WHICH ONE IS RESPONSIBLE FOR WIP IN MANUFACTURING: THE WORKSTATIONS OR THE MATERIAL HANDLING SYSTEM?

Mandyam M. Srinivasan

Yavuz A. Bozer

Department of Industrial & Operations Engineering
University of Michigan
Ann Arbor, MI 48109-2117

November 1990

Technical Report 90-28

WHICH ONE IS RESPONSIBLE FOR WIP IN MANUFACTURING: THE WORKSTATIONS OR THE MATERIAL HANDLING SYSTEM?†

Mandyam M. Srinivasan

Yavuz A. Bozer

Department of
Industrial and Operations Engineering
The University of Michigan
Ann Arbor, MI 48109, USA

ABSTRACT

The answer to the above question may seem obvious: whichever of the two is the bottleneck, and if there is none, whichever of the two has longer "service times" and/or larger variation in "service times." As we empirically show in this study, the answer is not quite so obvious. In fact, our simulation results suggest that, as long as there are no bottlenecks in the system, the material handling system should almost never be responsible for work-in-process even if it has "service time" parameters that are comparable to processing times at the workstations. In the paper we first show the simulation results, and subsequently we provide further insight based on well-known analytical results in queueing theory. We also discuss certain properties of the workstations and the handling system that affect the work-in-process levels.

1. INTRODUCTION

In this paper we address a fundamental issue which, to our knowledge, has not been studied before at the conceptual level in a general setting. Namely, given a manufacturing system that consists basically of two components, that is, a set of workstations and a "trip-based" material handling system, we are interested in identifying the component responsible for work-in-process (WIP). In other words, we are interested in studying where the queues develop and the reasons behind it, assuming that neither component is a bottleneck.

A number of researchers have examined the behavior of WIP in manufacturing systems. However, such studies tend to either oversimplify (or disregard) the material handling system and focus on the workstations, or they concentrate only on the handling system and disregard the workstations. For example, one may obtain the expected queue

---

†Dr. Bozer is supported by a Presidential Young Investigator Award under NSF Grant DMC 8858562.

lengths by the models presented by Yao and Buzacott [12,13], and Toro Ramos and McGinnis [11], where the material handling system is modeled as a central server. With a central server model, the empty device dispatching rule — which is known to affect the throughput capacity of the handling system (see [2] and [8] among others) — is not explicitly represented in the model. Furthermore, the service time parameters for the handling system are expressed only as an overall distribution. That is, when a device travels from one workstation to another, the specific travel time is not directly used in the model. In some other studies that fall into this category, the material handling system is simply not considered when studying the distribution of WIP in the system (see, for example, Shanthikumar and Stecke [7]).

On the other hand, in the single-device material handling model developed by Chow [1], for example, the material handling system is modeled as a single server queue, and the device is assumed to be dispatched according to the First-Come-First-Served (FCFS) dispatching rule. Although one can obtain the expected waiting time (and the average queue length) for all the "move requests" with such a model, one cannot obtain the average queue lengths at the workstations since the workstations themselves and the queues that form ahead of them are not considered in Chow's model.

There are a number of other models one can use to examine expected queue lengths or waiting times in manufacturing systems. (See, for example, Suri and Hildebrant [9], and Hodgson, King, and Monteith [4], among many others.) While some of these models are analytical models, most of them are simulation models. In both cases, however, these models either do not explicitly capture material handling travel times, that is, they do not differentiate between empty and loaded travel (which we describe in more detail in section 3), or they do not explicitly consider the workstations.

In addition to the above studies, another study which has been cited often with regard to WIP behavior in manufacturing is presented by Merchant [6]. According to the author, the average workpiece in a batch-type metal cutting shop spends only about 5% of its time on machine tools. For the remaining 95% of the time, the workpiece is either moving or waiting.

In this study we examine the behavior of WIP in a manufacturing system where both the workstations and the handling system are explicitly represented in the model. We also differentiate between jobs waiting to be processed and jobs waiting to be moved. In section 2 we describe the problem setting and the assumptions we make for the study. In section 3 we describe the simulation model and discuss the results obtained from it. In section 4, using well-known analytical results, we provide further insight as to why the material handling system should not be responsible for WIP. We also empirically show that, even if the WIP associated with the workstations is reduced by increasing the processing capacity at the workstations (while keeping the workload constant), the handling system would still not be responsible for WIP. In section 5 we present our conclusions.

2. PROBLEM ENVIRONMENT AND ASSUMPTIONS

The manufacturing system of interest is assumed to be composed of two primary components: the workstations and the material handling system. Each workstation may represent an assembly station, a single machine (or a "processor"), or a group of machines (such as a cell). The material handling system, on the other hand, is assumed to be a trip-based handling system, where self-powered devices, which operate independently and asynchronously, perform trips to move loads one at a time from one workstation to another (see Srinivasan, Bozer, and Cho [8]). Manual handling systems (with people pushing carts), lift trucks, pick-and-drop automated guided vehicle (AGV) systems, and bridge cranes, to name a few, are examples of trip-based handling systems. (The reader may refer to Tompkins and White [10] for more information on the above and other handling systems.)

As shown in Figure 1, each workstation is assumed to have an input queue and an output queue. These queues may be "informal" queues or "formal" queues. For example, if a lift truck driver simply deposits the container in any open space (on the floor) near the workstation, then the queue can be viewed as an "informal" queue. In contrast, if the driver has to deposit the container on a specific conveyor of given length, then the queue can be considered to be a "formal" one.

We assume that sufficient queueing space is provided so that no blocking of any form can occur. (Note that this assumption is easier to satisfy with "informal" queues.) The loads destined to a workstation are dropped at its input queue while the loads leaving that workstation are picked up from its output queue. It is assumed that the device travel time between the input and output queues at a workstation is negligible. (If this assumption is not satisfied, one can always increase the pick up or deposit time associated with a workstation to capture the additional travel time required between the two queues.)

There are two types of workstations. The first type is an *input/output (I/O) station.* Loads that arrive from outside the system, arrive directly at the *output queue* of an I/O station where they wait to be picked up by a device. Likewise, loads that require no further processing in the system are dropped off at the *input queue* of an I/O station where they are assumed to instantly leave the system. Note that no processing takes place at an I/O station. Also, flow is not necessarily conserved at an I/O station because a load may enter the system from one I/O station and exit from another. Although it is not essential for the study, we assume that arrival of loads from outside the system follows a Poisson process.

The second type of workstation is a *processor station* where actual processing takes place. At a processor station, the loads are removed from the corresponding input queue and after a certain period of time, which reflects the processing time, they are deposited at the corresponding output queue. It is assumed that flow is conserved at each processor station. For the purposes of this study, it is also assumed that there is a single processor at each processor station. The utilization of the processor is designated by ρ_p. (In section 4.4 we will discuss the impact of relaxing this assumption and replacing a single processor with parallel processors at a workstation.) The material handling requirement *within* a processor station is beyond the scope of the study.

The WIP in the system is attributed to the two system components as follows: all the containers (i.e., loads) that are waiting in the input queues, plus those that are being processed, are considered to be "WIP due to the processors" or "PR-WIP" for short. All

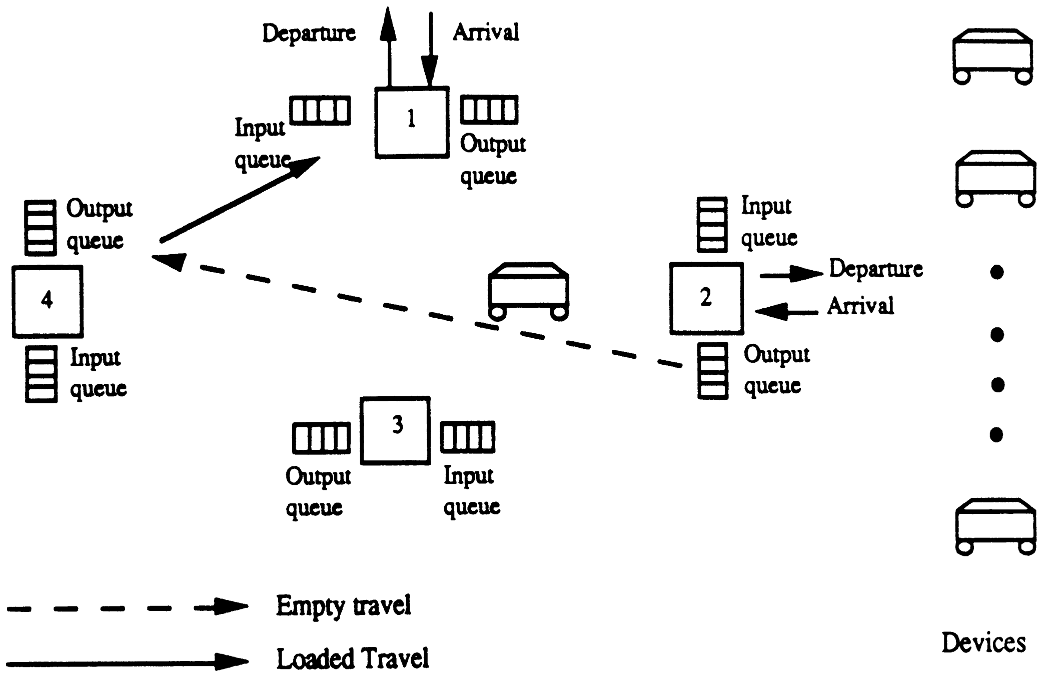

Figure 1: Trip-Based Material Handling System

the loads that are waiting in the output queues, plus those that are being moved, are considered to be "WIP due to the material handling system" or "MH-WIP" for short. (We also use the term *move request* for a load that is waiting in the output queue of a workstation.)

In serving the move requests, each device is assumed to follow the Modified-First-Come-First-Served (MOD FCFS) rule. Under this rule, a device which has just delivered a load to its destination first checks the output queue of the destination station. If there are one or more unassigned loads in the output queue, the device is assumed to pick up the oldest unassigned load in that queue. Otherwise, the (empty) device is dispatched to the oldest (unassigned) move request in the system. If there are no unassigned move requests in the system, the device stays idle at its last delivery point, i.e., the destination station, until a new move request arrives. If an arriving move request finds two or more devices idle, it is assigned to the device which has been idle for the longest time period.

The MOD FCFS rule, which was proposed and evaluated by Srinivasan, Bozer and Cho [8], is a *centralized* (empty device) dispatching rule since an empty device needs to be "told" where the oldest unassigned move request is located in the system.[1] According to the results reported in [8], the MOD FCFS rule is comparable in throughput performance to the Shortest-Travel-Time-First (STTF) rule, which was shown empirically to be a reasonably efficient dispatching rule (see Egbelu and Tanchoco [2]).

Under the STTF rule, an empty device is always dispatched to the closest unassigned move request in the system. We use the MOD FCFS rule instead of the STTF rule since the performance of the latter is more likely to be layout dependent. Furthermore, since the two dispatching rules in question are comparable in overall performance, we do not believe that the main results we present in this study will change significantly if one uses the STTF rule instead. In fact, a number of simulation runs we made under the STTF rule clearly indicate that the conclusions we draw under the MOD FCFS rule would be more or less equally valid under the STTF rule (and perhaps other centralized dispatching rules that have comparable throughput performance).

Let M denote the number of workstations in the system. Further let ϑ and Ω denote the set of processor stations and the set of I/O stations, respectively. Assuming that the flow data is given, let λ_i denote the rate at which loads arrive at the output queue of workstation i. Likewise, let Λ_i denote the rate at which loads are delivered to the input queue of workstation i. We assume that $\Lambda_i = \lambda_i$ in steady state for $i \in \vartheta$. (This also implies that a processor station may not be a bottleneck.) For I/O stations, on the other hand, recall that Λ_i need not equal λ_i in general. However, from conservation of flow, we must have $\sum_{i \in \Omega} \Lambda_i = \sum_{i \in \Omega} \lambda_i$ since we assume that the material handling system may not be a bottleneck either.

Let f_{ij} denote the number of loads per time unit that flow from (the output queue of) workstation i to (the input queue of) workstation j. Recall that a device handles only

---

[1] Centralized dispatching rules can be implemented in various ways. A well-known approach is to use radio frequency (RF) communication. Radio dispatched AGVs and lift trucks, for example, are fairly common in industry.

one load on each trip. (It is implicit that $f_{ii} = 0$.) Given the f_{ij} values, we have:

$$\lambda_i = \sum_{j=1}^{M} f_{ij} \qquad \text{and} \qquad \Lambda_i = \sum_{j=1}^{M} f_{ji}, \tag{2.1}$$

by definition.

Flow can also be expressed in terms of the routing matrix where p_{ij} denotes the probability that a load picked up at workstation i is destined to workstation j. The p_{ij} values can be easily obtained from the relationship $p_{ij} = f_{ij}/\lambda_i$. Note that the routing matrix and the λ_i values ($i \in \Omega$) fully define the flow requirement in the system.

Consider next the travel time associated with a *trip*. Suppose a device picks up a load from the output queue of workstation i, transports it to workstation j, unloads it at the input queue of workstation j, and subsequently inspects the output queue of workstation j. The total time required to perform the above operations — including the pick-up time at workstation i, the deposit time at workstation j, and the time required to inspect workstation j — is assumed to be a random variable with mean τ_{ij}, which represents the loaded travel time from station i to j.

In contrast, if a device has just delivered a load at workstation i and upon inspection finds its output queue empty, then it must travel empty from workstation i to workstation j assuming that the oldest unassigned move request is located at workstation j. The resulting empty travel time is assumed to be a random variable with mean σ_{ij}, which represents the empty travel time from station i to j. (It is implicit that $\sigma_{ii} = 0$.)

Hence, to serve one move request, a device must perform one *trip* which consists of two parts: empty and loaded travel. Obviously, the travel time associated with the first part (i.e., empty travel) may be equal to zero for some move requests.

Let α_f denote the expected proportion of time that a device has to travel loaded. (This includes the time required to pick up and deposit loads, by definition.) Since the flow data and the travel times are given, it is straightforward to compute α_f from the following expression:

$$\alpha_f = \frac{\sum_{i=1}^{M} \sum_{j=1}^{M} f_{ij}\tau_{ij}}{N} = \frac{\sum_{i=1}^{M} \lambda_i \sum_{j=1}^{M} p_{ij}\tau_{ij}}{N}, \tag{2.2}$$

where N denotes the number of devices in the system. Note that α_f is independent of the empty device dispatching rule. Also note that, if $\alpha_f \geq 1$, the material handling system will not be able to meet the required throughput (regardless of the empty device dispatching rule used) and α_f can no longer be defined as a proportion.

Although the condition $\alpha_f < 1.0$ is a necessary condition for the material handling system to meet the required throughput, it is not a sufficient condition. This is primarily due to empty device travel. Note that, in order to serve the next move request, a device may have to first travel empty to the workstation where the load is waiting. Hence, some empty travel is "mandatory" under MOD FCFS (as well as STTF). Letting α_e denote the proportion of time that a device is traveling empty, we define $\rho_h = \alpha_f + \alpha_e$ as the device utilization. (We will also refer to ρ_h as the utilization of the material handling system since the devices are assumed to be homogeneous.) Obviously, the material handling system is stable, that is, it will meet the required throughput, if ρ_h is less than 1.0.

Lastly, the "service time" for the two components of interest are defined as follows: for a processor station, the service time is simply equal to the processing time at that workstation. For the material handling system, the service time is equal to the trip time (which consists of empty plus loaded travel). In section 3, we experiment with various service time distributions for the two components. For example, a uniform service time for a processor implies that the processing time is uniformly distributed. For the material handling system, however, a "uniform service time" is assumed to imply uniformly distributed empty travel (with mean σ_{ij}) and uniformly distributed loaded travel (with mean τ_{ij}). (Since a trip consists of empty and loaded travel, controlling the distribution of the two parts is easier than controlling the resulting distribution of the trip time.)

We do not explicitly capture any interaction that might exist among the devices and the impact it may have on the trip times. For example, the type of interaction that occurs among lift trucks is quite different than the type of interaction that occurs among vehicles in an AGV system. In order to keep the model general, rather than simulate the details of various interactions that might occur among the devices, we simply sample the trip times from the appropriate distribution.

3. THE SIMULATION RESULTS

The system described in the previous section was simulated using SIMLIB (see Law and Kelton [5], p. 141). Following a warm-up period, each replication was simulated long enough to allow an average of 5,000 loaded trips per device. That is, in a system with say, 8 devices, a replication is completed when a total of 40,000 move requests have been served. The results shown in the following paragraphs are based on 10 successive replications.

Three different layouts were analyzed with the above simulation model. As shown in Figure 2a, the first layout, i.e., layout 1, has 5 processor and 2 I/O stations. Layout 2 (shown in Figure 2b) has 7 processor and 4 I/O stations. Lastly, layout 3 (shown in Figure 2c) has 15 processor and 5 I/O stations. (Recall that each load enters or exits through an I/O station where no processing takes place.) The detailed data associated with each layout — that is, the routing matrix, the travel distances, and the load interarrival times — are shown in the Appendix. The (mean) empty travel time for a given trip is obtained simply by dividing the corresponding travel distance (shown in the Appendix) by the travel speed. The (mean) loaded travel time is obtained by adding the pick-up and deposit times to the empty travel time. For all three layouts, it is assumed that it always takes 1/3 minutes to pick-up or deposit a load.

Three types of distributions were examined for the processing times and the handling times: constant (C), uniform (U), and exponential (E). More specifically, we examined the following combinations: C/U, U/U, U/E, and E/E, where the first and second letters represent the handling time distribution and the processing time distribution, respectively. Whenever the uniform distribution is used, it is assumed to have a coefficient of variation of 0.40. Lastly, we let γ designate the ratio of the overall mean processing time per job to the overall mean trip time (or handling time) per job.

Although one would generally expect the processing time to be greater than the han-

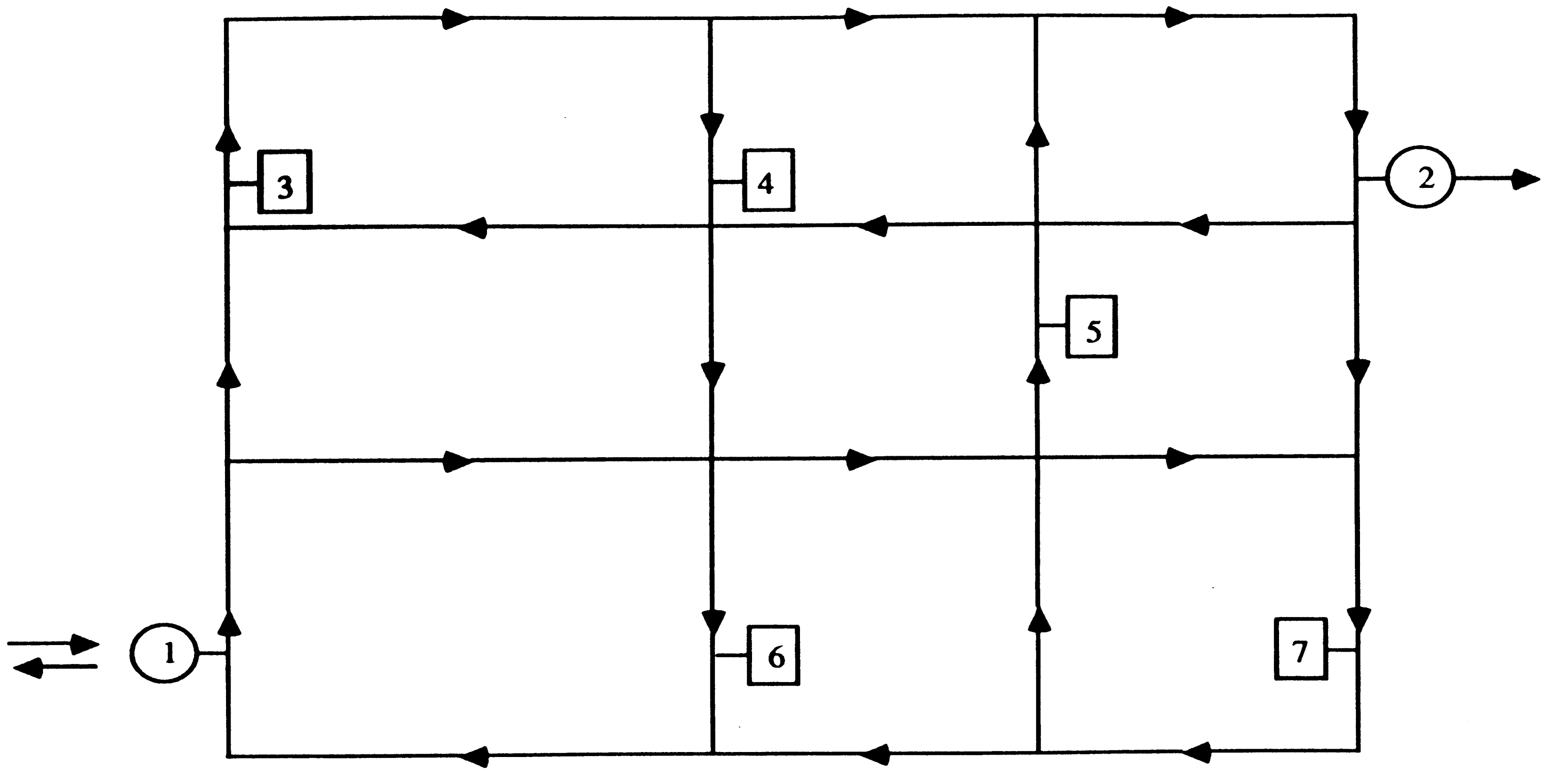

Figure 2a. Layout 1 with 2 I/O and 5 processing stations.

○ : I/O station

□ : Processing station

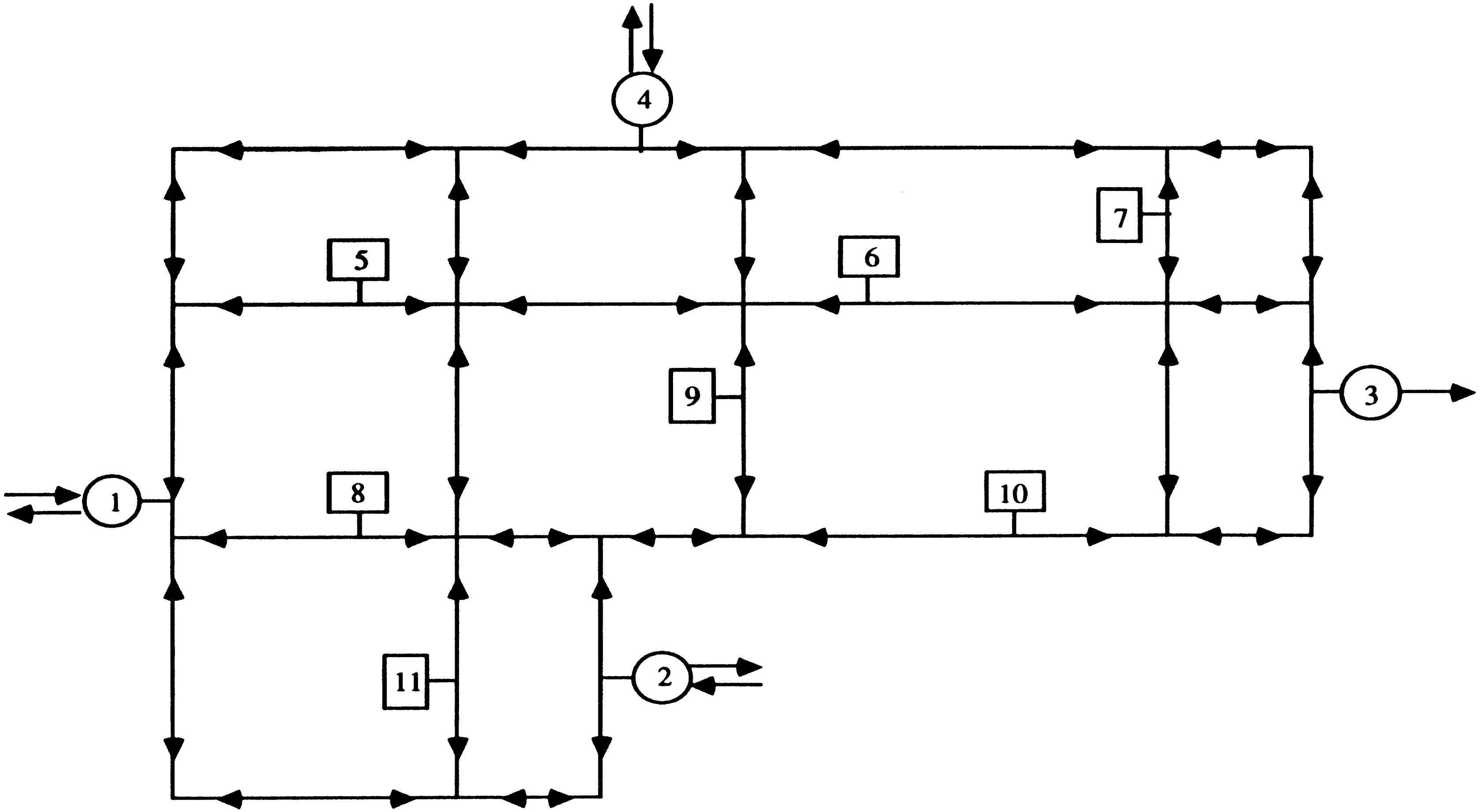

Figure 2b. Layout 2 with 4 I/O and 7 processing stations.

○ : I/O station

□ : Processing station

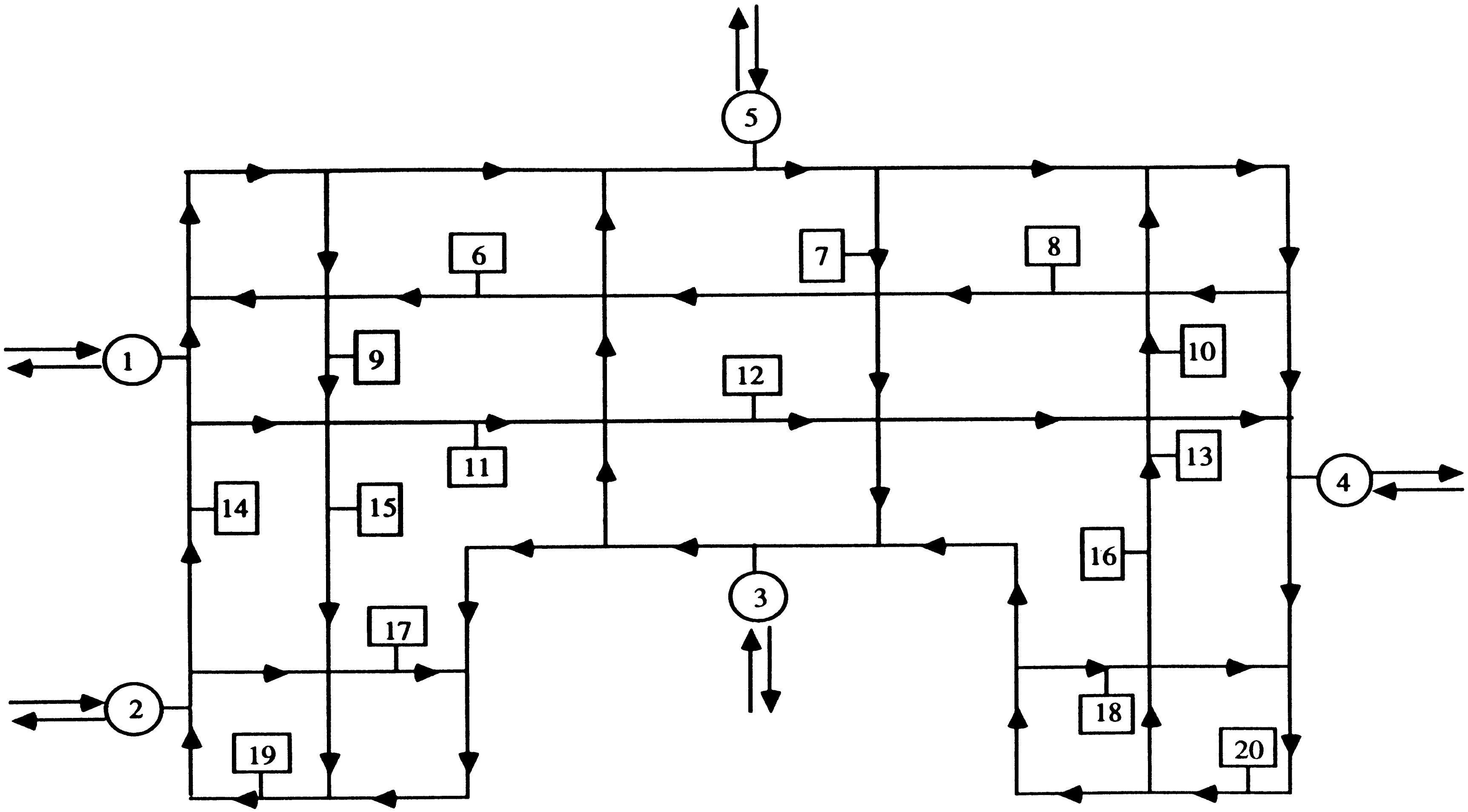

Figure 2c. Layout 3 with 5 I/O and 15 processing stations.

○ : I/O station

□ : Processing station

dling time in most manufacturing systems, in the first set of runs we used a γ value of approximately 1.0. We adjusted the number of devices, N, the device travel speed, s (expressed in distance units/minute for *both* empty and loaded device travel), and the processing times to obtain $\gamma \approx 1.0$ and $\rho_h \approx \rho_p$. The results obtained from the simulation model are shown in Tables 1a, 1b, and 1c, for layouts 1, 2, and 3, respectively. (Also see Figure 3 for the results shown in Table 1b.)

The distribution of WIP is measured by a break down of the time a load spends 1. waiting in an input queue, 2. being processed, 3. waiting in an output queue, and 4. being transported, as shown by the column labeled "time" in Table 1. The expected number of loads associated with each one of the above four categories is shown by the column labeled "WIP" in Table 1. The column which shows the percent break down for the above four categories naturally applies to both the "time" column and the "WIP" column. That is, on the average, if all loads spend, say, 30% of their time waiting in an input queue, then all the loads waiting to be processed account for 30% of the WIP in the system. The two subtotals highlighted in Table 1 reflect the *percent* PR-WIP and *percent* MH-WIP.

Note that in all the cases that were simulated, although the utilization of the handling system is comparable to the utilization of the processors, and $\gamma \approx 1.0$, a significant portion of the WIP is associated with the processors; that is, as reflected by the two "Sub-Totals" in Table 1, the PR-WIP is considerably larger than the MH-WIP. In all three layouts, the difference between PR-WIP and MH-WIP becomes more pronounced as the utilization of the processors *and* the devices increase. For example, in layout 1, with $\rho_p \approx \rho_h \approx 0.75$ and uniform processing times, approximately 70% of the time a load is part of the PR-WIP. For the same processing time distribution, when the values of ρ_p *and* ρ_h are increased to approximately 0.90, then a load is part of the PR-WIP approximately 80% of the time. Considering that the two components have comparable service times and comparable utilizations, we find the above results to be quite remarkable. However, as shown in section 4.1, we also believe that the explanation is quite straightforward.

The expected number of loads being processed (which is shown under the column labeled "WIP" and the row labeled "Processor" in Table 1) can be easily obtained by multiplying the number of processor stations with ρ_p. However, the above does not hold for the expected number of loads being transported (which is shown under "WIP" and "Device" in Table 1). This is simply because a device is considered to be utilized while its traveling empty to pick up the next move request. That is, as explained earlier, a portion of the "service time" (or handling time) is spent traveling empty.

Such a definition is necessary for the material handling system since empty travel is mandatory. Note that, as far as WIP is concerned, this does not bias the results since the load that is waiting to be picked up is still associated with the MH-WIP while an empty device is on its way. Also note that the definition of "service time" adopted for the material handling system is somewhat unconventional because during a portion of the service time — which represents empty travel — the load being "served" is still waiting in the output queue. If we assume that only loaded travel (which occurs *after* the load is removed from the output queue) represents the service time, then empty travel can be viewed as the "switch-over time" or the "set-up time" that the device incurs before serving a move request. As shown in section 4.2, this interpretation of empty travel plays

Table 1a. Simulation results for layout 1 ($\gamma = 1$)

Handling/Processing Time Distribution				Constant/Uniform			Uniform/Uniform			Uniform/Exponential			Exponential/Exponential		
				Time in system			Time in system			Time in system			Time in system		
ρ_P	N	s	Component	time	%	WIP	time	%	WIP	time	%	WIP	time	%	WIP
0.60	5	3	**Total**	246.47		8.32	245.75		8.29	312.89		10.53	312.58		10.53
			Inp. Queue	64.77	26.28	2.18	65.19	26.53	2.19	128.41	41.04	4.31	129.24	41.35	4.34
			Processor	88.18	35.78	3.00	88.18	35.88	3.00	88.53	28.29	3.00	88.53	28.32	3.00
			Sub-Total	152.95	62.06	5.18	153.37	62.41	5.19	216.94	69.33	7.31	217.77	69.67	7.34
			Out. Queue	40.84	16.57	1.37	40.05	16.30	1.34	43.62	13.94	1.46	42.57	13.62	1.43
			Device	52.68	21.37	1.77	52.33	21.29	1.76	52.33	16.73	1.76	52.24	16.71	1.76
			Sub-Total	93.52	37.94	3.14	92.38	37.59	3.10	95.95	30.67	3.22	94.81	30.33	3.19
				$\rho_h = 0.636, \gamma = 1.112$			$\rho_h = 0.631, \gamma = 1.123$			$\rho_h = 0.629, \gamma = 1.132$			$\rho_h = 0.628, \gamma = 1.134$		
0.75	5	2.5	Total	386.05		13.01	386.28		13.01	549.79		18.50	551.26		18.55
			Inp. Queue	159.10	41.20	5.34	159.91	41.40	5.37	314.90	57.28	10.58	315.32	57.20	10.60
			Processor	110.45	28.60	3.75	110.44	28.60	3.75	110.88	20.17	3.75	110.88	20.10	3.75
			Sub-Total	269.55	69.80	9.09	270.35	70.00	9.12	425.78	77.45	14.33	426.20	77.30	14.35
			Out. Queue	53.50	13.90	1.80	53.15	13.76	1.78	61.23	11.13	2.06	62.40	11.30	2.09
			Device	63.00	16.30	2.12	62.78	16.24	2.11	62.78	11.42	2.11	62.66	11.40	2.11
			Sub-Total	116.50	30.20	3.92	115.93	30.00	3.89	122.01	22.55	4.17	125.06	22.70	4.20
				$\rho_h = 0.758, \gamma = 1.172$			$\rho_h = 0.753, \gamma = 1.181$			$\rho_h = 0.746, \gamma = 1.195$			$\rho_h = 0.742, \gamma = 1.202$		
0.90	5	2	Total	850.55		28.64	848.96		28.59	1410.40		47.43	1428.77		48.04
			Inp. Queue	554.86	65.24	18.66	556.17	65.51	18.71	1095.16	77.65	36.81	1105.11	77.34	37.14
			Processor	132.68	15.60	4.50	132.69	15.63	4.50	133.21	9.45	4.50	133.21	9.32	4.50
			Sub-Total	687.54	80.84	23.16	688.86	81.14	23.21	1228.37	87.10	41.31	1238.32	86.66	41.64
			Out. Queue	83.42	9.80	2.80	81.65	9.62	2.74	103.58	7.34	3.48	112.15	7.85	3.76
			Device	79.59	9.36	2.68	78.45	9.24	2.64	78.45	5.56	2.64	78.30	5.49	2.64
			Sub-Total	163.01	19.16	5.48	160.10	18.86	5.38	182.03	12.90	6.12	190.45	13.34	6.40
				$\rho_h = 0.910, \gamma = 1.176$			$\rho_h = 0.897, \gamma = 1.192$			$\rho_h = 0.885, \gamma = 1.213$			$\rho_h = 0.876, \gamma = 1.225$		

Table 1b. Simulation results for layout 2 ($\gamma = 1$)

Handling/Processing Time Distribution				Constant/Uniform			Uniform/Uniform			Uniform/Exponential			Exponential/Exponential		
				Time in system			Time in system			Time in system			Time in system		
ρ_P	$\bar{N}$	s	Component	time	%	WIP	time	%	WIP	time	%	WIP	time	%	WIP
0.60	10	12	**Total**	34.82		13.06	34.71		13.01	42.02		15.76	41.66		15.63
			Inp. Queue	9.45	27.13	3.54	9.47	27.28	3.54	16.64	39.60	6.23	16.72	40.13	6.26
			Processor	11.15	32.03	4.20	11.15	32.13	4.20	11.14	26.51	4.20	11.14	26.74	4.20
			Sub-Total	20.60	59.16	7.74	20.62	59.40	7.74	27.78	66.11	10.43	27.86	66.87	10.46
			Out. Queue	5.46	15.68	2.04	5.35	15.42	2.00	5.50	13.10	2.06	5.06	12.14	1.89
			Device	8.76	25.16	3.28	8.74	25.18	3.27	8.74	20.79	3.27	8.74	20.99	3.28
			Sub-Total	14.22	40.84	5.32	14.09	40.60	5.27	14.24	33.89	5.33	13.80	33.13	5.17
				$\rho_h = 0.56, \gamma = 0.905$			$\rho_h = 0.561, \gamma = 0.906$			$\rho_h = 0.562, \gamma = 0.903$			$\rho_h = 0.562, \gamma = 0.903$		
0.75	10	9	**Total**	55.70		20.88	55.58		20.83	74.00		27.73	73.59		27.58
			Inp. Queue	23.49	42.16	8.79	23.52	42.30	8.80	41.44	56.00	15.51	41.53	56.43	15.54
			Processor	13.96	25.07	5.25	13.96	25.12	5.25	13.95	18.84	5.25	13.95	18.95	5.25
			Sub-Total	37.45	67.23	14.04	37.48	67.42	14.05	55.39	74.84	20.76	55.48	75.38	20.79
			Out. Queue	7.37	13.23	2.76	7.23	13.01	2.71	7.74	10.46	2.90	7.23	9.83	2.71
			Device	10.88	19.54	4.08	10.88	19.57	4.07	10.87	14.70	4.07	10.88	14.79	4.08
			Sub-Total	18.25	32.77	6.84	18.10	32.58	6.78	18.61	25.16	6.97	18.11	24.62	6.79
				$\rho_h = 0.721, \gamma = 0.885$			$\rho_h = 0.720, \gamma = 0.885$			$\rho_h = 0.716, \gamma = 0.890$			$\rho_h = 0.715, \gamma = 0.891$		
0.90	10	7	**Total**	121.43		45.49	121.30		45.44	186.23		69.95	186.22		69.95
			Inp. Queue	80.77	66.52	30.24	80.82	66.63	30.26	144.20	77.43	54.18	144.35	77.52	54.24
			Processor	16.77	13.81	6.30	16.77	13.82	6.30	16.75	9.00	6.30	16.75	9.00	6.30
			Sub-Total	97.54	80.33	36.54	97.59	80.45	36.56	160.95	86.43	60.48	161.10	86.52	60.54
			Out. Queue	10.54	8.68	3.95	10.39	8.57	3.89	11.96	6.42	4.48	11.79	6.33	4.41
			Device	13.35	10.99	5.00	13.32	10.98	4.99	13.32	7.15	4.99	13.33	7.15	5.00
			Sub-Total	23.89	19.67	8.95	23.71	19.55	8.88	25.28	13.57	9.47	25.12	13.48	9.41
				$\rho_h = 0.882, \gamma = 0.870$			$\rho_h = 0.878, \gamma = 0.874$			$\rho_h = 0.866, \gamma = 0.885$			$\rho_h = 0.863, \gamma = 0.888$		

Table 1c. Simulation results for layout 3 ($\gamma = 1$)

Handling/Processing Time Distribution				Constant/Uniform			Uniform/Uniform			Uniform/Exponential			Exponential/Exponential		
				Time in system			Time in system			Time in system			Time in system		
ρ_P	N	s	Component	time	%	WIP	time	%	WIP	time	%	WIP	time	%	WIP
0.60	15	45	**Total**	29.89		25.00	29.83		24.95	37.12		31.02	36.84		30.78
			Inp. Queue	9.18	30.72	7.66	9.19	30.82	7.67	16.41	44.20	13.69	16.42	44.57	13.70
			Processor	10.74	35.92	9.00	10.74	35.98	9.00	10.75	28.96	9.00	10.75	29.18	9.00
			Sub-Total	19.92	66.64	16.66	19.93	66.80	16.67	27.16	73.16	22.69	27.17	73.75	22.70
			Out. Queue	3.85	12.88	3.21	3.78	12.68	3.15	3.84	10.35	3.20	3.54	9.62	2.95
			Device	6.12	20.48	5.13	6.12	20.52	5.13	6.12	16.49	5.13	6.13	16.63	5.13
			Sub-Total	9.97	33.36	8.34	9.90	33.20	8.28	9.96	26.84	8.33	9.67	26.25	8.08
				$\rho_h = 0.588, \gamma = 1.156$			$\rho_h = 0.588, \gamma = 1.156$			$\rho_h = 0.588, \gamma = 1.156$			$\rho_h = 0.589, \gamma = 1.162$		
0.75	15	35	Total	48.28		40.35	48.18		40.27	66.53		55.56	66.17		55.26
			Inp. Queue	22.88	47.39	19.10	22.87	47.47	19.09	41.05	61.70	34.25	41.06	62.05	34.27
			Processor	13.45	27.85	11.25	13.45	27.91	11.25	13.46	20.24	11.25	13.46	20.35	11.25
			Sub-Total	36.33	75.24	30.35	36.32	75.38	30.34	54.51	81.94	45.50	54.52	82.40	45.52
			Out. Queue	4.92	10.20	4.11	4.83	10.01	4.03	4.99	7.50	4.16	4.61	6.96	3.84
			Device	7.03	14.56	5.89	7.03	14.60	5.90	7.03	10.57	5.90	7.04	10.64	5.90
			Sub-Total	11.95	24.76	10.00	11.86	24.62	9.93	12.02	18.06	10.06	11.65	17.60	9.74
				$\rho_h = 0.713, \gamma = 1.194$			$\rho_h = 0.713, \gamma = 1.199$			$\rho_h = 0.710, \gamma = 1.204$			$\rho_h = 0.710, \gamma = 1.203$		
0.90	15	25	Total	114.94		96.23	114.92		96.21	181.50		152.23	181.22		151.98
			Inp. Queue	82.67	71.92	69.25	82.72	71.98	69.29	148.42	81.77	124.59	148.47	81.93	124.62
			Processor	16.16	14.06	13.50	16.16	14.06	13.50	16.18	8.91	13.50	16.18	8.93	13.50
			Sub-Total	98.83	85.98	82.75	98.88	86.04	82.79	164.60	90.68	138.09	164.65	90.86	138.12
			Out. Queue	7.45	6.48	6.22	7.36	6.41	6.15	8.22	4.53	6.87	7.89	4.35	6.59
			Device	8.67	7.54	7.26	8.68	7.55	7.27	8.68	4.78	7.27	8.68	4.79	7.27
			Sub-Total	16.11	14.02	13.48	16.04	13.96	13.42	16.90	9.32	14.14	16.57	9.14	13.86
				$\rho_h = 0.909, \gamma = 1.131$			$\rho_h = 0.908, \gamma = 1.132$			$\rho_h = 0.896, \gamma = 1.150$			$\rho_h = 0.893, \gamma = 1.152$		

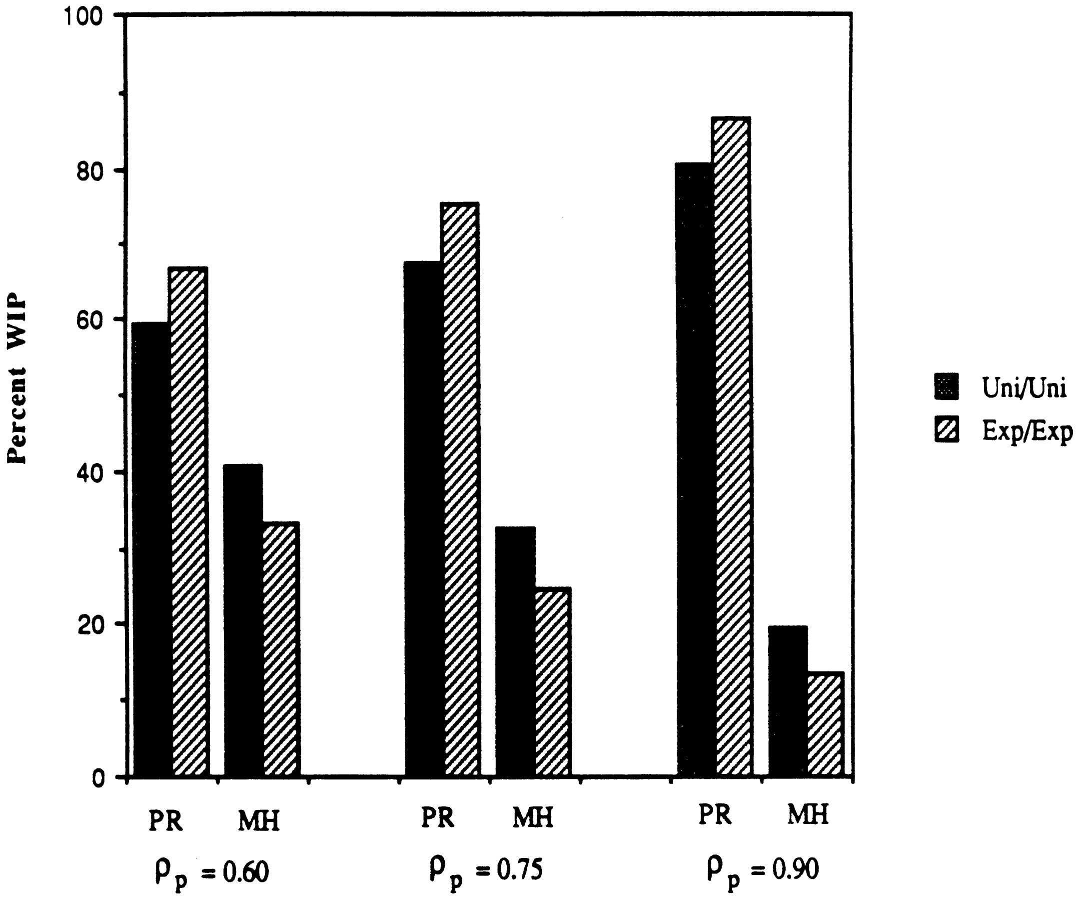

Figure 3: Break down of WIP for Table 1b

an important role in explaining one of the fundamental observations we make for MH-WIP.

One can also observe in Table 1 that MH-WIP is remarkably insensitive to the handling time distribution. If the processing time is uniformly distributed, changing the handling time from constant to uniform has no noticeable effect on the MH-WIP (or the PR-WIP). The same result is observed if the processing time is exponentially distributed and the handling time is changed from uniform to exponential.

In contrast, the distribution of the processing time appears to have a significant impact on the PR-WIP (and a marginal impact on the MH-WIP). As one would anticipate, since most of the WIP is associated with the processors, changing the processing time distribution from uniform to exponential (while maintaining uniform handling times) significantly increases the time spent waiting to be processed. As a result, the total time spent in the system and total WIP increase significantly. Obviously, the *percent* of WIP attributable to the processors increases as well. Note that, when exponential processing times are used, there is also an increase in the expected output queue length. However, compared to the increase in the expected input queue lengths, this increase is quite small.

Hence, if one has the option of reducing WIP by reducing the variability in the system, it is clear that significant reductions in WIP can be obtained by reducing the variability of the processing times, rather than the handling times (provided that neither component is a bottleneck, and that both components have non-negligible variance in their service times).

The second set of runs we made are based on a more realistic γ value of 3.0; that is, the overall mean processing time is three times longer than the overall mean handling time. The larger γ value was obtained simply by increasing the speed of the material handling devices and at the same time reducing the number of devices to maintain $\rho_p \approx \rho_h$. The results for layouts 1, 2, and 3 are presented in Tables 2a, 2b, and 2c, respectively. (Also see Figure 4 for the results shown in Table 2b.) From Table 2 it can be observed that, increasing the γ value from approximately 1.0 to 3.0, further increases the WIP associated with the processors. Although the increase in PR-WIP is less noticeable if the processing time is exponential and $\rho_p = 0.90 \approx \rho_h$, PR-WIP accounts for 80% to 94% of the WIP in the system despite comparable utilizations for the processors and the handling system! (Other observations we made with $\gamma = 1.0$ still remain valid with $\gamma = 3.0$.)

Comparing Tables 1 and 2 for each layout, one can observe that the above increase in *percent* PR-WIP is primarily due to a reduction in MH-WIP, and not due to an absolute increase in PR-WIP. In other words, when γ is changed from 1.0 to 3.0, i.e., when the device speed is increased and the number of devices is reduced, no significant changes in the number of loads waiting to be processed is observed. (This also indicates that the expected queue length at the input queues is much more sensitive to the processing time distribution than it is to the distribution of the load arrival times.) As one would anticipate, there are also no significant changes in the number of loads being processed (since γ was changed without changing the processing times).

However, when γ is increased from 1.0 to 3.0, i.e., when the device speed is increased and the number of devices is reduced, we clearly observe a decrease in both the number of loads being transported and the number of loads waiting in the output queues. Since the device speed was increased, the former observation was anticipated. The latter observation,

Table 2a. Simulation results for layout 1 ($\gamma = 3$)

Handling/Processing Time Distribution				Constant/Uniform			Uniform/Uniform			Uniform/Exponential			Exponential/Exponential		
				Time in system			Time in system			Time in system			Time in system		
ρ_P	N	s	Component	time	%	WIP	time	%	WIP	time	%	WIP	time	%	WIP
0.60	2	9	**Total**	197.23		6.59	197.72		6.60	263.68		8.79	265.73		8.86
			Inp. Queue	67.68	34.31	2.26	67.74	34.26	2.26	130.42	49.46	4.35	130.59	49.14	4.36
			Processor	89.83	45.55	3.00	89.83	45.43	3.00	90.15	34.19	3.00	90.14	33.92	3.00
			Sub-Total	157.51	79.86	5.26	157.57	79.69	5.26	220.57	83.65	7.35	220.73	83.06	7.36
			Out. Queue	19.73	10.00	0.66	20.15	10.19	0.67	23.11	8.77	0.77	25.02	9.42	0.83
			Device	19.99	10.14	0.67	20.00	10.12	0.67	20.00	7.58	0.67	19.98	7.52	0.67
			Sub-Total	39.72	20.14	1.33	40.15	20.31	1.34	43.11	16.35	1.44	45.00	16.94	1.50
				$\rho_h = 0.570, \gamma = 3.144$			$\rho_h = 0.570, \gamma = 3.146$			$\rho_h = 0.566, \gamma = 3.190$			$\rho_h = 0.564, \gamma = 3.190$		
0.75	2	6	Total	348.88		11.65	351.24		11.74	512.90		17.11	522.51		17.44
			Inp. Queue	164.23	47.07	5.49	165.00	46.98	5.52	314.12	61.24	10.49	315.49	60.38	10.54
			Processor	112.31	32.19	3.75	112.32	31.98	3.75	112.73	21.98	3.75	112.72	21.57	3.75
			Sub-Total	276.54	79.26	9.24	277.32	78.96	9.27	426.85	83.22	14.24	428.21	81.95	14.29
			Out. Queue	43.53	12.48	1.45	45.14	12.85	1.51	57.27	11.17	1.91	65.55	12.55	2.19
			Device	28.81	8.26	0.96	28.78	8.19	0.96	28.78	5.61	0.96	28.75	5.50	0.96
			Sub-Total	72.03	20.74	2.41	73.92	21.04	2.47	86.05	16.78	2.87	94.30	18.05	3.15
				$\rho_h = 0.809, \gamma = 2.780$			$\rho_h = 0.805, \gamma = 2.787$			$\rho_h = 0.799, \gamma = 2.816$			$\rho_h = 0.791, \gamma = 2.850$		
0.90	2	5	Total	820.93		27.49	825.22		27.63	1332.60		44.58	1352.65		45.25
			Inp. Queue	584.92	71.25	19.61	586.03	71.01	19.65	1061.88	79.68	35.56	1065.70	78.79	35.69
			Processor	134.74	16.41	4.50	134.76	16.33	4.50	135.26	10.15	4.50	135.27	10.00	4.50
			Sub-Total	719.66	87.66	24.11	720.79	87.34	24.15	1197.14	89.83	40.06	1200.97	88.79	40.19
			Out. Queue	67.20	8.19	2.24	70.39	8.54	2.35	101.42	7.61	3.39	117.70	8.70	3.93
			Device	34.07	4.15	1.14	34.04	4.12	1.13	34.04	2.56	1.13	33.98	2.51	1.13
			Sub-Total	101.27	12.34	3.38	104.43	12.66	3.48	135.46	10.17	4.52	151.68	11.21	5.06
				$\rho_h = 0.920, \gamma = 2.928$			$\rho_h = 0.913, \gamma = 2.953$			$\rho_h = 0.903, \gamma = 2.994$			$\rho_h = 0.89, \gamma = 3.030$		

Table 2b. Simulation results for layout 2 ($\gamma = 3$)

Handling/Processing Time Distribution				Constant/Uniform			Uniform/Uniform			Uniform/Exponential			Exponential/Exponential		
				Time in system			Time in system			Time in system			Time in system		
ρ_p	N	s	Component	time	%	WIP	time	%	WIP	time	%	WIP	time	%	WIP
0.60	3	70	**Total**	25.60		9.59	25.58		9.59	33.49		12.53	33.52		12.55
			Inp. Queue	9.39	36.67	3.51	9.38	36.67	3.51	17.06	50.94	6.38	17.05	50.86	6.38
			Processor	11.19	43.72	4.20	11.19	43.75	4.20	11.22	33.50	4.20	11.22	33.48	4.20
			Sub-Total	20.58	80.39	7.71	20.57	80.42	7.71	28.28	84.44	10.58	28.27	84.34	10.58
			Out. Queue	1.47	5.73	0.55	1.46	5.71	0.55	1.66	4.97	0.62	1.70	5.07	0.64
			Device	3.55	13.88	1.33	3.55	13.87	1.33	3.55	10.59	1.33	3.55	10.59	1.33
			Sub-Total	5.02	19.61	1.88	5.01	19.58	1.88	5.21	15.56	1.95	5.25	15.66	1.97
				$\rho_h = 0.575, \gamma = 2.963$			$\rho_h = 0.574, \gamma = 2.964$			$\rho_h = 0.570, \gamma = 2.993$			$\rho_h = 0.570, \gamma = 2.999$		
0.75	3	45	**Total**	44.40		16.64	44.49		16.67	63.60		23.83	63.79		23.89
			Inp. Queue	23.48	52.88	8.79	23.51	52.85	8.81	41.98	65.99	15.73	42.01	65.86	15.74
			Processor	14.00	31.54	5.25	14.00	31.47	5.25	14.03	22.07	5.25	14.04	22.01	5.25
			Sub-Total	37.48	84.42	14.04	37.51	84.32	14.06	56.01	88.06	20.98	56.05	87.87	20.99
			Out. Queue	2.77	6.24	1.04	2.82	6.33	1.05	3.44	5.41	1.29	3.58	5.62	1.34
			Device	4.15	9.34	1.56	4.16	9.35	1.56	4.15	6.53	1.56	4.16	6.51	1.56
			Sub-Total	6.92	15.58	2.60	6.98	15.68	2.61	7.59	11.94	2.85	7.74	12.13	2.90
				$\rho_h = 0.722, \gamma = 2.951$			$\rho_h = 0.723, \gamma = 2.946$			$\rho_h = 0.717, \gamma = 2.984$			$\rho_h = 0.715, \gamma = 2.991$		
0.90	3	30	**Total**	111.02		41.65	110.99		41.65	188.09		70.56	189.04		70.91
			Inp. Queue	82.53	74.34	30.98	82.43	74.27	30.95	157.11	83.53	58.97	157.44	83.29	59.09
			Processor	16.81	15.14	6.30	16.81	15.15	6.30	16.86	8.96	6.30	16.86	8.91	6.30
			Sub-Total	99.34	89.48	37.28	99.24	89.42	37.25	173.97	92.49	65.27	174.30	92.20	65.39
			Out. Queue	6.65	5.99	2.49	6.73	6.07	2.52	9.11	4.84	3.41	9.73	5.15	3.64
			Device	5.03	4.53	1.88	5.02	4.51	1.88	5.01	2.67	1.88	5.01	2.65	1.88
			Sub-Total	11.68	10.52	4.37	11.75	10.58	4.40	14.12	7.51	5.29	14.74	7.80	5.52
				$\rho_h = 0.907, \gamma = 2.825$			$\rho_h = 0.904, \gamma = 2.837$			$\rho_h = 0.893, \gamma = 2.880$			$\rho_h = 0.888, \gamma = 2.900$		

Table 2c. Simulation results for layout 3 ($\gamma = 3$)

Handling/Processing Time Distribution				Constant/Uniform			Uniform/Uniform			Uniform/Exponential			Exponential/Exponential		
				Time in system			Time in system			Time in system			Time in system		
ρ_p	N	s	Component	time	%	WIP	time	%	WIP	time	%	WIP	time	%	WIP
0.60	5	400	**Total**	24.28		20.25	24.28		20.24	31.81		26.47	31.81		26.47
			Inp. Queue	9.23	37.99	7.67	9.22	37.96	7.66	16.53	51.97	13.75	16.53	51.95	13.75
			Processor	10.77	44.35	9.00	10.77	44.36	9.00	10.83	34.02	9.00	10.82	34.03	9.00
			Sub-Total	20.00	82.34	16.67	19.99	82.32	16.66	27.36	85.99	22.75	27.35	85.98	22.75
			Out. Queue	0.78	3.21	0.65	0.78	3.22	0.65	0.94	2.97	0.79	0.95	2.99	0.79
			Device	3.50	14.45	2.93	3.51	14.46	2.93	3.51	11.04	2.93	3.51	11.03	2.93
			Sub-Total	4.28	17.66	3.58	4.29	17.68	3.58	4.45	14.01	3.72	4.46	14.02	3.72
				$\rho_h = 0.669, \gamma = 3.060$			$\rho_h = 0.670, \gamma = 3.060$			$\rho_h = 0.667, \gamma = 3.090$			$\rho_h = 0.668, \gamma = 3.088$		
0.75	5	200	Total	41.98		35.01	42.02		35.04	60.96		50.76	61.02		50.81
			Inp. Queue	22.77	54.25	18.97	22.80	54.25	18.99	41.18	67.56	34.30	41.18	67.49	34.30
			Processor	13.46	32.06	11.25	13.46	32.03	11.25	13.53	22.19	11.25	13.53	22.17	11.25
			Sub-Total	36.23	86.31	30.22	36.26	86.28	30.24	54.71	89.75	45.55	54.71	89.66	45.55
			Out. Queue	1.88	4.47	1.56	1.89	4.51	1.57	2.38	3.92	1.98	2.44	4.00	2.03
			Device	3.87	9.22	3.23	3.87	9.21	3.23	3.87	6.35	3.23	3.87	6.34	3.23
			Sub-Total	5.75	13.69	4.79	5.76	13.72	4.80	6.25	10.25	5.21	6.31	10.34	5.26
				$\rho_h = 0.811, \gamma = 3.164$			$\rho_h = 0.811, \gamma = 3.161$			$\rho_h = 0.805, \gamma = 3.200$			$\rho_h = 0.806, \gamma = 3.199$		
0.90	5	125	Total	104.10		86.77	104.14		86.81	180.79		150.34	181.10		150.61
			Inp. Queue	78.67	75.58	65.55	78.66	75.54	65.54	153.55	84.93	127.69	153.57	84.80	127.72
			Processor	16.15	15.51	13.50	16.15	15.50	13.50	16.23	8.98	13.50	16.23	8.96	13.50
			Sub-Total	94.82	91.09	79.05	94.81	91.04	79.04	169.78	93.91	141.19	169.80	93.76	141.22
			Out. Queue	4.98	4.78	4.13	5.03	4.83	4.18	6.71	3.71	5.57	7.00	3.87	5.81
			Device	4.30	4.13	3.59	4.30	4.13	3.59	4.30	2.38	3.58	4.30	2.37	3.58
			Sub-Total	9.28	8.91	7.72	9.33	8.96	7.77	11.01	6.09	9.15	11.30	6.14	9.39
				$\rho_h = 0.953, \gamma = 3.228$			$\rho_h = 0.953, \gamma = 3.229$			$\rho_h = 0.940, \gamma = 3.286$			$\rho_h = 0.938, \gamma = 3.293$		

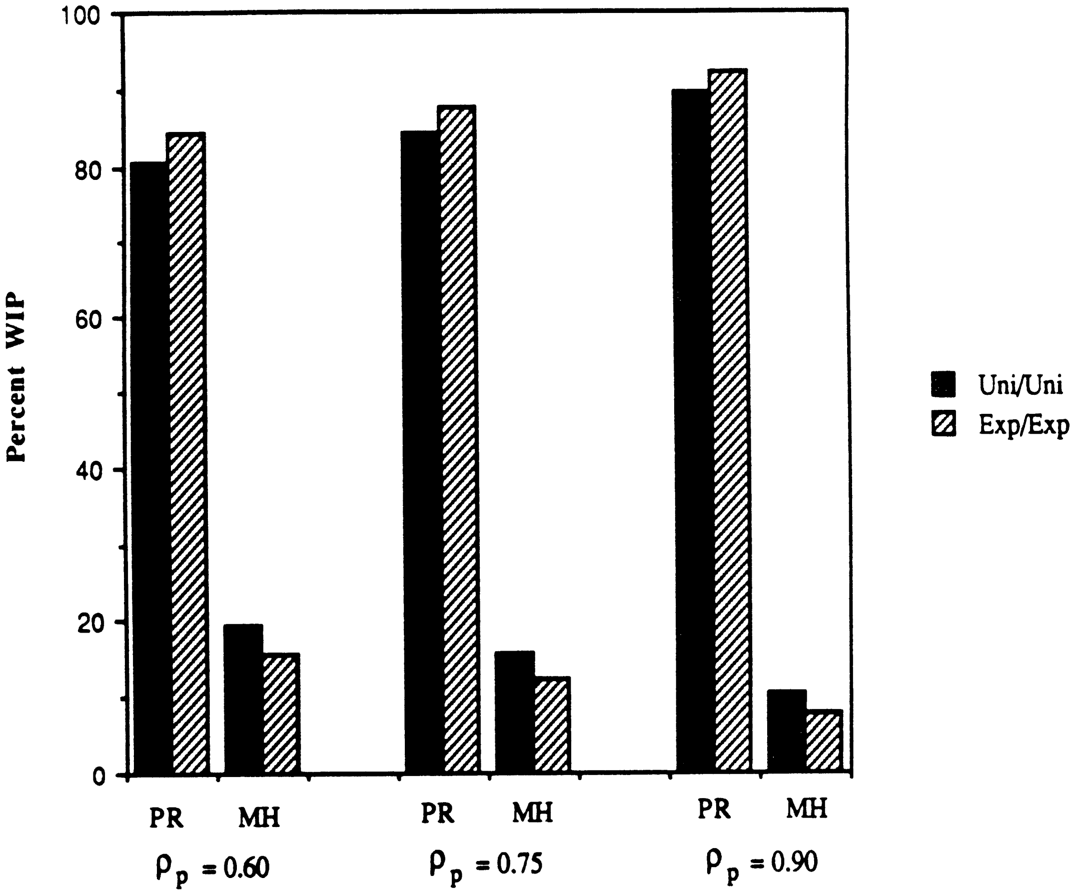

Figure 4: Break down of WIP for Table 2b

however, appears to contradict a well-known result in queueing. We will discuss this apparent conflict in section 4.2. The simulation results clearly indicate, however, that MH-WIP decreases as one uses fewer but faster devices. Cost and safety considerations aside, this result is a significant one for trip-based material handling systems.

4. DISCUSSIONS AND FURTHER RESULTS

In this section, based on some well-known results in queueing, we will present some intuitive explanations for the results we observed in section 3. We must stress that the explanations we offer are far from exact ones in a theoretical sense. That is, we still need to rely on the simulation model to substantiate our conclusions.

4.1. Processor WIP versus Material Handling WIP. The results in section 3 clearly indicate that, for all three layouts and for all handling/processing time distributions tested, the processors are responsible for WIP if $\rho_p \approx \rho_h$. Obviously, if $\rho_p > \rho_h$, PR-WIP will further increase relative to MH-WIP.

The above result can perhaps be best explained by first noting the "conceptual structure" of the queues and servers involved in the manufacturing system in question. As described in section 2, although each station has its own input queue and output queue, a load waiting in an input queue can be served only by the corresponding processor whereas a load waiting at an output queue can be served by any one of the devices (depending on the state of the system and the empty device dispatching rule used). That is, as shown in Figure 5, as far as the material handling system is concerned, there is actually a single (conceptual) queue where all the move requests wait for service. Note that, under MOD FCFS, the move requests are not served strictly on a FCFS basis.

Let us momentarily disregard the I/O stations and the move requests that queue up at the output queues of the I/O stations where no processing takes place. (Recall that loads which arrive from outside the system arrive at the output queue of an I/O station, and that loads delivered to the input queue of an I/O station instantaneously leave the system.) Also, let us suppose that the number of workstations, M, is equal to the number of devices, N. That is, $M = N = K$. Then, in reference to Figure 5, if all the arrival processes were Poisson and the service times exponential, the processors (and the input queues) would represent K M/M/1 queues, while the material handling devices (and the conceptual move request queue) would represent an M/M/K queue.

If the utilization of each processor is equal to ρ_p, the utilization of each device is equal to ρ_h, and $\rho_p = \rho_h$, then it is well-known in queueing that the total expected number of "customers" in the K M/M/1 queues, that is, PR-WIP, will be greater than the expected number of "customers" in the M/M/K queue, that is, MH-WIP. (See, for example, Gross and Harris [3, p. 149]. Recall that the number of "customers" include those that are being served.) An intuitive explanation for the above result (which is also known as "server pooling"), is that, in an M/M/K queue no server will be idle as long as the (conceptual) move request queue is non-empty. In contrast, with K M/M/1 queues, one or more servers may be idle although other servers in the system have non-empty queues.

In the above discussion we assumed that the number of processors is equal to the number of devices. In manufacturing applications where trip-based handling systems are

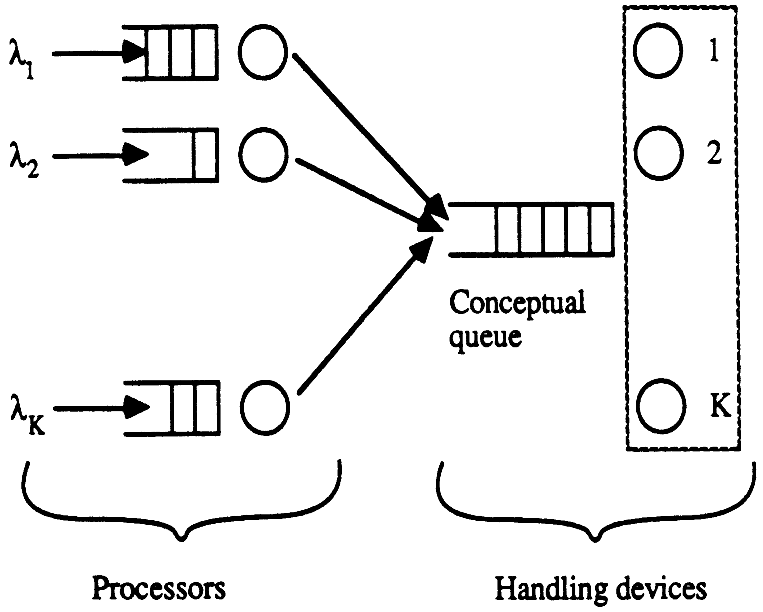

Figure 5: Depiction of K M/M/1 queues versus one M/M/K queue

used, such a situation would be somewhat unusual because the number of handling devices is usually less than the number of processors in most systems. However, one may use less devices than processors only if the average handling time is shorter than the average processing time; that is, $\gamma > 1$. Rather than change the layout (to reduce the average handling time), in section 3 we simply increased the speed of each device, and to maintain $\rho_p \approx \rho_h$ we reduced the number of devices.

In other words, in reference to Figure 5, we replaced the K servers of the material handling system with fewer but proportionally faster devices. Suppose we increase the speed of the device so that we need only one device to maintain approximately the same ρ_h value. It is again well-known in queueing that the expected number of "customers" in an M/M/1 queue is less than the expected number of "customers" in an M/M/K queue if the two systems have the same traffic intensity. (See, for example, Gross and Harris [3, p. 149]. Recall that the number of "customers" include those that are being served.)

Hence, as we reduce the number of devices by increasing their speed relative to the processing times, then the total WIP associated with the material handling system should decrease. Although we do not have Poisson arrivals (except for loads that arrive from outside the system, which we have excluded from our discussion), and we do not necessarily have exponential service times, in section 3 we clearly observed an increase in *percent* PR-WIP due to an absolute decrease in MH-WIP.

4.2. Number of Devices and Device Speed. From the results presented in section 3, and the arguments presented in section 4.1, it is clear that MH-WIP decreases when fewer and faster devices are used. However, as we remarked at the end of section 3, using fewer and faster devices leads to an apparent conflict with a well-known result in queueing.

Given an M/M/1 queue and an M/M/K queue, if the two queues have the same traffic intensity, that is, the single server is K times faster than each server in the multi-server queue, then the expected number of "customers" in the M/M/1 queue will be *less than* the expected number of "customers" in the M/M/K queue as we remarked earlier. However, the expected number of "customers" *waiting for service* in the M/M/1 queue will be *greater than* the expected number of "customers" waiting for service in the M/M/K queue. Hence, if the M/M/1 versus M/M/K analogy "applies" to our system, when fewer and faster devices are used, one will observe a decrease in the overall MH-WIP, and an *increase* in the number of loads waiting to be moved. (Recall that, in section 3, although we observed a decrease in the overall MH-WIP, we also observed a *decrease* in the number of loads waiting to be moved.)

This apparent conflict can be explained as follows. Recall that the material handling service time involves empty device travel during which the load being "served" is waiting at the output queue of a station. In order to avoid this unusual situation where the "customer" being served spends part of the service time waiting in the queue, one can redefine the service time as loaded travel only. As we noted earlier, with this definition, empty travel can be viewed as the "set-up time" incurred by the server before it can serve the next move request.

In an M/M/1 or M/M/K queue, as soon as service begins, the customer is removed from the queue, by definition, and there is no "set-up time" involved for the server(s).

Hence, we may test the "consistency" of our results simply by letting an empty device travel infinitely fast. With such an approach we would have no "set-up time" for the server and a load can be removed from the output queue as soon as service begins, i.e., as soon as the device is dispatched to pick up the load.[2] When we reran the simulation model with infinitely fast *empty* devices, the results we observed matched those that one would expect from the M/M/1 versus M/M/K analogy. That is, the overall MH-WIP still decreased when we used fewer but faster devices, while the number of loads waiting to be moved showed an increase. Hence, once the "set-up time" is removed from the simulation model, then the M/M/1 versus M/M/K analogy "applies," and our results are consistent with what is known for these two queues (although we do not have Poisson arrivals in our system and the service times are not necessarily exponential).

Of course, the "set-up time" (or empty travel) is an integral part of the system we are studying and therefore the M/M/1 versus M/M/K analogy does not hold. (We provided the preceding discussion to clarify what initially seemed to be a contradiction, and to explain the role of empty device travel.) One may still question why the number of loads waiting to be moved decreases as faster but fewer devices are used. We believe the following explanation provides the necessary insight.

Recall that empty travel occurs either when a loaded device, upon arriving at its destination, must pick up a job elsewhere in the system (i.e., device-initiated empty travel), or a job entering an output queue finds one or more idle devices (i.e., station-initiated empty travel). Unless excess material handling capacity has been provided, most of the empty travel performed by the devices is device-initiated.

Let v denote the device speed in a system with K devices. (By definition, if a single, faster device is used, its speed will be given by Kv.) Consider first device-initiated empty travel. Without loss of generality, suppose a loaded device delivers a load at station 1 and finds other jobs to pick up in the system except at output queue 1. Let P_i, $i \neq 1$, denote the probability that output queue i has the oldest (unassigned) move request. Also, let d_{1i} denote the distance from station 1 to station i. Then the "set-up time" for the single (faster) device is equal to $\sum_{i\neq 1} P_i d_{1i}/Kv$, while the "set-up time" for the multi-device system is equal to $\sum_{i\neq 1} P_i d_{1i}/v$. Hence, *provided that the P_i values remain approximately the same,* the "set-up time" for a single (faster) device is K times shorter, given that a non-zero set-up time is involved. (Obviously, if the output queue of station 1 contains an unassigned move request, then the "set-up time" under both systems will be equal to zero.)

Consider next station-initiated empty travel. Suppose a job enters output queue 1 and finds one or more idle devices. Let P'_i denote the probability that an idle device is located at station i. Since the locations of the idle devices can be considered to be independent in a multiple device system, the location of the longest idle device is assumed to be given by P'_i as well. Then the "set-up time" for the single (faster) device is equal to $\sum_i P'_i d_{i1}/Kv$, while it is equal to $\sum_i P'_i d_{i1}/v$ for the multi-device system. (It is implicit that $d_{11} = 0$

---

[2] As an alternate approach, we could have retained finite empty travel speed for the device and "remove" the load from the output queue as soon as the device was dispatched to pick it up. However, this would require a "logical" change in the simulation model, whereas letting empty devices travel infinitely fast requires a change only in the data file.

and that only one device needs to be assigned to the newly arriving move request.)

Thus, provided that the P_i and P_i' values do not change significantly from one system to the other, one would expect shorter "set-up times" in a system with fewer but faster devices. Since the "set-up time" (i.e., the empty device travel time) is closely related to the waiting times at the output queues, one would also expect shorter waiting times with fewer but faster devices.

The above result is partly based on the empty device dispatching rule we used for the study, namely, the MOD FCFS rule. Note that, under MOD FCFS, as one increases the number of devices, the probability of finding a move request that is closer (to the current location of the device) than the oldest move request is not likely to increase. This is primarily because, regardless of how far the oldest move request is located, a device will serve the oldest move request in the system if the current location has no unassigned move requests.

Note that the same is not true for the STTF rule which attempts to assign the closest unassigned move request to an empty device. Under the STTF rule, as one increases the number of devices, the probability of being able to find a move request closer than the oldest one is likely to increase. Hence, under the STTF rule, as the number of devices is increased (and their speed is proportionally reduced), the "set-up time" may not change (as much as it did with MOD FCFS) since the device travels slower but it is assigned to a closer move request.

We have not investigated the behavior of the "set-up time" under any dispatching rule other than the MOD FCFS rule, and our "intuition" about the above probabilities may not be correct. However, unless a significant increase in the expected "set-up time" is incurred, using faster and fewer devices is likely to reduce MH-WIP under most centralized dispatching rules. (Our simulation results indicate that this certainly holds true for the MOD FCFS rule.) Note that, when fewer and faster devices are used, the number of loads being moved — which is, by definition, part of MH-WIP — will always decrease regardless of the empty device dispatching rule used.

4.3. Increase in the Material Handling WIP. In this section we provide additional simulation results to investigate the values of ρ_p and ρ_h that yield an almost even break down between PR-WIP and MH-WIP. The average handling time is comparable to or slightly *longer* than the average processing time; that is, γ (which designates the ratio of the average processing time to the average handling time) varies between 0.80 and 0.95.

The results are shown in Tables 3a, 3b, and 3c, for layouts 1, 2, and 3, respectively. For the above range of γ values, we note that if $\rho_p = 0.50$, then ρ_h is approximately between 0.63 and 0.73 to have MH-WIP $\approx$ PR-WIP. That is, the utilization of the handling system would have to be approximately 26% to 46% larger than the utilization of the processors in order to have MH-WIP $\approx$ PR-WIP. If $\rho_p = 0.60$, then ρ_h can be as large as approximately 0.88 (i.e., 47% larger) in order to have MH-WIP $\approx$ PR-WIP.

In the above runs, the devices are fairly slow relative to the processing times since $0.80 < \gamma < 0.95$. Therefore, in a more realistic setting where one would use faster but fewer devices, one can further increase the above values reported for ρ_h and still have MH-WIP $\approx$ PR-WIP. (Recall that faster and fewer devices tend to reduce MH-WIP.)

It is important to stress that, the smallest γ value we used in Tables 1 through 3 is

Table 3a. Simulation results for layout 1 ($\gamma < 1$, $\rho_h > \rho_p$)

Handling/Processing Time Distribution				Constant/Uniform			Uniform/Uniform			Uniform/Exponential			Exponential/Exponential		
				Time in system			Time in system			Time in system			Time in system		
ρ_p	N	s	Component	time	%	WIP	time	%	WIP	time	%	WIP	time	%	WIP
0.5	5	3	**Total**	202.58		6.84	202.10		6.83	240.27		8.10	239.66		8.08
			Inp. Queue	35.75	17.65	1.20	36.36	17.99	1.22	71.09	29.59	2.39	71.67	29.90	2.41
			Processor	73.34	36.20	2.50	73.34	36.29	2.50	73.64	30.65	2.50	73.64	30.73	2.50
			Sub-Total	109.09	53.85	3.70	109.70	54.28	3.72	144.73	60.24	4.89	145.31	60.63	4.91
			Out. Queue	40.81	20.15	1.37	40.07	19.83	1.35	43.21	17.98	1.45	42.11	17.57	1.41
			Device	52.68	26.00	1.77	52.33	25.89	1.76	52.33	21.78	1.76	52.24	21.80	1.76
			Sub-Total	93.49	46.15	3.14	92.40	45.72	3.11	95.54	39.76	3.21	94.35	39.37	3.17
				$\rho_h = 0.631, \gamma = 0.934$			$\rho_h = 0.625, \gamma = 0.942$			$\rho_h = 0.626, \gamma = 0.945$			$\rho_h = 0.624, \gamma = 0.947$		
0.60	5	2	**Total**	317.09		10.69	315.94		10.65	392.98		13.23	399.88		13.46
			Inp. Queue	59.57	18.79	2.00	61.19	19.37	2.05	124.74	31.74	4.19	126.91	31.74	4.26
			Processor	88.18	27.81	3.00	88.18	27.91	3.00	88.53	22.53	3.00	88.53	22.14	3.00
			Sub-Total	147.75	46.60	5.00	149.37	47.28	5.05	213.27	54.27	7.19	215.44	53.88	7.26
			Out. Queue	89.75	28.30	3.01	88.11	27.89	2.96	101.25	25.77	3.40	106.14	26.54	3.56
			Device	79.59	25.10	2.68	78.46	24.83	2.64	78.46	19.96	2.64	78.30	19.58	2.64
			Sub-Total	169.34	53.40	5.69	166.57	52.72	5.60	179.71	45.73	6.04	184.44	46.12	6.20
				$\rho_h = 0.884, \gamma = 0.804$			$\rho_h = 0.875, \gamma = 0.812$			$\rho_h = 0.874, \gamma = 0.815$			$\rho_h = 0.868, \gamma = 0.821$		

Table 3b. Simulation results for layout 2 ($\gamma < 1$, $\rho_h > \rho_p$)

Handling/Processing Time Distribution				Constant/Uniform			Uniform/Uniform			Uniform/Exponential			Exponential/Exponential		
				Time in system			Time in system			Time in system			Time in system		
ρ_p	N	s	component	time	%	WIP	time	%	WIP	time	%	WIP	time	%	WIP
0.50	7	13	**Total**	28.74		10.78	28.71		10.76	33.09		12.41	32.99		12.38
			Inp. Queue	5.11	17.77	1.91	5.13	17.88	1.92	9.18	27.74	3.43	9.21	27.93	3.45
			Processor	9.29	32.32	3.50	9.29	32.34	3.50	9.28	28.06	3.50	9.29	28.15	3.50
			Sub-Total	14.40	50.09	5.41	14.42	50.22	5.42	18.46	55.80	6.93	18.50	56.08	6.95
			Out. Queue	6.07	21.13	2.27	6.03	21.00	2.25	6.37	19.23	2.38	6.22	18.86	2.33
			Device	8.27	28.78	3.10	8.26	28.78	3.09	8.26	24.97	3.10	8.27	25.06	3.10
			Sub-Total	14.34	49.91	5.37	14.29	49.78	5.34	14.63	44.20	5.48	14.49	43.92	5.43
				$\rho_h = 0.738, \gamma = 0.820$			$\rho_h = 0.737, \gamma = 0.822$			$\rho_h = 0.734, \gamma = 0.826$			$\rho_h = 0.733, \gamma = 0.825$		
0.60	7	10	**Total**	40.56		15.20	40.55		15.20	48.90		18.33	49.16		18.43
			Inp. Queue	8.99	22.16	3.36	9.04	22.29	3.38	16.41	33.54	6.14	16.52	33.61	6.18
			Processor	11.16	27.51	4.20	11.16	27.52	4.20	11.16	22.82	4.20	11.16	22.70	4.20
			Sub-Total	20.15	49.67	7.56	20.20	49.81	7.58	27.55	56.36	10.34	27.68	56.31	10.38
			Out. Queue	10.36	25.54	3.88	10.32	25.44	3.86	11.30	23.11	4.23	11.43	23.26	4.29
			Device	10.05	24.79	3.76	10.04	24.75	3.76	10.05	20.53	3.76	10.05	20.43	3.76
			Sub-Total	20.41	50.33	7.64	20.36	50.19	7.62	21.35	43.64	7.99	21.48	43.69	8.05
				$\rho_h = 0.890, \gamma = 0.819$			$\rho_h = 0.887, \gamma = 0.821$			$\rho_h = 0.882, \gamma = 0.826$			$\rho_h = 0.879, \gamma = 0.830$		

Table 3c. Simulation results for layout 3 ($\gamma < 1$, $\rho_h > \rho_p$)

Handling/Processing Time Distribution				Constant/Uniform			Uniform/Uniform			Uniform/Exponential			Exponential/Exponential		
				Time in system			Time in system			Time in system			Time in system		
ρ_p	N	s	Component	time	%	WIP	time	%	WIP	time	%	WIP	time	%	WIP
0.50	13	40	Total	25.04		20.95	24.92		20.84	29.05		24.29	28.82		24.09
			Inp. Queue	5.04	20.11	4.19	5.04	20.22	4.19	9.02	31.06	7.52	9.05	31.41	7.54
			Processor	8.92	35.64	7.50	8.92	35.82	7.50	8.93	30.75	7.50	8.93	31.00	7.50
			Sub-Total	13.96	55.75	11.69	13.96	56.04	11.69	17.95	61.81	15.02	17.98	62.41	15.04
			Out. Queue	4.52	18.04	3.76	4.43	17.75	3.68	4.56	15.71	3.80	4.30	14.92	3.58
			Device	6.56	26.21	5.50	6.53	26.21	5.47	6.54	22.48	5.47	6.54	22.67	5.47
			Sub-Total	11.08	44.25	9.26	10.96	43.96	9.15	11.10	38.19	9.27	10.84	37.59	9.09
				$\rho_h = 0.738$, $\gamma = 0.884$			$\rho_h = 0.734$, $\gamma = 0.889$			$\rho_h = 0.731$, $\gamma = 0.893$			$\rho_h = 0.732$, $\gamma = 0.893$		
0.60	13	30	Total	34.39		28.76	34.38		28.74	41.95		35.04	41.80		34.91
			Inp. Queue	9.03	26.26	7.53	9.08	26.42	7.57	16.16	38.52	13.46	16.21	38.77	13.50
			Processor	10.73	31.20	9.00	10.73	31.21	9.00	10.74	25.60	9.00	10.74	25.69	9.00
			Sub-Total	19.76	57.46	16.53	19.81	57.63	16.57	26.90	64.12	22.46	26.95	64.46	22.50
			Out. Queue	6.91	20.09	5.76	6.84	19.89	5.70	7.32	17.46	6.11	7.12	17.05	5.94
			Device	7.72	22.45	6.47	7.73	22.48	6.47	7.73	18.42	6.47	7.73	18.49	6.47
			Sub-Total	14.63	42.54	12.23	14.57	42.37	12.17	15.05	35.88	12.58	14.85	35.54	12.41
				$\rho_h = 0.893$, $\gamma = 0.881$			$\rho_h = 0.891$, $\gamma = 0.882$			$\rho_h = 0.886$, $\gamma = 0.889$			$\rho_h = 0.885$, $\gamma = 0.890$		

equal to approximately 0.80. If the processing times are considerably shorter than the handling times, i.e., $\gamma < 0.80$, then the number of devices in the system will exceed the number of processors. In such cases, even if $\rho_p \approx \rho_h$, the MH-WIP may very well exceed the PR-WIP since the number of "customers" in K_1 M/M/1 queues may be less than the number of "customers" in an M/M/K_2 queue, if K_2 is sufficiently larger than K_1. However, if the fastest devices available are still slow relative to the processors, then a trip-based handling system is perhaps not a suitable handling system in such cases. One must consider using conveyors or stop-and-go AGV systems. (In the latter case, every load is permanently assigned to an AGV which functions as a "work platform" that travels with the load from one processor station to another. Among other applications, stop-and-go AGVs are being used in the automotive industry.)

4.4. Reducing the Workstation WIP. A remaining question is the following: if the processors are responsible for WIP, what happens when one increases the processing speeds? Does one obtain a "real" reduction in total WIP, or does the WIP simply get transferred from the input queues to the output queues at each processor station? (Note that we are *not* increasing the rate at which loads arrive from outside the system; i.e., we are not increasing the "workload" in the system. We are only increasing the processing speeds in order to reduce PR-WIP.)

Since the "workload" has not changed, the average number of loads that must be moved per time unit by the material handling system will not change. However, the distribution of the loads arriving at the output queues of the processing stations will change when the processing speeds are increased. Results presented in previous sections indicate that the performance of the handling system is generally not sensitive to the processing time distribution. Hence, maintaining the same processing time distribution while increasing the processing speeds may not affect MH-WIP.

The results shown in Table 4 (and Figure 6) clearly indicate that the above is true. That is, MH-WIP does not increase as the processing speed is increased. Consequently, a "real" reduction in total WIP is obtained since PR-WIP decreases with increased processing speed. (Obviously *percent* MH-WIP increases because PR-WIP decreases while MH-WIP remains approximately the same. This result also supports our argument in section 4.3 that, as the processors operate faster and faster relative to the devices, MH-WIP may eventually exceed PR-WIP.)

Recall that we assumed only one processor at each processing station. One may question the impact of this assumption on the results obtained in this study. The M/M/1 versus M/M/K analogy implies that, if we use slower but parallel processors at each processor station, then the number of loads waiting to be processed will decrease, but the number of loads being processed will increase, leading to an overall increase in PR-WIP. Hence, assuming one processor at each processor station, gives the workstations the "maximum benefit" in terms of PR-WIP.

Lastly, we would like to stress that we used the M/M/1 versus M/M/K analogy (as well as the K M/M/1 versus M/M/K analogy) only in an approximate manner. That is, our arguments are far from rigorous in a theoretical sense. However, the simulation model confirmed and shaped our expectations which were mostly guided by these well-known queueing models.

Table 4. Simulation results with reduced processing times

Handlg./Processg. Time Distn.	Constant/Uniform			Uniform/Uniform			Uniform/Exponential			Exponential/Exponential		
	Time in system			Time in system			Time in system			Time in system		
Component	time	%	WIP	time	%	WIP	time	%	WIP	time	%	WIP
Total	431.27 (850.55)		14.53 (28.64)	430.04 (848.96)		14.49 (28.59)	80.29 (186.23)		30.09 (69.95)	70.94 (181.22)		59.23 (151.98)
Input Queue	154.87	35.91	5.20	156.02	36.28	5.24	41.19	51.30	15.42	41.01	57.81	34.22
Processor	110.44	25.61	3.75	110.44	25.68	3.75	13.95	17.36	5.25	13.46	18.98	11.25
Sub-Total	265.31	61.52 (80.84)	8.95 (23.16)	266.46	61.96 (81.14)	8.99 (23.21)	55.14	68.66 (86.43)	20.67 (60.48)	54.47	76.79 (90.86)	45.47 (138.12)
Output Queue	86.36	20.02	2.90	85.12	19.80	2.86	11.83	14.75	4.43	7.79	10.97	6.49
Device	79.60	18.46	2.68	78.46	18.24	2.64	13.32	16.59	4.99	8.68	12.24	7.27
Sub-Total	165.96	38.48 (19.16)	5.58 (5.48)	163.58	38.04 (18.86)	5.50 (5.38)	25.15	31.34 (13.57)	9.42 (9.47)	16.47	23.21 (9.14)	13.76 (13.86)
	$\rho_h = 0.897$, $\gamma = 0.991$ (0.910) (1.176)			$\rho_h = 0.887$, $\gamma = 1.002$ (0.897) (1.192)			$\rho_h = 0.863$, $\gamma = 0.739$ (0.866) (0.885)			$\rho_h = 0.891$, $\gamma = 0.962$ (0.893) (1.152)		
	$\rho_p = 0.90$, $N = 5$, $s = 2$			$\rho_p = 0.90$, $N = 5$, $s = 2$			$\rho_p = 0.90$, $N = 10$, $s = 7$			$\rho_p = 0.90$, $N = 15$, $s = 25$		
	Layout 1			Layout 1			Layout 2			Layout 3		

Note: Numbers in parantheses are the values in Table 1 that were observed before the processing times were reduced.

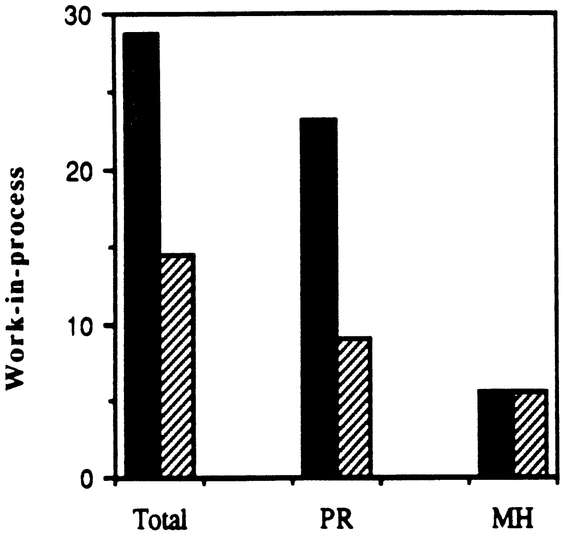

Figure 6a: Layout 1 (Constant/Uniform)

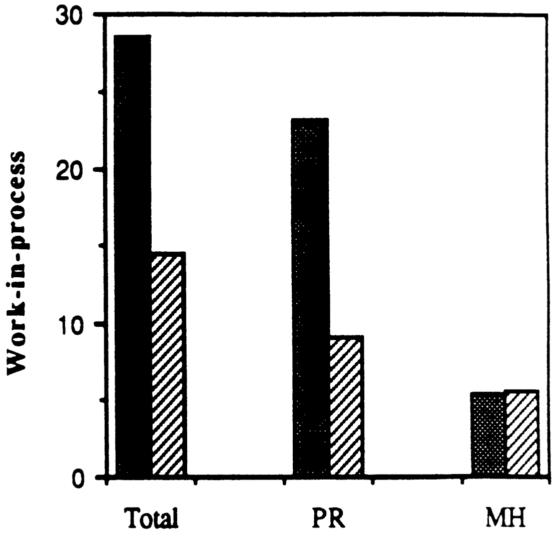

Figure 6b: Layout 1 (Uniform/Uniform)

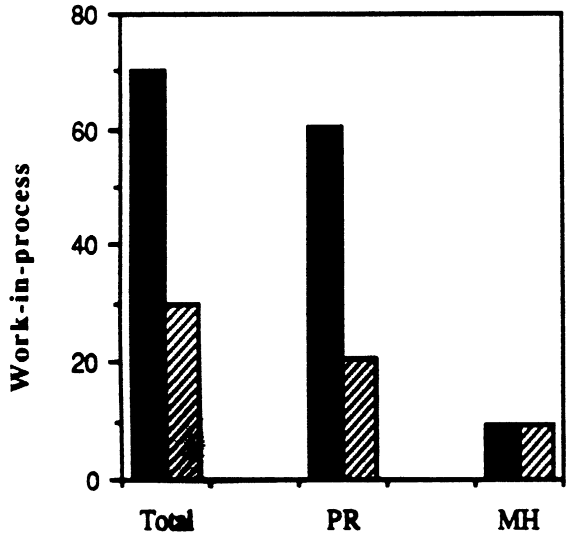

Figure 6c: Layout 2 (Uniform/Expontl.)

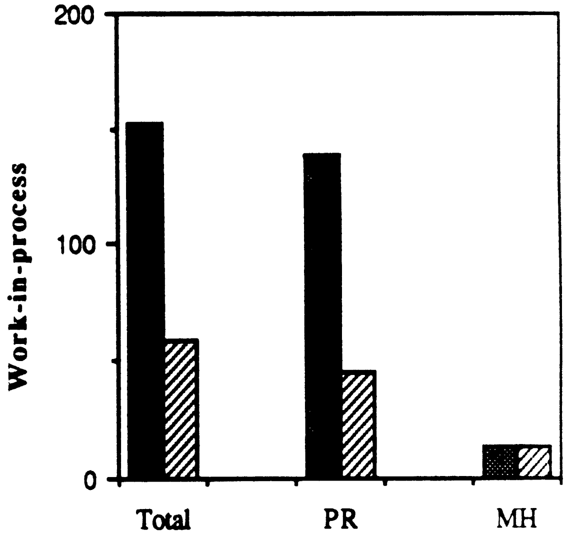

Figure 6d: Layout 3 (Expontl./Expontl.)

■ Before Reduction
▨ After Reduction

Figure 6: Effect of processing time on WIP (from Table 4)

5. CONCLUSIONS

In this paper we studied a manufacturing system that consists of two components: the workstations (or processors) where processing takes place, and a trip-based material handling system which moves the loads from one processor station to another.

Assuming infinite space for the queues, we used a simulation model to show empirically that the processors are largely responsible for WIP even if the utilization of the processors is approximately equal to the utilization of the trip-based handling system. That is, we empirically showed that the WIP associated with the processors (i.e., PR-WIP, which is measured by the number of loads waiting to be processed plus the number of loads being processed) usually far exceeds the WIP associated with the material handling system (i.e., MH-WIP, which is measured by the number of loads waiting to be transported plus the number of loads being transported).

We also empirically showed that: 1. reducing the variance of the processing times has a greater impact on overall WIP than reducing the variance of the handling times, 2. aside from cost and safety considerations, using fewer and faster handling devices reduces MH-WIP, and 3. increasing the processor speeds reduces PR-WIP without increasing MH-WIP.

We explained the above results and offered additional insight on the role that empty device travel plays in MH-WIP. We also attempted to relate some of our observations to well-known tradeoffs between M/M/1 versus M/M/K queues, as well as K M/M/1 queues versus an M/M/K queue. Although the arguments we used are far from rigorous in a theoretical sense, the results we observed from the simulation model mostly agree with the results known for the above queues.

Our results regarding the distribution of WIP does *not* imply that one need not be concerned with the material handling system in studying or designing a manufacturing system. Note that the material handling system should not be responsible for WIP *as long as it is not a bottleneck.* Designing a trip-based material handling system with the right type and number of devices, the right dispatching rule, and the right flow patterns (such as the guidepath of an AGV system) is still an area where more research is needed.

APPENDIX

Table A1. Routing matrix of jobs in Layout 1.

Station No.	1	2	3	4	5	6	7
1	0.0	0.0	0.5	0.5	0.0	0.0	0.0
2	0.0	0.0	0.0	0.0	0.0	0.0	0.0
3	0.5	0.0	0.0	0.0	0.5	0.0	0.0
4	0.0	0.0	0.0	0.0	0.3	0.7	0.0
5	0.0	0.5	0.1	0.0	0.0	0.4	0.0
6	0.2	0.0	0.5	0.0	0.0	0.0	0.3
7	0.0	0.3	0.1	0.6	0.0	0.0	0.0

Table A2. Vehicle travel distance between stations in Layout 1, distance units.

Station No.	1	2	3	4	5	6	7
1	0	62	16	42	36	28	48
2	58	0	38	64	44	36	16
3	64	46	0	26	50	42	62
4	38	50	18	0	24	16	36
5	50	26	30	56	0	28	42
6	22	84	38	64	58	0	70
7	42	54	58	84	28	56	0

Table A3. Mean interarrival time of jobs at the input station in Layout 1, mins/job.

Input station	**Interarrival time**
1	30
2	∞

Table A4. Routing matrix of jobs in Layout 2.

Station No.	1	2	3	4	5	6	7	8	9	10	11
1	0.0	0.0	0.0	0.0	0.2	0.2	0.0	0.2	0.2	0.0	0.2
2	0.0	0.0	0.0	0.0	0.2	0.1	0.1	0.2	0.1	0.1	0.2
3	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
4	0.0	0.0	0.0	0.0	0.2	0.4	0.0	0.1	0.1	0.1	0.1
5	0.1	0.1	0.1	0.2	0.0	0.1	0.0	0.1	0.1	0.1	0.1
6	0.0	0.0	0.1	0.1	0.1	0.0	0.2	0.1	0.2	0.2	0.0
7	0.0	0.1	0.2	0.1	0.0	0.2	0.0	0.1	0.2	0.0	0.1
8	0.2	0.0	0.0	0.0	0.2	0.1	0.1	0.0	0.2	0.1	0.1
9	0.0	0.2	0.1	0.0	0.1	0.2	0.1	0.2	0.0	0.1	0.0
10	0.0	0.2	0.2	0.1	0.0	0.1	0.0	0.1	0.1	0.0	0.2
11	0.2	0.0	0.1	0.1	0.2	0.0	0.1	0.0	0.1	0.2	0.0

Table A5. Vehicle travel distance between stations in Layout 2, distance units.

Station No.	1	2	3	4	5	6	7	8	9	10	11
1	0	22	47	30	14	32	46	8	27	31	17
2	22	0	36	29	23	24	38	14	16	20	14
3	47	36	0	33	37	19	12	39	27	16	41
4	30	29	33	0	16	14	21	25	13	28	27
5	14	23	37	16	0	18	32	16	17	32	18
6	32	24	19	14	18	0	14	27	8	23	29
7	46	38	12	21	32	14	0	41	22	18	43
8	8	14	39	25	16	27	41	0	19	23	9
9	27	16	27	13	17	8	22	19	0	15	21
10	31	20	16	28	32	23	18	23	15	0	25
11	7	14	41	27	18	29	43	9	21	25	0

Table A6. Mean interarrival time of jobs at the input stations in Layout 2, mins/job.

Input station	Interarrival time
1	4.9
2	9.8
3	∞
4	14.7

Table A7. Routing matrix of jobs in Layout 3.

Station No.	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
1	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.2	0.0	0.2	0.0	0.0	0.2	0.2	0.0	0.2	0.0	0.0	0.0
2	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.2	0.0	0.0	0.2	0.2	0.0	0.2	0.0	0.2	0.0
3	0.0	0.0	0.0	0.0	0.0	0.2	0.0	0.0	0.0	0.0	0.2	0.2	0.2	0.0	0.0	0.0	0.0	0.0	0.2	0.0
4	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.2	0.0	0.2	0.0	0.0	0.2	0.0	0.0	0.2	0.0	0.2	0.0	0.0
5	0.0	0.0	0.0	0.0	0.0	0.2	0.2	0.2	0.2	0.2	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
6	0.15	0.0	0.2	0.0	0.0	0.0	0.0	0.2	0.0	0.0	0.2	0.05	0.1	0.0	0.0	0.0	0.0	0.0	0.1	0.0
7	0.0	0.0	0.2	0.1	0.0	0.1	0.0	0.2	0.0	0.1	0.0	0.1	0.1	0.0	0.0	0.0	0.0	0.0	0.0	0.1
8	0.0	0.0	0.0	0.1	0.2	0.0	0.1	0.0	0.0	0.0	0.1	0.0	0.1	0.0	0.0	0.1	0.0	0.1	0.0	0.2
9	0.2	0.05	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.2	0.1	0.15	0.0	0.1	0.0	0.1	0.0	0.1	0.0
10	0.0	0.0	0.1	0.2	0.0	0.0	0.1	0.1	0.0	0.0	0.0	0.1	0.1	0.0	0.0	0.1	0.0	0.1	0.0	0.1
11	0.0	0.1	0.1	0.0	0.0	0.1	0.0	0.0	0.0	0.0	0.0	0.2	0.0	0.2	0.1	0.0	0.0	0.0	0.2	0.0
12	0.0	0.0	0.15	0.15	0.0	0.05	0.0	0.1	0.0	0.1	0.0	0.0	0.1	0.05	0.0	0.1	0.0	0.1	0.0	0.1
13	0.0	0.0	0.0	0.05	0.2	0.0	0.1	0.1	0.1	0.1	0.0	0.0	0.0	0.05	0.0	0.0	0.0	0.2	0.0	0.1
14	0.2	0.0	0.0	0.0	0.1	0.1	0.1	0.0	0.1	0.0	0.1	0.1	0.0	0.0	0.0	0.0	0.1	0.0	0.1	0.0
15	0.1	0.05	0.05	0.0	0.0	0.1	0.1	0.0	0.1	0.0	0.0	0.0	0.0	0.2	0.0	0.0	0.2	0.0	0.1	0.0
16	0.0	0.0	0.0	0.2	0.1	0.0	0.0	0.1	0.0	0.1	0.0	0.0	0.1	0.0	0.0	0.0	0.0	0.2	0.0	0.2
17	0.1	0.1	0.0	0.0	0.0	0.05	0.0	0.0	0.0	0.0	0.15	0.0	0.0	0.2	0.2	0.0	0.0	0.0	0.2	0.0
18	0.0	0.0	0.2	0.1	0.0	0.0	0.1	0.0	0.0	0.1	0.0	0.0	0.1	0.0	0.0	0.1	0.0	0.0	0.0	0.3
19	0.2	0.1	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.25	0.1	0.1	0.0	0.05	0.1	0.0	0.1	0.0	0.0	0.0
20	0.0	0.0	0.0	0.14	0.0	0.0	0.0	0.0	0.0	0.2	0.0	0.0	0.2	0.06	0.0	0.2	0.0	0.2	0.0	0.0

Table A8. Vehicle travel distance between stations in Layout 3, distance units.

station No.	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
1	0	46	52	60	28	42	36	61	20	55	28	38	91	54	26	87	35	91	40	74
2	14	0	46	54	42	36	50	55	34	49	22	32	85	8	20	81	9	85	14	68
3	48	34	0	38	26	20	34	39	28	33	36	16	69	42	34	65	43	69	28	52
4	90	76	42	0	68	62	76	41	70	35	78	58	31	84	76	27	85	31	70	14
5	72	58	24	32	0	24	8	33	32	27	40	40	63	66	38	59	47	63	52	46
6	48	34	40	48	36	0	44	49	8	43	16	26	79	42	14	75	23	79	28	62
7	64	50	16	24	22	16	0	25	24	19	32	32	55	58	30	51	39	55	44	38
8	69	55	21	29	27	21	35	0	29	24	37	37	60	63	35	56	44	60	49	43
9	40	26	32	40	28	22	36	41	0	35	8	18	71	34	6	67	15	71	20	54
10	75	61	27	25	33	27	41	6	35	0	43	43	56	69	41	52	50	56	55	39
11	62	48	24	32	20	14	28	33	22	27	0	10	63	56	28	59	37	63	42	46
12	62	48	14	22	40	34	48	23	42	17	50	0	53	56	48	49	57	53	42	36
13	79	65	31	9	37	31	45	10	39	4	47	47	0	73	45	36	54	40	59	23
14	6	32	38	46	34	28	42	47	26	41	14	24	77	0	12	73	21	77	26	60
15	34	20	66	74	62	56	70	75	54	69	42	52	105	28	0	101	9	105	14	88
16	83	69	35	13	41	35	49	14	43	8	51	51	4	77	49	0	58	44	63	27
17	35	21	67	75	63	57	71	76	55	70	43	53	106	29	41	102	0	106	15	89
18	89	75	41	19	47	41	55	20	49	14	57	57	10	83	55	6	64	0	69	13
19	20	6	52	60	48	42	56	61	40	55	28	38	91	14	26	87	15	91	0	74
20	76	62	28	26	54	48	62	27	56	21	64	44	17	70	62	13	71	17	56	0

Table A9. Mean interarrival time of jobs at the input stations in Layout 3, mins/job.

Input station	Interarrival time
1	2.75
2	5.50
3	8.25
4	11.00
5	13.75

BIBLIOGRAPHY

[1] Chow, W. M., "Design for Line Flexibility," *IIE Transactions*, Vol. 18, No. 1, 1986, pp. 95–108.

[2] Egbelu, P. J. and Tanchoco, J. M. A., "Characterization of Automatic Guided Vehicle Dispatching Rules," *International Journal of Production Research*, Vol. 22, No. 3, 1984, pp. 359–374.

[3] Gross, D. and Harris, C. M., *Fundamentals of Queueing Theory*, Second Edition, Wiley, 1985.

[4] Hodgson, T. J., King, R. E., and Monteith, S. K., "Extracting Heuristic Control Rules for AGV's using Markov Decision Processes," *Material Flow*, Vol. 4, 1987, pp. 85–96.

[5] Law, A. M. and Kelton, W. D., *Simulation Modeling and Analysis*, Second Edition, McGraw-Hill, 1991.

[6] Merchant, M. E., "The Factory of the Future," paper presented at the Annual Meeting of the American Association of Engineering Societies, Port St. Lucie, Florida, 1982.

[7] Shanthikumar, J. G. and Stecke, K. E., "Reducing Work-In-Process Inventory in Certain Classes of Flexible Manufacturing Systems," *European Journal of Operational Research*, Vol. 26, 1986, pp. 266–271.

[8] Srinivasan, M. M., Bozer, Y. A., and Cho, M. S., "Trip-Based Material Handling Systems: Throughput Capacity Analysis," Technical Report 89–32, Department of Industrial and Operations Engineering, The University of Michigan, Ann Arbor, MI 48109-2117, 1989.

[9] Suri, R. and Hildebrant, R. R., "Modeling Flexible Manufacturing Systems using Mean-Value Analysis," *SME Journal of Manufacturing Systems*, Vol. 3, No. 1, 1984, pp. 27–38.

[10] Tompkins, J. A. and White, J. A., *Facilities Planning*, John Wiley & Sons, NY, 1984.

[11] Toro Ramos, Z. R. and McGinnis, L. F., "Performance Modeling for a Single Material Handling Device with Random Service Requests: Part I: Pure Blocking," *Proceedings of the 1990 Material Handling Research Colloquium*, Hebron, Kentucky, 1990, pp. 271–283.

[12] Yao, D. D. and Buzacott, J. A., "Modeling the Performance of Flexible Manufacturing Systems," *International Journal of Production Research*, Vol. 23, No. 5, 1985, pp. 945–959.

[13] Yao, D. D. and Buzacott, J. A., "Modeling a Class of Flexible Manufacturing Systems with Reversible Routing, *Operations Research*, Vol. 35, No. 1, 1987, pp. 87–93.