CHAPTER V

SINGLE LOCATION

SPARE PARTS

INVENTORY SYSTEM

- A CASE STUDY
CHAPTER V

SINGLE LOCATION SPARE PARTS INVENTORY SYSTEM - A CASE STUDY

5.1 Introduction

In this chapter a case study of leading Semiconductor equipment Manufacturer Company is taken where single location spare part inventory system is considered. It manufactures systems that perform most of the primary steps in the chip fabrication process. The main customers of the company are semiconductor wafer manufacturers and semiconductor integrated circuit manufacturers, which either uses the chips they manufacture in their own products or sell them to other companies downstream. The company is at the top of the supply chain for most personal computers and other high technology products.

Semiconductor systems are very expensive investments and are very critical to operations of many high technology companies. Unused semiconductor manufacturing capacity due to equipment failures is very costly. In order to provide spare parts and service to customers for equipment failures and scheduled maintenances, the company has an extensive spare parts network. The company also has agreements with its leading customers where it manages the stock rooms (for all or a group of spare parts) in customer facilities (some of these are consignments). Three distribution centers: at various locations [63]. The depot locations are such that they can provide a 24-hours service to customers (those who do not have stock rooms operated by the company) for equipment failures ("down orders"). However, the distribution centers may also be used as a primary source for down orders for certain customers. In addition, the
distribution center provides a second level of support for down orders that cannot be satisfied from the local depots. The customers also demand spare parts to be used in their scheduled maintenance activities ("lead time orders").

The primary source to meet these demands is usually the distribution centers. However local depots can also be used for this purpose for certain customers.

Both types of customer orders (down and lead time) go through an order fulfillment engine which searches for available inventory in different locations according to a search sequence specific to each customer. However, the down orders need to be satisfied immediately (their request date is the date of order creation), while the lead time orders need to be satisfied at a future date. A depot may be facing down and lead time demand from a variety of customers, while a distribution center may be facing down and lead time demand from external customers in addition to the "replenishment orders" requested by internal customers: the depots and stock rooms managed by the company. The operations of this complex network is further complicated by a vast number of parts composed of consumables and non-consumables (more than 50,000 active parts need to be managed) and varying service level requirements by different customers.

While providing an implementable and "good" solution for the whole spares network is a proven challenge. We focus on an important issue where improvements can provide immediate and significant benefits. In the existing practice, locations that are facing different types of demand (down, lead time or replenishment), the company targets to achieve the maximum of the service level requirements while considering the aggregated demand. Moreover, the company does not recognize the possible demand lead times (the difference between requested date and
ship date in excess of transportation time) for lead time orders and possible slacks (the difference between the replenishment lead time the company uses for planning downstream locations and transportation lead time) for replenishment orders. Obviously, this approach is inefficient. We suggest an inventory model that recognizes both the demand lead times and multiple demand classes, and allows for providing differentiated service levels through rationing.

We consider a single location spare part inventory system which faces two classes of demand arrivals with different criticality. The down orders which result from the equipment failures of customers are assumed to constitute the high priority, i.e., critical class, whereas the maintenance orders are assumed to constitute the low priority, i.e., non-critical class. Demand arrivals of the critical and non-critical class are both assumed to be Poisson with rates of \( \lambda_c \) and \( \lambda_n \), respectively. Both arrivals are satisfied from the same pool of inventory which is controlled by a base stock policy with a base stock level \( S \). Therefore, each demand arrival triggers a replenishment order with a deterministic lead time of \( L \). In addition, the demand from the non-critical class allows a deterministic demand lead time of \( T \), which is called the demand lead time.

Given this system, our purpose is to determine the minimum inventory investment which satisfies the service requirements for both classes. Furthermore, we assume the ownership of on-order inventory and minimize expected inventory on hand plus on expected inventory on order. Note that unlike the case in a standard continuous review (\( S - 1, S \)) policy, the inventory position is not always equal to \( S \) in this system with demand lead times. The expected inventory position is in fact equal to \( S + \lambda_n \times T \), where the second term is due to the outstanding replenishment orders for the non-critical demand class that are yet not due.
5.2 Simulation Model

In this section we present the model of our simulation study. We have coded a discrete event simulation algorithm in pictorial form with the next-event time advance mechanism to advance the simulation clock. The input parameters are $S$, $Sc$, $c$, $n$, $L$ and $T$ and the random output parameters are $\beta c$ and $\beta n$.

We have modeled the simulation with five events. Besides the end simulation event which terminates the simulation run, we have four other events which are represented by the associated functions in the C++ code. Next we present the flow charts of these events.
Critical demand over

Internal Counter

Is $|O|>0$

Document hand inventory by one

Increase critical backorders by one

Scheduled an ordered arrival

Schedule next critical order

Return
Figure describes the critical demand event function. After a critical demand arrival first the counter for the cumulative number of critical demand arrivals is incremented by one. Then the on-hand inventory is checked to see whether or not this arrival can be satisfied immediately. If on-hand inventory is greater than zero and the critical arrival can be satisfied immediately the counter for satisfied critical customers is incremented by one while the on-hand inventory is decremented by one. Otherwise, the counter for critical backorders is incremented by one. Observe that every critical demand arrival event schedules a replenishment order arrival event for $L$ time units after since the inventory is controlled by a base stock policy. Also, it schedules the next critical arrival event.

A non-critical arrival event is similar to a critical arrival event. However the only counter that is updated is the counter for the cumulative number of non-critical arrivals (since all non-critical demand arrivals are accepted as they arrive).
This is because the due date of such an arrival is $T$ time units after its arrival. Thus another difference of the non-critical arrival event is that it also schedules this evaluation event.

A replenishment event merely represents the arrival of a replenishment order. Thus if there are any critical backorders the counter for critical backorders is decremented by one. If there is a non-critical backorder and the inventory on hand is at $S_c$ the counter for non-critical backorders is decremented by one. Otherwise this replenishment order is used to increment the on-hand inventory by one. A replenishment arrival event also schedules the next replenishment arrival event. An evaluation event merely determines whether a non-critical arrival from $T$ time units
before (i.e., one whose due date has arrived) will be satisfied or not. If
inventory on-hand is above $S_c$ the counter for satisfied non-critical
customers is incremented by one while the on-hand inventory is
decremented by one. Otherwise the counter for non-critical backorders is
incremented by one. An evaluation event also schedules the next
evaluation event.
Flow diagram of the replenishment order arrival event
Flow diagram of the evaluation event
The run time of the simulation is $10^7$ time units and there is one replication. We also tested our simulation model with batch-mean method with $10^5$ run time and 100 replications and show that the confidence intervals of our related output parameters are in the order of $0^5$. To verify the accuracy of our simulation with a single replication we choose

$$S = 12, \ S_c = 3,$$

$$\lambda = 4,$$

$$n = 8, \ L = 0.5$$

and $T = 0.1$

with the batch-mean method. To do this we divided the simulation into 100 replications of $10^5$ each. We assume independence of successive simulation runs which is acceptable considering the relatively long individual placation lengths of 105. As a result we see that the associated confidence intervals of our simulation outputs are

$$3.26 \times 10^{-5} \text{ and } 5.60 \times 10^{-4}$$

For the critical and non-critical service levels respectively. Having verified that these confidence intervals are indeed small we conclude that we can confidently use our output from the simulation model with one replication as an approximation for the associated service levels.

 Attempt to verify the spare part inventory system at a specific location. We have selected a depot that is serving a number of customers for both down demand and lead time demand. As described earlier, the depot is positioned to provide a 24-hours service to a specific list of customer locations for their down orders. Considering the shipment time around 24 hours to its customer locations, this means that the down orders need to be satisfied immediately from on stock inventory (i.e., demand lead time is zero). The depot is also used to support maintenance orders from the same list of customers. However, in this case, the parts do
not have to be shipped right away. The customers usually plan maintenance activities in advance, and demand lead times of 2 weeks are common and acceptable for such orders.

We selected 64 spare parts for our study. A summary of characteristics for these parts is given in Table 1. In order to ensure the appropriateness of the (S -1, S) inventory policy and the validity of the Poisson demand assumption, we include rather expensive and slow moving parts in our study. We used the demand history of a 12 months period in years 2004 and 2005 and include all requested orders (these could include orders that were not satisfied or canceled later) whose primary source is the depot we have selected. The ratio of critical orders to total orders varies at the part level. On the average, 52.2 % of a part’s demand is from down orders (i.e., critical demand)

<table>
<thead>
<tr>
<th></th>
<th>MIN</th>
<th>MAX</th>
<th>AVERAGE</th>
</tr>
</thead>
<tbody>
<tr>
<td>Part Cost($)</td>
<td>1,104</td>
<td>40,451</td>
<td>8,681</td>
</tr>
<tr>
<td>Critical annual demand</td>
<td>1</td>
<td>166</td>
<td>43.19</td>
</tr>
<tr>
<td>Non- critical annual demand</td>
<td>2</td>
<td>120</td>
<td>39.59</td>
</tr>
<tr>
<td>Total annual demand</td>
<td>41</td>
<td>212</td>
<td>82.78</td>
</tr>
<tr>
<td>Percentage of critical Demand</td>
<td>1.18</td>
<td>96.77</td>
<td>52.20</td>
</tr>
<tr>
<td>COGS($)</td>
<td>94,985</td>
<td>3,318,438</td>
<td>643,600</td>
</tr>
<tr>
<td>Lead time (days)</td>
<td>19</td>
<td>120</td>
<td>68.06</td>
</tr>
<tr>
<td>Lead time demand</td>
<td>10.06</td>
<td>19.65</td>
<td>13.99</td>
</tr>
</tbody>
</table>

Table 1
5.3 Part Characteristics

In the same 12 months period, these 64 spare parts had a sales volume of $41.2 million (in cost). $24.3 million (59.1 %) of this is generated by orders that are denoted by customers as down orders; $16.9 million (40.9 %) of this is generated by orders that are denoted by customers as lead time orders. We note again that with company’s current practice, the demand lead times are not recognized by the company and the safety stocks are set to satisfy service level requirements for the down orders while considering the total demand (down orders and lead time orders).

Testing number of service level targets for down demand as the company may change these targets depending on its negotiation with its customers that are served through this particular depot. Noticeable thing is that setting service level targets for lead time demand alone is not an established practice for the company, as the current practice provides a service level which is same as the service level targeted for down demand. Therefore, we also test a number of service level targets for lead time demand. However, we set the service level targets for the lead time demand always less than the service level targets for the down demand which is in line with company’s and customers’ expectations.

The analysis is done in three steps.

Step 1: We do not recognize the demand lead time for lead time orders and we do not apply any rationing. We simply calculate the minimum base stock levels that will satisfy the target service level requirement for the down orders considering the total demand (down demand plus lead time demand). This reflects the current practice in the company.

Step 2: We recognize the demand lead time for lead time orders; however do not use any rationing to provide differentiated service to the
two types of demand classes. We calculate the minimum base stock levels that will satisfy the target service level requirement for the down orders. This is similar to the model in and in fact fed time orders and down orders gets the same service level in this case.

**Step 3:** We recognize the demand lead times and also use rationing to provide differentiated service to two demand classes. In this analysis, we use the approximation and found that the procedure is also easy to implement which is important for a company that needs to manage 50,000 or more parts across more than 70 locations.

We demonstrate our analysis for a particular part in Table 2.

<table>
<thead>
<tr>
<th>Service Level (%)</th>
<th>No demand lead time</th>
<th>No rationing</th>
<th>Rationing</th>
</tr>
</thead>
<tbody>
<tr>
<td>Critical savings</td>
<td>Non critical</td>
<td>S</td>
<td>S</td>
</tr>
<tr>
<td>90</td>
<td>80</td>
<td>18</td>
<td>16</td>
</tr>
<tr>
<td>95</td>
<td>80</td>
<td>20</td>
<td>18</td>
</tr>
<tr>
<td>97</td>
<td>80</td>
<td>21</td>
<td>19</td>
</tr>
<tr>
<td>99</td>
<td>80</td>
<td>23</td>
<td>20</td>
</tr>
<tr>
<td>99</td>
<td>85</td>
<td>23</td>
<td>20</td>
</tr>
<tr>
<td>99</td>
<td>90</td>
<td>23</td>
<td>20</td>
</tr>
<tr>
<td>99</td>
<td>95</td>
<td>23</td>
<td>20</td>
</tr>
</tbody>
</table>

**Table 2**

The third column in the table shows the minimum base stock levels to satisfy the service level requirement for the critical demand class when one does not recognize the demand lead times and does not apply any rationing. The fourth column in the table shows the minimum base stock levels to satisfy the service level requirement for the critical demand class when one recognizes the demand lead times, but does not apply any rationing. The fifth column in the table shows the percentage saving for this case over the no demand lead time case, that is, $100 \times (\text{column3-column4})/\text{column3}$. The sixth and seventh columns in the table show the
minimum base stock levels and the corresponding critical levels to satisfy the service level requirements for the critical demand class and non-critical demand class, individually when one recognizes the demand lead times and also uses rationing (which uses the approximation for the critical service level). The eighth column in the table shows the percentage saving for this case over the no demand lead time case, that is, \[ 100 \times \frac{\text{column3} - \text{column6}}{\text{column3}}. \] This part is a $3,530 part with a lead time of 86 days. For the 12 month period in this analysis, the down demand was 12 units, and the lead time demand was 41 units.

When the critical and non-critical service levels are 90% and 80%, respectively, recognizing demand lead time for the non-critical demand class generates 2 units savings in base stock levels. The use of rationing is not very useful here as the service level difference is not significant in this setting. However as the service level is increased for the critical demand class, we see savings through rationing and continue to see savings through recognition of demand lead times for the non-critical class. The maximum saving through recognition of demand lead times is achieved when the service level for the critical demand classes is highest. The maximum saving through rationing is achieved when the service level difference between critical and non-critical demand classes is highest. After I reach the critical and non-critical service levels of 99% and 80%, respectively, start to increase the service level of the non-critical demand class, while keeping the critical service level constant. We observe that base stock levels are same if we do not apply any rationing, and the impact of rationing disappears as the non-critical service level approaches critical service level.

A similar analysis is done for all 64 parts. Tables 3 and 4 show the dollar value of base stock levels (in thousands dollars) for three different approaches. We see that recognizing the demand lead times and using
rationing to differentiate service levels generate significant savings to the company for these 64 parts. When the critical service level is 99 % and the non-critical service level is 80 %, recognizing demand lead times saves 7.4 % on base stock levels (which is equal to the inventory investment, once we assume that the pipeline stocks are owned by the company); additional 4.1 % is saved once the company starts rationing (even though we use an approximation for the service level of the critical demand class) to provide differentiated services to two types of demand.

<table>
<thead>
<tr>
<th>SL %</th>
<th>No Demand Lead Time ($000)</th>
<th>No Rationing ($000)</th>
<th>% saving</th>
<th>Approx ($000)</th>
<th>% saving</th>
<th>Exact ($000)</th>
<th>% saving</th>
</tr>
</thead>
<tbody>
<tr>
<td>99</td>
<td>80</td>
<td>14,050</td>
<td>13,009</td>
<td>7.41</td>
<td>12,432</td>
<td>11.52</td>
<td>11778</td>
</tr>
<tr>
<td>97</td>
<td>80</td>
<td>12,930</td>
<td>12,007</td>
<td>7.41</td>
<td>11,652</td>
<td>9.88</td>
<td>11041</td>
</tr>
<tr>
<td>95</td>
<td>80</td>
<td>12,294</td>
<td>11,450</td>
<td>6.87</td>
<td>11,233</td>
<td>8.63</td>
<td>10,762</td>
</tr>
<tr>
<td>90</td>
<td>80</td>
<td>11,449</td>
<td>10,636</td>
<td>7.10</td>
<td>10,563</td>
<td>7.74</td>
<td>10,197</td>
</tr>
</tbody>
</table>

Table 3

Impact of critical service level

<table>
<thead>
<tr>
<th>SL %</th>
<th>No Demand Lead Time ($000)</th>
<th>No Rationing ($000)</th>
<th>% saving</th>
<th>Approx ($000)</th>
<th>% saving</th>
<th>Exact ($000)</th>
<th>% saving</th>
</tr>
</thead>
<tbody>
<tr>
<td>99</td>
<td>80</td>
<td>14,050</td>
<td>13,009</td>
<td>7.41</td>
<td>12,432</td>
<td>11.52</td>
<td>11778</td>
</tr>
<tr>
<td>99</td>
<td>85</td>
<td>14,050</td>
<td>13,009</td>
<td>7.41</td>
<td>12,591</td>
<td>10.38</td>
<td>11952</td>
</tr>
<tr>
<td>99</td>
<td>90</td>
<td>14,050</td>
<td>13,009</td>
<td>7.41</td>
<td>12,691</td>
<td>9.67</td>
<td>12140</td>
</tr>
<tr>
<td>99</td>
<td>95</td>
<td>14,050</td>
<td>13,009</td>
<td>7.41</td>
<td>12,804</td>
<td>8.87</td>
<td>12511</td>
</tr>
</tbody>
</table>

Table 4

Impact of Non-Critical level

As the critical service level declines to approach the non-critical
service level, we see that savings due to the recognition of demand lead times are still significant, while the impact of rationing is less pronounced. In the last two columns of Tables 3 and 4, we also report the dollar value of base stocks when we use rationing with the exact values of the service level for the critical demand class derived through simulation and percentage savings over the no demand lead time case.

5.4 Conclusion

We conclude that the recognition of the demand leads times and the use of rationing create significant savings for the company. This is true even when we use an approximation to estimate the service level for the critical demand class. More savings are obviously possible if we can accurately determine the service level for the critical demand class. However, the approximation is easy to implement (which is necessary for this particular company) and as it is shown here, its performance is quite reasonable.