Semiconductor Capacity Planning: Stochastic Modeling and Computational Studies

Robert M. E. Christie
Delphi Technical Center

S. David Wu
Lehigh University

Report No. 00T-001

Semiconductor Capacity Planning: Stochastic Modeling and Computational Studies

Robert M.E. Christie
Delphi Technical Center

S. David Wu
Manufacturing Logistics Institute
Department of Industrial and Manufacturing Systems Engineering
Lehigh University

Abstract

This paper presents a multistage stochastic programming model for strategic capacity planning at a major U.S. semiconductor manufacturer. Main sources of uncertainty in this multi-year planning problem include *demand* of different technologies and *capacity estimations* for each fabrication (fab) facility. We test the model using real-world scenarios requiring the determination of capacity configuration for twenty-nine technology categories among five fab facilities. The objective of the model is to minimize the gaps between *product demands* and the *capacity* allocated to the technology specified by each product. We consider two different scenario-analysis constructs: first, an *independent* scenario structure where we assume no *prior information* and the model systematically enumerate possible states in each period. The states from one period to another are independent from each other. Second, we consider an *arbitrary* scenario construct which allows the planner to sample/evaluate arbitrary multi-period scenarios that captures the dependency between periods. In both cases, a *scenario* is defined as a multi-period *path* from the root to a leaf in the scenario tree. We conduct intensive computational experiments on these models using real data supplied by the semiconductor manufacturer. The purpose of our experiments is two folds: first to determine the number of scenarios necessary for the *independent* model to achieve high-quality capacity configuration. Using this as a benchmark, we then compare the results from the *arbitrary* model and illustrate the different uses of the two scenario constructions. The modeling and empirical testing provide some basic insight on the complex decision problem of semiconductor capacity planning.

1. Introduction

Capacity planning is central to the competitiveness of a semiconductor firm. Many industries are confronted with high capital requirements of capacity, but few are competing in the environment of short product life cycles, near-continuous technological innovation, and long manufacturing lead-time as in semiconductor. At any one time, a semiconductor manufacturer must produce a variety of products in a number of different production facilities as they endeavor to capture market share while meeting the requirements of their current customers. The flexibility necessary to produce different technologies at the same fab results in another puzzle: how should the existing capacity be configured in order to meet customer demand? This problem of *capacity configuration* becomes especially challenging when coupled with inherently inaccurate demand forecasts and imprecise measures of capacity.

We will organize our presentation using real scenarios and considerations encountered during the annual capacity planning process for a major U.S. semiconductor manufacturer. The planning process involves determining the production location and quantity for some twenty-nine product technologies among five wafer fabrication facilities for each year in a five-year planning horizon. The main sources of uncertainty include demand forecast and capacity estimation, among others. To explain the current capacity planning process, we use a case where five facilities need to be configured for nine technologies. The process begins with determining the capacity estimates for each fab and demand forecasts for each technology. Table 1 details a disguised five-year demand forecast for each of the nine technologies, where demand is expressed in wafer starts per week (WSPW).

Table 1: Sample 5-Year Demand Forecast

Technology Code	Demand (WSPW)				
	Year				
	1	2	3	4	5
1	-	-	114	845	1310
2	-	-	51	792	1353
3	156	1348	2001	1616	1307
4	-	165	1550	1668	1366
5	849	747	485	417	395
6	359	572	657	359	443
7	2708	1982	1092	763	614
8	684	669	433	290	250
9	175	120	75	56	19

Since a wafer in each technology category does not consume the same amount of capacity (due to different masking layers, yields, etc.), a normalization factor is used to equalize the capacity statement from each fab to the same unit. These normalization factors generally range from 0.7 to 4.0, depending on the manufacturing complexity of the technology. The capacity planner then determines how the fabs should be configured to best meet the projected demand for product technologies. In practice, this is an iterative process via manual scenario analysis. That is, with the help of a spreadsheet, the planner establishes an initial configuration for all fabs to match capacity with technology demand based on a set of criteria justified by managerial preferences (such as present capacity configurations and the desire to configure the new fabs for the latest technologies). Table 2 lists an example configuration by location based on the data given in Table 1. Table 3 summarizes this same configuration grouped by technology. Given a particular configuration, a "gap" is then calculated as the difference between *allocated capacity* and *product demand*. From this initial configuration, the planner adjusts the configuration with the goal of minimizing this gap. These adjustments continue until the planner is satisfied that a good configuration has been determined and a "minimum" gap has been reached. The gap for the example configuration is provided in Table 4. A positive gap indicates excess capacity for a particular technology while a negative gap value (indicated with parentheses) indicates a shortage of capacity for a particular technology.

Table 2. Example Capacity Configuration Grouped by Facility

Location Code	Technology Code	Capacity (WSPW) Year 1	2	3	4	5
	5	-	58	202	62	-
	6	-	-	-	141	202
A	7	1842	2087	1835	1835	1835
	8	19	-	-	-	-
	7	680	730	728	728	728
B	8	325	294	270	270	270
	9	0	0	30	117	117
	3	99	421	426	398	507
	4	0	240	643	670	601
	5	647	409	38	-	-
C	6	297	348	301	224	194
	7	338	-	-	-	-
	3	-	295	939	447	-
	4	-	-	383	378	6
	1	-	-	-	569	659
D	2	-	-	-	62	683
	3	-	396	360	764	738
	4	-	-	751	868	894
	5	271	381	48	-	-
E	6	35	309	457	-	-
	7	1187	465	-	-	-

Table 3. Example Capacity Configuration Grouped by Technology

Technology Code	Capacity (WSPW) Year 1	2	3	4	5
1	-	-	30	685	775
2	-	-	-	62	683
3	99	1113	1725	1671	1929
4	0	240	1777	1915	1500
5	918	848	288	62	-
6	332	657	758	365	397
7	3367	2552	1835	1835	1835
8	699	730	728	728	728
9	325	294	270	270	270

Table 4. Gap Between Demand and Capacity for Example Configuration

Technology Code	(Demand – Capacity) WSPW Year				
	1	2	3	4	5
1	-	-	84	159	535
2	-	-	51	731	670
3	57	235	276	(55)	(622)
4	-	(76)	(227)	(248)	(134)
5	(69)	(100)	196	355	395
6	27	(85)	(101)	(6)	46
7	(659)	(570)	(743)	(1072)	(1221)
8	(14)	(61)	(295)	(438)	(479)
9	(150)	(174)	(194)	(213)	(251)

The above manual capacity planning process is not uncommon in the semiconductor industry, and it presents an opportunity for improvement. The planner's use of spreadsheets allows for quick calculations during the planning process, however, to consider a sufficient number of configurations through manual scenario analysis would be an overwhelming task. The quality of the final configuration is thus dependent on the quality of the configuration considered by the planner in the first place. A more systematic approach to scenario analysis would be of great benefit. Furthermore, the demand data are generated from sales forecasts, which, over a five-year period, are notoriously inaccurate. Capacity estimation is also considered imprecise because the amount of capacity consumed by each technology is not easily quantified. The current capacity measurement, wafer starts per week (WSPW), attempts to normalize each of the technologies according to the respective capacity requirements. Different normalizing factors are used in the five facilities and are often difficult to standardize because each facility has different yields, equipment capabilities, and production policies.

The capacity planning process is a crucial activity for semiconductor manufacturer. Long lead times, high capital costs, and strict outsourcing arrangements present a challenging environment for capacity planners who must cope with the difficulty of manually considering different capacity configurations and the stochastic nature of the process inputs. Providing a set of tools to enhance capacity planning would allow the firm to make better use of its existing capacity and to minimize its need for outsourcing. In the remainder of the paper, we present a stochastic programming approach to the semiconductro capacity planning.

2. A Multistage Stochastic Programming Model for the Capacity Configuration Problem

We propose a multistage stochastic linear program (MSLP) with scenario analysis for the capacity-planning problem in semiconductor manufacturing. The model uses scenarios constructed from the expected variation from *demand forecasts* over the planning period and *capacity variations* from the initial estimate. Both demand and capacity allocation are defined based on aggregate product technologies. As a convention, technologies are first aggregated according to the line width, which ranges from 0.18μm to greater than 2.0μm, then further classified by an additional technology category, such as linear or digital.

Considering possible scenarios of demand and/or capacity variations over the planning period, the stochastic program determines the capacity configuration for each facility in each time period. The objective is to minimize the sum of all differences (closing the gap) between the capacity allocated for a particular technology and that technology's corresponding demand. Most customers require the fab (or fabs) to submit to a certification process for each technology category. This process involves a trial run of a specified batch of wafers under the specific technology, subject to yield and quality requirements. Since this process typically involves significant cost and each facility has different production and equipment capability restrictions, not all technologies can be produced at all fabs in each time period. In the case where demand exceeds manufacturing capacity for a particular technology, The manufacturer typically outsources the production to one or many of its manufacturing partners (referred to as foundries).

The usefulness of this MSLP model extends beyond the determination of configurations for each facility in each time period. The model can be used to evaluate the robustness of a particular capacity configuration via scenario analysis. For example, given their capacity configuration plans for a particular period, the capacity planners may be interested to know how a specific set of demand scenarios for a particular group of technologies would affect the firm's ability to meet that demand. On the other hand, when making capital expenditure decisions, the planner may be interested to know investing in which capacity category would likely to produce the most beneficial rate of return.

2.1 Model Formulation

We adopt the following notation to construct a base MSLP similar that of Gassmann and Ireland (1995).

Notation:

t = 1,…,T denotes time periods (years)
i = 1,…, M denotes product technologies
j = 1,…, F denotes wafer fabrication facilities or fabs
s = 0,…, S denotes scenarios as described by an underlying *scenario tree*
g_i = 0,…,G^i denotes technology groups
g_j = 0,…, G^j denotes facility groups
e_s = $(e_{s1}, e_{s2}, \ldots, e_{sT})$ denotes a sequence of events over different periods under scenario s
$(e_s)_t$ = indicates that a variable or parameter is under event sequence e_s up to and including period t

Parameters:

w_{ij} = weighting factor for producing technology i at facility j
N_{ijt} = normalization factor for amount of capacity per wafer consumed by producing technology i at facility j, where $N_{ijt} = -1$ if facility j is incapable of producing technology i in time period t
D_{its} = demand for technology i in time period t in scenario s
C_{jt} = capacity (stated in equivalent units) of facility j in time period t
p_{st} = probability of having realized event sequence $(e_j)_t$
$\delta^{C}_{jtsg_i}$ = capacity change at facility j in period t in scenario s for group g_i (as a percentage)
$\delta^{D}_{jtsg_i}$ = demand change for technology i in period t in scenario s for group g_i (as

a percentage)

Decision Variable:

x_{ijts} = configure facility i to produce quantity of technology j in time period t under scenario s

Objective:

$$\text{Minimize } \sum_{i=1}^{M}\sum_{s=0}^{S}\sum_{t=1}^{T} p_{st}\cdot w_{ij}(D_{its}-\sum_{j=1}^{F} x_{ijts})$$

Demand Constraints:

$$\sum_{j=1}^{F} x_{ijts} \le (1+\delta_{tsg_i}^{C})\cdot D_{it} \qquad \forall i,t,s \qquad (1)$$

Capacity Constraints:

$$\sum x_{ijts} \le (1+\delta_{jstg_j}^{C})\cdot C_{jt} \qquad \forall j,t,s \qquad (2)$$

Fab-Technology Compatibility Constraints:

$$x_{ijts}\cdot N_{ijt} \ge 0 \qquad \forall i,j,t,s \qquad (3)$$

Consistent Configuration Constraints Over Multiple Periods:

$$\sum_{i=2}^{M} x_{ijts} \ge pct\cdot \sum_{i=2}^{M} x_{ij(t-1)s} \qquad \forall j,t,s \qquad (4)$$

Uniform First Period Configuration Constraint (*Nonanticipativity* Constraints):

$$x_{ijts} = x_{ijt(s+s')} \qquad \forall i,j,s \qquad (5)$$

Nonnegativity Constraints:

$$x_{ijts} \ge 0 \qquad \forall i,j,t,s \qquad (6)$$

The objective of this formulation is to minimize the total expected demand and capacity gap under all scenarios. The objective

$$\text{minimize } \sum_{i=1}^{M}\sum_{s=0}^{S}\sum_{t=1}^{T}(p_{st}\cdot w_{ij}(D_{its}-\sum_{j=1}^{F} x_{ijts}))$$

$$\equiv \text{minimize } (\sum_{i=1}^{M}\sum_{s=0}^{S}\sum_{t=1}^{T} p_{st}\cdot w_{ij}\cdot D_{its} - \sum_{i=1}^{M}\sum_{j=1}^{F}\sum_{s=0}^{S}\sum_{t=1}^{T} p_{st}\cdot w_{ij}\cdot x_{ijts})$$

$$\equiv \text{minimize } \sum_{i=1}^{M}\sum_{s=0}^{S}\sum_{t=1}^{T} p_{st}\cdot w_{ij}\cdot D_{its} - (\text{ maximize } \sum_{i=1}^{M}\sum_{j=1}^{F}\sum_{s=0}^{S}\sum_{t=1}^{T} p_{st} w_{ij} x_{ijts})$$

Since the first term is a constant, the problem is equivalent to the maximization problem defined by the second term. Or, in words, to maximize the total expected capacity allocation to all fabs. The weighting factor is used to encourage the model to impose preferences for assigning technology i to facility j. For example, the manufacturer prefers to assign more units to the company-owned facilities before assigning units to outside manufacturing partners. The weighting factor w_{ij} can also be used to express preferences of assigning a certain technology i to fab j. The probability value p_{st} specifies the probability of realizing a particular scenario s in period t, hereafter referred to as the path probability. Constraint set (1) state that the total number of units configured to produce technology i over all fabs shall not exceed the demand for that technology. The percentage for the demand change in time period t in scenario s, $\delta_{stg_i}^{D}$, adjusts the quantity demanded according to the scenario construction. Similarly, Constraint set (2) restrict the model to configuring a facility with total allocation no more than total capacity available. The capacity change value, δ^{C}_{stgl}, operates like its demand change counterpart by adjusting the actual capacity available at a particular facility under specific scenarios. The third set of constraints consider the compatibility between each technology and fab facility. If a facility is not capable of producing a particular technology, a value of -1 is assigned to the normalization factor, N_{ijt}. Thus, for an incompatible pair of technology i and facility j, the corresponding allocation x_{ijts} will be set to 0. Constraint set (4) require the model to suggest configurations that lead to consistent utilization of facility capacity over time. In the case of an individual facility, the model must assign a total configuration for year t that is at least a certain percentage (normally pct is set at 80%) of the previous year's configuration for that facility. This set of constraints is necessary because it is not practical for an individual facility to be utilizing a widely varying amount of its capacity from year to year. Constraint set (5) is a set of "*nonanticipativity*" constraints requiring the configuration variables for all scenarios in the first time period to be identical. This constraint is essential for the stochastic programming solution to be implementable. In particular, the configuration variables, x_{ij1s}, would be the same for all technology-fab pairs across all scenarios in time period 1. The configuration variables are not subject to this constraint after the first period. Because the configuration variables are different for each scenario after the first period, it is necessary to resolve the model each year after updating the demand forecast and capacity estimation data. The MSLP model determines a configuration for the first period given what is expected to occur over the five-year period.

3. Scenario Construction Using an Algebraic Modeling Language

Two types of scenario construction are used to implement the MSLP model. Both implementations use the same general model formulation except for the construction of the demand change parameter, δ^{D}_{stgm}, and the capacity change parameter, δ^{C}_{stgl}. A detailed explanation of each model is presented in Sections 3.1 and 3.2. A scenario is a sequence of events or outcomes over a specified time horizon. The events are characterized by a demand and capacity realization. A scenario tree is a collection of scenarios beginning from the same origin. The AMPL implementation of our MSLP model uses Gassmann and Ireland's (1995) data structure for the scenario tree construction.

To begin the explanation of the scenario tree data structures used in AMPL, we introduce a scenario tree with two outcomes from each node, resulting in 8 scenarios over T=3 time periods (Figure 1). Gassmann and Ireland uses the parameter $(e_s)_t$ to refer to a representative event. In Figure 1, each node in the scenario tree is a representative event, indexed by scenario, s, and time, t. In the context of our MSLP model, a representative event is a demand and capacity realization in a particular scenario. Accordingly, a scenario is a sequence of representative events. For example, scenario 0 is represented by the set of event sequences $\{e_{00}, e_{01}, e_{02}, e_{03}\}$ while scenario 3 is represented by the set $\{e_{00}, e_{01}, e_{22}, e_{33}\}$.

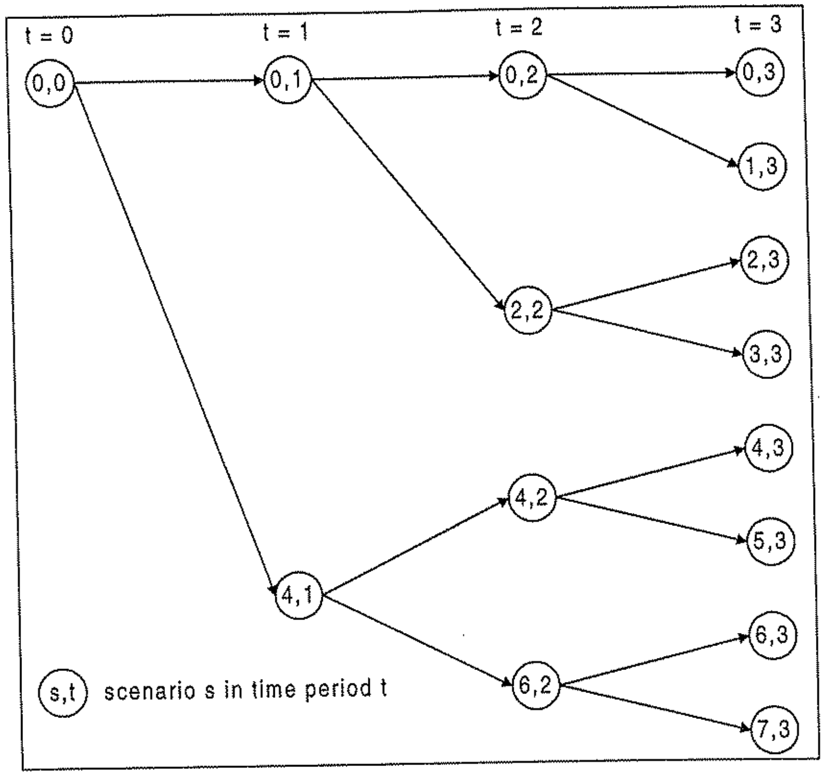

Figure 1: Scenario Tree

Each time period, *t*, is associated with the parameters *realizations(t)*and *branches(t).* The user-defined parameter *realizations(t)* indicates the number of outcomes from each node in a particular time period. The value of the parameter *realizations(t)* for the scenario tree in Figure 1 is 2 for all time periods because there are two outcomes from each node in this scenario tree. The parameter *branches(t)* is the number of branches stemming from a each node in that particular time period, calculated by AMPL as the product of all realizations beyond *t*. The value of *branches(1)* is thus 4.

The set scenarios is used to index all scenarios in the scenario tree. In addition to a set of event sequences (called *event_nodes(s,t)*), each scenario has associated with it the parameters starttime(s) and parent(s). The parameter parent(s) indicates the path from which a scenario branches. *parent(s)* is calculated in AMPL as:

parent{j in scenarios} := (if j=0 then 0 else j-(j mod branches[t])=0}t

In Figure 1, the parent of scenario 1 would be scenario 0, and the parent of scenario 3 would be scenario 2. The parameter *starttime(s)* indicates the time period in which a particular scenario branches off from its parent and continues with a different series of event sequences. AMPL calculates each value of *starttime(s)* as:

starttime{j in scenarios} := min{t in 1..T+1: (j mod branches[t])=0}t

The value for the *starttime(s)* parameter for scenario 2 in Figure 1 would be period 2. In the AMPL implementation of our MSLP model, each representative event is associated with parameter values for demand change, *demand_change(t,s,group),* and capacity change, *capacity_change(t,s,group)*. Each of these parameters is associated with a particular representative event in the scenario tree. To manage the demand change and capacity change parameters, we make further use of Gassmann's scenario tree structure by utilizing the parameters *rv_address(s,t)*, *previous(s,t)*, and *cond_prob(s,t)*. The parameter *rv_address(s,t)* associates each event node in the tree with the relevant user-defined distribution. *rv_address(s,t)* is calculated by AMPL as:

$$rv\_address[s,t] := 1 + ((j \text{ div } branches[t]) \text{ mod } realizations[t-1])$$

The parameter previous(s,t), indexed over event_nodes is calculated as:

$$previous[s,t] := \begin{cases} Parent[j] & \text{If } t = starttime[j] \\ J & \text{Otherwise} \end{cases}$$

The parameter *cond_prob(s,t)* is the conditional probability of each outcome according to the user-specified distribution. If each outcome is equally likely in Figure 1, the values of *cond_prob(s,t)* for representative event nodes (0,2) and (2,2) are both 0.5. *cond_prob(s,t)* is used to calculate *path_prob(s,t)*, which is the probability of arriving at a particular event node in the scenario tree. *path_prob(s,t)* for a particular representative event is calculated as the product of the *cond_prob(s,t)* values across the event nodes leading up to and including the particular representative event.

3.1 Independent Scenario Construction

The independent scenario construction considers a *fixed* number of outcomes from each node in the scenario tree. The purpose of the independent scenario construction is to consider the period-to-period fluctuations in product demand and facility capacity. The label "independent" stems from the fact that scenarios considered in each period are independent from the scenarios in previous periods. Thus, the scenario probabilities remain the same for each period, regardless of the previous outcomes. Use of this design does not require the user to make assumptions regarding future trends in how either the capacity or demand adheres to the forecast.

The distinguishing feature of the independent scenario construction is the structure of the demand change, $\delta^{D}_{stg_i}$, and capacity change, $\delta^{C}_{stg_j}$, parameters. Because these two parameters function similarly, only the demand change parameter is presented as follows: As previously explained, each node in the scenario tree represents a realization of either a demand or a capacity value. The demand change parameter is the percentage demand change for technology i in period t in scenario s for group g_i. In the AMPL implementation, the demand change parameter is represented by *demand_change(t,s,g)*, indexed across time, t, and realization, s, and the set called *tech_group*. The *tech_group* set is used because, in addition to aggregating the forecast data for the last three years of the planning period, the independent scenario construction applies the demand change and capacity change parameters to groups of technologies as another means of controlling the growth of scenarios. By assigning each technology to a particular group, the consideration of every possible outcome for every technology is avoided. The decision of assigning technologies to particular groups is based on an estimation of the expected adherence to forecasted values. For example, of the twenty-nine technologies, it is likely that the older technologies will conform to their respective forecasts in a similar fashion. Therefore, these older technologies would be assigned to the same group. The same process would be followed until each technology is assigned to a group.

Consider again Constraint set (1):

$$\sum_{j=1}^{F} x_{ijts} \leq (1+\delta^{C}_{tsg_i}) \cdot D_{it} \qquad \forall i,t,s \qquad (1)$$

The demand change parameter affects the right-hand side of each constraint in this set by altering the forecasted demand value for a particular period by a specified percentage. For example, the demand change parameter for a particular technology group may have a value of 15 percent. Thus, the total allocation for any one of the technologies in that particular group must be less than 115 percent of the forecasted demand value. The capacity change parameters in Constraint set (2) function similarly. Figure 2 illustrates a scenario tree that uses three separate technology groups: 1, 2, and 3. The demand change vector representing the possible outcomes from each event node is thus ($\delta^{D}_{stg1}, \delta^{D}_{stg2}, \delta^{D}_{stg3}$). If each element in this vector can be either ±15 percent, the number of outcomes from each node will be 2^3, or 8.

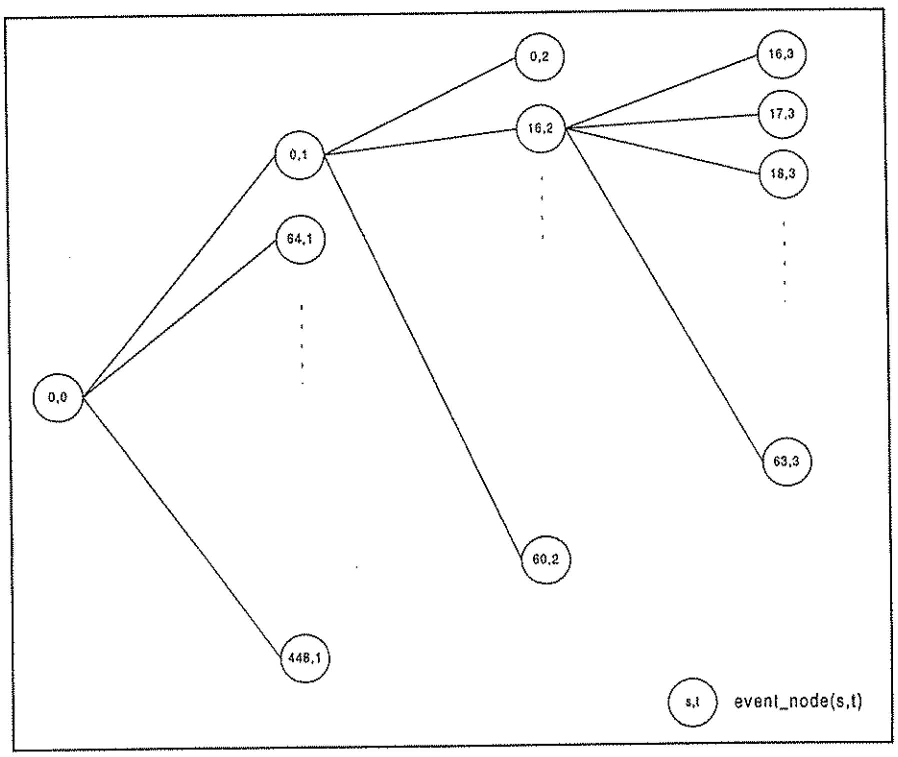

Figure 2: 3-Technology Group Scenario Tree

The distinguishing characteristic of the independent scenario construction is the period-to-period independence of the outcome probabilities. In the case of the 3-technology group construction, each node in a particular time period, *t*, spawns the same set of 8 outcomes – regardless of the path leading up to the particular event node in that time period.

3.2 Arbitrary Scenario Construction

The second type of scenario tree structure is the *arbitrary* scenario construction suggested by Gassmann and Ireland's (1995). We use this scenario construction to represent a particular set of demand/capacity scenarios chosen by the planner for consideration or a sampled subset of scenarios. Each *path* in scenario tree represents a demand/capacity scenario over multiple periods. By this definition, the outcomes considered in each consecutive period are no longer independent, and the scenario tree no longer has a fixed node degree. Figure 3 illustrates an arbitrary scenario tree with a varying number of outcomes for the event nodes in each period.

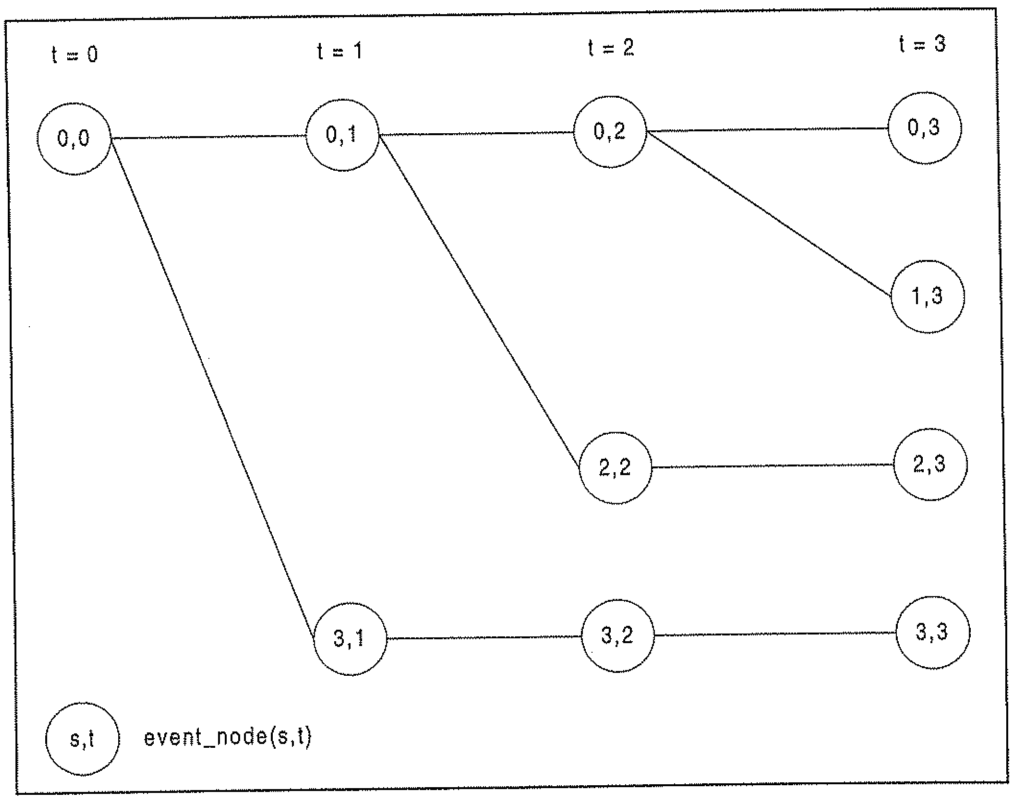

Figure 3: Arbitrary Scenario Tree

In contrast to the independent scenario construction, the *arbitrary* construction does not try to enumerate all possible outcomes of each scenario in each period. Instead, it considers a set of scenarios in each period that are considered to be representative of likely events given what has happened up to this point. In our testing, we implemented an arbitrary scenario tree where the outcomes in each period are a *randomly sampled* set of outcomes of demand or capacity variations. As in the independent scenario construction, values for each demand (capacity) change value are considered. Because this demand change value is applied to an *individual technology* instead of a *group* of technologies, the notation of the demand change parameter is changed from δ^D_{stgi} to δ^D_{sti}, representing the demand change for technology i in time period t in scenario s. Again, take the case where δ^D_{sti} can assume a value of ±15 %. Each constraint in constraint set (1) references a demand change vector, $(\delta^D_{st1}, \delta^D_{st2}, \ldots, \delta^D_{stM})$ that has a dimension of *M=29 technologies*. Each element in the vector assumes a value of either +15 percent or −15 percent, corresponding to the demand change value applied to the respective twenty-nine technologies involved in the capacity planning decision. The number of combinations of outcomes for each period is thus 2^{29}. The method of sampling from this set of outcome combinations first requires the choice of sample size. For the purpose of comparison, we test the arbitrary model using the same number of scenarios used in the independent model. In the 3-technology group case, the independent model uses $(2^3)^3$ or 512 scenarios. This corresponds to the arbitrary scenario construction where eight possible outcomes are sampled in each period.

4. Computational Experiments

The intent of the computational experiments performed with both the independent and arbitrary scenario construction is two-fold. In the case of the independent scenario construction, we intend to determine the number of scenarios required to make a "good" capacity configuration. The experiments with the arbitrary model are performed with the intent of providing a comparison between the independent and arbitrary constructions. It is important to note that in practice, the choice between independent and arbitrary scenario constructions is based on the type of statistical information available and the intention of the decision-maker. If the decision-maker is interested in incorporating random perturbation due to the inaccuracy of demand/capacity data, the independent construction shall be preferred. If, on the other hand, the decision-maker is interested in "what-if" or sensitivity analysis of the effect of various future scenarios, the arbitrary scenario construction is more appropriate.

4.1 Independent Scenario Model Experiments

The essential factor in the independent scenario construction is the number of groups to which technologies are assigned. The purpose of grouping is to stem the growth of scenarios by aggregating the number of possible outcomes from each node. We conduct experiments with the independent model using three levels of aggregation as follows:

1. Three technology groups: the 29 technologies are first assigned to one of 3 technology groups based on the assumptions that the group members will share a common demand characteristic over time. Scenario analysis is then performed on the groups based on demand and capacity variations.
2. Two technology groups: same as above except that only two groups are used.
3. One technology group: all technologies are considered using identical scenarios.

Ten runs for each level were performed. Table 5 details each run.

Table 5: Demand/Capacity Change Outcomes for the Independent Construction

Run Number	Demand Change	Capacity Change
1	Increase or decrease 15% from forecast	No change
2	Increase or decrease 30% from forecast	No change
3	Increase or decrease 45% from forecast	No change
4	No change	Increase or decrease 10% from forecast

5	No change	Increase or decrease 20% from forecast
6	No change	Increase or decrease 30% from forecast
7	Biased +30% from forecast	No change
8	Biased –30% from forecast	No change
9	No change	Biased +20% from forecast
10	No change	Biased –20% from forecast
11	Increase or decrease 15% from forecast	Increase or decrease 10% from forecast
12	Increase or decrease 15% from forecast	Increase or decrease 20% from forecast
13	Increase or decrease 30% from forecast	Increase or decrease 10% from forecast
14	Increase or decrease 30% from forecast	Increase or decrease 20% from forecast

Runs 1 through 6 have two possible outcomes for each technology group. Each outcome is equally likely. For example, in the case of two technology groups for run 1, there are four possible outcomes:

Table 6. Possible Outcomes and Associated Probabilities for Run One in Two-Group Independent Scenario Construction

Outcome One	Outcome Two	Outcome Three	Outcome Four
$\{g_1 = 15\%, g_2 = 15\%\}$	$\{g_1 = 15\%, g_2 = -15\%\}$	$\{g_1 = -15\%, g_2 = 15\%\}$	$\{g_1 = -15\%, g_2 = -15\%\}$
$P_{one} = 0.25$	$p_{two} = 0.25$	$p_{three} = 0.25$	$p_{four} = 0.25$

From each node in the scenario tree, each of these four outcomes has a probability of 0.25 of occurring. Four outcomes from each node over three periods tree in the two-technology group case result in 64 scenarios. The one- and three-technology group each result in 8 and 512 scenarios, respectively. Runs 7 through 10 use biased probabilities to emphasize an over- or under-estimate of either the demand or capacity forecast. Consider run 7 with two technology groups:

Table 7: Possible Outcomes and Associated Probabilities for Run Seven in Two-Group Independent Scenario Construction

Outcome One	Outcome Two	Outcome Three	Outcome Four
$\{g_1 = 30\%, g_2 = 30\%\}$	$\{g_1 = 30\%, g_2 = -30\%\}$	$\{g_1 = -30\%, g_2 = 30\%\}$	$\{g_1 = -30\%, g_2 = -30\%\}$
$P_{one} = 0.7 * 0.7 = 0.49$	$p_{two} = 0.7 * 0.3 = 0.21$	$p_{three} = 0.3 * 0.7 = 0.21$	$p_{four} = 0.3 * 0.3 = 0.09$

The +30 percent coordinate points are weighted with a probability of 0.7, while the -30 percent coordinate points are assigned a probability of 0.3. Thus, outcome one, $\{g_1 = 30\%, g_2 = 30\%\}$, receives a probability of occurrence of 0.49, while outcome three, $\{g_1 = -30\%, g_2 = 30\%\}$, receives a probability of occurrence of 0.21. The outcomes for runs eight, nine, and ten are assigned probabilities in a similar fashion.

Three sets of capacity and demand forecast data were used in the experiments. The first data set uses disguised real data provided by the semiconductor manufacturer. The second and third data sets, referred to henceforth as Replicate 1 and Replicate 2, respectively, are randomly generated based on characteristics of the original data set. The capacity forecast values for Replicate 1 are generated by randomly perturbing each value in the original data set within a range of 50 percent. The demand forecast value perturbs within a range of 40 percent to complete the data set for Replicate 1. The data set for Replicate 2 uses a similar perturbation method; using a range of 20 percent for both the capacity and demand forecast values.

4.2 Arbitrary Scenario Model Experiments

Unlike the independent scenario construction, the arbitrary scenario construction does not use the technology grouping technique. Instead, it considers the technologies individually when applying the demand and capacity change parameters. The outcome sampling technique discussed in Section 3.2 is introduced as a means of choosing only a fraction of the possible outcomes from each node in the scenario tree. To establish comparison against the independent construction, we set the number of possible outcomes from each event node as eight for both the random capacity and random demand cases. This results in 512 scenarios (8^3) in the scenario tree. The same data sets that were used in the independent scenario model experiments are used by the arbitrary scenario model. The design of the experiment runs is similar to that used for the independent scenario model. The runs are summarized in Table 8.

Table 8. Demand/Capacity Change Outcomes for the Arbitrary Construction

Run Number	Demand Change	Capacity Change
1	Increase or decrease 15% from forecast	No change
2	Increase or decrease 30% from forecast	No change
3	Increase or decrease 45% from forecast	No change
4	No change	Increase or decrease 10% from forecast
5	No change	Increase or decrease 20% from forecast
6	No change	Increase or decrease 30% from forecast
7	Biased +30% from forecast	No change
8	Biased –30% from forecast	No change
9	No change	Biased +20% from forecast
10	No change	Biased –20% from forecast
11	Increase or decrease 15% from forecast	Increase or decrease 10% from forecast

12	Increase or decrease 15% from forecast	Increase or decrease 20% from forecast
13	Increase or decrease 30% from forecast	Increase or decrease 10% from forecast
14	Increase or decrease 30% from forecast	Increase or decrease 20% from forecast

As with the independent scenario model experiments, the demand change and capacity change columns refer to the two possible values that can be applied to either the capacity or a technology demand at a particular node. The dependent outcome probabilities for each period used in runs one through six are listed in Table 9. The terms "high" and "low" refer to a positive and negative demand or capacity change, respectively.

Table 9: Outcome Probabilities for Arbitrary Scenario Construction for Runs 1-6

Period	Probability of Outcome	
One	P{ δ^D_{s1m} = high} = 0.5	P{ δ^D_{s1m} = low} = 0.5
Two	P{ δ^D_{s2m} = high | δ^D_{s1m} = high} = 0.35	P{ δ^D_{s2m} = high | δ^D_{s1m} = low} = 0.65
	P{ δ^D_{s2m} = low | δ^D_{s1m} = high} = 0.65	P{ δ^D_{s2m} = low | δ^D_{s1m} = low} = 0.35
Three	P{ δ^D_{s3m} = high | δ^D_{s2m} = high, δ^D_{s1m} = high} = 0.2	P{ δ^D_{s3m} = high | δ^D_{s2m} = low, δ^D_{s1m} = low} = 0.8
	P{ δ^D_{s3m} = low | δ^D_{s2m} = high, δ^D_{s1m} = high} = 0.8	P{ δ^D_{s3m} = low | δ^D_{s2m} = low, δ^D_{s1m} = low} = 0.2
	P{ δ^D_{s3m} = high | δ^D_{s2m} = high, δ^D_{s1m} = low} = 0.5	P{ δ^D_{s3m} = high | δ^D_{s2m} = low, δ^D_{s1m} = high} = 0.5
	P{ δ^D_{s3m} = low | δ^D_{s2m} = high, δ^D_{s1m} = low} = 0.5	P{ δ^D_{s3m} = low | δ^D_{s2m} = low, δ^D_{s1m} = high} = 0.5

For runs 7 through 10, biased probabilities are used to construct the outcome vectors. Table 10 lists the biased probabilities for the positively biased cases in runs 7 and 9:

Table 10: Outcome Probabilities for for Biased Runs 7 -9

Period	Probability of Outcome	
One	P{ δ^D_{s1m} = high} = 0.7	P{ δ^D_{s1m} = low} = 0.3
Two	P{ δ^D_{s2m} = high | δ^D_{s1m} = high} = 0.55	P{ δ^D_{s2m} = high | δ^D_{s1m} = low} = 0.85
	P{ δ^D_{s2m} = low | δ^D_{s1m} = high} = 0.45	P{ δ^D_{s2m} = low | δ^D_{s1m} = low} = 0.15
Three	P{ δ^D_{s3m} = high | δ^D_{s2m} = high, δ^D_{s1m} = high} = 0.4	P{ δ^D_{s3m} = high | δ^D_{s2m} = low, δ^D_{s1m} = low} = 1.0
	P{ δ^D_{s3m} = low | δ^D_{s2m} = high, δ^D_{s1m} = high} = 0.6	P{ δ^D_{s3m} = low | δ^D_{s2m} = low, δ^D_{s1m} = low} = 0.0
	P{ δ^D_{s3m} = high | δ^D_{s2m} = high, δ^D_{s1m} = low} = 0.7	P{ δ^D_{s3m} = high | δ^D_{s2m} = low, δ^D_{s1m} = high} = 0.7

	P{ δ^D_{s3m} = low | δ^D_{s2m} = high, δ^D_{s1m} = low} = 0.3	P{ δ^D_{s3m} = low | δ^D_{s2m} = low, δ^D_{s1m} = high} = 0.3

The computational experiments were performed on a Pentium personal computer. The (3 groups, 3 periods), (2 groups, 3 periods), and (1 group, 3 periods) cases result in approximately 190,000, 26,000, and 4,000 variables, respectively.

4.3 Results from the Independent Scenario Model

Since the number of scenarios grows exponentially when a new outcome is considered at each node, it is important to determine the number of outcomes from each node required for a good capacity configuration decision. Following the design detailed in the previous section, the assessment of the required number of outcomes from each node begins with the establishment of a benchmark objective value for each of the demand and capacity ranges. The eight-outcome per node, 512-scenario case is assigned the role of the benchmark setting for this node-degree factor.

The benchmark results for the original data set are listed in Table 11. The average computation time required for a single run in the 512-scenario independent benchmark set is 544 seconds. The 64-scenario case requires an average of only 5.32 seconds for each run. Table 12 lists the results of the 64-scenario computation runs for the original data.

Table 11: Independent Model Benchmark (3-Technology Groups) Results

Run Number	Demand Change	Capacity Change	Objective Value	CPU Time (seconds)
1	Increase or decrease 15% from forecast	No change	93392.24	599.29
2	Increase or decrease 30% from forecast	No change	89091.281	538.45
3	Increase or decrease 45% from forecast	No change	84386.66	539.05
4	No change	Increase or decrease 10% from forecast	95717.935	539.84
5	No change	Increase or decrease 20% from forecast	94351.893	539.75
6	No change	Increase or decrease 30% from forecast	92169.973	537.76
7	Biased +30% from forecast	No change	95075.694	535.02
8	Biased –30% from forecast	No change	81980.957	537.85
9	No change	Biased +20% from forecast	95604.193	539.70
10	No change	Biased –20% from forecast	92436.235	539.51
11	Increase or decrease 15% from forecast	Increase or decrease 10% from forecast	93467.44	552.45
12	Increase or decrease 15% from forecast	Increase or decrease 20% from forecast	93050.469	539.94

13	Increase or decrease 30% from forecast	Increase or decrease 10% from forecast	89703.503	539.46
14	Increase or decrease 30% from forecast	Increase or decrease 20% from forecast	89703.503	542.49

Table 12: Independent Model Results – 2-Technology Groups Results Original Data Set

Run Number	Demand Change	Capacity Change	Objective Value	% From Benchmark	CPU Time (seconds)
1	Increase or decrease 15% from forecast	No change	93048.989	0.368%	5.30
2	Increase or decrease 30% from forecast	No change	88468.795	0.699%	5.33
3	Increase or decrease 45% from forecast	No change	83487.824	1.065%	5.33
4	No change	Increase or decrease 10% from forecast	95652.494	0.068%	5.32
5	No change	Increase or decrease 20% from forecast	93723.042	0.666%	5.31
6	No change	Increase or decrease 30% from forecast	91634.614	0.581%	5.33
7	Biased +30% from forecast	No change	93955.498	1.178%	5.32
8	Biased –30% from forecast	No change	81642.316	0.413%	5.33
9	No change	Biased +20% from forecast	95148.574	0.477%	5.32
10	No change	Biased –20% from forecast	91737.535	0.756%	5.32
11	Increase or decrease 15% from forecast	Increase or decrease 10% from forecast	93545.616	-0.084%	5.30
12	Increase or decrease 15% from forecast	Increase or decrease 20% from forecast	93471.622	-0.453%	5.30
13	Increase or decrease 30% from forecast	Increase or decrease 10% from forecast	89944.783	-0.269%	5.30
14	Increase or decrease 30% from forecast	Increase or decrease 20% from forecast	90864.182	-1.294%	5.75

When compared to the benchmark values, it would appear that the smaller number of scenarios does not affect the objective value by a significant amount. This may be because, although three technology groups are used, their demand and capacity fluctuations behave in the same way. That is, it is equally likely that group A has a demand increase of 30 percent while groups B and C have the same increase.

The results for the 8-scenario original data set case, listed in Table 13, further illustrates that, when all groups behave in a similar way, a larger number of scenarios can achieve close approximation to the more computationally demanding large scenario cases. The 8-scenario case required an average of less than one second for computation time.

Table 13: Independent Model Results – 1-Technology Group Original Data Set

Run Number	Demand Change	Capacity Change	Objective Value	% From Benchmark	% From 2-Group	CPU Time (seconds)
1	Increase or decrease 15% from forecast	No change	91990.976	1.500%	1.137%	0.86
2	Increase or decrease 30% from forecast	No change	85504.644	4.026%	3.351%	0.86
3	Increase or decrease 45% from forecast	No change	78207.896	7.322%	6.324%	0.87
4	No change	Increase or decrease 10% from forecast	95516.608	0.210%	0.142%	0.87
5	No change	Increase or decrease 20% from forecast	92262.786	2.214%	1.558%	0.86
6	No change	Increase or decrease 30% from forecast	88706.190	3.758%	3.196%	0.88
7	Biased +30% from forecast	No change	90463.806	4.851%	3.716%	0.87
8	Biased –30% from forecast	No change	79644.820	2.850%	2.447%	0.87
9	No change	Biased +20% from forecast	93684.614	2.008%	1.539%	0.86
10	No change	Biased –20% from forecast	90627.576	1.957%	1.210%	0.87
11	Increase or decrease 15% from forecast	Increase or decrease 10% from forecast	92583.434	0.946%	1.029%	0.87
12	Increase or decrease 15% from forecast	Increase or decrease 20% from forecast	92382.874	0.717%	1.165%	0.88
13	Increase or decrease 30% from forecast	Increase or decrease 10% from forecast	87672.864	2.264%	2.526%	0.88
14	Increase or decrease 30% from forecast	Increase or decrease 20% from forecast	88944.045	0.847%	2.113%	0.87

The above comparison uses the objective value as the sole measure for evaluating the differences among the different scenario tree sizes. The results from the computations using the original data set lead to the conclusion that there is not a significant difference among the 3-, 2-, and 1-group cases. The two additional replicates of this same experiment support this conclusion.

To further investigate the effect of scenario aggregation, the experiments to consider the cases where the three technology groups each experience a different pattern of variation as follows:

Group 1 has a biased demand forecast of +30 percent

Group 2 has a biased demand forecast of –30 percent
Group 3 has a demand deviation of ±30 percent

Under the 512-scenario case, each of these different patterns of demand variations can be captured. This is compared to the 8-scenario case in Table 13 where a demand variation of ±30 percent is used.

Table 14 : Comparison of 512-Scenario Case with Mixed Demand Variation Patterns and 8-Scenario Case with One Demand Variation Pattern

Number of Tech Groups	Demand Change	Capacity Change	Objective Value	CPU Time (Seconds)
3	Group 1: Biased +30% from forecast Group 2: Biased –30% from forecast Group 3: Increase or decrease 30% from forecast	No change	88168.486	538.31
1	Increase or decrease 30% from forecast	No change	85504.644	0.86
Percentage Difference in Objective Value			3.021%	

Obviously, when distinctions among groups are more drastic, the difference between the 512-scenario and the aggregate cases will be more significant. Thus, the choice for decision-makers depends on the resolution of the information available for future scenarios and the amount of computer time one can afford.

4.4 Results from the Arbitrary Scenario Model

As detailed earlier in Section 4.1, the computational experiments with the arbitrary scenario model involved the use of scenario trees with 8 outcomes from each node, resulting in 512 scenarios. This set of experiments was run on the same machine used for the independent model experiments. The number of variables was approximately the same as that in the 512-scenario independent model case (190,000).

Each run detailed in the design of experiments was executed three times, with each execution using a different set of scenario samples. Table 16 lists the average objective value for the three executions performed for each run. The average run time for each execution was approximately 590 seconds. When evaluating the average objective value for the three executions, the sum of the run time for the three executions is considered as the sum of each execution's run time. This sum averaged approximately 1770 seconds.

The percentage difference between the objective values for the respective runs in the independent and arbitrary models is quite insignificant. One would expect the respective values to converge as more scenario trees are generated for each of the runs in the arbitrary model.

Table 15 : Arbitrary Model 512-Scenario Results - Original Data Set

Run Number	Demand Change	Capacity Change	Average Objective Value	% Difference from Independent Results	CPU Time
1	Increase or decrease 15% from forecast	No change	93690.798	0.319%	1739.71
2	Increase or decrease 30% from forecast	No change	90828.273	1.912%	1736.91
3	Increase or decrease 45% from forecast	No change	87263.405	3.297%	1744.72
4	No change	Increase or decrease 10% from forecast	95709.445	-0.009%	1744.59
5	No change	Increase or decrease 20% from forecast	94550.501	0.210%	1742.08
6	No change	Increase or decrease 30% from forecast	90671.830	-1.652%	1743.21
7	Biased +30% from forecast	No change	95094.381	0.020%	1928.80
8	Biased –30% from forecast	No change	86924.849	5.688%	1724.81
9	No change	Biased +20% from forecast	95384.911	-0.230%	1731.28
10	No change	Biased –20% from forecast	93347.325	0.976%	1728.39
11	Increase or decrease 15% from forecast	Increase or decrease 10% from forecast	93547.892	0.086%	1772.39
12	Increase or decrease 15% from forecast	Increase or decrease 20% from forecast	92960.118	-0.097%	1774.33
13	Increase or decrease 30% from forecast	Increase or decrease 10% from forecast	90537.183	0.921%	1775.60
14	Increase or decrease 30% from forecast	Increase or decrease 20% from forecast	90101.860	0.442%	1774.04

A more insightful comparison of the two scenario constructions requires the analysis of the actual allocation variables. Of greatest interest are those classifications of allocations that have cost implications. Such measures include the average number of technologies allocated to each facility, the average quantity of product assigned for production at outside foundries, and a comparison of the capacity utilized for each model.

Figure 4 illustrates the average number of technologies allocated to each facility in the first year of allocations for both scenario constructions. With the exception of runs 4 and 13, the arbitrary scenario construction assigns a larger number of technology types to facilities than the independent construction. A desirable feature of a capacity plan is fewer technologies assigned to each facility. This affords easier production planning and higher equipment utilization. Figure 5 illustrates the average allocations to foundries, outside producers contracted by the manufacturer. Run 6 represents the only instance of a significant difference between the two models using this measure.

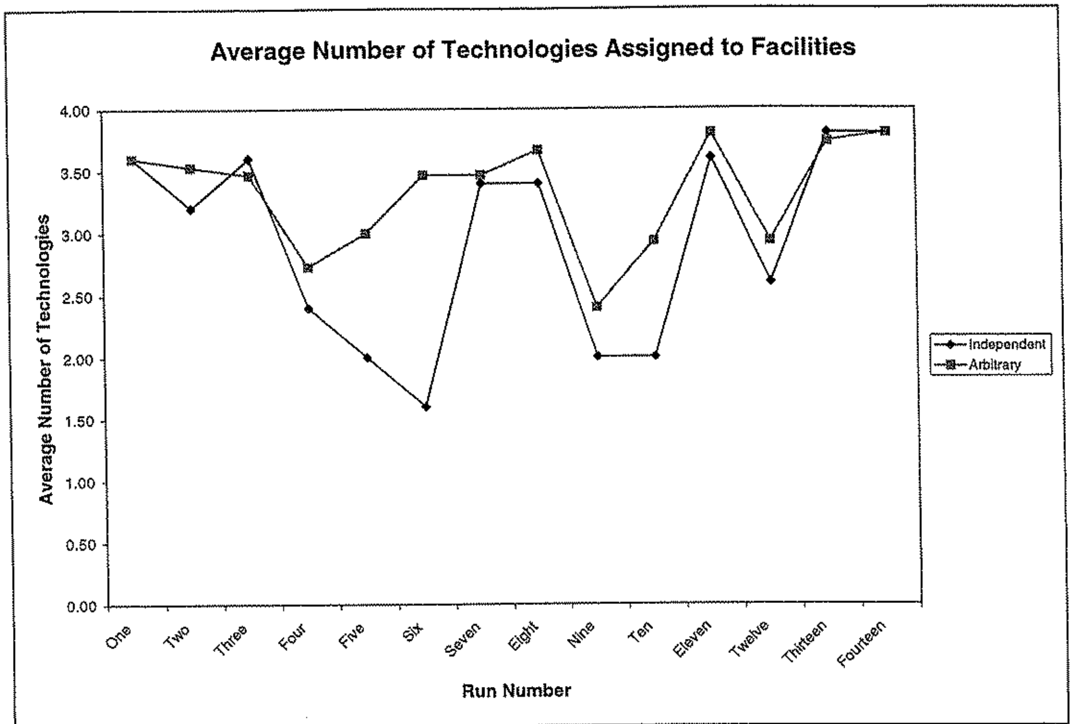

Figure 4: Average Number of Technologies Assigned to Facilities in Period 1

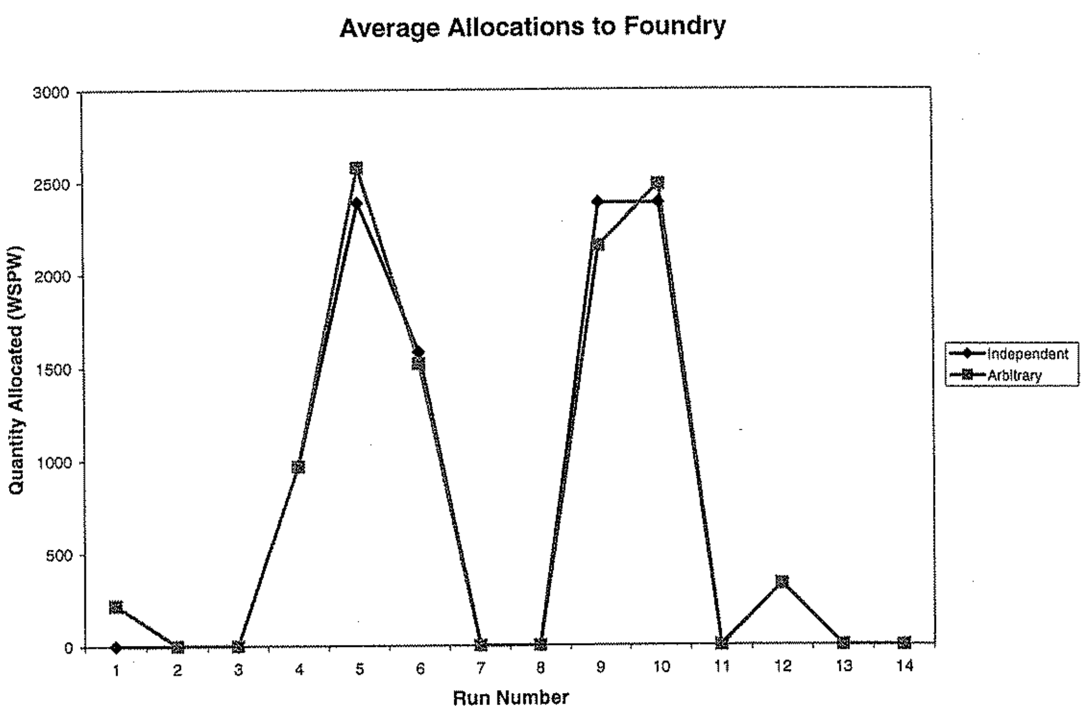

Figure 5: Average Allocations to Foundries in Period 1

Figure 6 details the average allocations to all five facilities in period 2 for both the independent and arbitrary scenario constructions. This measure is similar to the measure of average allocations to foundries in that it determines the extent to which the capacity is being utilized. The average capacity allocations to facilities is very similar between the two models.

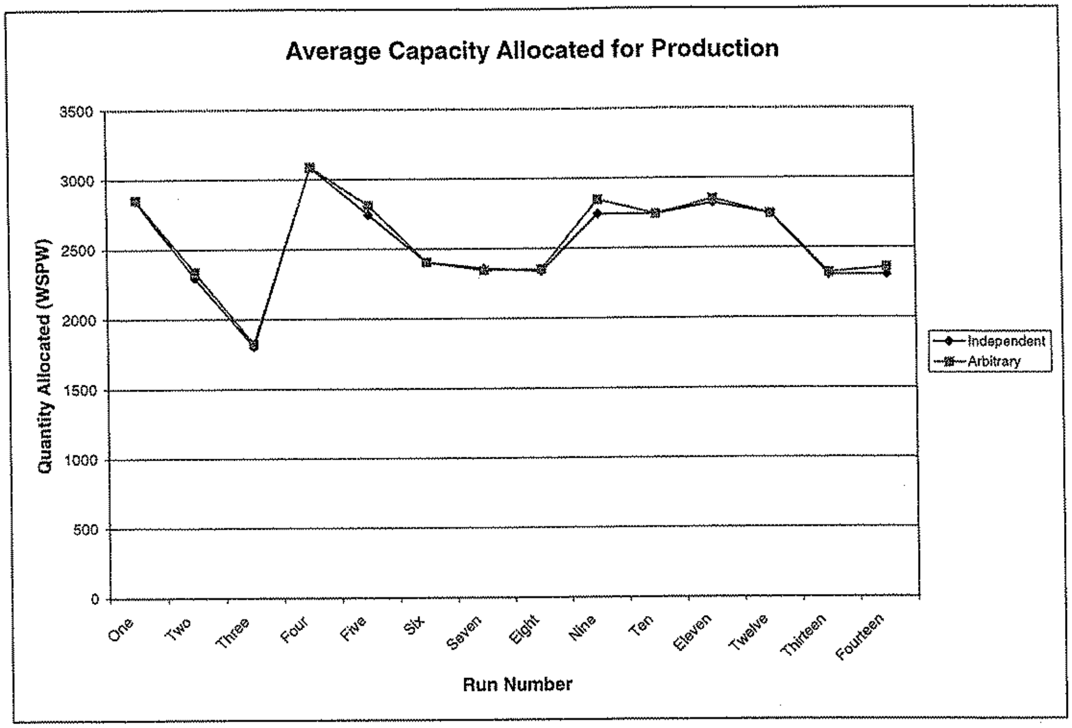

Figure 6: Average Capacity Allocation for Production in Period 2

5. Conclusions

We present an MSLP model for annual capacity planning using two implementations of scenario analysis. The decision variables represent quantities of capacity, measured in wafer starts per week (WSPW), configured for particular technologies across five different fabrication facilities. The two types of scenario constructions have different purposes in the planning decision process. The independent scenario construction has a fixed-node scenario tree with independent probabilities from period to period. This construct is intended for the evaluation of random perturbation due to the inaccuracy of demand/capacity data. The arbitrary scenario construction does not have the fixed-node scenario tree constraint and instead allows for path probabilities which are dependent upon previous outcomes. This model is suited for "what-if" or sensitivity analysis. Our implementation and testing of this model was performed using AMPL/CPLEX. The experiments using the independent model determined that there is not a significant difference in the objective function when fewer total scenarios are used.

The results of our arbitrary scenario implementation were used as a comparison to those of the independent scenario model. In addition to comparing the objective function values for each model, new metrics, such as the average number of different technologies assigned to facilities in Period 2 and the average capacity utilized in Period 2, we used to further contrast the two scenario constructions.

References

Bonser, Jewel S., S. David Wu, Robert H. Storer, "A Multi-Stage Stochastic Programming Model for Fuel Procurement Problems," Technical Report No. 96T-014, Lehigh University, 1996.

Eppen, Gary D., R. Kipp Martin, and Linus Schrage, "A Scenario Approach to Capacity Planning," *Operations Research*, v37, n4, July-August 1989.

Fourer, Robert, David M. Gay, and Brian W. Kernighan, *AMPL: A Modeling Language for Mathematical Programming*, Scientific Press, 1993.

Hillier, Frederick S. and Gerald J. Lieberman, *Introduction to Mathematical Programming*, McGraw-Hill, Inc., 1995.

Gassmann, H.I. and A.M. Ireland, "Scenario Formulation in an Algebraic Modelling Language," *Annals of Operations Research*, v59, 1995.

Parker, R. Gary and Ronald L. Rardin, *Discrete Optimization*, Academic Press, 1988.

Wagner, Harvey M., *Principles of Operations Research*, Prentice-Hall, Inc., 1975.

Winston, Wayne L., *Operations Research: Applications and Algorithms*, Duxbury Press, 1994.