2002 Deliverable Report:
Forecasting methodologies for multidimensional aggregated demands

Task 879.001: Intelligent Demand Aggregation and Forecast Solutions
Project 879: Demand Data Mining and Planning in Semiconductor Manufacturing Networks

Task Leader: Argon Chen
Co-PI's: Ruey-Shan Guo and Shi-Chung Chang
Students: Kenk Chen, Ziv Hsia, and Jakey Lan

1. Abstract and Summary

The demand uncertainty propagated and magnified over the semiconductor demand-supply network is the crucial cause of poor manufacturing/logistic plans. To manage the demand variability, appropriate demand aggregation/forecasting approaches are known to be effective. In the second year of this research task, there are three main research accomplishments: optimum demand aggregation/disaggregation hierarchy, and forecasting by dynamic EWMA demand disaggregation. In the first accomplishment, we have defined and proposed optimum demand planning hierarchy that can greatly improve the quality of demand plans. Finally, the dynamic EWMA demand disaggregation approaches improve the demand forecasts by taking into account the dynamic changes of product mix.

2. Technical Results

2.1 Optimum Demand Planning Hierarchy

An On-Line Analytical Processing (OLAP) tool is useful for analysis of multi-perspective (multi-dimensional) demand aggregation and disaggregation. Demand planners can use the tool to quickly roll up demands to an aggregated level for a total demand or drill down a total demand to detailed demands from different perspectives. For example, a semiconductor demand planner can roll up (or aggregate) the detailed demand to calculate the total demand for logic IC in North America and Europe during the first during the last two quarters of the year. The demand planner can also drill down (or disaggregate) the total demand to see, for example, the proportion of the North American market. OLAP technologies, however, do not provide users the most effective way of analyzing different levels of demand aggregation. We first describe and define different types of demand planning views and demand planning hierarchy. A symbolic representation system to express these views and the demand planning hierarchies is also proposed. Based on the degree of demand fluctuation or the forecast accuracy, the evaluation metrics are defined to evaluate the performance of demand planning hierarchy. Finally, an optimization method based on the dynamic programming approach is also developed to find the optimum demand planning hierarchy. A demand data set from a semiconductor manufacturing company is used to test the algorithms proposed in this research.

2.1.1 Representation of Demand Views and Planning Hierarchies

Product demands can be analyzed from different views such as time, product, customer, etc., and there are usually more than one hierarchical levels or attributes in these views. Take the time view as an example. The highest level could be the year; the quarter could be the next, then the month; and the week could the most detailed time level.

In this research, we develop notations to represent the demand views, different hierarchical levels, and the relationships among these levels and attributes. The view names are in capital letters like TIME, PRODUCT, etc. The hierarchical levels of a view are represented by the subscripts following the view name, for example, $\text{TIME}_{\text{quarter}}$ represents the time view at the quarter level. In addition, "all" is used to denote the highest aggregated level of this view. For example, $\text{PRODUCT}_{\text{all}}$ represents the demand for all products. Also we classified the demand views into the following three categories: view with hierarchical levels, view with attributes and view with mixed attributes.

View with Hierarchical Levels

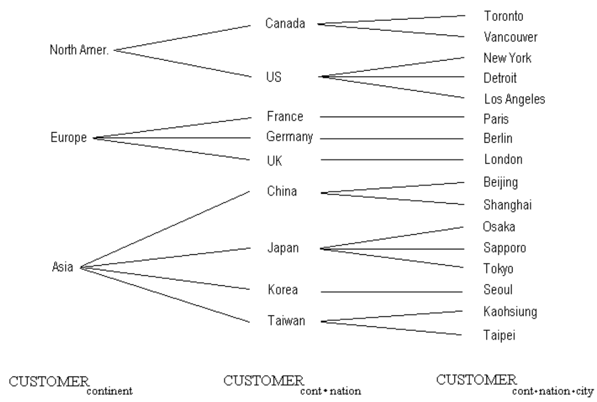

Fig. 1 Hierarchical relationship in the customer view

The view of customer geography can be an example. Enterprises may have the demand data from each individual customer, and these individual customer data can be added up to a higher geographical level like city or nation. And, these higher-level data can be summed to an even higher level such as continent. To describe such a hierarchical relationship, a symbol " • " would be adopted in this research. Take the customer view for an example. $\mathrm{CUSTOMER}_{\mathrm{nation}}$ means that we're now looking at the level "nation" for our demand data, and $\mathrm{CUSTOMER}_{\mathrm{city}}$ means the nation demand is broken down to the city demand. Therefore, the notation $\mathrm{CUSTOMER}_{\mathrm{nation} \bullet \mathrm{city}}$ describes the hierarchical relationship between "nation" and "city". Figure 1 provides a detailed example to illustrat the customer view.

View with Attributes

Besides the views with hierarchical levels, there are also views with attributes that don't have hierarchical relationships. These attributes are independent, and each one can be an independent demand perspective. Take PC memory modules for example. One attribute may be the functional type of the module. It could be SDRAM, DDR, or RDRAM. We can use this functional feature to view the product demands. Another possible attribute may be the memory access rate. We can look at the demand data by different access rates, such as 10ns, 12ns, or 16ns. Also we can use the combination of the access rate and the function type to look at the demands. This kind of views becomes more complex as the product can be characterized by more attributes. We use capital letters to specify the view name, and the subscript is used to denote the attributes as well. What's different is that a symbol " ×" is placed in between two attributes to represent the perspective with a combination of the two independent attributes. In the example above, $\mathrm{PRODUCT}_{\mathrm{type}}$ means that we're considering demand data of different types of memory module, and $\mathrm{PRODUCT}_{\mathrm{type} \times \mathrm{size}}$ means that now the combination of the type and size is a perspective of interest. Also, we use parentheses () behind the name of an attribute to represent its attribute value. $\mathrm{PRODUCT}_{\mathrm{rate}\,(10\mathrm{ns})}$ and $\mathrm{PRODUCT}_{\mathrm{rate}\,(12\mathrm{ns})}$ are the examples. The former points out the demand is for product with "rate = 10ns", and the latter indicate the demand data is for product with "rate = 12ns". Figure 2 shows the example.

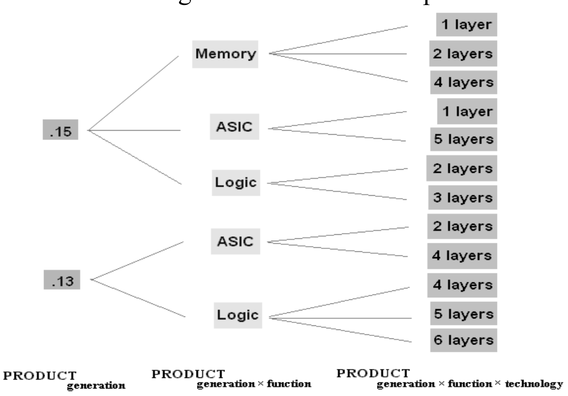

Fig 2 Product view with attributes

View with Mixed Attributes

There may be many other forms of "views of mixed attributes". Here is another example: $\mathrm{PRODUCT}_{(A \cdot B) \times C}$. In this case, the attribute B is hierarchical to the attribute A, and then the combination of B and C is considered. In brief, with the proposed notation system, any hierarchical and/or combinational relationships among attributes can be expressed.

Demand Planning Hierarchy

the demand planning hierarchy is composed of several steps. Each step represents a demand perspective with each demand view specified at a certain aggregation level. Suppose that there are three views to look at the demand data: TIME (with two levels: quarter and month) and CUSTOMER (with two levels: region and city) are hierarchical-level views, while PRODUCT (with two attributes: function and size) is an attribute view. If we want to see the monthly demands of different regions, it will be expressed as $\mathrm{TIME}_{\mathrm{quarter} \cdot \mathrm{month}} \times \mathrm{CUSTOMER}_{\mathrm{region}} \times \mathrm{PRODUCT}_{\mathrm{all}}$. Remember that a demand planning hierarchy is a sequence of demand planning steps that starts from the highest hierarchical level, and ends at the most detailed levels of each view. And, the difference between two steps is only one aggregation level difference in a demand view. We can then express the demand planning hierarchy by a demand planning path (DPP) with all steps listed and connected by "|" symbol. For instance, a possible DPP for the example above can be expressed as follows:

$$\mathrm{TIME}_{\mathrm{quarter}} \times \mathrm{CUSTOMER}_{\mathrm{all}} \times \mathrm{PRODUCT}_{\mathrm{all}}$$
$$|$$
$$\mathrm{TIME}_{\mathrm{quarter}} \times \mathrm{CUSTOMER}_{\mathrm{region}} \times \mathrm{PRODUCT}_{\mathrm{all}}$$
$$|$$
$$\mathrm{TIME}_{\mathrm{quarter}} \times \mathrm{CUSTOMER}_{\mathrm{region} \cdot \mathrm{city}} \times \mathrm{PRODUCT}_{\mathrm{all}}$$
$$|$$
$$\mathrm{TIME}_{\mathrm{quarter} \cdot \mathrm{month}} \times \mathrm{CUSTOMER}_{\mathrm{region} \cdot \mathrm{city}} \times \mathrm{PRODUCT}_{\mathrm{all}}$$
$$|$$
$$\mathrm{TIME}_{\mathrm{quarter} \cdot \mathrm{month}} \times \mathrm{CUSTOMER}_{\mathrm{region} \cdot \mathrm{city}} \times \mathrm{PRODUCT}_{\mathrm{function}}$$
$$|$$
$$\mathrm{TIME}_{\mathrm{quarter} \cdot \mathrm{month}} \times \mathrm{CUSTOMER}_{\mathrm{region} \cdot \mathrm{city}} \times \mathrm{PRODUCT}_{\mathrm{function} \times \mathrm{size}}$$

2.1.2 Evaluation and Optimization of Demand Planning Hierarchy

To evaluate demand planning hierarchy, we use two metrics: weighted-average CV and weighted-average CFE. The weighted-average CV is defined as:

$$CV_1 \cdot \frac{\bar{x}_1}{\sum_{k=1}^{i} \bar{x}_k} + CV_2 \cdot \frac{\bar{x}_2}{\sum_{k=1}^{i} \bar{x}_k} + \cdots + CV_i \cdot \frac{\bar{x}_i}{\sum_{k=1}^{i} \bar{x}_k} \text{ To}$$

where i stands for the number of demand series under an aggregation level, CV_i is the coefficient of variation of each demand series, $\bar{x}_i$ is the average size of the demand series, and $\sum_{k=1}^{i} \bar{x}_k$ is the sum of all sample means. The weighted-average CV is to evaluate the demand variability. For evaluating forecast accuracy, we use $\sqrt{MSE} / \bar{x}$ (referred to as coefficient of forecast error, CFE), where *MSE* stands for the mean square error of the forecast. The weighted-average CFE is then chosen to evaluate the overall forecast performance of each demand planning hierarchy.

Searching for an optimum demand planning path (DPP) is similar to a network's shortest path problem. The different aggregation levels are different nodes in the network. These nodes can be grouped into different stages, and each stage contains nodes having the same degree of aggregation, i.e. having the same total number of hierarchical levels and attributes. The arcs connecting nodes are the feasible paths for the demand planning hierarchy to take. An arc is connecting a node in a stage with a higher degree of aggregation to a node in a stage with a lower degree of aggregation. Each arc has its length. Here, the arc length will be the calculated evaluation metric of the connected node with a lower degree of aggregation. Figure 3 shows the constitution of the Demand Planning Path Dynamic Programming.

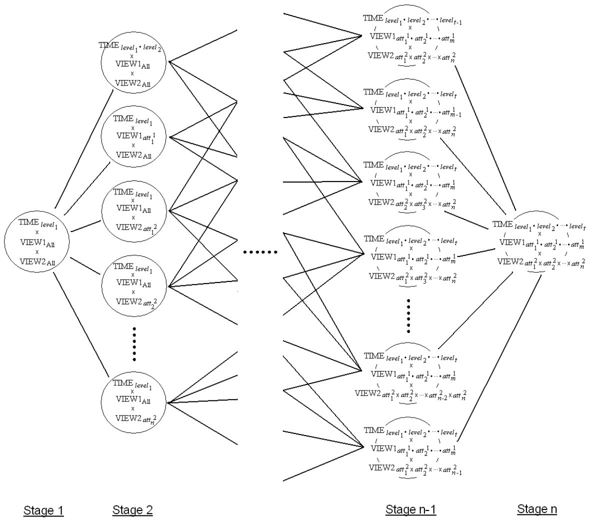

Fig. 3 Dynamic programming for optimum DPP

2.1.3 Case Study: Semiconductor Demand

The demand data set from a semiconductor manufacturing company is used construct the optimum demand planning hierarchy. This data set consists of three views: TIME and CUSTOMER are hierarchical views, and PRODUCT view has mixed attributes. The demand data is across 157 weeks, and there are three levels in the time view: quarter, month and week. Therefore this data set can be also viewed as 37-month or 13-quarter demand data. As to the customer view, there are two levels: GG and GC. GC stands for the individual customer code, and GG is a higher-leveled geographical code. A GC code belongs to only one GG category and a GG code contains 1 to 40 GC codes. In this data set, there are 4 GG codes and 73 GC codes. The product view is more complicated. There are four attributes in this view: T, L, P and PartNum. T stands for "technology", which indicates the type of manufacturing technology. L stands for "number of metal layers", and P stands for "package", which indicates what packaging technology is used to encapsulate this product. These attributes are independent, and each one can be an independent demand perspective. But a fourth attribute "PartNum", which stands for "part number" of the product, is nested to the combination of T, L and P. That is, one PartNum belongs to a specific combination of T, L and P. If PartNum 113 belongs to the combination of T(1), L(2) and P(3), it's impossible to see PartNum 113 appearing under the combination of T(2), L(3), and P(4). This is why we call this view "the view with mixed attributes", because the relationships among T, L and P are independent, while there's another attribute PartNum nested to the combination of the above three. In this data set, there are 21 types of technology, 8 numbers of metal layers, 17 types of packaging technology, and 941 part numbers.

First we applied our dynamic programming algorithm based on the weighted-average CV. The goal is to find a DPP that is the smoothest, i.e., smallest weighted-average CV's throughout the path. The result is shown in Fig. 4.

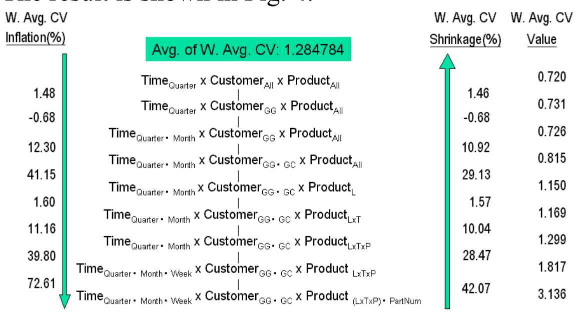

Fig. 4 Least-fluctuation demand planning path

We then use the weighted-average CFE to find the least-forecast-error demand planning path in Figure 5.

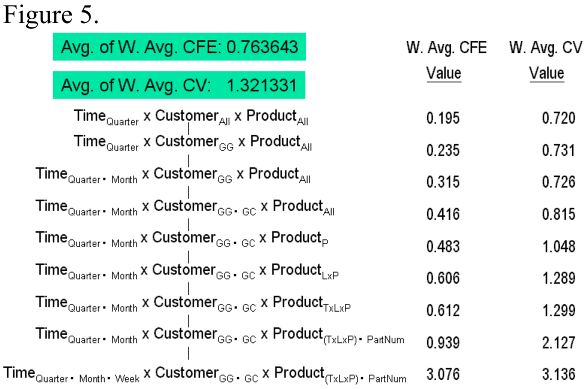

Fig. 5 Least-forecast-error demand planning path

2.2 Dynamic EWMA Demand Disaggregation

The famous Moore's law points out the importance of semiconductor product life cycle (PLC). PLC, therefore, is an important factor affecting demand forecasting. However, demand disaggregation, one of the most important tasks in demand planning, seldom takes PLC effect into consideration. Here, the exponentially weighted moving average (EWMA) statistic is utilized to develop a disaggregation method that captures the PLC effectsptured by dynamically adjusting the smoothing constant of the EWMA statistic.

2.2.1 EWMA Disaggregation Method

Consider a n-week product demand dataset, the estimated proportion of week "n+1" with the EWMA disaggregation formula is defined as:

$$\hat{P}_{i,n+1} = \frac{\sum_{t=1}^{n} w_{i,t} \cdot d_{i,t}}{\sum_{j=1}^{m} \sum_{t=1}^{n} w_{j,t} \cdot d_{j,t}}$$

and

$$\sum_{t=1}^{n} w_{i,t} = \sum_{t=1}^{n} \frac{\alpha_i (1-\alpha_i)^{n-t}}{1-(1-\alpha_i)^n} = 1$$

where

$d_{i,k}$ = Demand of product "i" at time "k"

$w_{i,k}$ = Weight of product "i" at time "k"

n = Number of total historical data

m = Number of total products

α_i = Smoothing constant of product "i"

Taking different PLC phases into considerations, when the PLC is going through a stable phase such as "introduction" or "maturity", trend of the data is steady and the demand variation is mostly due to the demand noise. In this case, a lower α value in the EWMA disaggregation formula is better because the smaller α value puts equal weights on historical data and the effect of noise will be minimized. On the other hand, a bigger α value should be used when the data is going through the "growth" or "decline" phases where the demand mean level changes rapidly. A higher α value puts more weights on the latest historical data and the proportion estimate be mainly determined by the latest observations. The relation between the smoothing constant α and the PLC pattern is shown in Figure 6.

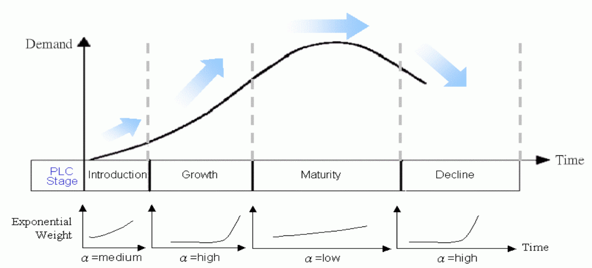

Fig. 6 Best-Fitting α at different PLC Phases

2.2.2 Dynamic Adjusting Algorithm

We first develop a metric to determine PLC phase transition. We refer to the metric as "PLC indicator". We then develop a methodology to adjust the value of α based on the PLC indicator. The PLC indicator proposed in this study is sample auto-correlation (SAC):

$$\hat{\rho}_j = \frac{\frac{1}{n}\sum_{t=1}^{n-j}(x_t - \bar{x})(x_{t+j} - \bar{x})}{\frac{1}{n}\sum_{t=1}^{n}(x_t - \bar{x})^2}$$

where

$\hat{\rho}_j$ = Sample auto-correlation with lag "j"

n = Current time period

x = observations of samples

Figure 7 shows the relationships among α, SAC and PLC phases.

(3.7)

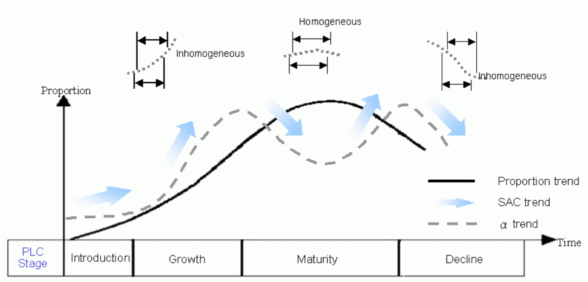

Fig. 7 α, SAC and PLC

Figure 8 shows the procedure of adjusting α based on SAC.

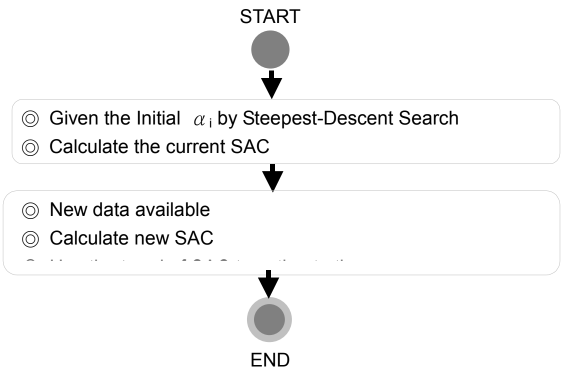

Fig. 8 α adjustment procedure

Experiments show that the dynamic EWMA disaggregation method has improvement up to 90% over the conventional methods.