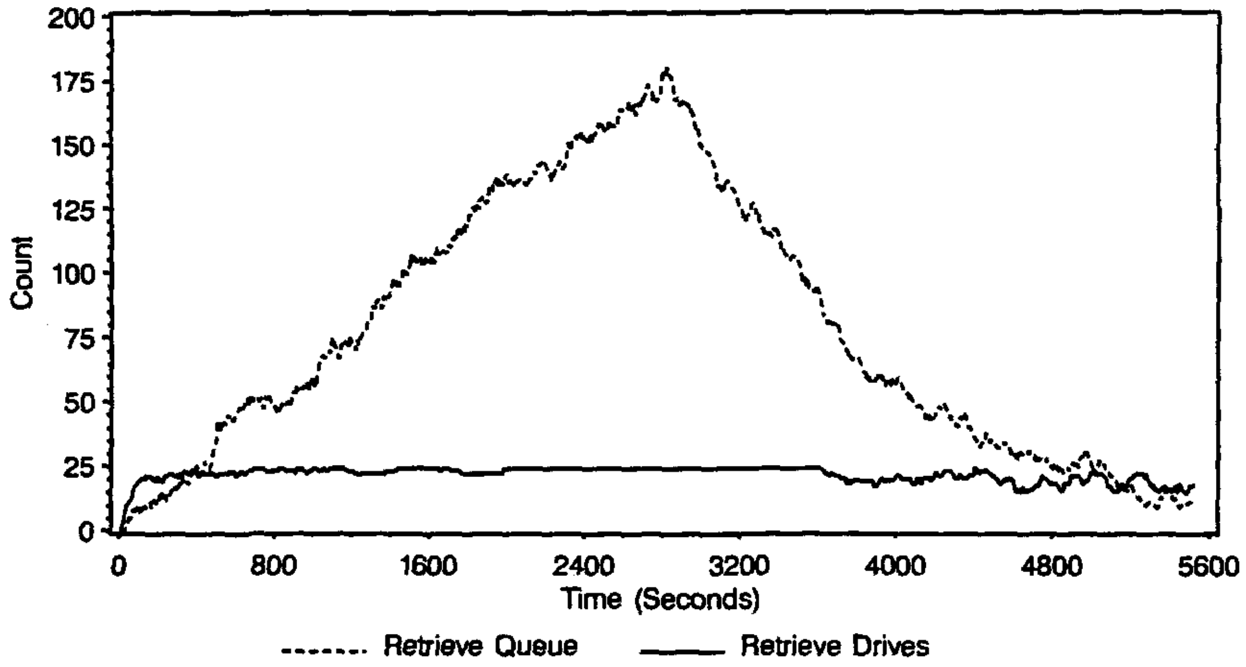


Figure 4.2.8-4 Store Tape Drive Utilization Over Time

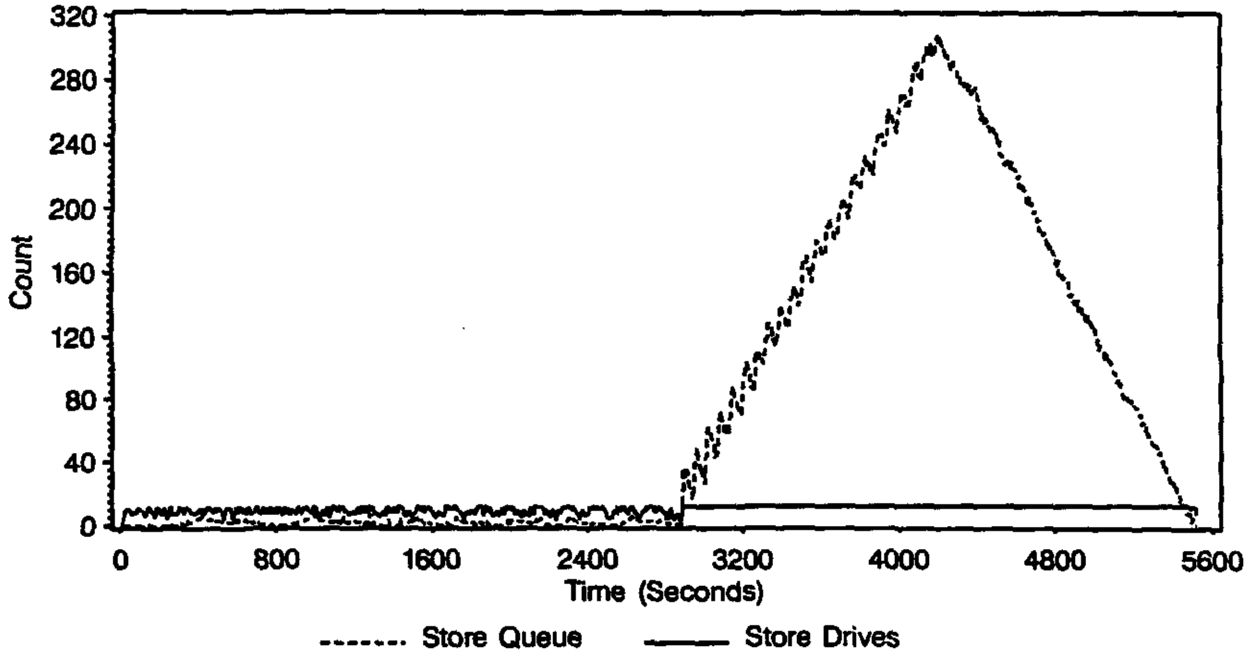
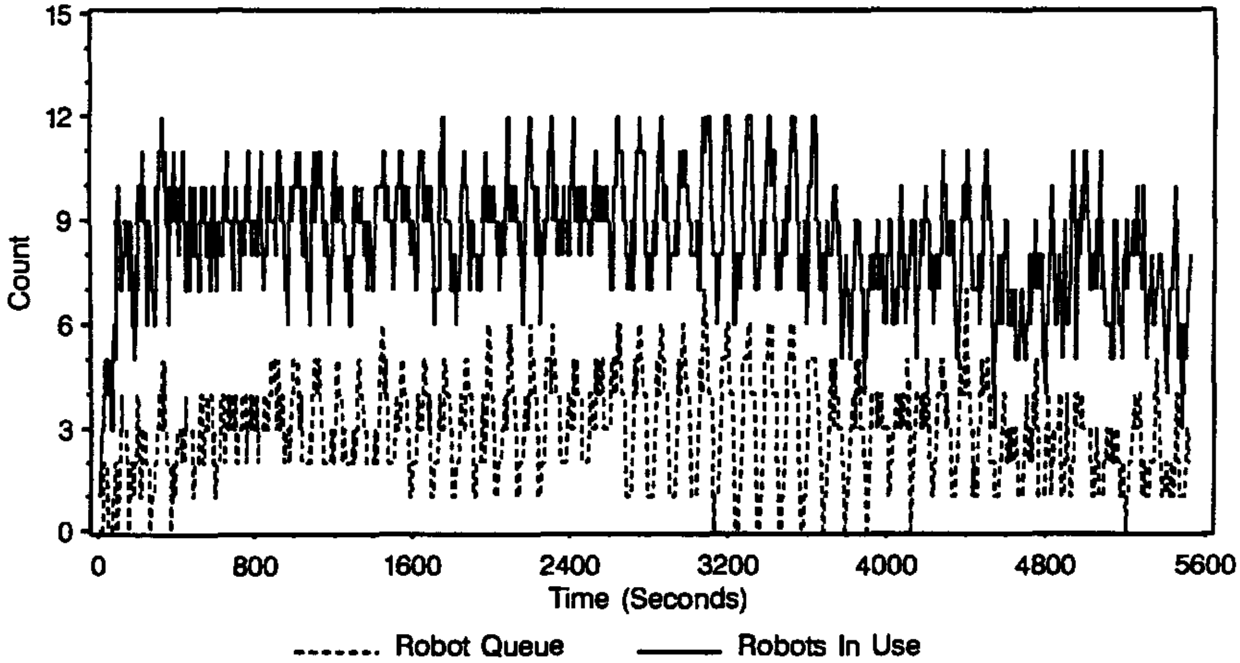


Figure 4.2.8-5 Robot Utilization Over Time



4.3 Observations

Some general observations from the simulation runs presented in the previous section are:

- Service Ring utilization drops when Pass 2 processing is being reduced.
- Store tape drive utilization drops when Pass 2 processing is being reduced.
- Unless CPUs are already a limited resource, the In Rings will be a bottleneck anytime *l3_arrival_rate / cpu_box_rows in_ring_rate.*
- If the In Ring is bottlenecked, poorer usage of the CPUs will result. Events spend more time getting to the CPU.
- Based on load of 3,000 MB/s where systems are in steady state, each control processor only needs 100 MB of memory to handle the arrival of Level 3 events. Additional memory is only needed for Pass 2 data, and 50 MB should be ample.

- Once each MSS Buffer Processor's fill begins to increase much over 2 GB, the system is behind. Unless the load decreases for some period of time, the system will never catch up and data loss may occur. Note that in each of the MSS Buffer Fill graphs, when the buffer fill was in the 12-16 GB range, this was the ending maximum value and was still increasing.
- Twelve robots can not handle loads greater than 3,000 MB/s

5.0 Summary

The results of the study are summarized as follows:

- Pass 1 can be performed immediately after Level 3 on the same machine thereby significantly reducing networking requirements.
- 1,280 workstations are required to support a trigger rate of 3,000 events per second.
- There must be at least 40 CPU processing rows to ensure throughput on the In Ring networks has not degraded to the point of poor utilization of the Computing Processors.
- At least 100 MB/s effective data throughput is required to support an architecture with 40 horizontal rows. If this rate can not be achieved then adding more rows would alleviate network contention.
- Assuming an aisle-based mass storage system, at least 12 robots and 42 tape drives are required.

This page intentionally left blank

A Simulation Model Diagrams

Figure A-1 Module Page

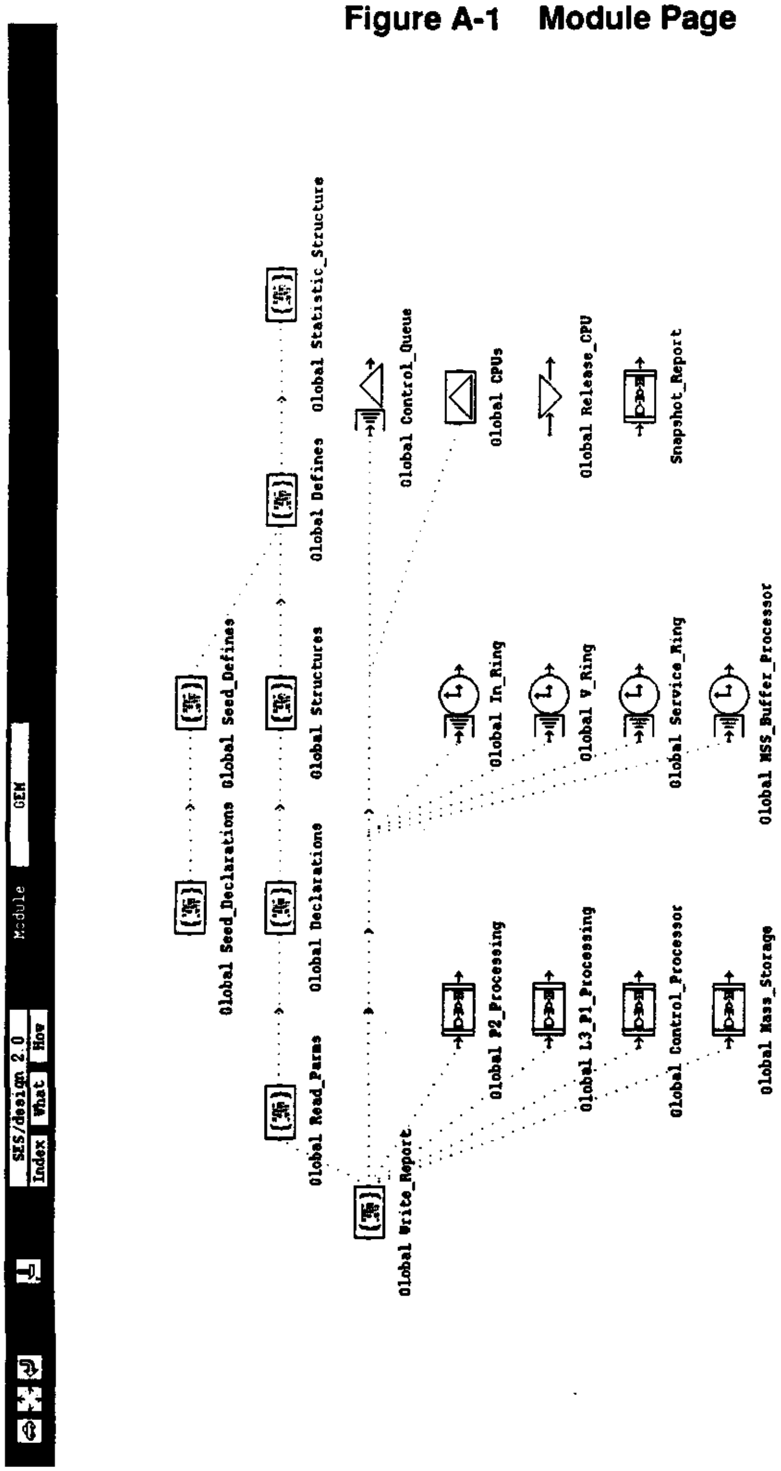
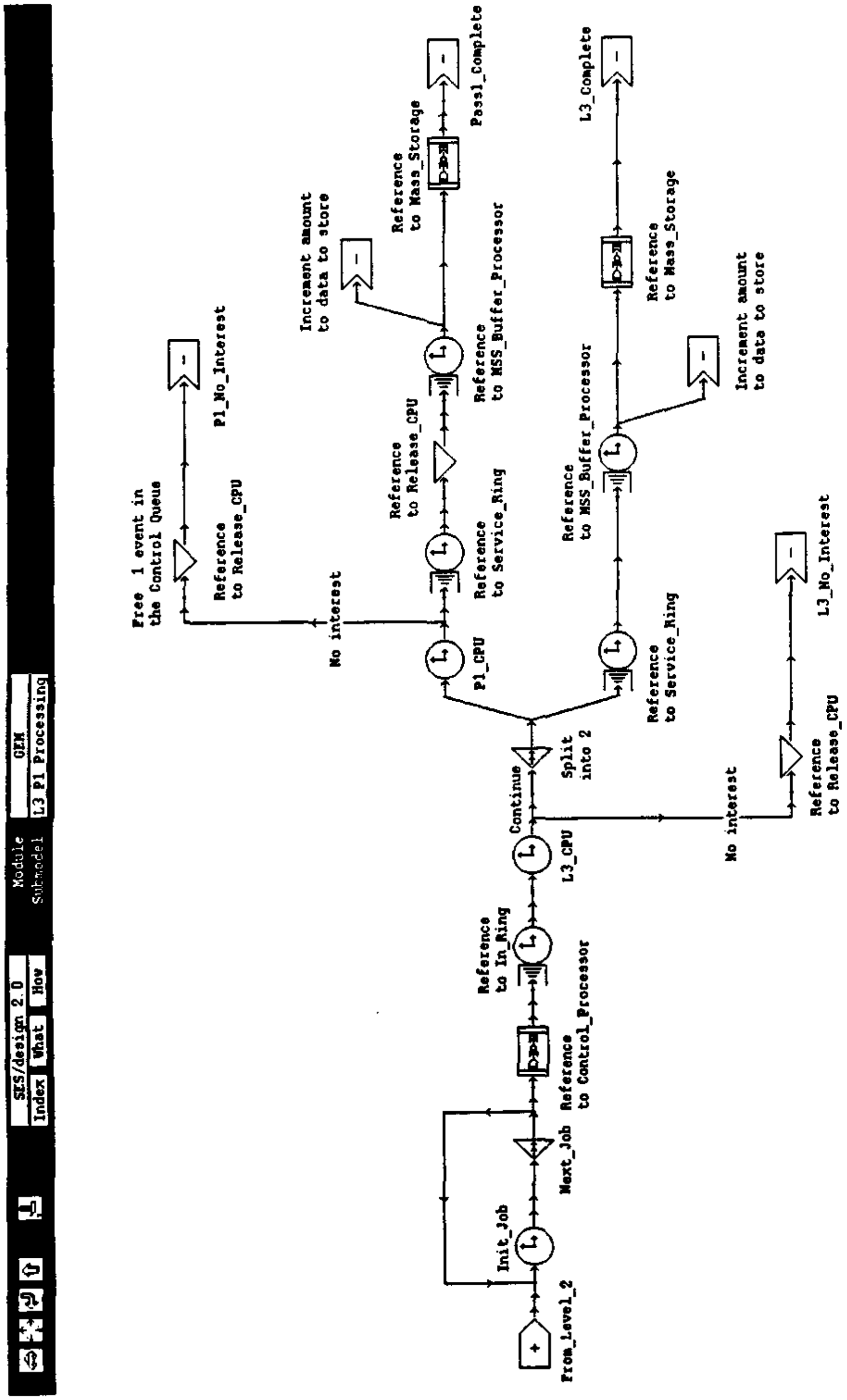


Figure A-2 Level 3 and Pass 1 Processing



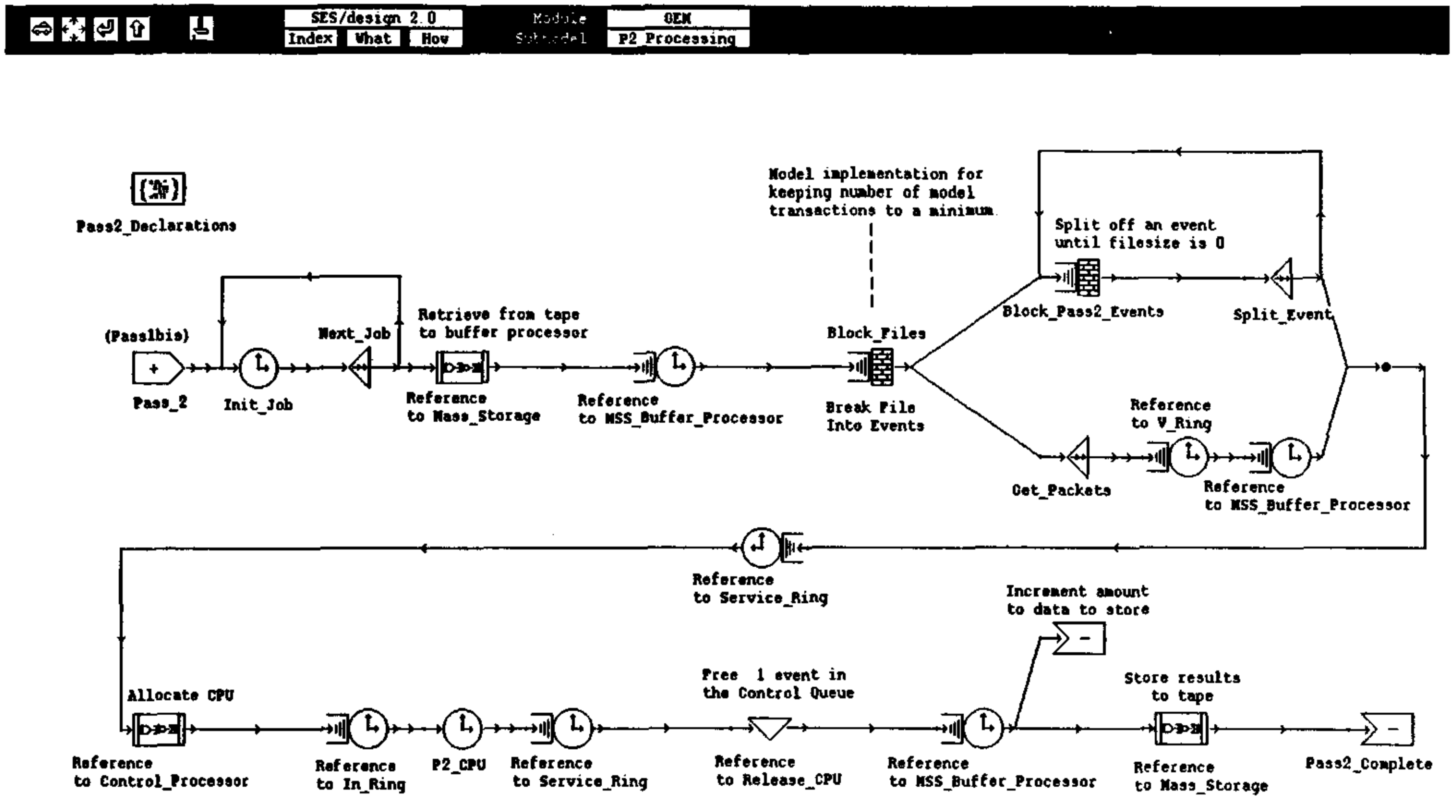

SES/design 2.0
Index
What
How
Module
GEM
Submodel
P2 Processing
Pass2_Declarations
Model implementation for keeping number of model transactions to a minimum
Split off an event until filesize is 0
(Pass1bis)
Pass_2
Init_Job
Next_Job
Retrieve from tape to buffer processor
Reference to Mass_Storage
Reference to MSS_Buffer_Processor
Block_Files
Break File Into Events
Block_Pass2_Events
Split_Event
Get_Packets
Reference to V_Ring
Reference to MSS_Buffer_Processor
Reference to Service_Ring
Increment amount to data to store
Allocate CPU
Reference to Control_Processor
Reference to In_Ring
P2_CPU
Reference to Service_Ring
Free 1 event in the Control Queue
Reference to Release_CPU
Reference to MSS_Buffer_Processor
Store results to tape
Reference to Mass_Storage
Pass2_Complete

Figure A-3 Pass 2 Processing

Figure A-4 Control Processor

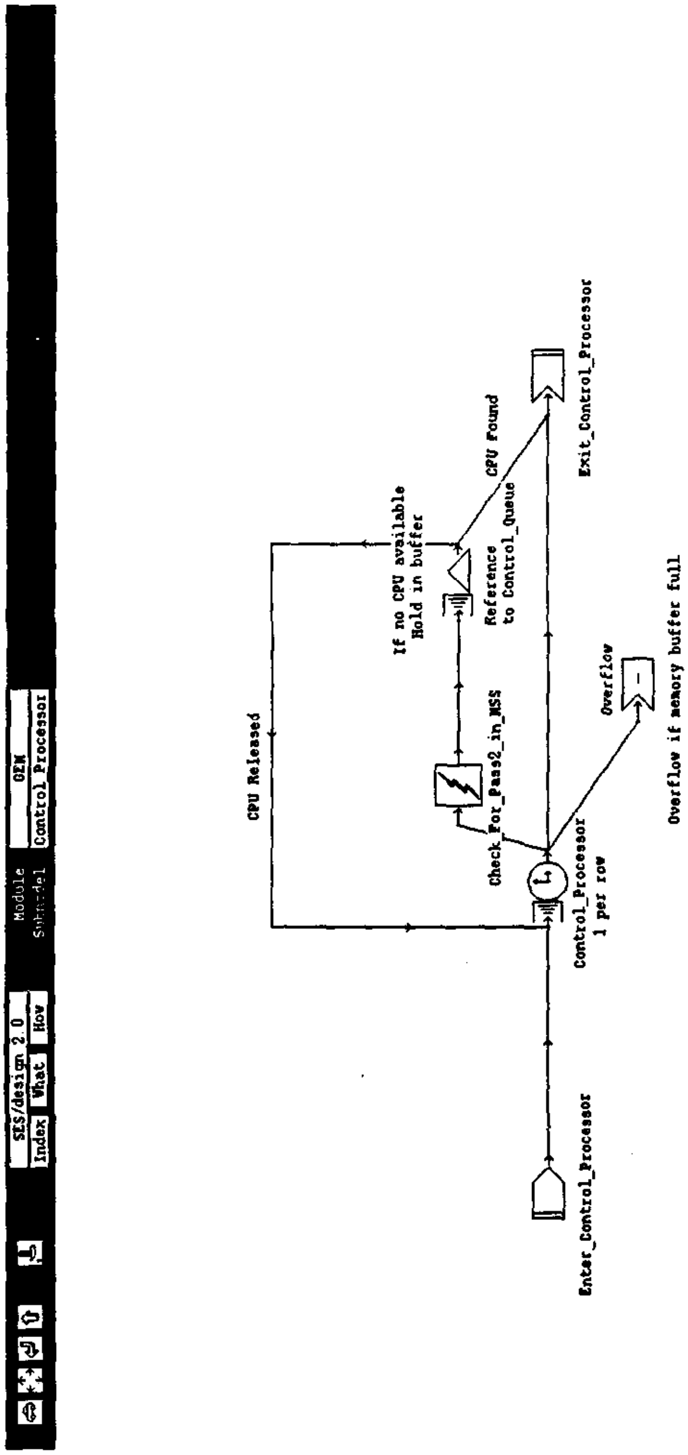
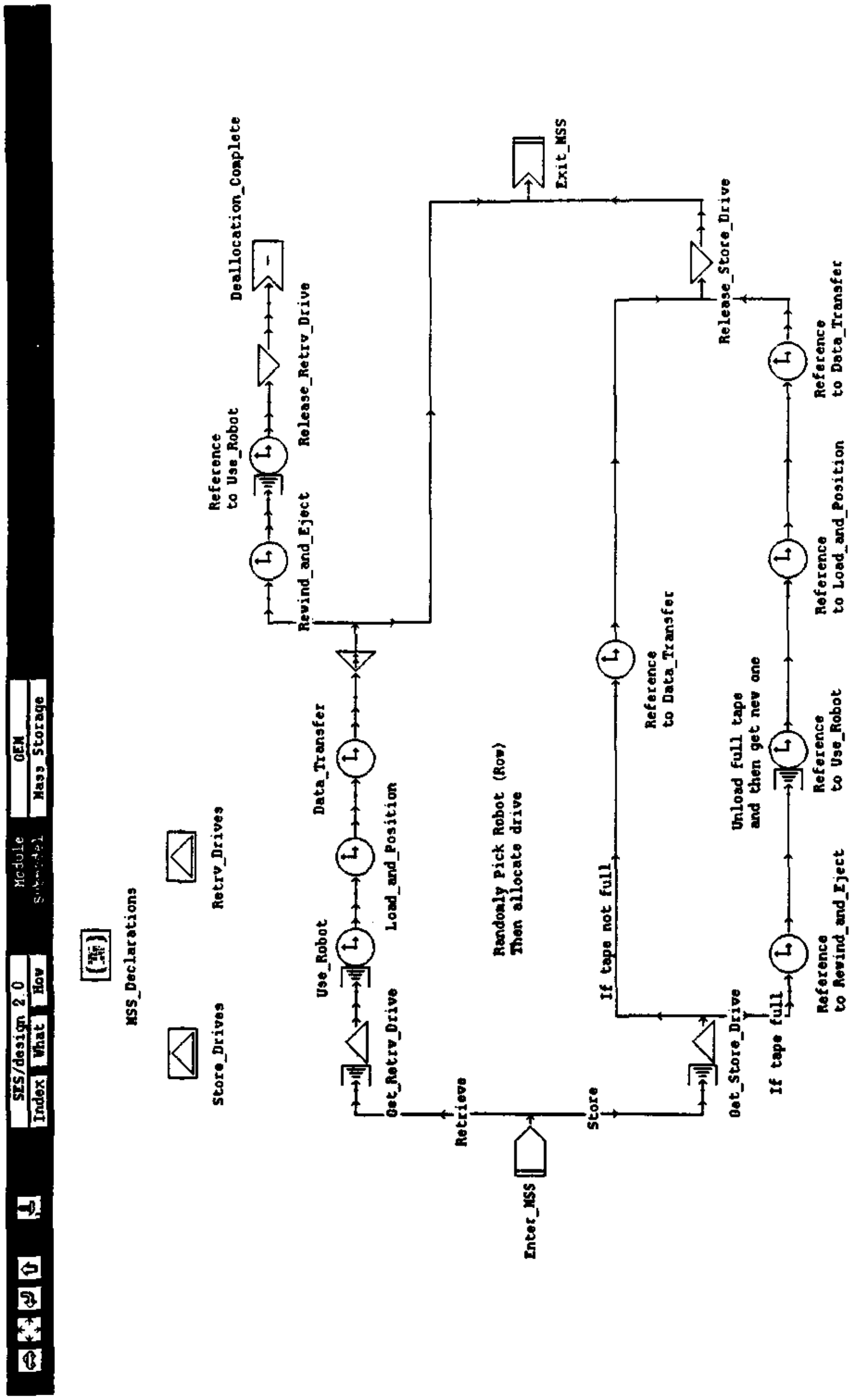
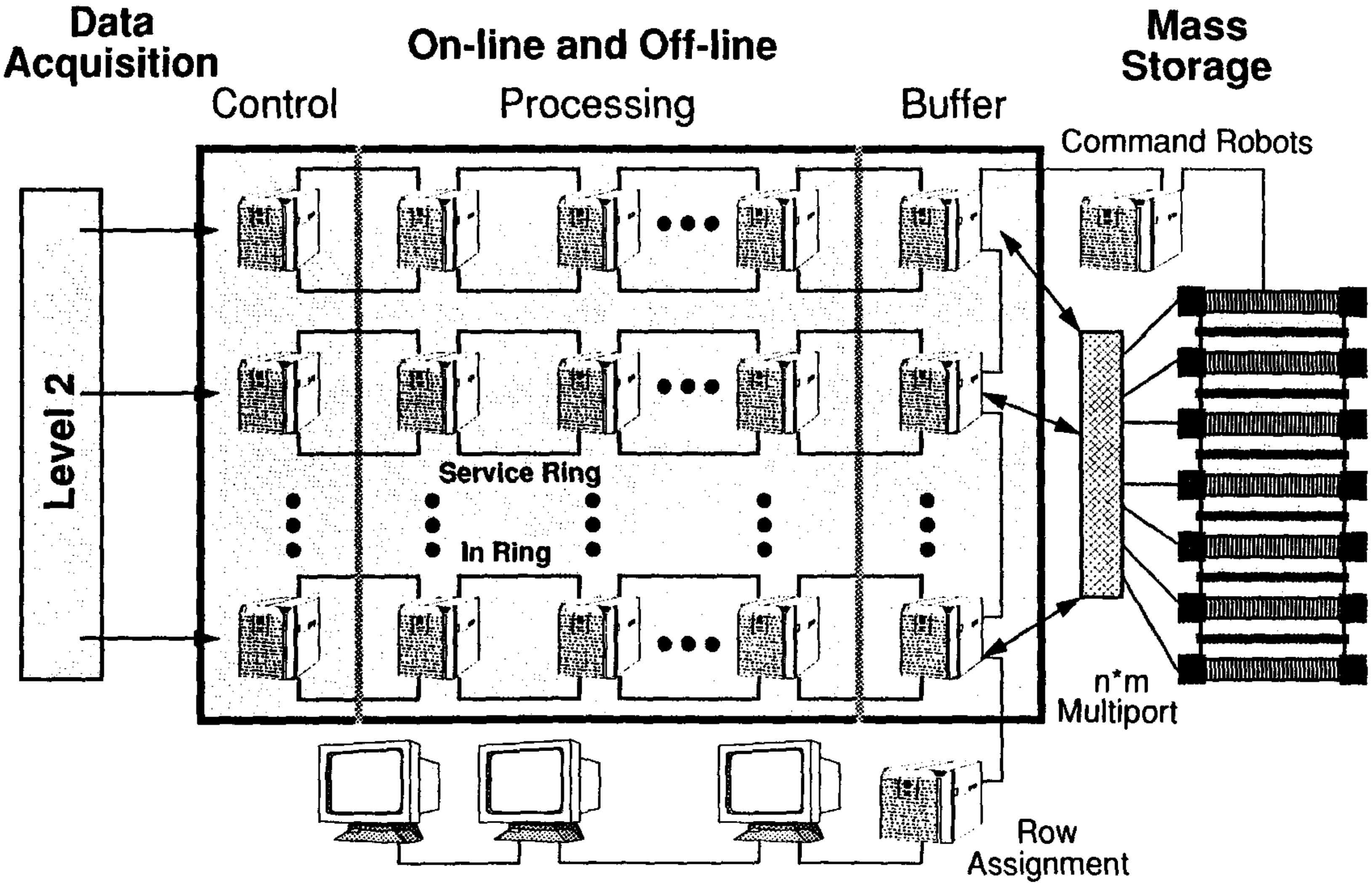


Figure A-5 Mass Storage System



GEM DETECTOR COMPUTING STUDY
FINAL RESULTS
SEPTEMBER 18, 1992


Data Acquisition
On-line and Off-line
Mass Storage
Control
Processing
Buffer
Command Robots
Level 2
Service Ring
In Ring
n*m Multiport
Row Assignment

TOPICS

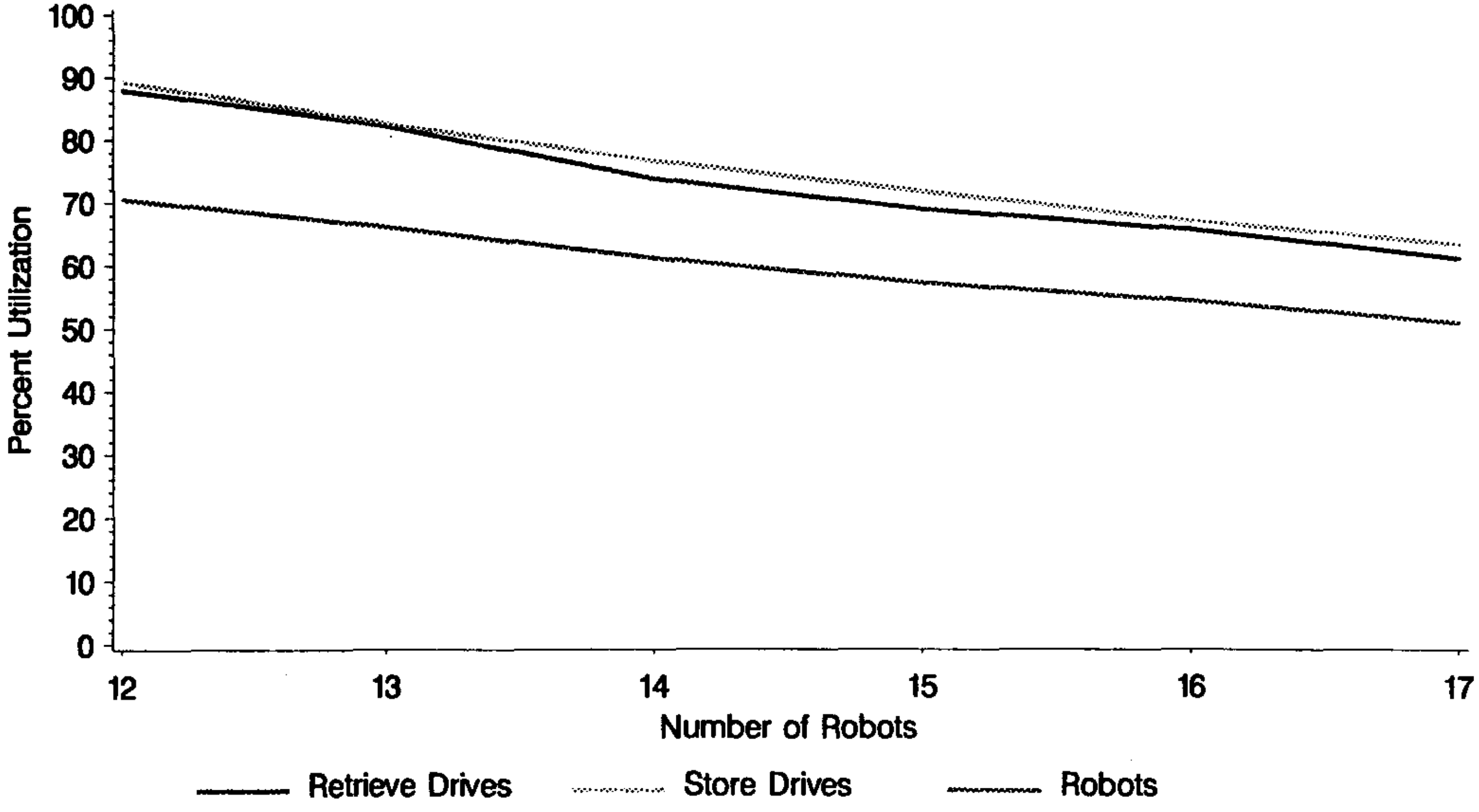
- Varying Number of Robots
- 2 Retrieve Drives Per Robot vs. 3 Retrieve Drives per Robot (12 Robots)
- 40 CPU Row Configuration
- 32 CPU Row Configuration
- 32 CPU Row Configuration with Level 3 Processing Time Doubled
- 24 CPU Row Configuration
- Trigger Rate Transition from 4000 MHz to 2000 MHz

Varying Number of Robots

- 3000 Level 3 events per second
- 32 CPU Rows
- 40 CPUs per Row
- 1280 Total CPUs
- 150 MB Control Processor Buffer containing no more than 33% Pass 2
- Control Processing Time of 0.001 seconds
- Effective Network Rates of 100 MB/s
- 12 to 17 Robots
- 3 Drives per Robot (1 for Stores and 2 for Retrieves)

MSS RESOURCE UTILIZATION

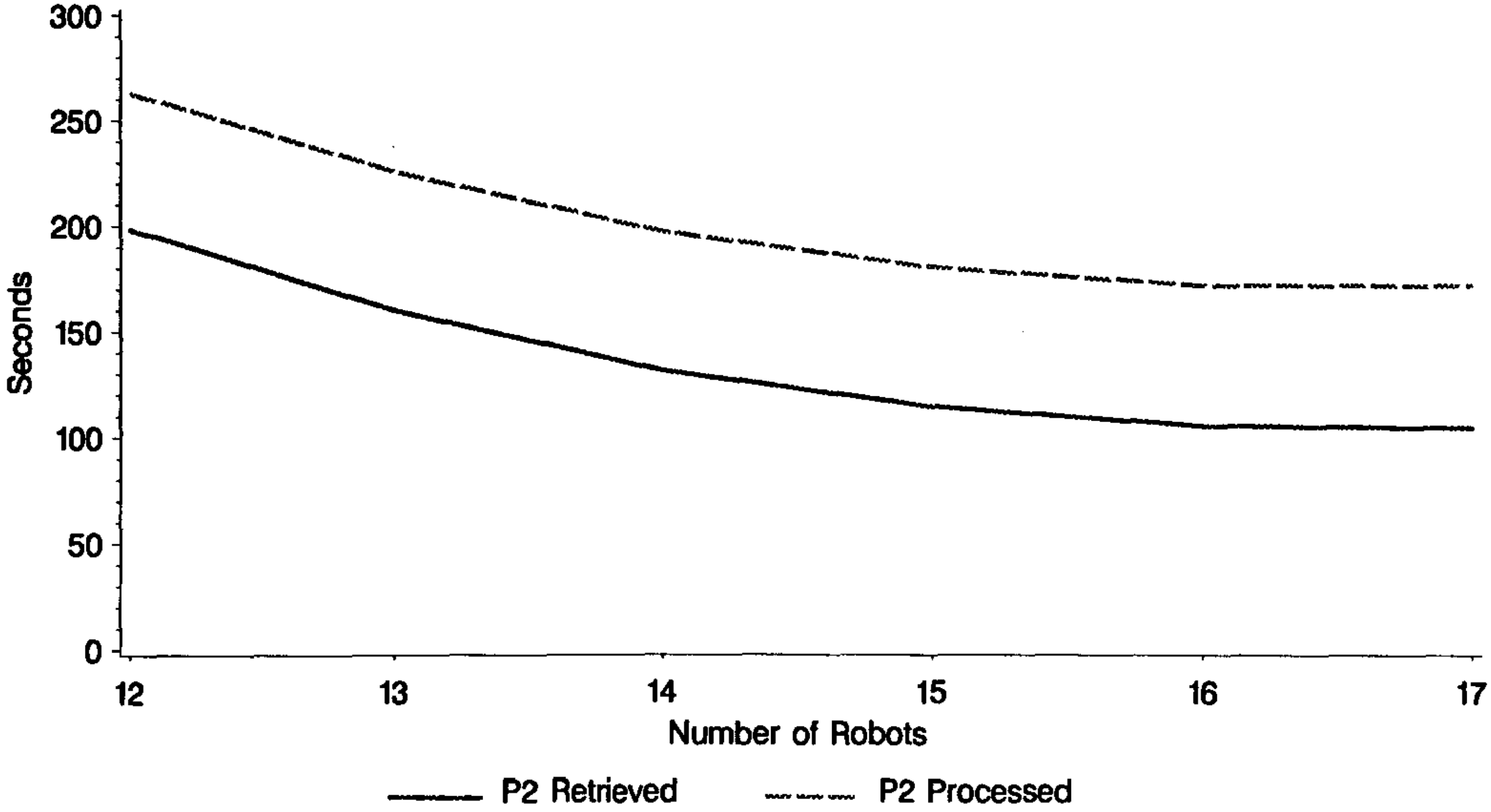
ROBOT VARIATION



NOTE: 3 Drives per Robot (1 for Stores, 2 for Retrieves)

PASS 2 COMPLETION TIMES

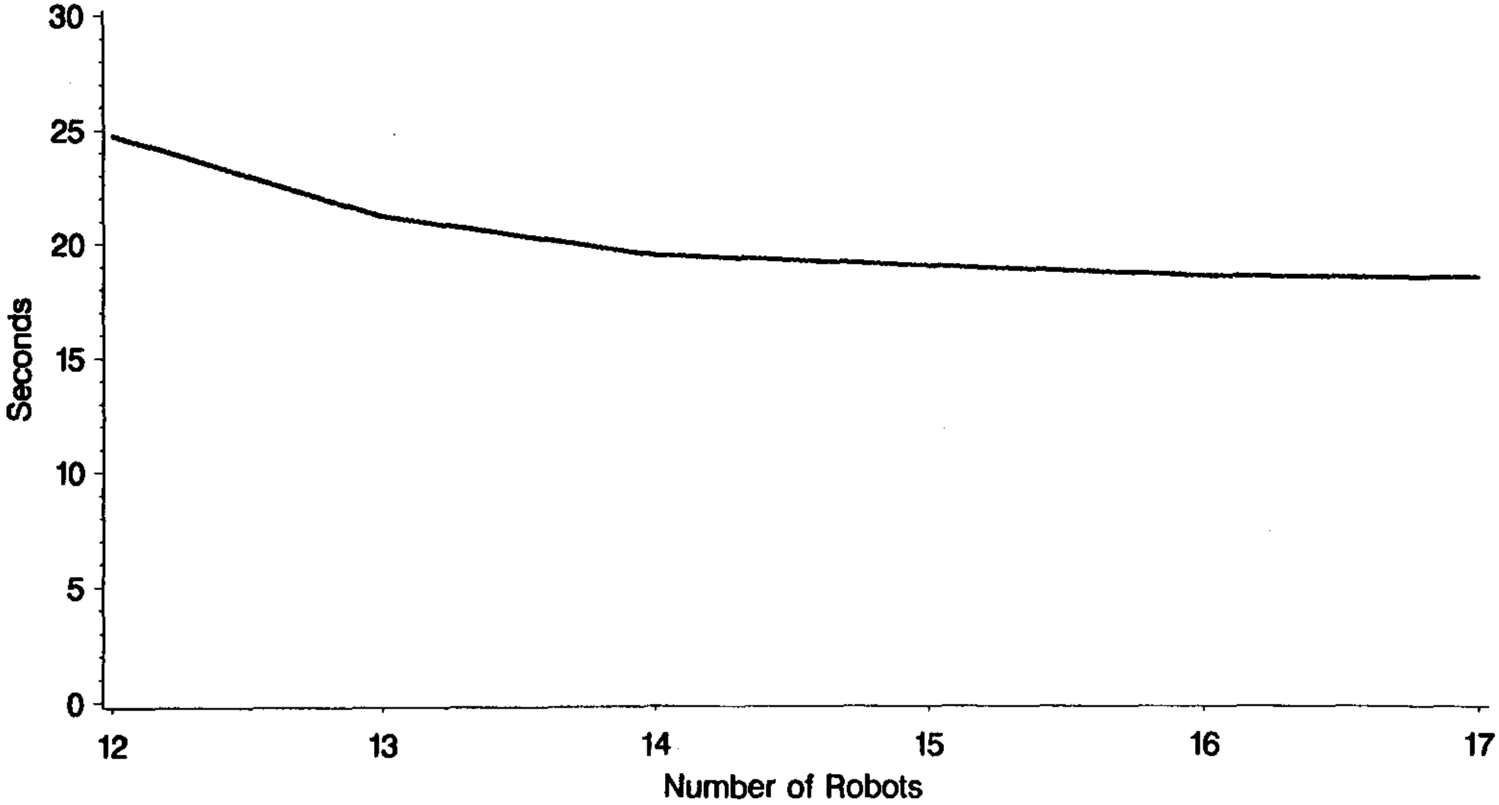
ROBOT VARIATION



NOTE: 3 Drives per Robot (1 for Stores, 2 for Retrieves)

MSS STORE DELAY TIMES

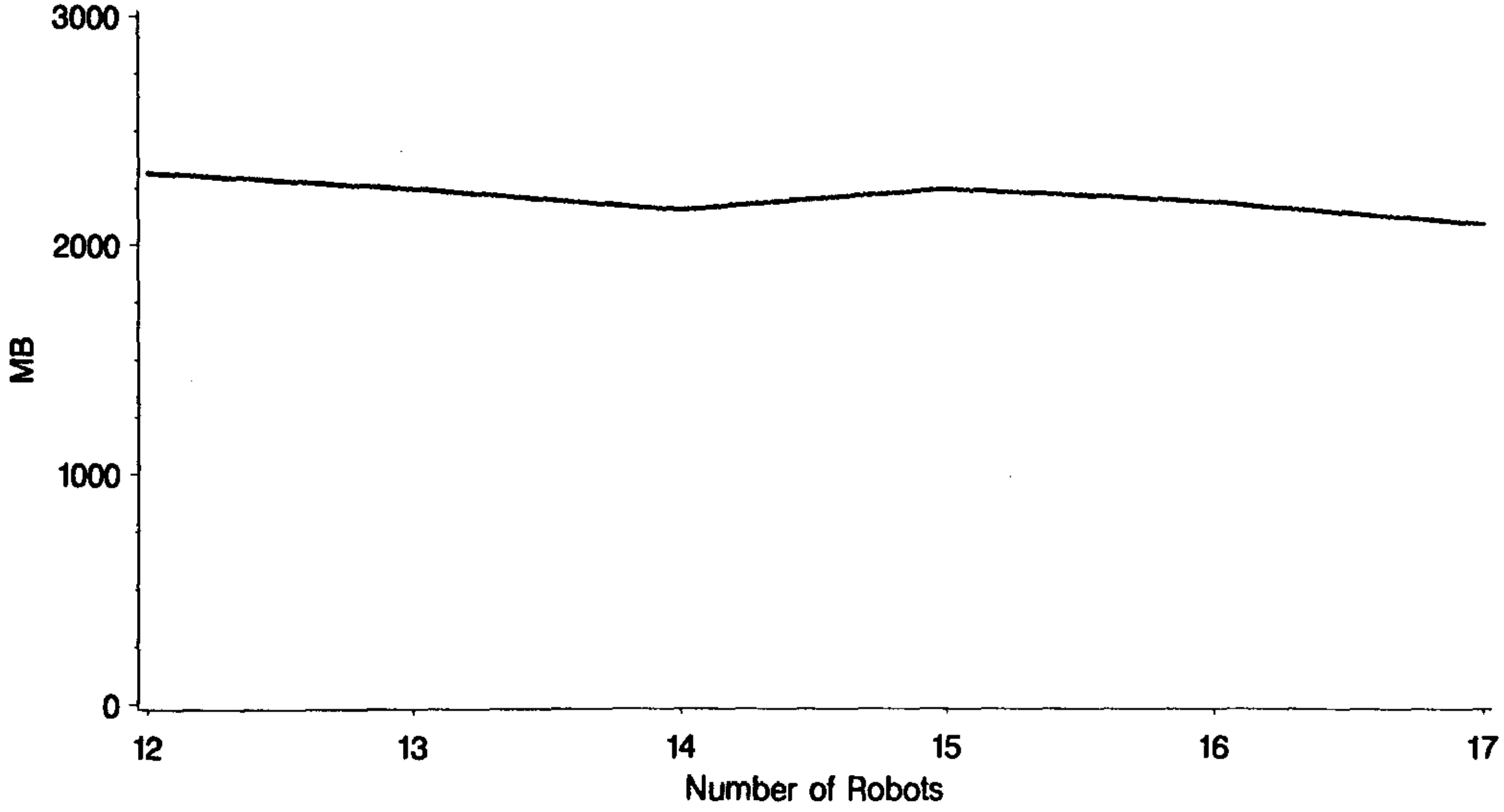
ROBOT VARIATION



NOTE: 3 Drives per Robot (1 for Stores, 2 for Retrieves)

MSS BUFFER FILL

ROBOT VARIATION



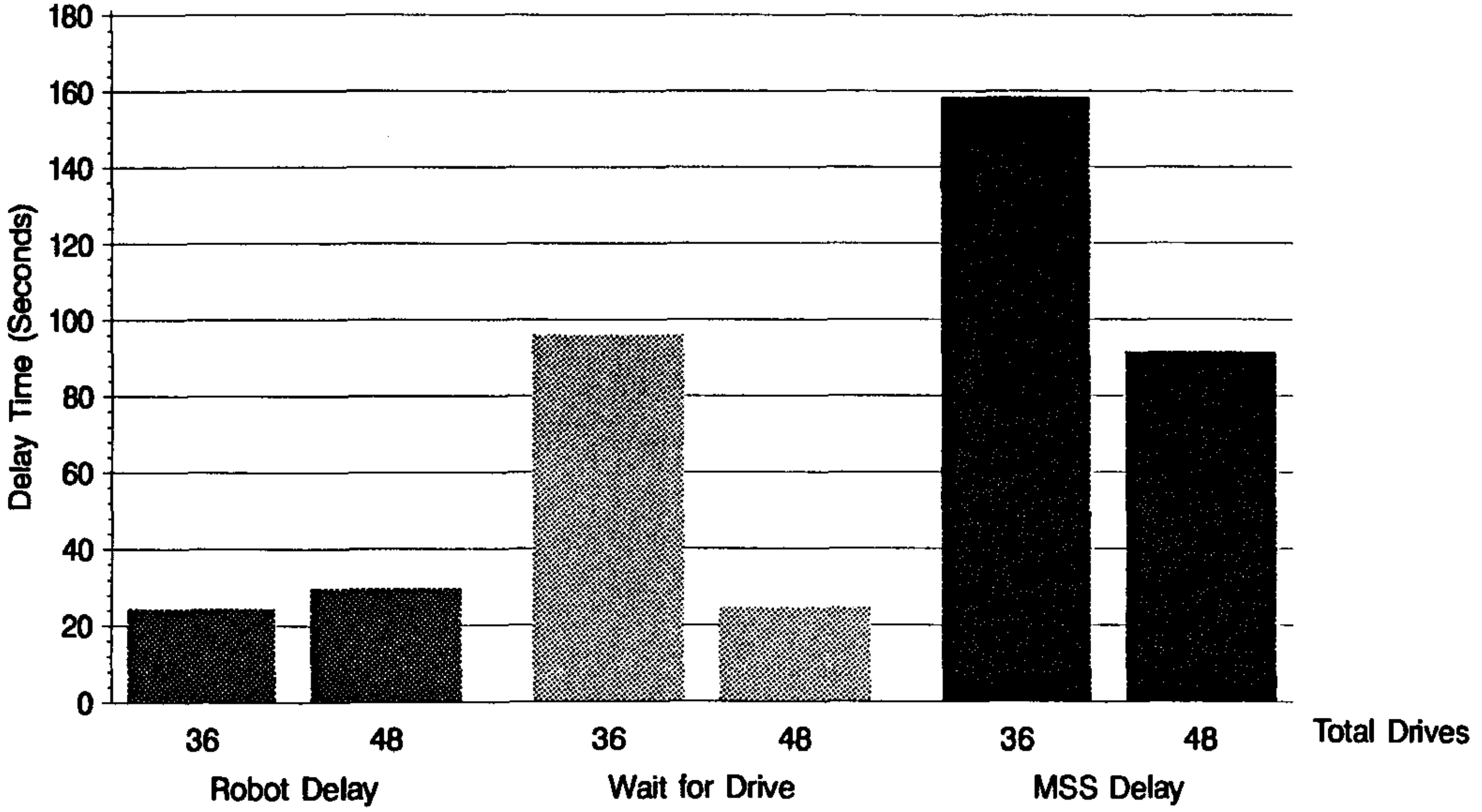
NOTE: 3 Drives per Robot (1 for Stores, 2 for Retrieves)

12 Robot Configuration
2 Retrieve Drives per Robot vs. 3 Retrieve Drives per Robot

- 3000 Level 3 events per second
- 40 CPU Rows
- 32 CPUs per Row
- 1280 Total CPUs
- 150 MB Control Processor Buffer containing no more than 33% Pass 2
- Control Processing Time of 0.001 seconds
- Effective Network Rates of 100 MB/s

RETRIEVE DELAY TIMES

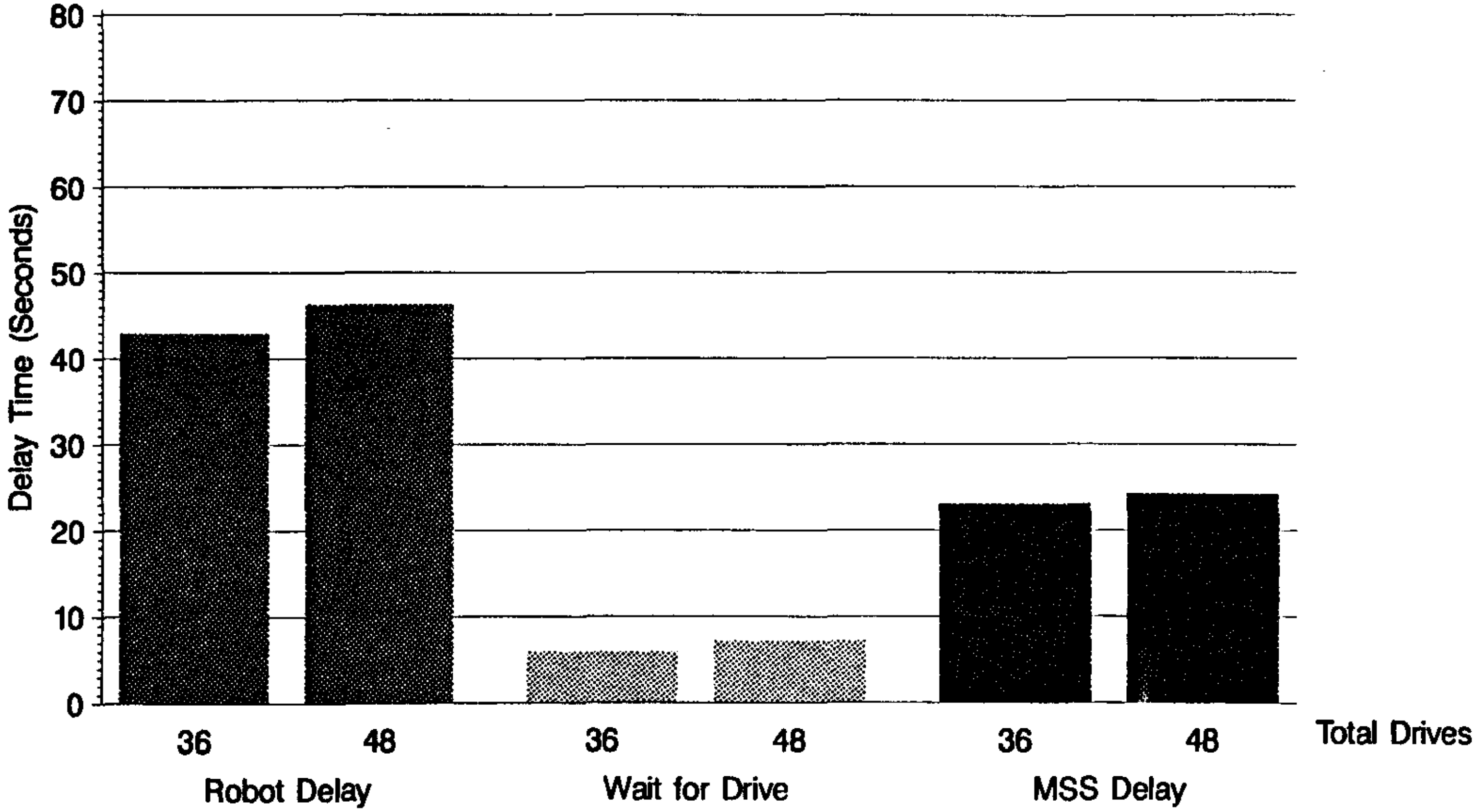
12 ROBOT CONFIGURATION



Drive Distribution: 12 Store Drives + 24 or 36 Retrieve Drives

STORE DELAY TIMES

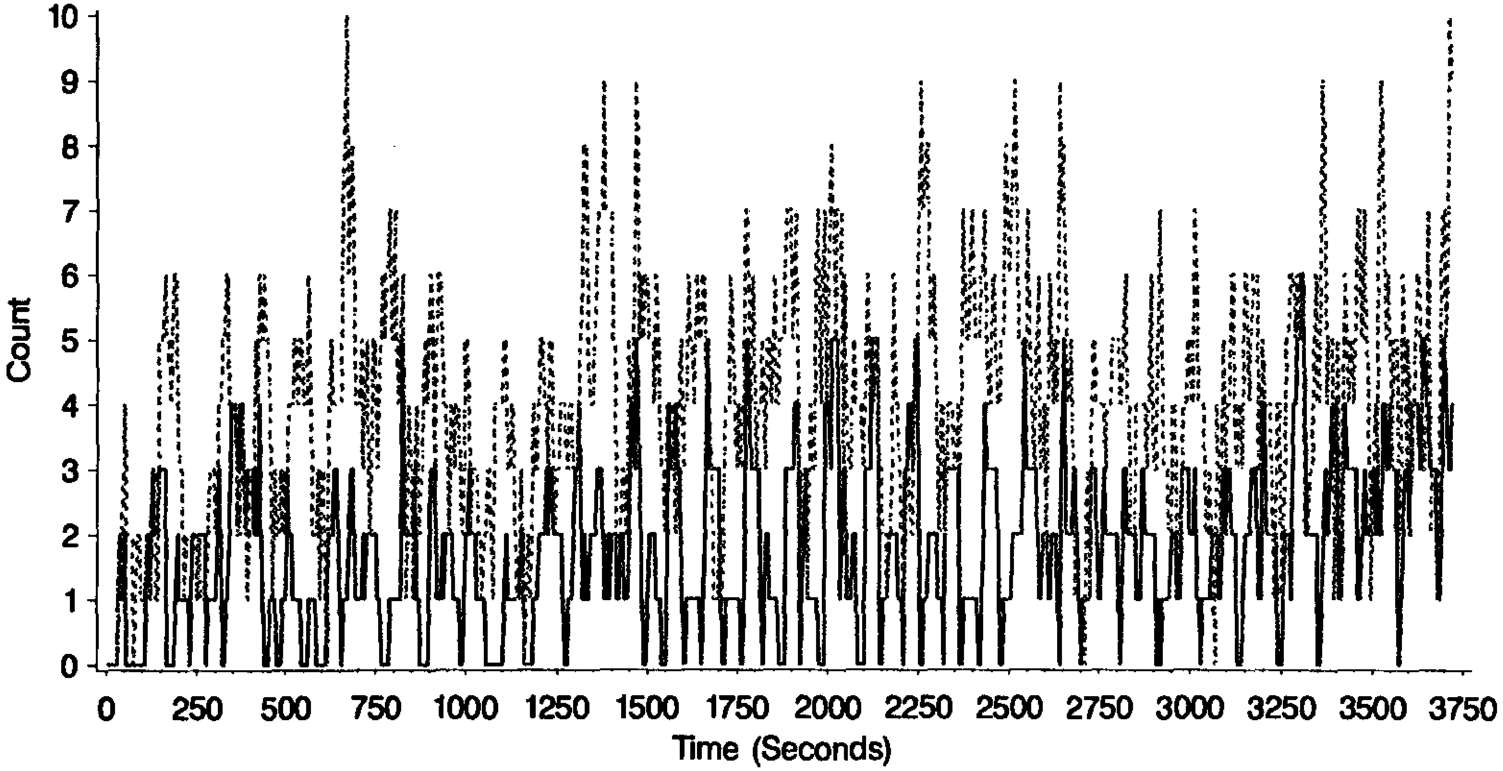
12 ROBOT CONFIGURATION



Drive Distribution: 12 Store Drives + 24 or 36 Retrieve Drives

COMPARISON OF ROBOT QUEUES

12 ROBOT CONFIGURATION

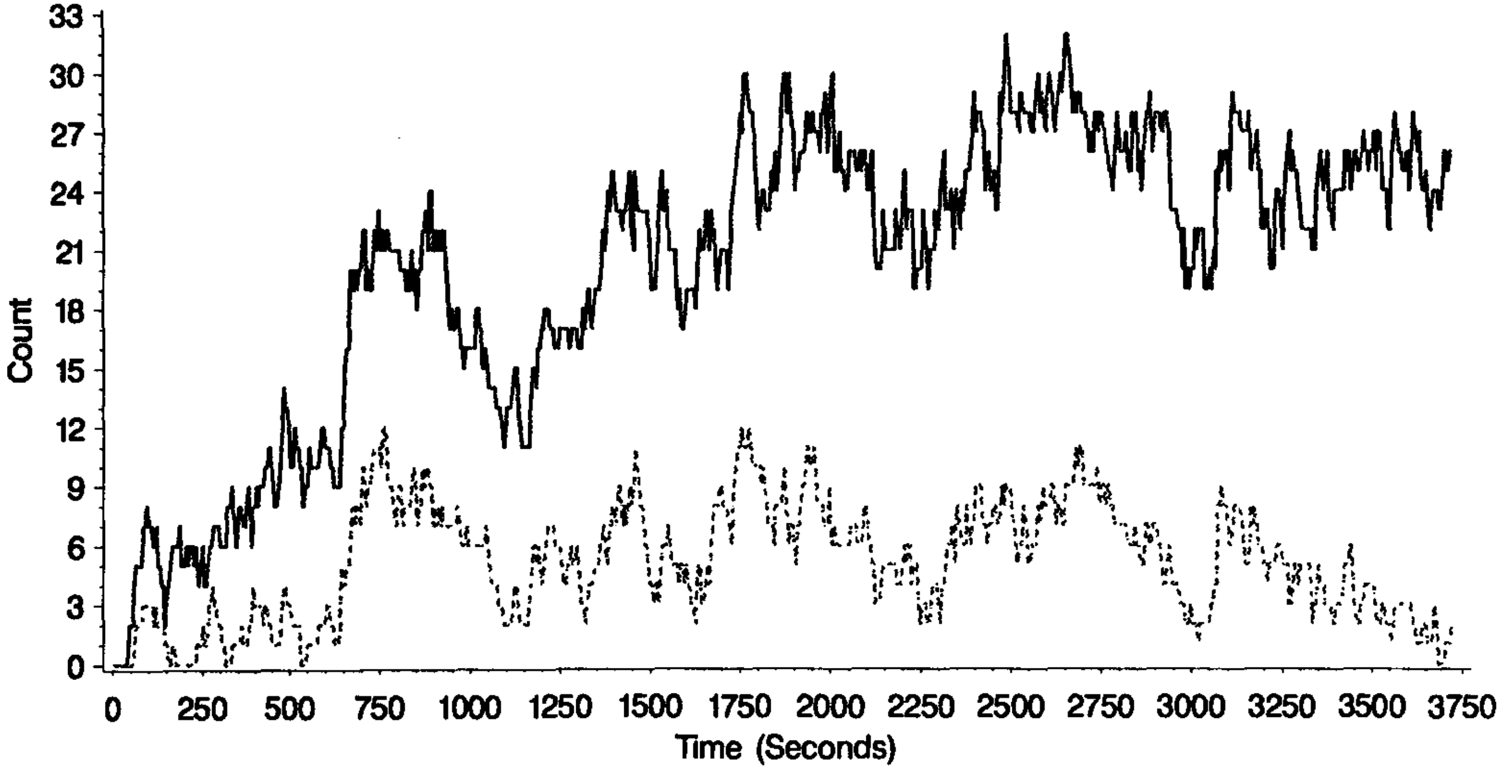


Total Drives ——— 36 48

Drive Distribution: 12 Store Drives + 24 or 36 Retrieve Drives

COMPARISON OF RETRIEVE DRIVE QUEUES

12 ROBOT CONFIGURATION



Total Drives ——— 36 48

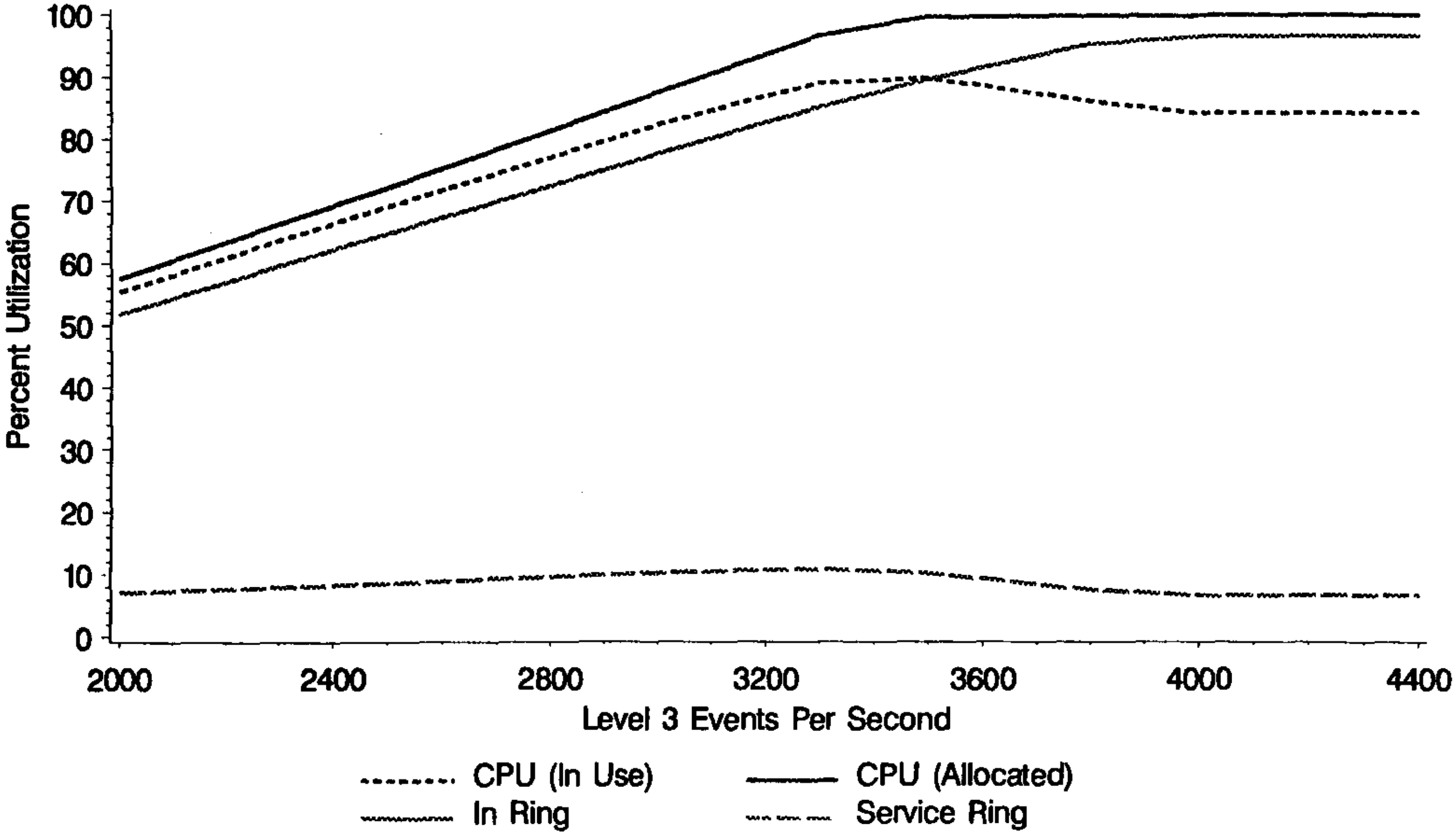
Drive Distribution: 12 Store Drives + 24 or 36 Retrieve Drives

40 CPU Row Configuration

- 40 CPU Rows
- 32 CPUs per Row
- 1280 Total CPUs
- 150 MB Control Processor Buffer containing no more than 33% Pass 2
- Control Processing Time of 0.001 seconds
- Effective Network Rates of 100 MB/s
- 12 Robots
- 3 Drives per Robot (1 for Stores and 2 for Retrieves)

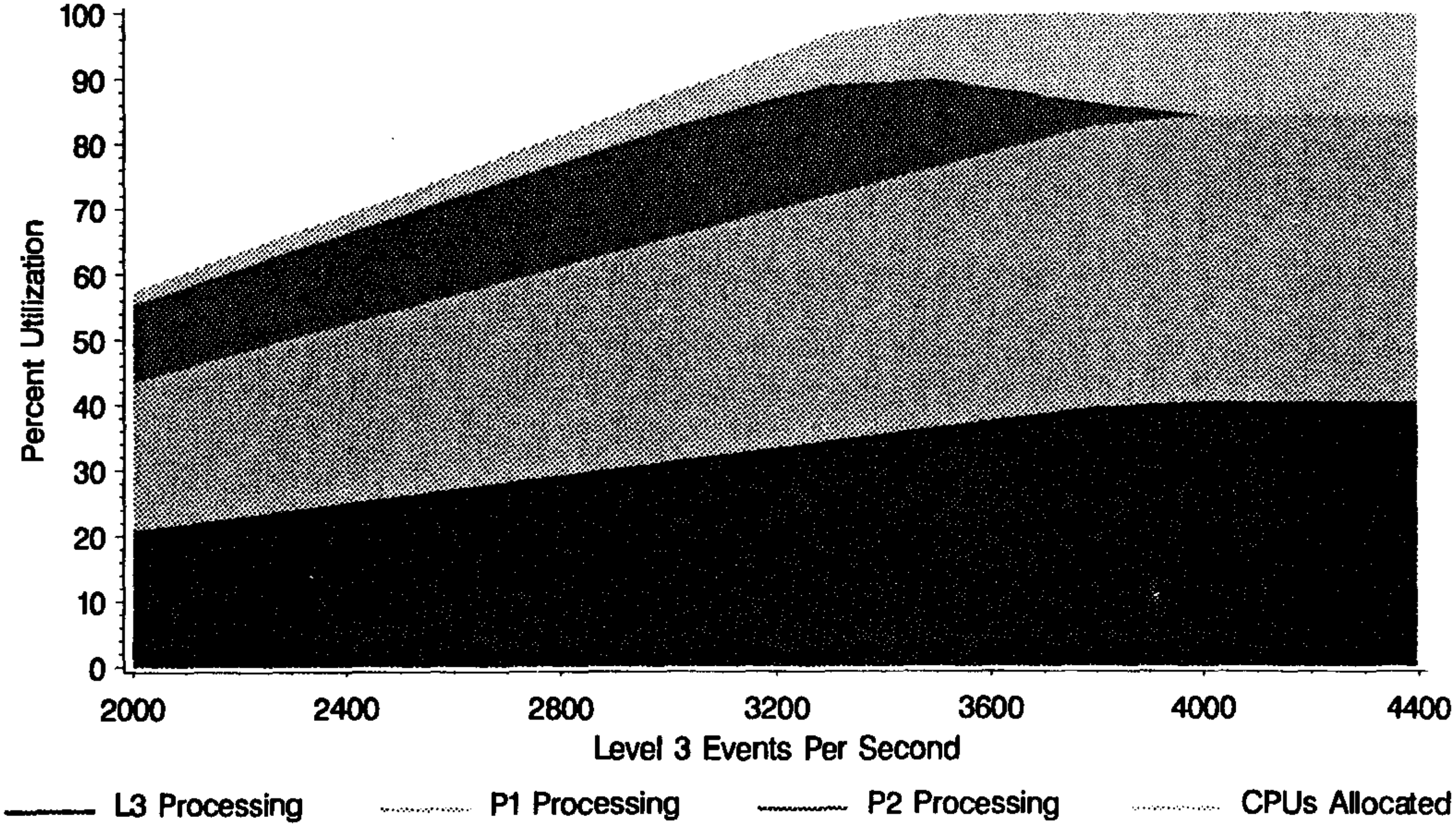
CPU AND NETWORK UTILIZATION

40 CPU ROWS



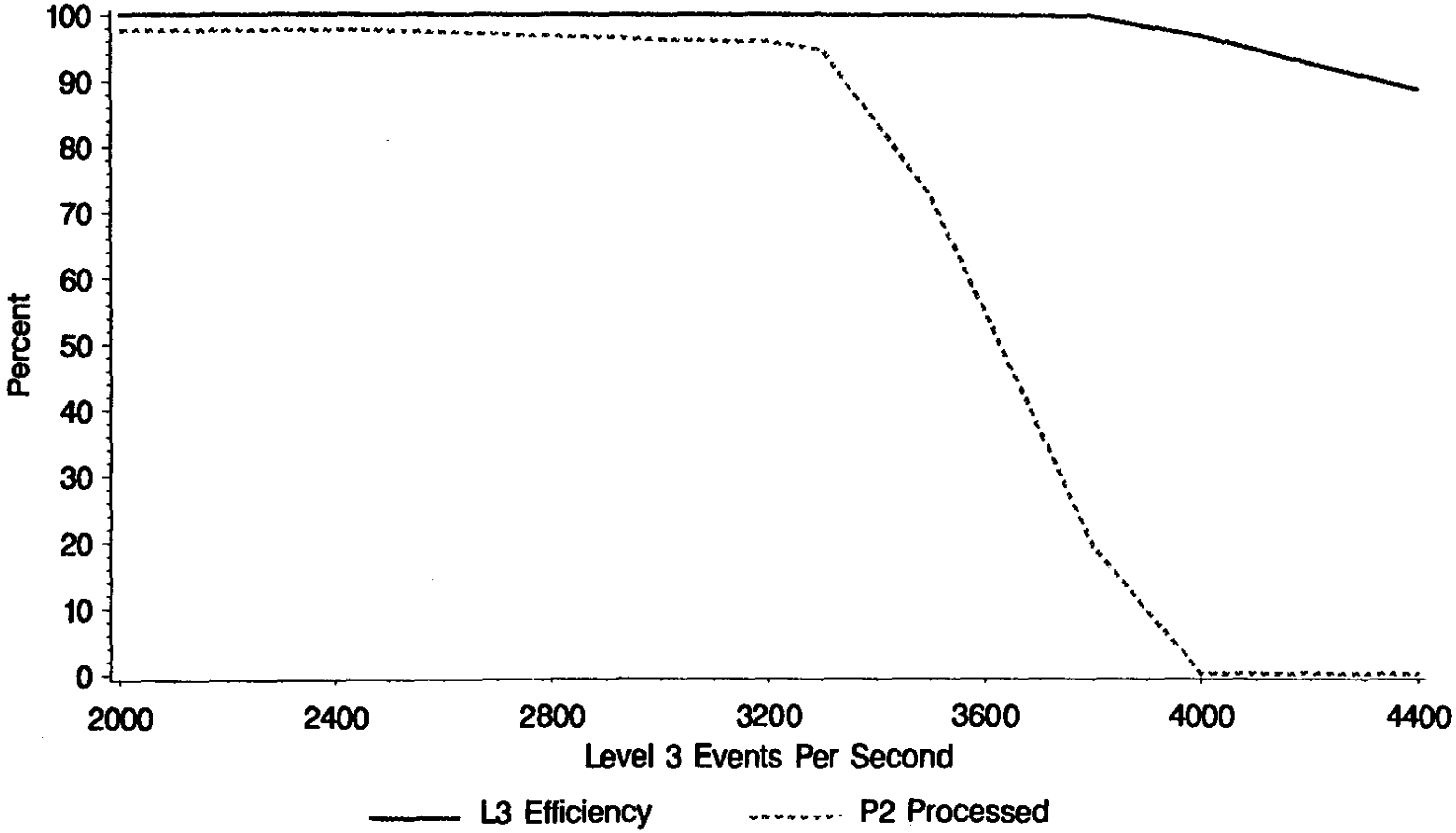
CPU UTILIZATION SUBDIVIDED BY EVENT PROCESSING

40 CPU ROWS



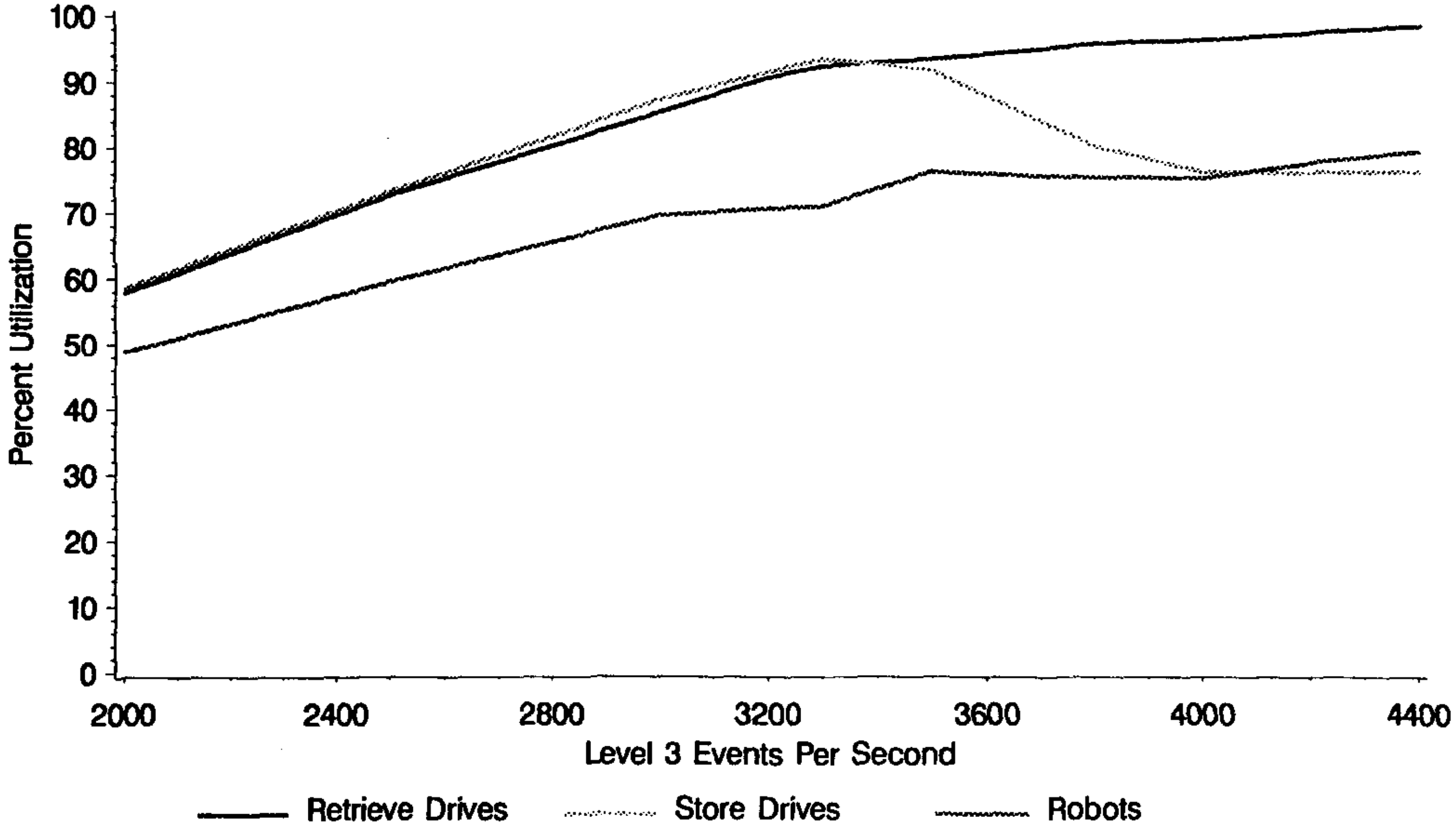
LEVEL 3 EFFICIENCY AND PASS 2 COMPLETION

40 CPU ROWS



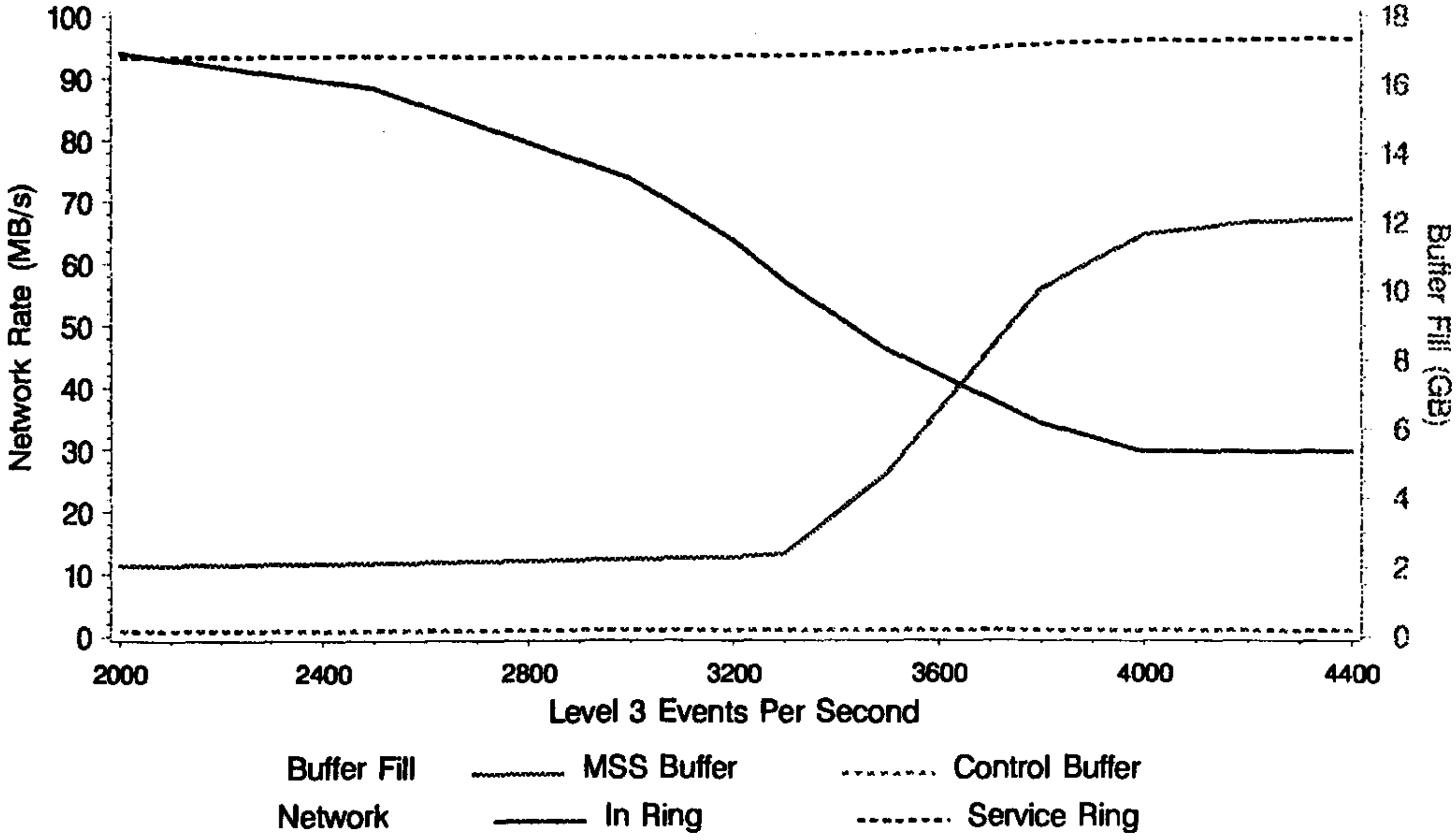
MSS RESOURCE UTILIZATION

40 CPU ROWS



NETWORK RATES AND BUFFER FILL

40 CPU ROWS

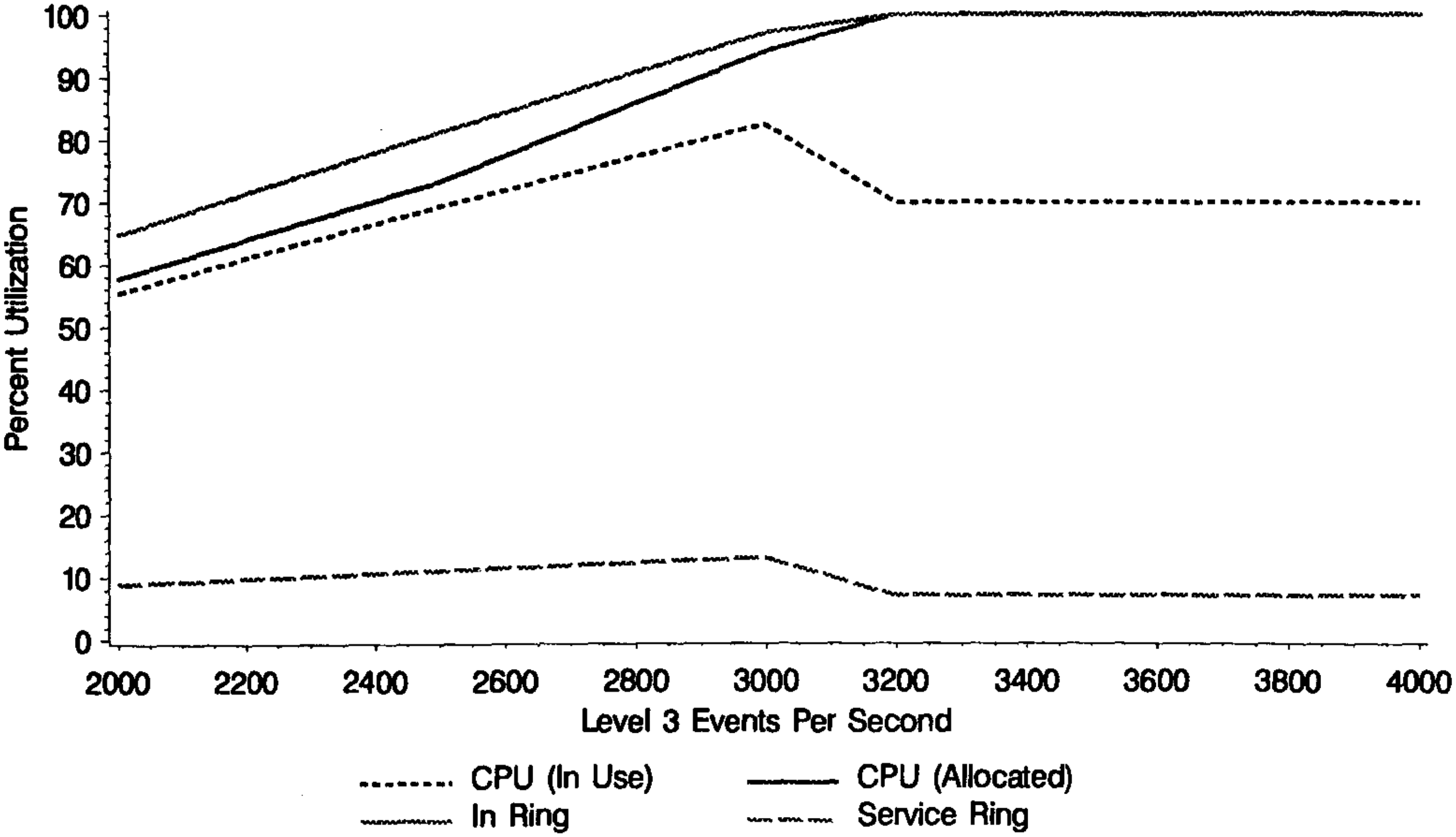


32 CPU Row Configuration

- 32 CPU Rows
- 40 CPUs per Row
- 1280 Total CPUs
- 150 MB Control Processor Buffer containing no more than 33% Pass 2
- Control Processing Time of 0.001 seconds
- Effective Network Rates of 100 MB/s
- 12 Robots
- 3 Drives per Robot (1 for Stores and 2 for Retrieves)
- Level 3 Processing - 67 SSCUPs

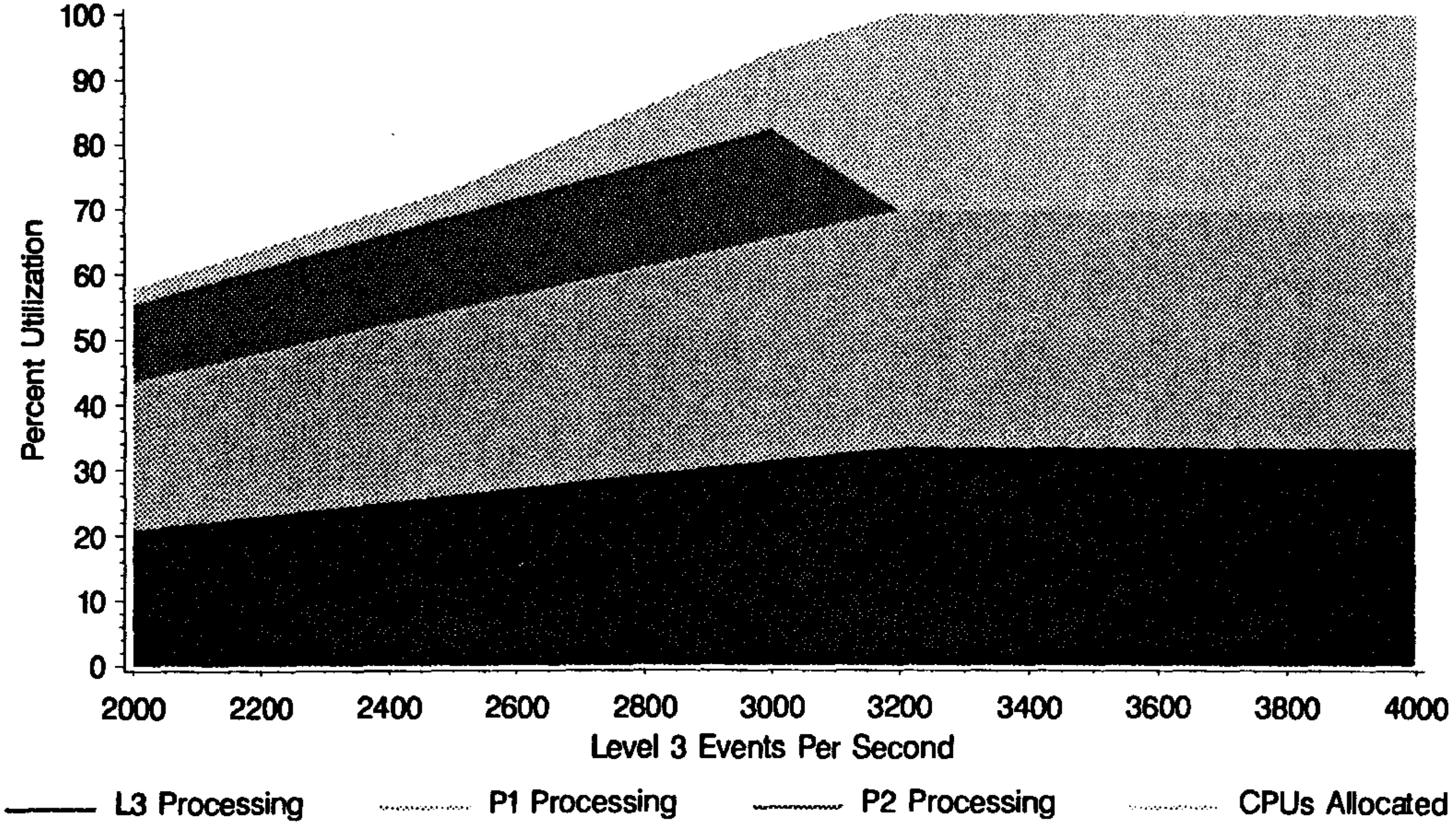
CPU AND NETWORK UTILIZATION

32 CPU ROWS



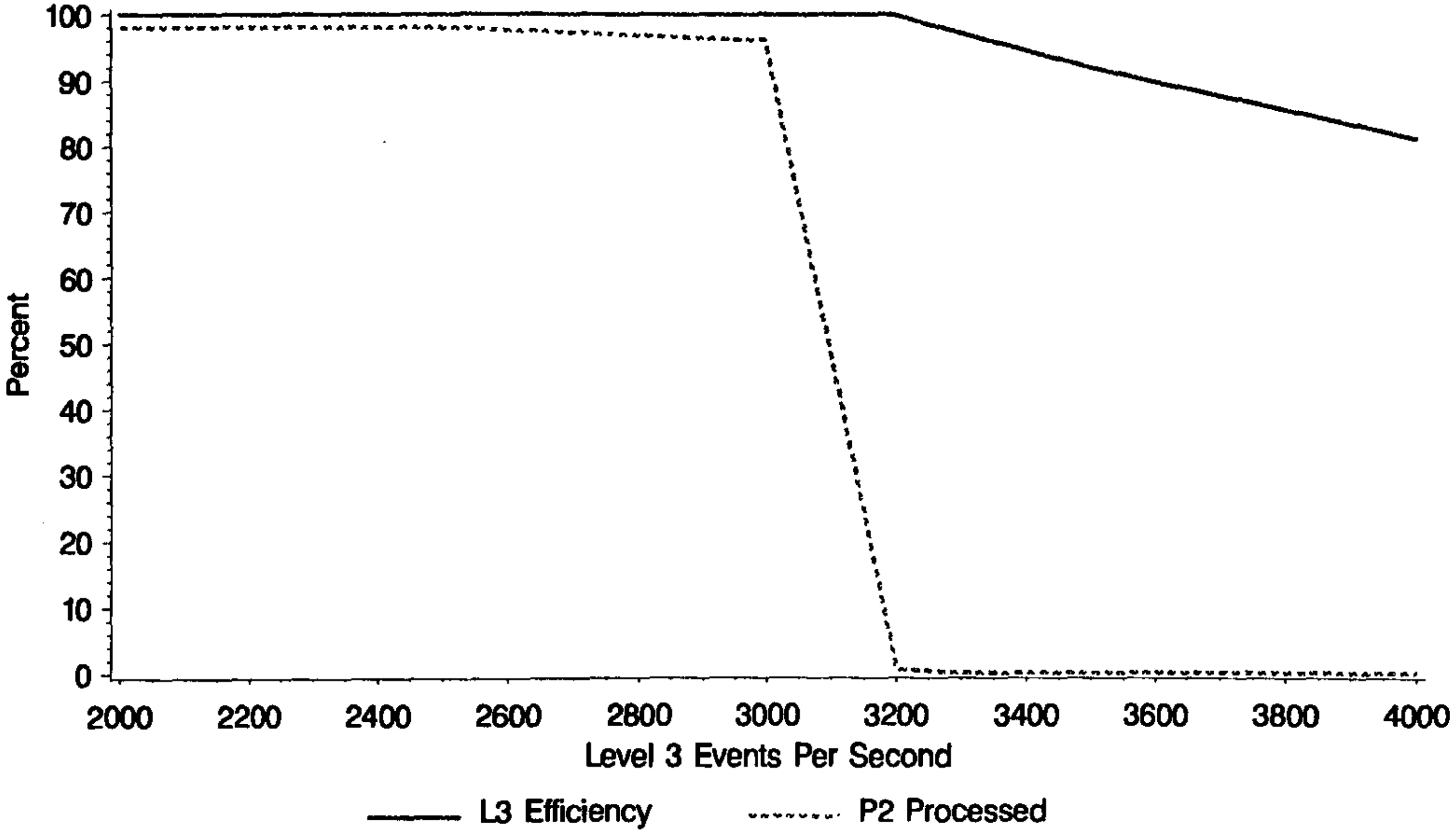
CPU UTILIZATION SUBDIVIDED BY EVENT PROCESSING

32 CPU ROWS



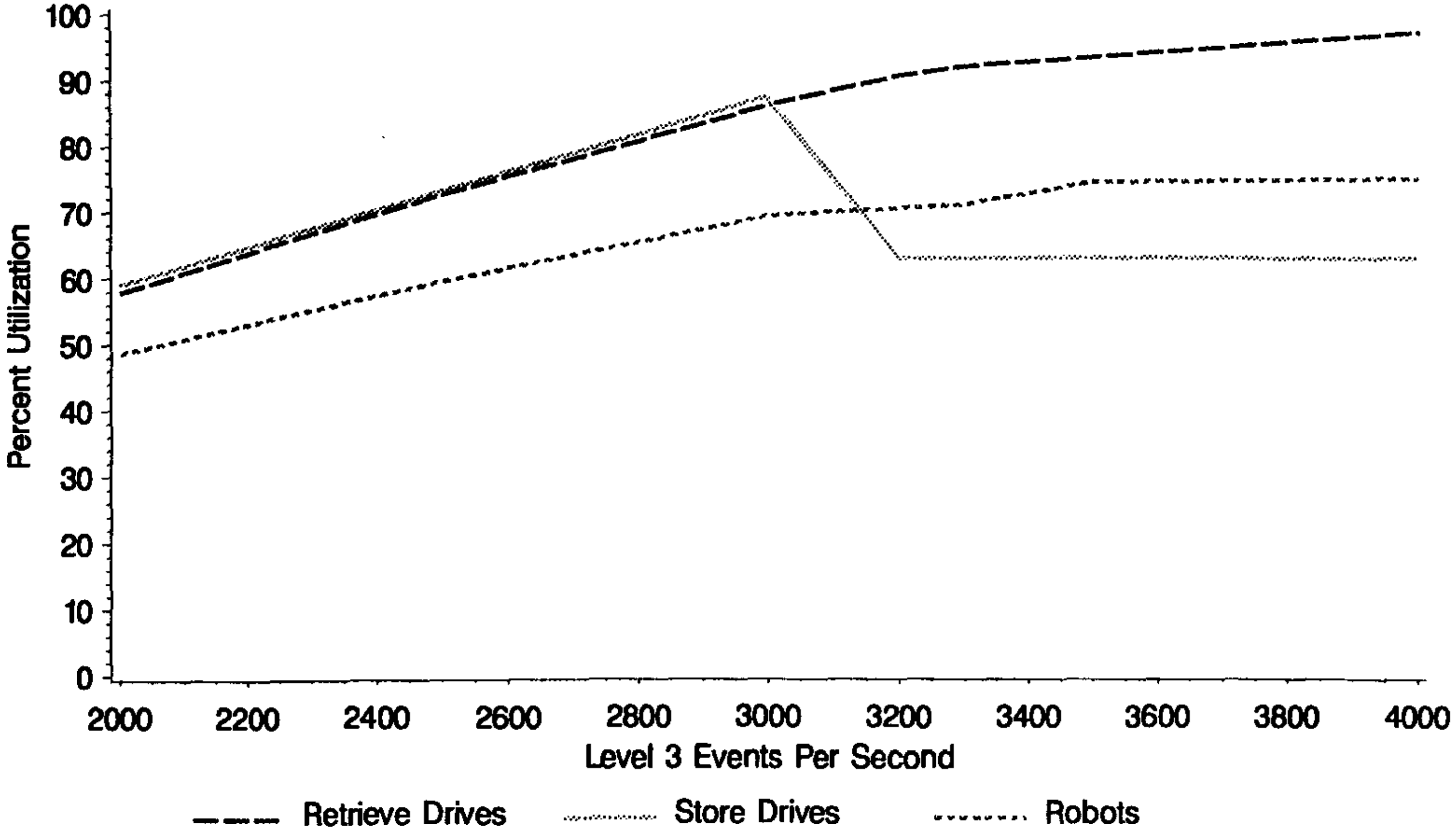
LEVEL 3 EFFICIENCY AND PASS 2 COMPLETION

32 CPU ROWS



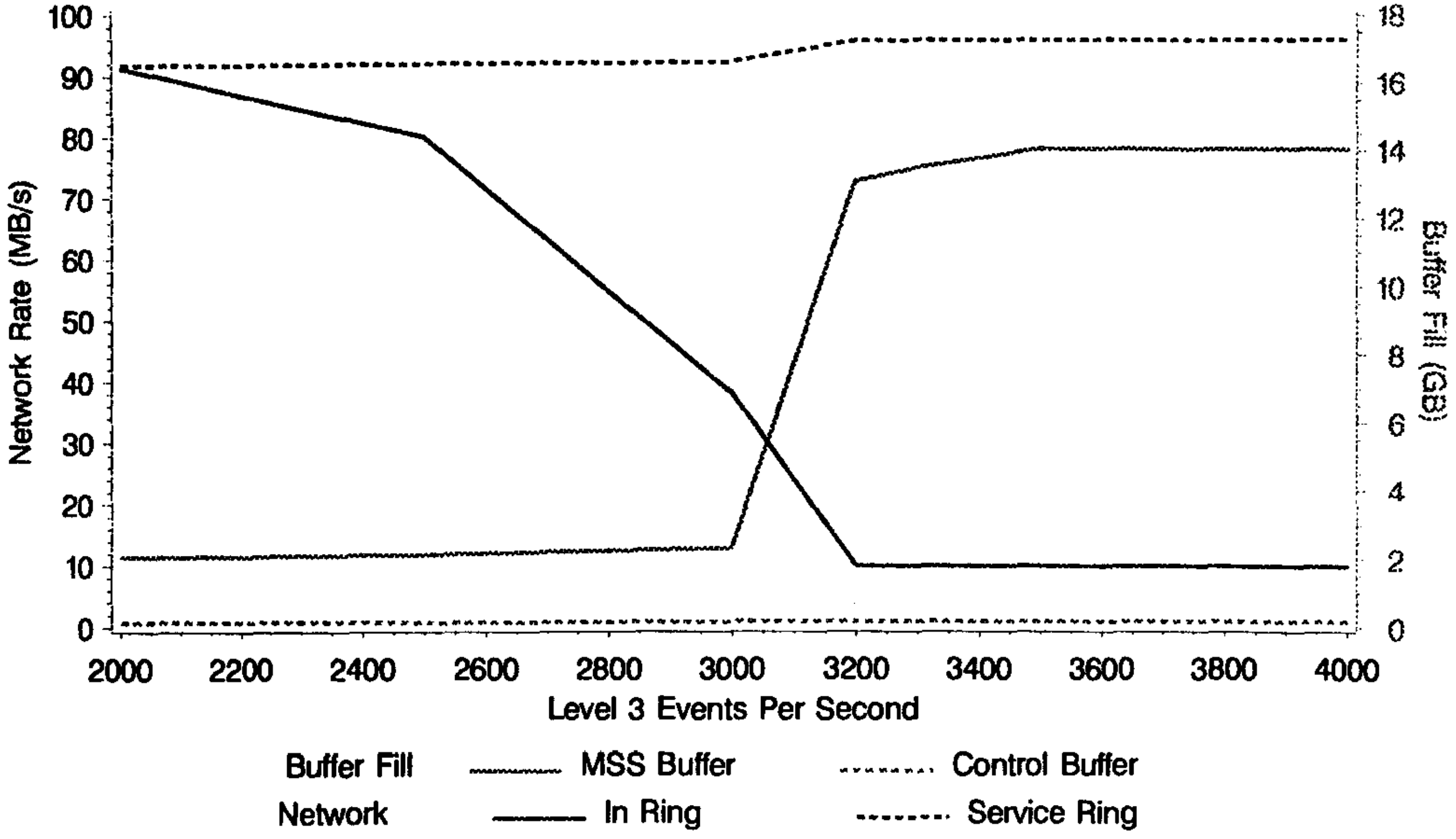
MSS RESOURCE UTILIZATION

32 CPU ROWS



NETWORK RATES AND BUFFER FILL

32 CPU ROWS

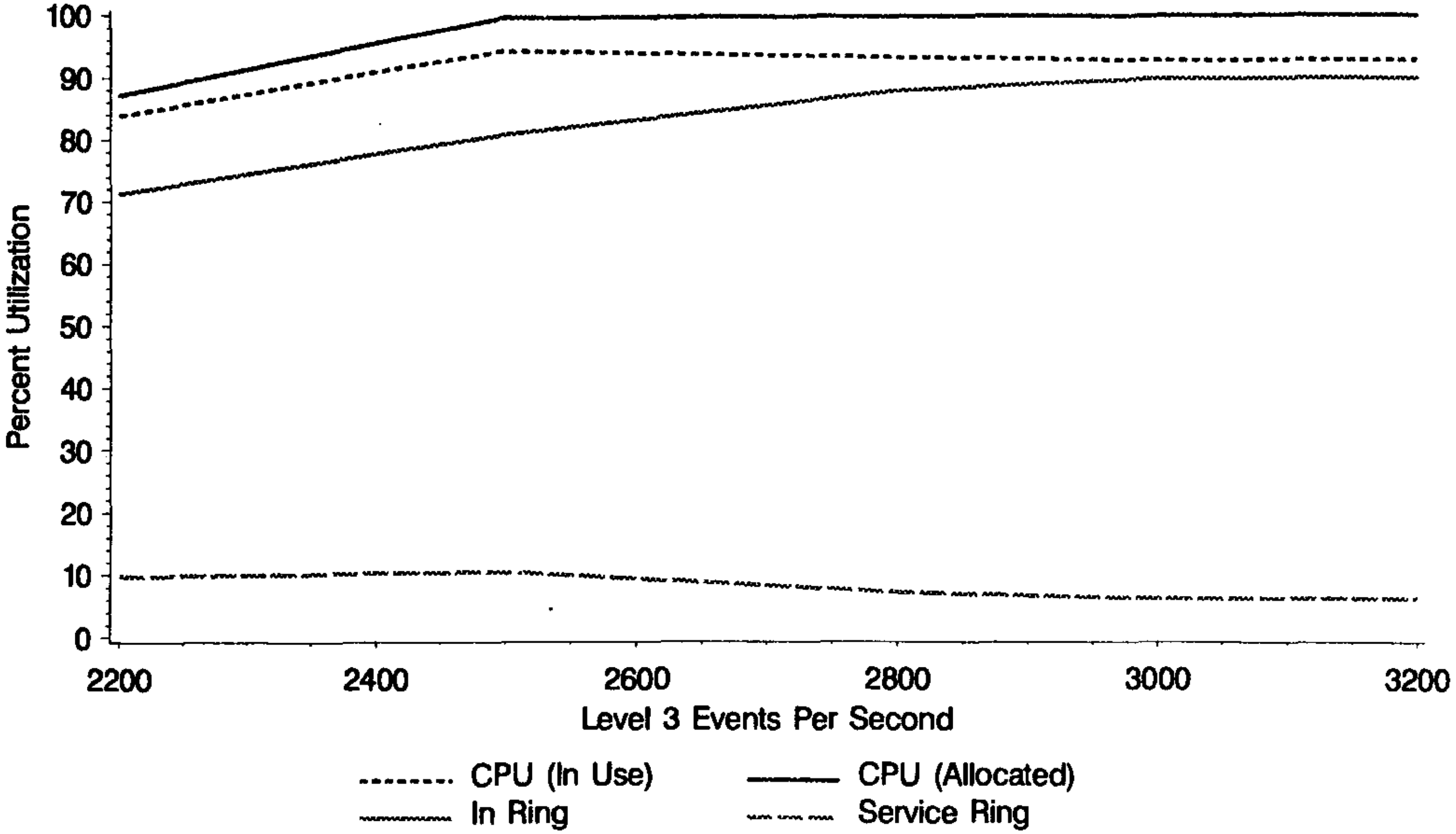


32 CPU Row Configuration
Level 3 Processing Time Doubled

- 32 CPU Rows
- 40 CPUs per Row
- 1280 Total CPUs
- 150 MB Control Processor Buffer containing no more than 33% Pass 2
- Control Processing Time of 0.001 seconds
- Effective Network Rates of 100 MB/s
- 12 Robots
- 3 Drives per Robot (1 for Stores and 2 for Retrieves)
- Level 3 Processing - 134 SSCUPs

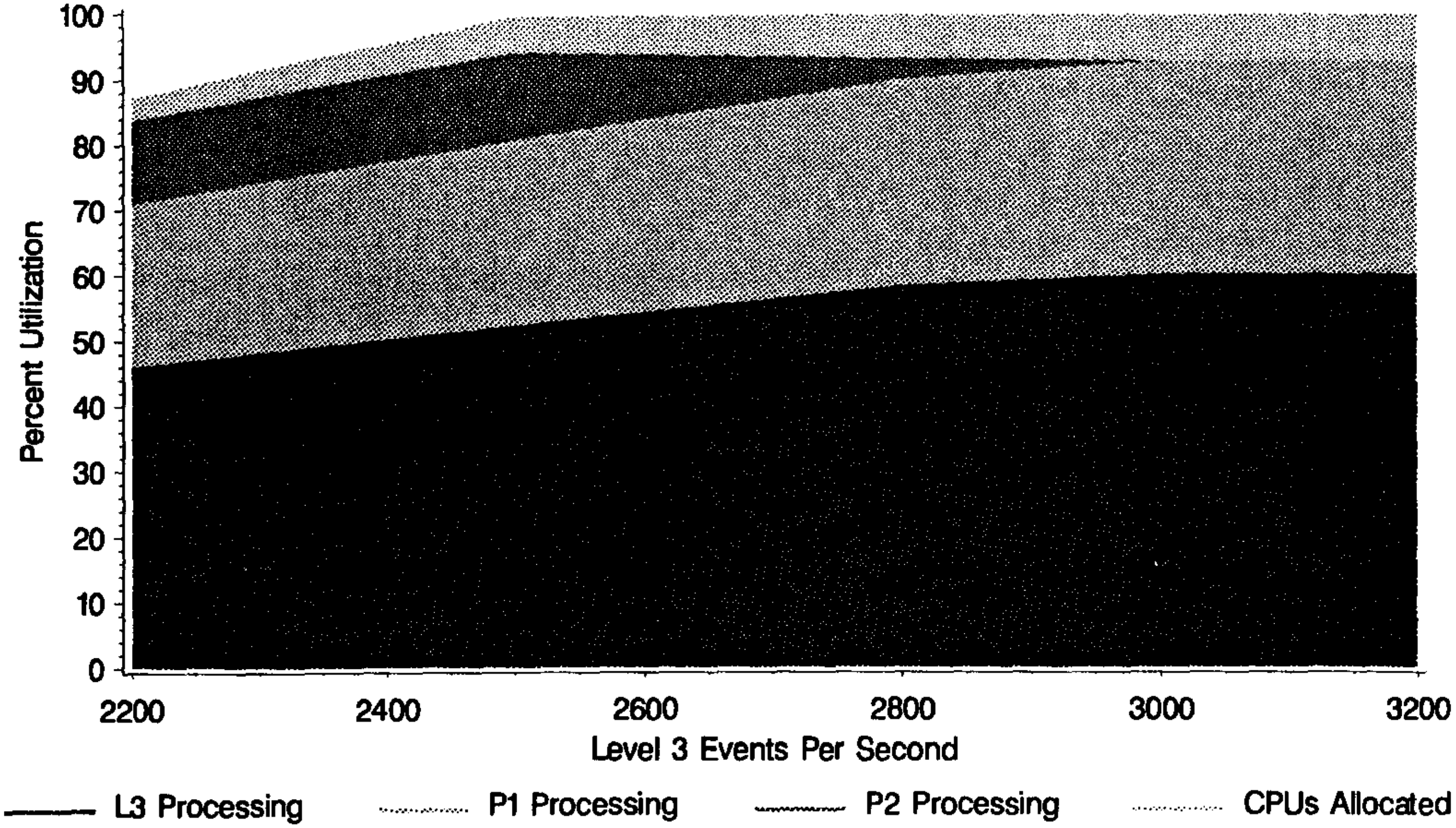
CPU AND NETWORK UTILIZATION

32 CPU ROWS (Level 3 Processing - 134 SSCUPs)



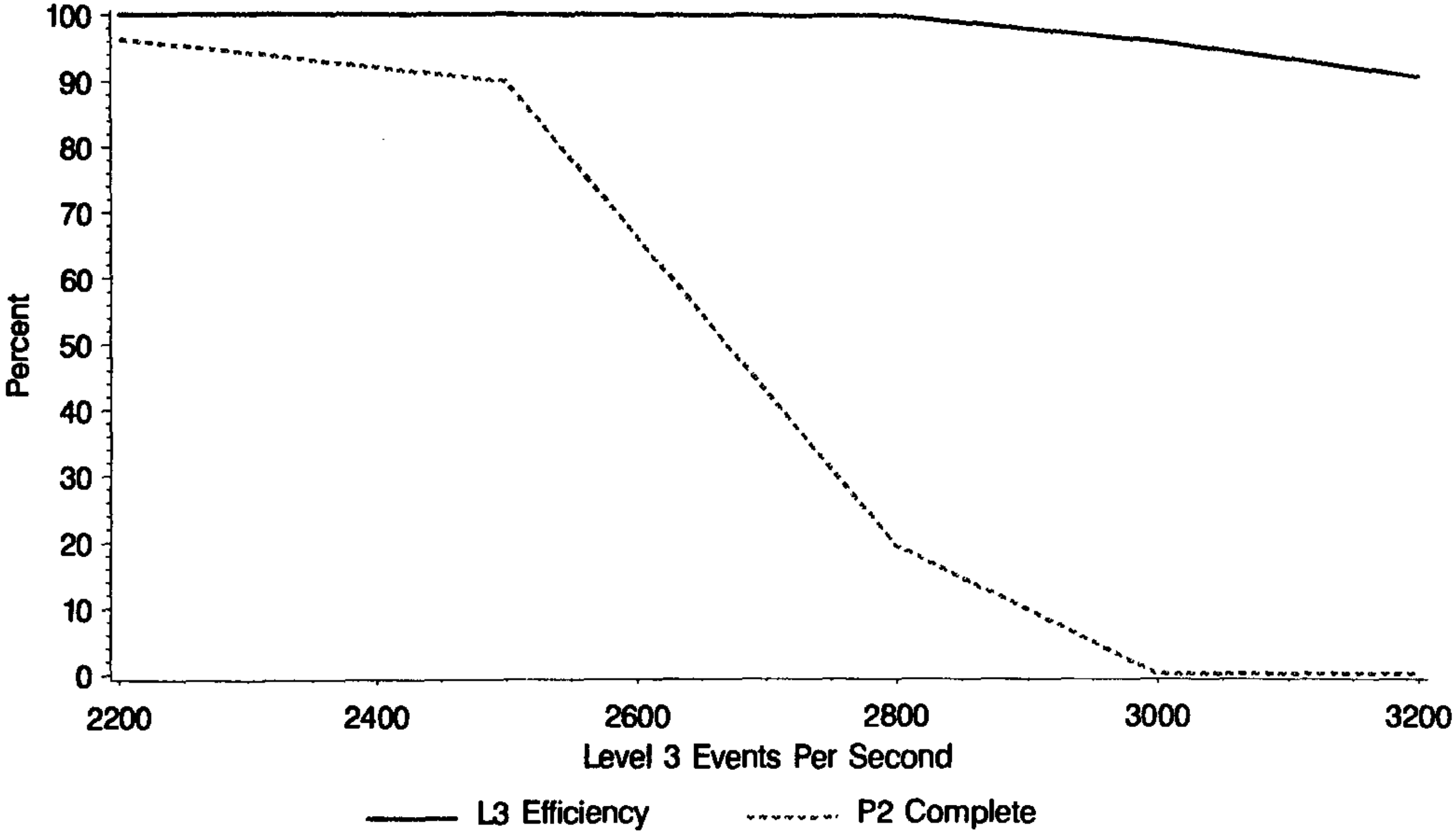
CPU UTILIZATION SUBDIVIDED BY EVENT PROCESSING

32 CPU ROWS (Level 3 Processing - 134 SSCUPs)



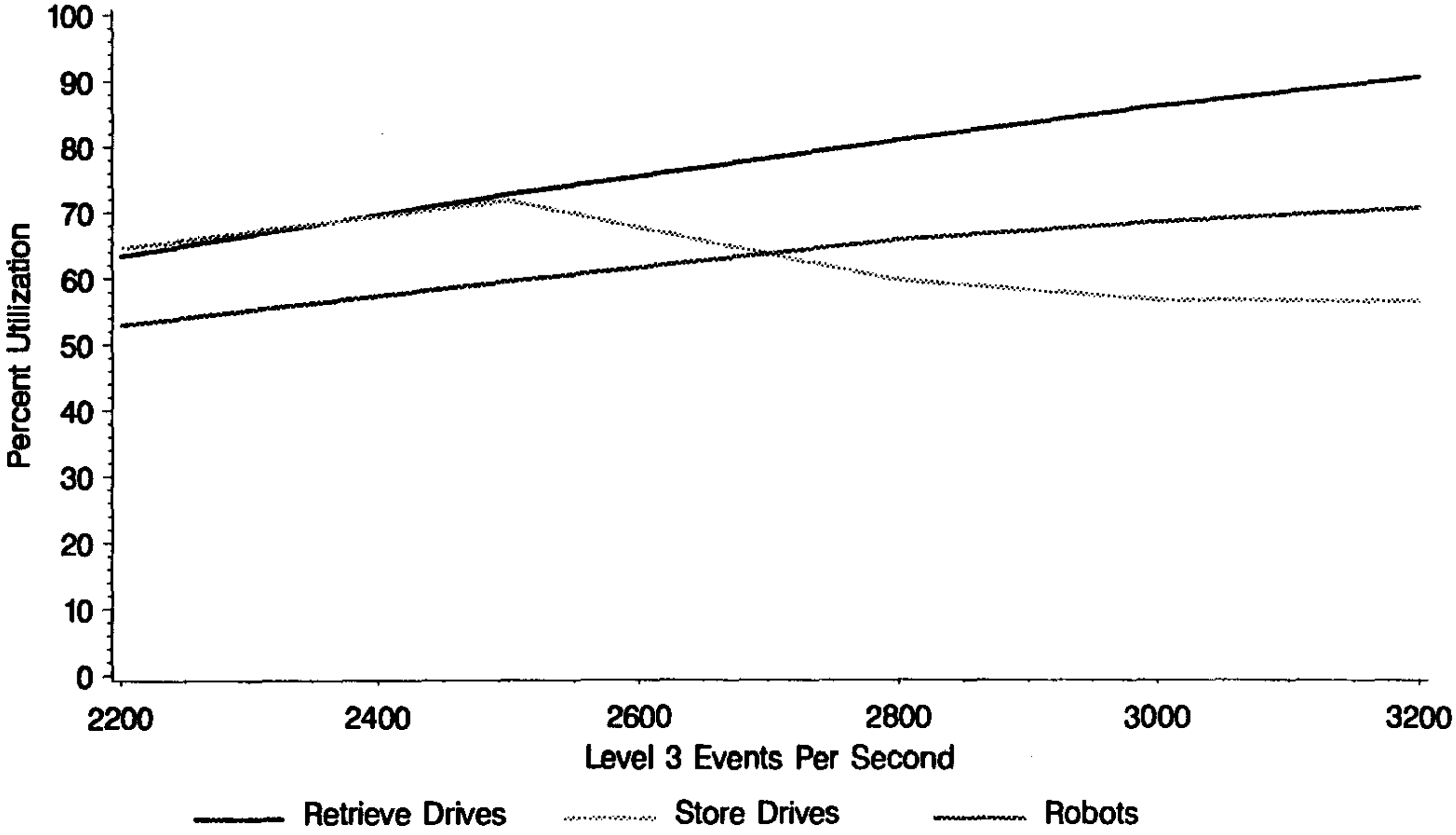
LEVEL 3 EFFICIENCY AND PASS 2 COMPLETION

32 CPU ROWS (Level 3 Processing - 134 SSCUPs)



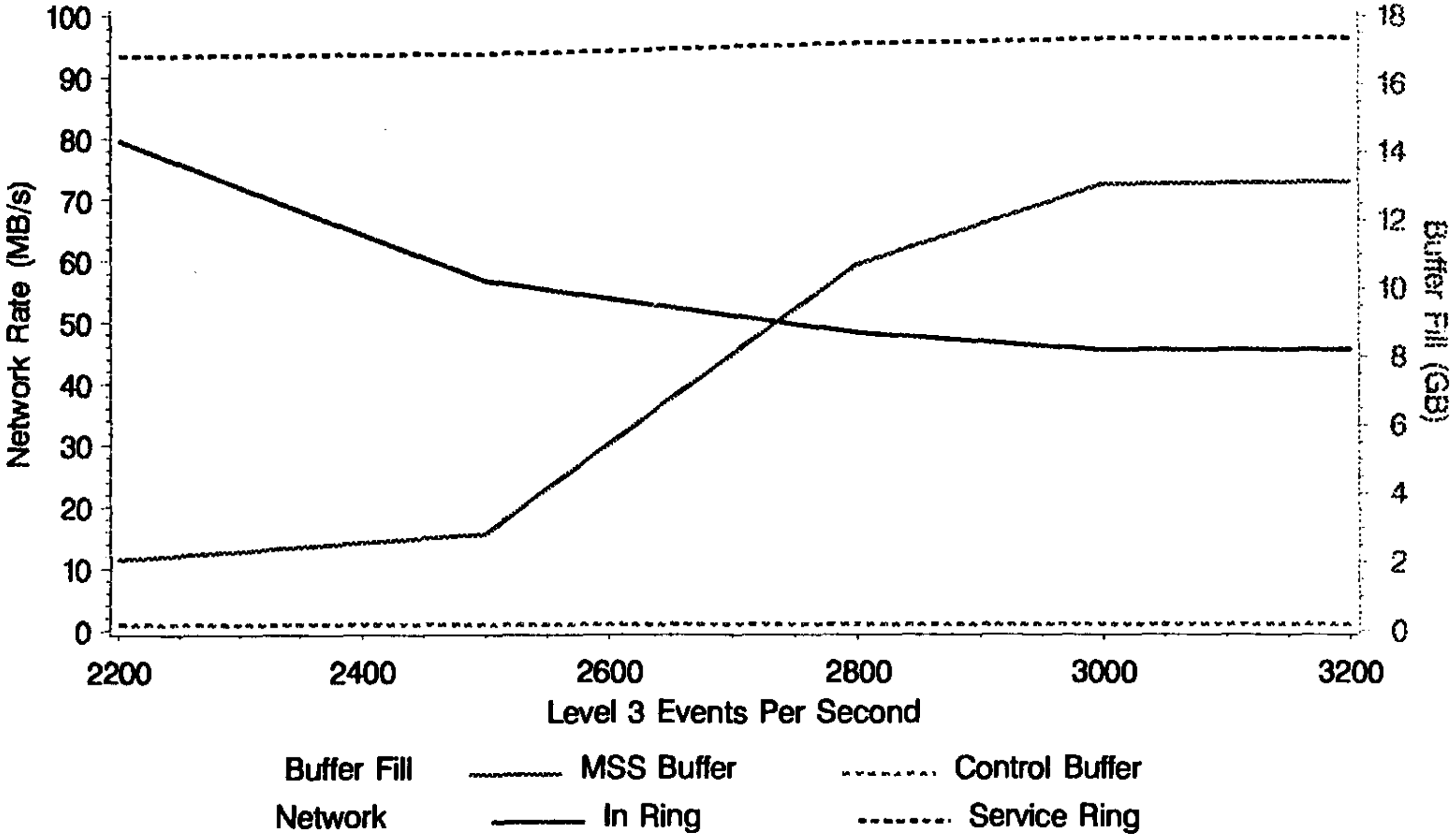
MSS RESOURCE UTILIZATION

32 CPU ROWS (Level 3 Processing - 134 SSCUPs)



NETWORK RATES AND BUFFER FILL

32 CPU ROWS (Level 3 Processing - 134 SSCUPs)

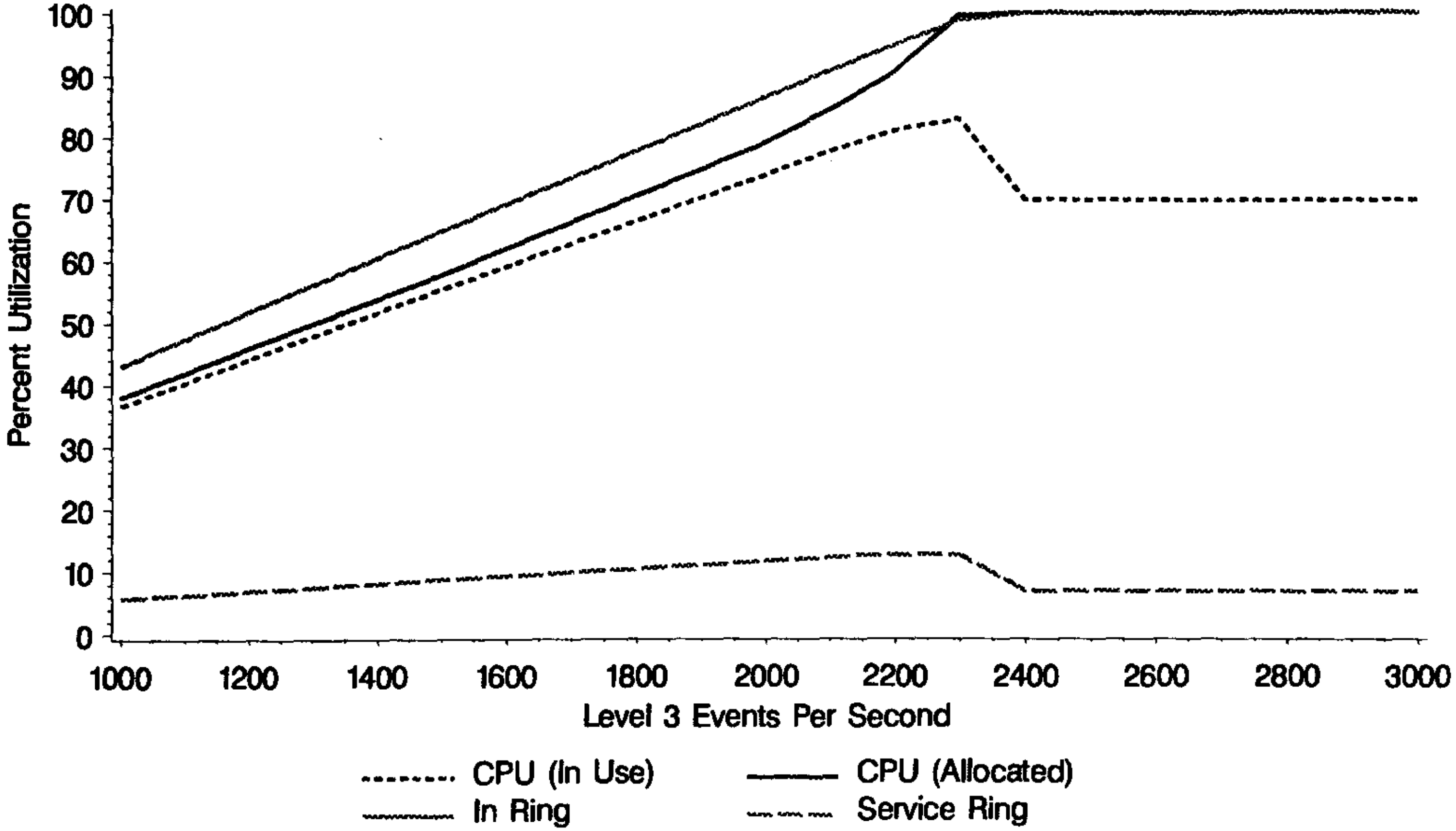


24 CPU Row Configuration

- 24 CPU Rows
- 40 CPUs per Row
- 960 Total CPUs
- 150 MB Control Processor Buffer containing no more than 33% Pass 2
- Control Processing Time of 0.001 seconds
- Effective Network Rates of 100 MB/s
- 12 Robots
- 3 Drives per Robot (1 for Stores and 2 for Retrieves)
- Level 3 Processing - 67 SSCUPs

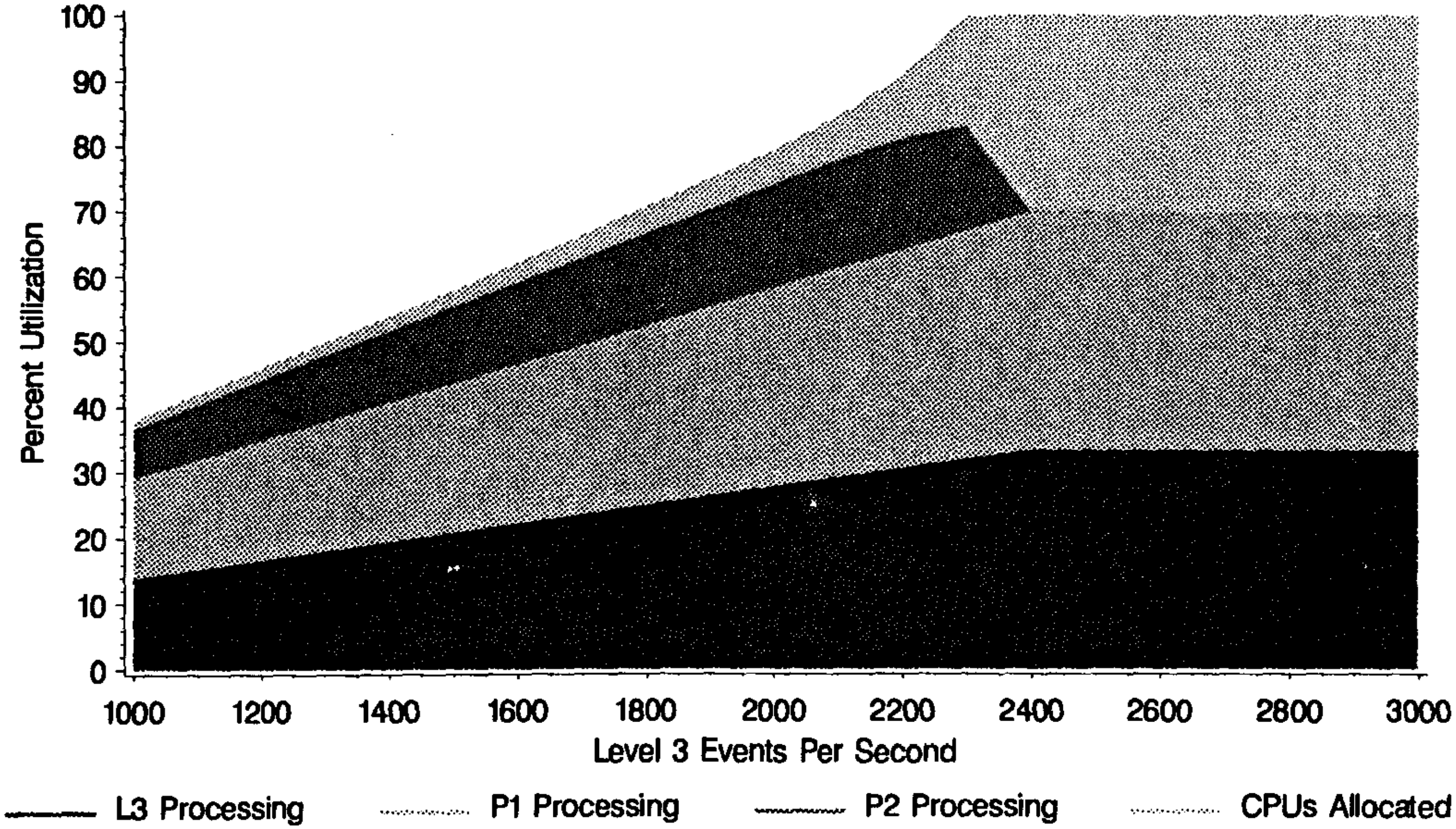
CPU AND NETWORK UTILIZATION

24 CPU ROWS



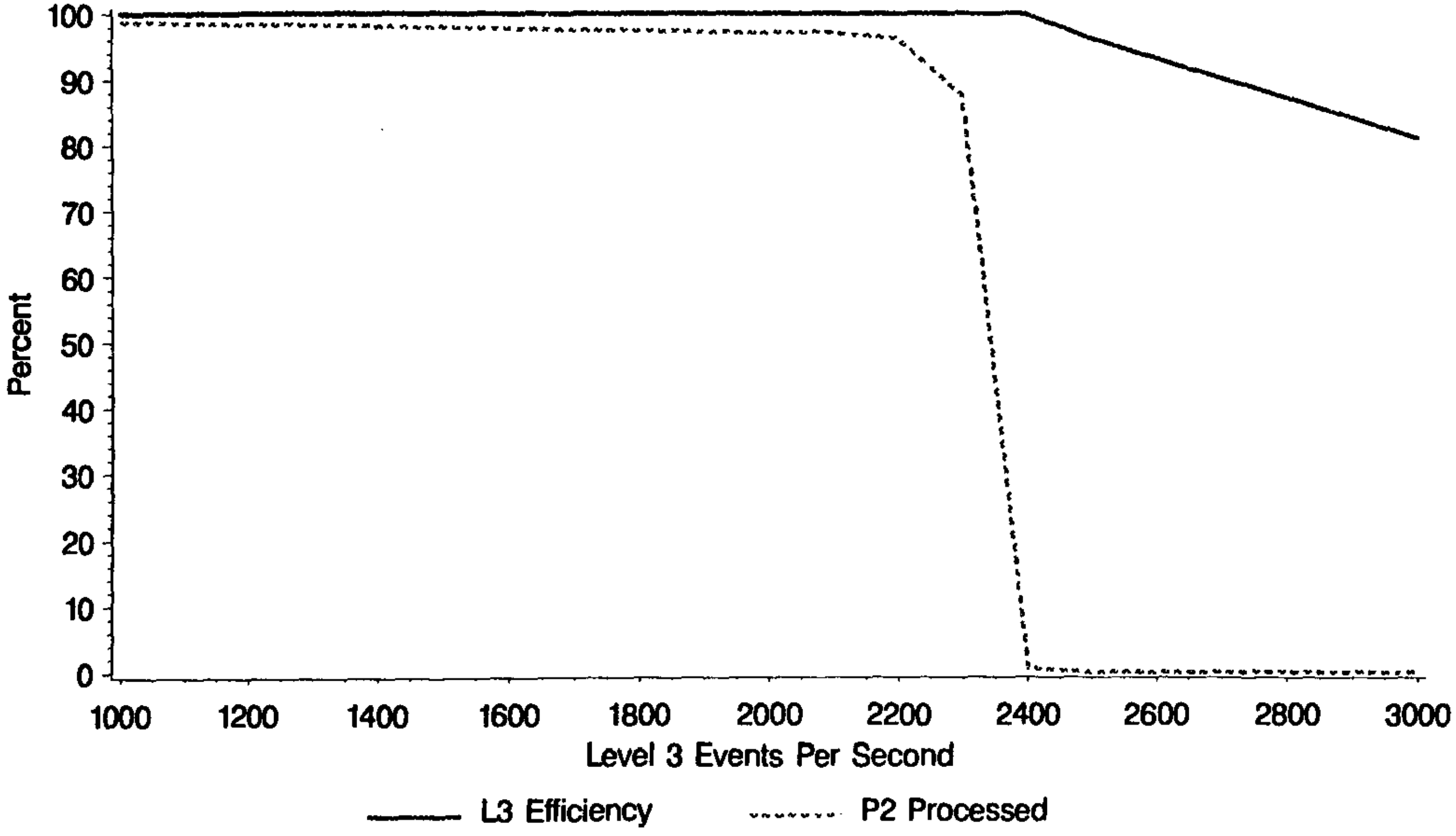
CPU UTILIZATION SUBDIVIDED BY EVENT PROCESSING

24 CPU ROWS



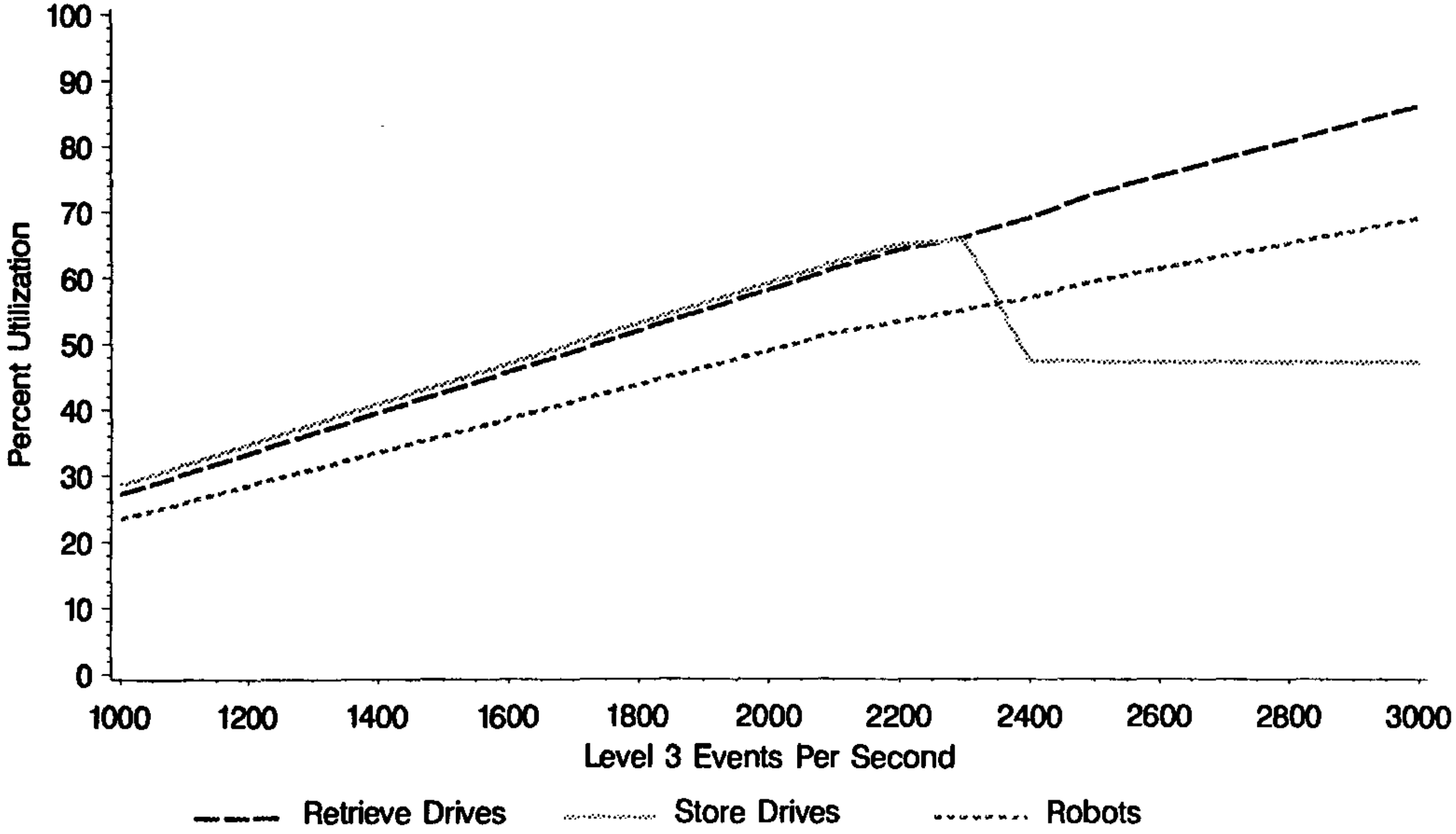
LEVEL 3 EFFICIENCY AND PASS 2 COMPLETION

24 CPU ROWS



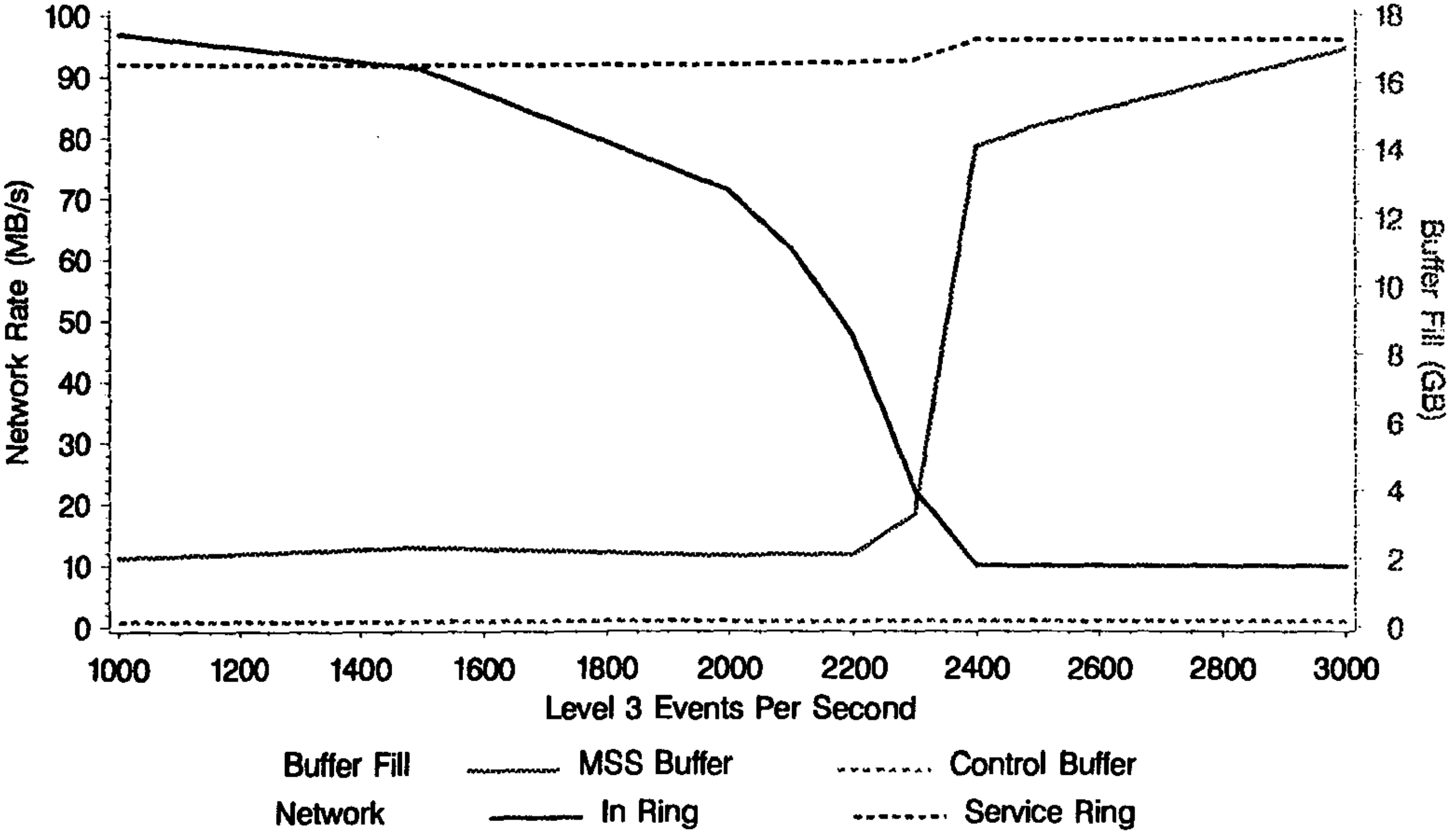
MSS RESOURCE UTILIZATION

24 CPU ROWS



NETWORK RATES AND BUFFER FILL

24 CPU ROWS

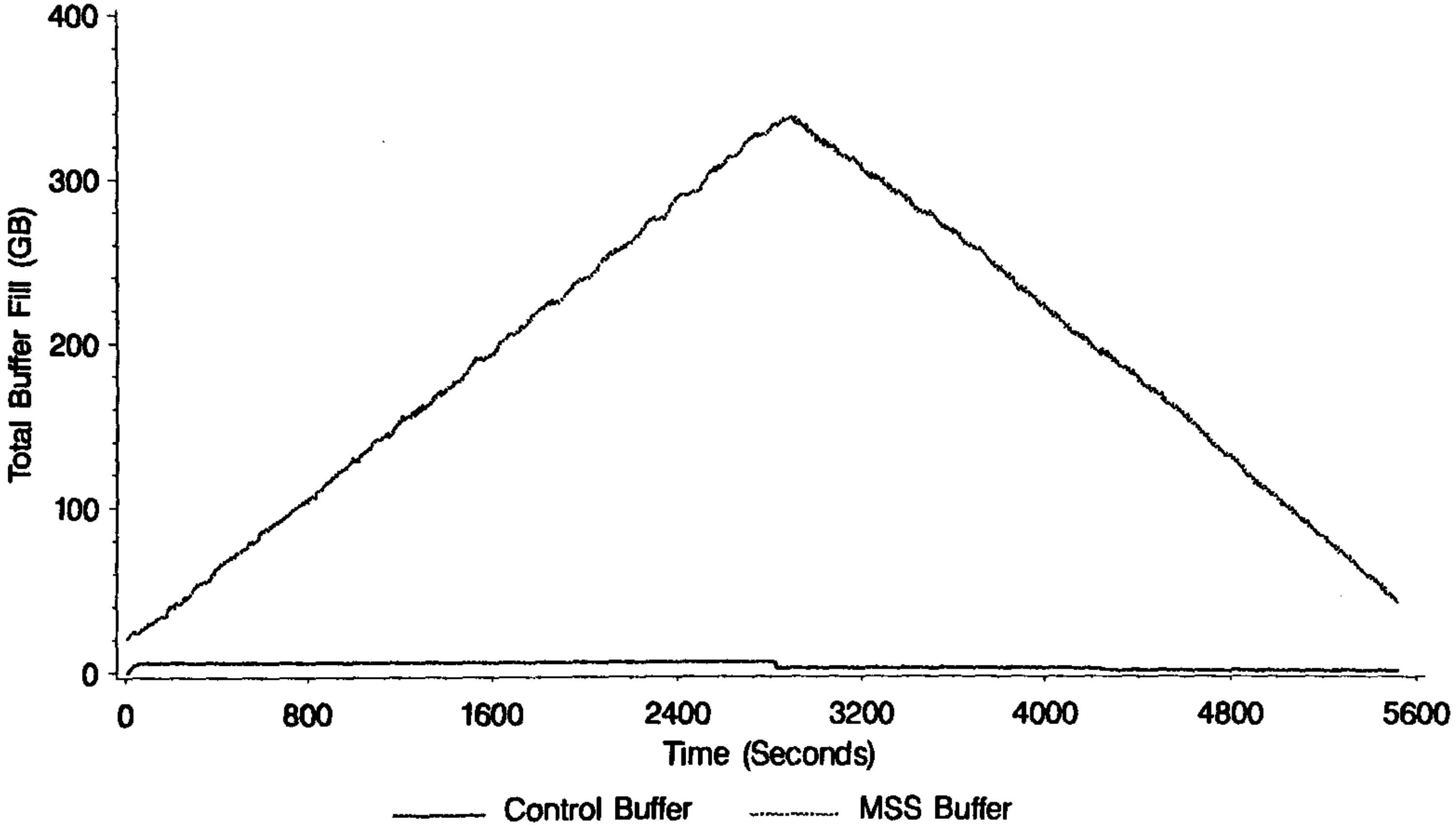


Trigger Rate Transition From
4000 MHz to 2000 MHz

- 40 CPU Rows
- 32 CPUs per Row
- 1280 Total CPUs
- 150 MB Control Processor Buffer containing no more than 33% Pass 2
- Control Processing Time of 0.001 seconds
- Effective Network Rates of 100 MB/s
- 12 Robots
- 3 Drives per Robot (1 for Stores and 2 for Retrieves)
- Level 3 Processing - 67 SSCUPs
- Changed trigger rate after 45 minutes

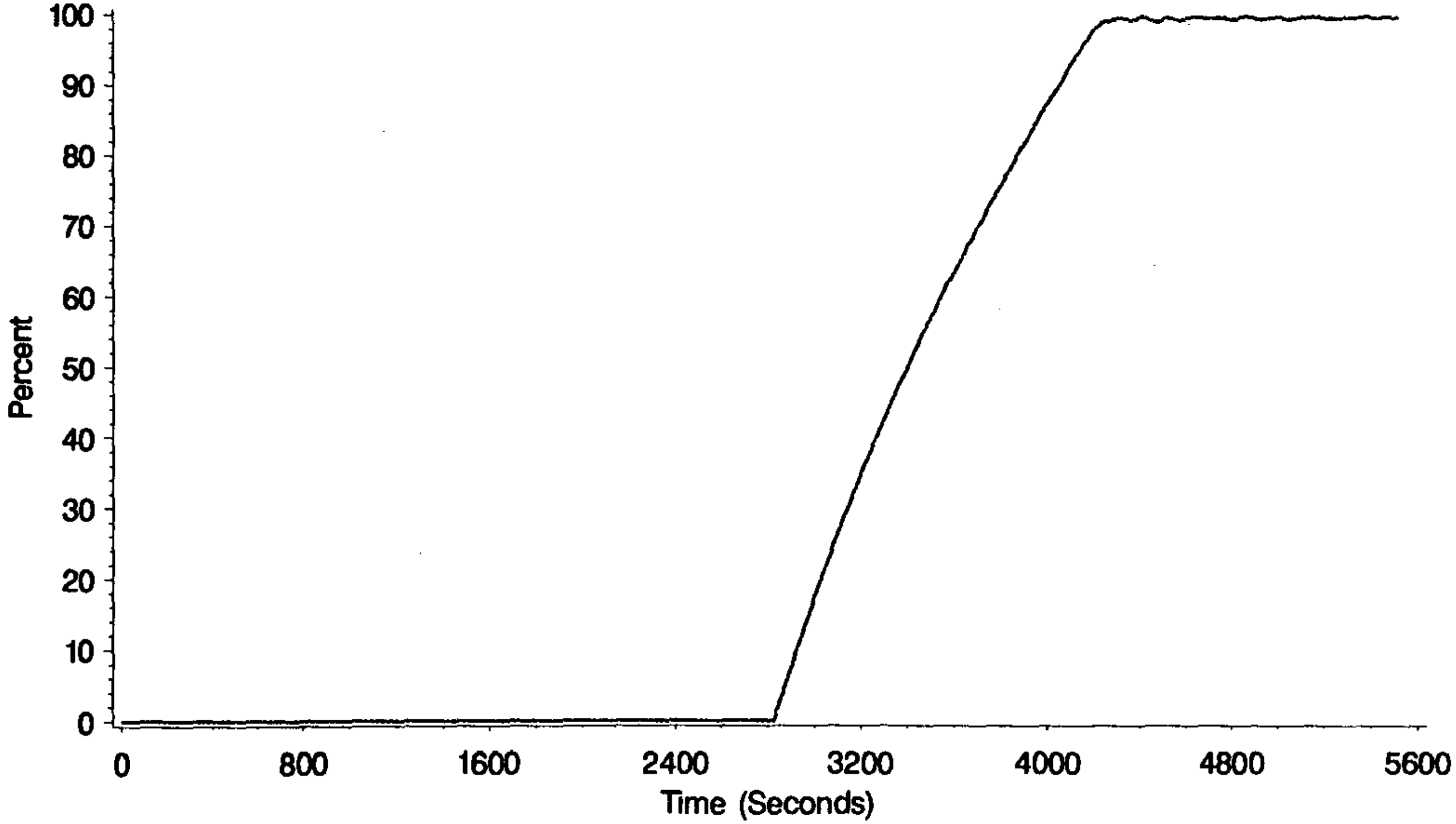
MSS BUFFER FILL OVER TIME

4000 MHz to 2000 MHz



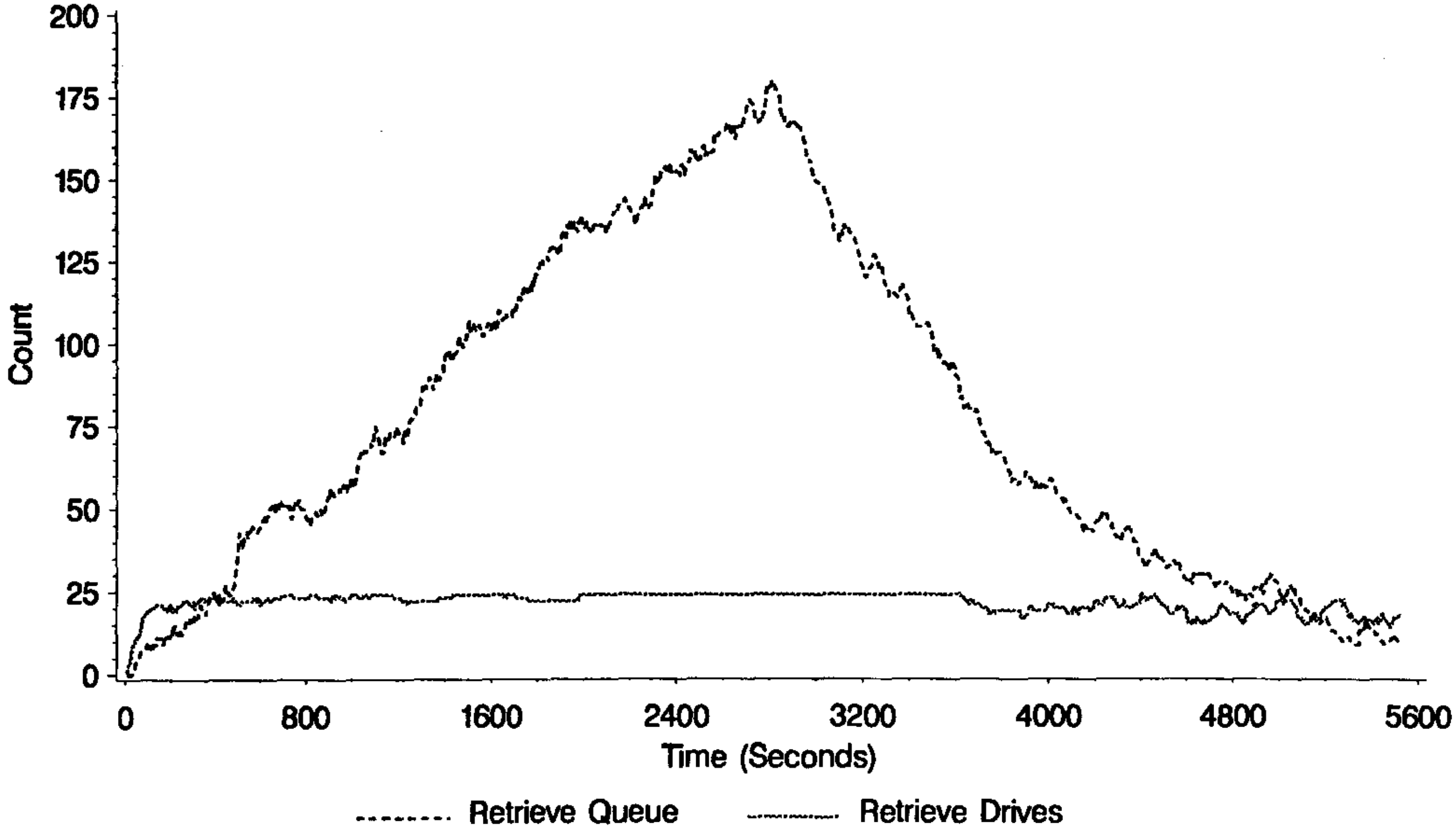
PASS 2 COMPLETION OVER TIME

4000 MHz to 2000 MHz



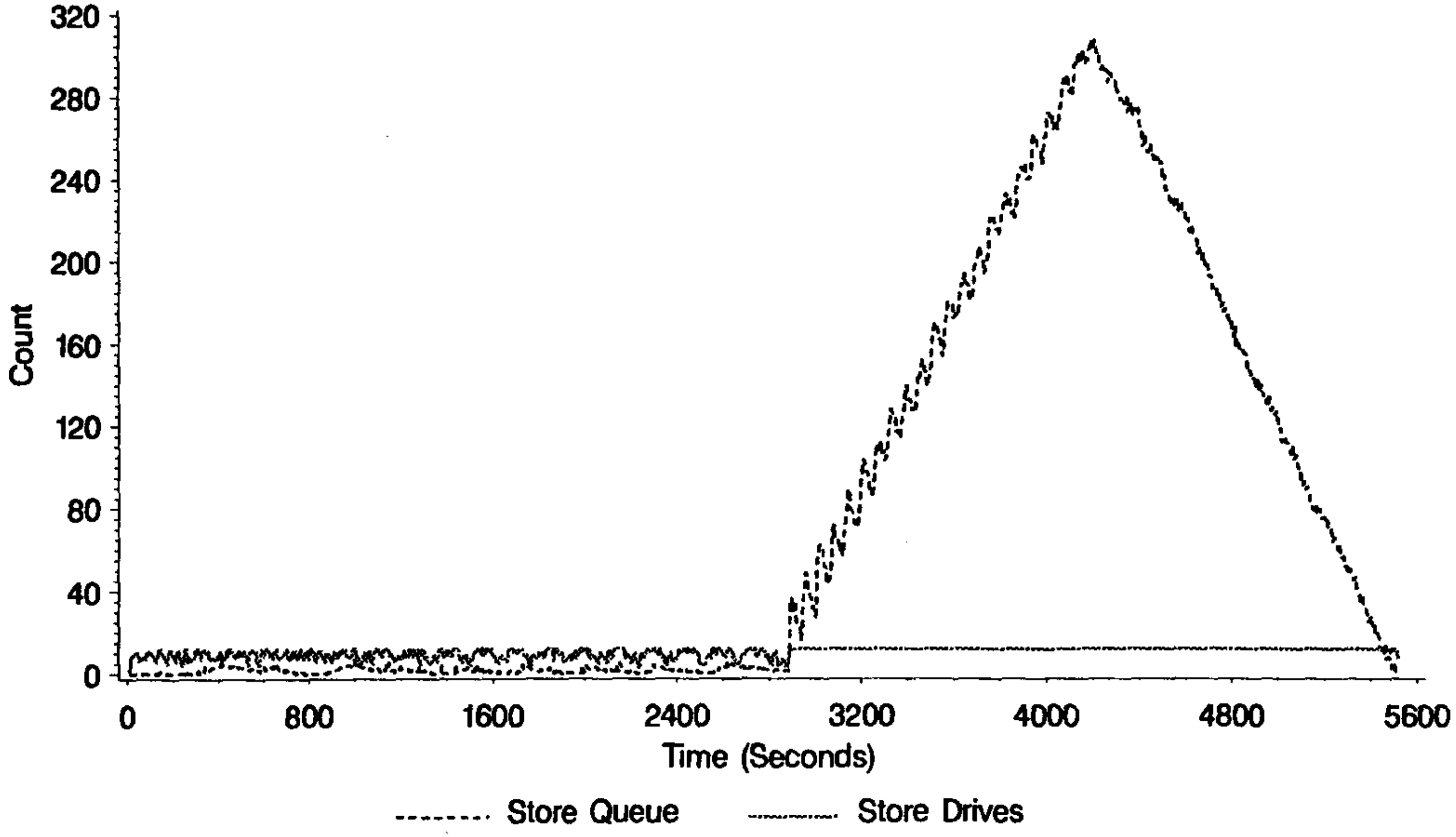
RETRIEVE DRIVE UTILIZATION

4000 MHz to 2000 MHz



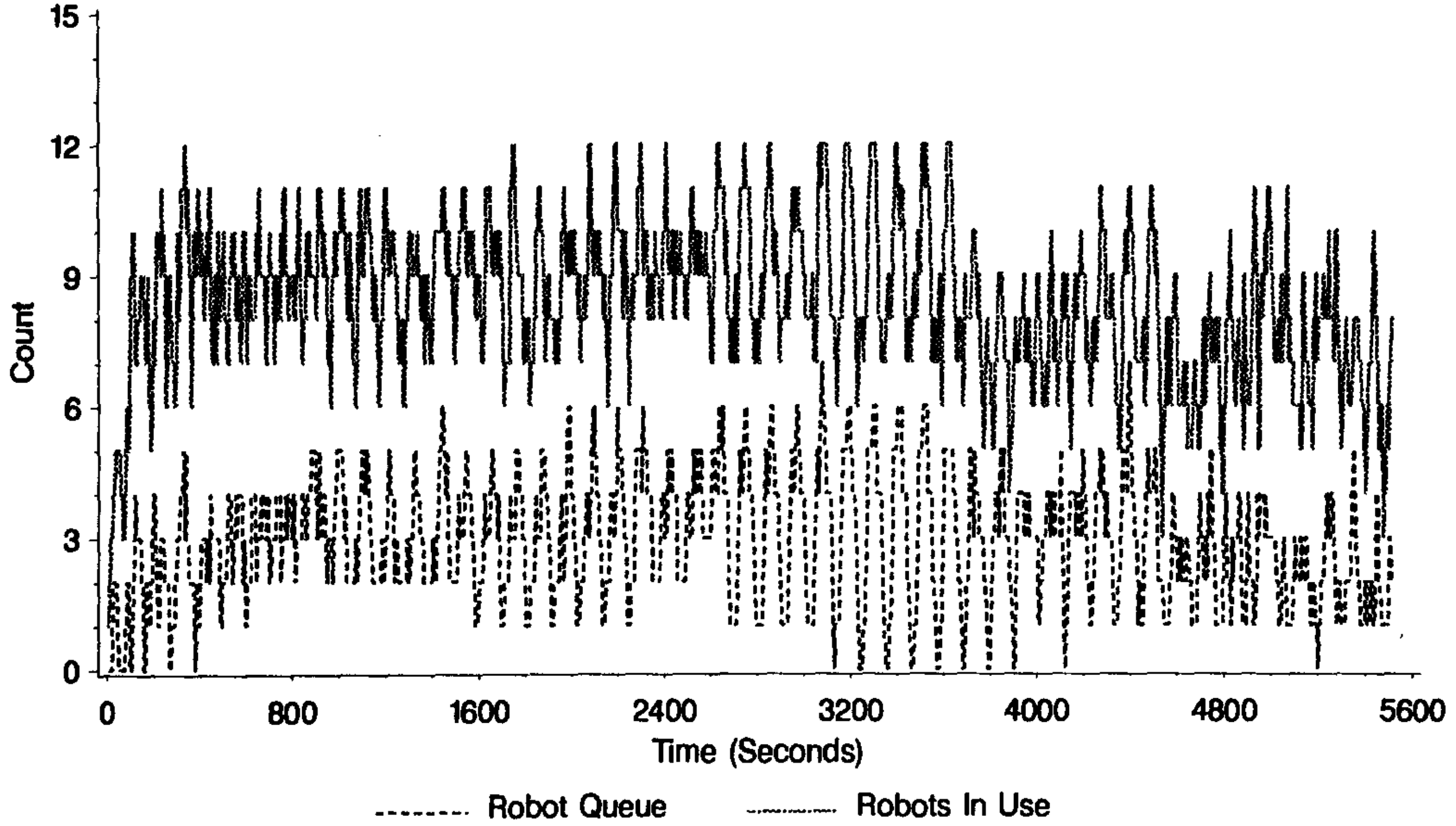
STORE DRIVE UTILIZATION

4000 MHz to 2000 MHz



ROBOT UTILIZATION

4000 MHz to 2000 MHz


Count
Time (Seconds)
-------- Robot Queue
———— Robots In Use