USING SIMULATION TO TEST THE ROBUSTNESS OF VARIOUS EXISTING PRODUCTION CONTROL POLICIES*

Sheldon Lou
Houmin Yan
Suresh Sethi

Faculty of Management
University of Toronto
246 Bloor Street W.
Toronto, Canada M5S 1V4

Anne Gardel
Prabhat Deosthali

Digital Equipment Corporation
HL02-1/BO7
75 Reed Road
Hudson, MA 01749

ABSTRACT

This paper compares four production control policies in terms of their robustness against random disturbances such as machine failures and demand fluctuations. A simulation model based on a VLSI wafer fabrication facility is used to test the performance of the policies. Several criteria including average total WIP, average backlog, and a cost function that is a combination of the above two are used to measure the performance.

1 INTRODUCTION

Robustness is one of the most important criteria when different production control policies are compared. Although a definition of a robust control policy, that is generally agreed upon, does not exist, we would like to define it as a policy that performs well when environmental conditions or system parameters are undergoing relatively large changes. More specifically, a robust production control policy should meet at least the following two requirements. First, it should perform well in presence of random disturbances, such as machine failures, random rework, demand fluctuations, etc. By good performance, we mean the system output should closely follow the production plan while keeping a low Work-In-Progress (WIP). Note that the production plan may be subject to abrupt changes due to unforeseen factors such as demand variations. When the production plan changes, the robust policy should swiftly adjust the production process so that the discrepancy between the actual output and the planned output can be kept as small as possible. Second, it should not be designed only for a set of fixed parameters. Rather, its performance should remain adequate when system parameters, such as machine capacities while fully operational, machine repair rates, buffer sizes, etc., change. It is this definition of ro-

bustness that forms the basis for our comparisons of various production control policies.

The production systems considered in this paper have the following two features. First is that of *mass production*. That is, they produce a large number of products but a small number of product types. Under the mass production environment, one does not have to keep track of individual parts. Instead, one uses production rates and inventory/backlog levels to measure system performance. The second feature of the production systems considered in this paper is that they have *sufficient average capacity* in relation to average demand. That is, they can meet the production plans in the long run, even though from time to time they may produce less or more than what the production plans require. Therefore, our goal is not to compare the throughput rates resulting from different policies (excess production is not desirable either), but to examine how well they follow the production plans and how well they adjust themselves when the plans change.

We use three different measures of performance in this paper. The first two are average WIP and average backlog levels over time. The third is the weighted average cost with weights associated with average inventory levels at different production stages and with the average backlog level at the last (output) stage. Note that the inventory at a production stage is the inventory in the buffer following that stage plus the parts being processed in that stage. Inventories at different stages are, therefore, the inventories of intermediate goods and of finished products.

The control policies considered in this paper are *WIP Control* (WC), which is essentially a Kanban System (Suzaki 1987), *Uniform-Loading* (UL), *WIP-To-Bottleneck Control* (WB), and *Two-Boundary Control* (TB). The WIP Control policy sets a WIP threshold level for each part type at each stage. If there is only one part type and its WIP in that stage is lower than the threshold level, only then may we load additional parts. The UL policy is a simple open-loop control policy which uniformly loads parts
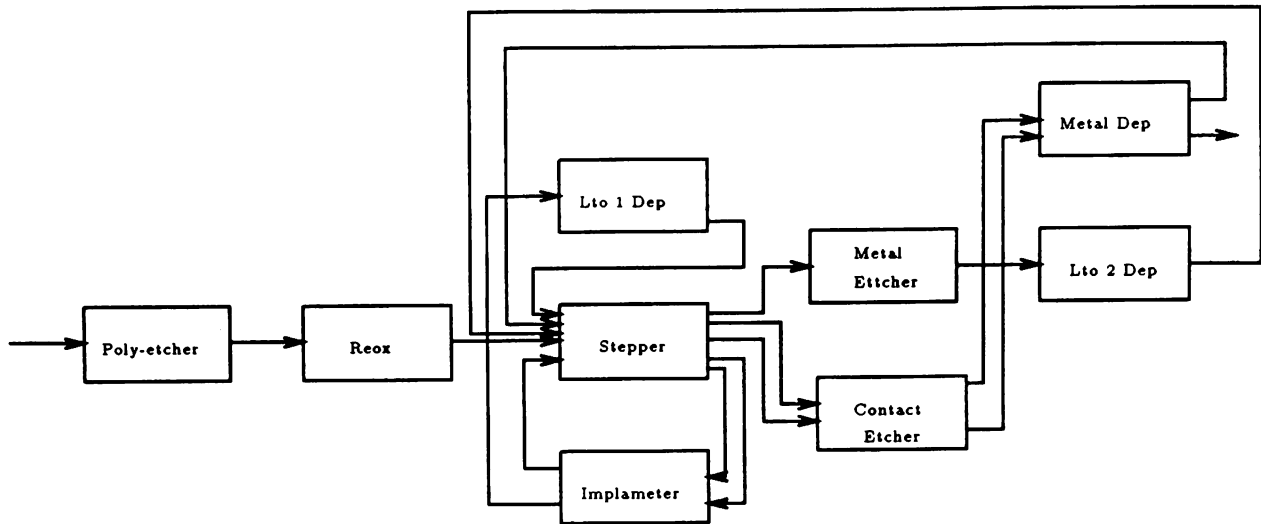
Figure 1: A wafer fabrication model

into the production system at the beginning of each week. The WB policy identifies the bottleneck of the system and sets a threshold level for the total WIP from the first work station to the bottleneck work station. Finally, the TB policy features two threshold levels, called WIP threshold and surplus threshold for each part type at each stage. One controls the WIP level as in the WIP control policy; the other controls the so-called *surplus* level, which is the difference between the actual and the planned productions. Only when both WIP and surplus levels are less than their corresponding thresholds may new parts be loaded.

There is a variety of production systems on which the above policies could be tested. We have selected a simplified model(See Fig. 1) derived from a semiconductor wafer [1] fabrication facility.

It involves unreliable machines, fluctuating demands, unstable yield, and a so-called re-entrant process. In our wafer fabrication model, each wafer needs to be processed in *Stepper*, (photolithography station), a key work station 5 times. Between two consecutive entries into the Stepper, wafers have to be processed in other work stations as well. This process is called the re-entrant process. Owing to the special features of the wafer fabrication, a so-called *hub-centered* approach, discussed in detail in Section 3, will be utilized.

Simulations are conducted to evaluate system performance. Simulation results show that the performance of the TB policy is the most robust of all. In particular, we have the following results:

1. If the demand rate changes abruptly, e.g., a sudden surge or an unexpected reduction in demand occurs, then only the TB policy performs reasonably well. The WC and WB would react slowly to a demand surge, if a set of relatively low threshold values were selected. On the other hand, if a set of relatively high threshold values were used, then while it would follow demand surge promptly, it would also result in a high steady state WIP level.

2. The performance of the TB policy is the most robust of all when system parameters such as machine capacities change.

3. The UL policy results in a periodically fluctuating surplus and WIP levels.

The rest of the paper is organized as follows. We review the related literature in the next section. In Section 3, we specify control policies applied in this study. Simulation results are presented in Section

_____

[1]Wafers are slices of silicon crystals on which semiconductor devices are built.

4. Finally, the paper is concluded in Section 5 with some remarks.

2 REVIEW OF THE LITERATURE

As described in Section 1, we use a wafer fabrication model to test the robustness of various control policies. Wafer fabrication planning and control is a complicated task due to the large number of wafers and machines involved. It is further exacerbated by the presence of a re-entrant process and random disturbances such as machine breakdown and random yields. Recently, several papers have been devoted to the study of production control of such facilities. Burman *et al.* (1986) use a simulation model to estimate WIP, the cycle time, and the production rate. Dayhoff and Atheron (1984) describe the dynamics of wafer movement and identify several problems that can be analyzed by discrete event simulation. Chen *et al.* (1986) present a class of queuing network models for the analysis of wafer fabrication facilities. Wein (1988) uses a Brownian network model to analyze different input control policies. He observes that the job release order plays a more important role in reducing cycle time than do the dispatching rules. Glassey *et al.* (1986) (1988) propose a starvation avoidance policy, which consists of an SRPT(shortest remaining process time) priority rule and an SA(starvation avoidance) rule to have the bottleneck machine produce as much as possible. Lou *et al.* (1989a) propose a shop floor control strategy known as the *two boundary control* policy. In their recent paper (Lou and Kager, 1989b), simulation results show that this strategy obtains a good performance in a shop with 16–job steps. Mozumder *et al.* (1988a) (1988b) describe a computer-aided manufacturing system which performs some of the on-line and off-line controls based on the current information and prior knowledge. The Kanban system concept is also being proposed for use in the semiconductor manufacturing facilities (Cory *et al.* 1986, Martin *et al.* 1989).

3 SPECIFICATION OF POLICIES

This paper employs the *hub-centered control* concept[2] and compare four different control policies in terms of their robustness against random disturbances and parameter changes. Two such disturbances, machine breakdowns and uncertain demands, are modeled in this paper. The demand rate, which is defined as the number of lots produced per shift, is constant for each part type except for abrupt changes from time to time. Machine breakdowns are assumed to

be Poisson and the machine repairs follow an exponential distribution. The rate at which demand changes is much smaller than the rate of machine breakdown and repair[3]. In the 2,000 hour simulation time (roughly 20 working weeks), there occurs only 3 demand rate changes in contrast to individual machines breaking down and getting repaired between 20 to 100 times, on average.

To explain our control policies, let us consider an N-stage production system processing one part type. Denote the WIP level and surplus level at stage i as b_i and x_i respectively.

WC Policy: For stage i, we define a threshold level h_{bi}. If $b_i(t) < h_{bi}$, then we load parts into the stage, provided there are finished parts from the previous stage. If the WIP level at that stage is too high, i.e., $b_i(t) \geq h_{bi}$, then we don't load parts, leaving the machine idle[4]. Note, that h_{bi} can be interpreted as the total number of Kanbans of stage i in a Kanban system.

WB Policy: Suppose that stage j has been identified as the bottleneck stage. Then we compute the total WIP before the bottleneck stage and compare it with a predetermined threshold level h_{WB}, i.e., we load parts into the system if and only if

$$\sum_{i=1}^{j-1} b_i(t) < h_{WB}.$$

TB Policy: For stage i, we compare $b_i(t)$ and $s_i(t)$ with two predetermined thresholds h_{bi} and h_{si}. We load parts into the stage only if $b_i(t) < h_{bi}$ and $s_i(t) < h_{si}$. Intuitively, the relation $s_i \geq h_{si}$ means that our cumulative production at stage i is sufficient, and $b_i(t) \geq h_{bi}$, implies that the inventory level at stage i is too high. In both cases, it is clear suggest that extra production would cause unnecessary WIP.

An actual production system is much more complex than the above-mentioned N-stage system. Multiple products and multiple entries into the hub machine, e.g. stepper, have to be considered. We then use the so-called *Hub-Centered* Control introduced in Lou *et. al.* (1990). The idea is that only the part release (the loading of parts into the first stage of the system) and part dispatch of the hub–the photolithography cell–should be controlled, all other work cells can then be governed simply by FIFO(first-in-first-out) rule.

To resolve the conflicts among parts of different types and different entries, we design a set of

---

[2]The formal development of our model, policies and the approach can be find in Lou, Yan, and Sethi (1991).

[3]See Sethi and Zhang (1991) for a theory of hierarchical control for such manufacturing systems with events occurring at different time scales.

[4]For multi-part-type systems presented in the next section, we switch to some other part type that has a lower WIP level.

weights and compare the weighted differences between the WIP and surplus levels and their corresponding threshold levels.

Remark: The TB policy is motivated by the analysis of a tandem two-machine system with unreliable machines (Ryzin, et al. (1991). The results suggest that the two-boundary control policy is a good approximation to the optimal policy. In the next section, we describe a more complex reentrant multistage production system with two part types. This is the system that we use for testing our policies. It should be noted that even though our system is more complex than those described in this section, the various control policies for it can be formulated as straightforward extension of policies described above for the N-stage serial system.

4 SIMULATION RESULTS

A number of simulations is performed to compare the four control policies described earlier on a simplified wafer fabrication system consisting of nine unreliable workstations. These workstations have an exponentially distributed up and down times defined by the parameters presented in Table 1 below. The numbers are derived from an actual wafer fabrication facility. The capacities, however, are multiples of real capacities for reasons of confidentiality.

Station name	Ave. down time	Ave. up time	Capacity
poly etcher	7	44	3
reox	6	115	5
stepper	3	15	35
implanter	5	15	15
lto 1 dep	21	74	8
contact etcher	10	51	28
metal dep	5	15	18
metal etcher	6	23	12
lto 2 dep	11	77	18

Table 1. Facility parameters

Simulations are performed on a DEC-RISC machine and an event-driven simulator designed for job shop production is used. Simulation runs are all started with an empty fab where machines are up and idle. The time horizon for each simulation run is 2000 time units representing a 20-week period. As stated earlier, each process step is as-

signed a cost coefficient c_i. We use $c_1 = 1.0$ and $c_i = 1.0 + 0.2(i - 1), 1 \le i \le 15$ to reflect the fact that inventory costs are higher at the later stages of the production process. For the last stage ($i = 16$), $c_{16}^+ = 4.0$ and $c_{16}^- = 10.0$ are used for inventory and backlog costs, respectively.

In the rest of this section, we use two groups of simulation experiments to compare the four control policies. In the first group, we plot the average of the 100 trajectories for the total WIP, the summation of the WIP levels of the nine work stations, and the average trajectories for surplus at the last stage. Recall that a negative surplus level at the last stage implies that the cumulative production of the job shop is less than what the production plan requires. A good policy should provide trajectories that have both small positive surplus and WIP levels.

In the second group of simulations, we compute Average Total WIP (ATW), Average Backlog (AB), and Average Total Cost (ATC) over the 2,000 hour time horizon. We then plot ATW, AB, and ATC as functions of the utilization ratio, which is defined as the ratio of the demand rate and the system capacity (we actually use the capacity of the bottleneck work cell– Stepper –as the system capacity). A robust policy should provide low and flat ATW, AB, and ATC curves.

4.1 WIP And Surplus Trajectories

Before we present the average trajectories, we first show the ATCs and their confidence intervals (at a level of 95%) for these four policies in Table 2.

Now let us look at the average trajectories. Figs. 2 and 3 show surplus and average WIP trajectories for a constant demand rate.

In the first period of the experiment (from $t = 0$ to $t = 500$), TB, WB, and WC policies outperform UL (see Fig. 2, where the average surplus trajectories of the last work center are plotted for different policies.). Their surplus levels become non-negative after a while, i.e., they catch up with the production plan, while that associated with UL does not. At the beginning of the experiment, all workstations have negative surplus levels and very small inventory levels. Therefore, only the local inventory thresholds are active for TB and WC policies. If these threshold values are large enough (we will discuss the impact of these values shortly), parts are pumped into the system in order to catch up with the production

Policies	TB	WC	WB	UL
ATC	607.75 ± 2.72	1323.1 ± 20.15	946.83 ± 15.35	1234.95 ± 7.54

Table 2. Average costs for different policies.

plan. Another important parameter is the WIP of the whole job shop. In Fig. 3, we plot

the average WIP trajectories under the four policies. The trajectories of TB and WC are relatively stable. The WB policy does not control the WIP level of the whole job shop directly; it only controls the WIP level up to the bottleneck work center. Since we assume that the job shop has sufficient capacity, its WIP level becomes very high.
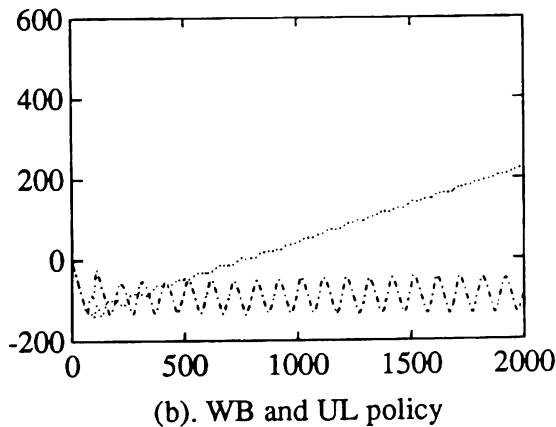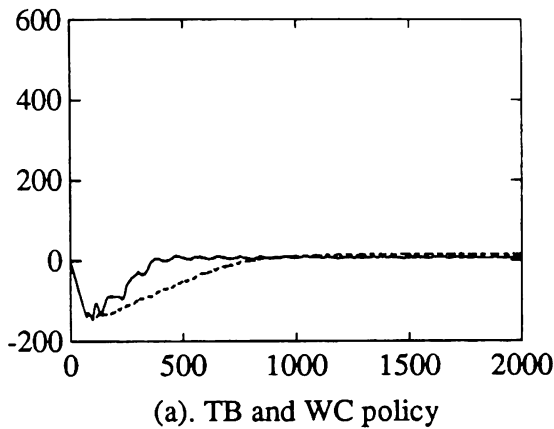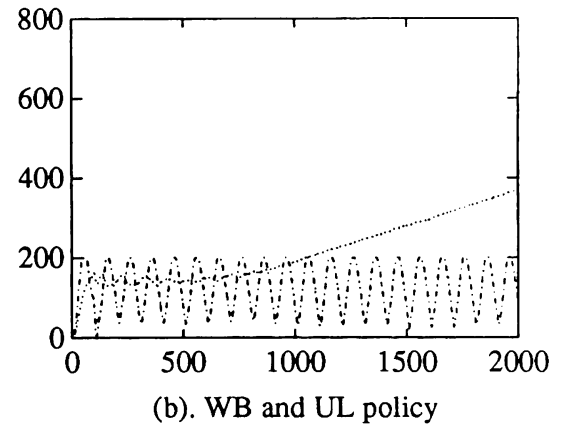

(a). TB and WC policy

(b). WB and UL policy

Figure 3: Average WIP trajectories

(a). TB and WC policy
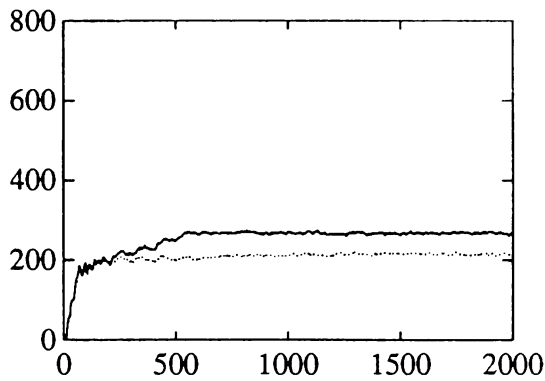

(b). WB and UL policy

Figure 2: Average surplus trajectories

To compare TB and WC, let us see Figs. 2(a) and 3(a). While the WC policy provides a WIP level which is about 40% higher than that of TB in the second period, it also has difficulty to catch up with the production plan in the first period. If we raise the threshold levels for WC, as shown in Fig. 4, where two sets of curves corresponding to two sets of threshold levels are plotted, this catching up process can be accelerated, but the WIP level will be even higher. The reason for TB to out perform the other policies is the presence of two sets of threshold levels.

In the first period, as we stated before, the surplus levels of different stages are negative, therefore only the $h_{b_{ij}}$ are active, and their values are high. Thus, in the first period, a large amount of parts are pumped into the system. In the second period, the surplus thresholds become active, and thus restricting the growth of the WIP levels.

To test the system behavior against demand changes, we first decrease the demand rate by 50% at 1300 hours and then increase it to to its original value at 1600 hours. As before, 100 simulation runs are performed for each policy. Average costs, their corresponding confidence intervals, and the cost changes relative to that presented in Table 2 are listed in Table 3. We see from this table that the influence of the demand fluctuation on TB is negligible. The average trajectories are presented in Figs. 5 and 6. It can be seen that the demand reduction between 1300 and 1600 hours results in an increase of WIP levels. While the increase is moderate for TB and WC, it is substantial for WB and UL. This is because both TB and WC use a feedback control mechanism, which automatically adjust the WIP to a prespecified level.
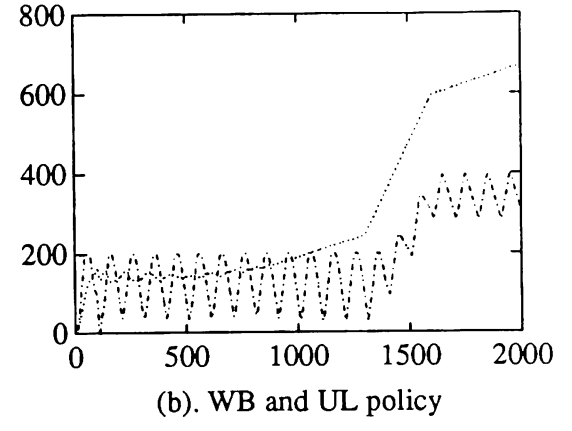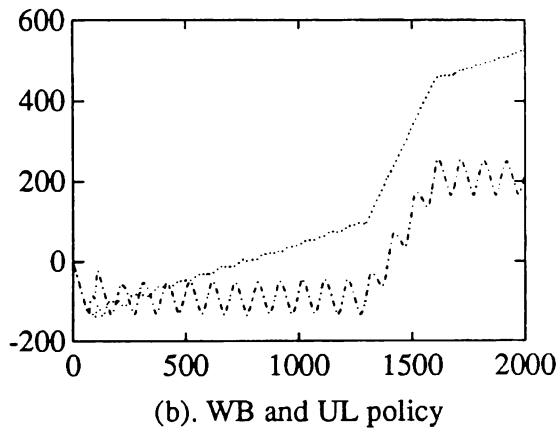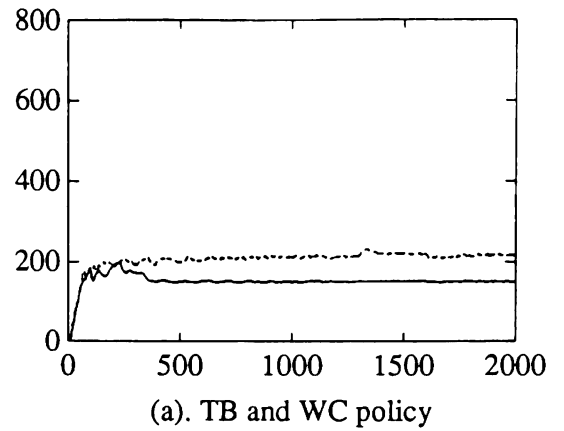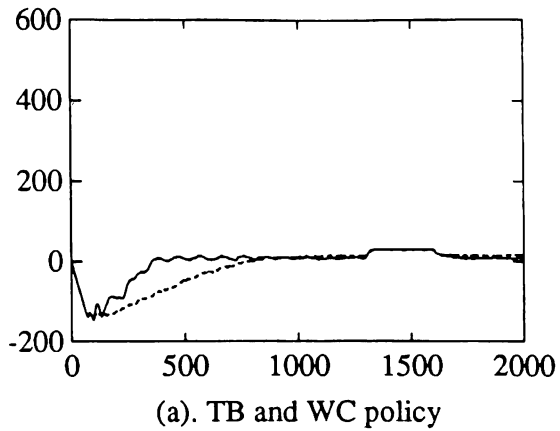

(a). Surplus

(b). Average WIP

——— Higher Threshold

············ Lower Threshold

Figure 4: Surplus and WIP trajectories for WC policy

4.2 Performance Changes With Respect To Changes In The Utilization ratio

The next set of simulations runs shown in Fig. 7 is designed to test the system performance against the changes in the utilization ratio. It should be noted that they are performed under a constant demand rate.

Figs. 7(a) and (b) show how ATW and AB change, respectively, when utilization ratio changes. It can be easily seen that when the system capacity is high (corresponding to a low utilization ratio), the WIP levels are high and backlog levels are low for WC and WB policies. However, when the utilization ratio becomes high, even though their WIP levels are reduced, the backlog levels increase sharply. In contrast to WC and WB, TB keeps relatively stable and low WIP and backlog levels. The increase in the backlog level is moderate when the utilization ratio increases. The UL policy always has a high backlog level.

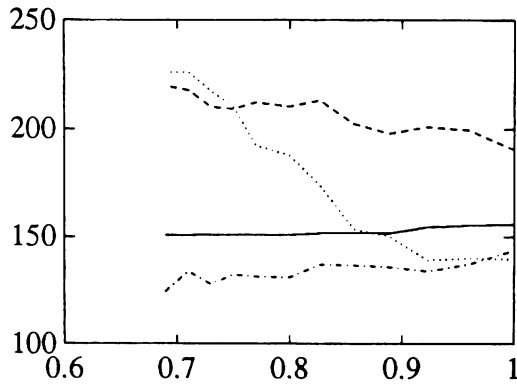| Policies	TB	WC	WB	UL
Cost	604.52 ± 3.40	1403.44 ± 17.93	1276.15 ± 15.24	1166.78 ± 6.22
Cost change	0.53%	6.08%	34.81%	5.67%

Table 3. Average cost when demand rate is changing.

(a). TB and WC policy

(b). WB and UL policy

——— TB   - - - - WC

············ WB   —·— UL

Figure 5: Surplus trajectories with demand change

(a). TB and WC policy

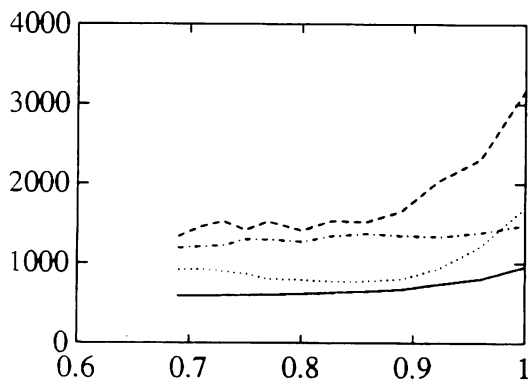(b). WB and UL policy

——— TB   - - - - WC

············ WB   —·— UL

Figure 6: Average WIP trajectory with demand change

(a). Average wip level
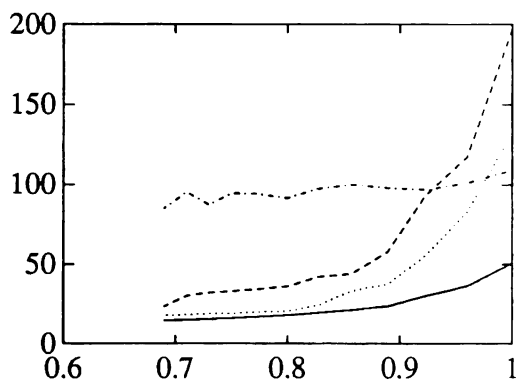
(b). Average backlog level

(c). Average costs

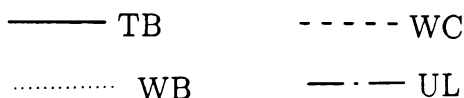——— TB   - - - - - WC

·············· WB   —·— UL

Figure 7: ATW, AB and ATC under different utilization rate

The ATC curves are plotted in Fig. 7(c). The result is similar to what we have just discussed. Once again, TB exhibits the best performance.

5 CONCLUSION

In this paper, we use simulation to compare the performance of four different production policies. Our findings support that the TB policy is the most robust of all when random interference, such as machine breakdowns and demand variations, exists. The WC policy (*i.e.*, the Kanban system), outperforms WB and UL policies. Since the WB policy does not directly control the system WIP, it relies on the system capacity. If the system capacity is relatively high, the WIP level can be out of control. Finally, the UL policy always presents a backlog. Both its WIP and backlog exhibit high variability.

REFERENCES

D.Y. Burman, J. Gurrola-Cal, A. Nozari, S. Sathaye, and J.P. Sitarik, "Performance analysis techniques for IC manufacturing line," *AT&T Bell Labs. Tech. J.*, 1986.

H. Chen, J.M. Harrison, A. Mandelbaum, A. van Ackere, and L.M. Wein, "Queueing network models of semiconductor wafer fabrication," Graduate School of Business, Stanford Univ. CA, 1986.

L. Cory, "Just-in-time approach to IC fabrication," *Solid State Technol.*, pp. 177-179, May, 1986.

J.E. Dayhoff and R.W. Atherton, "Simulation of VLSI manufacturing areas," *VLSI Design*, pp. 84-92, Dec. 1984.

C.R. Glassey, and G.C. Resende, "Closed-loop job release control for VLSI circuit manufacturing,", *IEEE Trans. on Semiconductor Manufacturing*, Vol. 1, No. 1, 1988.

S.X.C. Lou and G. Van Ryzin, "Optimal Control Rules for Scheduling Job Shops", *Annals of O.R.*, Vol. 17, 1989a, pp. 233-248.

S.X.C. Lou, and P. Kager, "A robust production control policy for VLSI wafer fabrication," *IEEE Trans. on Semiconductor Manufacturing*, Vol. 2, No. 4, 1989b.

S.X.C. Lou, H. Yan, S. Sethi, A. Gardel, and P. Deosthali, "Hub-Centered Production Control of Wafer Fabrication", *Proceedings of Advanced Semi-Conductor Manufacturing Conference and Workshop*, Danvers, MA, Sept. 1990.

S.X.C. Lou, H. Yan, S. Sethi, "A robustness study of production control policies", *Working Paper*, University of Totonto, 1991.

C. Lozinski, and C.R. Glassey, "Bottleneck starvation indicators for shop floor control," *IEEE Trans on Semiconductor Manufacturing*, Vol.2 No. 4, 1988.

L.M. Martin-vega, M. Pippin, E. Gerdon, and R. Burcham, " Applying just-in-time in a wafer fab: a Case study", *IEEE Trans. on Semiconductor Manufacturing*, Vol. 2, No. 1, 1989.

P.K.Mozumder, C.R. Shyamsundar, and A.J. Strojwas, "Statistical control of VLSI fabrication Processes: a framework," *IEEE Trans. on Semiconductor Manufacturing*, Vol. 1, No.2, 1988a.

C.R. Shyamsundar, P.K. Mozumder, and A.J. Strojwas,"Statistical control of VLSI fabrication processes: software system," *IEEE Trans. on Semiconductor Manufacturing*, Vol. 1, No. 2, 1988b.

G.V. Ryzin, "Control of manufacturing system with delay," LIDS Report, Lab. for Information and Decision Systems, MIT, 1987.

G. Van Ryzin, S.X.C. Lou, and S. B. Gershwin, "Production Control for a Tandem Two-Machine System", forthcoming in *IIE Transactions*.

S. Sethi and Q. Zhang, "Multilevel hierarchical control in dynamic stochastic marketing–production systems", Working Paper, University of Toronto, 1991.

K. Suzaki, *The New Manufacturing Challenge*, The Free Press, New York, 1987.

L.M Wein,"Scheduling semiconductor wafer fabrication," *IEEE Trans. on Semiconductor Manufacturing*, Vol. 1, No. 3, 1988.

AUTHOR BIOGRAPHIES

SHELDON X. C. LOU received his M.S. and Ph.D. degrees in electrical engineering from the Massachusetts Institute of Technology, Cambridge, in 1982 and 1985, respectively. He was a Research Scientist at the Department of Electrical Engineering of MIT from 1985 to 1987. Currently, he is an Assistant Professor at the Faculty of Management, University of Toronto. His research interests include stochastic control, optimization, production planning and control, and system failure detection and identification.

HOUMIN YAN is a Ph.D. candidate in the Faculty of Management, University of Toronto. He received his B.S. and M.S. degrees in Electrical Engineering from Qinghua University, Beijing, China, in 1982 and 1985, respectively. His current research interests include scheduling of manufacturing systems, production control and planning, simulation methodology and simulation applications.

SURESH SETHI is General Motor Research Professor at the Faculty of Management, University of Toronto, and a principal investigator for the Manufacturing Research Corporation, Ontario. His research interests are production planning under uncertainty, scheduling in manufacturing, and stochastic dynamic optimization.

ANNE GARDEL is a Software Principal Engineer at semiconductor facility of Digital Equipment Corporation. She has a BS in industrial and Management Engineering from Rensselaer Polytechnic Institute and an in Management Engineering from the University of Massachusetts at Amherst. Her interest lies in the area of finite scheduling using simulation and expert system based software.

PRABHAT DEOSTHALI is a Software Engineer in the CIM Technology group in the semiconductor organization at Digital Equipment Corporation. Since joining DEC in 1988, he has had an extensive experience in the areas of modeling and simulation, capacity planning and finite capacity scheduling. He has provided analysis and solutions to the capacity and scheduling problems on the wafer fabs. Prabhat received the Bachelor of Science degree in Mechanical Engineering from the University of Bombay, India. He received the Master of Science degree in Industrial Engineering from Louisiana State University in 1988.