IIE Transactions (1995) **27**, 321–331

Economic determination of specification levels and allocation priorities of semiconductor products

K. JO MIN[1] and SHMUEL S. OREN[2]

[1] *Department of Industrial and Manufacturing Systems Engineering, 205 Engineering Annex, Iowa State University, Ames, IA 50011, USA*

[2] *Department of Industrial Engineering and Operations Research, 4135 Etcheverry Hall, University of California, Berkeley, CA 94720, USA*

Received October 1991 and accepted May 1993

We construct and analyze an economically efficient way of pricing and allocating semiconductor chips of which production technology is characterized by persistent quality variations and of which production capacity is exceeded by potential demand. In our model, specification levels and allocation priorities of competing orders from customers are systematically determined for a single profit maximizing producer. In the proposed scheme, the producer offers a 'product line' of priority classes under an allocation rule that always supplies higher priority classes with higher spec. level chips. This product line design and allocation rule enable us to cast the producer's profit maximization problem as a nonlinear programming formulation. Also, we investigate the optimality of the proposed allocation rule and derive conditions under which the profitability of downgrading is determined.

1. Introduction

Integration of manufacturing and marketing strategies is a subject of growing interest to academics and practitioners alike. This subject is particularly relevant in the semiconductor industry owing to the unique nature of its production process. This paper introduces a marketing strategy for custom semiconductor chips, which incorporates the unique nature of the production process and offers a market-based solution to some outstanding distribution problems plaguing the semiconductor industry.

1.1. *Overview of semiconductor production*

We begin by briefly describing some aspects of semiconductor production that are relevant to this paper (for a more thorough description, see Leachman, 1986 or Intel, 1985). Once semiconductor chips have been manufactured, they are extensively tested to determine the level of certain critical attribute(s) such as speed or power consumption. According to the test results, the chips are classified into bins where each bin has a pre-determined specific range of the attribute level. For example, a bin may be for the chips with speed levels between 10 MHz and 15 MHz, and any chip in this bin is said to have the spec. (specification or minimum quality) level of 10 MHz.

Higher spec. level chips can always substitute for lower spec. level chips. Thus, the producer can fill orders of lower spec. chips by supplying higher spec. chips (labelled as such). Another option, however, is to relabel higher spec. chips as lower spec. chips. The latter approach is often referred to as downgrading or down-binning. Downgrading is currently employed as an emergency measure to fulfill delivery commitments to customers when there is a shortage of lower spec. chips. However, when the producer has a certain degree of market power, downgrading can also serve as a device to maintain scarcity and higher prices for higher quality chips in order to maximize his profit (see Section 1.5 for further details).

Another important economic aspect in the above manufacturing process is the determination of the spec. levels that define the bins. In general by setting spec. levels higher (lower), the quality of chips to be sold increases (decreases) and the price the producer charges also increases (decreases) while the quantity to be sold decreases (increases). Unfortunately, this important decision is often carried out without systematic analysis of economic consequences.

From an entire industry perspective, the semiconductor industry is generally characterized as either capacity constrained or cyclic (i.e., there are periods during which potential demands exceed production capacities). Even in recession periods, owing to the diversity in their usage, there are certain products whose demand exceeds the production capacity. And in expansion periods, the production capacity rarely catches up with the potential market for the products (see for example, Leachman, 1986). In this paper, we will consider semiconductor chips for which the producer enjoys a certain degree of market

power and for which potential demand exceeds the production capacity. Because of this production capacity limitation as well as the random factors in the production process (which affect the actual number of chips falling into each bin), delivery commitments are frequently not met. This poor on-time-delivery performance often results in over-ordering by customers and over-booking by producers, and haphazard allocation of chips in case of shortages.

In this paper, we limit our investigation to a single critical attribute (e.g., either speed or power consumption). In addition, we consider semiconductor products for which the producer has monopoly power (e.g., certain brands of microprocessors). Our analysis is not applicable to competitive semiconductor products (e.g., DRAM chips). Finally, throughout this paper, we note that the terms 'allocation' and 'delivery' as well as (semiconductor) 'chips' and 'products' are used interchangeably.

1.2. *Economic determination of spec. levels and allocation priorities of semiconductor products*

The primary objective of this paper is to construct an economically efficient allocation mechanism so as to maximize the producer's profit under the constrained production capacity and quality variations described above. Furthermore, unless prescribed by a special consideration, the corresponding optimal spec. levels can also be determined for the producer as an output of the profit maximization process.

Our approach is based on the premise that there are numerous potential customers with heterogeneous preferences with respect to quality levels (see, for example, Smith, 1986), where *several (more than one) quality levels are acceptable to potential customers*. Under this assumption, we construct a product line of priority classes defined in terms of product spec. levels and their corresponding delivery probabilities. The product line of priority classes is based on an allocation rule to which we refer as the *priority supply rule*. In simple language, the priority supply rule is designed to *provide higher spec. level products to higher valued consumption units regardless of production outcomes* (we will formally define the priority supply rule in the next subsection). The profit maximizing allocation is realized via non-uniform pricing strategies of the producer who exploits the heterogeneous preferences of the potential customers in order to achieve his desired market segmentation.

Customers are assumed to self-select priority classes so as to maximize their net benefits (i.e., total benefits minus purchasing prices) from the price table. The spec. levels, delivery probabilities, and prices corresponding to the priority classes are endogenously derived by the producer so as to maximize his profit subject to constraints implied by the customers' self-selection behavior as well as other production and demand data.

Under the general framework of the model described above, the producer announces the prices for priority classes for a duration of T periods (e.g., T production runs), which can be viewed as a contingent forward delivery contract. The duration T plays an important role in the implementation phase of the model with regard to the verifiability of the priority supply rule and the credibility of the delivery contract. As T gets larger, the average quality mix supplied in each priority class will conform with the delivery probabilities specified in the price table. Throughout this paper, we will assume that the critical economic aspects of both producer and customers remain constant over time. That is, the model environments are essentially stationary and dynamic effects such as the learning effect in chip production technology are assumed to be negligible for the duration in so far as the producer and customers are concerned.

From the perspective of allocative efficiency, several papers have focused on rationing mechanisms and their economic implications. Antle and Eppen (1985) investigate economic inefficiencies in excess allocation as well as under-allocation of resources when asymmetry of information exists among economic agents. Wilson (1989), among others, studies an efficient allocation mechanism under limited supply, which is referred to as priority rationing or priority pricing. According to this approach, available supply is allocated on the basis of contracts that specify each customer's priority. Priority rationing (or priority pricing) has been shown to achieve efficiency gains in the case of non-storable commodities or goods and services with congested demands or queues due to limited supply capacities. In the next three subsections, we discuss in detail three unique features of this paper; the priority supply rule, the ordering/allocation process, and the downgrading.

1.3. *Priority supply rule*

We formally state the priority supply rule as follows: Suppose that the spec. levels have been determined and all products are classified into bins with spec. $u_1, \ldots, u_N$ (for a customer, spec. level u_i is preferable to spec. level $u_{i+1}, i = 1, \ldots, N-1$). The priority supply rule determines how the bin contents are assigned to customers whose orders are specified in terms of priority classes 1, $\ldots, M$. According to this rule, regardless of the number of products in each bin, the producer allocates spec. u_1 products to the first class customers. If there is a shortage of u_1 products, they are allocated randomly to the first class and the balance of the first class orders is met with the next best products, spec. u_2. If spec. u_2 products are also exhausted, then the next best spec. u_3 are allocated, and so on until the demand for first class is met. Only then does the allocation to the second class customers start in the same manner, and only after the second class, the

allocation for the third class, and so on. This rationing process terminates when either all the bins are exhausted or when all the demands are satisfied. We will assume that any surplus product, after satisfying all demands, has negligible salvage value/ cost and alternative usage as well as storage are not permitted.

We note that one of the advantages of the priority supply rule is that it is verifiable in the following sense. For instance, an independent auditing organization can easily verify that the highest spec. level chips in a lower class order are no better than the lowest spec. level chips in a higher class order under any contingency. Such verification is essential for establishing the producer's credibility and inducing the desired customer response to the pricing scheme employed in this paper.

1.4. *Ordering/allocation process of the model*

In order to explain the ordering/allocation process of this paper, we first state our assumptions. We assume that customers as well as the producer are expected-value decision makers. We also assume that the producer has no incentive to deviate from what he has announced to do, considering economic consequences such as reputation effects after the producer's deviation. We further assume that, owing to sufficiently high cancellation penalties, the customers have no incentive to cancel, once orders have been made. For simplicity, we also assume that there is only one production run (including classification of chips according to spec. levels) between two consecutive deliveries and an instantaneous delivery occurs right after a production run.

Under these assumptions, a typical price table for our model is shown in Table 1. In the price table, given spec. levels $u_j, j = 1, \ldots, N$, an expected price per delivery for ordering one unit of priority class i, p_i, is specified for $i = 1, \ldots, M$. Also specified are Pr_{ij}, the probability of delivery for one unit of priority class i of a spec. u_j chip, for $i = 1, \ldots, M$ and $j = 1, \ldots, N$.

Given this type of price table, a 'product' to be sold to customers is a unit of priority class $i (i = 1, \ldots, M)$ where a unit of priority class i is defined in terms of spec. levels u_j $(j = 1, \ldots, N)$ and their corresponding delivery probabilities Pr_{ij}. By ordering one unit of priority class i, a customer is expected to receive on average (averaged over deliveries) a sum of product mix of Pr_{ij} units of spec. j chips, $j = 1, \ldots, N$. The time sequence of events and decisions of our model is summarized as follows.

Table 1. Price table for priority and spec. level pricing

Class	*Expected price*	u_1	u_2		u_N
1	p_1	Pr_{11}	Pr_{12}		Pr_{1N}
2	p_2	Pr_{21}	Pr_{22}		Pr_{2N}
M	p_M	Pr_{M1}	Pr_{M2}		Pr_{MN}

1. First, given the contract duration of T periods (i.e., T production runs), the producer designs the price table shown in Table 1 taking the customers' net benefit maximizing behavior into consideration (the producer may exercise his downgrading option).
2. Next, the producer announces the price table to all customers, and the customers will self-select their optimal priority classes after evaluating all priority classes and their corresponding prices.
3. Finally, after each production run, the producer classifies each chip according to its spec. level (and downgrades if the downgrading option has been chosen) and delivers to customers according to the priority supply rule.

We note that the payment from customers to the producer will consist of a nonrefundable 'priority charge' paid at the time of ordering and a delivery charge ex post paid at the time of delivery. This two-part pricing strategy is further justified and illustrated via a numerical example later in this paper.

1.5. *Downgrading*

Downgrading refers to the practice of relabelling higher spec. chips as lower spec. chips. In the semiconductor industry, downgrading is currently employed as an emergency measure to fulfill delivery commitments to customers. This objective could be accomplished, however, by providing higher spec. chips (explicitly labelled as such) in lieu of lower spec. chips that are originally ordered. The latter approach is not so prevalent, and one plausible reason for that is the producer's fear of undercutting the price of higher spec. chips. Hence, there is a strategic motive for downgrading rather than simple downward substituting. In this paper, we take this rationale further and consider the case when downgrading is viewed as a full-fledged strategic instrument for the producer. As such, it can be used to improve profit by maintaining scarcity and higher prices for higher spec. chips to the extent supported by the demand. Since we deviate here from the current industry practice, our analysis of downgrading at this point in time is more of an academic investigation than an implementation proposal. Furthermore, even if such strategy were adopted by a producer, it is unlikely that we would be able to verify or validate it owing to its proprietary nature.

For our analysis, we assume that, for the producer, the cost incurred in relabelling is negligible. Once relabelled (i.e., the face value is now set at a lower spec. level), we assume that it is impossible for customers to distinguish chips originally classified as lower spec. from relabelled as lower spec. chips (without extensive retesting and reclassifying chips themselves). We also assume that customers

will find it economically detrimental to bypass (e.g., by retesting and reclassifying chips) the downgrading practice of the producer. That is, customers will accept the spec. level (whether it is originally labelled or relabelled by the producer) shown on a chip. Under these assumptions, by downgrading, the producer may be able to create the scarcity of higher spec. level chips artificially and enhance his ability to differentiate/exploit the customers' heterogeneous preferences so as to improve his profitability.

In order to study the downgrading option analytically, we first investigate the market segmentation (i.e., a realization of heterogeneous customers' purchase decisions based on the priority classes and the corresponding prices offered by the producer) by employing the 'single exchange' perturbation analysis. By single exchange, we mean that one unit of given spec. level chips from a priority class is exchanged with another chip from another priority class. This perturbation analysis results in a set of conditions that can be used to determine *a priori* the profitability of downgrading. This analysis also provides a set of conditions from which one can infer the optimality of the priority supply rule among a broad class of allocation rules with respect to the profit maximization objective.

The rest of the paper is organized as follows. First, we characterize the market environments mathematically and formulate the producer's profit maximization problem. Next, the market segmentation is analyzed via single exchange perturbations. Based on these perturbations, the profitability of downgrading and the optimality of the priority supply rule are investigated. Finally, a simple two-part tariff for the purpose of implementation and an illustrative example are presented.

2. Basic model

2.1. *Market characterization*

The characterizations of customers and quality levels closely follow those in Smith (1986). The customer heterogeneity is represented by a customer's type index $t \in [0, 1]$. This index serves as the preference ranking of a customer relative to other customers in terms of preference for higher qualities. In this paper, larger t corresponds to higher quality preferences. The variations in products' quality level are characterized in terms of a quality level index $u \in [\rho, \sigma]$ where $\rho > 0$ and ρ and σ denote the lower limit and upper limit of the quality, respectively. The index u indicates the quality level of a critical attribute such as speed or power consumption. We will let larger u correspond to higher quality levels. Given a customer type t and a chip of quality level u, we assume that there is a corresponding utility function $U(u, t)$ for all u and t. We assume that this utility is additive over quantity and each utility for a chip can be treated separately. Hence, each unit utility can be considered as a separate customer and each customer can be assumed to purchase either one or zero unit of the product (see, for example, Wilson, 1989). The term utility function is identical to 'Willingness To Pay (WTP)' function in the pricing literature (see, for example, Oren *et al.*, 1982, 1983).

According to our definitions of the customer type index t, the quality level u, and the utility function $U(u, t)$, we have $\partial U/\partial u > 0$ and $\partial U/\partial t > 0$. In addition, we assume that $\partial^2 U/\partial u^2 \leq 0$ and $\partial^2 U/\partial u \partial t > 0$; i.e., the utility function is concave in u while the marginal utility with respect to the quality level u increases with the customer index t. The cumulative demand of customers is characterized by $D(t)$, the total amount of demand of customers ranked t or higher. We will assume that $\mathrm{d}D/\mathrm{d}t < 0$.

The contingency supply quantity function is characterized by $S(u, \boldsymbol{B})$, the total number of products manufactured in a given period with quality level u or higher under contingency $\boldsymbol{B}$. We will assume that $\partial S/\partial u < 0$. The vector $\boldsymbol{B} = (B_1, \ldots, B_K)$ denotes a random vector whose elements correspond to factors that affect the quality distribution of production outcomes. We assume that the sample space β for $\boldsymbol{B}$ and the corresponding joint probability distribution $\mathrm{Prob}\{\boldsymbol{b}\}$ over all possible realizations $\boldsymbol{b} \in \beta$ are known to the producer. Also, the producer is assumed to have complete knowledge of the customers' type distribution $D(t)$ and the form of the utility function $U(u, t)$, but he cannot identify the particular type of a customer. We assume that the production cost is lump sum constant because, regardless of the number of defects or the actual distribution of products over the quality level, the manufacturing and testing costs have already been incurred. We further assume that the constant production cost is set equal to zero because the constant cannot affect the optimal decisions of our model.

In order to implement this allocation mechanism, discretization schemes for the continuous quality levels and customer ranks are necessary. For the discretization of u, we will employ the concept of specification levels introduced earlier. Let $u_1, \ldots, u_N$ ($u_1 > \ldots > u_N$) denote N specification levels. Then, all products with quality level $u \in [u_j, u_{j-1}]$ will be labelled as u_j spec. products for $j = 1, \ldots, N$. Throughout the rest of this paper, we will assume that a customer's utility depends on the spec. level rather than on the true quality of the product. This is a reasonable assumption when customer testing for the precise quality level and utilizing subsequent results of the tests are uneconomical. In such cases, a customer must rely on the spec. level which is the (minimum) guaranteed quality level of a product.

For the discretization of customer type t, we employ the concept of customer blocks, or classes. Specifically,

we will assume that there are M customer blocks and customer block i consists of customers type $t \in [t_i, t_{i-1}]$ where $i = 1, \ldots, M$ and $t_0 = 1$. The customers of type $t_i, i = 1, \ldots, M$, will be referred to as boundary customers. The corresponding quantity demanded for block i is given by $D(t_i) - D(t_{i-1})$ while the boundary customer t_i's utility given a spec. u_j product will be $U(u_j, t_i)$. For notational convenience, we will denote $D(t_i)$ by D_i, $U(u_j, t_i)$ by U_{ji}, $\boldsymbol{B}$ by B, and $\boldsymbol{b}$ by b.

In what follows, because of our assumption of stationary model environments, we present the decision processes and decisions of both producer and customers for the case of $T = 1$. For the cases of $T \geq 2$, the optimal decision processes and decisions remain the same. Only the economic quantities resulting from the optimal decisions such as the total profits and the net customer benefits during T periods change (such economic quantities in a single period are multiplied by T).

2.2. *Derivation of the basic model and its formulation*

We first derive mathematical expressions for the priority supply rule and the delivery probabilities Pr_{ij} without downgrading (the downgrading option will be easily incorporated later). For this purpose we introduce variables denoting the amount of products available and the amount of shortage/surplus under each contingency $b \in \beta$ as follows:

V_{jb}: the total number of products of spec. level u_j when $B = b$.

Q_{ijb}: the remaining demand in class i after using up spec. j products when $B = b$.

R_{ijb}: the remaining supply of spec. j products after supplying class i when $B = b$.

The relation between V_{jb} and the supply quantity function is as follows: $\forall\, b \in \beta, j = 1, \ldots, N$,

$$V_{jb} = S(u_j, b) - S(u_{j-1}, b). \tag{1}$$

The demand for each priority class is given by $D_i - D_{i-1}$ for $i = 1, \ldots, M$, where $D_0 = 0$. Consequently, we can express the shortage and surplus relations with respect to each class and spec. level under the priority supply recursively as follows: $\forall\, b \in \beta, j = 1, \ldots, N$, and $i = 1, \ldots, M$,

$$Q_{ijb} = \max\,[Q_{ij-1b} - R_{i-1jb}, 0], \tag{2}$$

$$R_{ijb} = R_{i-1jb} - (Q_{ij-1b} - Q_{ijb}), \tag{3}$$

where $R_{0jb} = V_{jb}, Q_{i0b} = D_i - D_{i-1}, t_0 = 1$, and $D_0 = 0$.

According to the priority supply rule, for any pair of (i,j), spec. u_j products allocated to class i are randomly distributed within class i. Hence, under any given contingency b, the conditional probability, $\mathrm{Pr}_{ij|b}$, is:

$$\mathrm{Pr}_{ij|b} = (Q_{ij-1b} - Q_{ijb})/(D_i - D_{i-1}). \tag{4}$$

From averaging $\mathrm{Pr}_{ij|b}$ over all possible contingencies, we have: for $i = 1, \ldots, M$ and $j = 1, \ldots, N$,

$$\mathrm{Pr}_{ij} = \sum_{b \in \beta} \mathrm{Prob}\,\{b\}((Q_{ij-1b} - Q_{ijb})/(D_i - D_{i-1})). \tag{5}$$

We now turn to modeling the customer's decisions. We assume that customers are expected-value decision makers and an identical price table is provided to all potential customers. Then, each customer t's expected utility when he orders a unit of priority class i are given by,

$$EU_i(t) \quad \sum_{j=1}^{N} \mathrm{Pr}_{ij} U(u_j, t). \tag{6}$$

The corresponding net expected utility, $NEU_i(t)$, is equal to $EU_i(t) - p_i$. The optimal customers' behavior or self-selection is simply to choose priority level $\hat{i}$, where $NEU_{\hat{i}}(t) = \max_i NEU_i(t)$.

Given the price schedule of the producer (see Table 1) and the optimal customers' behavior, market segmentation is the result of the aggregate responses by the customers optimizing their net benefits (i.e., net expected utility). In order for there to be successful market segmentation, a certain monotonicity condition on utility functions has been a standard assumption in the literature dealing with product line type pricing (see, for example, Smith, 1986). A set of such monotonicity conditions that our model satisfies (it can be easily verified) is as follows: for $i = 1, \ldots, M$,

$$(EU_i(t_h) - EU_{i+1}(t_h)) - (EU_i(t_k) - EU_{i+1}(t_k)) > 0 \quad \text{for } 1 \geq t_h > t_k > 0, \tag{7}$$

where $EU_i > EU_{i+1}$ for any $t \in [0, 1]$ is assumed.

The above condition states that the valuation difference between the priority classes increases in t.

Because the basic model satisfies the monotonicity conditions (7), we can represent the market segmentation of all customers in terms of the following boundary customers relations, given appropriate prices, $p_1, \ldots, p_M$: for $i = 1, \ldots, M-1$,

$$\sum_{j=1}^{N} \mathrm{Pr}_{ij} U(u_j, t_i) - p_i = \sum_{j=1}^{N} \mathrm{Pr}_{i+1j} U(u_j, t_i) - p_{i+1}, \tag{8}$$

and for $i = M$,

$$\sum_{j=1}^{N} \mathrm{Pr}_{Mj} U(u_j, t_M) - p_M = 0. \tag{9}$$

The above relations state that the boundary customer $t_i, i = 1, \ldots, M-1$, is indifferent between purchasing priority class i and $i+1$, and the last boundary customer t_M is indifferent between subscribing to priority level M

or withdrawing from the market. Also it can be easily verified that all non-boundary customers of type $t \in [t_i, t_{i-1}]$ will purchase priority class $i, i = 1, \ldots, M$. Throughout the rest of the paper, we will refer to the monotonicity conditions (7) as the efficient monotonicity conditions because, given appropriate prices, higher ranked customers will receive higher spec. level chips by self-selecting higher priority classes.

Employing the various relations of (1)–(6) and (8)–(9), the entire formulation for the producer's profit maximization problem is shown as follows.

$$\max \pi = \sum_{i=1}^{M} p_i(D_i - D_{i-1}) \tag{10}$$

s.t. $1 = t_0 \geq t_1 \geq \ldots \geq t_M; \sigma = u_0 \geq u_1 \geq \ldots \geq u_N \geq \rho;$ $p_1 \geq p_2 \geq \ldots \geq p_M;$

all variables $(t_i, u_j, p_i, \mathrm{Pr}_{ij}, U(u_j, t_i), S(u_j, b), Q_{ijb}, R_{ijb},$ and $D_i, \forall\, i, j,$ and b defined) ≥ 0.

Boundary customers relations: for $i = 1, \ldots, M-1$,

$$\sum_{j=1}^{N} \mathrm{Pr}_{ij} U(u_j, t_i) - p_i = \sum_{j=1}^{N} \mathrm{Pr}_{i+1\,j} U(u_j, t_i) - p_{i+1}, \tag{11}$$

and for $i = M$,

$$\sum_{j=1}^{N} \mathrm{Pr}_{Mj} U(u_j, t_M) - p_M = 0. \tag{12}$$

Production-classification relations: $\forall\, b \in \beta, j = 1, \ldots, N,$

$$V_{jb} = S(u_j, b) - S(u_{j-1}, b). \tag{13}$$

Priority supply relations: $\forall\, b \in \beta, j = 1, \ldots, N,$ and $i = 1, \ldots, M,$

$$Q_{ijb} = max[Q_{ij-1b} - R_{i-1jb}, 0] \tag{14}$$

$$R_{ijb} = R_{i-1jb} - (Q_{ij-1b} - Q_{ijb}). \tag{15}$$

Allocation probability relations: for $i = 1, \ldots, M$ and $j = 1, \ldots, N,$

$$\mathrm{Pr}_{ij} = \sum_{b \in \beta} \mathrm{Prob}\{b\}((Q_{ij-1b} - Q_{ijb})/(D_i - D_{i-1})). \tag{16}$$

The downgrading feature discussed earlier can be easily incorporated into the basic model as follows. Let W_{ijb} be the amount of originally u_i chips $(u_i > u_j)$ relabelled as u_j chips when $B = b$, while W_{iib} be the amount of u_i chips retaining the original spec. level u_i when $B = b$. Then the relations among W_{ijb}, $S(u_j, b)$, $S(u_{j-1}, b)$, and V_{jb} are as follows: $\forall\, b \in \beta, j = 1, \ldots, N,$

$$S(u_j, b) - S(u_{j-1}, b) = W_{jjb} + \ldots + W_{jNb}, \tag{17}$$

$$V_{jb} = W_{1jb} + \quad + W_{jjb}. \tag{18}$$

By simply replacing the production-classification relations of (13) with the production-downgrading-classification relations of (17) and (18), the downgrading feature is incorporated. We now proceed to analyze the market segmentation resulting from the above profit maximization formulation.

3. Market segmentation analysis

As mentioned in the introduction, we employ the single exchange perturbations in the priority supply rule so as to analyze the market segmentation under our pricing scheme for the profit-maximizing producer. Such an analysis of the market segmentation, in turn, enables us to examine the profitability of downgrading and the optimality of the priority supply rule. To facilitate the single exchange perturbation analysis, which is the main instrument of analysis in this section, we first define two economic quantities, $\bar{p}_i$ and a_{ijb}, as follows.

$\bar{p}_i$: the incremental price paid for upgrading priority class $i+1$ to class i, $i = 1, \ldots, M-1$;

a_{ijb}: the number of chips j delivered to priority class i, after downgrading, under the priority supply rule when $B = b$.

We will let $\bar{p}_M = p_M$ be the 'base' price that customers of all priority classes must pay and refer to $\bar{p}_i$, $i = 1, \ldots, M$ as the premium price for priority class i. From the boundary customers relations (8)–(9), for premium prices $\bar{p}_i$, we have: for $i = 1, \ldots, M-1$,

$$\bar{p}_i = EU_{ii} - EU_{i+1i}$$

and,

$$\bar{p}_M = EU_{MM}, \tag{19}$$

where $EU_{ii}(EU_{i+1i})$ is the expected utility of the boundary customer i selecting priority class $i(i+1)$. The corresponding profit π is given by

$$\pi = \bar{p}_1 D_1 + \bar{p}_2 D_2 + \ldots + \bar{p}_M D_M. \tag{20}$$

By employing relations (19) and (20), it can be verified that the following proposition regarding the relation between the producer's profit and a_{ijb} holds.

Proposition 1. *For the profit maximization model under the priority supply rule, the profit is linear in a_{ijb} and the corresponding per unit contribution to the profit is given by*

$$c_{ijb} = \mathrm{Pr}\{b\}((-U_{ji-1}D_{i-1} + U_{ji}D_i)/(D_i - D_{i-1})). \tag{21}$$

Since c_{ijb} of (21) is proportional to the probability of contingency b, it is convenient to define the conditional per unit contribution to the profit that is independent of any contingency. We denote the conditional per unit contribution by c_{ij}, where $c_{ij} = c_{ijb}/\mathrm{Pr}\{b\}$. We note that the c_{ij} depends on the market segment sizes and utility

functions. The conditional per unit contribution c_{ij} may or may not be positive. This is the motivation for studying single exchange perturbations in the priority supply rule. We will show that such perturbations in the priority supply rule may or may not yield a positive change in profit. This in turn may indicate the profitability of downgrading as well as the optimality of the priority supply rule. We now present the following corollary concerning the single exchange perturbations. The proof can be easily obtained from Proposition 1.

Corollary 1. *Let us suppose that we perturb the priority supply rule of our model by exchanging one unit of u_i chip from class h with one unit of u_j chip from class k under contingency $b \in \beta$ and adjust prices accordingly so as to preserve the original market segmentation (i.e., the boundary customer $t_i, i = 1, \ldots, M$ remains the same). We further assume that this perturbation is sufficiently small that the efficient monotonicity conditions (shown in (7)) are still satisfied. Then $\delta(\pi)$, the change in the profit, is given by:*

$$\begin{aligned}\delta(\pi) =& \Pr\{b\}[(U_{ih-1} - U_{jh-1})D_{h-1} \\ &+ (-U_{ih} + U_{jh})D_h]/(D_h - D_{h-1}) \\ &+ \Pr\{b\}[(-U_{ik-1} + U_{jk-1})D_{k-1} \\ &+ (U_{ik} - U_{jk})D_k]/(D_k - D_{k-1})\end{aligned} \tag{22}$$

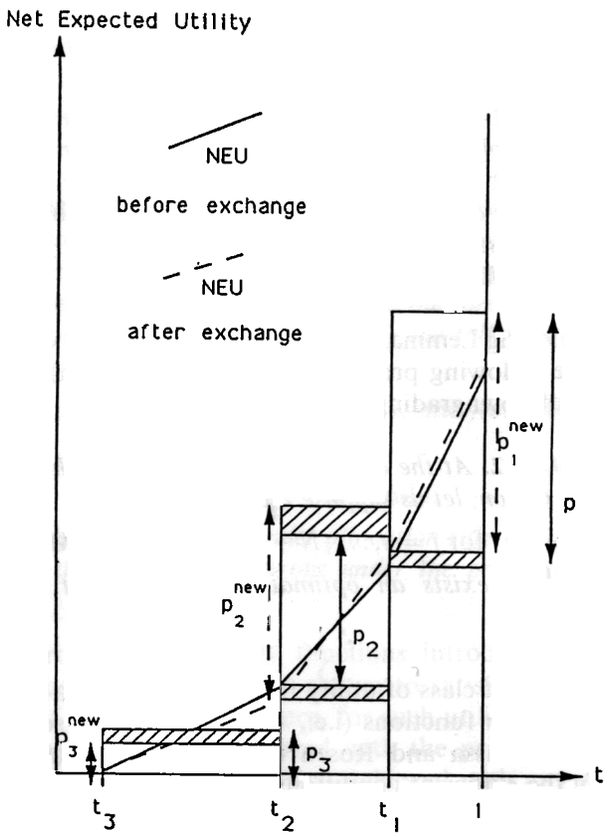

Fig. 1. Impact of single exchange on the market segmentation.

or equivalently,

$$\delta(\pi) = -c_{hib} + c_{hjb} + c_{kib} - c_{kjb}.$$

Fig. 1 illustrates the implication of Corollary 1 for the case of $M = 3$. For simplicity the expected utility functions are assumed to be linear in t. We show the effect of exchanging one unit of u_i chip ($u_i > u_j$) from market segment 3 with one unit of u_j chip from market segment 2 while maintaining the same market segmentation. Such an exchange increases the expected utility of priority class 2 while it decreases the expected utility of priority class 3. Consequently the producer is able to increase the premium price $\bar{p}_2$. However, in order to prevent customer crossovers (i.e., without price reductions, some customers of class 1 and class 3 may switch to class 2), it may be necessary to reduce the premium prices $\bar{p}_1$ and $\bar{p}_3$. The change in the prices and the profit can be seen in Fig. 1. Prices p_1, p_2, and p_3 (not premium prices $\bar{p}_1$, $\bar{p}_2$, and $\bar{p}_3$) are represented by the arrows, and the changes in profit are represented by the shaded rectangles. We note that the total change in profit can be either positive or negative.

Now consider a case where one unit of u_i chip ($u_i > u_j$) is downgraded as u_j chip in class k when $B = b$. This transaction can be viewed as a degenerate single exchange where a u_i chip is exchanged with a u_j chip from an external imaginary source. Then the incremental profit $\delta(\pi)$ is given by

$$\begin{aligned}\delta(\pi) = \Pr\{b\}&[(U_{ik-1} - U_{jk-1})D_{k-1} \\ &+ (-U_{ik} + U_{jk})D_k]/(D_k - D_{k-1}),\end{aligned} \tag{24}$$

which is the difference between the per unit contribution of a_{kjb} and a_{kib}. Again the resulting $\delta(\pi)$ can be either positive or negative. Fig. 2 depicts the net effect when a u_i chip is downgraded as a u_j chip in the second market segment. Here the example is drawn in such a way that for the producer, the extra profit [A] is much smaller than the loss [B]. Thus, in this case, downgrading decreases profit. This situation, however, is reversed in Fig. 3, which shows larger [A] relative to [B] resulting in an increased profit.

3.1. *Analysis of downgrading*

As shown in Fig. 3, in general the producer may be able to achieve a higher level of profit by downgrading without cannibalizing the neighboring market segments. An important question arises as to under what condition(s) downgrading is (un)profitable. In this subsection we focus on the unprofitability of downgrading and develop sufficient conditions under which profit maximization will be achieved without downgrading. Such investigation is of importance because whenever these sufficiency conditions are met, the producer may rule out down-

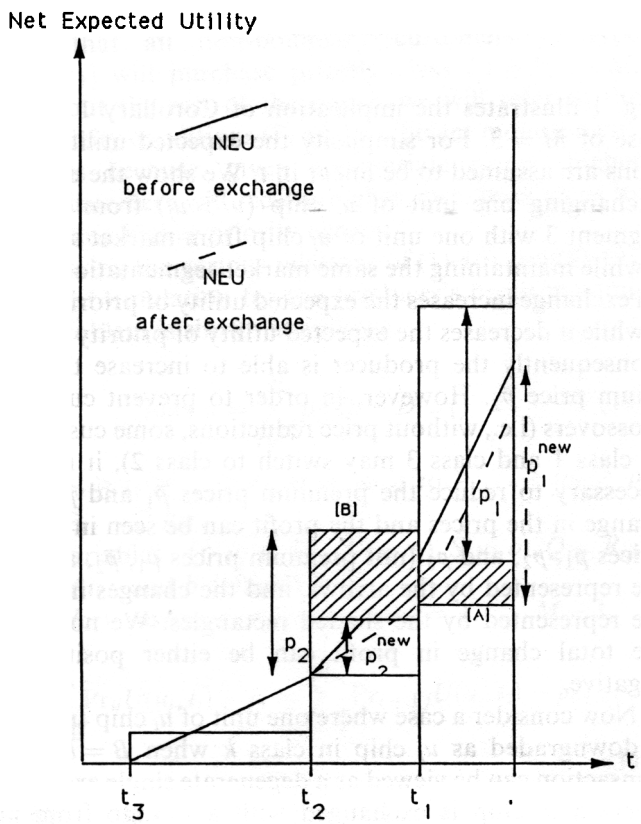

Fig. 2. Impact of downgrading on the market segmentation: decreased profit.

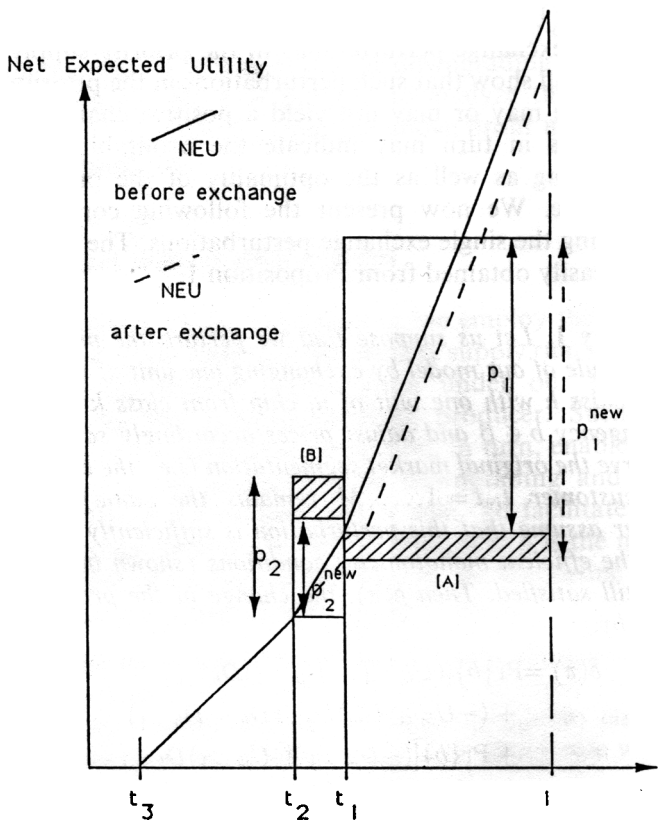

Fig. 3. Impact of downgrading on the marked segmentation: increased profit.

grading as a strategic option. This in turn will eliminate the possibility of 'customer bypass' on the downgraded chips (i.e., reinspection and reclassification of spec. levels by customers) even when price differentials among spec. levels make such bypass economically attractive. It will also simplify the formulation by reducing the number of variables and thus facilitate numerical solutions.

We now turn our attention to analytic characterization of (un)profitability of downgrading. From numerical calculations of various profit maximization examples with downgrading, it is observed that profitability of downgrading may be inferred from the values of c_{ij}. The explicit relation between the profitability of downgrading and the values of c_{ij} is formalized in Proposition 2 below. The proof of Proposition 2 (as well as Proposition 3 in the following subsection) relies on the following lemma and corollary concerning the priority supply rule. The proofs can be obtained by simply applying the definition of the priority supply rule.

Lemma 1. *Given any allocation outcome under the priority supply rule, no single exchange or unilateral transfer (i.e., unilateral reallocation of a chip from one class to another class) that upgrades the average quality of the product mix of the higher priority class can be made between any two priority classes k and l ($k < l$) for all possible contingencies.*

Corollary 2. *Given any allocation outcome under the priority supply rule, no single exchange or unilateral transfer that upgrades the product mix of the higher priority class can be made between any two* neighboring *priority classes k and $k+1$, $k = 1, \ldots, M-1$, for all possible contingencies.*

By employing Lemma 1 and Corollary 2, it can be verified that the following proposition regarding the (un)profitability of downgrading holds.

Proposition 2. *At the optimality of our model with downgrading option, let us suppose that*

$$c_{ij} \geq c_{ij+1} \quad \text{for } j = 1, \ldots, N-1 \text{ and } i = 1, \ldots, M. \quad (25)$$

Then, there exists an optimal solution without downgrading.

An important class of utility functions referred to as the product form functions (i.e., $U(u,t) = g(u)h(t)$; see, for example, Mussa and Rosen (1978)) can be shown to satisfy the sufficient conditions of Proposition 2. Therefore, for such utility functions, the maximum profit is achievable without downgrading.

3.2. *Optimality of the priority supply rule*

The priority supply rule of our model, which provides higher spec. level chips to higher ranked customers, is the most desirable allocation rule given our product line format from an economic efficiency point of view. The question still remains, however, whether there are alternative allocation rules that will result in higher profits than the priority supply rule. In addressing this question, we will restrict our investigation to contingency allocation rules that will satisfy the efficient monotonicity conditions. The definitions of a_{ijb} and c_{ijb}, shown in the beginning of Section 3, are extended so that they are applicable to *any* contingency allocation rule that satisfies the efficient monotonicity conditions. Furthermore, Lemma 1 and Corollary 2 of the previous subsection are true for the allocation outcomes under any alternative allocation rules that are identical to the allocation outcomes under the priority supply rule. We note that even though the allocation outcomes under an alternative allocation rule cannot be identical to those under the priority supply rule for all model parameters, they can be identical for a specific, given set of model parameters (e.g., the number of spec. levels and priority classes and/or the distribution of Prob$\{b\}$).

As in the previous section, numerical calculations of various examples of profit maximization models suggest that the relative profitability of a particular contingency allocation rule (as compared to the priority supply rule) may be inferred from c_{ij} values. The explicit relation of the profitability of a particular contingency allocation rule and the values of c_{ij} is formalized in the following proposition, which can be verified by employing Lemma 1 and Corollary 2.

Proposition 3. *Let us suppose that under an allocation rule R, which differs from the priority supply rule, the profit is maximized and the resulting solution satisfies the efficient monotonicity conditions. Suppose that the corresponding c_{ij} values at the optimal solution satisfy the following conditions:*

$$c_{ki} - c_{kj} - c_{k+1i} + c_{k+1j} \geq 0 \quad \forall\, k, i, \text{ and } j(u_i > u_j) \quad (26)$$

and,

$$c_{ki} \geq c_{k+1i} \quad \forall\, k \text{ and } i. \quad (27)$$

Then, the optimal profit under allocation rule R cannot be higher than the optimal profit under the priority supply rule.

The product form utility functions introduced in the previous subsection can be shown to satisfy the conditions of Proposition 3. Hence, for such utility functions, the maximum profit achieved with the priority supply rule is at least as high as that achieved with any other contingency allocation rule that satisfies the efficient monotonicity condition.

4. Two-part pricing and an illustrative example

We turn our attention to the problem of how the producer actually charges the customers. Until now the prices of the priority classes are defined to be the total charges to be collected from each customer according to the priority class he selects. However, in practice, if the total charge is collected after the delivery, customers may be tempted to place multiple orders with the intention of canceling some orders depending on the production outcome. On the other hand, if the total charge is collected before the delivery, it might be objectionable to most customers.

A reasonable approach to resolve this problem is to decompose the price of each priority class into two parts. That is, the price consists of a nonrefundable priority charge paid in advance to secure delivery of a given priority, and a delivery charge ex post for the chips actually delivered. To discourage multiple orders, the delivery charge paid by customers upon delivery should depend on the actual chips delivered, i.e., the same chips delivered under different priorities cost the same. The format of such two-part pricing is analogous to two-part tariffs discussed in capacity pricing (e.g., installation charge and usage charges of Oren *et al.*, (1985)) or in electric power pricing (e.g., demand charge and energy charge of Chao *et al.* (1986)).

The customers are expected-value decision makers as we have assumed throughout this paper. Thus, a customer's priority selection depends on the expected total price he pays, which is given by

$$p_i = \hat{p}_i + \sum \mathrm{Pr}_{ij}\tilde{p}_j, \quad (28)$$

where $\hat{p}_i$ is the priority charge for class i while $\tilde{p}_j$ is the delivery charge for spec. u_j.

The decomposition of these prices into priority and delivery prices may not be unique unless the probabilities Pr_{ij} have some special structure. When such ambiguity arises, additional criteria may be imposed. For example, in the electric power industry, the demand charge (ex-ante) corresponds to the cost of fixed capital whereas the energy (delivery) charge attempts to reflect the marginal cost of producing electric power.

In the following example, we elucidate some important features of our model. The production data of a certain kind of chip called product 10c from an anonymous

Table 2. Actual production data between October and December 1987

Production run	*Total*	*Bin 1*	*Bin 2*	*Bin 3*	*Reject*
1	616	162	353	101	0
2	618	115	423	78	2
3	617	137	378	101	1
4	618	111	416	91	0

Table 3. Price table of the example

Class	Price	Pr_{ij}		
		$u_1 = 0.89$	$u_2 = 0.778$	$u_3 = 0.585$
1	$p_1 = 0.51$			
2	$p_2 = 0.319$			

semiconductor manufacturer (R. Leachman, personal communication, 1988) are as follows.

In the period of October–December 1987, there were four production runs for this particular chip. After testing, each chip was classified into one of the four bins according to the test results. The bins are numbered so that bin i chips have higher spec. level than bin $i+1$ chips $\forall\, i$. Because chips of this particular kind are graded according to their overall performance on the tests, the corresponding spec. levels do not have any physical unit (e.g., nanosecond). In this example we assume that the spec. levels of bins 1, 2, and 3 have been set at 0.65, 0.35, and 0.05 respectively (these spec. levels are by no means optimal). In order to fit the available data to obtain a supply quantity function, we choose a quadratic functional form shown below.

$$S(u, B) = 1 - (1 - B)u^2 - Bu.$$

Using least-square error fitting for each production run and denoting the realization of B by b_1, b_2, b_3, and b_4, we obtain the resulting estimates: $b_1 = 0.7691$, $b_2 = 0.8545$, $b_3 = 0.857$, and $b_4 = 0.9044$. The demand data are assumed to be as follows: $D(t) = 1 - t$ and $U(u, t) = ut$ for $t \in [0, 1]$ and $u \in [0, 1]$. We also assume that the number of priority classes $M = 2$ and the number of spec. levels $N = 3$. The resulting optimal solutions for this problem are summarized in Tables 3 and 4.

From Tables 3 and 4, we note the following:

1) the corresponding optimal profit from (10) is calculated to be $\pi = 0.191$;

2) since the utility function is of product form, without any calculation, we deduce that at the optimal profit under the priority supply rule is higher than or equal to the optimal profit under any contingency supply rule that satisfies the efficient monotonicity conditions (by Proposition 3). Furthermore, again owing to the product form utility function, we deduce that at the optimal solution, downgrading is unprofitable (by Proposition 2). Therefore, *a priori*, we could proceed with the formulation of Section 2 without downgrading option.

The price table (Table 3) lists only the expected total price for each priority. Using the numerical values obtained above, we have from (28) the following system of linear equations:

Table 4. Contingency delivery table of the example (a_{ij} denotes the number of u_j chips delivered to class i)

B value	Class 1	Class 2
b_2		
b_3		
b_4		

$$p_1 = 0.51 = \hat{p}_1 + 0.508\,\tilde{p}_1 + 0.484\,\tilde{p}_2 + 0.008\,\tilde{p}_3,$$
$$p_2 = 0.319 = \hat{p}_2 + 0.019\,\tilde{p}_2 + 0.967\,\tilde{p}_3.$$

The solution for the above equations is not unique because the system of equations is under-determined (two equations and five unknowns). A set of prices that satisfy the above relations is as follows: the delivery charges, $\tilde{p}_1 = 0.459$, $\tilde{p}_2 = 0.401$, and $\tilde{p}_3 = 0.322$; the priority charges, $\hat{p}_1 = 0.08$ and $\hat{p}_2 = 0$.

Acknowledgement

The authors thank two anonymous referees and the associate editor for comprehensive and constructive comments. The second author's research is supported by a National Science Foundation grant IRI-8902813, and the first author's research is supported by a National Science Foundation grant DDM-9110945. Also the first author thanks Cheng-Kang Chen, Ming-Jong Yao, and Tzong-Yih Chiu for helpful comments.

References

Antle, R. and Eppen, G. (1985) Capital rationing and organizational slack in capital budgeting. *Management Science*, **31**, 163–174.

Chao, H., Oren, S., Smith, S. and Wilson, R. (1986) Multiple demand subscription pricing for electric power. *Energy Economics*, **4**, 199–217.

Intel (1985) *Component Quality/Reliability Handbook*.

Leachman, R. (1986) Preliminary design and development of a corporate-level production planning system for the semiconductor industry. Report No. ORC 86-11, Operations Research Center, University of California at Berkeley.

Mussa, M. and Rosen, S. (1978) Monopoly and product quality. *Journal of Economic Theory*, **18** (2), 301–317.

Oren, S., Smith, S. and Wilson, R. (1982) Linear tariffs with quality discrimination. *The Bell Journal of Economics*, **13**, 455–471.

Oren, S., Smith, S. and Wilson, R. (1983) Competitive nonlinear tariffs. *Journal of Economic Theory*, **29** (1), 49–71.

Oren, S., Smith, S. and Wilson, R. (1985) Capacity pricing. *Econometrica*, **53** (3), 545–567.
Smith, S. (1986) New product pricing in quality sensitive markets. *Marketing Science*, **5** (1), 70–87.
Wilson, R. (1989) Efficient and competitive rationing. *Econometrica*, 57 (1), 1–40.

Biographies

K. Jo Min is Assistant Professor in the Department of Industrial and Manufacturing Systems Engineering at Iowa State University. He holds a Ph.D. in Industrial Engineering and Operations Research from the University of California at Berkeley. His research include production systems/inventory control, design and analysis of allocation systems, and pricing theory.

Dr Shmuel S. Oren is Professor of Industrial Engineering and Operations Research at the University of California at Berkeley. He holds a B.S. and M.S. in Mechanical Engineering from the Technion in Israel and an M.S. and Ph.D. in Engineering Economic Systems from Stanford. Before his current position he was on the faculty of The Engineering Economic Systems Department at Stanford University and he worked for eight years as a research scientist at the Xerox Palo Alto Research Center. His research interests include optimization theory, modeling and analysis of economic systems, coordination and decentralization through market mechanisms and electric utility planning and operation.