Determining Transfer Barch Sizes to Minimize Work-in-Process in Manufacturing Systems

Jonghwa Kim
Department of Industrial Engineering
Konkuk University

ABSTRACT

Trip-based material handling systems such as AGV systems, lift trucks, etc. are often designed with a given flow matrix (or FROM-TO chart) which is usually treated as the number of loaded trips that the devices must perform per unit time between the stations. In reality, the number of trips that would result from parts flow in a facility is dictated by the transfer batch size, i.e., the number of parts that are transferred from one station to the next in one trip. In this paper, we present analytical and simulation results aimed at determining optimal or near-optimal transfer batch sizes in manufacturing systems.

1. INTRODUCTION

In this study, we are considering trip-based material handling systems. They consist of one or more handling devices that are self-powered and operate independently from each other [16]. Examples include Unit Load AGV systems, lift trucks, microload AS/RS, bridge cranes, and so on, which are used in (flexible) manufacturing systems. In a trip-based handling system, it is assumed that a device performs a trip to move the unit loads one at a time. The *transfer batch size* (TBS) is defined as the number of parts in a unit load. In this study, we show that the TBS has a significant impact on the performance of the handling devices and the work-in-process (WIP) level in the system.

2 PROBLEM DESCRIPTION AND MOTIVATION

In describing the transfer batch sizing problem, a *part* refers to the smallest unit processed individually in the system. Parts are handled in a *container*, which is the smallest possible unit moved by a device on one trip. A container may hold one or more parts in it (depending on the size and weight of the part). For simplicity, we assume that a container holds only one part. If a container holds more than one part, the above assumption can be relaxed simply by setting the TBS of that part type equal to an integer multiple of the container size. A *unit load* refers to a collection of containers moved together by a device on one trip. Hence, the TBS is the number of parts and/or containers in a unit load, which may vary depending on the part type.

Consider next the manufacturing system. We define two types of stations: input/output(I/O) stations and processing stations. Each station has a dedicated *input queue* and *output queue* of infinite capacity. Parts arriving from outside the system are directly placed in the output queue of an I/O station, while parts that require no further processing leave the system instantly through the input queue of an I/O station. Each external arrival occurs according to a Poisson process and consists of one or more parts depending on the TBS of that part type. At a *processing station*, parts are removed from the input queue, one part at a time, by the FCFS rule and processed for a given period. After a part is processed, it is *staged* by the processor until the desired TBS is reached, at which point the parts are placed in the output queue as a unit load.

A unit load placed in an output queue (i.e., a "move request") must wait for a device. When assigned to a particular move request, the handling device first travels empty to the output queue where the move request is located and then it delivers the load to the appropriate input queue. The next move request to be served by an empty device is determined by the FCFS rule. When a device becomes empty, if there are no unassigned move requests, the device becomes idle at its last delivery point. Also, at the time a move request occurs, if more than one idle device is available, the oldest idle device is dispatched.

In this study, we show that, for a fixed throughput rate and fixed number of devices, the expected WIP in the input queues increases with the TBS. This is primarily due to "bulk arrivals," i.e., as the TBS increases, the expected number of parts per arrival instance at each input queue increases. On the other hand, the expected WIP in the output queues generally decreases with the TBS since the number of trips that the devices must perform per unit time decreases as the TBS is increased. Given the above tradeoff, we determine the optimal TBS to minimize the total expected WIP in the system. For this purpose, we develop separate analytical models to estimate the expected WIP in the input and output queues. For the latter, we explicitly capture the empty device travel time with the FCFS device dispatching rule.

The assumptions for the study are as follows:

1. Multiple part types are processed in the system and the TBS of each part is determined independently.
2. No set up times are considered.
3. Each station has sufficient processing capacity and is utilized less than 100.
4. At each input queue, unit loads are delivered according to a Poisson process.
5. Move requests occur according to a Poisson process.
6. The layout of the system, the production route for each part, the throughput requirement for each part, and the speed of the handling devices are assumed to be given.

7. The first and second moment of the travel time distribution from one station to another is given.
8. The devices are homogeneous and move one unit load at a time; a unit load includes only one part type.

Assumptions 4 and 5 are perhaps not entirely valid. However, given a sufficient number of stations and the randomness induced by different production routes, our results indicate that the coefficient of variation for the interarrival times at the input queue is consistently close to one, which is a necessary but not sufficient condition for a Poisson process. Superimposing the move requests generated at many output queues, on the other hand, would approach a Poisson process (see Kuehn[10]).

3. LITERATURE REVIEW

The expected WIP associated with the handling system depends in part on the dispatching rule. There are few analytical models that take into account a specific dispatching rule. Chow [4,5] presents an analytical model to approximate the device utilization and the expected (output queue) waiting times in a single-device system using M/G/1 queue. However, the expected waiting time is obtained only for a single "conceptual" queue (as opposed to individual output queues). The author does not extend the model to multiple device systems.

Cho[3] also uses the M/G/1 queue with FCFS service to model single-device systems. The author empirically shows that the above model works well in single-device systems. For systems with N devices, he uses the same model and increases the speed of the device N times. Although the "speed up" works well for estimating the utilization of the devices, it underestimates the expected waiting times in the output queues.

Yao and Buzacott[17] model the material handling system as a central server station. As in Chow's model, the expected waiting times in the output queues are estimated only for a single queue that forms ahead of the central station. Furthermore, the travel time distribution does not vary by the origin and destination of a particular move request. Also, the probability that a unit load will be routed to a particular station does not depend on where the unit load is picked up. Solberg[14] and Solot[15] also models the material handling system as another station.

Bertrand[1] uses CAN-Q model developed by Solberg[13] to estimate the expected time in the system for each job. The author develops an iterative algorithm to determine the optimal production batch size and shows that ignoring the WIP carrying cost may result in considerably larger production batch sizes. He models the material handling system as another station.

Bozer, *et. al.*[2] develop an iterative algorithm to estimate the expected waiting times in the output queues for a single device operating under the MOD FCFS rule. They observe that the algorithm fails to converge only when the utilization of the device is around 0.99 or higher. The authors do not extend the model to multiple-device systems.

Egbelue[6] presents an optimization model to determine the container size (i.e., the TBS) and the number of handling devices. He assumes that only one container type is selected for all the part types. Thus, the weight capacity of the selected container dictates the TBS for each part type; that is,

$$\textit{TBS of part type } j = \left\lfloor \frac{\textit{(the weight capacity of selected container)}}{\textit{(the weight of part type } j\textit{)}} \right\rfloor$$

For example, if the container holds 500 pounds and a part weighs 100 pounds, then the TBS of that part type is set equal to five and TBSs of one through four are not evaluated. For each candidate container size, the author first uses a simulation model to estimate the expected WIP in the system as a function of the number of devices. He then uses the above estimates in an optimization model to determine the optimal container size based on various cost elements. In a subsequent model, Egbelu[7] extends his approach to determine the number of processors required.

The problem we address here is similar to the one studied by Egbelu[6]. However, instead of simulation, we use analytical models to estimate the expected WIP in each queue for given TBSs. For each part type we allow any integer number as the TBS.

4. ANALYTICAL MODEL TO ESTIMATE WIP

In this section, we develp an analytical model, which we call "the WIP model" to estimate the expected WIP in each queue. The following notation is used throughout the paper. Let

M = number of stations in the system
JT = number of part types in the system
D_k = demand for part type k (parts/time unit) (= production rate)
Q_k = transfer batch size of part type k
R_i = set of part types which require processing at station i = $\{k \mid$ part type k visits station $i\}$
Ω_k = the set of stations at which part type k requires processing = $\{i \mid$ station i is visited by part type $k\}$
Λ_i = arrival rate at the input queue of station i (unit loads/time unit)
Λ_T = sum of the arrival rates at the input queues of stations $(\sum_{i=1}^{M} \Lambda_i)$
$E(S_i)$ = expected processing time at station i (time units/part)
$E(S_i^{(2)})$ = second moment of processing time at station i
$E(N_i)$ = expected number of parts in a unit load arriving at the input queue of station i
$E(J_i)$ = expected number of parts in a unit load arriving at the output queue of station i

λ_i	= arrival rate at the output queue of station i (unit loads/time unit)
λ_T	= sum of the arrival rates at the output queues of stations $(=\sum_{i=1}^{M}\lambda_i)$
ND	= number of handling devices in the system
p_{ij}	= fraction of unit loads routed from station i to j
P	= time required for a device to pick-up or deposit a unit load (constant)
σ_{ij}	= expected empty travel time from station i to j
$\sigma_{ij}^{(2)}$	= second moment of the empty travel time from station i to j
τ_{ij}	= expected loaded travel time from station i to j $(\sigma_{ij}+2P)$
$\tau_{ij}^{(2)}$	= second moment of the loaded travel time from station i to j

4.1 Expected Waiting Times in the Input Queues

Parts arrive in bulk at the input queue of each station and the number of parts in each arriving unit load varies depending on the TBS of each part type. Using the $M^{(b)}/G/1$ results given by Ross[12], the expected waiting time of a *part* in the input queue of station i, WI_i^P, can be obtained as follows:

$$WI_i^P=\frac{E(S_i)[E(N_i^{(2)})-E(N_i)]/E(N_i)+\Lambda_i E(N_i)E(S_i^{(2)})}{2[1-\Lambda_i E(N_i)E(S_i)]} \quad (1)$$

In equation (1), Λ_i, is given by

$$\Lambda_i=\sum_{k\in R_i} D_k/Q_k \quad (2)$$

The first and the second moment of the number of parts in a unit load arriving at each input queue are given as follows:

$$E(N_i)=\sum_{k\in R_i}(D_k/Q_k)Q_k\Big/\sum_{l\in R_i}D_l/Q_l=\sum_{k\in R_i}D_k/\Lambda_i \quad (3)$$

$$E(N_i^{(2)})=\sum_{k\in R_i}(D_k/Q_k)Q_k^2\Big/\sum_{l\in R_i}D_l/Q_l=\sum_{k\in R_i}D_kQ_k/\Lambda_i \quad (4)$$

Note that $E(N_i)$ and $E(N_i^{(2)})$ are assumed to be zero for an I/O station, since all the unit loads arriving at the input queue of an I/O station leave the system immediately, By substituting equations (2), (3) and (4) into equation (1), we obtain

$$WI_i^P=\frac{E(S_i)[\sum_{k\in R_i}D_k(Q_k-1)/\sum_{k\in R_i}D_k]+E(S_i^{(2)})\sum_{k\in R_i}D_k}{2[1-\sum_{k\in R_i}D_kE(S_i)]} \quad (5)$$

On the other hand, the expected waiting time of a *unit load* in the input queue of station i is expressed as follows[12]:

$$WI_i^u=WI_i^P-\text{waiting time dueto the parts in the same unit load}$$
$$=WI_i^P-E(S_i)[E(N_i^{(2)})-E(N_i)]\Big/2E(N_i) \quad (6)$$

By substituting equations (2), (3) and (4) into equation (6), and simplifying, we obtain

$$WI_i^u=\frac{E(S_i)^2[\sum_{k\in R_i}D_k(Q_k-1)]+E(S_i^{(2)})\sum_{k\in R}D_k}{2(\sum_{k\in R}D_k)[1-\sum_{k\in R}D_kE(S_i)]} \quad (7)$$

4.2 Expected Waiting Times in the Output Queues

Under the FCFS dispatching rule, the expected empty travel time to station i, E_i, is estimated as follows[3]:

$$E_i=\sum_{h=1}^{M}\sum_{j=1}^{M}p_{hj}\sigma_{ji}\frac{\lambda_h}{\lambda_T}=\sum_{j=1}^{M}\sigma_{ji}, \quad (8)$$

since $\sum_{h=1}^{M}p_{hj}\lambda_h=\Lambda_j$. The second moment of the expected empty travel time to station i, $E_i^{(2)}$, is given by

$$E_i^{(2)}=\sum_{j=1}^{M}\frac{\lambda_j}{\lambda_T}\sigma_{ji}^{(2)} \quad (9)$$

The expected loaded travel time from station i, L_i, and its second moment, $L_i^{(2)}$, are easily obtained as follows:

$$L_i=\sum_{j=1}^{M}p_{ij}\tau_{ij} \quad (10)$$

$$L_i^{(2)}=\sum_{j=1}^{M}p_{ij}\tau_{ij}^{(2)} \quad (11)$$

where $\tau_{ij}=\sigma_{ij}+2P$. Note that we include pickup/deposit time, P, as part of the expected loaded travel time.

Based on equations (8) and (10), the expected service time for a device to serve a move request at station i is given by

$$T_i=E_i+L_i, \quad (12)$$

and the second moment of the service time is given by

$$T_i^{(2)}=E_i^{(2)}+L_i^{(2)}+2E_iL_i \quad (13)$$

The total workload for the handling system is given by

$$WL=\sum_{i=1}^{M}\lambda_iT_i \quad (14)$$

Thus, the expected device utilization is given by

$$\rho=\sum_{i=1}^{M}\frac{\lambda_iT_i}{ND} \quad (15)$$

since the handling system consists of ND devices.

The device utilization consists of two components: the expected fraction of loaded travel and the expected fraction of empty travel. The former is easily obtained by

$$\alpha_f=\frac{1}{ND}\sum_{i=1}^{M}\lambda_i\sum_{j=1}^{M}p_{ij}\tau_{ij} \quad (16)$$

Hence, the expected fraction of empty travel is given by

$$\alpha_e=\rho-\alpha_f \quad (17)$$

We now estimate the expected waiting time of a *unit load* in each output queue. The move requests form a single "conceptual queue" which is served on a FCFS basis. Given Poisson arrivals for the move requests, we can use a M/G/c model to estimate the above expected waiting time. Unlike central server models, we account for empty travel explicitly. That is, the service time parameters are based on the origin and destination of the move requests.

Using the M/G/c model given by Nozaki[11], the expected waiting time for a M/G/c queue is obtained as follows:

$$W_q = \frac{\sum_{i=1}^{M} \lambda_i T_i^{(2)} [\sum_{i=1}^{M} \lambda_i T_i]^{ND-1}}{2(ND-1)![ND - \sum_{i=1}^{M} \lambda_i T_i]^2 K} \tag{18}$$

where $K = \sum_{n=0}^{ND-1} \frac{[\sum_{i=1}^{M} \lambda_i T_i]^n}{n!} + \frac{[\sum_{i=1}^{M} \lambda_i T_i]^{ND}}{(ND-1)!(ND - \sum_{i=1}^{M} \lambda_i T_i)}$

In deriving equation (18), the waiting time in the output queue is defined as the time spent by a move request from the instance of arrival to the instance of receiving service. In a trip-based handling system, while the device travels empty, the move request physically remains in the output queue. Therefore, the actual expected waiting time of a move request in the output queue of station i is given by

$$WO_i^u = W_q + E_i \tag{19}$$

4.3 Expected Time in the System

The expected time in the system of a unit load for part type k, TW_k^u, is given by

$$TW_k^u = \sum_{(iPj) \in \Omega_k} (WI_i^u + PR_{ik}^u + WO_i^u + \tau_{ij}) \ , \tag{20}$$

where PR_{ik}^u is the expected "time at the processor" at station i, which is defined as the time spent by the first part of a unit load from the instance of its removal from the input queue to the instance of its placement in the output queue of station i. Since parts are staged by the server until all the parts in the unit load are processed, PR_{ik}^u is given by $Q_k E(S_i)$.

Note that the expected time a unit load spends at station i is equal to the expected time a part in that unit load spends at station i, since parts are always handled as a unit load. The same holds true for travel times. Hence, the expected time a unit load spends in the system is equal to the expected time a part in that unit load spends in the system.

4.4 Expected WIP in the System

In this section, the expected WIP (in terms of the total number of parts) in the system is estimated by Little's formula. First, the expected total WIP in the input queues, WIP_T^{in}, is given by

$$WIP_T^{in} = \sum_{i=1}^{M} \Lambda_i E(N_i) WI_i^P = \sum_{i=1}^{M} \sum_{k \in R_i} D_k WI_i^P \tag{21}$$

Second, to estimate the expected total WIP in the output queues, we need the expected number of parts in a move request. The expected number of parts in a move request at station i, $E(J_i)$, is obtained as

$$E(J_i) = (\sum_{k \in R_i} D_k) / \lambda_i \tag{22}$$

Therefore, the total expected WIP in the output queues, WIP_T^{out}, is obtained as

$$WIP_T^{out} = \sum_{i=1}^{M} \lambda_i E(J_i) WO_i^P = \sum_{i=1}^{M} \sum_{k \in R_i} D_k WO_i^u \tag{23}$$

Recall that $E(J_i)$ is equal to $E(N_i)$ for the processing stations since flow is conserved. However, at I/O stations, $E(N_i)$ is assumed to be zero, while $E(J_i)$ may be non-zero depending on the part routing.

Finally, the total expected WIP in the system, WIP_T^{sys}, is obtained as

$$WIP_T^{sys} = \sum_{k=1}^{JT} Q_k \frac{D_k}{Q_k} TW_k^u = \sum_{k=1}^{JT} D_k TW_k^u \tag{25}$$

5. EVALUATION OF THE WIP MODEL

To test the WIP model, we used six layouts. In evaluating the WIP model, we observed similar results for the other layouts. Therefore, in this paper we present only the results obtained with layout 1. (See [9] for the reults and data related with other layouts.) Layout 1 from Srinivasan, *et al.*[16] is shown in Figure 1. The distance matrix and additional data are shown in Tables 1 and 2. For evaluation purposes, the processing time at each station is assumed to be exponentially distributed with the same

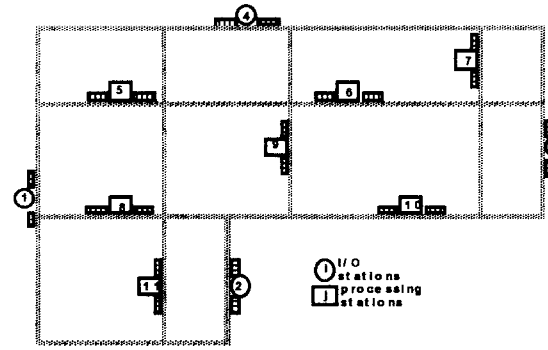

Figure 1: Layout 1

Table 1 : Distance matrix for Layout 1

station number	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	17	14	41	27	18	29	43	9	21	25	0

Table 2: Part routes and throughput requirements #1

part	parts/mir	prrt route 1.L1
1	0.1	1 - 5 - 6 - 7 - 9 - 10 - 3
2	0.2	1 - 8 - 11 - 10 - 6 - 9 - 3
3	0.1	1 - 11 - 5 - 8 - 7 - 4
4	0.1	2 - 10 - 9 - 7 - 6 - 4
5	0.1	2 - 8 - 5 - 10 - 7 - 6 - 3
6	0.2	2 - 11 - 8 - 9 - 5 - 4

Device speed:from 200(with 1 device) to 50(with 4 devices)
Pickup/deposit time is negligible.

mean regardless of part type. The mean processing time is selected such that the utilization of each processing station is equal to 0.70. Also, the travel times between the stations are assumed to be exponentially distributed.

We compare the results obtained by the WIP model with those obtained from simulation. The simulation model simulates the "actual system," i.e., we do not force Poisson arrivals except for the arrivals from outside the system. Simulation results are obtained from 10 replications, where at least 1,000 unit loads of each part are processed through the system per replication. For simplicity, we varied only the TBS of part type 6 in Table 2. Also, as we increase the number of devices, we proportionally reduce the device speed and increase the pick-up/deposit times to maintain a comparable device utilization.

5.1 Expected Waiting Times in the Output Queues

The expected output queue waiting times obtained with 1, 2, and 4 devices are shown in Table A1 in Appendix. In general, the WIP model estimates the expected output queue waiting times reasonably well, regardless of the TBS or the number of devices. The maximum relative error is less than 15%, and in most cases it is less than 10. Also, the absolute errors are fairly small. For a fixed throughput and fixed number of devices, the expected waiting time in each output queue decreases as the TBS increases since the number of unit loads that must be moved by the devices per time unit decreases when the TBS increases.

5.2 Expected Device Utilization

The overall performance of the handling system depends on the device utilization. In Table4, we present the expected device utilizations obtained from the WIP model and by simulation for TBSs of 1, 5, and 10 for part type 6. The results obtained from the WIP model do not vary with the number of devices since we adjust the device speed and the pickup/deposit times as the number of devices varies. The results in Table 4 indicate that the WIP model slightly overestimates the expected device utilization. However, the error is less than 2% in all cases.

Table 3 : Device utilization

TBS of part 6	1 device WIP	1 device sim 95% C. I.	2 devices WIP	2 devices sim 95% C. I.	4 devices WIP	4 devices sim 95% C. I.
1	.8536	.8507 (.0051)	.8536	.8494 (.0050)	.8536	.8494 (.0053)
5	.7150	.7130 (.0037)	.7150	.7112 (.0038)	.7150	.7101 (.0046)
10	.6972	.6901 (.0036)	.6972	.6891 (.0032)	.6972	.6885 (.0038)

5.3 Total Expected WIP in the System

The original data for layout 1 (see Table 2) was presented by Srinivasan et al. [16]. In the remainder of the paper, however, for layout 1 we use the alternate data shown in Table 4, which is designed to emphasize the throughput differences between the part types and also to control the stations visited by each part type. Using six devices and increasing the TBS from 1 through 10 for each part type, one at a time, we obtained the expected WIP results shown in Figure A1 in Appendix. (The results for part types4 and5 are not shown since they are almost identical to those obtained for part type1.)

Table 4: Part routes and throughput requirements #1

part	parts/mir	prrt route 2.L1
1	0.01	1 - 5 - 6 - 9 - 11 - 10 - 3
2	0.03	1 - 7 - 5 - 8 - 10 - 11 - 3
3	0.08	1 - 6 - 7 - 10 - 9 - 4
4	0.01	2 - 5 - 7 - 8 - 11 - 6 - 3
5	0.01	2 - 7 - 9 - 8 - 6 - 11 - 4
6	0.02	2 - 5 - 8 - 9 - 10 - 4

Device speed:from 100(with 1 device) to 20(with 4 devices)

Pickup/deposit time is negligible.

Depending on the TBS, the WIP model underestimates the expected output queue WIP, and it overestimates the expected input queue WIP. However, the above errors are not significant, and as shown in Figure A1, the WIP model tracks the expected total WIP curve (obtained from simulation) rather closely over a wide range of TBSs and device utilizations. Hence, for optimization purposes, the WIP model appears to be satisfactory. In the next section, we will formally present the objective function and evaluate the robustness of the TBSs obtained via the WIP model.

6. DETERMINING THE "OPTIMAL" TBS

In this section, within the context of transfer batch sizing, we present alternative formulations to minimize the WIP and/or material handling related costs in a manufacturing system. We also present the "optimal" TBSs obtained via exhaustive enumeration and simulation. Lastly, we present a heuristic optimization scheme based on a genetic algorithm (GA).

6.1 Alternative Formulations

Assuming a fixed number of devices, a simple formulation of the problem may be presented as follows:

(P1) $MIN \quad C_W \; WIP_T^{sys}$

$s.t. \quad \rho < 1$

$Q_k \leq Ub_k \qquad k=1, \ldots, JT$

Q_k : positive integer.

where C_W is the WIP cost per part per time unit and Ub_k is the upper bound on the TBS of part type k. Of course, we must also ensure that the expected device utilization is less than 1.0.

In the above formulation (P1), we use a single estimate for the WIP carrying cost. However, in most manufacturing systems, since more "value" is added as the parts are processed, the WIP carrying cost at a particular station may depend on the part type. The WIP model we present can accommodate such a case since the expected input queue waiting time at a station is independent of the part type (due to FCFS service) and the arrival rate (in parts) of a part type at a station is given. Hence, the WIP model can estimate the expected WIP associated with each station for each part type.

In P1, the number of devices is assumed to be fixed. However, the number of devices (i.e., the fleet size) is an important design variable since it is likely to affect the overall cost. The transfer batch sizing problem with a variable fleet size can be formulated as follows:

$$
\begin{aligned}
\text{(P2)} \quad & MIN \quad C_W \; WIP_T^{sys} + C_D N_D \\
& s.t. \quad \rho < 1 \\
& \qquad Q_k \leq Ub_k \qquad k=1, \ldots, JT \\
& \qquad C_D N_D \leq B \\
& \qquad Q_k : \text{positive integer.}
\end{aligned}
$$

where B is the budget for the material handling system, and C_D is the equivalent cost per time unit per device.

Using the WIP model we present, the cost associated with each unit space in the input or output queues may also be included in the objective function (as in Grasso and Tanchoco[8]) provided that one is designing with *average* queue lengths in mind. One can also add cost elements similar to those shown by Egbelu[6]. For simplicity, we will include only the WIP carrying cost and the material handling cost.

6.2 Computational Results

To demonstrate the type of solutions obtained from the WIP model, we set $C_W = 1$ and $Ub_k = 10$ for all k in P1. We used exhaustive enumeration to find solutions by the WIP model.

Given that our main concern is to determine "good" TBSs, we compared the quality of the solutions determined by the WIP model with those obtained via simulation. Table 5 summarizes five best solutions obtained by the WIP model and the corresponding objective function values obtained by simulating those TBSs. Also, five best solutions and objective function values determined by simulation alone (within the ±2 range of the best solution obtained by the WIP model) are presented. The "optimal" solution obtained from the WIP model is not always the best solution obtained via simulation. However, taking into account that the error in the objective function value (see column9 in Table 5) is quite small (less than 6% in all test problems) and that simulation results contain random variation, we conclude that TBSs obtained from the WIP model are reasonably good. Although we were unable to prove convexity, the overall structure of the total expected WIP (shown earlier in Figure A1) strongly suggests that one is highly unlikely to find a much better solution outside the ±2 range that we investigated.

Table 5 :Solutions obtained by the WIP model and simulation

	Best 5 solutions							
	by the WIP				by simulation			
Layout	Rank	T.B.S.	Obj. value	sim. results	T.B.S.	Obj. value	95% C.I.	error
1	1	2 3 3 2 2 2	44.01	39.53	2 3 3 2 2 2	39.53	2.1	0.00
	2	2 3 3 2 3 2	44.16	39.84	2 3 3 3 2 2	39.53	1.9	0.0
	3	2 3 3 3 2 2	44.22	39.53	2 3 3 2 2 3	39.80	1.9	0.0
	4	3 3 3 2 2 2	44.27	40.24	2 3 3 2 3 2	39.84	2.40	0.0
	5	2 3 3 2 2 3	44.41	39.80	2 4 3 2 2 2	39.97	1.3	0.00
2	1	3 2 2 3 2	99.35	87.49	2 3 2 3	86.73	4.20	0.0
	2	2 3 2 3 2	100.11	86.73	4 2 2 3	87.12	2.74	0.00
	3	2 2 3 3 2	100.41	87.58	3 2 2 3	87.49	3.7	0.00
	4	4 2 2 3 2	100.63	87.12	3 2 3 3	87.56	3.06	0.0
	5	3 2 3 3 2	100.73	87.56	2 2 3 3	87.58	3.57	0.00
3	1	4 4 4 4 3	62.50	56.37	3 3 4 3	53.60	4.35	0.05
	2	3 4 4 4 3	62.55	56.09	3 4 3 3	53.76	5.69	0.04
	3	3 4 3 3 4	62.65	53.76	2 4 3 3	54.20	5.62	0.0
	4	4 4 3 4 3	62.69	57.04	4 4 3 3	54.45	5.77	0.05
	5	3 3 4 4 4	62.75	55.18	2 5 3 3	54.52	6.1	0.0

6.3 Genetic Algorithm

To avoid exhaustive enumeration and search the solution space more systematically, we developed a heuristic based on a genetic algorithm (GA), which seems suitable for our problem structure. Of course, a GA-based heuristic is not guaranteed to find the global optimal solution and the quality of the solution depends on the input parameters.

After testing various problems, we constructed a simple GA with "elitist reproduction" and "biased mutation." We stop the algorithm either when it reaches the maximum number of generations (1,000 in our test cases) or the 15th current best solution is not improved for 50 consecutive generations. For all the test problems, using the above settings, the GA we developed obtained the same "optimal" solution we obtained earlier via exhaustive enumeration (see Table 6).

Table 6 : "Optimal" solutions obtained by GA

Lay-out	No. of devices	Enumeration		Genetic Algorithm	
		TBS	Total WIP	TBS	Total WIP
1	1	7,8,8,7,7,7	110.79	7,8,8,7,7,7	110.79
	3	2,3,3,2,2,2	44.01	2,3,3,2,2,2	44.01
	5	1,2,2,1,1,1	31.54	1,2,2,1,1,1	31.54
2	1	8,7,7,8,6	252.93	8,7,7,8,6	252.93
	3	3,2,2,3,2	99.35	3,2,2,3,2	99.35
	5	2,1,2,2,1	70.89	2,1,2,2,1	70.89
3	1	11,11,11,11,10	160.88	11,11,11,11,10	160.88
	3	4,4,4,4,3	62.5	4,4,4,4,3	62.5
	5	2,2,3,2,2	42.91	2,2,3,2,2	42.91

Population size=50,Ellitist reproduction=4,
Crossover prob.=0.7, Mutation prob.=0.01

7 CONCLUSIONS

The WIP model we present here is significant because it establishes a formal, analytical relationship between the expected WIP level in a manufacturing system and the capacity of the material handling system that supports it. Traditional thinking (or "conventional wisdom") dictates that, being a "non-value added" operation, investment in material handling should be minimized. Such thinking led to research and analytical/simulation models where the objective is to determine the minimum fleet size to meet a given throughput requirement. In fact, it is fair to say that even material handling equipment vendors follow the same line of thinking and use simulation to design systems with minimum required number of devices.

Our results based on the WIP model clearly suggest that such thinking is flawed. For example, in Table 6, we show that even with a single device one can meet the required throughput (provided, of course, that the TBSs are sufficiently large). If two devices are added to the system, a significant reduction in total expected WIP is obtained (since the TBSs are reduced). Given all the known manufacturing problems and costs associated with excessive WIP, it is very likely that the additional investment required to purchase two devices would be well-justified. Of course, further increases in the fleet size may be unnecessary; such diminishing returns would be indicated by the WIP model and the GA-based algorithm we present here.

The WIP model and results we present here are also significant for facility layout. When an existing layout is improved (through department relocations), it typically reduces the workload on the handling system. However, one is often not in a position to reduce the handling work force or to readily dispose of handling equipment. We show that savings can be still realized by reducing the TBS for all the jobs while maintaining the same handling work force; i.e., savings can be derived from reduced WIP levels rather than a reduced handling work force. This option may prove to be critical in justifying layout improvements with a "fixed" handling capacity/work force.

References

[1] Bertrand, J. W. M., ``Multiproduct Optimal Batch Size with In-Process Inventories and Multi Work Centers," *IIE Transactions*, Vol.17, No.2, pp.157—163, 1985.

[2] Bozer, Y. A., M. S. Cho and M. M. Srinivasan, ``Expected Waiting Times in Single-Device Trip-Based Material Handling Systems," to appear in the *European Journal of Operational Research.*

[3] Cho, M. S., ``Design and Performance Analysis of Trip-Based Material Handling Systems in Manufacturing," Ph.D. Dissertation, Department of Industrial and Operations Engineering, University of Michigan, Ann Arbor, 1990.

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

[5] Chow, W. M., ``An Analysis of Automated Storage and Retrieval Systems in Manufacturing Assembly Lines," *IIE Transactions*, Vol.18, No.2, pp.204—214, 1986.

[6] Egbelu, P. J., ``Economic Design of Unit Load-Based FMSs Employing AGVs for Transport," *Intl. Journal of Production Research*, Vol.31, No.12, pp.2753—2775, 1993.

[7] Egbelu, P. J., ``Concurrent Specification of Unit Load Sizes and Automated Guided Vehicle Fleet Size in Manufacturing System," *Intl. Journal of Production Economics*, Vol.29, pp.49—64, 1993.

[8] Grasso, E. T. and J. M. A. Tanchoco, ``Unit Load and Material Handling Considerations in Material Requirements Systems," *Material Flow*, Vol.1, pp.79—87, 1983.

[9] Kim, J. ``Determining Transfer Batch Sizing in Trip-Based Material Handling Systems," Ph.D. Dissertation, Department of Industrial and Operations Engineering, University of Michigan, Ann Arbor, 1994.

[10] Kuehn, P. J., ``Approximate Analysis of general Queueing Networks by Decomposition," *IEEE Transactions on Communications*, Vol.COM-27, No.1, pp.113—126, 1979.

[11] Nozaki, S. and S. Ross, ``Approximations in Finite Capacity Multiserver Queues with Poisson Arrivals," *Journal of Applied Probability*, 15, pp.826—834, 1978.

[12] Ross, S. M., *Introduction to Probability Theory*, Academic Press Inc., FL, 1985.

[13] Solberg, J. J., ``The Optimal Planning of Computerized Manufacturing Systems: CAN-Q User's Guide," School of Industrial Engineering, Purdue University, West Lafayette, IN, 1980.

[14] Solberg, J. J., ``Capacity Planning with a Stochastic Workflow Model," *AIIE Transactions*, Vol.13, No.2, pp.116—122, 1981.

[15] Solot, P. and J. M. Bastos, ``MULTIQ: A Queueing Model for FMSs with Several Paallet Types," *Journal of the Operational Research Society*, Vol.39, No.9, pp.811—821, 1988.

[16] Srinivasan, M. M., Y. A. Bozer and M. S. Cho, ``Trip-Based Material Handling Systems: Throughput Capacity Analysis," *IIE Transactions*, Vol.26 No.1, pp.70—89, 1994.

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

Appendix

Table A1 : Expected waiting times in the output queues, Layout 1

	TBS	ST. NO	1	2	3	4	5	6	7	8	9	10	11
1 device	1	WIP	1.106	1.086	0.000	0.000	1.072	1.057	1.091	1.076	1.048	1.072	1.084
		sim	1.121	1.105	0.000	0.000	1.092	1.092	1.135	1.095	1.066	1.106	1.107
		95%C.I.	0.050	0.053	0.000	0.000	0.049	0.048	0.047	0.052	0.053	0.049	0.051
	5	WIP	0.572	0.547	0.000	0.000	0.537	0.513	0.540	0.541	0.507	0.524	0.550
		sim	0.613	0.588	0.000	0.000	0.555	0.562	0.594	0.546	0.523	0.578	0.553
		95%C.I.	0.012	0.016	0.000	0.000	0.010	0.011	0.012	0.011	0.010	0.012	0.010
	10	WIP	0.539	0.514	0.000	0.000	0.504	0.478	0.505	0.508	0.474	0.490	0.518
		sim	0.590	0.569	0.000	0.000	0.524	0.540	0.579	0.523	0.505	0.563	0.518
		95%C.I.	0.009	0.016	0.000	0.000	0.010	0.008	0.012	0.009	0.007	0.010	0.009
2 devices	1	WIP	1.162	1.122	0.000	0.000	1.095	1.066	1.133	1.102	1.047	1.095	1.119
		sim	1.174	1.144	0.000	0.000	1.114	1.093	1.171	1.121	1.062	1.128	1.139
		95%C.I.	0.043	0.047	0.000	0.000	0.044	0.041	0.043	0.048	0..048	0.042	0.049
	5	WIP	0.642	0.593	0.000	0.000	0.573	0.524	0.578	0.580	0.513	0.547	0.599
		sim	0.672	0.624	0.000	0.000	0.577	0.560	0.617	0.576	0.516	0.586	0.590
		95%C.I.	0.010	0.014	0.000	0.000	0.011	0.010	0.011	0.007	0.012	0.010	0.010
	10	WIP	0.611	0.561	0.000	0.000	0.542	0.490	0.542	0.550	0.480	0.513	0.568
		sim	0.671	0.622	0.000	0.000	0.564	0.552	0.620	0.565	0.513	0.588	0.570
		95%C.I.	0.007	0.013	0.000	0.000	0.007	0.009	0.013	0.008	0.010	0.010	0.007
4 devices	1	WIP	1.326	1.246	0.000	0.000	1.192	1.133	1.268	1.206	1.096	1.191	1.240
		sim	1.343	1.270	0.000	0.000	1.209	1.162	1.299	1.223	1.111	1.223	1.258
		95%C.I.	0.044	0.046	0.000	0.000	0.041	0.039	0.042	0.048	0.045	0.039	0.041
	5	WIP	0.838	0.741	0.000	0.000	0.700	0.602	0.711	0.716	0.580	0.649	0.752
		sim	0.870	0.780	0.000	0.000	0.689	0.631	0.744	0.700	0.575	0.684	0.729
		95%C.I.	0.010	0.011	0.000	0.000	0.008	0.011	0.010	0.008	0.006	0.010	0.008
	10	WIP	0.812	0.712	0.000	0.000	0.674	0.570	0.675	0.690	0.550	0.615	0.726
		sim	0.863	0.778	0.000	0.000	0.666	0.621	0.729	0.675	0.570	0.675	0.694
		95%C.I.	0.007	0.018	0.000	0.000	0.007	0.007	0.008	0.007	0.008	0.006	0.006

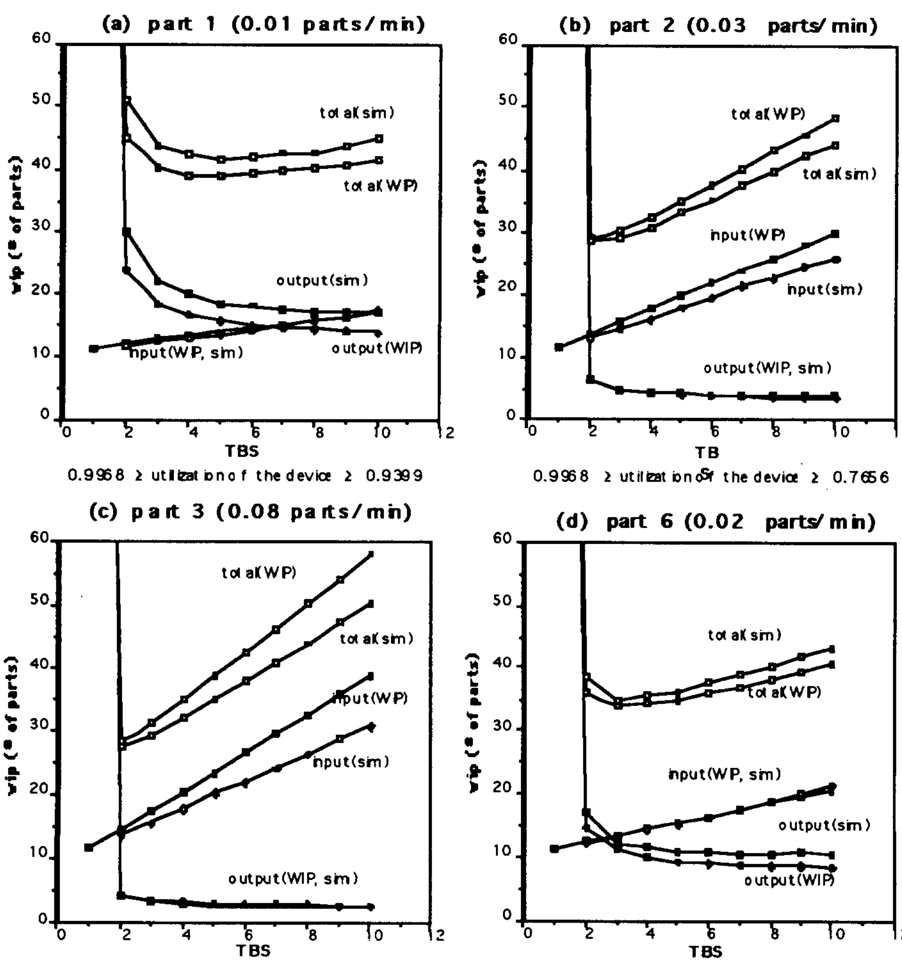

Figure A1: Expected WIP in the system, Layout 1