TECHNICAL RESEARCH REPORT

Evaluating the Impact of Process Changes on Cluster Tool Performance

by Jeffrey W. Herrmann, Niranjan Chandrasekaran,
Brian F. Conaghan, Manh-Quan Nguyen,
Gary W. Rubloff, Rock Z. Shi

T.R. 99-30

Evaluating the Impact of Process Changes on Cluster Tool Performance

Jeffrey W. Herrmann, Niranjan Chandrasekaran, Brian F. Conaghan,

Manh-Quan Nguyen, Gary W. Rubloff, and Rock Z. Shi

Institute for Systems Research

University of Maryland

College Park, MD 20742

ABSTRACT

Cluster tools are highly integrated machines that can perform a sequence of semiconductor manufacturing processes. Their integrated nature can complicate analysis when evaluating how process changes affect the overall tool performance.

This paper presents two integrated models for understanding cluster tool behavior. The first model is a network model that evaluates the total lot processing time for a given sequence of activities. By including a manufacturing process model (in the form of a response surface model, or RSM), the model calculates the total lot processing time as a function of the process parameter values and other operation times. This model allows one to quantify the sensitivity of total lot processing time with respect to process parameters and times.

In addition, we present an integrated simulation model that includes a process model. For a given scheduling rule that the cluster tool uses to sequence wafer movements, one can use the simulation to evaluate the impact of process changes including changes to product characteristics and changes to process parameter values. In addition, one can construct an integrated network model to quantify the sensitivity of total lot processing time with respect to process times and process parameters in a specific scenario.

The examples presented here illustrate the types of insights that one can gain from using such methods. Namely, the total lot processing time is a function not simply of each operation's process time, but specifically of the chosen process parameter values. Modifying the process parameter values may have significant impacts on the manufacturing system performance, a consequence of importance which is not readily obvious to a process engineer when tuning a process (though in some cases, reducing process times may not change the total lot processing time much). Additionally, since the cluster tool's maximum throughput depends upon the process parameters, the tradeoffs between process performance and throughput should be considered when evaluating potential process changes and their manufacturing impact .

1. INTRODUCTION

Importance of cluster tools in semiconductor manufacturing

A cluster tool is a manufacturing system with integrated processing modules linked mechanically. Typical cluster tools include load locks, process modules, and a wafer handler. A cluster tool can process multiple wafers simultaneously. Sequential cluster tools integrate a sequence of processes, while other tools have two or more identical modules that are used in parallel.

A sequential cluster tool can improve yield and device performance since wafers are exposed to fewer contaminants between process steps. The tool can include an in-situ metrology step that provides real-time feedback on process performance. In addition, sequential cluster tools reduce queueing and thus cycle time. Finally, because they may require less operator intervention, they can reduce operating costs. A cluster tool with multiple parallel modules can increase throughput and reduce cycle times by reducing the total time needed to process a lot of wafers. In addition, it may be more reliable, since a single module's failure doesn't necessarily stop production. And a cluster tool uses less space than the standalone tools that it replaces.

Semiconductor manufacturers are increasingly using cluster tools. Annual sales of cluster tools will increase from $11.2 billion in 1997 to $21.9 billion in 2000 (*Semiconductor Business News*, 1998).

Cluster tool performance

As with other tools, manufacturers use metrics such as overall equipment effectiveness (OEE) and cost-of-ownership (COO) to evaluate cluster tool performance (Murphy *et al.*, 1996; Dance *et al.*, 1998). These measures require, among many things, the cluster tool's maximum throughput, which is inversely related to the time needed to process an entire lot of wafers. In addition, the total lot processing time is an important input to sector-level or factory-level discrete event simulation models and shop floor scheduling methods. However, lot processing time in a cluster tool is not a constant, but rathera function of batch size, product characteristics, and the individual process times. Moreover, the process times are functions of the process parameters, which change the achievable process rate and thus the time required.

Unlike single-process tools, the complex behavior of a cluster tool makes determining this relationship a difficult task. There are simulation models that describe cluster tool behavior (see, for instance, Wood, 1994; Mauer & Schelasin, 1994; LeBaron and Pool, 1994; and Atherton *et al.*, 1990). These tools are very useful for evaluating performance of a specific tool operating under specific conditions. Existing discrete event simulation models do not, however, yield insight into how process changes affect cluster tool performance, since they take fixed values for each process step without describing the relationship between process parameters and process step times.

Prior work

Wood (1996) presents simple models that relate the total lot processing time to the number of wafers in the lot. He derives formulas for ideal sequential and parallel tools. After carefully considering the transitions at the beginning and the end of the lot, Perkinson *et al.* (1994) derive a model that relates the total lot processing time to the number of wafers. Both papers present linear models and identify two operating regions: in one region, the total lot processing time is constrained by the wafer handling time; in the second region, the total lot processing time is constrained by the module process time. Venkatesh *et al.* (1997) analyze the throughput of a sequential cluster tool with a dual-blade wafer handler. They also identify conditions when the tool operation is constrained by the wafer handler. Srinivasan (1998) presents more detailed Petri net models for sequential tools and parallel tools and uses these to determine the steady state behavior of the tool.

Lopez & Wood (1996) compare two systems of cluster tools, (1) a series of tools that have multiple parallel modules and (2) a set of sequential cluster tools. They conclude that equipment reliability affects the configurations' relative performance. As reliability improves, the sequential cluster tools reduce cycle time. Lopez & Wood (1998) present analytical models for these systems when the tools are perfectly reliable. Dhudshia & Hepner (1996) address issues related to measuring cluster tool reliability.

Contributions of this paper

This paper moves beyond the previous work by considering how process changes affect cluster tool performance. The performance measure of interest is the total lot processing time.

To do this, we must consider two types of cluster tool control. The cluster tool controller has to sequence the wafer movements. It may use some rule or ***scheduling algorithm*** to dispatch the wafer handler dynamically as modules become available or finish processing. Or, the controller may follow a ***prespecified sequence*** of wafer moves. The controller's approach influences how a process change affects cluster tool performance.

When the controller follows a ***prespecified sequence***, we use an analytical model of cluster tool behavior that quantifies how process changes affect cluster tool performance. This model extends the simple models presented before by more precisely describing the cluster tool behavior, especially when the cluster tool has a mix of sequential and parallel modules. In addition, this model incorporates manufacturing process models that relate process performance to process parameter values. This allows one to understand how process parameter changes affect cluster tool performance. This goes beyond the traditional use of manufacturing process models for process optimization and control. For a specific example, we use our analytical model to evaluate how process changes affect lot makespan (i.e., the total time required to fully process a wafer lot). The changes include deposition thickness and process

parameters such as temperature and pressure. In addition, this model allows one to quantify the sensitivity of cluster tool performance to small changes that don't affect the sequence of activities.

When the controller uses a rule or ***scheduling algorithm*** to sequence the wafer moves, we use cluster tool simulation software to evaluate cluster tool performance. The simulation can use the rule to determine the sequence of wafer movements and activities. To evaluate the impact of changing process parameter values, we incorporate a process model into the simulation software. For an example, we see how reductions to the deposition thickness affect the cluster tool performance, and how the sequencing rule and the tool configuration affect this result.

These results promise to help process engineers understand how process changes affect the tool performance. Although each cluster tool configuration is different, our results provide some basic insights. The methodology presented here can be applied to other cluster tools. With these results, process engineers can develop better processes, equipment purchasers can make better procurement decisions, and fab managers can improve factory performance.

Organization of Paper

The remainder of the paper is structured as follows: Section 2 presents an analytical model that describes total lot processing time for a given sequence of activities. Section 3 discusses the impact of process time changes on total lot processing time. Section 4 explains how process parameter changes affect cluster tool performance. Section 5 discusses an integrated simulation model and shows how process changes impact the sequence of activities and the cluster tool performance. Section 6 summarizes our results and concludes the paper.

2. A NETWORK MODEL FOR PRESPECIFIED SEQUENCES

To process a lot of wafers, the components of a cluster tool must perform a large number of tasks. The wafer handler must move wafers between the load lock and the process modules. Each process module must process wafers. The sequence of wafer moves leads to a certain sequence of activities. Consider a cluster tool that follows a prespecified sequence of wafer moves. That is, the cluster tool controller does not determine the sequence of wafer moves dynamically, based on the state of the system; instead, it follows this prespecified script. In this case, a process change will not modify the sequence of activities, but it will alter the process times and lot makespan.

For the given sequence of activities, one can model the cluster tool behavior using a network. A network is a collection of nodes and directed arcs. In our model, each node represents an activity, and directed arcs between nodes describe the precedence constraints between activities. For each wafer, certain activities must occur in a certain order, and this defines some precedence constraints. Since a resource (a wafer handler or a process module) can perform only one activity at a time, this defines other

precedence constraints. (The sequence of activities determines these relationships.) Figure 1 shows an example for a lot of two wafers. Each wafer requires processing in Orientation and in Deposition. Figure 1(a) shows the Gantt chart (or timing diagram) for each resource in the cluster tool. Figure 1(b) shows the corresponding network. (Note that this example includes no anticipatory wafer handler moves.)

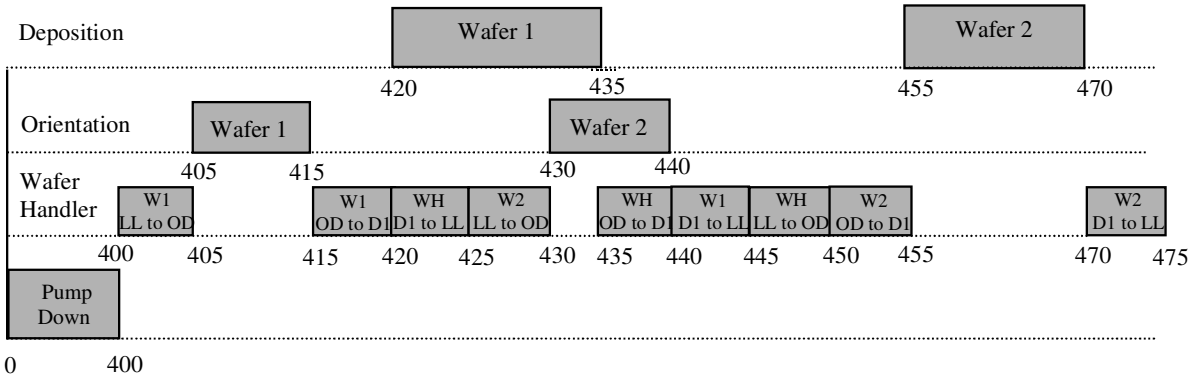


Figure 1(a). Gantt chart for two wafers. The horizontal axis is time in seconds.
Note that WH = Wafer Handler, W1 = Wafer 1, W2 = Wafer 2, LL = Loadlock, OD = Orientation and Degas, D1 = Deposition Chamber 1.

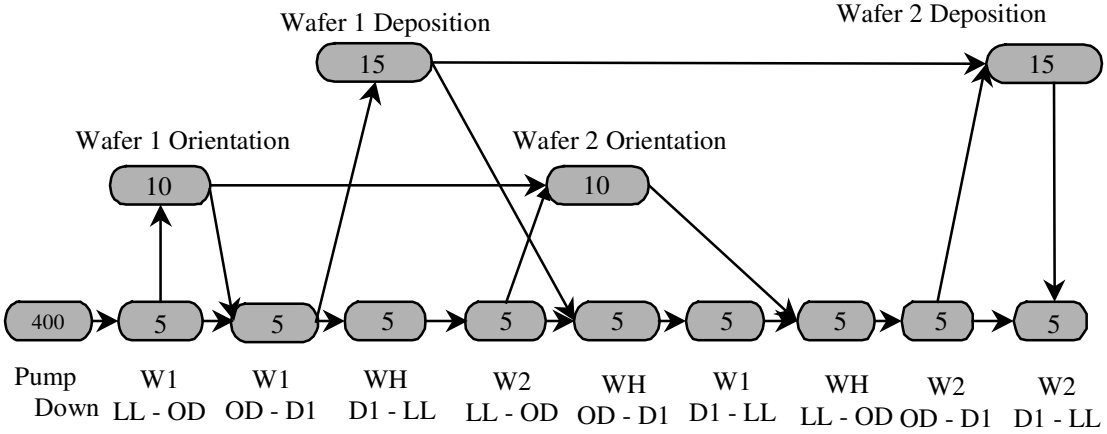


Figure 1(b). Corresponding Network.

Associated with each node is the time that the corresponding activity requires. A path through the network is a sequence of nodes connected by arcs. The length of a path is defined as the sum of the components nodes' times. The network's critical path is the longest path through the graph. (Note that there might be more than one critical path, and this critical path is unrelated to the critical path of the device being constructed.) The critical path's length equals the total time needed to process the lot. This is the total lot processing time, which we call the *lot makespan*. This total lot processing time affects the throughput of the cluster tool. Figure 2 shows the critical path for the network of Figure 1.

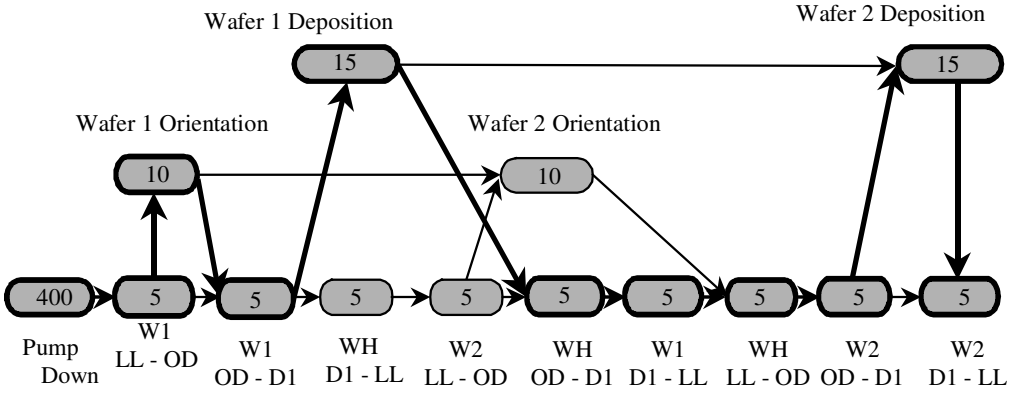


Figure 2. Critical Path for Network

The following algorithm will identify the critical path and calculate its length. Given a network $N = (V, A)$, where $V = \{v_1, ..., v_n\}$ is the set of nodes and A is the set of directed arcs $\{(j, k)\}$. Each node v_k represents a distinct activity in processing the lot. Arc (j, k) is in A if activity j must precede activity k. Associated with each node v_k is the time t_k that the activity requires. Let $P(v_k)$ be the set of immediate predecessors of node v_k. That is, v_j is in $P(v_k)$ if and only if (j, k) is in A. Similarly, let $S(v_k)$ be the set of immediate successors of node v_k. That is, v_j is in $S(v_k)$ if and only if (k, j) is in A.

Then, calculate the earliest completion time C_k of each node v_k as follows:

$C_k = t_k$ if $P(v_k)$ is empty. Otherwise $C_k = \max\{C_j: j \text{ is in } P(v_k)\} + t_k$. Repeat for all v_k in V.

Let $T = \max\{C_k: v_k \text{ is in } V\}$. Calculate the latest completion time D_k of each node v_k as follows:

$D_k = T$ if $S(v_k)$ is empty. Otherwise $D_k = \min\{ D_j - t_j: j \text{ is in } S(v_k)\}$. Repeat for all v_k in V.

All nodes v_k such that $C_k = D_k$ are on a critical path. The length of any critical path is T.

From this, we can write MS = T = $f(t_1, ..., t_n)$. That is, the makespan is a function of the activity times.

3. PROCESS TIME CHANGES

The network model provides some initial insight. Increasing the time of an activity on a critical path will increase the length of the critical path. Increasing the time of an activity not on a critical path will not increase the length of the critical path if the activity remains off the critical path. However, a large increase will yield a different and longer critical path. Figure 3(a) presents a network and its critical path. Figure 3(b) illustrates an increase in the time of an activity that is on the critical path. Figure 3(c) illustrates a small increase in the time of an activity that is not on the critical path. Figure 3(d) illustrates a large increase in the time of that same activity, which creates a new critical path.

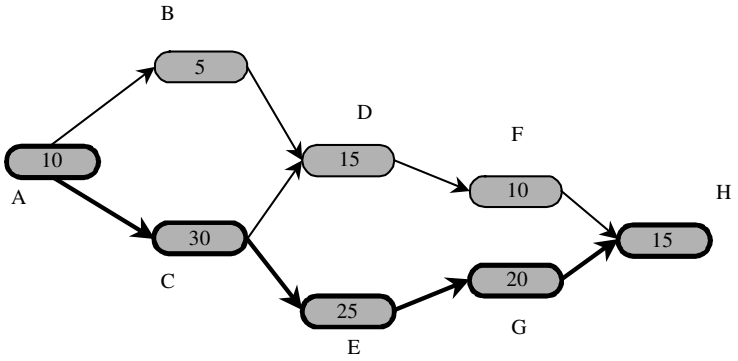


Figure 3(a). A Network with a Critical Path of Length 100

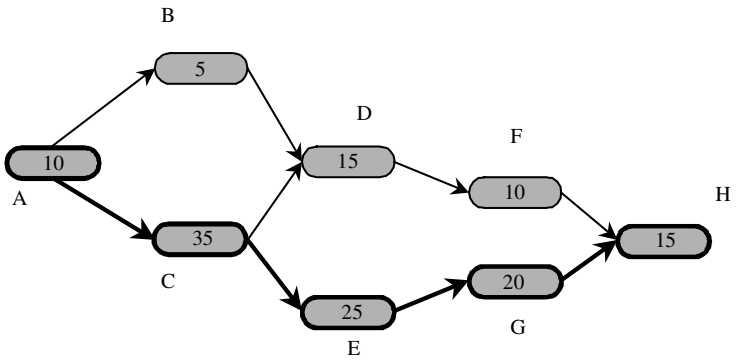


Figure 3(b). The Network after Increasing Activity C by 5.
The critical path is now 105.

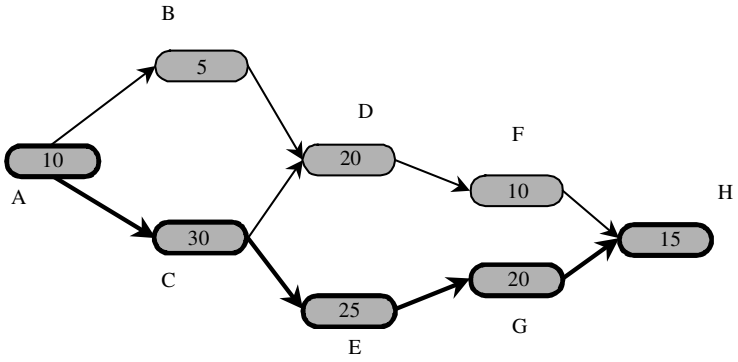


Figure 3(c). The Network after Increasing Activity D by 5.
The critical path remains 100.

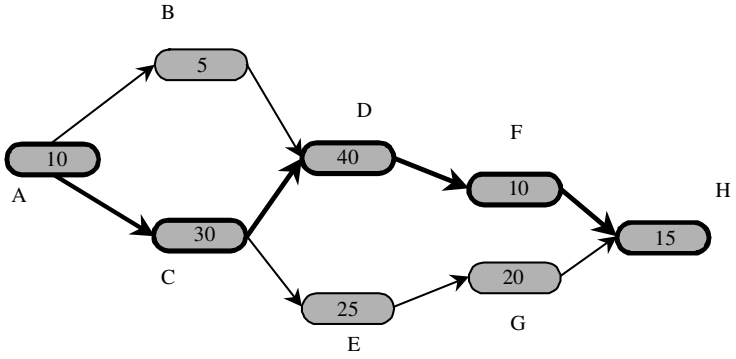


Figure 3(d). The Network after Increasing Activity D by 25.
The critical path is now 105.

Cluster tool processing involves repeating certain types of activities many times as it processes a lot of wafers. We will use the term "operation" to describe a type of activity that can occur multiple times. This includes processes such as deposition as well as other activities such as wafer handling. If there are m different operations, and operation k requires p_k time units, then $t_j = p_k$ if activity j is an operation of type k. Thus, we can write MS = $f(p_1, ..., p_m)$. The lot makespan is a function of the operation times.

Consider an increase to the time that an operation requires. The more often this operation occurs on a critical path, the more it will affect the total critical path length and thus increase the lot makespan. If an operation occurs *n* times on the critical path, increasing its time by *t* time units will increase the lot makespan by *nt* time units.

Consider, for example, the network in Figure 4(a). The deposition process time equals 7 seconds, and the critical path includes the deposition process for the second wafer. Increasing the process time by

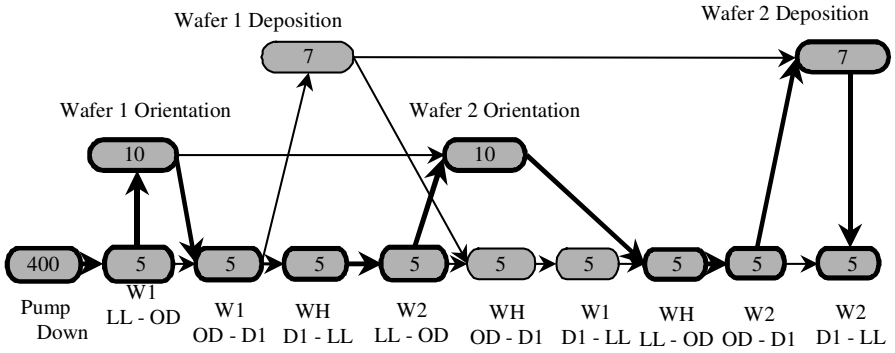


Figure 4(a). The critical path includes one deposition process.

one second increases the lot makespan by one second. The other instance of this operation also requires more time, but it is not on the critical path. If the increase is large enough, the other activity may change the critical path, making it even longer (unless the sequence of activities changes). Consider the network in Figure 4(b). In this case, the deposition process time has increased to 15 seconds. This changes the critical path, which now includes both deposition processes. Increasing the deposition process time by one second now increases the lot makespan by two seconds.

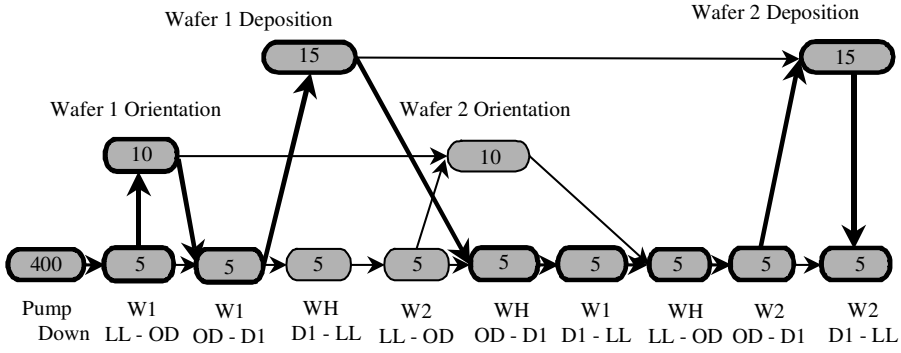


Figure 4(b). The critical path includes both deposition processes.

In general, as the operation time continues to increase, the operation will occur on the critical path more often, thus amplifying the impact of each unit of increase. In addition, the critical path includes multiple operations. Increasing any operation's time will increase the lot makespan. Calling one

operation the “bottleneck” may be inappropriate. However, it would be reasonable to say that one operation is “more critical” than another if the critical path includes the first operation more often than the second.

Using this methodology, we examined the impact of process time changes on lot makespan for a specific cluster tool. The cluster tool has one loadlock that holds the entire lot of 20 wafers, one single-wafer chamber for orientation and degassing (OD), and two single-wafer process modules that perform tungsten chemical vapor deposition (W CVD). Each wafer requires processing at the OD chamber and at one of the two W CVD chambers. The tool has a single wafer handler that moves wafers between the loadlock and the chambers.

We constructed a network that corresponds to a specific sequence of activities, as depicted in Appendix A. By analyzing this network we can evaluate how the lot makespan (MS) is sensitive to the deposition process time (D). (All other parameters remain constant.) Thus, we can write MS = f(D) = 640 + max {D, 10} + max {D, 35} + 8 max {D, 40}.

Table 1 summarizes the results, and Figure 5 illustrates the relationship. The sensitivity dMS/dD is the derivative of lot makespan with respect to deposition process time. This clearly shows that, under certain conditions, increasing the process time barely changes the lot makespan. In other conditions, however, increasing the process time causes a much greater change, as evidenced by the sharp upturn in lot makespan for deposition process times above 40 sec. in Fig. 5. Note that these results provide a more exact analysis than the simple models presented in previous research.

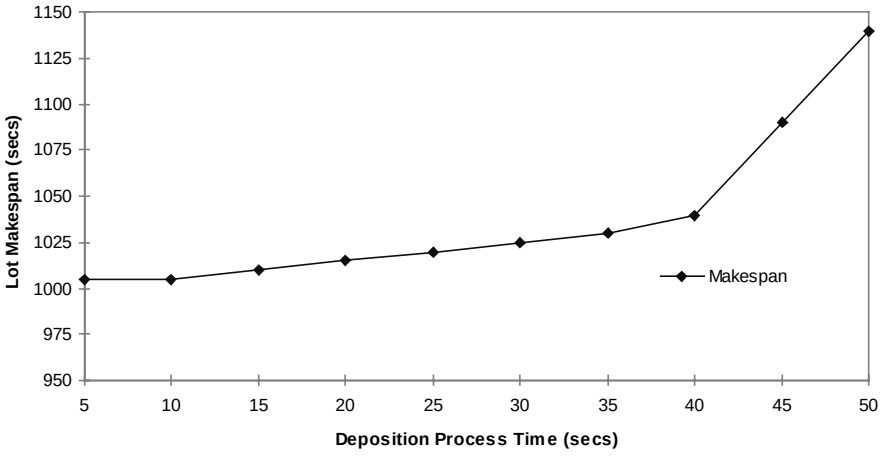


Figure 5. Lot makespan versus deposition process time.
(Note that the makespan axis does not start at zero.)

Likewise, reducing an operation's process time will have a more significant impact at first. As the process time continues to decrease, the operation occurs less often on the critical path, so the reductions don't have the same benefit.

Also, note that these results depend upon the times chosen for the other operations and upon the sequence of wafer moves. Different process times or sequences yield different results.

Deposition Process Time D (seconds)	Lot Makespan (MS)	Sensitivity dMS/dD	Lot Makespan Range(seconds)
$0 \le D \le 10$	MS = 1005	0	MS = 1005
$10 \le D \le 35$	MS = 1005+ (D - 10)	1	$1005 \le MS \le 1030$
$35 \le D \le 40$	MS = 1030 +2 (D - 35)	2	$1030 \le MS \le 1040$
$D > 40$	MS = 1040 + 10 (D - 40)	10	$MS > 1040$

Table 1. Relationship between lot makespan and deposition process time.

4. INTEGRATING PROCESS MODELS

In semiconductor manufacturing, as in other manufacturing environments, a manufacturing process is governed by a number of process parameters. When executing the process, the operator (or the computer controlling the process) sets the process parameters to prescribed settings so that the process will run effectively and efficiently. Determining good settings involves many tradeoffs between product quality, product performance, consumables cost, and nominal processing time. Often it is necessary to change the process parameter settings to improve process performance, to enhance process integration with other process steps, to restore process performance after a disturbance, or to shift technology design points in accordance with scaling toward more aggressive technology nodes.

When attempting to determine new settings, a process engineer may conduct a set of experiments to evaluate how the process parameters affect the process performance. Each experiment is a lot of wafers processed under a specific combination of parameter values. In theory, an engineer could conduct an experiment for every possible combination of parameter values. Since there may exist a large number of possible combinations, however, in practice the engineer selects a small subset of the combinations and runs these experiments. Then, using statistical software (like ECHIP), the engineer can construct a response surface model (RSM) that fits the experimental results. The RSM is an empirical (often quadratic) mathematical formula that relates process performance to the process parameter values. (For more information on designing experiments and forming RSMs, see Box & Draper, 1987.) The RSM

gives the engineer insight into how the process parameters affect the process performance. The engineer can then select the new process parameter settings that best meet the process performance goals.

This experimental approach to optimizing manufacturing processes by changing the process parameters has been very successful (e.g., Stefani *et al.*, 1996). However, process engineers often focus on the process itself and may find it difficult to consider how process parameters changes affect the overall manufacturing system performance. One significant impact of changing process parameters is a change to the process time. If a cluster tool performs a multi-step process sequence, the process parameter change may affect the lot makespan in more complex ways, which affects tool throughput. However, process engineers usually develop RSMs for process rate (like etch rate or deposition rate). Although a higher rate should improve throughput by decreasing the nominal process time, quantifying the impact can be a difficult task. As we saw in the previous section, sometimes a small change to the process time changes the lot makespan drastically, and sometimes it does not. Process "improvements" that significantly lower a cluster tool's throughput (especially if that tool is a bottleneck tool) can seriously degrade manufacturing system performance by increasing cycle time and decreasing maximum throughput.

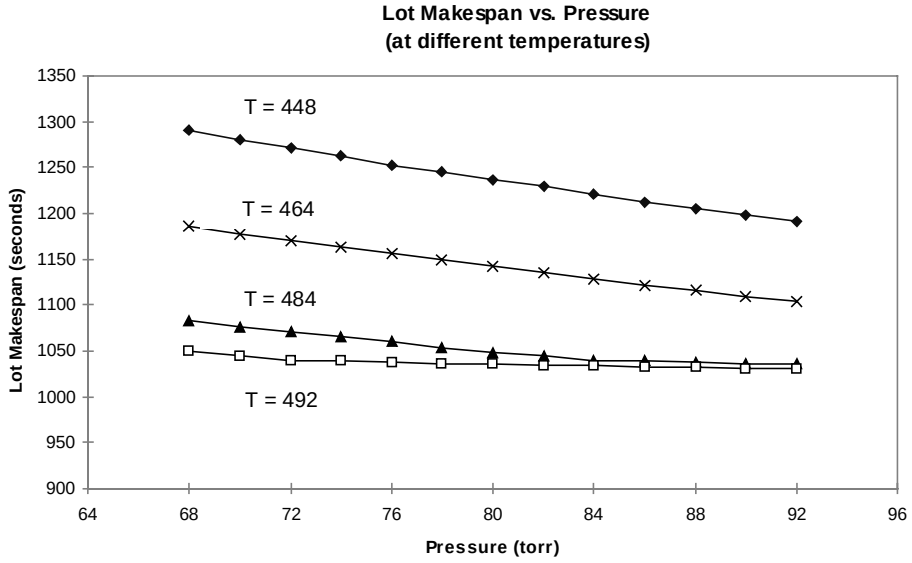
To acquire better feedback, the process engineer needs an integrated cluster tool model that incorporates process models. Specifically, we need to relate the lot makespan to the process parameters and relevant product attributes. To do this, we first use a manufacturing process model to relate the operation time p_k to these parameters, denoted by the vector x_k. If we ignore the other operations, then $p_k = g_k(x_k)$, and $MS = f(p_k) = f(g_k(x_k)) = h(x_k)$.

Consider the cluster tool example presented earlier. We included an RSM for W CVD in the network model to understand how process parameter changes affect lot makespan. For analyzing the W CVD cluster tool, we used an RSM for the W CVD process that was based on data collected by Stefani *et al.* (1996). The deposition rate RSM has the following four process parameters: reactor pressure, deposition temperature, the mole fraction of WF_6, and the mole fraction of H_2. The output is the average deposition rate (A/sec). The process is a H_2 reduction of WF_6, run in an Applied Materials Centura reactor, preceded by a short SiH_4 and WF_6 nucleation step that deposits a 400 Å seed layer.

Specifically, let *DR* be the actual deposition rate in Å per second, *P* the reactor pressure in torr, and *T* the deposition temperature in degrees Kelvin. Then the RSM *DR*(*P*, *T*) can be expressed as follows (the mole fractions were set to their median values used in the experiments):

$$DR = 63.55 + 0.4248(P-80) - 364.6\left(\frac{1000}{T} - 1.346\right) - 2.079(P-80)\left(\frac{1000}{T} - 1.346\right)$$
$$+ 1.297x10^{-4}(P-80)^2 + 945.5\left(\frac{1000}{T} - 1.346\right)^2$$

The deposition process time $D = g(Th, P, T) = Th/DR(P, T)$, where Th equals the deposition thickness. From this we can construct an integrated network model that expresses the lot makespan as a function of pressure and temperature. $MS = h(Th, P, T) = 640 + \max\{Th/DR(P, T), 10\} + \max\{Th/DR(P, T), 35\} + 8 \max\{Th/DR(P, T), 40\}$. If we set $Th = 3000$ Å and calculate a few values (see Figure 6), we find that, as expected, the lot makespan decreases as temperature and pressure increase. These changes increase the deposition rate. The mole fractions do not affect the deposition rate or lot makespan as significantly. At lower temperatures, the impact of pressure and temperature changes is



large, because the deposition process time is large, and thus the deposition process is on the critical path many times. However, at higher temperatures, increasing the pressure does not decrease lot makespan as much.

Figure 6. Lot makespan versus pressure at different temperatures
(note that the makespan axis does not start at zero).

With the integrated network model, we can model how process parameter changes affect tool performance and find the partial derivatives of lot makespan with respect to the process parameters. For our example, we first differentiate the deposition process time with respect to the pressure and temperature:

$$\frac{\partial D}{\partial T} = \frac{-3x10^6}{(DRxT)^2}\left[2743 + 2.079P - \frac{1.9x10^6}{T}\right]$$

$$\frac{\partial D}{\partial P} = \frac{-3000}{DR^2}\left[3.2 + 2.6x10^{-4}P - \frac{2080}{T}\right]$$

The derivative of lot makespan with respect to *D* is the sensitivity dMS/dD (as shown in Table 1). Thus, if we multiply the above terms by this sensitivity, we have the partial derivatives of lot makespan with respect to the pressure and temperature. Figure 7 graphs the derivative of lot makespan with respect to pressure (dMS/dP) as pressure and temperature change. This also shows that, at lower pressures and temperatures, the lot makespan is more sensitive to changes in pressure (the derivative is more negative). At higher pressures and temperatures, the lot makespan is less sensitive to pressure changes because the critical path has shifted and does not include the deposition process as often.

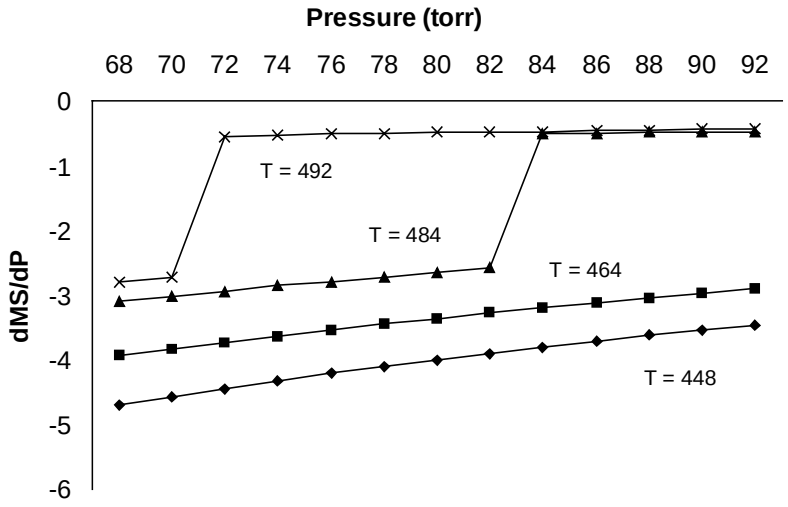


Figure 7. Derivative of lot makespan with respect to pressure

These results yield two important benefits. First, they demonstrate that with such analysis tools (i.e., incorporating process RSM's into operations simulators) an engineer changing the process parameters can determine if the proposed change will significantly increase the lot makespan. If so, it would be prudent to calculate the impact that this increase will have on the entire manufacturing system

and perhaps consider a less drastic change. Second, these results show that the lot makespan is sensitive to process parameters, and the analysis tools provide a mechanism to assess this sensitivity both qualitatively and quantitatively. An equipment engineer considering whether to purchase the tool might normally calculate the cluster tool's maximum throughput; but now it is possible to use these results to determine a lower bound and an upper bound on the tool's maximum throughput, which can help determine bounds on the tool's cost effectiveness. In addition, if the lot makespan is very sensitive to an operation process time, then that operation is on the critical path often. Thus, it may be worthwhile to consider adding to the cluster tool another process module that performs that operation.

Finally, note that the integrated network model shows how the lot makespan depends on a product attribute, such as the deposition thickness for a W plug height and aspect ratio. This can be important in technology shift or shrink. As an example, consider the impact of technology changes on our W CVD cluster tool performance. As gate widths decrease, the interconnect diameter decreases, which lowers the required W deposition thickness. If the process parameters remain the same, the deposition process time will decrease, and the lot makespan will decrease. Using the integrated network model, we can evaluate the impact on lot makespan. The W CVD process parameters are set to their median values. Table 2 shows the results. The most important conclusion is that large reductions to deposition thickness do not generally cause corresponding, or even significant, reductions in lot makespan (or great increases in throughput). In these regions, the deposition process affects the lot makespan, but not very much. Specifically, under these conditions, the sensitivity dMS/dD = 1, so the lot makespan is not very sensitive to deposition process changes.

Technology Node (nm), Layer	Deposition Thickness Th (angstroms)	Deposition Process Time D (seconds)	Percent Reduction in Process Time	Lot Makespan MS (seconds)	Percent Reduction in Lot Makespan
250, Contact	2100	33	-	1028	-
180, Contact	1500	24	27	1019	0.9
150, Contact	1275	20	39	1015	1.3
130, Contact	1050	17	48	1012	1.6
250, Via 1-2	2700	42	-	1060	-
180, Via 1-2	1950	31	35	1026	3.2
150, Via 1-2	1575	25	40	1020	3.8
130, Via 1-2	1350	21	50	1016	4.2
250, Via 3-4	3750	59	-	1230	-
180, Via 3-4	2625	41	31	1050	15
150, Via 3-4	2188	34	42	1029	16

130, Via 3-4	1896	30	49	1025	17
250, Via 5	7500	118	-	1820	-
180, Via 5	5250	83	30	1470	19
150, Via 5	4375	69	42	1330	27
130, Via 5	3792	60	49	1240	32
150, Via 6	5250	83	-	1470	-
130, Via 6	4550	72	13	1360	7.5

Table 2. The impact of deposition thickness on lot makespan.

(Under a fixed sequence)

5. SIMULATION MODELS AND DISPATCHING RULES (SCHEDULING ALGORITHMS)

When the cluster tool controller uses a rule or scheduling algorithm to dispatch the wafer handler dynamically as the tool processes the lot, large changes in process times will lead to different sequences of activities. Analytical models like the network model presented above cannot model these complex sequencing decisions. In this case, discrete-event simulation models are a natural tool, since the simulation can be programmed to use the same rule. The simulation can determine the sequence of activities and the total lot processing time.

Typically, the input to a cluster tool simulation model includes information about the tool configuration, the number of wafers in a lot, the sequence of processes that each wafer should undergo, the time that each operation (e.g., wafer handler move or process) requires, and a rule for moving the wafers within the tool. An example is the Cluster Tool Performance Simulator (CTPS) software that Lee Schruben developed at Cornell University.

This information is sufficient for evaluating the cluster tool performance in a given scenario. If the process time changes, one can change the input parameters and recalculate the tool performance. However, in practice, process times themselves are a function of the product characteristics and the process parameter values. Thus, as described above, it would be desirable to include these attributes as the input to the simulation model. Then, one can use the simulation to evaluate the cluster tool performance when the product characteristics or process parameter values change.

We have created this type of integrated simulation model. We started with the CTPS software mentioned above. To model the cluster tool described in Section 3, we added the W CVD RSM to the software. The user enters the product's deposition thickness and the process parameter values (reactor pressure, deposition temperature, the mole fraction of WF_6, and the mole fraction of H_2), in addition to the other model inputs. The software then uses the RSM to calculate the deposition process time and continues by simulating the cluster tool behavior.

As an example, consider the impact of technology changes on our W CVD cluster tool performance. As gate widths decrease and the interconnect diameter decreases, the required deposition thickness decreases and the deposition process time decreases. Thus, it might seem possible to reduce the number of process modules required to achieve a certain throughput. For instance, if the deposition thickness decreases by 50%, perhaps one module can do the work of two. We can use the integrated simulation model to check this.

The cluster tool can use a push dispatching rule or a pull dispatching rule. The dispatching rule helps the controller sequence wafer moves when two different wafers are waiting for the wafer handler. Under the pull rule, the wafer that has fewer remaining process steps is moved first. Under the push rule, the wafer that has more remaining process steps is moved first. Thus, in our example, suppose there are unprocessed wafers in the loadlock, an empty OD chamber, and a deposition chamber is holding a processed wafer. The pull rule will give priority to the wafer that has finished deposition. The push rule will give priority to the next unprocessed wafer that needs to visit the OD chamber.

As the technology shifts, the deposition thickness at each layer will decrease, as shown in Table 3. The W CVD process parameters are set to their median values. Under either dispatching rule, the integrated simulation model can determine a sequence of activities and determine the lot makespan. In addition, we can evaluate the impact of removing the second deposition chamber.

Table 3 shows the results, which are similar to those in Table 2. Figures 8a and 8b graph the results for the one-chamber configuration and the two-chamber configuration. The most important conclusion is that large reductions (up to 2X) in deposition thickness do not necessarily cause comparable reductions in lot makespan (or large increases in throughput). The impact is greater when the tool has only one deposition chamber, but the performance with one chamber never reaches the two chamber performance.

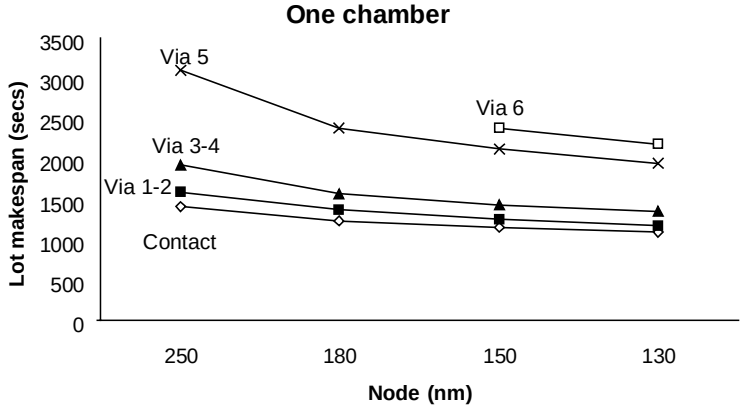


Figure 8a. Lot makespan as a function of technology node
(one deposition chamber).

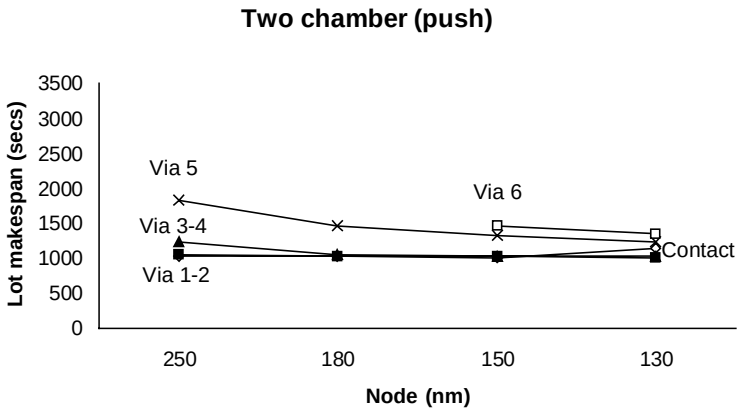


Figure 8b. Lot makespan as a function of technology node
(two deposition chambers, push).

Technology Node (nm), Layer	Deposition Thickness Th (angstroms)	Deposition Process Time D (seconds)	Lot Makespan MS (seconds)		
			Two chambers, push	Two chambers, pull	One chamber
250, Contact	2100	33	1028	1028	1465
180, Contact	1500	24	1019	1100	1285
150, Contact	1275	20	1015	1060	1205
130, Contact	1050	17	1145	1145	1145
250, Via 1-2	2700	42	1060	1060	1645
180, Via 1-2	1950	31	1026	1026	1425
150, Via 1-2	1575	25	1020	1020	1305
130, Via 1-2	1350	21	1016	1070	1225
250, Via 3-4	3750	59	1230	1230	1985

180, Via 3-4	2625	41	1050	1050	1625
150, Via 3-4	2188	34	1029	1029	1485
130, Via 3-4	1896	30	1025	1025	1405
250, Via 5	7500	118	1820	1820	3165
180, Via 5	5250	83	1470	1470	2445
150, Via 5	4375	69	1330	1330	2185
130, Via 5	3792	60	1240	1240	2005
150, Via 6	5250	83	1470	1470	2445
130, Via 6	4550	72	1360	1360	2245

Table 3. The impact of dispatching rules and chambers on lot makespan.

Note: when Th = 1050, the tool uses only one deposition chamber.

Note that both rules can cause the lot makespan to increase when the deposition thickness decreases although the deposition process time decreases (e.g., two chambers, pull, for 150, Via 1-2 vs. 130, Via 1-2). This can happen because the controller doesn't use the second deposition chamber when the wafer handler empties a deposition chamber before moving a wafer from OD to deposition. (When the controller follows a prespecified sequence as in Section 3, the sequence forces the tool to use both chambers.) Also, in some cases, the pull rule performs worse than the push rule.

From these results, one can see that the one-chamber configuration is worse than the existing scenario even after technology shifts and the deposition thickness decreases. Moreover, the decreasing thickness can lead to worse performance, so a new control scheme may be necessary. These results illustrate the potential of the integrated simulation model. And they illustrate how the dispatching rule and tool configuration influence the impact that a process change has. Finally, they demonstrate a methodology for evaluating throughput changes with technology evolution, which may in turn be incorporated into cost-of-ownership assessments in order to advise equipment purchase decisions.

Finally we note that the changing sequences affect how much the lot makespan is sensitive to the deposition process time. From each sequence that the simulation model creates, we can construct the corresponding integrated network model (as discussed in Sections 2 and 3), which is valid for changes that do not modify the sequence of activities. From this model, we can calculate the sensitivity. If a critical path includes deposition processes more often, the lot makespan is more sensitive to the deposition process time. For the one chamber configuration, the sensitivity is 20, since the critical path includes every deposition process. For the two chamber configuration, the sensitivity ranges from 10 (when the deposition thickness is large) to 1 (when the deposition thickness is smaller). Thus, in this example, the two chamber configuration leads to better and less sensitive tool performance.

6. SUMMARY AND CONCLUSIONS

This paper has presented two integrated models for understanding cluster tool behavior. The first model is a network model that evaluates the total lot processing time for a given sequence of activities. By including a manufacturing process model (in the form of an RSM), the model calculates the total lot processing time as a function of the operation times and the process parameter values. This model allows one to quantify the sensitivity of total lot processing time with respect to process times and process parameters.

In addition, we have constructed an integrated simulation model that includes a process model. If the cluster tool uses a rule to sequence wafer movements, one can use the simulation to evaluate the impact of process changes, including changes to product characteristics and changes to process parameter values. In addition, one can construct an integrated network model to quantify the sensitivity of total lot processing time with respect to process times and process parameters in a specific scenario.

The examples presented here illustrate the types of insights that one can gain from using such methods. Under certain conditions, modifying the process parameter values will change the total lot processing time very little, while under other conditions, the change will be great. This approach provides a vehicle for direct feedback of manufacturing metrics to process engineers involved in process alterations or tuning. Additionally, since the cluster tool's maximum throughput depends upon the process parameters, the range of possible process recipes with acceptable throughput can be directly identified and using when evaluating a tool's potential performance and cost-effectiveness.

Also, note that the critical path includes multiple operations (e.g., wafer moves and orientations and depositions). Increasing any operation's time will increase the lot makespan. Calling one operation (or the corresponding resource) the "bottleneck" may be inappropriate. However, it would be reasonable to say that one operation is "more critical" than another if the critical path includes the first operation more often than the second.

The models presented here considered only changes to a single process. However, in some cases, multiple characteristics may change simultaneously. For instance, changing process parameter values can also change the product quality and the tool overhead time (e.g., the time for establishing the correct pressure and temperature and for venting gases and cooling). One could modify the models presented here to analyze such types of behavior. In addition, we have presented a single example of a specific cluster tool configuration. We emphasize that one can apply this methodology to evaluate the impact of process changes in other tool configurations.

Future work will consider methods that can find good operation sequences for cluster tool configurations. In addition, we look to construct more sophisticated models (for larger and more complex

systems) that describe the relationship between process parameters and manufacturing system performance.

Acknowledgements

This research was supported by the National Science Foundation and the Semiconductor Research Corporation. We thank our collaborators Michael Fu and Steven Marcus for their useful insights.

Appendix A. Activity Sequence

Tool configuration: One loadlock (LL) with 20 wafers. One chamber (OD) for orientation and degassing. Two chambers (CVDA and CVDB) for W CVD. Pump down time = 400 seconds. Each wafer handler move requires 5 seconds. Each OD requires 10 seconds.

Pump down tool
Move wafer 1 from LL to OD. Begin wafer 1 OD.
When wafer 1 OD ends, move wafer 1 from OD to CVDA. Begin wafer 1 CVD.
Move wafer handler to LL.
Move wafer 2 from LL to OD. Begin wafer 2 OD.
When wafer 2 OD ends, move wafer 2 from OD to CVDB. Begin wafer 2 CVD.

REPEAT NEXT 12 MOVES UNTIL WAFER 20 BEGINS CVD.

Move wafer handler to LL.
Move wafer 3 from LL to OD. Begin wafer 3 OD.
When wafer 1 CVD ends, move wafer handler to CVDA.
Move wafer 1 from CVDA to LL.
When wafer 3 OD ends, move wafer handler to OD.
Move wafer 3 from OD to CVDA. Begin wafer 3 CVD.
Move wafer handler to LL.
Move wafer 4 from LL to OD. Begin wafer 4 OD.
When wafer 2 CVD ends, move wafer handler to CVDB.
Move wafer 2 from CVDB to LL.
When wafer 4 OD ends, move wafer handler to OD.
Move wafer 4 from OD to CVDB. Begin wafer 4 CVD.

When wafer 19 CVD ends, move wafer handler to CVDA.
Move wafer 19 from CVDA to LL.
When wafer 20 CVD ends, move wafer handler to CVDB.
Move wafer 20 from CVDB to LL.
End.

Bibliography

Atherton, R.W., F.T. Turner, L.F. Atherton, and M.A. Pool, "Performance analysis of multi-process semiconductor manufacturing equipment," Proceedings, pages 131-136, IEEE/SEMI Advanced Semiconductor Manufacturing Conference, 1990.

Box, G.E.P., and N.R. Draper, *Empirical Model-Building and Response Surfaces*, Wiley, New York, 1987.

Connors, Daniel P., Gerald E. Feigin, and David D. Yao, "A queueing network model for semiconductor manufacturing," *IEEE Transactions on Semiconductor Manufacturing*, Volume 9, Number 3, pages 412-427, 1996.

Dance, Daren L., Devid W. Jimenez, and Alan L. Levine, "Understanding equipment cost-of-ownership," *Semiconductor International*, pages 117-122, July, 1998.

Dhudshia, Vallabh H., and Clyde Hepner, "Cluster tool performance tracking," *Future Fab International*, Volume 1, pages 173-175, Technology Publishing Ltd., London, 1996

LeBaron, H.T., and M. Pool, "The simulation of cluster tools: a new semiconductor manufacturing technology," Proceedings, pages 907-912, Winter Simulation Conference, Buena Vista, Florida, December, 1994.

Lopez, Marcel J., and Samuel C. Wood, "Performance models of systems of multiple cluster tools," 1996 *IEEE/CPMT International Electronics Manufacturing Technology Symposium.*

Lopez, Marcel J., and Samuel C. Wood, "Systems of multiple cluster tools: configuration and performance under perfect reliability," *IEEE Transactions on Semiconductor Manufacturing*, Volume 11, Number 3, pages 465-474, 1998.

Mauer, J., and R. Schelasin, "Using simulation to analyze integrated tool performance in semiconductor manufacturing," *Microelectronics Engineering*, Volume 25, pages 239-246, 1994.

Murphy, Robert, Puneet Saxena, William Levinson, "Use OEE; don't let OEE use you," *Semiconductor International*, pages 125-132, September, 1996.

Perkinson, Terry L., Peter K. McLary, Ronald S. Gyurcsik, and Ralph K. Cavin, "Single-wafer cluster tool performance: an analysis of throughput," *IEEE Transactions on Semiconductor Manufacturing*, Volume 7, Number 3, pages 369-373, 1994.

Semiconductor Business News, "Applied Materials, Novellus, LAM Research lead cluster tool market," CMP Media Inc., 1998.

Stefani, Jerry A., Scott Poarch, Sharad Saxena, Purnendu K. Mozumder, "Advanced process control of a CVD tungsten reactor," *IEEE Transactions on Semiconductor Manufacturing*, Volume 9, Number 3, pages 366-383, 1996.

Srinivasan, R.S., "Modeling and performance analysis of cluster tools using petri nets," *IEEE Transactions on Semiconductor Manufacturing*, Volume 11, Number 3, pages 394-403, 1998.

Venkatesh, Srilakshmi, Rob Davenport, Pattie Foxhoven, and Jaim Nulman, "A steady-state throughput analysis of cluster tools: dual-blade versus single-blade robots," *IEEE Transactions on Semiconductor Manufacturing*, Volume 10, Number 4, pages 418-424, 1997.

Wood, S.C., "Adaptable manufacturing systems for integrated circuit production," Technical Report ICL 94-032, Integrated Circuits Laboratory, Stanford University, 1994.

Wood, Samuel C., "Simple performance models for integrated processing tools," *IEEE Transactions on Semiconductor Manufacturing*, Volume 9, Number 3, pages 320-328, 1996.