Journal of the Korean
OR/ MS Society
Vol. 21, No. 2, August 1996

Coordinating Production Order and Scheduling Policy under Capacity Imbalance

Seung-Kyu Rhee*

Abstract

This paper comes from an observation that overemphasis on capacity utilization measure, which is usual under capacity shortage, can seriously hurt the firm's profit and potential process improvement. We suggest a model that can be used in designing a coordination scheme for decentralized marketing and manufacturing activities. Using a price- and time-sensitive demand and capacitated lotsizing model, we derive an effective communication medium between marketing and manufacturing. This Balance Indicator of process capacity and flexibility also implies that the increase in capacity availability and setup time reduction should be balanced by its market requirements. This is particularly important when a firm tries to improve its process capability by *kaizen*. Further, the model can be used to show the comparative performances of scheduling policies under capacity imbalance. We show that shortening the scheduling cycle can improve the firm profit without changing the simple scheduling rule.

I. Introduction

When a manufacturing company experiences capacity shortage or surplus, it is very common for top management to become too sensitive about the utilization of capacity. When market explodes, the profit is out there only if they have more capacity available. In this situation, it is quite understandable that the top managers doubt if what they got is really the most

* Assistant Professor, Graduate School of Management, KAIST

they can get. On the other hand, while JIT and flexibility philosophy prevail in academic and professional journals, managers still fear the under-utilization of the expensive manufacturing capacity. We find this phenomenon very serious while observing a major Korean tire manufacturing company. Korean tire industry is virtually duopolistic and the two major firms heavily invested in manufacturing facilities during the fierce competition of the last decade. Now the total manufacturing capacity of the industry is almost doubled in 6 years, from 28 million units per year of 1988.

Since the market demand for tire is seasonal, they begin to experience under-utilization of the capacity during off-peak season. Late spring to early summer is an important season in Korean tire business not only because demand is growing, but it is the season of industrial disputes. Korean labor relationship is not settled yet after the nation's democratization of late eighties, and the labor union of the other major company is very strong. While the rival firm is struggling in strike, this firm now has to face a huge gap between demand and capacity. In this volatile environment, the planning staff of the company uses only one production scheduling policy, namely, to maximize the utilization of the capacity. This practice has served well because the firm has never experienced real excess in capacity. Actually the fifty year history of the company is that of ceaseless expansion. This is not an exceptional anecdote of one company, for most of major Korean manufacturing companies have experienced the similar growth.

Sticking to inflexible scheduling policy, however, cannot always work well without cost. Customers are more conscious about the lead time and quality performance than ever. The scheduling practice not only makes it hard for the firm to fully exploit the changing market opportunity, but can cause more complicated problems in the plant. Lot sizes grow in reducing the capacity loss due to setups, and hence lead time gets longer. Even the preventive maintenance activities could be deferred. The overemphasis on the capacity utilization discourages quality- and flexibility-improvement efforts.

According to Kaplan and Norton[5], heavy emphasis on either a financial performance measure or an operational one cannot help managers get the desirable competitive advantage. The strategic priorities of a firm should be translated into specific performance measures in various dimensions. Kaplan and Norton introduced the "Balanced Scorecard" concept of four different performance perspectives: financial, customer, internal business, and innovation and learning. An operational performance measure like capacity utilization in a plant might be a translation of the top's interest in the financial performance of capacity investment. There are, however, other dimensions to consider in making decisions related to capacity. Figure 1 briefly shows the four groups of performance measures interdependent on capacity utilization.

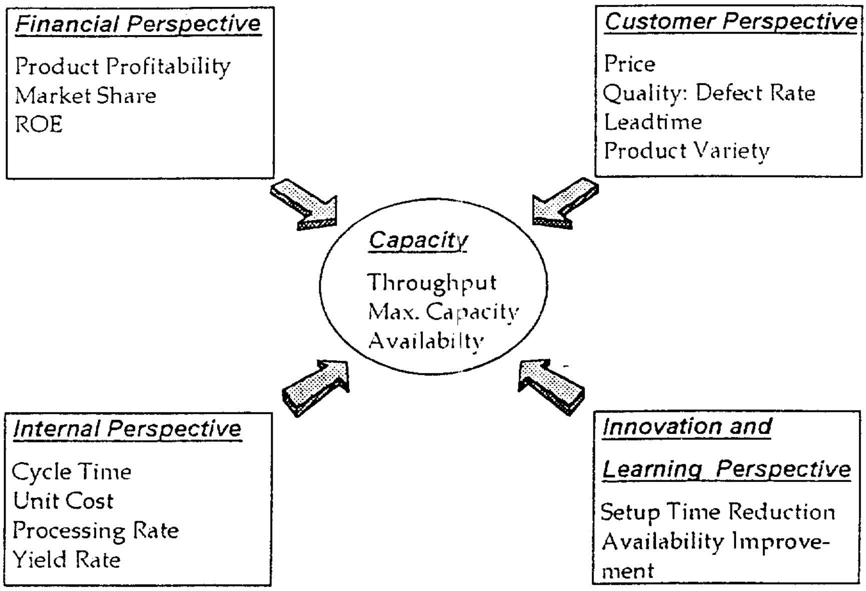

Figure 1. Capacity Variables Viewed from Four Perspectives

In this paper, capacity utilization means the ratio of effective output (or throughput) to the nominal processing capacity of a production process. The process can be a workstation in a complex network, a manufacturing cell, or the entire plant. When the top management or the plant manager makes any decision that changes a variable depicted in the Figure 1, it will affect the capacity utilization. If the firm is organized in a decentralized fashion, then there must be delegation of decision rights and information asymmetry between the headquarters, marketing, and manufacturing. Examining the variables and performance measures shown above, we can safely assume that the financial and customer measures are under the control of headquarters and marketing. The manufacturing division controls the variables in the capacity factor, the internal business measures, and innovation and learning perspectives.

Kaplan[4] also suggested that companies have to improve their management accounting systems under the changing competitive environment. The new systems for measurement and control must be able to help managers i) to motivate the learning and improvement activities, ii) to calculate accurately the profitability of individual products and customers, and iii) to guide decisions on acquiring advanced technological capabilities. We consider the three requirements can be a good guideline to design a decentralized decision structure and performance measurement system related to the capacity problem.

Existing literature on capacity related decisions and throughput analysis is very much di verse and huge, but most of them explain the capacity problem in a static cost minimization setting. It was only recent that we could find literature dealing capacity problem from a comprehensive viewpoint as above. Bitran and Sakar[1] proposed decision models that include reductions of variability and capacity change. They studied the tradeoffs between capacity and process improvements considering WIP constraint (and hence lead time effects). Spence and Porteus[9] was among the first to investigate the capacitated lotsizing problem in view of setup reduction possibilities and options of attaining more expensive additional capacity(overtime). Using their frameworks, we can incorporate the learning aspects of the capacity related decisions. Karmarkar[6] introduced the relationships among lot sizes, lead times and WIPs in various manufacturing settings. Later he extended the results to the relationship between effective capacity and WIP (Karmarkar[7]). This line of works introduced the time dimension as a serious performance criterion in making capacity-related decisions.

Recently de Groote[2] extended the interpretation of the Spence and Porteus model. He focused on the demand variety effects on the total cost of capacitated lotsizing problem. The demand variety is an aggregated function of direct manufacturing costs and setup times so that it can represent the burden laid on the manufacturing to produce diverse products. In applying the concept to managerial decision problems, he related the capacity variable with the product mix (but not the profitability of each product), process improvement, and the flexibility of technology. Although we see some performance measures from financial and customer perspectives could be included in this approach, the market value of the lead time, capacity and flexibility cannot be handled directly in the model.

In this paper we study the capacity problem in a whole firm setting, so that we can evaluate the manufacturing's alternatives from more balanced perspectives. In section 2 we extend the Spence and Porteus model to capture the market measures that must be used in making capacity decisions. Section 3 gives the example applications to utilize the results from the basic model. Next we consider scheduling policies and their profit performance under demand sensitive to lead time and capacity shortage. This is followed by a conclusion and future research directions discussed in Section 5.

II. The Basic Model

Consider a firm that produces n products, each of which faces a linear demand curve. The customers appreciate the shorter lead time. This kind of demand can be represented by $p_j=\alpha_j-$

$\beta_j m_j-\gamma_j T_j$, where p_j denotes the price of prduct j, m_j the demand quantity, and T_j the lead time, for each $j=1,\cdots,n$. The coefficients α_j, β_j, and γ_j are positive constants. The firm's problem is to maximize the profit. In this situation, marketing can sell more if it sets price lower. If the lead time to complete an order gets shorter then the firm can charge more to its customers. With this simple construction, we can integrate the market information into the production decision making. Marketing decides the quantity to be produced while manufacturing controls the production decision under capacity constraint and hence lead time. We are assuming here that the firm's product line is fixed in relevant time span. This is contrasted with that of de Groote[3], where the marketing decides how many different products to be offered and the manufacturing makes lotsizing decisions without capacity constraint.

As in Spence and Porteus[9] we consider a simple deterministic capacitated lot sizing model for the manufacturing. In order to explain the lead time effects both on demand and cost sides, we simply assume that the cycle time (or time between orders: TBO) for product j ($T_j=Q_j/m_j$) can be used, where Q_j denotes the lot size of product j. That is, the scheduling of the lot size is assumed to be feasible (see Karmarkar[8]). The cost to meet the demand includes direct cost of production($c_j m_j$, $\forall j$), setup costs and inventory holding costs, $\sum_{j=1}^{n}(m_j q_j c_s S/Q_j+ic_jQ_j/2)$, where S denotes the nominal setup time, c_s the direct setup cost per unit time, q_j the proportionality constant for product j, and i is the fractional opportunity cost of capital. The processing rate of each product j is the r_j units per time period. The maximum available capacity of the production process is p, which is a fraction of unit time period. For now we assume that setup time S and the availability of the capacity p are given constants. The firm's problem is as follows:

$$\underset{m_j,\,Q_j\geq 0}{\text{Maximize}}\ \prod(m,\,Q)$$

$$=\underset{m_j,\,Q_j\geq 0}{\text{Maximize}}\ \sum_{j=1}^{n}\left\{\left(\alpha_j-\beta_j m_j-\gamma_j\frac{Q_j}{m_j}\right)m_j-\frac{m_j q_j c_s S}{Q_j}-\frac{ic_jQ_j}{2}-c_j m_j\right\} \tag{1}$$

$$\text{subject to}\ \sum_{j=1}^{n} m_j\left(\frac{q_j S}{Q_j}+\frac{1}{r_j}\right)\leq p \tag{2}$$

Now we assume that the firm is decentralized as headquarters (HQ: includes planning and marketing functions) and a plant (P). HQ has the information on $\{\alpha_j,\ \beta_j,\ \gamma_j|j=1,\cdots,n\}$ and i, and also has decision rights over production plan, say, selecting m_j values. On the other hand, P has the information on $\{c_j,\ q_j,\ \gamma_j|j=1,\cdots,n\}$ and process parameters, c_s, S, and p. Also, P has the decision rights on selecting lot size levels (Q_j) given demand m. Further we assume that the capital cost i, the market value of cycle time γ_j's and direct cost of production c_j's are effectively communicated between HQ and P so that they are common knowledge. We think this is not impractical assumption, because those parameters are usually treated as very important and sensitive ones in a manufacturing company. Now the firm(HQ)'s problem becomes:

$$\underset{m_j, Q_j\geq 0}{\text{Maximize}} \sum_{j=1}^{n}\left\{(\alpha_j-\beta_j m_j-c_j)m_j-\frac{m_j q_j c_s S}{Q_j}-\left(\gamma_j+\frac{ic_j}{2}\right)Q_j\right\}$$

$$=\underset{m_j\geq 0}{\text{Maximize}}\left[\sum_{j=1}^{n}(\alpha_j-\beta_j m_j-c_j)m_j-\underset{Q_j\geq 0 \text{ and } (2)}{\text{Min}}\sum_{j=1}^{n}\left\{\frac{m_j q_j c_s S}{Q_j}+\left(\gamma_j+\frac{ic_j}{2}\right)Q_j\right\}\right] \tag{3}$$

$$=\underset{m_j\geq 0}{\text{Maximize}}[CM(m)-IC(m)]$$

The bottom line of (3) means that HQ should select the production plan of its diverse products such that the total contribution margins $CM(m)$ minus indirect cost $IC(m)$ will be maximized. The indirect cost function $IC(m)$ is a reaction function of the plant to the given production plan $\{m_j|j=1,\cdots, n\}$ under current process capability (S and p). First we try to solve the P's problem to obtain the reaction function. Letting $\delta(m)=\sum_{j=1}^{n}m_j/r_j$ leaves us the available machine time for setups by $(p-\delta(m))$. The resulting problem is almost same as that of Spence and Porteus[9].

$$\underset{Q_j\geq 0}{\text{Minimize}}\sum_{j=1}^{n}\left\{\frac{m_j q_j c_s S}{Q_j}+\left(\gamma_j+\frac{ic_j}{2}\right)Q_j\right\} \tag{4}$$

$$\text{subject to } \sum_{j=1}^{n} m_j\left(\frac{q_j S}{Q_j}\right)\leq p-\delta(m) \tag{5}$$

The solution to the problem (4)-(5) can be derived by a straightforward application of the Kuhn-Tucker theorem. There are two solution possibilities, according to whether the constraint (5) is binding or not (that is, the dual variable u is positive or not). Let $L_j=2\gamma_j+ic_j$ (lotsize cost) for notational convenience.

THEOREM 1. The optimal lot sizes in the basic model are, for each j:

$$Q_j(m|p,\ S)=\begin{cases}\sqrt{\dfrac{2m_j q_j c_s S}{L_j}} & \text{if } u(m|p,\ S)\leq 0\\[2ex] \dfrac{S\sqrt{K(m)}}{2(p-\delta(m))}\sqrt{\dfrac{2m_j q_j}{L_j}} & \text{if } u(m|p,\ S)\geq 0\end{cases} \tag{6}$$

Total indirect cost of production is:

$$IC(m|p,\ S)=\begin{cases}\sqrt{c_s K(m) S} & \text{if } u(m|p,\ S)\leq 0,\\ c_s(p-\delta(m))+\dfrac{K(m)S}{4\,(p-\delta(m))} & \text{if } u(m|p,\ S)\geq 0,\end{cases} \tag{7}$$

where $u(m|p,\ S)=\dfrac{K(m)S}{4\,(p-\delta(m))^2}-c_s$ and $K(m)=\left(\sum_{j=1}^{n}\sqrt{2\,L_j q_j m_j}\right)^2$. (8)

With the reaction function shown in (7) the HQ can make the most profitable product mix decision. Note that the optimality condition of the problem (1)-(2) requires that the solutions from (6)-(8) must also satisfy:

$$\frac{\partial\Pi}{\partial m_j}-u\left(\frac{q_j S}{Q_j}+\frac{1}{r_j}\right)=\alpha_j-2\,\beta_j m_j-c_j-\frac{q_j c_s S}{Q_j(m|p,\ S)}-u\left(\frac{q_j S}{Q_j(m|p,\ S)}+\frac{1}{r_j}\right)=0\ \ \forall j=1,\ \cdots,\ n. \tag{9}$$

Using $u=\text{Max}\{0,\ u(m|p,\ S)\}$, (9) is the same condition for the HQ's optimality condition for the problem (3), as follows:

$$\frac{\partial\Pi}{\partial m_j}=\frac{\partial CM(m)}{\partial m_j}-\frac{\partial\, IC(m|p,\ S)}{\partial m_j}=0\ \ \forall j=1,\ \cdots,\ n. \tag{10}$$

We discuss this problem in two different situations. When the plant capacity is more than enough for the market demand considering profitability ($u=0$), (9) and (10) become:

$$\frac{\partial\Pi}{\partial m_j}=\alpha_j-2\,\beta_j m_j-c_j-\sqrt{\frac{L_j q_j c_s S}{2\,m_j}}=0\ \ \forall j=1,\ \cdots,\ n. \tag{11}$$

On the other hand when the plant capacity is tight ($u>0$), (9) and (10) are the same condition as follows. Further, we can rewrite the optimality condition to get a new interpretation.

$$\frac{\partial\Pi}{\partial m_j}=\alpha_j-2\,\beta_j m_j-c_j-\frac{S}{2\,(p-\delta(m))}\sqrt{\frac{L_j q_j K(m)}{2\,m_j}}-\frac{K(m)\,S}{4\,r_j(p-\delta(m))^2}+\frac{c_s}{r_j}$$

$$=\alpha_j-2\,\beta_j m_j-c_j-v(m|p,\ S)\sqrt{\frac{L_j q_j c_s S}{2\,m_j}}-\frac{c_s}{r_j}(v(m|p,\ S)^2-1)=0\ \ \forall j=1,\ \cdots,\ n. \tag{12}$$

where $v(m|p,\ S)$ denotes the ratio of the total setup time desirable to the actual available time:

$$v(m|p,\ S)=\frac{1}{2\,(p-\delta(m))}\sqrt{\frac{K(m)S}{c_s}}=\frac{\sum_{j=1}^{n}\sqrt{L_j q_j m_j S/\ 2c_s}}{(p-\delta(m))}=\frac{\sum_{j=1}^{n} m_j q_j S/Q_j}{\left(p-\sum_{j=1}^{n}\frac{m_j}{r_j}\right)} \tag{13}$$

$$=\frac{\text{(Total setup time required without capacity constraint)}}{\text{(Available time for setup due to capacity constraint)}}$$

THEOREM 2. When the plant capacity is tight, then the lot sizes of all the products for given demand increases by the same rate ($v(m|p,\ S)\geq 1$) from the sizes if there were no capacity constraint. Further, the HQ must use higher hurdle rate for product contribution margins in making a master production schedule.

Proof The first part is trivial from (6), because

$$Q_j(m|p,\ S)=\frac{S\sqrt{K(m)}}{2\,(p-\delta(m))}\sqrt{\frac{2\,m_j q_j}{L_j}}=v(m|p,\ S)\sqrt{\frac{2\,m_j q_j c_s S}{L_j}}\quad \forall j=1,\ \cdots,\ n. \tag{14}$$

and $v(m|p,\ S)\geq 1$ (otherwise, from (13) the capacity constraint cannot be binding). For the product contribution hurdle rate, we use the marginal indirect cost parts in (11) and (12) as follows:

$$HR_j(m|p,\ S)=\begin{cases}\sqrt{\dfrac{L_j q_j c_s S}{2\,m_j}} & \text{if } u(m|p,\ S)\leq 0,\\[2ex] v(m|p,\ S)\sqrt{\dfrac{L_j q_j c_s S}{2\,m_j}}+\dfrac{c_s}{r_j}\left(v(m|p,\ S)^2-1\right) & \text{if } u(m|p,\ S)\geq 0.\end{cases}\quad \forall j. \tag{15}$$

Since $v(m|p,\ S)\geq 1$, the hurdle rate for product contribution margin gets higher when the capacity becomes tight. Further, from (8) $u(m|p,\ S)=c_s\left(v(m|p,\ S)^2-1\right)$. So we know that the hurdle rate for each product accounts for the elevated marginal indirect cost $v(m|p,\ S)\sqrt{L_j q_j c_s S/\,2\,m_j}$ and the marginal capacity cost per unit $u(m|p,\ S)/r_j$.

Numerical Example

Here we consider two-product case as follows :

$$\alpha=\begin{bmatrix}30\\25\end{bmatrix},\ \beta=\begin{bmatrix}0.004\\0.001\end{bmatrix},\ \gamma=\begin{bmatrix}3\\3\end{bmatrix},\ c=\begin{bmatrix}18\\18\end{bmatrix},\ r=\begin{bmatrix}2000\\3000\end{bmatrix},\ q=\begin{bmatrix}0.2\\0.1\end{bmatrix},\ S=0.1,\ c_S=10000,\ p=0.7$$

We can get the following solutions from basic model.

$$m=\begin{bmatrix} 525.61 \\ 1052.37 \end{bmatrix},\ Q=\begin{bmatrix} 243.39 \\ 243.53 \end{bmatrix},\ Setups=\begin{bmatrix} 2.16 \\ 4.32 \end{bmatrix},\ v=1.4830,\ u=11978.30,\ \pi=8697.35$$

Consider the situation where the firm does not coordinate the marketing (what to produce) and manufacturing (lotsizing and inventory) decisions. In the example above, maximal profitable demand level (assuming lead time=1) is (1125, 2000) without considering the indirect cost information. If the manufacturing division is given this order, which requires well beyond the available capacity, then it may seem reasonable to reduce the production quantity proportionally. Considering the setup time (one setup for each product), the plant can produce only (613, 1090) units. Comparing this result (π=5053.98) with the optimum, we can see that the lack of coordination brings the huge distortion of product mix and capacity utilization. Now we turn to other managerial applications of the basic model.

Ⅲ. The Applications of the Basic Model

Coordination of Marketing and Manufacturing Functions

As we see in the numerical example above, the real profitability of a product mix cannot be determined without evaluating its indirect cost consequences. Reversely, a plant manager cannot fully appreciate the profit contribution of its flexibility or responsiveness if the market information is not available. The basic model can be a conceptual backbone of the communication structure for this case. Contribution margin of each product should be compared against the hurdle rate from the marginal indirect cost. When the plant capacity is tight, the plant manager must feed the indirect cost information to the marketing to facilitate the product mix decision. If the headquarters wants the plant to improve the capability (higher available capacity and shorter setup time), it should give the profitability information to the plant manager.

The aggregate function $v(m|p, S)$ of the basic model can be viewed from two different perspectives. From the plant manager's viewpoint, it represents the challenge from the product variety of the production order. If the production order from the HQ raises the v value significantly, the plant manager will complain about the unrealistic planning practice of HQ. On the other hand, even though the top managers do not have precise information about the high function value, they may see the large lot sizes and high inventory level as an evidence of poor plant management. This observation partly makes sense as the ratio is a function of p and S, which are under the control of the plant managers. They can improve p by reducing the contingency factors of a process such as defect or rework rate, maintenance and down time of machines, and unstable material flows. By the same token, vast literature on JIT and TQM addresses that they can improve the S value, and hence reduce the value of $v(m|p, S)$.

The above argument makes the aggregated function very good communication medium in a decentralized manufacturing company that experiences tight capacity problem. When the HQ and P meet to confirm a master production schedule or aggregate production plan, they can examine the function value to see whether the product variety is too high and/or whether the flexibility of the plant is sufficient. Note that the value's p and S have certain limits, or p cannot exceed 1 and S cannot be reduced below 0. So if the plant manager's parameters approach to their limit, and HQ still needs more capacity, it is time to consider capacity expansion. Summing up the meanings of the factors comprising the v function, we suggest that the inverse of $v(m|p, S)$ can be a representative indicator of process balance between flexibility and capacity.

Let $f=1/v(m|p, S)$ denote the process balance indicator of the firm at a specific product mix and process configuration. It follows that since $0<f\leq 1$, f value near 1 means perfect (or sufficient) balance. In other words, in that case the product diversity (imposed by m) and process capability (p, S and other parameters) are perfectly balanced. Given the product order, if the process gets more flexible (smaller S) or more available (larger p), then the balance will improve. As we will see later in a numerical example, however, if the demand of the firm responds to the improved performance of the process, then this balance indicator can give more helpful information about the firm's capability. This measure can be also useful when combined with de Groote's product variety measure and appropriate incentive to increase the flexibility of the process.

Effects of Process Improvement

Now we look into the effects of plant parameter improvement. From the basic model, the parameters p and S can be best candidates for continuous improvement. The available capacity can be increased by total preventive maintenance program or by reduction of rework. The setup time reduction has been the single most popular topic of manufacturing management of the last decade. Here we examine the effects of the improvements using the previous numerical example. First we set up a benchmark case of increasing the capacity availability upto its potential limit of 95%, which yields the firm profit of $\pi=11,438$. This is 31.5% increase from the base case. Next we find the necessary level of setup time reduction to attain the same profit increase. The S value should be reduced to 0.000457 from the current 0.01, which is more than 90% reduction.

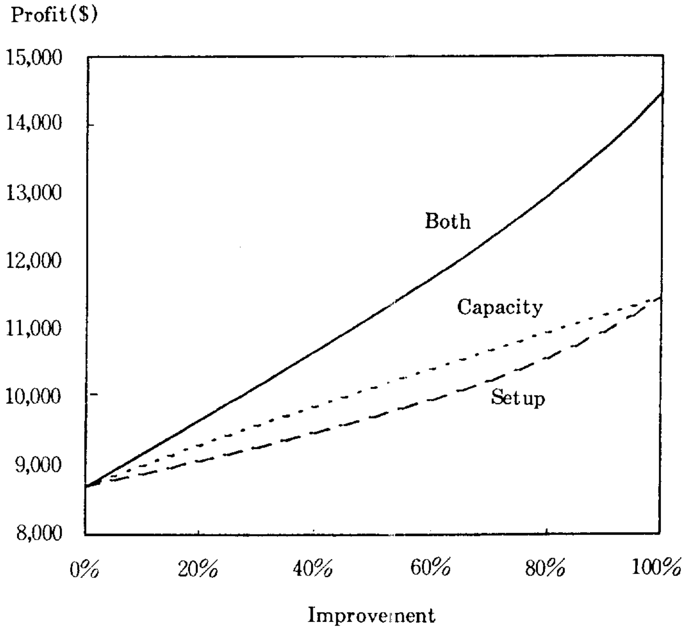

Figure 2. The Profit Increase from Process Improvement

Figure 2 shows that the profit impacts of the improvements in process parameters. The 100% in the horizontal axis means the upper bounds of improvement in the two parameters. First we can see that the profit increase from improvement in the availability is concave and that from setup time reduction is convex. This is interesting because most of process improvement literature emphasizes the importance of setup time reduction, and not much attention has been paid to the capacity availability. When the capacity is tight, however, the continuous improvement effort should be directed to improve the capacity availability first, because it can increase the profitable production and save more time to setup change. On the other hand, when the setup time alone is improved, the additional capacity gained may be very little because the process time allotted to the setup change is very small. This explains the convexity of the profit change from the setup time reduction in the graph.

The second implication of the figure 2 is that the combined efforts to improve both process characteristics can result in far better profit performance than those for one dimensional improvement. Considering the increasing difficulty of improving a process parameter, the steep profit increase from the joint improvement shows the way to go. We can confirm this point by observing the major outputs of the computation. Figure 3 shows impacts of the process improvements for the throughput m, the lotsize, and the process balance indicator f.

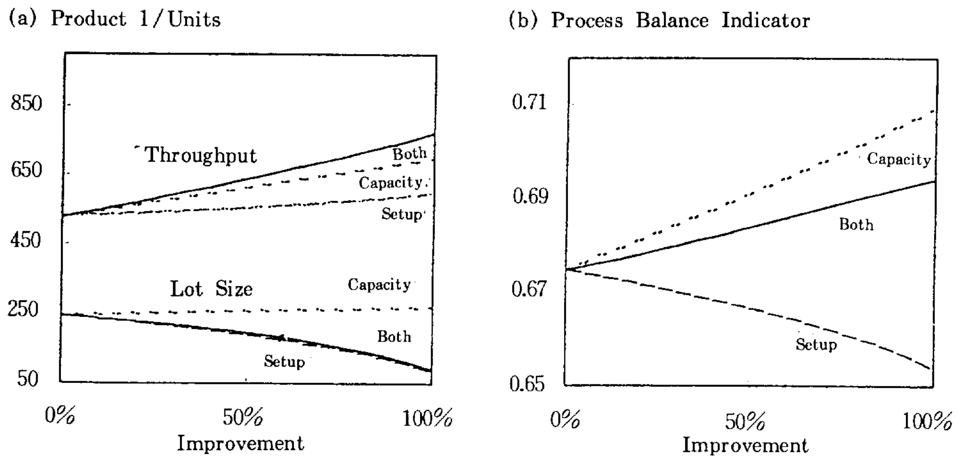

Figure 3. The Plant Performance Change from Process Improvement

As the available capacity increases, (a) of figure 3 shows that the firm increases the production without decreasing the lotsize. Behind this, however, the total capacity used for setup is increased by 23% from 0.086404 to 0.10623. This explains the improved process balance of the availability improvement case of (b). Also of interest revealed in figure (3-b) is that the current process wants capacity more than setup time improvement. This is the reason why the balance improvement is slower in the case of simultaneous improvement than in the capacity-only option.

IV. Scheduling Policy under Capacity Imbalance

Evaluation of simple production scheduling policy

The basic model represents the normative approach to the firm's problem. A real company or a decision maker does not necessarily act as an optimizer. So if we have a descriptive model of a decentralized firm then we can compare the performance variables of the two different settings. In the tire manufacturing company we mentioned in the introduction, the production scheduling is done as follows:

$$\begin{aligned} &\underset{m_j \geq 0}{\text{Maximize}}\ \Pi = \sum_{j=1}^{n} (\alpha_j - \beta_j m_j - \gamma_j - L_j) m_j - \sum_{j=1}^{n} c_s q_j S \\ &\text{subject to } \sum_{j=1}^{n} \frac{m_j}{r_j} \leq p - \sum_{j=1}^{n} q_j S \end{aligned} \tag{16}$$

The constraint in (16) means that the setup for each product type is done only once in a scheduling period, $Q_j = m_j$, for each j, so the firm can produce as many products as possible. The lotsizing possibility is eliminated so that this approach makes sense only when there are plenty of demands the price of which is well beyond the supply cost. Now we assume that this is the case so the shadow price of the capacity constraint is positive. The optimal solution to the problem (16) is:

$$m_j = \frac{\alpha_j - c_j - L_j}{2\beta_j} - \frac{u(p,\ S)}{2\beta_j r_j} \text{ for each } j = 1,\ \cdots,\ n \text{ and}$$

$$u(p,\ S) = \left[\sum_{j=1}^{n}\left(\frac{\alpha_j - c_j - L_j}{2\beta_j r_j} + q_j S\right) - p\right] \Big/ \sum_{j=1}^{n}\frac{1}{2\beta_j r_j^2} \tag{17}$$

We can easily interpret the shadow value $u(p,\ S)$ of the capacity given market demand and process parameters above. The numerator means the gap between the required capacity to support potentially profitable demand and available capacity. The denominator represents the required marginal capacity to generate marginal profit through producing additional products. From this we can understand that the shadow value means the monetary evaluation of capacity shortage. Further, the second term in m_j solution means that each product should reduce its production amount from its market potential by a criterion of profitable capacity usage.

Evaluation of improved scheduling policy by reducing the scheduling period

The simple scheduling policy above has its own merit, that is its simplicity. Highly complicated scheduling policy that requires large amount of information and coordination could be very difficult to implement in a decentralized firm. When a firm faces diverse product demand and experiences capacity shortage, it is often unreal to expect to see more frequent setups and smaller lotsizes in view of adequate profitability information. If the headquarters of the firm knows that the capacity is wasted by the simple throughput maximizing policy, there is easier way to improve the lead time performance and profit without changing scheduling rule. Simply reducing the scheduling period length could solve the most of the plant's large lotsizing practice.

Let z be the ratio of the reduced period length to the original one: if z is 0.5 and the formal scheduling cycle is 4 weeks then the new scheduling cycle is 2 weeks. With this change the plant's problem in (16) changes only in the period-related costs and time data. So $L_j \Rightarrow zL_j$ and $S \Rightarrow S/z$. Applying this to (17) we can derive the results for the whole period to compare the performance with those from (16).

Numerical Example

Max. Throughput Model(z=1):

$$m=\begin{bmatrix}758.40\\872.40\end{bmatrix},\ Q=\begin{bmatrix}758.40\\872.40\end{bmatrix},\ Setups=\begin{bmatrix}1\\1\end{bmatrix},\ v=4.5081,\ u=4065.50,\ \pi=5485.72$$

Max. Throughput Model (z=0.5):

$$m=\begin{bmatrix}658.80\\931.80\end{bmatrix},\ Q=\begin{bmatrix}329.40\\465.90\end{bmatrix},\ Setups=\begin{bmatrix}2\\2\end{bmatrix},\ v=4.3989,\ u=9559.20,\ \pi=8122.21$$

Max. Throughput Model (z=0.25):

$$m=\begin{bmatrix}576.60\\875.10\end{bmatrix},\ Q=\begin{bmatrix}144.15\\218.78\end{bmatrix},\ Setups=\begin{bmatrix}4\\4\end{bmatrix},\ v=4.1828,\ u=12824.4,\ \pi=8333.82$$

Recall that the base case optimal profit of the firm was $\pi=8697.35$. Comparing the value with the result from the Max. Throughput Model (z=1), we see the poor usage of the production process costs the firm about 37% profit loss. The difference mainly comes from the price decrease from long lead time performance. Of course in practice, we may not be able to observe this kind of dramatic price cut due to poor lead time performance. In highly competitive market of the products that are in their later stage of life cycle, however, the lead time performance may be major order-winning criterion for profitable accounts. The lead time penalty coefficient γ in our model represents this point. In this situation, the profit performance of the throughput maximizing schedule is radically improved by reducing its scheduling cycle time. When z=0.25, the plant manager will clear the orders four times in a month. This enforces the manager to reduce the lotsizes by a similar scale, and hence improve the lead time performance. Note that the profit in this case (π=8333.82) is only 4.18% less than the optimum. We can see that the exact calculation of optimal lotsize is not important in this situation. What matters most here is that reducing the lead time by enforcing frequent setups.

V. Conclusions

Matching the plant capability with the market requirements is no easy matter. There have been so many arguments to help building effective marketing/manufacturing coordination mechanism, but so far little have been successful. The work reported here has been tried to provide the necessary information to make a profitable production order and an improved lotsizing decision. We suggest that the marginal indirect cost should be used as a hurdle rate in making production order. Then with the given order, the plant manager should work to improve its process balance between capacity availability and flexibility.

Using the balance indicator f, the plant manager gets the information about which process parameter is more important to support the firm's profitable order winning. The planning staff in HQ can be more realistic in deciding what to produce with given plant capability. The modeling framework used in this paper is compatible with the criteria suggested by Kaplan[4]. The balance indicator gives the plant manager incentive to improve its process capability. The hurdle rate gives the marketing manager information of product profitability. Finally, although we do not include the comparison of different process technologies, the basic model can certainly handle the issues as in de Groote [2].

We do not think the lotsizing practice without explicitly considering the cost tradeoff will disappear easily. When this is a serious obstacle to improve the lead time performance, enforcing the reduced planning cycle can certainly help. If the plant managers have to try more frequent setups, there is a natural incentive to improve it.

There are several ways to extend ideas from the basic model. The first and most obvious one is to applying the process balance concept to more general manufacturing models. More detailed and realistic scheduling research can be done to investigate the coordination mechanism.

References

[1] Bitran, G.R. and D. Sakar, "Throughput Analysis in Manufacturing Networks," *European Journal of Operational Research*, 74, 448-465, 1994.

[2] de Groote, X., "Flexibility and Product Variety in Lot-sizing Models," *European Journal of Operational Research*, 75, 264-274, 1994.

[3] de Groote, X., "Flexibility and Marketing/Manufacturing Coordination," *International Journal of Production Economics*, 36, 153-167, 1994.

[4] Kaplan, R.S., "New Systems for Measurement and Control," *The Engineering Economist*, 36, 201-218, 1991.

[5] Kaplan, R.S. and D.P. Norton, "The Balanced Scorecard-Measures That Drive Performance," *Harvard Business Review*, Jan.-Feb., 71-79, 1992.

[6] Karmarkar, U.S., "Lot Sizes, Lead Times and In-Process Inventories," *Management Science*, 33, 409-418, 1987.

[7] Karmarkar, U.S., "Capacity Loading, Release Planning and Master Scheduling with WIP and Lead Times," *Journal of Manufacturing Operations Management*, 2, 105-123, 1989.

[8] Karmarkar, U.S., "Manufacturing Lead Times," *Logistics of Production and Inventory*, (ed.) by S.C. Graves, A.H.G. Rinnoy Kan and P.H.Zipkin, North-Holland, Chapter 6, 1993.

[9] Spence, A.M. and E.L. Porteus, "Setup Reduction and Increased Effective Capacity," *Management Science*, 33, 1291-1301, 1987.