

# Job Sequencing and Optimal Scheduling for Tri-Temperature Lot Sizes in Semiconductor Testing

**Yoshiki B. Kurata** †

Industrial Engineering Department  
Technological Institute of the Philippines, Cubao, Quezon City, Philippines  
Tel: (+63)02-911-09-64 loc 357, Email: [yoshiki\\_kurata@yahoo.com](mailto:yoshiki_kurata@yahoo.com)

**Herwina Richelle C. Andres**

Operations Research – Equipment and Facilities Engineering  
Analog Devices, Inc., Cavite, Philippines  
Tel: (+63)02-867-70-30 loc 7855, Email: [HerwinaRichelle.Andres@analog.com](mailto:HerwinaRichelle.Andres@analog.com)

**Abstract.** A strategy in reducing production costs is to keep the capital costs low as semiconductor test equipment cost millions of dollars. It is essential that optimal usage of each is practiced through robust scheduling and sequencing of lots. The study emphasizes on solving the job sequencing and optimal scheduling for tri-temperature lot sizes arising in semiconductor testing. It aims to create an optimal model and matrix for the given processing of the tri-temperature lot sizes, specifically the ambient-hot-cold temperature sequence to minimize the time allocated in every lot size (processing time) and the cycle time encountered from which the number of machines can be calculated. The study utilized the evolutionary solver in getting the minimum cycle time for the 75 lots. Significant result is that the best option for the tri-temperature lots is to have a dedicated machine per temperature stage. At this optimal scenario, the number of machines needed to be used to complete the processing of the selected 75 lots in a week was determined to be at 3 machines per stage. Sensitivity analysis was made to give the operations a look of the impact once machines availability swings.

**Keywords:** scheduling, sequencing, waiting times, evolutionary solver, makespan

## 1. INTRODUCTION

The rapid growth of semiconductor manufacturing company is one of the main cause for an increasing demand of variant products that are commonly used in most of all modern devices. Semiconductors plays a significant impact for most electronic products even when modest number of products are being produced in the factory. Computer chips and microprocessors are some products that most semiconductor manufacturers are producing that uses etching thousands or even millions of transistors on several layers of semiconductor wafers. Although there are different approaches being implemented for every type of electronic device, wafer manufacturing, water fabrication, packaging and final testing are the common processes done in the production.

At computer-controlled test stations and at various temperatures, semiconductor devices undergo series of test processes. Each test process comprises of setup operations and processing operations which occurs in a specified order for

devices that result to some precedence constraints for the schedule. The assignment of devices to test stations and the process of conducting test operations greatly affects the required time in finishing the entire process which results in sequence dependent setup times. This study aims to develop realistic model of the semiconductor test scheduling problem and provide heuristics for scheduling the precedence constrained test operations with sequence dependent setup times.

Lot sequencing in the production lines are also of a high concern to the planners and production superiors. Through the study, an optimal sequencing will be established to allow the concerned manpower in dealing with constrained capacities.

### 1.1 Problem Statement

In the company's quality assurance department, finished semiconductors are subjected to testing operations of three different temperature conditions – ambient, cold, and hot temperatures. The semiconductors are tested in lots and each

lot must be tested under each temperature condition strictly in that sequence. The company currently tests the lots using three machines, one for each temperature set with a span of 168 hours or 7 days for 75 lots being tested.

Scheduling is a proper way of allocating shared resources to competing activities over a period of time which is one of the significant literature in the field of operations research. This study highlights the on-going investigation of a specific machine scheduling problem wherein jobs and machine represents activities and resources, respectively. In addition, each machine is considered to process one job at a time.

### 1.2 Objectives of the Study

Job sequencing and scheduling is an important issue in increasing the efficiency of manufacturing systems. However, the gap between theoretical models and industry practices that are being identified by many researches which make them conclude that by using realistic scheduling model with better algorithms can be beneficial for a fast production processes.

For the study, the researchers aim to create an optimal model and matrix for the given processing of the tri-temperature lot sizes, specifically ambient, cold, and hot temperatures. Its purpose is to minimize the time it takes to complete the entire schedule (i.e. schedule makespan of less than 168 hours).

Specifically, the study also aims to:

- Minimize the time allocated on every lot size, processing time, and the cycle time encountered in the tri-temperature lot size;
- Translate the existing model and consider using one machine or three machines and determine if it will lead to a more optimal testing operations;
- Reduce the production costs and increase the throughput while meeting delivery dates; and
- Conclude the optimal set of configurations with the set of machines with respect to the existing temperature sequence to compare on the machine requirements being recommended.

### 1.3 Scope and Limitations of the Study

In order for the researchers to successfully model the problem, test environment characteristics were considered. These characteristics included in this study are (1) precedence constraints, (2) setup times, (3) equipment utilized during the testing, and (4) production demands. Also, the study is bounded by lots with part names requiring three temperature tests in the sequence of Ambient, Cold, and Hot.

## 2. LITERATURE REVIEW

Lu (2001) explained the process of testing semiconductors consists of several test stations with one brand-work center. Each task is done at each test station at different temperatures. After performing each task, it will be transferred to the brand-work center for decomposition method. The method first decomposes the job shop into individual work centers each shop are assumed to be known in advance.

Job-shop scheduling is one technique that can be used to manage the ordering between operations that can be processed using one machine which includes fix precedence between operations (Fleming, 1997). For job scheduling problems that are being processed using different machines, minimization of makespan must be used wherein it is the sum of job processing and setup times while utilizing the machine (Amiya, 2006).

Pinedo (2008) define scheduling as often done interactively via a decision support system that is installed on workstations linked to the ERP system as scheduling plays a vital role in the manufacturing sector. Figure 1 shows the importance of scheduling in the sequence of manufacturing system's information flow.

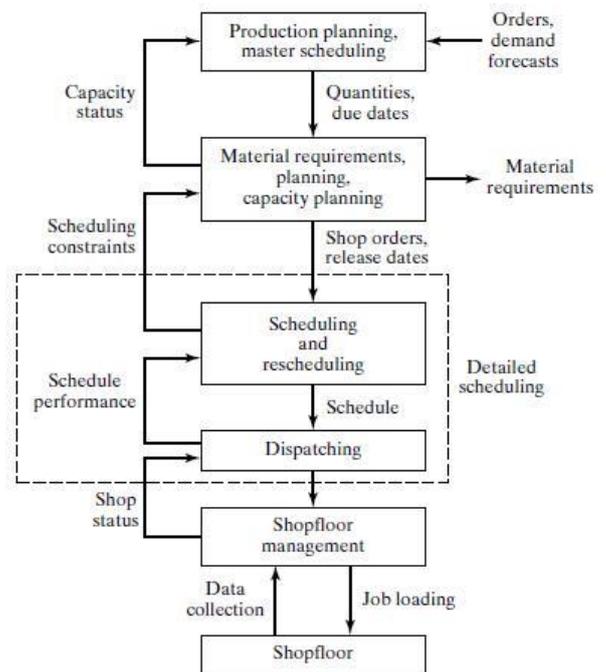


Figure 1. Information Flow Diagram in a Manufacturing System

According to the study made by Ullah (2000), a typical job shop are the ones with m machines and n jobs that should be processed. For each job, it requires one operation for each machine that are put in specific order but may vary depending

on the job being performed (Fleming 1997). On the other hand, real job shops are considered to be more complicated since some jobs are not being processed using machines but still, machines are utilized more than once. In effect, it is presumed that workflow is not unidirectional in a typical job shop. Any given machine may observe new jobs arriving from outside the shop (as new inputs), and from other machines within the shop (WIP), the same machine may be the last machine for a particular job, or it may be an intermediate processing step. Thus, the workflow can be illustrated as in figure 2 below.

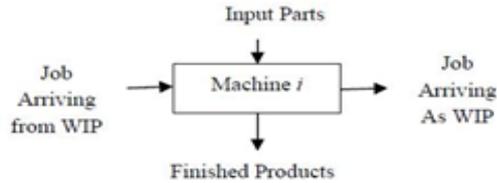


Figure 2. Job Workflow

Sequencing is a process of having the job orders set into priorities which can be useful in manufacturing process. These priority rules help provide direction for jobs to be performed efficiently which can lead to ranking job loading decisions in manufacturing centers. These rules include the following:

- **DD - Due Date of a job.** The job having the earliest due date are prioritized.
- **FCFS - First Come, First Served.** The first job that reaches the production center are processed first.
- **LPT - Longest Processing Time.** Focus on the job with the longest processing time.
- **PCO - Preferred Customer Order.** Jobs considered as customer preference should be prioritized.
- **SPT - Shortest Processing Time.** Focus on the job with the shortest processing time.

## 2.1 Theoretical Framework

The research focuses on scheduling machines and sequencing lots within the testing operations. These two concepts have a direct impact on the number of machines required for testing, the tri-temperature test sequence, each lot's waiting time, and ultimately on the makespan. The intermediate variable number of machines currently amounts to three machines or one machine per temperature test. The intermediate variable test sequence is affected by a policy which is not necessarily part of the system and dictates that a lot must first undergo the ambient temperature test, the cold temperature second, and the hot temperature last. The intermediate variable waiting time also can be affected by a

factor not necessarily part of the system, specifically downtime. The study approximates that 30% of the time, downtime is experienced.

The dependent variable makespan will be the study's measure of improvement. The problem is taken from the testing operations, spanning 168 hours. The lots taken from the observation will be the primary concern in scheduling and sequencing. The model built for this particular scenario can be translated to the scheduling and sequencing of all future testing operations.

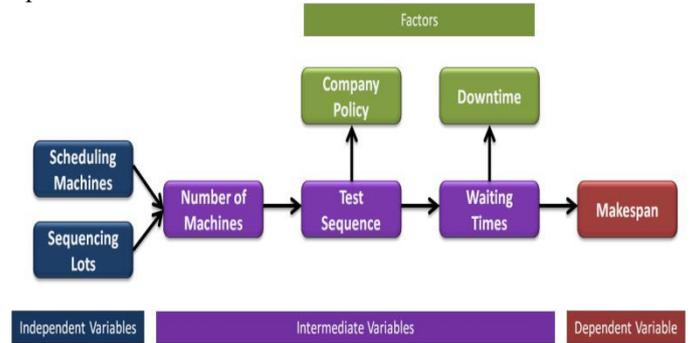


Figure 3. Theoretical Framework

## 3. RESEARCH DESIGN AND METHODOLOGY

The following formulas were utilized in the study.

$$\text{Mean Flow Time: } \bar{F} = \frac{l}{n} \sum_{j=1}^n F_j \quad (1)$$

$$\text{Mean Tardiness: } \bar{T} = \frac{l}{n} \sum_{j=1}^n T_j \quad (2)$$

$$\text{Maximum Flow Time: } F_{max} = \max_{1 \leq j \leq n} \{F_j\} \quad (3)$$

$$\text{Maximum Tardiness: } T_{max} = \max_{1 \leq j \leq n} \{T_j\} \quad (4)$$

$$\text{Number of Tardy Job: } N_T = \sum_{j=1}^n f(T_j) \quad (5)$$

where  $f(T_j) = 1$ , If  $T_j > 0$  and  $f(T_j) = 0$  otherwise.

**Flow time ( $F_j$ )** - Total time allotted for job  $j$  in a system.

**Makespan** - Total processing time for all jobs.

**Lateness ( $L_j$ )** - Total time spent in completing the job that differs from the due date. It is called Positive lateness if the job is completed after the due date.

**Tardiness ( $T_j$ )** - It is the lateness of the job  $j$  if it fails to meet its due date, or zero. *i.e*  $T_j = \max \{0, L_j\}$ .

The critical ratio method is the ration between the time spent until due date and the required job processing time. It is considered to be a relative measure of critical job order priority when jobs are combined with other jobs that are put on hold. Thus, critical ratio prioritizes jobs that must be done in order to meet the predetermined shipping schedule. If the critical ratio is less than 1, then the job is considered to be falling behind the shipping schedule but if the critical ratio is greater than 1, then the job is considered to be ahead of the shipping schedule. If the critical ration is equal to 1, then the job is considered to be done and processed on time. Critical ratio uses the following formula:

$$\text{Critical Ratio} = \frac{\text{Remaining Time}}{\text{Remaining Process Time}} \quad (6)$$

$$\text{Critical Ratio} = \frac{(\text{Due Date} - \text{Today's Date})}{\text{Days of Remaining Process Time}} \quad (7)$$

The static flow shop-sequencing problem is a process of determining the best sequence on each machine in the flow shop. It is called 'Permutation' flow shop if the jobs have the same order of sequences on all the machines. In this case, the first machine is considered to be the main cause of the problem with the addition of extra constraint of same job sequence at each machine. Ironically, this problem is a little harder to address than the more general case, even though this might seem as a small part (sub problem) of the general case.

Various objectives can be used to determine the quality of the sequence, but the majority of the research considers the minimization of makespan (*i.e.*, the total completion time of the entire list of jobs) as the primary objective but for some other researches, objectives are flow time related (*e.g.*, minimal mean flow time), due-date related (*e.g.*, minimal maximum lateness), and cost related (*e.g.*, minimal total production cost). (Palekar, 2001).

In comparison, both sequencing and scheduling are concerned with the optimal allocation of resources to activities over a period of time that could be infinite or finite. Since early 1950's, scheduling and sequencing has been the main subject of extensive research that created an impressive amount of literature for further studies. Thus, any discussion of the available material is bound to be selective. (Palekar, 2001).

Semiconductor testing for device characterization is one

way of evaluating the future performance of devices under near failure environmental conditions. In addition, burn-in is also one way to eliminate failure on their lots that can be used by semiconductor manufacturers. Maintaining the temperature of the device at a certain period which must be switched quickly to ensure quality of work done by the machine.

Stringent testing is one solution for proper conditioning of semiconductors at higher temperature ranges. This test which is sometimes called Tri-temperature test requires 40°C to 150°C to maintain the absolute maximum temperatures. Tri-temperatures are of three parts – the Ambient Temperature with 25oC, Hot Temperature with a temperature of 150oC and the Cold Temperature with (-40oC).

Environmental Thermal Chamber is used to test devices at varying temperatures but should control the temperatures by passing moving air over heating and cooling elements. The change in temperature were slow to respond which causes the temperature to vary in the chamber which may lead to inefficiency of temperature testing and may limit throughput in a manufacturing environment.

#### 4. RESULTS AND DISCUSSION

Minimizing the cycle time of the lots to be processed is the main objective of the study. This led the researchers to conduct brainstorming on what would be the best approach to deal with the problem and to provide optimal and timely solutions and recommendations for the problem of the company. Upon discussion and consideration of different alternatives, the group had decided to do a mathematical approach on solving the problem for best results.

The research formulated the sequencing model to contain non-smooth functions that can be solved using evolutionary solver. The non-smooth function weakens the performance of linear and nonlinear solvers but the existence solution procedure suitable for non-smooth functions can be more flexible in building models than that of linear and nonlinear solvers. In particular, Excel functions such as the IF function represents some logical choices. Another inclusion are some mathematical functions like as ABS, MIN, MAX, CEILING, FLOOR, ROUND, and INT. (Although it is sometimes possible to avoid using these functions directly, doing so may require the use of binary variables or auxiliary variables in cumbersome or unusual ways). Also, spreadsheet-oriented functions like CHOOSE, COUNTIF, INDEX, and LOOKUP can be beneficial in solving spreadsheet calculations which are good ways to interpret non-smooth functions.

The modeling flexibility are costly and since evolutionary solver gives no assumptions about the objective function, it is incapable of identifying the optimal solution. Generally, it only

conducts searching which is followed by comparing the solutions encountered and automatically stops if the results do not make any difference or does not find any improvements. Although evolutionary solver does not provide optimal solution but for some other problems, it may deliver a good solution which is called heuristic procedure.

#### 4.1 Model Assumptions

The assumptions used in this study are consistent with general studies in scheduling and are as follows:

1. Different solutions may be calculated when running the evolutionary solver twice due to some randomized steps.
2. The evolutionary solver may provide multiple solutions to the problem but if the solutions are not improving, it automatically stops and considers the best member among all solutions to be the optimal solution.
3. The evolutionary solver does not make any difference when finding for improvements or solution even with the presence of non-smooth functions which are likely different in the case of linear and nonlinear solvers.
4. User-controlled options are evident when evolutionary solver are being used.
5. The evolutionary solver does not always provide optimal solution so some judgment must be considered.
6. A job may not be processed on more than one stage at a time.
7. The number of jobs,  $X_i$ , is known.
8. All lots arrive at time zero.
9. One-lot processing is done.
10. All lots follow the same sequence. Only the sequence of testing operations under temperature conditions of ambient, cold, and hot will be considered in the study.
11. No operation may be pre-empted (once started it must be processed to completion).
12. No machine may process more than one job at a time.
13. In the case of using two machines or one machine, a stabilization time is incorporated into the model (i.e. added to the testing time of the preceding temperature test).
14. Available hours are assumed to be 70% of the total number of hours.

#### 4.2 Nomenclatures

##### 4.2.1 Sequence of Lots

Decision Variables

Let:  $X_i$  = lot  $i$ , ( $i = 1, 2, \dots, 75$ )

Objective Function

$$\text{Minimize } Z = M$$

$$x_i = \text{alldifferent}$$

$$M_j \geq 1$$

$$x_i \geq 0$$

$$x_i \text{ is an integer}$$

wherein:

$Z$  = total time to complete the testing of lot parameters

*alldifferent* = evolutionary solver parameter

$M$  = makespan

$i$  = number of lots

$j$  = number of stages

$M_j$  = number of machines at stage  $j$  ( $j = 1, 2, 3$ )

The testing times and waiting times of each lot are indexed by  $x_i$  in the excel worksheet. The INDEX function is used to reference testing times and waiting times with its corresponding lot. The constraints reflect how lots are ensured to take on integer values with no duplicates through the *alldifferent* constraint, that there is at least one machine in each stage, and that no lot can be assumed a negative value.

##### 4.2.2 Machine Requirement

Decision Variables

Let:  $X_i$  = number of machines per option

Objective Function

$$\text{Minimize } X = C_1X_1 + C_2X_2 + C_3X_3$$

*integer* = lots must be an integer value

MWAH = 117.6 hours

Machine Availability = 70%

wherein:

$X$  = total cost of testing

$C_1, C_2,$  and  $C_3$  = costs per machine configuration

MWAH = machine weekly available hours

##### 4.3 Model Scenarios

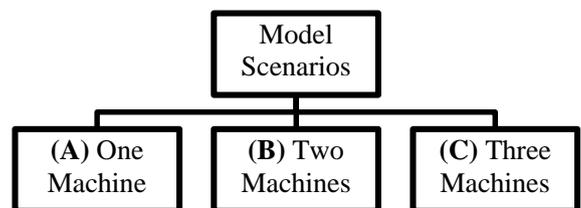


Figure 4: Model Scenarios

#### 4.4 Model Results

The target output of this study is to determine and to minimize the total cost of testing for the ambient-cold-hot sequence. Upon solving, the researchers were able to compute the optimal number of machines and the costs incurred per option given the number of lots per week.

In addition, the model created will calculate the minimum time to complete the testing of the lots at the different temperatures. Makespan time is the summation of waiting time, stabilization time and processing time. This is summarized in Figure 5. Total time for Scenario A is significantly higher and can be attributed mostly to the waiting times and the stabilization. Option A to B reduced the total time by 98%; option B to C further reduced the total time by 60%.

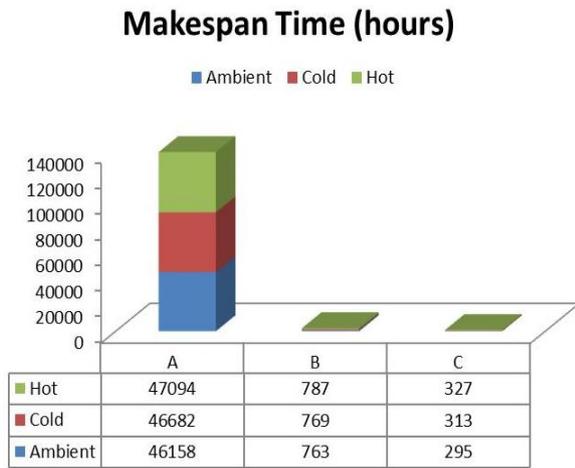


Figure 5. Makespan Time

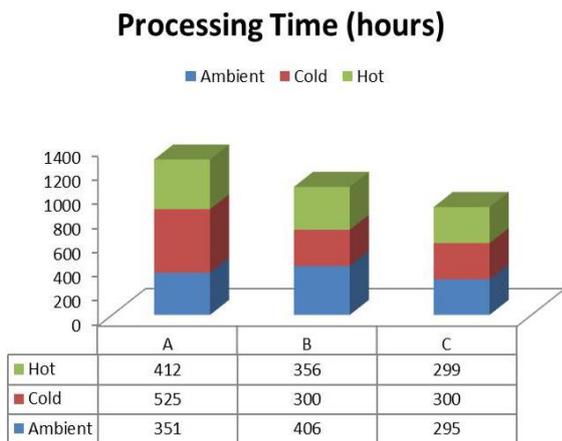


Figure 6: Total Processing Time

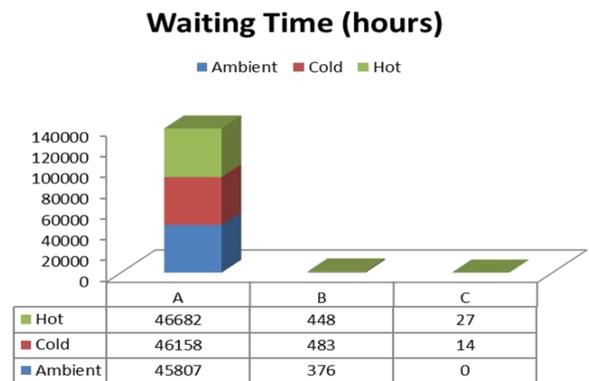
Figure 6 presents the total processing times per temperature of the 75 lots. Option A has the highest as this

takes into account the three temperature stabilizations for each lot. Option B has the second highest since only Ambient and Hot tests are being shared in one machine. Option C has the lowest as no temperature stabilization occurs as each temperature has its own machine. Processing time from A to B reduced by 18%; from B to C reduced by 16%.

Figure 7. Total Waiting Time

Figure 7 shows the data for the waiting time. Waiting time as a function of queue has the highest time for option A. presents the total processing times per temperature of the 75 lots. The sharing of Ambient and Hot affects the waiting time even for Cold as stabilization time lengthens the whole process. Option C has the least value as the waiting only occurs when the processing of the previous lot of the preceding stage is longer than the previous lot being processed by the succeeding stage. Total waiting time from option A to B reduced by 99%; from B to C, by 97%. The sequence of each scenario did not vary since their lot processing times has a standard deviation only of 1.50 and a variance of 2.3. These were calculated using Stat:Fit. See Appendix A.

The calculated makespan time of the three options, exceeded the 168 hours for one-week processing. The manufacturing operations have the choice to calculate the required number of machines to complete the 75 lots in one-



week time. Figure 8 presents the calculated number of machines for each stage in each of the options using the machine requirement linear programming model. Total time per week is multiplied by 70% of machine availability resulting to 117.6 hours per week per machine. Option A will need 13 machines to have the 75 lots be completed in one-week time, 10 machines for option B, and 9 machines for option C.

Table 1 presents the cost of the three options considered. This is on the assumption that single functionality equipment will be cost only 0.5 Million \$ and as it increases another functionality, the cost will be 0.7 Million \$ and for the three temperature machine capability, the cost will be 0.9 Million \$. Among the three options, Option C will incur the cheapest cost with the minimum number of machines. This count will give

the manufacturing floor the minimum machine space allotment as well.

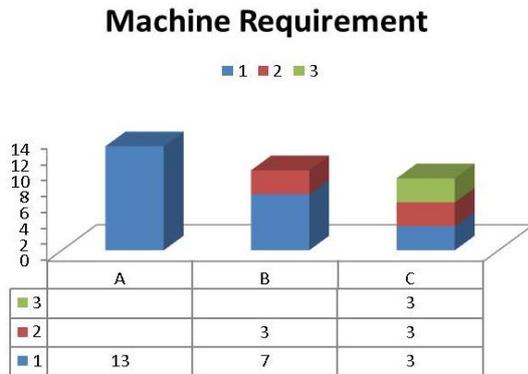


Figure 8. Machine Requirement

Table 1. Number of Machines per Option with Allocation of Costs to be Incurred

Availability	Options	No. of Machines			Cost (Million \$)			Total Machines	Total Cost (Million \$)
		1	2	3	1	2	3		
75%	A	11	-	-	0.9			11	9.9
	B	7	3	-	0.7	0.5		10	6.4
	C	3	3	3	0.5	0.5	0.5	9	4.5
70%	A	13	-	-	0.9			13	11.7
	B	7	3	-	0.7	0.5		10	6.4
	C	3	3	3	0.5	0.5	0.5	9	4.5
65%	A	13	-	-	0.9			13	11.7
	B	8	3	-	0.7	0.5		11	7.1
	C	3	3	3	0.5	0.5	0.5	9	4.5
60%	A	15	-	-	0.9			15	13.5
	B	9	3	-	0.7	0.5		12	7.8
	C	3	3	3	0.5	0.5	0.5	9	4.5
55%	A	15	-	-	0.9			15	13.5
	B	9	4	-	0.7	0.5		13	8.3
	C	4	4	4	0.5	0.5	0.5	12	6

probability of occurrence is being handled in descending order. In practice, the assessment process can be difficult but balancing resources used to eliminate between risks that has high probability of occurrence but lower loss and high loss but lower probability of occurrence can often be mishandled.

This study is mainly focusing on the availability of machine performance and the lots to be scheduled for testing. According to research, in order to determine the average cycle time of lots, the operational time variability must be used. Machine breakdowns, setup, and operator availability may be factors of variability, with 70% machine availability used. The other parameter, volume, is highly variable on a week-to-week basis.

A sensitivity analysis was made for the machine availability where it can go as high as 75% and as low as 55%. This will have a significant impact on the weekly available hours and eventually on the number of machines that will be

needed. Table 3 shows the simulated data should machine availability value vary. This allows the management how significant it would be to maintain at 70% or improve it even.

Table 3. Machine Requirement – Sensitivity

Availability	Options	No. of Machines			Cost (Million \$)			Total Machines	Total Cost (Million \$)
		1	2	3	1	2	3		
70%	A	13	-	-	0.9			13	11.7
	B	7	3	-	0.7	0.5		10	6.4
	C	3	3	3	0.5	0.5	0.5	9	4.5

Table 4 presents the percentage differences of each option on different machine availability using the total machines as the baseline. Improving the machine availability by 5% will only have a 15% reduction in the number of machine usage for Option A, but there will be no effect for Options B and C. Bringing the availability down by 5% from the baseline will increase the machine requirement for Option B by 10%. Moreover, this highlights that Option C would still have an insignificant effect on the total number of machines should the availability go down up to 60%.

Table 4: Machine Requirement Percentage Difference - Sensitivity

Availability	A	B	C
75%	-15%	0%	0%
70%	0%	0%	0%
65%	0%	10%	0%
60%	15%	20%	0%
55%	15%	30%	33%

Composite Risk Index (CRI) is the product of the Impact of Risk Event (IRI) and the Probability of Occurrence (PO) and is given by formula shown below. It quantifies the impact by integrating the impact and the occurrence. It is suggested that both be rated from 1 to 5 where 5 is the highest. IRI was based on the sensitivity analysis on Table 3 and the probability of occurrence was based on the actual performances of the machines.

$$CRI = IRI \times PO$$

wherein:

CRI – Composite Risk Index

IRI – Impact of Risk Event

PO – Probability of Occurrence

Table 5 summarizes the calculated CRI values for each of the availability analyzed. This study assumes that both IRI and PO will be true for Options A, B and C. It is noted that both 65% and 60% have the highest value of 12, followed by 55% with 10; next is the baseline for the availability, 70% with a score of 5; lastly, 75% with a score of 2.

Table 5. Composite Risk Index

Availability	Composite Risk Index (CRI)		
	IRI	PO	CRI
75%	2	1	2
70%	1	5	5
65%	3	4	12
60%	4	3	12
55%	5	2	10

Presenting the sensitivity analysis allows the management to have a quantifiable understanding on the impact to the total number of machines. Activities on Availability improvement can be generated to keep the values at an acceptable level.

## REFERENCES

- Amiya, K., Job Sequencing Rules for Minimizing the Expected Makespan in Flexible Machines. *European Journal of Operations Research* (pp. 274-288), 1996
- Booz, Allen and Hamilton, Earned Value Management Tutorial Module 2: Work Breakdown Structure. Office of Science, Tools and Resources for Project Management., 2011
- Fleming, A.M.S., Job-shop Scheduling – Generic Algorithms in Engineering Systems. The Institute of Electrical Engineers, 1997
- French, S. (1982). Sequencing and Scheduling: An Introduction to the Mathematics of the Job-Shop, (Ellis Horwood Limited: Chichester, England).
- Hubbard, Douglas, The Failure of Risk Management: Why It's Broken and How to Fix It. John Wiley and Sons, Inc., 2009
- Jacobs, J. H., Etman, F. P., van Campen, E. J. J., Rooda, J. E., Characterization of Operational Time Variability Using Effective Process Times. *IEEE Transactions on Semiconductor Manufacturing*, Vol. 16, No. 3, 2003.
- Johnson and Montgomery, Operations Research in Production Planning, Scheduling and Control. John Wiley and Sons, Inc., 2008
- Lu, Yufang, Scheduling of Wafer Test Processes in Semiconductor Manufacturing, 2001
- Palekar, N., Job Sequencing and WIP Level Determination in a Cycle., 2001
- Pinedo, M., Scheduling: Theory, Algorithms, and Systems (3rd Edition). Prentice Hall, 2008

Stevenson, W., Operations Management: An Asian Perspective

Ullah, S., Review on Job-shop and Flow-shop Scheduling using Multi-Criteria Decision. Bangladesh University of Engineering and Technology., 2000

## Appendix A. Stat:Fit Data Description

