MINIMIZING WORK-IN-PROCESS IN DESIGN OF FACILITY LAYOUTS

Saifallah Benjaafar  
Division of Industrial Engineering  
Department of Mechanical Engineering  
University of Minnesota  
Minneapolis, MN 55455

Revised November, 2000

Abstract

In this paper, we present a formulation of the facility layout design problem where the objective is to minimize work-in-process (WIP). In contrast to some recent research, we show that layouts obtained using a WIP-based formulation can be very different from those obtained using the conventional quadratic assignment problem (QAP) formulation. For example, we show that a QAP-optimal layout can be WIP-infeasible. Similarly, we show that two QAP-optimal layouts can have vastly different WIP values. In general, we show that WIP is not monotonic in material handling travel distances. This leads to a number of surprising results. For instance, we show that it is possible to reduce overall distances between departments but increase WIP. Because WIP is affected by both mean and variance of travel times, we find that reducing variance can be as important as reducing average travel time. Furthermore, we find that the relative desirability of a layout can be affected by changes in material handling capacity even when travel distances remain the same. More importantly, we show that the relative desirability of a layout can be affected by non-material handling factors, such as department utilization levels, variability in department processing times, and variability in product demands. Finally, we study the effect of system parameters, such as flow symmetry, layout geometry, and material handling capacity on the difference in WIP between QAP-optimal and WIP-optimal layouts. We find that although there are conditions under which the difference in WIP is significant, there are those under which both layouts are WIP-equivalent.
1. Introduction

In two recent papers [5, 6], Fu and Kaku presented a plant layout problem formulation for job shop-like manufacturing systems where the objective is to minimize average work-in-process. In particular, they investigated conditions under which the familiar quadratic assignment problem (QAP) formulation, where the objective is to minimize average material handling costs, also minimizes average work-in-process. By modeling the plant as an open queueing network, they showed that under a set of assumptions, the problem reduces to the quadratic assignment problem. Using a simulation of an example system, they found that the result apparently holds under much more general conditions than are assumed in the analytical model.

To obtain a closed a form expression of expected WIP, Fu and Kaku [6] made the following assumptions: (1) external part type arrival processes into the system are Poisson; (2) processing times at a department are i.i.d. exponential; (3) material handling is carried out via discrete material handling devices, such as forklifts or automated guided vehicles (AGV), (4) travel times of the material handling devices are exponentially distributed, (5) input and output buffer sizes at departments are sufficiently large so that blocking is negligible, and (6) service discipline is first-come, first-served (FCFS). In modeling travel times, they ignored empty travel by the material handling devices and accounted only for full trips. These assumptions allowed them to treat the network as a Jackson queueing network - i.e., a network of independent M/M/1 and M/M/n queues - for which a closed form analytical expression of average WIP is available. They showed that WIP accumulation at the processing departments - at both the input and output buffers - is always independent of the layout and that travel times are a linear transformation of the average distance traveled by the material handling system when full. Since the measure of material handling cost used in the QAP formulation is itself a linear function of the same average travel distance, they showed that the queueing and QAP formulations are equivalent.

In this paper, we show that when some of the assumptions used by Fu and Kaku are relaxed, their key observation regarding the equivalence of the two formulations is not always valid. In fact, under general conditions, we show that layouts generated using the queueing-based model can be very different from those obtained using the conventional QAP formulation. More importantly, we show that the choice of layout does have a direct impact on WIP accumulation at both the material handling system and at the individual departments, and that the behavior of expected WIP is not necessarily monotonic in the average distance traveled by the material handling device. This leads to a number of surprising and counter-intuitive results. In particular,
we show that reducing overall distances between departments can increase WIP. We also show that the desirability of a layout can be affected by non-material handling factors, such as department utilization levels, variability in processing times at departments and variability in product demands. In general, we find the objective function used in the QAP formulation to be a poor indicator of WIP. For example, we show that a QAP-optimal layout can be WIP-infeasible - i.e., it results in infinite WIP. Similarly, we show that two QAP-optimal layouts can have vastly different WIP values. Furthermore, we find that the QAP formulation, by accounting only for full travel by the material handling system, ignores the important role that empty travel plays. For example, we find that minimizing full travel, as the QAP formulation does, can cause empty travel to increase which, in turn, can increase WIP. This leads to some additional counterintuitive results. For instance, we find that it can be highly desirable to place departments in neighboring locations even though there is no direct material flow between them. Likewise, we show that it can be beneficial to place departments with high inter material flows in distant locations from each other. On the other hand, we also show that there are conditions on flow distribution, layout geometry and material handling capacity under which both a QAP-optimal and a WIP-optimal layout are WIP-equivalent.

In our model, we relax several of the assumptions used by Fu and Kaku. In particular, we let part inter-arrival times and processing times be generally distributed and determined by the number of product types and their routings. We also allow the distances traveled by the material handling devices to be determined by the layout configuration. This allows us to characterize exactly the distribution of travel times and to capture both empty and full travel by the material handling system. We show that relaxing these assumptions enable us to capture important interactions between the layout configuration, the distribution of travel times, and several operating characteristic of the processing departments. These interactions are absent in the Fu and Kaku model which, in part, explains the results we obtain. Our results are applicable to systems where a shared material handling system consisting of discrete devices is used. This excludes systems with continuous conveyors and systems with dedicated material handling for each segment of the flow.

Although there is an extensive literature on design of facility layouts (see Meller and Gau [15] for a recent review), the design criterion in the majority of this literature is material handling cost, measured either directly as a function of material handling distances or indirectly through an adjacency score [20]. Few papers consider operational performance measures, such as WIP,
throughput, or cycle time, as a design criterion or a design constraint. Among those that do, we note the previously mentioned papers by Fu and Kaku [5, 6] and papers by Kouvelis and Kiran [13, 14]. Kouvelis and Kiran introduce a modified formulation of the quadratic assignment problem where the objective is to minimize the sum of material handling and WIP holding costs subject to a constraint on throughput. In modeling travel times, they however ignore empty travel and consider only the mean of the full travel time distribution. A similar approach is also used by Solberg and Nof [18] in evaluating different layout configurations. Outside the layout literature, there is a related body of research on design of material handling systems. Although some of this literature addresses the modeling of empty travel, especially as it pertains to design of automated guided vehicles, it generally assumes a fixed layout (see Johnson and Brandeau [10] for a recent review).

We should note here that using operational performance as a criterion in the design of manufacturing facilities is not new. In fact, reducing WIP and cycle time while increasing throughput has driven much of the process improvement efforts of the last two decades [9]. Surprisingly, layout design has continued to be carried out using a mostly traditional measure of cost. In this paper, we show that layouts could also be designed to enhance operational performance and support a firm’s strategic objectives of smaller WIP and shorter cycle time. More importantly, we offer guidelines as to when using an operational performance criterion, such as WIP, is particularly useful.

2. Model Formulation

We use the following assumptions and notation.
i) The plant produces $N$ products. Product demands are independently distributed random variables. Unit orders arrive according to a renewal process with rate $D_i$ (average demand per unit time) and a squared coefficient of variation $C_i^2$ for $i = 1, 2, \ldots, N$. The squared coefficient of variation denotes the ratio of the variance over the squared mean of unit order inter-arrival times.

ii) Material handling is carried out by a single discrete material handling device, or a transporter (an extension to the multi-device case is discussed in appendix 1). In responding to a request, a material handling device travels empty from the department location of its last delivery to the department location of the current request. Material transfer requests are serviced on a first
come-first served (FCFS) basis. In the absence of any requests, the material handling device remains at the location of its last delivery.

iii) The travel time between any pair of locations \( k \) and \( l \), \( t_{kl} \), is assumed to be deterministic and is given by \( t_{kl} = d_{kl}/v \), where \( d_{kl} \) is the distance between locations \( k \) and \( l \) and \( v \) is the speed of the material handling transporter.

iv) Products are released to the plant from a loading department and exit the plant through an unloading (or shipping) department. Departments are indexed from \( i = 0 \) to \( M + 1 \), with the indices \( i = 0 \) and \( M + 1 \) denoting, respectively, the loading and unloading departments.

v) The plant consists of \( M \) processing departments, with each department consisting of a single server (e.g., a machine) with ample storage for work-in-process. Jobs in the queue are processed in first come-first served order. The amount of material flow, \( \lambda_{ij} \), between a pair of departments \( i \) and \( j \) is determined from the product routing sequence and the product demand information. The total amount of workload at each department is given by:

\[
\lambda_i = \sum_{k=0}^{M} \lambda_{ki} = \sum_{j=1}^{M+1} \lambda_{ij} \quad \text{for} \ i = 1, 2, \ldots, M,
\]

\[
\lambda_0 = \lambda_{M+1} = \sum_{i=1}^{N} D_i, \quad \text{and}
\]

\[
\lambda_{t} = \sum_{i=0}^{M} \sum_{j=1}^{M+1} \lambda_{ij},
\]

where \( \lambda_{t} \) is the workload for the material handling system.

vi) Processing times at each department are independent and identically distributed with an expected processing time \( E(S_i) \) and a squared coefficient of variation \( C_{s_i}^2 \) for \( i = 0, 1, \ldots, M + 1 \) (the processing time distribution is determined from the processing times of the individual products).

vii) There are \( K \) locations to which departments can be assigned. A layout configuration corresponds to a unique assignment of departments to locations. We use the vector notation \( \mathbf{x} = \{x_{ik}\} \), where \( x_{ik} = 1 \) if department \( i \) is assigned to location \( k \) and \( x_{ik} = 0 \) otherwise, to differentiate between different layout configurations. The number of locations is assumed to be greater than or equal to the number of departments.

We model the plant as an open network of GI/G/1 queues, with the material handling system being a central server queue. Note that because parts are delivered to the departments by the
material handling system, the operating characteristics of the material handling system, such as utilization and travel time distribution, directly affect the inter-arrival time distribution of parts to the departments. Similarly, since the queue for the material handling system consists of the department output buffers, the inter-arrival time distribution to this queue is determined by the departure process from the departments, which is, in turn, determined by the operating characteristics of the departments. Therefore, there is a close coupling between the inputs and outputs of the processing departments and the material handling system. In our model, we explicitly capture this coupling and show that there exists a three way interaction between the department operating characteristics, the operating characteristics of the material handling system, and the layout configuration, and that this interaction has a direct effect on WIP accumulation.

In order to show this effect, let us first characterize the travel time distribution. In responding to a material transfer request, the material handling device performs an empty trip from its current location (the location of its last delivery), at some department $r$, followed by a full trip from the origin of the current request, say department $i$, to the destination of the transfer request at a specified department $j$ (see figure 1). The probability distribution $p_{rij}$ of an empty trip from $r$ to $i$ followed by a full trip from $i$ to $j$ is, therefore, given by:

$$p_{rij} = \sum_{k=0}^{M} p_{kr}p_{ij},$$

(4)

where $p_{ij}$ is the probability of a full trip from department $i$ to department $j$ which can be obtained as

$$p_{ij} = \frac{\lambda_{ij}}{\sum_{k=0}^{M} \sum_{j=1}^{M+1} \lambda_{ij}}.$$  

(5)

Given a layout configuration $x$, the time to perform an empty trip from department $r$ to department $i$ followed by a full trip to department $j$ is given by $t_{rij}(x) = t_{ri}(x) + t_{ij}(x)$, where

$$t_{ij}(x) = \sum_{k=1}^{K} \sum_{l=1}^{K} x_{ik} x_{jl}d_{kl}/v$$

(6)

and is the travel time from department $i$ to department $j$. From (4)-(6), we can obtain the mean and variance of travel time as follows:
\[
E(S_t) = \sum_{r=0}^{M+1} \sum_{i=0}^{M} \sum_{j=1}^{M+1} p_{rij} t_{rij}(x) = \sum_{r=1}^{M+1} \sum_{i=0}^{M} \sum_{j=1}^{M+1} (\lambda_{kr} \lambda_{ij} / \lambda_t^2) t_{rij}(x), \tag{7}
\]

and

\[
\text{Var}(S_t) = E(S_t^2) - E(S_t)^2, \tag{8}
\]

where,

\[
E(S_t^2) = \sum_{r=1}^{M+1} \sum_{i=0}^{M} \sum_{j=1}^{M+1} p_{rij} (t_{rij}(x))^2 = \sum_{r=1}^{M+1} \sum_{i=0}^{M} \sum_{j=1}^{M+1} (\lambda_{kr} \lambda_{ij} / \lambda_t^2)(t_{rij}(x))^2, \tag{9}
\]

and

\[
t_{rij}(x) = \sum_{k=1}^{M+1} \sum_{l=1}^{M} \sum_{s=1}^{M+1} x_{rk} x_{il} x_{js} (d_{kl} + d_{ks}) / v = \sum_{k=1}^{M+1} \sum_{l=1}^{M} x_{rk} x_{il} d_{kl} / v + \sum_{l=1}^{M+1} x_{il} x_{ls} d_{ls} / v. \tag{10}
\]

We can also obtain average utilization of the material handling system, \(\rho_t\), as follows:

\[
\rho_t = \lambda_t E(S_t) = \sum_{r=1}^{M+1} \sum_{i=0}^{M} \sum_{j=1}^{M+1} (\lambda_{kr} \lambda_{ij} / \lambda_t) (t_{ri}(x) + t_{ij}(x)), \tag{11}
\]

which can be simplified as:

\[
\rho_t = \sum_{r=1}^{M+1} \sum_{i=0}^{M} (\lambda_r \lambda_i / \lambda_t) t_{ri}(x) + \sum_{i=0}^{M} \sum_{j=1}^{M+1} \lambda_{ij} t_{ij}(x) \tag{12}
\]

or equivalently,

\[
\rho_t = \rho_t^e + \rho_t^f \tag{13}
\]

where

\[
\rho_t^e = \sum_{r=1}^{M+1} \sum_{i=0}^{M} (\lambda_r \lambda_i / \lambda_t) t_{ri}(x) \tag{14}
\]

corresponds to the utilization of the material handling system due to empty travel, and

\[
\rho_t^f = \sum_{i=0}^{M} \sum_{j=1}^{M+1} \lambda_{ij} t_{ij}(x) \tag{15}
\]

is the utilization of the material handling system due to full travel.

From the above expressions, we can see that the travel time distribution is determined by the layout configuration and that this distribution is not necessarily exponential. As a result, the arrival process to the departments is not always Poisson distributed, even if external arrivals are Poisson and processing times are exponential. This means that our system cannot be treated, in
Figure 1 - Empty and full travel in a system with discrete material handling devices
general, as a network of M/M/1 queues. Unfortunately, exact analytical expressions of expected WIP are difficult to obtain for queues with general inter-arrival and processing time distributions. Therefore, in order to estimate average work-in-process, we resort to network decomposition and approximation techniques, where each department, as well as the material handling system, is treated as being stochastically independent, with the arrival process to and the departure process from each department and the material handling system being approximated by renewal processes. Furthermore, we assume that two parameters, mean and variance, of the job inter-arrival and processing time distributions are sufficient to obtain average WIP at each department. The decomposition and approximation approach has been widely used to analyze queueing networks in a variety of contexts [1, 2, 3, 22]. A number of good approximations have been proposed by several authors (for example, see [1] for a recent review). In this paper, the approximations we use have been first proposed by Kraemer and Langenbach-Belz [11] and later refined by Whitt [22, 23] and shown to perform well over a wide range of parameters [3, 23]. The approximations coincide with the exact analytical results obtained by Fu and Kaku for the special case of Poisson arrival and exponential processing/travel times. Since in layout design our objective is to primarily obtain a ranked ordering of different layout alternatives, approximations are sufficient, as long as they guarantee accuracy in the ordering of these alternatives. Approximations are also adequate when we are primarily interested, as we are in this paper, in the qualitative behavior of WIP. Comparisons of our analytical results with results obtained using simulation are discussed in section 3.

Under a given layout, expected WIP at each department $i$ ($i = 0, 1, ..., M+1$), is approximated as follows:

$$ E(WIP_i) = \frac{\rho_i^2(C_{a_i}^2 + C_{s_i}^2)g_i}{2(1 - \rho_i)} + \rho_i, $$

where $\rho_i = \lambda_i E(S_i)$ is the average utilization of department $i$, $C_{a_i}^2$ and $C_{s_i}^2$ are, respectively, the squared coefficients of variation of job inter-arrival and processing times, and

$$ g_i \equiv g_i(C_{a_i}^2, C_{s_i}^2, \rho_i) = \begin{cases} 
\exp\left[-\frac{2(1 - \rho_i)(1 - C_{a_i}^2)^2}{3\rho_i(C_{a_i}^2 + C_{s_i}^2)}\right] & \text{if } C_{a_i}^2 < 1 \\
1 & \text{if } C_{a_i}^2 \geq 1.
\end{cases} $$

Similarly, expected WIP at the material handling system is approximated by:

$$ E(WIP_t) = \frac{\rho_t^2(C_{a_t}^2 + C_{s_t}^2)g_t}{2(1 - \rho_t)} + \rho_t. $$
Note that $\rho_t$ and $\rho_i$ must be less than one for expected work-in-process to be finite. The squared coefficients of variation can be approximated as follows [3, 10]:

$$C_{a_i}^2 = \sum_{j \neq i} \frac{\lambda_j p_{ji}}{\lambda_i} (p_{ji} C_{d_i}^2 + (1 - p_{ji})) + \frac{\lambda_0}{\lambda_i} (\gamma_i C_{a_0}^2 + (1 - \gamma_i)), \text{ and}$$

$$C_{a_i}^2 = \rho_i^2 C_{a_i}^2 + (1 - \rho_i^2) C_{a_i}^2,$$  \hspace{1cm} (20)

where $C_{d_i}^2$ is the squared coefficient of inter-departure time from department $i$, $p_{ij}$ is the routing probability from node $i$ to node $j$ (nodes include departments and the material handling device), $\gamma_i$ is the fraction of external arrivals that enter the network through node $i$, and $1/\lambda_0$ and $C_{a_0}^2$ are, respectively, the mean and squared coefficient of variation of the external job inter-arrival times.

In our case, $\gamma_0 = 1$ and $\gamma_i = 0$ for all others since all jobs enter the cell at the loading department. The routing probability from departments $i = 0$ through $M$ to the material handling system is always one, that from the material handling system to departments $j = 1$ through $M + 1$ is

$$p_{ij} = \frac{\sum_{i=0}^{M+1} \lambda_{ij}}{\sum_{i=0}^{M} \sum_{j=0}^{M+1} \lambda_{ij}},$$ \hspace{1cm} (21)

and to the loading department ($j = 0$) is zero. Parts exit the cell from department $M + 1$ (unloading department) so that all the routing probabilities from that department are zero. Substituting these probabilities in the above expression, we obtain:

$$C_{a_0}^2 = \sum_{i=1}^{N} (D_i / \sum_{i=1}^{N} D_i) C_{i}^2,$$ \hspace{1cm} (22)

$$C_{a_i}^2 = \sum_{i=0}^{M} (\lambda_i / \lambda_i) C_{a_i}^2 = \sum_{i=0}^{M} \pi_i C_{d_i}^2, \text{ and}$$

$$C_{a_i}^2 = \pi_i C_{a_i}^2 + 1 - \pi_i \text{ for } i = 1, 2, \ldots, M + 1,$$ \hspace{1cm} (23)

where $\pi_i = \lambda_i / \lambda$. Equalities (18)-(20), along with (10), can be simultaneously solved to yield:

$$C_{a_i}^2 = \pi_i (\rho_i^2 C_{s_i}^2 + (1 - \rho_i^2) C_{a_i}^2 ) + 1 - \pi_i, \text{ for } i = 1, 2, \ldots, M + 1, \text{ and}$$

$$C_{a_i}^2 = \pi_i (\rho_i^2 C_{a_i}^2 + (1 - \rho_i^2) C_{a_i}^2 ) + 1 - \pi_i, \text{ for } i = 1, 2, \ldots, M + 1, \text{ and}$$

$$C_{a_i}^2 = \pi_i (\rho_i^2 C_{a_i}^2 + (1 - \rho_i^2) C_{a_i}^2 ) + 1 - \pi_i,$$ \hspace{1cm} (25)
The layout design problem can now be formulated as:

Minimize \( E(WIP) = \sum_{i=0}^{M+1} E(WIP_i) + E(WIP_i) \) \( (27) \)

subject to:

\[
\sum_{k=1}^{M+1} x_{ik} = 1 \quad i = 0, 2, \ldots, M + 1 \) \( (28) \)

\[
\sum_{i=0}^{M+1} x_{ik} = 1 \quad k = 1, 2, \ldots, K \) \( (29) \)

\[
\rho_i < 1 \) \( (30) \)

\[
x_{ik} = 0, 1 \quad i = 0, 2, \ldots, M + 1; k = 1, 2, \ldots, K \) \( (31) \)

The above formulation shares the same constraints, constraints (28), (29) and (31), as the QAP formulation. Constraints (28) and (29) ensure, respectively, that each department is assigned to one location and each location is assigned to one department. We require an additional constraint, constraint (30), to ensure that a selected layout is feasible and will not result in infinite work-in-process. As in the QAP formulation, we assume \( K = M+2 \). The case where \( K > M+2 \) can be handled by introducing dummy departments with zero input and output flows. The objective function is however different from that of the QAP. In the conventional QAP, the objective function is a positive linear transformation of the expected full travel time and is of the form:

Minimize \( z = \sum_i \sum_j \sum_k \sum_l x_{ij} x_{kl} \hat{\lambda}_{ij} d_{kl} \). \( (32) \)

Therefore, a solution that minimizes average full travel time between departments is optimal. Because expected WIP is not, in general, a linear function of average full travel time, the solutions obtained by the two formulations, as we show in the next section, can be different. However, a special case where the two formulations lead to the same solution is the one considered by Fu and Kaku, where all inter-arrival, processing, and transportation times are
exponentially distributed and empty travel time is negligible. In this case, we have $C_{a_i}^2 = C_{s_i}^2 = 1$, for $i = 0, 1, \ldots, M+1$, which when substituted in the expression of expected WIP, while ignoring empty travel, leads to:

$$E(WIP) = \sum_{i=0}^{M+1} \frac{\rho_i}{1 - \rho_i} + \frac{\rho_i}{1 - \rho_i},$$

with

$$\rho_t = \sum_{i} \sum_{j} \sum_{k} \sum_{l} \lambda_{ij} x_{ik} x_{il} d_{kl} / v.$$  

Since only $E(WIP_t)$ is a function of the layout and since $E(WIP_t)$ is strictly increasing in $\rho_t$, any solution that minimizes $\rho_t$ also minimizes the overall WIP. Noting that $\rho_t$ is minimized by minimizing $z = \sum_{i} \sum_{j} \sum_{k} \sum_{l} \lambda_{ij} x_{ik} x_{il} d_{kl}$ we can see that minimizing $z$ also travel or (2) relax the exponential assumption regarding inter-arrival, processing, or travel times, the equivalence between the QAP and the queueing-based model does not hold any longer.

The quadratic assignment problem has been shown to be NP-hard [16]. Since the objective function in (23) is a non-linear transformation of that of the QAP, the formulation in (27)-(31) also leads to an NP-hard problem. Although, for relatively small problems, implicit enumeration (e.g., branch and bound) can be used to solve the problem to optimality [16], for most problems, we must resort to a heuristic solution approach. Several heuristics have been proposed for solving the QAP (see Pardalos and Wolkowicz [16] for a recent review) and any of these could be used to solve our model as well. In a recent software implementation of the formulation in (23)-(31), Yang and Benjaafar [24] used both implicit enumeration and a modified 2-opt heuristic, similar to the one proposed by Fu and Kaku [5], to solve the problem. In this paper, we limit our discussion to mostly layouts where the QAP-optimal layout is easily identified.

### 3. Model Analysis and Insights

We first consider the case where the assumption of exponential inter-arrival, processing, and transportation times still holds but we explicitly account for empty travel. In section 3.2, we consider the more general case.
3.1 The Exponential Case

In this case, expected WIP can still be obtained exactly as:

\[ E(WIP) = \sum_{i=0}^{M+1} \frac{\rho_i}{1 - \rho_i} + \frac{\rho_t}{1 - \rho_t} \]

with,

\[ \rho_t = \rho_t^e + \rho_t^f = \sum_{r=1}^{M+1} \sum_{i=0}^{M} \frac{\lambda_r \lambda_i}{\lambda_t} t_{ri}(x) + \sum_{i=0}^{M} \sum_{j=1}^{M+1} \lambda_{ij} t_{ij}(x) \]

We can see that, once again, only \( E(WIP_t) \) is a function of the layout. We can also see that \( E(WIP_t) \) is strictly decreasing in \( \rho_t \), and, therefore, a solution that minimizes \( \rho_t \) minimizes \( E(WIP_t) \). However, in this case, \( \rho_t \) is the sum of two components, \( \rho_e \) and \( \rho_f \). In the following set of observations, we show that a layout that minimizes \( \rho_f \) does not necessarily minimize \( \rho_e \), and consequently, a layout that minimizes \( \rho_f \) does not necessarily minimize WIP. In fact, we show that a QAP-optimal layout (i.e., a layout that minimizes \( \rho_f \)) is not even guaranteed to be feasible. More generally, we show that two QAP-optimal layouts can result in different WIP values. Furthermore, under certain conditions, we find that WIP is reduced more effectively by reducing empty travel, even if this increases full travel. This means that it can be sometimes desirable to place departments in neighboring locations even though there is no direct material flow between them. This also means that it can be beneficial to place departments with high inter material flows in distant locations from each other.

**Observation 1:** A layout that minimizes full travel does not necessarily minimize WIP.

The result follows from noting that reducing \( \rho_f \) can increase \( \rho_e \). If the increase in \( \rho_e \) is sufficiently large, an increase in expected WIP can then follow. We illustrate this result using the following example. Consider a system consisting of 12 locations and 12 departments arranged in a 3x4 grid as shown in figure 2(a). Departments are always visited by all products in the following sequence: \( 0 \to 1 \to 2 \to 3 \to 2 \to 3 \to 4 \to 5 \to 6 \to 7 \to 8 \to 9 \to 8 \to 9 \to 8 \to 9 \to 8 \to 9 \to 10 \to 11 \). The distance matrix between locations is shown in figure 2(b) - we assume rectilinear distances with unit distance separating adjacent locations. Average processing times at the departments are shown in figure 2(c). Material handling speed is 1.65 (units of distance per unit of time) and overall demand rate is 0.027 (unit loads per unit time). Let us consider the two layouts shown in Figures 3(a) and 3(b), denoted respectively by \( x_1 \) and \( x_2 \) (the arrows are used to indicate the direction of material flow). It is easy to verify that layout \( x_1 \) is QAP-optimal and...
(a) Available department locations

From/To  0  1  2  3  4  5  6  7  8  9  10  11
0        0  1  2  3  4  3  2  1  2  3  4  5
1        1  0  1  2  3  2  1  2  3  2  3  4
2        2  1  0  1  2  1  2  3  4  3  2  3
3        3  2  1  0  1  2  3  4  5  4  3  2
4        4  3  2  1  0  1  2  3  4  3  2  1
5        3  2  1  2  1  0  1  2  3  2  1  2
6        2  1  2  3  2  1  0  1  2  1  2  3
7        1  2  3  4  3  2  1  0  1  2  3  4
8        2  3  4  5  4  3  2  1  0  1  2  3
9        3  2  3  4  3  2  1  2  1  0  1  2
10       4  3  2  3  2  1  2  3  2  1  0  1
11       5  4  3  2  1  2  3  4  3  2  1  0

(b) Distances between department locations

<table>
<thead>
<tr>
<th>Departments</th>
<th>Average processing time</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>18</td>
</tr>
<tr>
<td>1</td>
<td>18</td>
</tr>
<tr>
<td>2</td>
<td>6</td>
</tr>
<tr>
<td>3</td>
<td>6</td>
</tr>
<tr>
<td>4</td>
<td>18</td>
</tr>
<tr>
<td>5</td>
<td>18</td>
</tr>
<tr>
<td>6</td>
<td>18</td>
</tr>
<tr>
<td>7</td>
<td>18</td>
</tr>
<tr>
<td>8</td>
<td>6</td>
</tr>
<tr>
<td>9</td>
<td>6</td>
</tr>
<tr>
<td>10</td>
<td>18</td>
</tr>
<tr>
<td>11</td>
<td>18</td>
</tr>
</tbody>
</table>

(c) Department average processing times

Figure 2 - Data for example layout
(a) Layout $x_1$
$\rho_t = 0.990$, $\rho_f = 0.311$, $\rho_e = 0.679$, $E(WIP_t) = 99.0$

(b) Layout $x_2$
$\rho_t = 0.951$, $\rho_f = 0.409$, $\rho_e = 0.542$, $E(WIP_t) = 19.41$

(c) Layout $x_3$
$\rho_t = 0.885$, $\rho_f = 0.344$, $\rho_e = 0.542$, $E(WIP_t) = 7.70$

Figure 3 - Example layouts
Figure 3 - Example layouts – continued

(d) Layout $x_4$
\[ \rho_t = 0.961, \rho_f = 0.327, \rho_e = 0.634, E(WIP_t) = 24.64 \]

(e) Layout $x_5$
\[ \rho_t = 1.04, \rho_f = 0.344, \rho_e = 0.695, E(WIP_t) = \infty \]

(f) Layout $x_6$
\[ \rho_t = 0.992, \rho_f = 0.360, \rho_e = 0.632, E(WIP_t) = 124.0 \]

(g) Layout $x_7$
\[ \rho_t = 0.853, \rho_f = 0.311, \rho_e = 0.542, E(WIP_t) = 5.80 \]
minimizes full travel. In contrast, layout $x_2$ is not QAP-optimal and, in fact, appears to be quite inefficient. Expected material handling system WIP for layout $x_1$ and $x_2$, as well as the corresponding full and empty material handling system utilizations, are shown in figures 3(a) and 3(b). For both layouts, the overall expected WIP due to the processing departments is the same and is equal to 11.346. We can see that, although layout $x_2$ does not minimize full travel, it results in significantly less empty travel which is sufficient to cause an overall reduction in material handling system utilization. As a consequence, expected WIP for layout $x_2$ is smaller than that of $x_1$. In fact, material handling system WIP is reduced by more than 80% (from 99.0 to 19.41) when layout $x_2$ is chosen over $x_1$!

This surprising result stems from the fact that the frequency with which a device makes empty trips to a particular department is proportional to the volume of outflow from that department. The likelihood of the material handling device being in a particular department is similarly proportional to the volume of inflow to that department. Therefore, if two department are highly loaded, the number of empty trips between them would be large even if no direct flow exists between these departments. In our example, departments 2, 3, 9, and 8 have 3 times the workload of any other department in the factory. Therefore, the likelihood of an empty trip between any 2 of the 4 departments is 3 times higher than between any other two departments. In layout $x_2$, by placing these four departments in neighboring locations, empty travel is significantly reduced. Note that this is realized despite the fact that there is no direct material flow between the department pairs 2-3 and 8-9.

The above result also leads us to the following more general observation which further highlights the fact that full travel is a poor indicator of WIP.

**Observation 2:** Expected WIP is not monotonic in $\rho_f$. Therefore, an increase in $\rho_f$ can either increase or decrease expected WIP.

Observation 2 follows from the fact that an increase in $\rho_f$ can result in either an increase or a decrease in $\rho_e$. Depending on how $\rho_e$ is affected, expected WIP may either increase or decrease. We illustrate this behavior by considering a series of layout configurations based on our previous example. The layouts, denoted $x_1$, $x_2$, ..., $x_7$, are shown in figure 3. The behavior of $\rho_f$, $\rho_e$, $\rho_t$, and $E(WIP_t)$ is graphically depicted in Figure 4. It is easy to see that $\rho_f$ can behave quite differently from $\rho_e$ and $\rho_t$. It is also easy to see that an increase or a decrease in $\rho_f$ does not always have predictable consequences on expected WIP.
(a) The effect of layout configuration on WIP

(b) The effect of layout configuration on material handling utilization

Figure 4 - The effect of layout configuration on utilization and WIP
The fact that $\rho_f$ can behave differently from $\rho_e$ means that it is possible to have layouts with similar values of $\rho_f$ but different values of $\rho_e$. This also means that layouts could have the same value of $\rho_f$ but different values of expected WIP. In fact, it is possible to have two QAP-optimal layouts with very different WIP values.

**Observation 3:** Two QAP-optimal layouts can have different WIP values.

The result follows from noting that two layouts can have the same $\rho_f$ but different values of $\rho_e$. For example, consider the two layouts, $x_1$ and $x_7$, shown in figure 3. Both layouts are QAP-optimal. However, $E(WIP|x_1) = 99$ and $E(WIP|x_7) = 5.8$. The above result shows that QAP-optimality can be a poor indicator of WIP performance. In fact, as we show in the next observation, a QAP-optimal layout can be, not only inefficient, but sometimes infeasible.

**Observation 4:** A QAP-optimal layout can be infeasible.

The result is due to the fact that, even though $\rho_f$ might be minimal, the corresponding $\rho_e$ can be sufficiently large to make $\rho_t$ greater than 1. We illustrate this result using the following example. Consider the same system description we used for the previous 3 observations except that material handling system speed is 1.6 instead of 1.65. Now consider the performance of the layout configurations $x_1$ and $x_3$ shown in figure 3. We have $E(WIP|x_1) = \infty$ while $E(WIP|x_3) = 10.5$. Thus, although layout $x_1$ is QAP-optimal, it is infeasible. Layout $x_3$ is non-optimal but produces a relatively small WIP. Clearly, QAP-optimality does not guarantee feasibility.

The previous four observations show that even if we retained the exponential assumption about inter-arrival, processing and transportation times, the QAP objective function can be a poor predictor of WIP. Therefore, there is a need to explicitly evaluate WIP if our objective is to design layouts that minimize it. In fact, regardless of the objective function, there is always a need to at least evaluate both empty and full travel by the material handling system since we must always generate feasible layouts. The fact that empty travel can be a significant portion of material handling system utilization also means that we need to design layouts that minimize it. This may sometimes result in going counter the common practice of favoring the placement of departments with large inter-material flows in neighboring locations. As we saw in the previous examples, reducing WIP could lead to departments being placed in adjacent locations although there is no direct material flow between them (e.g., the department pairs 2-3 and 9-8). Since empty travel is more frequent from and to departments that are popular destinations (i.e., departments with high flow rates), placing these departments in neighboring locations can significantly reduce empty travel even when there is no direct flow between these departments.
Therefore, the need to reduce full travel by placing department with large inter material flows in neighboring locations must be balanced by the need to reduce empty travel by placing departments that are popular destinations also close proximity.

3.2 The General Case

Let us now consider the general case where inter-arrival times, processing times, and transportation times are not necessarily exponentially distributed. In this case, expected WIP is given by equalities (16) and (18). Examining the expression of expected WIP, we can see that WIP accumulation is determined by (1) the variability in the arrival process, (2) the variability in the processing/transportation times, and (3) the utilization of the departments and the material handling system. We can also see that, because the material handling system provides input to all the processing departments, variability in transportation time, as well as the material handling system utilization, directly affect the variability in the arrival process to all the departments. In turn, this variability, along with the variability of the department processing times and the department utilizations, determine the input variability to the material handling system. Because of this close coupling, the variability of any resource, and its utilization, affect the WIP at all other resources. This effect is not captured by the exponential model and, as we show next, can lead to very different results with regard to layout WIP performance.

In the previous section we showed that expected WIP, although not monotonic in full travel utilization, is monotonic in overall material handling system utilization. In other words, we showed that a layout that minimizes average travel time (both full and empty) would also minimize WIP. Here, we show that, when we relax the exponential assumption, this is not necessarily true. In fact, we show that reducing average travel time (i.e., reducing $\rho_t$) can increase WIP. As a result, increasing the average distance between departments could, in fact, reduce WIP. Moreover, we show that the relative desirability of a layout can be highly sensitive to changes in material handling capacity even when travel distances remain the same. We also find that WIP accumulation at the material handling system can be affected by non material handling factors, such as the utilization of the processing departments or variability in the department processing time, which means that the relative desirability of two layouts could be affected by these factors.
**Observation 5:** A smaller average travel time (full + empty) does not always lead to a smaller WIP.

The proof of observation 5 follows by noting that the expression of expected WIP is a function of both $\rho_t$ and $C_{s_i}^2$. Since $C_{s_i}^2$ is not necessarily decreasing in $\rho_t$, a reduction in $\rho_t$ may indeed cause an increase in $C_{s_i}^2$, which could be sufficient to either increase material handling WIP or increase the arrival variability at the processing departments, which, in turn, could increase their WIP. We illustrate this behavior using the following example. Consider a facility with four departments ($i = 0, 1, 2, \text{and } 3$). Products in the facility are always manufactured in the following sequence: $0 \rightarrow 1 \rightarrow 2 \rightarrow 1 \rightarrow 2 \rightarrow 1 \rightarrow 2 \rightarrow 3$.

Other relevant data is as follows: $D_1 = 0.027$; $E(S_0) = E(S_3) = 30 \text{ and } E(S_1) = E(S_2) = 10$, $C_{a_0}^2 = 1.0, C_{s_i}^2 = 0.5 \text{ for } i = 0, 1, \ldots, 3, \text{ and } v = 0.68$.

We consider two layout scenarios, $x_8$ and $x_9$. The distances between departments are as follows, layout $x_8$: $d_{01}(x_8) = d_{02}(x_8) = d_{03}(x_8) = d_{12}(x_8) = d_{13}(x_8) = d_{23}(x_8) = 2$; and layout 2: $d_{01}(x_9) = 1, d_{02}(x_9) = 2, d_{03}(x_9) = 8, d_{12}(x_9) = 1, d_{13}(x_9) = 7, \text{ and } d_{23}(x_9) = 6$. The two layouts are graphically depicted in figure 5. Since $E(WIP(x_8)) = 15.87 < E(WIP(x_9)) = 19.26$ although $\rho_t(x_8) = 0.907 > \rho_t(x_9) = 0.896$, our result is proven. The effect of variability can be even more pronounced for facilities with asymmetric distances. For example, consider the layout $x_{10}$ similar to $x_9$ with the exception of the following distances $d_{23}(x_{10}) = 2, d_{30}(x_{10}) = 7, d_{31}(x_{10}) = 14 \text{ and } d_{32}(x_{10}) = 7$. In this case, we have $E(WIP(x_8)) = 15.87 < E(WIP(x_{10})) = 20.39$ although $\rho_t(x_8) = 0.907 > \rho_t(x_{10}) = 0.867$. Distance asymmetry is frequently encountered in practice and is due to factors such as the configuration of the material handling system (e.g., bi-directional versus uni-directional AGV’s), location of department pick-up and drop-off points, and zoning restrictions (e.g., material flow is forbidden through certain regions of the layout due to safety hazards) [19, 20].

The above results show the important effect that variability in travel times can play in determining overall WIP. In each of the above examples, the smaller value of average travel time is associated with higher travel time variability. This higher variability causes not only an increase in material handling WIP but also in department WIP (by increasing variability in the arrival process to the departments). These results point to the need for explicitly accounting for travel time variance when selecting a layout. A layout that exhibits a small variance may, indeed, be more desirable than one with a smaller travel time average. In practice, travel time variance is often dictated by the material handling system configuration. Therefore, special attention should be devoted to identifying configurations that minimize not only average travel...
\[ \rho_t = 0.907, \rho_f = 0.555, \rho_e = 0.352, C_{s_1}^2 = 0.087, C_{a_1}^2 = 0.567, C_{a_2}^2 = C_{a_2}^2 = 0.645, C_{a_3}^2 = 0.882, \]
\[ E(WIP_t) = 3.78, E(WIP) = 15.87 \]

**Layout x_8**

\[ \rho_t = 0.896, \rho_f = 0.476, \rho_e = 0.420, C_{s_1}^2 = 0.735, C_{a_1}^2 = 0.636, C_{a_2}^2 = C_{a_2}^2 = 0.878, C_{a_3}^2 = 0.959, \]
\[ E(WIP_t) = 6.16, E(WIP) = 19.26 \]

**Figure 5 - Example layouts for observation 5**
time but also its variance. For example, the star-layout configuration, shown in Figure 6(a) has a significantly smaller variance than the loop layout of 6(b), which itself has a smaller variance than the linear layout of 6(c).

Although, in the above examples, the layout with the lower variance is more desirable, we should caution that this relative desirability can be sensitive to the available material handling capacity. For example, from the stability condition ($\rho_t < 1$), we can see that the minimum feasible material handling speed is higher for layout $x_8$ than for layouts $x_9$ and $x_{10}$. This means that for certain material handling speeds layout $x_8$ is, indeed, infeasible while layouts $x_9$ and $x_{10}$ still result in finite WIP. More generally, as shown in Figure 7, the relative ranking of layouts can be affected by changes in material handling capacity. For example, layout $x_8$ is superior to layout $x_{10}$ when material handling speed is greater than 0.63 but it is clearly inferior for lower speeds. These results lead to the following important observation.

**Observation 6:** Layout desirability (relative ranking) is non-monotonic in material handling capacity.

Observation 6 highlights the fact that material handling capacity can have an unpredictable impact on layout desirability. It also points to the complex relationship between distribution of travel time, material handling capacity and WIP performance.

Travel distances and material handling capacity are not, however, the only factors that affect the relative desirability of a layout. Non-material handling factors such as department utilization levels, variability in department processing times, and variability in demand levels could determine whether one layout configuration is more desirable than another. For example, in the following observation, we show that variability in processing times and demand can significantly affect the relative desirability of a layout.

**Observation 7:** The relative desirability of a layout can be affected by non-material handling factors.

Since the arrival variability to the processing departments and the material handling system is affected by the utilization of the processing departments, the processing time variability, and the variability in product demands, it is possible that changes in these parameters could affect the relative desirability of a particular layout. We illustrate this behavior by considering layouts $x_8$ and $x_{10}$ described in observation 5 (in this case, we let material handling speed be 0.632). In Table 1, we show the effect of processing time and demand variability on the performance of the
(a) Star layout

(a) Loop layout

(a) Linear layout

Figure 6 - Star, loop and linear layouts
Figure 7 - The effect of material handling capacity on WIP performance
two layouts. As we can see, the same layout can be superior under one set of parameters and inferior under another.

Table 1 - The effect of variability on layout performance

<table>
<thead>
<tr>
<th>Variability</th>
<th>( E(WIP(x_8)) )</th>
<th>( E(WIP(x_{10})) )</th>
</tr>
</thead>
<tbody>
<tr>
<td>( C_{a_0}^2 = 0.2, C_{s_i}^2 = 0.2 )</td>
<td>17.48</td>
<td>22.80</td>
</tr>
<tr>
<td>( C_{a_0}^2 = 0.5, C_{s_i}^2 = 0.5 )</td>
<td>24.57</td>
<td>26.99</td>
</tr>
<tr>
<td>( C_{a_0}^2 = 1.0, C_{s_i}^2 = 1.0 )</td>
<td>36.24</td>
<td>33.85</td>
</tr>
<tr>
<td>( C_{a_0}^2 = 1.2, C_{s_i}^2 = 1.2 )</td>
<td>40.86</td>
<td>36.55</td>
</tr>
</tbody>
</table>

Since the results of observations 5-7 are based on approximations for both average WIP and the arrival processes to the various departments, we used computer simulation to confirm them. For each of the example layouts, we constructed a stochastic simulation model using the discrete event simulation language Arena [12]. The simulated models are identical to the analytical ones, except that the travel time distribution is not pre-specified. Instead, we provide the simulation model with the distances between departments, material handling speed and product routings. In contrast with the analytical approximations, the simulation model does away with the probabilistic routing assumption and captures dependencies between the length of consecutive trips that tend to occur in real systems (e.g., a long trip that takes the material handling device to the outer edges of the layout tend to be followed by another long trip). For each case, we collected statistics on average WIP at the different processing departments and material handling system. For each case, we also obtained a 95% confidence interval with a sufficiently small half width.

Although specific values of the approximated average WIP are not always within the simulation 95% confidence interval, the simulated results confirmed each of the three observations (in each case, the relative ranking of the simulated layouts is consistent with the one obtained analytically; also, in each case, the differences between ranked layouts are found to be
statistically significant). We found the inaccuracy in estimating overall WIP to be mostly due to inaccuracies in estimating the variability in the arrival process to the departments and the material handling system. This is especially significant when both demand and processing time variability are small. In this case, variability is over-estimated which, in turn, results in higher estimates of WIP. This effect is due to the probabilistic approximation used in determining the origin of material handling requests. This limitation can be, in part, addressed by extending the queuing network model to account for multi-product deterministic job routings (see, for example, Bitran and Tirupati [2]).

4. When Does Minimizing WIP Matter?

We have so far highlighted instances where the WIP formulation leads to a different layout from the one obtained using the QAP formulation. In this section, we examine conditions where the two formulations are \textit{WIP-equivalent}. We also highlight conditions that cause the difference in WIP between the two formulations to decrease. We focus primarily on the impact of three system characteristics: flow asymmetry, dimensional asymmetry, and material handling capacity, whose values we found to play an important role in determining WIP difference.

4.1 The Effect of Flow Asymmetry

The term flow asymmetry refers to the unbalance in workload among departments. In an asymmetric system, some departments are more visited than others, leading to more empty trips ending and originating at these departments. Consequently, even when there are no direct flows between these departments, empty travel utilization is affected by their placement relative to each other. As we saw in previous examples, this can lead to a significant difference in WIP between a QAP-optimal and a WIP-optimal layout. In contrast, in a symmetric system where department workloads are equal, empty travel is \textit{layout-independent} since there is equal likelihood of an empty trip originating at any department and ending at any other department. As a result, the difference in expected WIP between a QAP-optimal and a WIP-optimal layout is always zero. This leads us to the following observation.

\textbf{Observation 8:} In a flow-symmetric system, a QAP-optimal and a WIP-optimal layout are \textit{WIP-equivalent}.

The above observation may lead us to assume that any decrease in flow asymmetry would always result in a smaller percentage difference in WIP between a QAP-optimal and a WIP-
optimal layout. However, this is not true. In fact, the effect of flow asymmetry – as measured, for example, by the variance in department workloads - is generally not monotonic.

**Observation 9:** A decrease in flow asymmetry does not necessarily reduce the percentage difference in expected WIP between a QAP-optimal and a WIP-optimal layout.

In order to prove this result, we consider a system consisting of 8 products \(P_1, P_2, \ldots, P_8\) and 16 departments arranged in a 4×4 block layout. The department routing sequence for each product is as follows: \(P_1: 1 \rightarrow 2 \rightarrow 3 \rightarrow 4\); \(P_2: 5 \rightarrow 6 \rightarrow 7 \rightarrow 8\); \(P_3: 9 \rightarrow 10 \rightarrow 11 \rightarrow 12\); \(P_4: 13 \rightarrow 14 \rightarrow 15 \rightarrow 16\); \(P_5: 1 \rightarrow 5 \rightarrow 9 \rightarrow 13\); \(P_6: 2 \rightarrow 6 \rightarrow 10 \rightarrow 14\); \(P_7: 3 \rightarrow 7 \rightarrow 11 \rightarrow 15\); and \(P_8: 4 \rightarrow 8 \rightarrow 12 \rightarrow 16\). Let us also consider the product demand scenarios shown in Table 2, where total demand is kept constant but the contribution of individual products is varied. This results in department flow scenarios with varying levels of asymmetry as shown in Table 3. We use the variance \(\sigma_f^2\), in the department flow rates to measure asymmetry under each scenario. We also use the ratio \(\delta = E(WIP|\text{QAP})/E(WIP|\text{WIP})\) to measure the relative magnitude of WIP under a QAP-optimal layout relative to that obtained from a WIP-optimal one. In order to allow for a fair comparison between scenarios, the capacity of a department is adjusted proportionally to its workload in order to keep a constant utilization per department (in our case, 0.8). We also let the squared coefficients of variation of processing times and external inter-arrival times all be equal to 1. The QAP-optimal layout is shown in Figure 8. As we can see, the QAP-optimality of this layout is independent of the demand scenarios.

The effect of flow asymmetry on the WIP ratio is shown in Figure 9 for varying levels of material handling capacity. We can see that the effect of \(\sigma_f\) is indeed not monotonic. Although initial increases in \(\sigma_f\) do lead to a larger \(\delta\), additional increases invariably result in a smaller WIP ratio. Thus, \(\delta\) is maximum when \(\sigma_f\) is in the midrange and is significantly smaller in the extreme cases of either high or low asymmetry. A possible explanation for this non-monotonic behavior is as follows. In a highly asymmetric system, the demand from one product dominates the demand from all others. Hence the departments that are most visited are those that are visited by the product with the highest demand. Since these departments are already in neighboring locations under the QAP-optimal layout, the additional reduction in empty travel due to using the WIP criteria is limited. This is in contrast to situations where asymmetry is due to two or more products having relatively higher demands than the others. In that case, rearranging the layout so that the departments visited by these products are in neighboring locations does significantly reduce empty travel.
### Table 2– Demand scenarios for example system

<table>
<thead>
<tr>
<th>Scenario</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
</tr>
</thead>
<tbody>
<tr>
<td>$D_1$</td>
<td>40</td>
<td>40</td>
<td>58</td>
<td>58</td>
<td>76</td>
<td>112</td>
<td>148</td>
<td>202</td>
<td>256</td>
<td>292</td>
</tr>
<tr>
<td>$D_2$</td>
<td>40</td>
<td>40</td>
<td>22</td>
<td>22</td>
<td>4</td>
<td>4</td>
<td>4</td>
<td>4</td>
<td>4</td>
<td>4</td>
</tr>
<tr>
<td>$D_3$</td>
<td>40</td>
<td>40</td>
<td>22</td>
<td>22</td>
<td>4</td>
<td>4</td>
<td>4</td>
<td>4</td>
<td>4</td>
<td>4</td>
</tr>
<tr>
<td>$D_4$</td>
<td>40</td>
<td>40</td>
<td>58</td>
<td>58</td>
<td>76</td>
<td>112</td>
<td>148</td>
<td>94</td>
<td>40</td>
<td>4</td>
</tr>
<tr>
<td>$D_5$</td>
<td>40</td>
<td>58</td>
<td>58</td>
<td>76</td>
<td>76</td>
<td>40</td>
<td>4</td>
<td>4</td>
<td>4</td>
<td>4</td>
</tr>
<tr>
<td>$D_6$</td>
<td>40</td>
<td>22</td>
<td>22</td>
<td>4</td>
<td>4</td>
<td>4</td>
<td>4</td>
<td>4</td>
<td>4</td>
<td>4</td>
</tr>
<tr>
<td>$D_7$</td>
<td>40</td>
<td>22</td>
<td>22</td>
<td>4</td>
<td>4</td>
<td>4</td>
<td>4</td>
<td>4</td>
<td>4</td>
<td>4</td>
</tr>
<tr>
<td>$D_8$</td>
<td>40</td>
<td>58</td>
<td>58</td>
<td>76</td>
<td>76</td>
<td>40</td>
<td>4</td>
<td>4</td>
<td>4</td>
<td>4</td>
</tr>
</tbody>
</table>

### Table 3– Department arrival rates per scenario for example system

<table>
<thead>
<tr>
<th>Scenario</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\lambda_1$</td>
<td>80</td>
<td>98</td>
<td>116</td>
<td>134</td>
<td>152</td>
<td>152</td>
<td>152</td>
<td>206</td>
<td>260</td>
<td>296</td>
</tr>
<tr>
<td>$\lambda_2$</td>
<td>80</td>
<td>62</td>
<td>80</td>
<td>62</td>
<td>80</td>
<td>116</td>
<td>152</td>
<td>206</td>
<td>260</td>
<td>296</td>
</tr>
<tr>
<td>$\lambda_3$</td>
<td>80</td>
<td>62</td>
<td>80</td>
<td>62</td>
<td>80</td>
<td>116</td>
<td>152</td>
<td>206</td>
<td>260</td>
<td>296</td>
</tr>
<tr>
<td>$\lambda_4$</td>
<td>80</td>
<td>98</td>
<td>116</td>
<td>134</td>
<td>152</td>
<td>152</td>
<td>152</td>
<td>206</td>
<td>260</td>
<td>296</td>
</tr>
<tr>
<td>$\lambda_5$</td>
<td>80</td>
<td>98</td>
<td>80</td>
<td>98</td>
<td>80</td>
<td>44</td>
<td>8</td>
<td>8</td>
<td>8</td>
<td>8</td>
</tr>
<tr>
<td>$\lambda_6$</td>
<td>80</td>
<td>62</td>
<td>44</td>
<td>26</td>
<td>8</td>
<td>8</td>
<td>8</td>
<td>8</td>
<td>8</td>
<td>8</td>
</tr>
<tr>
<td>$\lambda_7$</td>
<td>80</td>
<td>62</td>
<td>44</td>
<td>26</td>
<td>8</td>
<td>8</td>
<td>8</td>
<td>8</td>
<td>8</td>
<td>8</td>
</tr>
<tr>
<td>$\lambda_8$</td>
<td>80</td>
<td>98</td>
<td>80</td>
<td>98</td>
<td>80</td>
<td>44</td>
<td>8</td>
<td>8</td>
<td>8</td>
<td>8</td>
</tr>
<tr>
<td>$\lambda_9$</td>
<td>80</td>
<td>98</td>
<td>80</td>
<td>98</td>
<td>80</td>
<td>44</td>
<td>8</td>
<td>8</td>
<td>8</td>
<td>8</td>
</tr>
<tr>
<td>$\lambda_{10}$</td>
<td>80</td>
<td>62</td>
<td>44</td>
<td>26</td>
<td>8</td>
<td>8</td>
<td>8</td>
<td>8</td>
<td>8</td>
<td>8</td>
</tr>
<tr>
<td>$\lambda_{11}$</td>
<td>80</td>
<td>62</td>
<td>44</td>
<td>26</td>
<td>8</td>
<td>8</td>
<td>8</td>
<td>8</td>
<td>8</td>
<td>8</td>
</tr>
<tr>
<td>$\lambda_{12}$</td>
<td>80</td>
<td>98</td>
<td>80</td>
<td>98</td>
<td>80</td>
<td>44</td>
<td>8</td>
<td>8</td>
<td>8</td>
<td>8</td>
</tr>
<tr>
<td>$\lambda_{13}$</td>
<td>80</td>
<td>98</td>
<td>116</td>
<td>134</td>
<td>152</td>
<td>152</td>
<td>152</td>
<td>98</td>
<td>44</td>
<td>8</td>
</tr>
<tr>
<td>$\lambda_{14}$</td>
<td>80</td>
<td>62</td>
<td>80</td>
<td>62</td>
<td>80</td>
<td>116</td>
<td>152</td>
<td>98</td>
<td>44</td>
<td>8</td>
</tr>
<tr>
<td>$\lambda_{15}$</td>
<td>80</td>
<td>62</td>
<td>80</td>
<td>62</td>
<td>80</td>
<td>116</td>
<td>152</td>
<td>98</td>
<td>44</td>
<td>8</td>
</tr>
<tr>
<td>$\lambda_{16}$</td>
<td>80</td>
<td>98</td>
<td>116</td>
<td>134</td>
<td>152</td>
<td>152</td>
<td>152</td>
<td>98</td>
<td>44</td>
<td>8</td>
</tr>
</tbody>
</table>

| $\lambda_t$ | 1280 | 1280 | 1280 | 1280 | 1280 | 1280 | 1280 | 1280 | 1280 | 1280 |
| $\sigma_f$  | 0    | 19   | 26   | 42   | 53   | 59   | 74   | 84   | 108  | 129  |
Figure 8 – The QAP-optimal layout

Figure 9 – The Effect of flow asymmetry
4.2 The Effect of Material Handling Capacity

As we can see from Figure 9, and more clearly from figure 10, the WIP ratio is also sensitive to the utilization of the material handling system. While for lightly loaded systems, the two layouts are practically WIP-equivalent (regardless of flow symmetry), the WIP ratio is most significant when system loading is high. In fact, as utilization approaches 1, $\delta$ grows without bound.

These results are of course not surprising. A WIP-optimal layout would generally result in a smaller fraction of utilization devoted to empty travel. This reduction in empty travel is of little consequence when there is excess material handling capacity. However, it becomes crucial when material handling capacity is tight. In fact, given that WIP grows exponentially in the utilization of the material handling system, even small decreases in empty travel would have a dramatic impact on WIP accumulation when utilization is high.

**Observation 10:** The difference in WIP between a QAP-optimal and a WIP-optimal layout is generally increasing in the utilization of material handling system.

Although the above result is generally true, there are instances when an increase in utilization could affect the variability of travel times sufficiently to cause a decrease in the WIP ratio. This does not occur often since the effect of utilization tends to dominate the effect of variability, particularly when material handling system utilization is high.

4.3 The Effect of Dimensional Asymmetry

The term dimensional asymmetry refers to asymmetry in the distances between different department locations. In a perfectly symmetric system, all department locations are equidistant from each other. In this case, full and empty travel are always the same regardless of department placement. Hence, a QAP-optimal layout (or any other layout) is also WIP-optimal. This leads to the following obvious observation, which we restate for the sake of completeness.

**Observation 11:** In a dimension-symmetric system, a QAP-optimal and a WIP-optimal layout are WIP-equivalent.

Although it is difficult to predict in general the impact of an increase in dimensional asymmetry on the WIP ratio - this would largely depend on the specific geometry of the layout and the distribution of the flow among departments - large increases in asymmetry tend to increase the difference in WIP between a QAP-optimal layout and a WIP-optimal layout. (a notable exception to this is a flow-symmetric system). In order to illustrate this effect, we
Figure 10 – The Effect of material handling capacity

Figure 11 – The Effect of dimensional asymmetry
consider the same example used in the previous two sections but allow the length to width ratio of the layout to vary. Specifically, we consider layouts with a 4x4, 2x8 and 1x16 footprint. Obviously, of the three layouts, the 1x16 is the most asymmetric and the 4x4 is the least. The value of the WIP ratio for each layout geometry is shown in Figure 11 for varying levels of flow asymmetry. As we can see, the WIP ratio is largest for the most asymmetric system and smallest for the most symmetric one. The effect of dimensional asymmetry is sensitive to flow asymmetry, with the WIP ratio being largest when flow asymmetry is highest.

We should note however that the effect of dimensional asymmetry is not always monotonic. Depending on the distribution of flow among departments, it is possible to see a reduction in the WIP ratio if an increase in dimensional asymmetry yields (unintentionally) a QAP solution where departments with the most flow are in neighboring locations. This tends to occur less frequently when both dimensional and flow asymmetry are high. In these cases, the QAP formulation does not usually favor placing departments in neighboring locations unless they have direct flows between them.

4.4 Managerial Implications

Table 4 provides a summary of our results and offers broad guidelines as to when using WIP as a design criterion is particularly valuable. The results suggest that a WIP-optimized layout is most beneficial when material handling capacity is limited, dimensional asymmetry is high, or when there is asymmetry in the flows. These results also point to strategies that managers and facility planners could pursue to increase the robustness of layouts with respect to WIP performance – i.e., investing in excess material handling capacity, adopting layout geometries that reduce travel distance variance, and ensuring that the most visited processes are centrally located.

Furthermore, these results draw attention to the importance of indirect interactions that take place between different areas of a facility. These indirect effects have implications to the way we should organize areas of a facility that may otherwise appear independent. For example, consider a system consisting of multiple cells that do not share any products or processes but are serviced by the same material handling system. Our results suggest that among these cells those that manufacture the products with the highest demand should be placed in neighboring locations (although they do not share any flows). Our results also show that organizing these cells into
parallel production lines, a common practice in many facilities, may lead to greater congestion and longer lead times. Instead adopting a configuration that minimizes distance asymmetries (e.g., using a layout where cells are configured into a U-shape and are arranged along a common corridor where most travel would take place) would maintain the efficient transfer of material within cells while freeing up additional material handling capacity to service the entire facility.

Many companies are beginning to realize the importance of these indirect effects and are increasingly designing layouts that minimize dimensional asymmetries and reduce empty travel. For example, GM built its new Cadillac plant in the form of a T to maximize supplier access to the factory floor and reduce the distance between loading docks and production stocking points [7]. Volvo designed its Kalmar plant as a collection of hexagon-shaped modules where material flows in concentric lines within each module [20]. Motorola is experimenting with layouts, where shared processors are centrally located in functional departments and are equidistant from multiple dedicated cells within the plant. Variations of the spine layout, where departments are placed along the sides of a common corridor, have been successfully implemented in industries ranging from electronic manufacturing to automotive assembly [17, 19, 20]. Layout configurations that minimize dimensional asymmetries and reduce empty travel are also found in non-manufacturing applications. For example, both the spine and star layouts are common configurations in airport designs. Spine and T-shaped layouts are also popular designs for freight and cross-docking terminals [8].

Table 4 - When is using the WIP criterion valuable?

<table>
<thead>
<tr>
<th></th>
<th>Low</th>
<th>Medium</th>
<th>High</th>
</tr>
</thead>
<tbody>
<tr>
<td>Flow asymmetry</td>
<td>Less valuable</td>
<td>More valuable</td>
<td>Moderately valuable</td>
</tr>
<tr>
<td>Dimensional asymmetry</td>
<td>Less valuable</td>
<td>Moderately valuable</td>
<td>More valuable</td>
</tr>
<tr>
<td>Material handling capacity</td>
<td>More valuable</td>
<td>Moderately valuable</td>
<td>Less valuable</td>
</tr>
</tbody>
</table>
5. Concluding Comments

In this paper, we showed that minimizing material handling travel distances does not always reduce WIP. Therefore, the criterion used in the QAP formulation of the layout design problem cannot be used as a reliable predictor of WIP. Because the QAP formulation accounts only for full travel, an optimal solution to the QAP problem tends to favor placing departments that have large inter-material flows in neighboring locations. In this paper, we showed that, when we account for empty travel, this may not always be desirable. Indeed, it can be more beneficial if departments that have no direct material flow between them are placed in neighboring locations. In particular, we found that empty travel can be significantly reduced by placing the most frequently visited departments in neighboring locations regardless of the amount of flows between these departments. Because, WIP is affected by both mean and variance of travel time, we found that reducing travel time variance can be as important as reducing average travel time. Equally important, we found that the relative desirability of a layout can be affected by non-material handling factors, such as department utilization levels, variability in department processing times and variability in product demands. We also identified instances where the QAP and WIP-based formulation are WIP-equivalent. This includes systems with flow/dimensional symmetry or systems with low material handling system utilization.

Several avenues for future research are possible. In this paper, the objective function was to minimize overall WIP in the system. In many applications, it is useful to differentiate between WIP at different departments and/or different stages of the production process. In fact, in most applications, the value of WIP tends to appreciate as more work is completed and more value is added to the product. Therefore, it is useful to assign different holding costs for WIP at different stages. This would lead to choosing layouts that reduce the most expensive WIP first (e.g., letting departments that participate in the last production steps be as centrally located as possible).

In addition to affecting WIP, the choice of layout determines production capacity. From the stability condition, \( \rho_t < 1 \), we can obtain the maximum feasible throughput rate:

\[
\lambda_{\text{max}}(x) = 1 / \sum_{i} \sum_{j} \sum_{k} \sum_{l} \sum_{s} p_{rij} (x_{rk} x_{il} x_{js} (d_{kl} + d_{ks}) / v),
\]  

Maximizing throughput by maximizing \( \lambda_{\text{max}} \) could be used as an alternative layout design criterion. In this case, layouts would be chosen so that the available material handling capacity is
maximized (i.e., $\rho_t$ is minimized). The stability condition can also be used to determine the minimum required number of material handling devices, $n_{min}$, for a given material handling workload, $\lambda_t$:

$$n_{min} = \lambda_t \sum_i \sum_j \sum_{l} \sum_{s} p_{rij} (x_{rk} x_{id} x_{js} (d_{kl} + d_{ls}) / v).$$

(36)

Minimizing $n_{min}$ can be used as yet another criterion in layout design. More generally, our modeling framework offers the possibility of integrating layout design with the design of the material handling system. For example, we could simultaneously decide on material handling capacity, such as number or speed of material handling devices, and department placement, with the objective of minimizing both WIP holding cost and capital investment costs.

**Acknowledgment:** I would like to thank Wally Hopp, the Associate Editor and two anonymous reviewers for many useful comments on an earlier version of the paper. I am grateful to Te Yang for help in generating computer code for the analytical models and carrying out the simulation experiments. This research is supported by the National Science Foundation under grant No. DMII-9908437 and the U.S. Department of Transportation under grant No. USDOT/DTRS93-G-0017.
Appendix 1: The Case of Multiple Transporters

For a system with multiple transporters, the travel time distribution is affected by the dispatching policy used to select a transporter whenever two or more are available to carry out the current material handling request. Analysis of most dispatching policies is difficult. In this appendix, we treat the mathematically tractable case of randomly selecting a device when two or more are idle. Although not optimal, this policy does yield a balanced workload allocation among the different devices. Assuming transfer requests are processed on a first come-first served basis, this policy also ensures an assignment of transporters to departments proportional to the departments’ workloads. As in the single transporter case, we assume that vehicles remain at the location at of their last delivery if there are no pending requests.

In order to characterize the probability distribution of travel time in a system with \( n_t \) transporters \((n_t > 1)\), we need to first obtain the probability \( p_{rij} \) of an empty trip from department \( r \) followed by a full trip from department \( i \) to \( j \). The probability of a full trip from \( i \) to \( j \) is still given by (27). The probability of an empty trip from \( r \) can be written as follows:

\[
Prob(\text{empty trip from } r) = \sum_{n_r=1}^{n_t} \sum_{n_s=1}^{n_t} \{Prob(\text{selecting transporter at } r \mid n_r \text{ and } n_s) \times \frac{Prob(n_r \mid n_s)Prob(n_s)}{Prob(selecting transporter at } r \mid n_r \text{ and } n_s)}
\]

where \( Prob(\text{selecting transporter at } r \mid n_r \text{ and } n_s) \) refers to the probability of selecting one of the idle vehicles at department \( r \) given that there are \( n_r \) idle vehicles at \( r \) and \( n_s \) total idle vehicles in the system, \( Prob(n_r \mid n_s) \) is the probability of having \( n_r \) idle vehicles at department \( r \) \((n_r = 1, 2, \ldots, n_t)\) given that there are \( n_s \) idle vehicles in the system, and \( Prob(n_s) \) is the probability of having \( n_s \) idle vehicles in the system \((n_s = 1, 2, \ldots, n_t)\). It is not too difficult to show that

\[
Prob(\text{selecting transporter at } r \mid n_r \text{ and } n_s) = n_r / n_s, \quad \text{and}
\]

\[
Prob(n_r \mid n_s) = \left( \frac{n_r}{n_s} \right) p_{r}^{n_r} (1 - p_{r}^{n_r}),
\]

where \( p_{r} \) is the probability of an idle vehicle being at department \( r \) which is given by:

\[
p_{r} = \sum_{i=0}^{M} \lambda_{ij} = \sum_{i=0}^{M} \lambda_{ij} / \sum_{i=0}^{M+1} \sum_{j=1}^{M} \lambda_{ij}.
\]

We can now write the probability \( p_{rij} \) as
\[ p_{ij} = \left\{ \sum_{n_r} \sum_{n_s} \binom{n_r}{n_s} p_r^{n_r} (1 - p_r^{n_r}) \text{Prob}(n_s) \right\} p_{ij} \]  \hspace{1cm} (A.5)

or equivalently as

\[ p_{ij} = \left\{ \sum_{n_r} \left(1/n_s\right) \text{Prob}(n_s) \sum_{n_r} \binom{n_r}{n_s} p_r^{n_r} (1 - p_r^{n_r}) \right\} p_{ij} . \]  \hspace{1cm} (A.6)

Noting that

\[ \sum_{n_r} \binom{n_r}{n_s} p_r^{n_r} (1 - p_r^{n_r}) = n_s p_r, \]  \hspace{1cm} (A.7)

yields to

\[ p_{ij} = \left\{ \sum_{n_r} \text{Prob}(n_s) p_r \right\} p_{ij} = p_r p_{ij} , \]  \hspace{1cm} (A.8)

which is the same as in the single transporter case (a result due to the random nature of the selection rule). The mean and variance of travel time can now obtained as in (24) and (26). Expected WIP due to the transporters can be obtained using approximations for a GI/G/n queue (see for example Whitt [21]). Similarly, the departure process from the transporters can be approximated as a departure process from a GI/G/n queue (see Whitt [21] and Buzacott and Shanthkumar [3]), which can then be used to characterize the arrival process to the departments and the transporters as in (25)-(26). A detailed analysis and software implementation of this approach can be found in Yang and Benjaafar [24]. The software is available from the authors upon request.
References


