PHILIPPINE ENGINEERING JOURNAL
PEJ 2009; Vol. 30, No. 1: 1-16

DEVELOPING THE BOUNDED INVENTORY LEVEL POLICY AS INVENTORY CONTROL MECHANISM FOR REENTRANT LINES

Rhey David S. Daway and Aura C. Matias, Ph.D.
Department of Industrial Engineering and Operations Research
University of the Philippines, Diliman

ABSTRACT

Manufacturing systems involving reentrant lines have become quite common especially in the semiconductor industry. A reentrant line can be thought of as a queueing network wherein work-in-process materials (WIP) visit one or more workstations, or service facilities, several times before eventually leaving the system. Now the complexity of the reentrant-line problem lies in the difficulty of determining the scheduling mechanism which results in the optimization of system performance.

Although several studies regarding the subject matter have already been conducted in the past, none were able to formulate a general model for every class of reentrant-line systems. As such, the study proposes an inventory control policy, called the Bounded Inventory Level Policy (BILP), which attempts to strike a balance between the frequency of setups and the variability in the internal flow processes for the purpose of reducing the long-run average total holding cost per unit output. It was initially found via mathematical analysis that there are instances wherein BILP is superior to existing inventory control policies. Therefore, the further development and usage of BILP were justified.

Keywords : Bounded Inventory Level Policy, Reentrant Lines, Scheduling Policies

1. INTRODUCTION

1.1 Background and Rationale of the Study

Manufacturing systems involving reentrant lines have become quite common especially in the semiconductor industry. A reentrant line can be thought of as a queueing network wherein work-in-process (WIP) materials visit one or more workstations, or service facilities, several times before eventually leaving the system. A WIP arriving at a service facility in a reentrant line may thus be grouped together into individual queues, or buffers, according to what production stage it is currently at. Now with the different buffers waiting to be processed at certain service facilities, and considering the fact that most of the service facilities have limited or finite capacities, the complexity of the reentrant-line problem lies in the difficulty of determining the scheduling mechanism which results in the optimization of system performance.

Although several studies regarding the subject matter have already been conducted in the past, the intricate structure of the problem somehow limits the researchers from being able to formulate a general model for every class of reentrant-line systems. The reentrant-line problem does not merely require the computation of the optimal levels of some real-valued decision variables, but rather involves the

Correspondence to: *Aura C. Matias, Ph.D., Department of Industrial Engineering and Operations Research. University of the Philippines, Diliman*

Received: June 17, 2010
Accepted: January 10, 2012

determination of the best scheduling mechanism from a multitude of possible decision rules. Hence, it is believed in this study that opportunities still exist for improving the current body of knowledge as it is deemed infeasible to enumerate all the scheduling policies for every reentrant-line configuration. As such, the study proposes an inventory control policy, called the Bounded Inventory Level Policy (BILP), which attempts to strike a balance between the frequency of setups and the variability in the internal flow processes for the purpose of reducing the long-run average total holding cost per unit output. BILP, along with the Last Buffer First Serve (LBFS) sequencing rule, constitutes the scheduling policy under study. It was initially found that there are instances wherein BILP is superior to existing inventory control policies, which therefore justifies its further development and usage.

1.2 Objective of the Study

The main objective of this article is to introduce BILP and present instances in which the proposed policy is superior to existing ones. The study should then serve as an initial step for promoting the further development and usage of BILP.

1.3 Scope and Limitations

The study of the proposed scheduling policy was limited to those reentrant-line systems characterized by having the following features: a single process technology; service facilities, each with a single-capacity server; a Markovian arrival processes; independent and identically-distributed service times and setup times; infinite buffers; a perfect process yield; and negligible machine failures.

1.4 Areas of Application

The scheduling policy developed in this study is applicable to reentrant-line systems which fit the description in the previous section. Such can be found in semiconductor manufacturing plants, particularly those of wafer fabrication where wafer lots are routed to the same photolithographic expose station several times before exiting from the system. However, the proposed scheduling policy may also be applicable to similarly-situated manufacturing systems, such as low-volume, high-mix production systems involving machining centers (i.e., advanced numerical control machines having multiple processing capabilities).

2. RESEARCH METHODOLOGY

The research methodology used in this study was carried out in a manner depicted in the following flowchart:

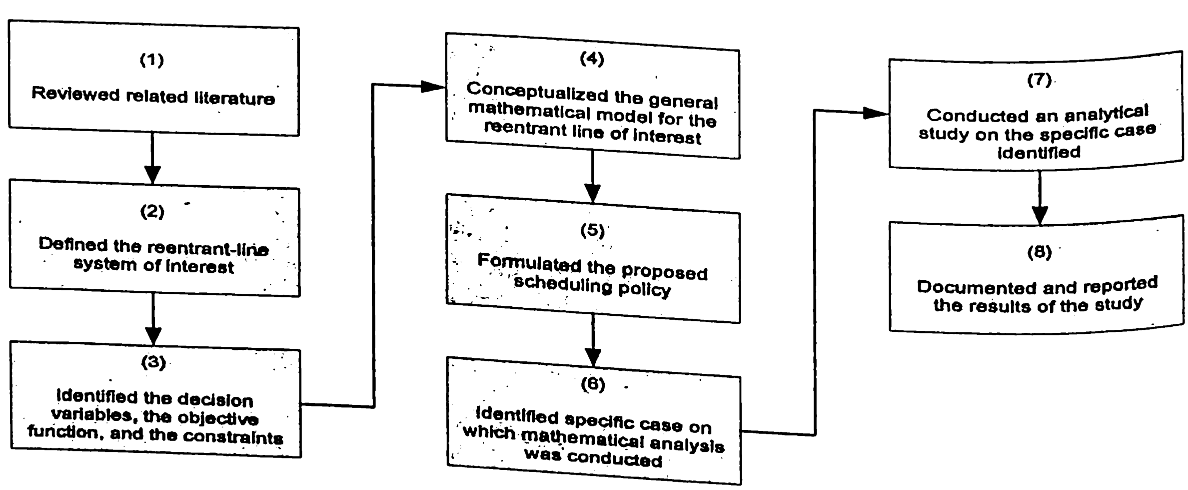

Figure 1. The Research Methodology Flowchart

3. REVIEW OF RELATED LITERATURE

Many researchers have already suggested policies for optimizing the performance of reentrant-line systems. The scheduling policies that were studied are basically in the fields of inventory control and sequencing. Inventory control policies answer the questions as to when and how much WIP are going to be processed during the production runs of each service facility. On the other hand, a sequencing rule refers to the collection of decision rules which choose what set of waiting jobs, or buffers, should be served first in each service facility once capacity becomes available. And so, it can be said that a scheduling policy is a combination of an inventory control policy and a sequencing rule.

Many of the past studies focused on certain subclasses of reentrant-line systems and then resorted to heuristics and numerical methods in looking for the scheduling policy which yields the best system performance. Even so, such studies were only successful in ensuring that the best policy was chosen among those which they enumerated. And so unless a general model that could be used to exhaust and evaluate every possible policy is formulated, it is unlikely for any study to be able to come up with the optimal solution to the problem. It is thus not surprising to see that every study reviewed in this article has its own limitations.

The studies by Wein[14] and by Lu et al.[10] focused on minimizing the mean and the variance of the production cycle times of certain fictitious wafer fabs. Although their models were incorporated with certain realistic features, such as the stochastic treatment of the arrival and the departure processes, some important aspects of real-world systems were neglected. In particular, the setup times were deemed as part of the processing times. Note that the amount of setup times incurred should be driven by the number of switchovers that took place during operations. However, having the setup times as part of the processing times would cause the study to neglect the effects of having frequent switchovers in the total amount of time that a WIP stays in the system. As such, this could have resulted in an inaccurate estimation of the true mean production cycle times, which may have led to an unreliable evaluation of scheduling policies.

On the other hand, the studies by Perkins and Kumar[12] and by Martinez et al.[11] were able to incorporate the separate effects of setups in their models. The two studies' main contributions are on the usage of their own sequencing rules in the optimization of their desired system performances. Perkins and Kumar came up with the CAF policies which aim to minimize the average weighted buffer level whereas Martinez et al. focused on minimizing the average total holding cost (per unit time) through critical buffer levels. Both studies somehow considered the negative impact of having excessive switchovers in their deterministic mathematical models; in which case, they came up with scheduling mechanisms employing the exhaustive inventory control policy. Note, however, that having the exhaustive policy for inventory control may even lead to a higher holding cost than when the production runs are limited. Although limiting the production runs may increase the frequency of setups, the resulting decrease in the variance of the material flow inside the system could reduce the average total holding cost of operating reentrant lines, as would be shown in this article.

The remaining studies that were reviewed in this article actually employed inventory control mechanisms aside from the exhaustive policy. The capacitated base stock policy by Bispo and Tayur[1] and the heuristic policy by Duenyas et al.[7] were shown in their respective works to be effective in reducing the holding cost. Nevertheless, their studies were only limited to specific configurations of reentrant-line systems: the simulation studies conducted by these authors were mainly centered on the single-server setting. (In addition, Bispo and Tayur considered the setup times to be negligible.) And so the analytical and numerical methods on which their optimal solutions were based may only be applicable to the reentrant lines of their interest. Further studies should be conducted in order to determine the applicability of their policies to reentrant-line systems with general process technologies.

A summary of some of the previous works related to reentrant lines is presented in Table 1.

Table 1. A Summary of Existing Literature Relating to Reentrant Lines

Policy	Author(s)	Description	Perfromance Measures	Limitations
Workload Regulating Input Policy (WR)	Wein (1988)	WR is an input control policy which only allows the release of materials into the reentrant-line system whenever the expected amount of work for the bottleneck workstation drops to some pre-specified level. It can be used with any sequencing rule.	(Minimize) Mean Production Cycle Times	The study conducted was limited to reentrant lines having: o a single process technology; o service facilities with several parallel single-capacity machines; o a general arrival process; o generally distributed service times; o generally distributed setup times which are already incorporated in the service times; o infinite buffers; o imperfect process yields which are already incorporated in the service times as rework; and o random machine failures.
Clear-A-Fraction Policy (CAF)	Perkins and Kumar (1989)	The CAF policy is a sequencing rule which prioritizes the buffers having current inventory levels that are at least some fraction of the total contents of all the buffers prior to the service facility of interest. It employs the exhaustive or clearing policy for inventory control.	(Minimize) Average Weighted Buffer Level	The study conducted was limited to reentrant lines having: o multiple process technologies; o service facilities, each with a single-capacity machine; o a deterministic arrival process; o deterministic service times; o deterministic setup or switchover times; o infinite buffers; o a perfect process yield at all production stages; and o negligible machine failures.
Fluctuation Smoothing Policies	Lu, Ramaswamy, and Kumar (1994)	A class of "Least Slack Policies" which prioritizes WIP in the buffer having the minimum "slack" value (i.e., the difference between the due date of the product and its estimated remaining production time). It is highly recommended that this sequencing rule be used with Wein's WR policy.	(Minimize) Mean and Variance of Production Cycle Times	The study conducted was limited to reentrant lines having: o a single process technology; o service facilities with several parallel single-capacity machines; o a Bursty arrival process; o generally distributed service times; o generally distributed setup times which are already incorporated in the service times; o infinite buffers; o imperfect process yields which are already incorporated in the service times as rework; and o random machine failures.
	Sohl and Kumar (1995)			The study conducted was limited to reentrant lines having the same characteristics as those initially studied by Lu, Ramaswamy and Kumar. However, the policy is extended to reentrant lines having multiple process technologies.
A Heuristic Policy For Single-Server Tandem Queueing Systems With Setups	Duenyas, Gupta and Olsen (1995)	A heuristic policy for single-server tandem queueing systems with setups, which controls the timing of the switchovers according to some batch splitting procedure. This policy prioritizes the buffer holding the WIP at the latest production stage.	(Minimize) Average Total Holding Cost	The study conducted was limited to tandem queueing systems having: o a single process technology; o a single-capacity server; o Markovian arrival process; o generally distributed service times; o generally distributed setup or switchover times; o infinite buffers; o a perfect process yield at all production stages; and o negligible machine failures. The policy also assumes that the unit holding cost rates increase as the products increase in value.
A Capacitated Multi-Echelon Base Stock Policy	Bispo and Tayur (2000)	A "base stock policy" wherein the amount of materials to be processed within each discrete period is determined for each buffer based on the difference between the assigned base stock value and the current echelon inventory level. The amount of available capacity is also considered in production decisions.	(Minimize) Infinite Horizon Holding Cost and (Maximize) Service Level	The study conducted was limited to cyclic reentrant flow lines having: o multiple process technologies; o service facilities, each with a multi-capacity machine; o general identically and independently distributed product demand; o deterministic service times which are the same for all products at each level and stage; o negligible setup or switchover times; o infinite buffers; o a perfect process yield at all production stages; and o random machine failures.
A Scheduling Policy Involving Critical Buffer Levels	Martinez, Balanggao, Guan Hing, and Maula (2007)	A scheduling policy that uses a sequencing rule which prioritizes the buffer with the largest difference between its current inventory level and an assigned parameter, called the "critical buffer level". It employs the exhaustive or clearing policy for inventory control.	(Minimize) Average Holding Cost per Unit Time	The study conducted was limited to reentrant lines having: o a single process technology; o service facilities, each with a single-capacity machine; o a general arrival process; o generally distributed service times; o deterministic setup or switchover times; o infinite buffers; o a perfect process yield at all production stages; and o negligible machine failures.

As shown in this review of related literature, research gaps still exist in the area of reentrant lines. And so, a scheduling policy was developed for the class of reentrant-line systems previously described.

4. THEORETICAL FRAMEWORK

4.1 Introduction to the General Model

Consider a queueing system composed of M service facilities (indexed from 1 to M), each consisting of a single-capacity server. All materials entering the said system are assumed to follow the same sequence of operations, which involves paths directing the WIP to visit certain service facilities several times for various processing requirements. Such is the characteristic of a reentrant-line system with a single process technology (refer to Fig. 2 for illustration).

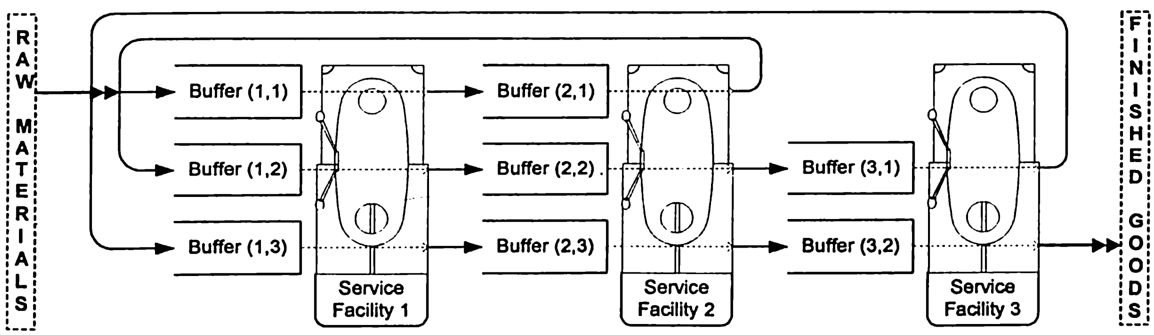

Figure 2. A Reentrant Line Having an Arbitrary Single Process Technology

As a consequence of having such feature, a reentrant line would have some of its service facilities responsible for serving different buffers. Let b be the reference of the b^{th} buffer prior to service facility m, for $m = 1$ to M. Also, let B_m be the total number of buffers prior to service facility m, for $m = 1$ to M. And so, each buffer in a reentrant line could be indexed by the ordered double (m, b) and be denoted as $B_{m,b}^{*}$, for $m = 1$ to M and $b = 1$ to B_m.

4.2 General Model Parameters

The parameters of the model, which have to be defined first in order to fully describe the reentrant line of interest, were listed as follows:

λ	=	the Poisson arrival rate of the input materials coming from the outside of the reentrant-line system
H_n	=	a random variable which represents the processing time of a WIP at the n^{th} production stage
$H_n(\cdot)$	=	the cumulative distribution function of H_n
G_m	=	a random variable which represents the setup time at service facility m
$G_m(\cdot)$	=	the cumulative distribution function of G_m
c_n	=	the unit holding cost rate incurred by the WIP at the n^{th} production stage

In this study, it was assumed that the input materials enter the system according to a Markovian process. With the additional premise that the interarrival times are identically distributed, the resultant arrival process would be a Poisson process having a rate of λ. Having a Poisson arrival process reflects the scenario wherein the input materials are introduced into the system according to the exogenous random occurrence of customer demand.

The processing times of the WIP at the n^{th} production stage (H_n), for $n = 1$ to N, were assumed to be independent and identically distributed continuous nonnegative random variables. This implies that the distribution of the processing times at each service facility is solely dependent on the nature of the WIP held at the buffer that is currently being served. Since the processing times were thought to be generally distributed, the additional parameters which need to be determined for H_n may vary depending on the distribution function, $H_n(\cdot)$.

Aside from the processing times, the other sources of uncertainties are the setup times. The setup times at service facility m (G_m), which are incurred whenever a switchover takes place, were also assumed to be independent and identically distributed continuous nonnegative random variables. However, the distribution of the setup times at each service facility was considered to be independent of the buffers that are being switched. The setup times were also deemed as generally distributed; and thus, the additional parameters required to fully characterize G_m may also vary with the distribution function, $G_m(\cdot)$.

Finally, the unit holding cost rates incurred by the WIP at each of the N production stages were assumed to be constant with respect to the length of time that a material stays in each of the buffers. In addition, it was also assumed that the c_n's, for $n = 1$ to N, are non-decreasing with respect to n, which means that each material does not get less expensive as it gains more value while advancing through the downstream production stages.

4.3 Formulation and Application of Decision Rules

The Inventory Control Policy. The proposed inventory control policy, which is hereby called the Bounded Inventory Level Policy (BILP), imposes an upper bound on the inventory level at each of the $N-1$ buffers in a reentrant line. Since the demand arrival, and consequently the entry of the input materials, is assumed to be an exogenous process, no attempt shall be made to impose bounds on the inventory level at the first buffer.

Under BILP, the reentrant line shall operate in a manner wherein production decisions for each service facility are made and carried out within consecutive intervals of time, called "production epochs". The start of a production epoch of service facility m should coincide with the point in time wherein the immediately preceding one has just ended; in which case, service facility m would become available. Now depending on the sequencing rule applied, a decision shall be made on which of the B_m buffers prior to service facility m should be served next. Subsequently, the necessary setups for serving the "current buffer" shall be performed, after which the actual processing would commence. The served materials from the current buffer would then be routed to the corresponding "immediate downstream buffer". (That is, the materials that are completely served during a production epoch shall be sent to the buffer carrying the WIP at the $(n+1)^{st}$ production stage if the current buffer happens to carry those at the n^{th} production stage.)

The amount of WIP served during the production epochs of service facility m shall depend upon the state of two buffers—the current buffer and its immediate downstream buffer. (It seems to be worth mentioning that the length of each production run under the heuristic policy of Duenyas et al.[7] only depends on the state of the current buffer.) Under BILP, materials from the current buffer shall be served in succession until either the "upper inventory level" of the immediate downstream buffer is reached or the current buffer is cleared of its contents, whichever comes first. Thus, the decision as to when the current production epoch is ended would be made every time a WIP departs from service facility m. Once either of the two said events takes place, the current production epoch of service facility m would end, and then a new one would start.

Of course, it shall be shown in this article that although bounding the amount of materials served during each production epoch entails an increase in the rate at which setups are performed, the ensuing decrease in the variability in the internal flow processes (i.e., the arrival and the departure processes in the point of view of each buffer) could actually cause the overall reduction in the expected waiting time of each WIP as long as the appropriate set of bounds are chosen.

To summarize the steps involved within each production epoch, a flowchart is presented in the following figure.

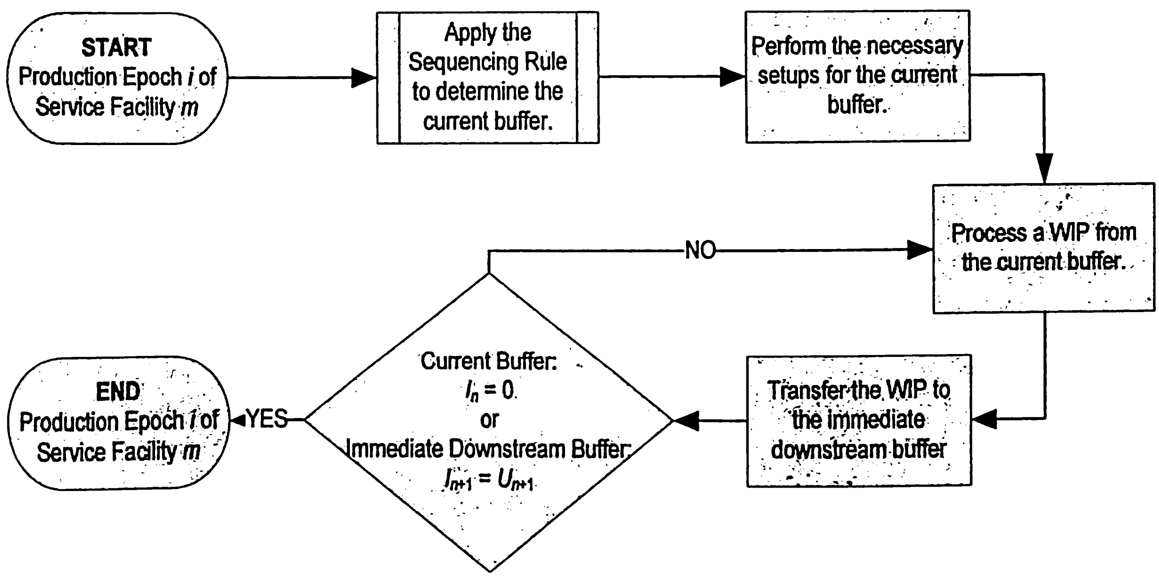

Figure 3. The BILP Inventory Control Process

In Fig. 3, n ($n + 1$) refers to the operation number of the current buffer (immediate downstream buffer). Thus, I_n (I_{n+1}) denotes the inventory level at the current buffer (immediate downstream buffer) whereas U_{n+1} is the upper inventory level assigned to the immediate downstream buffer.

The Sequencing Rule. Since the unit holding cost rates are further assumed to be non-decreasing with respect to the operation number, the buffer that would be processed next upon the completion of a production epoch is the non-empty one carrying the WIP at the latest production stage. In this way, the costliest materials prior to each service facility shall be forced to leave the system first. And so, the sequencing rule that shall be used along with BILP is a priority rule called the Last Buffer First Serve (LBFS) policy, which chooses the buffer having the highest operation number.

The Policy Parameters. Let **U** be a vector with $N - 1$ components. If **U** = [U_2 U_3 … U_N], then it could be used to represent the collection of upper bounds that would be sequentially imposed on the buffers carrying WIP from the 2nd up to the N^{th} production stage. And so, **U** is the decision vector that needs to be determined to be able to implement the proposed policy.

The Decision Criterion. It was implied earlier that the performance measure used in the study is the long-run average total holding cost per unit output. If C is defined as the random variable representing such performance measure, then it could be mathematically defined as

$$C = \sum_{n=1}^{N} c_n W_n$$

where W_n is a random variable which corresponds to the amount of time that a WIP stays at the n^{th} production stage. Since C just happens to be a linear combination of the W_n's,

$$E[C] = \sum_{n=1}^{N} c_n E[W_n] \quad (1)$$

And so, the decision criterion, which should be used in evaluating the scheduling policies of interest in this study, is that of minimizing $E[C]$.

4.4 The Conceptual Framework

Incorporating the model parameters and the decision rules introduced and discussed above into the sample reentrant-line system shown in Fig. 2, the conceptual framework of the study was established as follows.

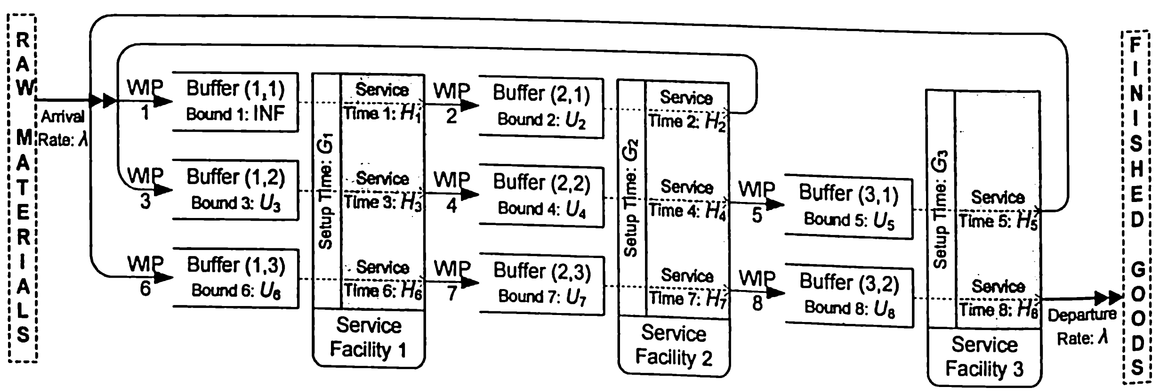

Figure 4. The Conceptual Framework

5. AN ANALYTICAL STUDY ON SPECIFIC CASES OF REENTRANT-LINE SYSTEMS

In order to demonstrate the performance improvements that could be achieved by implementing BILP, along with LBFS, mathematical analysis was conducted on a specific case of reentrant lines, namely the SRL. By showing that BILP could be superior to the traditional exhaustive policy, at least in the simple scenario identified, its further development and usage can be justified.

5.1 Introduction to the Specific Case

Consider a queueing system composed of one service facility, consisting of a single-capacity server. All raw materials entering the system are required to visit the service facility twice before eventually leaving as finished goods. Such is the characteristics of the "two-stage, single-server tandem queuing system with setups", which is illustrated in the following figure. (Notice that the system just described is a particular instance of those studied by Duenyas et al.[7])

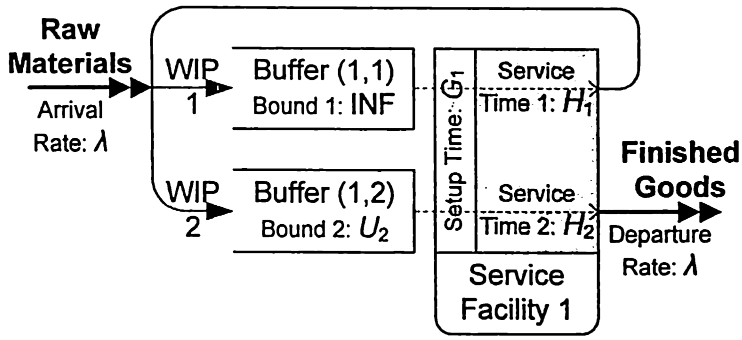

Figure 5. A Two-Stage, Single-Server Tandem Queuing System with Setups

It should be clear that the above system, which is hereby called the "SRL" for brevity, would still fall under the classification of reentrant lines that is of interest to this study.

5.2 Specific Model Parameters

In addition to what were defined in the general model, the input parameters specific to the SRL model were summarized in the following:

ρ_n = $\lambda E[H_n]$; the traffic intensity at $B_{1,n}^{\bullet}$, for $n = 1, 2$

ρ = $\rho_1 + \rho_2$; the service facility utilization

$G_1, G_1^{'}$ = independent and identically-distributed random variables, which represent the setup times incurred whenever the service facility switches from $B_{1,1}^{\bullet}$ to $B_{1,2}^{\bullet}$, and $B_{1,2}^{\bullet}$ to $B_{1,1}^{\bullet}$, respectively

G = $G_1 + G_1^{'}$

ρ_G = $\lambda E[G]$

U = U_2; the upper inventory level that is imposed on $B_{1,2}^{\bullet}$ whenever the SRL is subjected to BILP

5.3 Formulation of an Analytical Model for Performance

Since the operations of the SRL involve two stages of production, the resulting expression for $E[C]$ obtained from Eq. (1) would be

$$E[C] = c_1 E[W_1] + c_2 E[W_2] \quad (2)$$

Of course, $E[C]$ would only be finite if the necessary condition for stability (i.e., $\rho < 1$) is met.

Under BILP, it was found that the performance of the SRL varies with U. To be able to express $E[C]$ as a function of U, the random variables, C, W_1 and W_2, were respectively replaced with C^U, W_1^U, and W_2^U in this analytical study. And so,

$$E[C^U] = c_1 E[W_1^U] + c_2 E[W_2^U] \quad (3)$$

Needless to say, U should be chosen in such a way that $E[C^U]$ is minimized. The optimal value of U, denoted as U^*, should then be determined to maximize whatsoever improvements can be achieved by having BILP for inventory control. Since the exhaustive policy could be thought of as an instance of BILP where $U\to\infty$, it must therefore be shown that at least in some scenarios of the SRL, there exist a finite U such that $E[C^U] < E[C^\infty]$ so as to justify the claim that the latter could be superior to the former.

Eq. (2) and Eq. (3) reveal that the expected waiting time of the materials at each buffer need to be determined first so that the performance of the SRL could be measured. As such, attempts were made to derive expressions for both $E[W_1]$ and $E[W_2]$.

The Expected Waiting at $B_{1,1}^*$. An expression for $E[W_1]$ could be obtained by initially focusing the analysis on $B_{1,1}^*$ alone. In the point of view of $B_{1,1}^*$, the SRL can actually be observed as an "M/G/1 queuing system with generalized vacations", or simply a "vacation system". A vacation system was defined in a paper written by Fuhrmann and Cooper[8] as an M/G/1 queuing system wherein the service facility occasionally becomes unavailable. In the case of the SRL, such events occur whenever the service facility leaves $B_{1,1}^*$ for $B_{1,2}^*$. The intervals of time within which the service facility is unavailable for $B_{1,1}^*$, or is idle, would then be called "vacation periods". In addition, the intervals of time between the start of each successive vacation periods could be defined as "cycles" whereas those within which the service facility is busy processing $B_{1,1}^*$ could be referred to as "sojourn periods".

As a vacation system, the SRL should exhibit the "M/G/1 Decomposition Property", according to Fuhrmann and Cooper[8]. By applying such property, the following expression for $E[W_1^U]$ could be derived:

$$E[W_1^U]=\frac{(\rho_2 U+\rho_G)E[Z]}{\lambda(1-\rho_1)E[K]}+\frac{\lambda^2\left(E[H_1^2]+E[H_2^2]\right)+2\rho_2\rho_G}{2\lambda(1-\rho_1)}+\frac{\rho_2^2\left(E[K^2]-E[K]\right)}{2\lambda(1-\rho_1)E[K]}+\frac{\lambda^2E[G^2]}{2\lambda(1-\rho_1)E[K]} \tag{4}$$

where Z is the amount of materials present in $B_{1,1}^*$ when a random vacation begins while K is the amount of materials served at $B_{1,1}^*$ during a random cycle. Also, the performance of the SRL under the exhaustive policy could be obtained from Eq. (4) by letting U approach infinity to arrive at

$$E[W_1^\infty]=\frac{\lambda^2\left(E[H_1^2]+E[H_2^2]\right)(1-\rho_1)+2\rho_1\rho_2^2}{2\lambda\left[(1-\rho_1)^2-\rho_2^2\right]}+\frac{2(1-\rho_1)\rho_2\rho_G+\lambda^2E[G^2]E^{-1}[K^\infty](1-\rho_1)}{2\lambda\left[(1-\rho_1)^2-\rho_2^2\right]} \tag{5}$$

where K^∞ is used instead of K to emphasize that the above relationship is only valid for the SRL when it is operating under the exhaustive policy.

The Expected Waiting at $B_{1,2}^*$. A general expression for the expected waiting time at $B_{1,2}^*$ (i.e., $E[W_2]$) could be derived by considering the fact that the materials, which are served at $B_{1,1}^*$ during the i^{th} sojourn period, are also the ones which arrive at $B_{1,2}^*$ within the $(i+1)^{st}$ vacation. Following this reasoning, it can be shown[7] that

$$E[W_2^U] = E[G_1] + \frac{\rho\left(E[K^2] - E[K]\right)}{2\lambda E[K]} \quad (6)$$

and

$$E[W_2^\infty] = E[G_1] + \frac{\lambda^2\left(E[H_1^2] + E[H_2^2]\right)\rho + 2\rho\rho_1(1-\rho_1)}{2\lambda\left[(1-\rho_1)^2 - \rho_2^2\right]} + \frac{2\rho\rho_2\rho_G + \lambda^2 E[G^2]\rho E^{-1}[K^\infty]}{2\lambda\left[(1-\rho_1)^2 - \rho_2^2\right]} \quad (7)$$

5.4 Evaluation of Alternative Policies

The scheduling policies evaluated in this study are basically those involving BILP and the exhaustive policy. As such, the individual performances of the SRL under these two inventory control policies must be determined for the purpose of comparison.

By substituting Eq. (4) and Eq. (6) into Eq. (3), the following expression was obtained for $E[C^U]$:

$$E[C^U] = c_1\left[\frac{(\rho_2 U + \rho_G)E[Z]}{\lambda(1-\rho_1)E[K]} + \frac{\lambda^2\left(E[H_1^2] + E[H_2^2]\right) + 2\rho_2\rho_G}{2\lambda(1-\rho_1)} + \frac{\rho_2^2\left(E[K^2] - E[K]\right)}{2\lambda(1-\rho_1)E[K]} + \frac{\lambda^2 E[G^2]}{2\lambda(1-\rho_1)E[K]}\right] + c_2\left[E[G_1] + \frac{\rho\left(E[K^2] - E[K]\right)}{2\lambda E[K]}\right] \quad (8)$$

Similarly, $E[C^\infty]$ was obtained by substituting Eq. (5) and Eq. (7) into Eq. (3) as follows:

$$E[C^\infty] = c_1\left[\frac{\lambda^2\left(E[H_1^2] + E[H_2^2]\right)(1-\rho_1) + 2\rho_1\rho_2^2}{2\lambda\left[(1-\rho_1)^2 - \rho_2^2\right]} + \frac{2(1-\rho_1)\rho_2\rho_G + \lambda^2 E[G^2]E^{-1}[K^\infty](1-\rho_1)}{2\lambda\left[(1-\rho_1)^2 - \rho_2^2\right]}\right] + c_2\left[E[G_1] + \frac{\lambda^2\left(E[H_1^2] + E[H_2^2]\right)\rho + 2\rho\rho_1(1-\rho_1)}{2\lambda\left[(1-\rho_1)^2 - \rho_2^2\right]} + \frac{2\rho\rho_2\rho_G + \lambda^2 E[G^2]\rho E^{-1}[K^\infty]}{2\lambda\left[(1-\rho_1)^2 - \rho_2^2\right]}\right] \quad (9)$$

Since the objective of this analytic study is to be able show instances where it would be better to have BILP for inventory control rather than the exhaustive policy, it must be demonstrated that there exist an SRL scenario wherein $E[C^{\infty}] - E[C^{U}] > 0$ for some finite value of U. For simplicity, the case where $U = 1$ was considered. So as long as $1 - \rho - \rho_G < 1$, it can be proven[4] that

$$E[C^1] = c_1\left[\frac{\lambda^2\left(E[H_1^2]+E[H_2^2]\right)+2\rho_1\rho_2+2\rho\rho_G+\lambda^2E[G^2]}{2\lambda(1-\rho-\rho_G)}\right]+c_2E[G_1] \quad (10)$$

Now it was found upon mathematical manipulation that

$$E[C^{\infty}]-E[C^1] = c_1\left(E[W_1^{\infty}]-E[W_1^1]\right)+c_2\left(E[W_2^{\infty}]-E[W_2^1]\right) \quad (11)$$

where

$$\begin{aligned}E[W_1^{\infty}]-E[W_1^1] = &-\frac{\lambda^2\left(E[H_1^2]+E[H_2^2]\right)\left[(1-\rho)\rho_2+(1-\rho_1)\rho_G\right]}{2\lambda\left[(1-\rho_1)^2-\rho_2^2\right](1-\rho-\rho_G)}\\ &-\frac{2(1-\rho)(1-\rho_1)\rho_1\rho_2+2(1-\rho_1)^2\rho_1\rho_G}{2\lambda\left[(1-\rho_1)^2-\rho_2^2\right](1-\rho-\rho_G)}\\ &-\frac{2(1-\rho)\rho_2^2\rho_G+2(1-\rho_1)\rho_2\rho_G^2}{2\lambda\left[(1-\rho_1)^2-\rho_2^2\right](1-\rho-\rho_G)}\\ &-\frac{\lambda^2E[G^2]\left[(1-\rho)(1-\rho_1+\rho_2)E[K^{\infty}]-(1-\rho-\rho_G)(1-\rho_1)\right]}{2\lambda\left[(1-\rho_1)^2-\rho_2^2\right](1-\rho-\rho_G)E[K^{\infty}]}\end{aligned} \quad (12)$$

and

$$\begin{aligned}E[W_2^{\infty}]-E[W_2^1] = &\frac{\lambda^2\left(E[H_1^2]+E[H_2^2]\right)\rho+2\rho\rho_1(1-\rho_1)}{2\lambda\left[(1-\rho_1)^2-\rho_2^2\right]}\\ &+\frac{2\rho\rho_2\rho_G+\lambda^2E[G^2]\rho E^{-1}[K^{\infty}]}{2\lambda\left[(1-\rho_1)^2-\rho_2^2\right]}\end{aligned} \quad (13)$$

It is easy to see that all of the terms on the right-hand side of Eq. (12) are negative, which means that the expected waiting times at $B_{1,1}^{*}$ would increase by setting U to one (1). This is not surprising since limiting the amount of materials served during each production run actually increases the number of vacations taken by the service facility; in which case, the materials at $B_{1,1}^{*}$ would effectively stay longer.

On the other hand, Eq. (13) shows that the contrary would happen to the expected waiting time of the materials at $B_{1,2}^{*}$. It can be observed from Eq. (6) that $E[W_2^{U}]$ is directly proportional to $(E[K^2] - E[K])E^{-1}[K]$. As a matter of fact, both quantities are minimized whenever $U = 1$. To prove this point, it must first be noted that $K \geq 1$ no matter what the value of U is so that $E^2[K]$ would never be less than $E[K]$.

Since $Var(K) = E[K^2] - E^2[K] \geq 0$, it must also be true that $E[K^2] - E[K] \geq 0$. The minimum value of $(E[K^2] - E[K])E^{-1}[K]$ would then be zero (0). Since the first and the second moments of K would only be identical to each other if K is limited to one (1), letting U be equal to one (1) would minimize $E[W_2^U]$, and consequently maximize $E[W_2^\infty] - E[W_2^1]$.

Although subjecting the SRL to BILP with $U = 1$ would increase $E[W_1]$, the resulting decrease in $E[W_2]$ could actually improve the overall system performance. It is therefore theoretically possible to have $E[C^1] < E[C^\infty]$ especially if c_2 happens to be much higher than c_1. However, it can be shown that $E[C^\infty] - E[C^1]$ could still be positive even in the scenario wherein c_1 is equal to c_2. Substituting Eq. (12) and Eq. (13) into Eq. (11) with $c_1 = c_2 = 1$,

$$\begin{aligned} E[C^\infty] - E[C^1] = & \frac{\lambda^2\left(E[H_1^2] + E[H_2^2]\right)\left[(1-\rho)\rho_1 - (1+\rho_2)\rho_G\right]}{2\lambda\left[(1-\rho_1)^2 - \rho_2^2\right]\left(1-\rho-\rho_G\right)} \\ & + \frac{2\rho_1(1-\rho_1)\left[(1-\rho)\rho_1 - (1+\rho_2)\rho_G\right]}{2\lambda\left[(1-\rho_1)^2 - \rho_2^2\right]\left(1-\rho-\rho_G\right)} \\ & + \frac{2\rho_2\rho_G\left[(1-\rho)\rho_1 - (1+\rho_2)\rho_G\right]}{2\lambda\left[(1-\rho_1)^2 - \rho_2^2\right]\left(1-\rho-\rho_G\right)} \\ & - \frac{\lambda^2 E[G^2]\left[(1-\rho)(1-\rho_1+\rho_2)E[K^\infty] - (1-\rho-\rho_G)(1+\rho_2)\right]}{2\lambda\left[(1-\rho_1)^2 - \rho_2^2\right]\left(1-\rho-\rho_G\right)E[K^\infty]} \end{aligned} \quad (14)$$

Except for the last one, it can directly be observed that all the terms in Eq. (14) would definitely be positive if $(1-\rho)\rho_1 - (1+\rho_2)\rho_G > 0$. This implies that having a higher ρ_1 and at the same time a lower ρ_2 and a lower ρ_G would increase the difference between the individual performances of the SRL under the exhaustive policy and BILP with $U = 1$ even when $c_1 = c_2$. Still, satisfying the above relationship does not ensure that $E[C^\infty] - E[C^1] > 0$ since the last term in Eq. (14) would most probably be negative. However, having a lower $E[G^2]$ may lessen whatever negative impact the last term has. This reveals that the variance of the setup times is also an important factor in determining what inventory control policy to apply.

Hence, it was shown in this analytical study that there exist cases wherein BILP is superior to the exhaustive policy. In particular, it would be better to use BILP with $U = 1$ for inventory control than the exhaustive policy whenever $E[C^\infty]$ and $E[C^1]$ have a positive difference, which would be larger if: $c_2 - c_1$ is higher; ρ_1 is higher; ρ_2 is lower; ρ_G is lower; and $E[G^2]$ is lower.

The increase in $E[W_1]$ sustained if the production runs are limited is partly due to the ensuing increase in the number of vacations, or switchovers. Nevertheless, increasing ρ_1 while decreasing ρ_2 could reduce the effects of having a higher vacation frequency on the total length of time that the service facility leaves $B_{1,1}^*$ behind. For the same reason, shortening the amount of time spent on each setup would also result in a larger positive difference between $E[C^\infty]$ and $E[C^1]$.

5.5 The Frequency of Setups Versus the Variability in the International Flow Processes

It can actually be shown that the cost savings, which may be realized by implementing BILP, would basically result from the ensuing decrease in the variability in the internal flow processes. To precisely grasp the meaning of such, consider the mathematical models that were formulated in the previous sections.

As can be observed in Eq. (8), $E[C^U]$ is dependent on two quantities involving $E[Z]$, $E[K]$ and $E[K^2]$—$(\rho_2 U + \rho_G)E[Z]E^{-1}[K]$ and $(E[K^2] - E[K])E^{-1}[K]$. Evidently, $E[C^U]$ decreases as $(\rho_2 U + \rho_G)E[Z]E^{-1}[K]$ and $(E[K^2] - E[K])E^{-1}[K]$ simultaneously decrease. Now $E[Z]$ should be expected to increase with the frequency of setups as it is more likely that a large amount of WIP would be left at $B_{1,1}^{\bullet}$ upon a random vacation of the service facility if the switchovers are performed at a higher rate. On the other hand, $(E[K^2] - E[K])E^{-1}[K]$ can be used to directly quantify the variability in the internal flow processes as it could analytically be established that

$$E[IDT_{1,1}^2] = E[IAT_{1,2}^2] = E[H_1^2] + E[H_2^2] + E^2[H_2]\left(\frac{E[K^2] - E[K]}{E[K]}\right) + 2E[H_1]E[H_2] + 2\left(\frac{E[H_1]}{E[K]} + E[H_2]\right)E[G] + \frac{E[G^2]}{E[K]} \quad (15)$$

where $IDT_{1,1}$ $(IAT_{1,2})$ represents the interdeparture (interarrival) times of the materials coming out of $B_{1,1}^{\bullet}$ (going into $B_{1,2}^{\bullet}$)[4]. (Note that both $E[IDT_{1,1}^2]$ and $E[IAT_{1,2}^2]$ are positively related to $(E[K^2] - E[K])E^{-1}[K]$.)

Of course, the functional relationship between the two said quantities and U must be determined in order to see how both the frequency of setups and the variability in the internal flow processes interact to influence the performance of the SRL. It should be understood that the lower the value of U gets, the more often the switchovers would have to be performed. This implies that tightening U would increase the frequency of setups, which would consequently increase $(\rho_2 U + \rho_G)E[Z]E^{-1}[K]$. However, the opposite seems to happen to the term involving $E[K^2]$ as the amount of materials served during each cycle (i.e., K) would be bounded by U. Thus, $(E[K^2] - E[K])E^{-1}[K]$ would decrease as U decreases.

Although implementing BILP would unavoidably increase the frequency of setups, the subsequent decrease in the variability in the internal flow processes could actually improve the overall performance of the system. Hence, a balance should be made between these two competing effects when deciding on what values to assign to each upper inventory level so that the long-run average total holding cost of operating the reentrant line of interest would be minimized with respect to the policy parameters.

6. CONCLUSIONS AND RECOMMENDATIONS

In the interest of minimizing the long-run average total holding cost per unit output, the Bounded Inventory Level Policy (BILP) was conceived as an inventory control mechanism for reentrant lines. And so BILP, along with the LBFS sequencing rule, constitutes the scheduling policy proposed in this study.

It was established through mathematical analysis conducted on a specific reentrant-line setting that BILP, which was found to work by striking a balance between the frequency of setups and the variability in the internal flow processes, could be superior to the exhaustive policy when it comes to minimizing the long-run average total holding cost per unit output. As such, the further development and usage of BILP were justified.

In the light of the above results of the study, the following recommendations are thus proposed:

- Reentrant lines characterized by having features that were previously described should be operated under the BILP inventory control policy, along with the LBFS sequencing rule, whenever the system performance is measured in terms of the long-run average total holding cost per unit output.
- The optimal level of the decision vector, **U**, must be obtained using simulation optimization software in order to maximize the potential benefits of having BILP for inventory control.
- Especially if a long sequence of operations is involved, BILP should be implemented with the aid of a decision support system that could keep track of the WIP level at every production stage in the reentrant line of interest.

REFERENCES

[1] Bispo, C. F., and S. Tayur. Managing Simple Re-entrant Flow Lines: Theoretical Foundation and Experimental Results. Pittsburgh, 2000.
[2] Cooper, R. B. Introduction to Queueing Theory. Second Edition. New York: North Holland, 1981.
[3] Daway, R. D. S. "A Proposed Cost-Driven, Multi-Stage, Scheduling Policy for Reentrant Lines." 3rd ERDT Conference: Post Graduate Multidisciplinary Approach to Solving Philippine Problems. 11 September 2009, Metro Manila. n.p., n.d. 184-189. Print.
[4] Daway, R. D. S. "Developing the Bounded Inventory Level Policy as an Inventory Control Mechanism for Reentrant Lines." MS thesis. University of the Philippines, Diliman, 2010. Print.
[5] Daway, R. D. S., and A. C. Matias. "A Simulation Study on the Performance of a Reentrant Line Operated Under the Bounded Inventory Level Policy." Proceedings of the 3rd ORSP National Conference. 20 November 2009, Metro Manila. n.p., n.d. Print.
[6] Daway, R. D. S., and A. C. Matias. "Developing the Bounded Inventory Level Policy as an Inventory Control Mechanism for Reentrant Lines." 4th National IE Research Conference.12 March 2010, Metro Manila. n.p., n.d. Print.
[7] Duenyas, I., Gupta, D., and T. V. Olsen. "Control of a Single-Server Tandem Queueing System with Setups." Operations Research 46.2 (1998): 218-230.
[8] Fuhrmann, S. W., and R. B. Cooper. "Stochastic Decomposition in the M/G/1 Queue with Generalized Vacations." Operations Research 33.5 (1985): 1117-1129.
[9] Kumar, P. R., and R. M. Richardson. "Scheduling System and Scheduling Method for Reentrant Line Processing." Illinois, 06 January 1998. Free Patents Online. 2004-2007. <http://www.freepatentsonline.com/5706200.html>.
[10] Lu, S. C. H., Ramaswamy, D., and P. R. Kumar. "Efficient Scheduling Policies to Reduce Mean and Variance of Cycle-Time in Semiconductor Manufacturing Plants." IEEE Transactions on Semiconductor Manufacturing 7 (1994): 374-385.

[11] Martinez, Iris Ann G., Balanggao, Dedrick, Guan Hing, Dianne, and Jennifer Maula. The Critical Buffer-Level in Scheduling Re-Entrant Lines: A Proposal for a Dispatching Policy to Minimize Average Holding Costs. Quezon City, 2006.

[12] Perkins, J., and P. R. Kumar. "Stable Distributed Real-Time Scheduling of Flexible Manufacturing / Assembly / Disassembly Systems." IEEE Transactions on Automatic Control 34.2 (1989): 139-148.

[13] Sohl, D. L., and P. R. Kumar. "Fluctuation Smoothing Scheduling Policies for Multiple Process Flow Fabrication Plants." IEEE/CPMT International Electronics Manufacturing Technology Symposium (1995): 190-198.

[14] Wein, L. M. "Scheduling Semiconductor Wafer Fabrication." IEEE Transaction on Semiconductor Manufacturing 1.3 (1988): 115-130.