(34)

Increased Semiconductor Fabrication Capacity Through Cross-Site Processing

by

Christopher L. Cowger

B.S. Electrical Engineering
Duke University, 1994

Submitted to the Sloan School of Management and the
Department of Electrical Engineering and Computer Science
in partial fulfillment of the requirements for the degrees of

Master of Science in *Management* and
Master of Science in *Electrical Engineering and Computer Science*

in conjunction with the
Leaders for Manufacturing Program
at the **Massachusetts Institute of Technology**
May 1998

Signature of Author _____
Department of Electrical Engineering and Computer Science
Sloan School of Management

Certified by _____
Professor Thomas L. Magnanti, Thesis Advisor
Institute Professor

Certified by _____
Associate Professor James E. Chung, Thesis Advisor
Department of Electrical Engineering and Computer Science

Accepted by _____
Arthur C. Smith, Chairman, Committee on
Graduate Students, Department of Electrical Engineering and Computer Science

Accepted by _____
Lawrence S. Abeln, Director
Sloan Master's Program

Increased Semiconductor Fabrication Capacity Through Cross-Site Processing

by

Christopher L. Cowger

Submitted to the Massachusetts Institute of Technology
Sloan School of Management and the Department of
Electrical Engineering and Computer Science
in partial fulfillment of the requirements for the degrees of

Master of Science in Management
and
Master of Science in Electrical Engineering and Computer Science

Abstract

The ongoing boom in demand for semiconductor based products has left the industry's largest manufacturer, Intel Corporation, in the position of having insufficient production capacity for current and future consumer needs. While Intel must and will continue to build new fabrication facilities to keep pace with this ever-increasing product demand, their "Virtual Factory" fabrication environment will likely allow for the increased production capacity of existing facilities through cross-site fabrication of identical process technologies.

This thesis develops a linear program model to optimize cross-site fabrication of identical semiconductor process technologies. Through the cross-site fabrication process flow paths generated by this model, excess idle capacity across several of the "Virtual Factories" studied can be utilized to increase production volume from 0.5% to 10% (depending on pre-existing capacity and tool allocation schemes) without capital additions to current production assets. Several specific examples and a comprehensive case study illustrate the methodology developed and demonstrate its practical applications in a modern semiconductor manufacturing environment. In addition, the thesis addresses many organizational, economic, and technical risks/benefits of cross-site fabrication. It concludes with recommendations for additional model uses and further refinement.

Thesis Advisors:

- Associate Professor James E. Chung, Department of Electrical Engineering and Computer Science.
- Professor Thomas L. Magnanti, Institute Professor.

Acknowledgments

The author gratefully acknowledges the support and resources made available to him through the MIT Leaders for Manufacturing program (LFM), a partnership between the Massachusetts Institute of Technology and several major U.S. manufacturing companies.

I would like to thank the Intel Corporation for sponsoring this work and graciously supporting the LFM program at MIT. In particular, I would like to thank my Intel project supervisor, Dennis Arnow (LFM '93), and company mentor, Don Myers, for the use of their limited time and innumerable project contributions. Their insights were truly invaluable and allowed me to better understand the various technical and organizational issues involved with semiconductor manufacturing at Intel.

In addition, I would like to thank all of the other Intel employees who either helped directly in the development of this thesis or assisted in the dissemination of its key findings. More specifically, I would like to thank Everett Lee, Kent Franklin, Jacob Cherian, Eric Munson, John Chu, Andy Stack, Malcolm Delaney, Hien Nguyen, and Melinda Keyser for their valuable assistance in helping me to better understand the technical side of Intel's semiconductor fabrication processes. I'd like to thank Sean Cunningham and all of the members of the Intel MO/SCS for their continual feedback and assistance in refining the direction of my project while also providing valuable forums for the disclosure of its findings.

Special thanks go to David Haughton of Intel as well for his assistance before, during, and after the internship. His continued dedication stands out in support of the MIT Leaders for Manufacturing Program.

I would also like to extent very special thanks to my thesis advisors at MIT, James Chung, EECS, and Thomas Magnanti, Sloan School of Management and EECS. Without their knowledge and assistance, this thesis would not have been possible.

More informal thanks and a heart-felt coffee toast go out to my LFM classmates and fellow interns Doug Fong, Kevin Farrelly and Pete Lancia. "The Breakfast Club" kept me laughing the whole way through.

Finally, I would like to thank my wonderful parents. I am forever appreciative of the encouragement, support, guidance, and enthusiasm that they have continually given me throughout my athletic, academic, and professional careers.

Table of Contents

List of Equations

List of Figures

List of Tables

1. Introduction.

The semiconductor fabrication industry operates in an environment with tool costs commonly in the millions, new fabrication facility costs in the billions, and product demand growing continuously. In this increasingly capital intensive business setting, industry players need to maximize the use of both existing and future production assets. This thesis presents one approach for increasing the overall wafer starts capacity (WSC) for a collection of collaborative fabrication facilities, or 'fabs', running an identical manufacturing process technology. Specifically, this thesis presents:

1. A linear programming model for determining an optimized cross-facility fabrication path for semiconductor products manufactured across a "virtual factory" of fabrication facilities, a concept discussed in Chapter 3.
2. An analysis of the inherent organizational, economic, and technical risks, as well as the potential benefits, of implementing a cross-facility fabrication program for a virtual factory network of facilities.
3. Additional applications and potential benefits, beyond that of cross-facility fabrication planning, of the model and its associated methodologies.

1.1 Increased Wafer Starts Capacity Through Cross-Site Fabrication.

This thesis develops a model to increase the overall wafer starts capacity (WSC) of a virtual factory network of fabrication facilities through the utilization of cross-facility wafer processing capability. First, a preliminary heuristic model assists in determining which fabrication steps in a given process flow should be considered for the cross-site manufacture of semiconductor products between multiple fabs. Next, a preliminary numerical methodology uses current virtual factory production goals and process equipment capacity metrics to determine the amount and process step location of excess virtual factory production capacity. Finally, a linear program uses the outputs of these two activities to determine the lowest cost/risk cross-facility processing path for a predetermined number of excess wafer lots across the virtual factory network.

This model has the demonstrated capability to increase the cumulative wafer starts capacity of a virtual factory network 0.5% to 10% (depending on the pre-existing capacity and tool allocation scheme) without additional capital investment in equipment and facilities. It does so by simply viewing the collection of component fabs as a true single entity. Specific production wafers are no longer allocated to only one fab's available capacity for every process step. Rather, these wafers can be assigned to *any* virtual factory facility for *any* process step. In this manner, the capacity of the virtual factory increases from the simple sum of its component fab capacities to the capacity of the virtual factory's overall constraint step or steps, which is typically greater by some amount Δ.

EQUATION 1.1.1: VIRTUAL FACTORY CAPACITY WITHOUT CROSS-FACILITY PROCESSING

$$\text{Virtual Factory Capacity} = \Sigma_{\text{All Fabs}} \text{ (Individual Fab Capacity)}$$

EQUATION 1.1.2: VIRTUAL FACTORY CAPACITY WITH CROSS-FACILITY PROCESSING

$$\text{Virtual Factory Capacity} = \text{Virtual Factory Constraint Step(s) Capacity}$$
$$= \Sigma_{\text{All Fabs}} \text{ (Individual Fab Capacity)} + \Delta$$

1.2. Thesis Outline.

This thesis is divided into eight chapters. Chapter 2 provides a brief overview of the semiconductor fabrication process and its related equipment capacity calculations. Chapter 3 discusses the historical and current state of semiconductor fabrication at the industry's virtual factory pioneer, Intel Corporation. Chapter 4 outlines in detail the development and operation of the linear optimization model developed to determine the lowest cost/risk cross-facility processing path for additional wafers added to the virtual factory. Chapter 5 provides a simplified case study, which further details the operation of the model. Chapter 6 presents a critical analysis of key concerns, and risks inherent in the implementation of a cross-facility fabrication plan. Chapter 7 suggests additional uses, beyond that of cross-facility fabrication planning, for the model developed. Finally, Chapter 8 summarizes all key findings.

2. Semiconductor Manufacturing Basics.

This chapter provides an overview of several technical aspects of semiconductor manufacturing relevant to the chapters that follow. The first section differentiates the three major phases of semiconductor manufacturing. Section 2.2 provides brief descriptions of the processes characteristic of the six main functional areas in a typical fabrication process flow. Section 2.3 characterizes fabrication process equipment and introduces the concept of 're-entrant flow' manufacturing. Finally, Section 2.4 explains the calculations used in computing semiconductor equipment capacities.

2.1. Semiconductor Fabrication, Assembly, and Test

Semiconductor manufacturing is characterized by three major phases: fabrication, assembly, and test. The first phase, fabrication, begins with the introduction of bare silicon wafers into an ultra-clean fabrication facility or 'fab.' While in the fab, the wafers are processed through a sequence of steps that produce a replicated series of functioning integrated circuits known as 'die.' Given the small size of these die (typically 0.5" X 0.5") and the typical size of the silicon wafer foundation used today (150 - 200 mm), it is not uncommon for each wafer to contain hundreds of identical die. At the end of this fabrication process, every wafer is subjected to a preliminary electrical test (often referred to as 'e-test'), followed by an individual testing of each die in a test procedure known as 'sort.' The sort process marks dysfunctional die so that they might be more easily separated from the functional die in subsequent steps.

The next phase in the manufacturing process is assembly. The beginning of the assembly process physically separates each die on the wafer from the others. The functional (unmarked) die from the fabricated wafer are then individually packaged in a housing, which will allow for external electrical connections to the die while at the same time physically protecting it from the environment.

The final phase of the manufacturing process is test. During this phase, each individually packaged integrated circuit die (or 'chip') undergoes a series of basic functionality and performance tests to determine if it can be shipped to the final customer.

Given this complete, albeit brief, description of the total process, it is important to note that this thesis focuses solely on the fabrication phase of semiconductor manufacturing.

2.2. The Fabrication Process Flow

Each of the die produced in the fabrication phase of semiconductor manufacturing contain multiple layers, which have been either deposited on or grown into the wafer. Commonly, these layers are referred to as either 'front end' layers or 'back end' layers. The front end layers form the functioning electrical devices (transistors, capacitors, etc.) of the integrated circuit, while the back end layers form the wires creating connections between the different devices.

The fabrication of each front and back end layer requires multiple steps. These complex fabrication steps can be grouped into six main functional areas: photolithography, etch, ion implantation, thin film deposition, diffusion, and planarization. The following discussion provides a brief description of each functional area.

2.2.1. Thin Film Deposition

Thin film deposition steps are those in which material layers are deposited directly on the wafer's surface. The two main deposition processes used in these steps are Chemical Vapor Deposition (CVD), typically used for silicon nitride films, poly-Si films, passivation nitride films, etc., and Physical Vapor Deposition (PVD) or 'sputtering,' typically used for aluminum, aluminum alloys, titanium, gold, and tungsten films. The application of photosensitive resist to the wafer's surface for photolithography patterning purposes can also technically be considered a film deposition, although this process is carried out by a high-speed spin-on process.

2.2.2. Etch

The etch steps remove material layers or selected portions of material layers from the wafer. Semiconductor fabrication uses two main etch processes, 'wet etching' and 'dry etching.' Wet etching involves immersion of the wafer in a liquid reagent (H_2SO_4-H_2O_2, HF-H_2O, etc.) to remove select materials, while dry etching utilizes plasma or reactive ions in the presence of a vacuum to remove materials. Etch processes are typically used to cleanse exposed wafer surfaces of unwanted contaminates before subsequent process steps are performed and to remove material layers left exposed by photolithography process steps.

2.2.3. Ion Implantation

Ion implantation is the fabrication process step that introduces impurity or 'dopant' atoms, in ion form, into selected areas of the silicon wafer substrate. These ions are introduced into the substrate by accelerating them through an electric field, typically ranging in energy from 10-200 keV, and colliding them with selectively exposed regions of the wafer. The dopant ions (Boron, Arsenic, etc.) are introduced into the substrate silicon lattice structure in order to modify its electrical properties in the designated locations. These locally altered electrical characteristics are crucial to the ultimate formation of silicon-substrate-based electronic devices such as transistors and capacitors.

2.2.4. Diffusion

In commercial semiconductor manufacturing, diffusion is the high temperature fabrication process step by which either 1) a material layer is formed through chemical reactions with the wafer or 2) the wafer and its accompanying layers are thermally treated. (Note: Only the thermal annealing steps are true diffusion processes. Formation or 'growth' of material layers on the wafer through high temperature processes in the presence of certain gases is technically a film growth process.) An example of the former would be the formation or 'growth' of silicon dioxide on the silicon wafer substrate in the presence of oxygen in a diffusion furnace. An example of the latter would be in an annealing step which routinely follows an ion implantation. In this annealing step, the

high temperature within the diffusion furnace will diffuse the impurity ions implanted further into the underlying substrate. In addition, the high temperature of the process will induce a recrystallization of any underlying atomic lattice structure damage, which may have been caused by the ion implantation.

2.2.5. Photolithography

Commonly referred to simply as 'litho,' the photolithography steps impose patterns on existing layers of the wafer. After the layer to be patterned has been deposited or formed on the wafer, a photosensitive film called a photoresist is deposited on the wafer in liquid form through a high-speed spin process. An image pattern is then transferred onto the resist by exposing it to ultraviolet light, which has been passed through a patterned reticle or mask. The wafer, with exposed resist intact, is then developed in a solution which has the effect of dissolving either the exposed resist (positive resist) or the unexposed resist (negative resist), depending on the type used. This patterned resist then selectively exposes designated regions of the uppermost layer(s) on the wafer for subsequent ion implantation, etch, or other steps.

2.2.6. Planarization

Chemical mechanical planarization (CMP), or simply 'planarization,' is the fabrication process that achieves universal planarity of the wafer surface before transferring patterns through photolithography or adding additional material layers. This process is the fastest growing process segment in the industry as a result of the increasing need for semiconductor manufacturers to produce devices with feature sizes less than 0.35 microns. The semiconductor industry currently uses several methods for meeting these planarity requirements. They include: Spin-On Photoresist, Spin-On Glass, Boron Phosphosilicate Glass (BPSG) Reflow, and Deposition and Sputter Etchback.

2.3. Fabrication Equipment and Re-entrant Flow Manufacturing

A typical semiconductor fabrication facility contains dozens of different equipment types even though there are only six main functional areas as discussed above. There are three reasons for this.

- Often, the fab contains different technology generations of the same process equipment type. This situation usually arises because of capital expenditure limitations, as the more critically dimensioned process steps can be performed only on the latest generation of process equipment. Other less critical layers can be processed on earlier generation tools.
- The six main functional areas typically contain different equipment types for steps requiring a fundamentally different process. For example, the etch area uses different tool types for liquid reagent 'wet' etching and plasma or ion 'dry' etching.
- Though it is possible for some equipment types within a functional area to run multiple process steps, it is often necessary to dedicate a specific tool to a certain step or number of steps. Reasons for this include: lengthy set-up times between steps, interactive effects between steps, tool qualifications or re-qualifications, and seasoning requirements.

With typically hundreds of process steps being performed by only a few dozen different equipment types, the semiconductor fabrication process is known as 're-entrant flow' manufacturing. In a re-entrant flow manufacturing process, the same equipment sets perform multiple steps at different stages in the flow. As will be demonstrated later in Chapters 4 and 5, the re-entrant nature of the semiconductor fabrication process is a key limiter to excess cross-facility capacity allocation.

2.4. Fabrication Capacity Calculations

As discussed in Section 2.3, the fabrication process requires several unique equipment types, or simply 'tools.' Key to a complete understanding of the optimization tool presented in this thesis is a familiarity with the procedure for calculating the processing

capacity of each tool or tool group. This section will present the background necessary to understand each tool group's aggregate processing capacity, expressed in the industry standard units of wafer starts per week (WSPW), which includes the concepts of tool runrate and tool availability for a single product in multiple fabs.

2.4.1. Process Tool Runrates

Every individual tool has specific runrates, expressed in wafer starts per week, that varies by both process step and by the product being run. Thus, the term runrate is tool, step, and product specific. Given, however, that the analytical tool presented in this thesis deals with only a single product, fabricated in multiple fabs, our analysis will assume the product effect on tool runrate is constant.

Wafers progress from step to step in a fabrication process flow in lots that range in size from 10 to 25 wafers, with 25 being the most common. Depending on the process tool, however, the wafers might be processed individually or in batches of one or more lots. For a tool that processes a single wafer at a time, the runrate is the inverse of the time required to process that wafer.

EQUATION 2.4.1: SINGLE TOOL RUNRATE (SINGLE WAFER PROCESSING)

Runrate (wafers/hr) = (1 wafer/ time to process (in hours))

For a tool that processes in lot size or multiple lot size batches, the tool runrate is simply the number of wafers per batch divided by the processing time for that batch.

EQUATION 2.4.2: SINGLE TOOL RUNRATE (BATCH PROCESSING)

Runrate (wafers/hr) = (Number of wafers in batch/time to process (in hours))

The example below assumes a lot size of 25 wafers, a batch size of three lots, and a processing time of 50 minutes per batch. For this particular process step, then, the resulting runrate is 90 wafers/hour as shown in Equation 2.4.3.

EQUATION 2.4.3: EXAMPLE SINGLE TOOL RUNRATE CALCULATION

Runrate = (75 wafers/50 min.)*(60 min./hour) = 90 wafers/hour

2.4.2. Aggregate Tool Runrates and Total Wafer Starts Capacity

In order to simplify numerical manipulations of fab capacity, the semiconductor industry typically aggregates the runrate data previously calculated into a single number called the wafer starts capacity (WSC). This aggregated WSC, calculated separately for each unique tool, incorporates three important concepts.

- First and foremost, for every unique tool, the aggregate runrate number must account for all of the process steps that are performed on that same tool for a given product.
- Second, for reasons such as preventative maintenance, unplanned downtime, test wafer monitoring, and tool configuration modifications, a specific tool will not be available for processing wafers 100% of the time. The concept of a tool utilization target (U), expressed as a percentage of total tool availability, measures a tool's availability. For example, a utilization target for a specific tool of 85% would account for the capacity limiting factors mentioned above.
- Finally, different products fabricated in the same facility commonly have a varied range of total material layers with variable tool runrates for each. Aggregate tool runrates will thus vary by product as well. The single product input restriction of the analytical tool presented in this thesis will ignore this problem feature.

Given these considerations, the aggregate runrate for a specific tool can be determined by assuming that each of the different process steps run on a single tool occur in immediate succession. Although technically inaccurate given the complex, re-entrant flow nature of the process, this assumption is made in order to simplify the calculation. The aggregate tool runrate, therefore, will be the total number of wafers per batch on that tool divided by the total time required to run all of the process steps performed on the tool back to back.

EQUATION 2.4.4: AGGREGATE TOOL RUNRATE

$$\text{Aggregate Runrate (wafers/hour)} = (\text{wafers per batch})/(\Sigma_{\#\text{ of steps}}\ (\text{time per batch}))$$

The tool's wafer starts capacity, expressed in units of wafer starts per week (WSPW), is then calculated by multiplying the tool's aggregate runrate by a maximum possible tool availability of 168 hours/week times the tool's utilization target.

EQUATION 2.4.5: TOOL WAFER STARTS CAPACITY

$$\text{Wafer Starts Capacity} = (\text{Aggregate Runrate})*(168\ \text{hr/wk})*(\text{Utilization Target})$$

As an illustrative example, consider a product that is processed on a given tool three times throughout a given process flow. As before, the tool processes wafers in batch sizes of three lots (75 wafers). The process times for each of the three different process steps run on the tool are 40 minutes, 50 minutes, and 60 minutes respectively. The aggregate runrate for this particular tool will thus be:

EQUATION 2.4.6: EXAMPLE AGGREGATE RUNRATE CALCULATION

$$\text{Aggr. Runrate} = ((75\ \text{wafers/batch})/(40+50+60\ \text{min/batch}))*60\ \text{min/hour} = 30\ \text{wafers/hr.}$$

A utilization target of 80% (U = 0.80) will therefore result in a WSC of roughly 4000 WSPW.

EQUATION 2.4.7: EXAMPLE WAFER STARTS CAPACITY CALCULATION

$$\text{Wafer Starts Capacity} = (30\ \text{wafers/hr})*(168\ \text{hr/wk})*(0.80) = 4{,}032\ \text{WSPW}$$

This calculation simply means that one unit of this specific tool, processing this specific product, can provide 4,032 new wafer starts of capacity each week if it is operated at a total availability utilization of 80%.

2.5. Equipment Requirements

For every fabrication tool type, a certain number of units are required to meet the desired capacity for the entire process. Based on the example in Equation 2.4.7, to achieve a

desired production goal for the fab of 10,000 wafer starts per week would require three units of that particular tool. Thus for every tool type, the number of units required at a given level of production is calculated by dividing the desired fab capacity by the wafer starts capacity for that specific tool, both expressed in wafer starts per week (WSPW). The resulting number must be rounded up to the nearest integer as fabrication tools are available only in whole units.

EQUATION 2.5.1: NUMBER OF UNITS REQUIRED FOR A SPECIFIC TOOL TYPE

Units Required = ROUND UP {Capacity Desired/Tool WSC}

The integer tool unit requirement of the fabrication process is of crucial importance to the development of this thesis as it is the source of the varying excess capacity in and across the individual fabs that are running identical fabrication processes. For example, our previous situation required three units of a tool capable of 4,032 WSPW for a fab operating at an overall capacity of 10,000 WSPW. The combination of these three tools, however, would also yield an excess, idle capacity of 2,096 WSPW for the sum of all the process steps run on them. This designed-in, idle capacity serves as a key driver for the linear program to be introduced in Chapter 4.

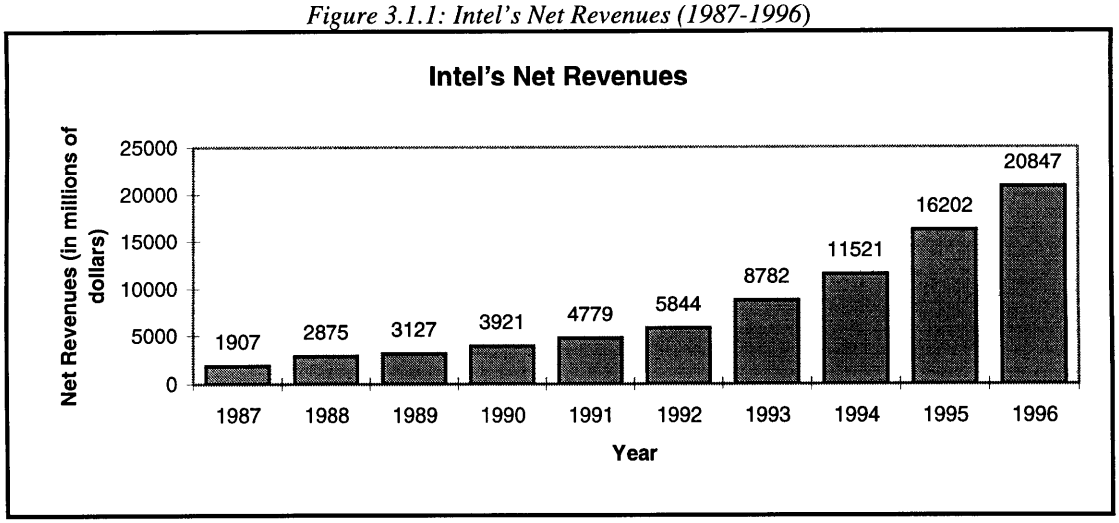
3. Semiconductor Fabrication at Intel.

Before addressing in detail the linear program model developed to determine optimized cross-facility processing paths, it is helpful to provide some background on the industry's virtual factory pioneer, Intel Corporation. This chapter introduces Intel and the current state of wafer fabrication manufacturing at Intel. The first section examines the historical growth in operations at the company. The second section explains the concept and operation of Intel's virtual factory. Finally, the last section addresses the current state of cross-facility processing at Intel.

3.1. Organizational Growth.

Best known for its Pentium™, Pentium Pro™, and Pentium II™ microprocessors, Intel Corporation is the world's largest producer of microprocessors. The company has more than 50,000 employees working in major production sites all over the world. Currently, the company has more than 15 wafer fabrication facilities, with more in planning or under construction, producing microcontrollers, microprocessors, and flash memories. The strong popularity of these products has helped Intel to grow its net revenues significantly over the past 10 years.

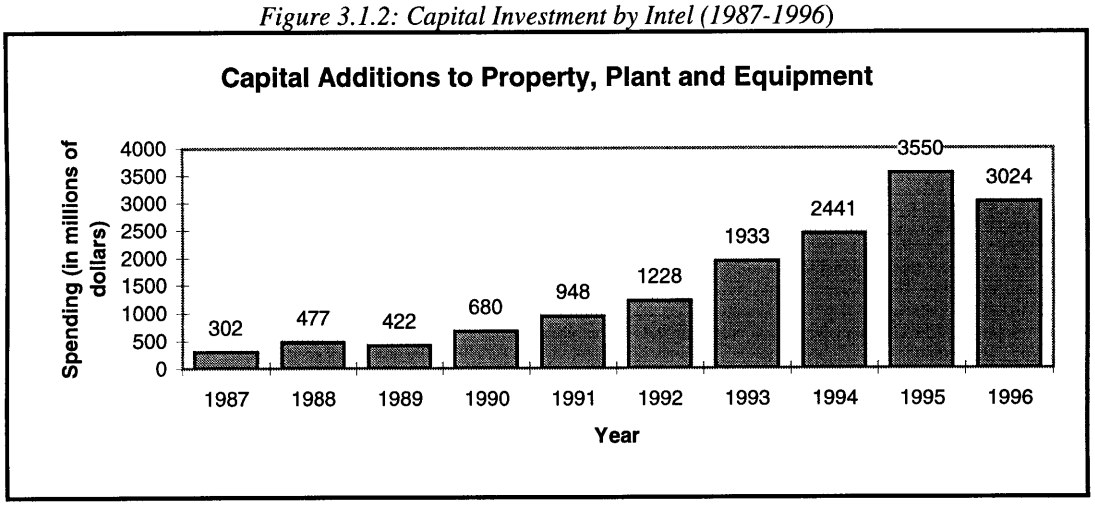
Figure 3.1.1: Intel's Net Revenues (1987-1996)



Spurred by the ever-increasing popularity of personal computers, the ongoing boom in demand for microprocessors has left Intel in the position of having insufficient fabrication capacity for current and future consumer needs. With the penalty for not producing enough product being lost revenues, the fabrication facilities that manufacture these devices are under tremendous pressure to maximize their output capability.

In attempting to keep pace with its expanding markets, Intel has invested heavily in fabrication equipment and facilities in the past ten years. On average, Intel invests between 13% and 22% of its net revenues on increasing production capacity through additions to property, plant and equipment, with one new fab being built or planned every year and a half.

Figure 3.1.2: Capital Investment by Intel (1987-1996)



3.2. Copy Exactly and the Virtual Factory.

The concept of a virtual factory in semiconductor fabrication at Intel refers to an operational collaboration of geographically distinct facilities, utilizing an identical process technology, to fabricate a variety of different products. On many levels, these different facilities will operate as a single entity.

For a given fabrication process technology, the various facilities utilize the same equipment types and process flow in an input-output matching methodology known

throughout the company as 'Copy Exactly!' Often, however, slight modifications need to made to one or several process step recipes at different facility locations to meet local operating conditions and standards. For example, certain processes steps at fabrication facilities located in zones of distinct environmental emission standards and/or at different altitudes might require slight temperature and/or pressure modifications. Process flow, in this case, refers to the actual sequence of fabrication steps from tool to tool in the process, while recipe refers to the exact fabrication parameters used for a particular step, such as time, temperature, pressure, etc. Both recipes and process flow are usually product specific.

3.3. The Current State of Cross-Facility Processing.

Intel currently has the ability to process wafers across geographically disparate fabs at several specific process steps for certain process technology flows. To date, however, this capability is mostly utilized only in case-specific, emergency situations. An example of such a situation would be the shipment of wafers from one fab to another to avoid a particular inoperative or under-performing tool at a particular process step(s). The wafers, in this case, would be shipped to and processed at the second fab for that particular step(s) and then shipped back to the original fab immediately thereafter. Often, certain other process steps, directly before and/or after the key step in question, would be grouped together for cross-facility processing due to technical concerns or work-in-progress (WIP) issues between the two participating facilities. Again, these decisions would be and are still currently being made on a case by case basis.

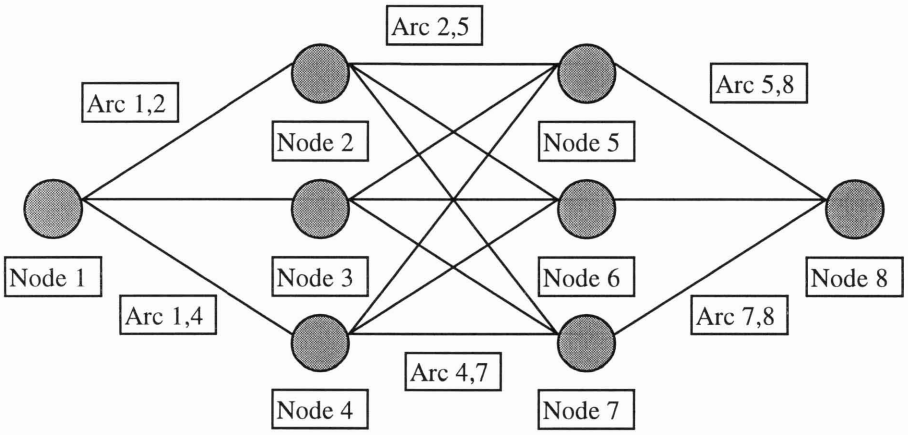
4. The Minimum Cost Network Flow Optimization Model.

We have developed a linear program to complete the complex task of identifying cross-facility processing paths that will permit Intel to introduce additional, over-capacity wafer lots into the virtual factory. The generic type of linear program problem solving methodology used for purposes proposed in this thesis is known as the Minimum Cost Network Flow (MCNF) Model. Given the constraining specifics of our particular model's intent, however, the complete model presented here is more appropriately described as a MCNF model with additional constraints. (See Section 4.3.2)

4.1. Generic MCNF Model Objectives and Operation.

Graphically, any MCNF model can be thought of as a network of nodes and arcs similar to the one in Figure 4.1.1:

Figure 4.1.1: Example MCNF Model Set-Up



Simplistically, the general objective of the model is to transport a given number of flow units from the input node of the network (Node 1 in Figure 4.1.1) to the output node of the network (Node 8) with a minimal cost of transportation through the network, subject to capacity and flow balance constraints. More specifically, to define a generic MCNF model, let:

- X_{ij} = The number of units of flow sent from node i to node j through arc (i,j).
- C_{ij} = The cost of transporting one unit of flow from node i to node j through arc (i,j).
- B_i = The net supply (outflow - inflow) at node i.
- L_{ij} = A lower bound on flow through arc (i,j). (Typically zero)
- U_{ij} = An upper bound on flow through arc (i,j).

The objective function of a MCNF model thus becomes:

EQUATION 4.1.1: MINIMUM COST NETWORK FLOW OBJECTIVE EQUATION

$$\text{Objective} = \text{Minimize} \sum_{\text{all arcs}} C_{ij}X_{ij}$$

Subject to the constraints:

EQUATION 4.1.2: MCNF MODEL FLOW BALANCE CONSTRAINT

$$\sum_j X_{ij} - \sum_k X_{ki} = B_i \quad \text{for all nodes i.}$$

EQUATION 4.1.3: MCNF MODEL ARC CAPACITY CONSTRAINT

$$L_{ij} \leq X_{ij} \leq U_{ij} \quad \text{for all arcs (i,j).}$$

The first generic, or 'flow balance,' constraint stipulates that the net flow of units out of any node i must be equal to some preset node net supply, B_i. The second generic, or 'arc capacity,' constraint ensures that the flow of units through any arc (i,j) does not violate that particular arc's upper and lower capacity limitations.

In some applications, such as the model developed in this thesis, a network flow model also contains capacity constraints imposed upon the throughput of its individual nodes. These constraints are commonly modeled using the concept of 'node splitting.' Node splitting, as the concept's name would indicate, is the simple process that divides one node into two nodes (e.g., node j into nodes j and j') and connects the two with an

intermediate arc (j,j'). Any capacity constraints through the original node (j) are then met by constraining the new intermediate arc (j,j') to be between the appropriate upper and lower bounds.

4.2. Semiconductor Fabrication MCNF Model Development & Objective.

As stated previously in Chapter 2, the integer tool requirement of semiconductor fabrication generates pockets of excess, idle production capacity at various process steps (or combinations of process steps) across the individual fabs in the virtual factory. If at least one of the virtual factory member fabs has excess capacity for every process step, then it is possible, through shipping wafer lots between fabs, to link these pockets together to form a complete fabrication process flow. This section will show how to identify these disparate pockets of excess capacity in the virtual factory and link them together, through an MCNF model, to form a complete fabrication process path, which seeks to minimize the shipping costs and absolute number of wafer lot shipments between separate fabs necessary to complete the flow.

4.2.1. Identification of Cross-Facility Windows as MCNF Model Nodes.

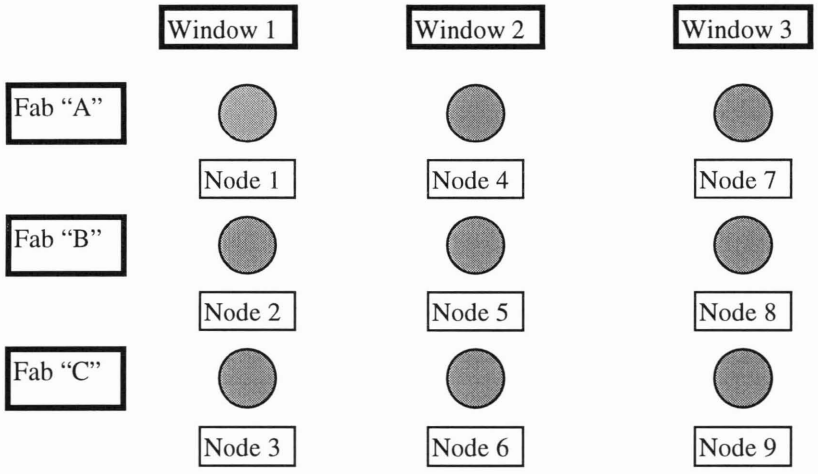
The first step necessary in translating any virtual-factory-based fabrication process into an MCNF model for cross-facility processing is to identify the process steps between which wafer lots should be optimally transported amongst fabs. All process steps between these optimal shipping steps (henceforth known as cross-processing 'windows') would thus be performed only in a single fab.

As previously described, a typical semiconductor fabrication process contains hundreds of complex steps in the six main functional areas. It is not surprising, then, that process experts often disagree as to where these optimal cross-processing windows are in any given process flow. Numerous interviews with experts in each of the functional areas, however, yielded several consensus guidelines for identifying cross-processing windows that would be applicable to any fabrication process.

- All measurement and/or analytical check steps should be performed in the same fabrication facility where the process step or steps being monitored were performed.
- All process steps with critical timing restrictions between them (i.e., steps that must be performed within 24 hours of each other) should be performed in a single fab.
- All process steps performed in immediate succession within a single functional area (i.e., photolithography, thin films, ion implant, etc.) should be conducted in the same fab.
- Wafer lots should not be shipped between fabs with photoresist (either patterned or unpatterned) intact on them.

Once identified for a given process flow, these grouped cross-processing window steps become one of the two defining factors for the nodes of the virtual factory MCNF model. The other defining factor is simply the virtual factory member fab itself. Thus, every node in the model is denoted by both a fab and a process window. For example, a three fab, three window fabrication process would contain the following nodes:

Figure 4.2.1: Example Node Network



4.2.2. Determination of Window/Node Capacities.

Once the nodes for a given virtual factory MCNF model have been established, it is necessary to determine the capacity of each of those nodes in order to ensure that no more wafers are sent to a particular node than it has excess capacity to process. The logic and calculations used to accomplish this task are quite simple. First, we calculate the total wafer starts processing capability for every process step (expressed in WSPW) in each node by multiplying the number of tools available to perform that step in that particular fab by the rated wafer starts capacity of that tool.

EQUATION 4.2.1: STEP CAPACITY WITHIN EACH NODE

Step Capacity = (Number of Tools Available for Process Step)*(WSC for Those Tools)

Next, we subtract the fab's overall WSC from each one of the individual process step capacities to determine the total amount of excess capacity available at each step in the process flow. This number would obviously be zero for each of the fab's constraint steps.

EQUATION 4.2.2: EXCESS STEP CAPACITY WITHIN EACH NODE

Excess Step Capacity = (Step Capacity) - (Overall WSC of the Fab)

Because experts commonly agree that wafers should be shipped between fabs only in whole lots, it is necessary to divide each step's excess capacity by the virtual factory's standard wafer lot size and round down to the nearest lot. Rounding down avoids shipping more wafers to a particular node than can be processed at a given utilization target.

EQUATION 4.2.3: EXCESS STEP CAPACITY IN LOTS

Excess Step Lots = ROUND DOWN{Excess Step Capacity/Lot Size}

Finally, the smallest of the excess step capacities (measured in lots) in each node becomes that particular node's capacity. Again, using the smallest of the step capacities avoids over-shipping or 'choking' any of the process steps in a node.

EQUATION 4.2.4: NODE CAPACITY

Node Capacity = Min [Individual Excess Step Lots in the Node]

These calculated node capacities define the upper limits, U_{ij}, for the arc capacity restrictions of the MCNF model. Namely, the sum of all the arcs (i.e., the sum of all X_{ij}'s) into a node cannot exceed that node's calculated capacity. Again, the individual node's capacity is modeled and constrained in practice by using the concept of 'node splitting,' presented in Section 4.1.

As an example of the entire node capacity calculation process, consider a node (i.e., an optimal group of process steps at a particular fab) that contains three process steps. Any of three tools, each with an individual WSC of 1,200 WSPW, can perform the first step. Either of two tools, each with an individual WSC of 1,750 WSPW, can perform the second step. Any one of four tools, each with an individual WSC of 800 WSPW, can perform the third step in the node. Assume that this node resides in a fab with an overall WSC of 3,000 WSPW, and that wafers are transported in lot sizes of 25. The ensuing node capacity calculations would be as follows:

EQUATION(S) 4.2.5: EXAMPLE NODE STEP CAPACITIES

Step 1 Capacity = 3*1,200 = 3,600 WSPW

Step 2 Capacity = 2*1,750 = 3,500 WSPW

Step 3 Capacity = 4*800 = 3,200 WSPW

EQUATION(S) 4.2.6: EXAMPLE EXCESS STEP CAPACITIES WITHIN THE NODE

Excess Step 1 Capacity = 3,600 - 3,000 = 600 WSPW

Excess Step 2 Capacity = 3,500 - 3,000 = 500 WSPW

Excess Step 3 Capacity = 3,200 - 3,000 = 200 WSPW

EQUATION(S) 4.2.7: EXAMPLE EXCESS STEP CAPACITIES IN LOTS

Excess Step 1 Lots = ROUND DOWN{600/25} = 24 Lots

Excess Step 2 Lots = ROUND DOWN{500/25} = 20 Lots

Excess Step 3 Lots = ROUND DOWN{200/25} = 8 Lots

EQUATION 4.2.8: EXAMPLE NODE CAPACITY

Node Capacity = Min [24,20,8] = 8 Lots

4.2.3. The Objective Function.

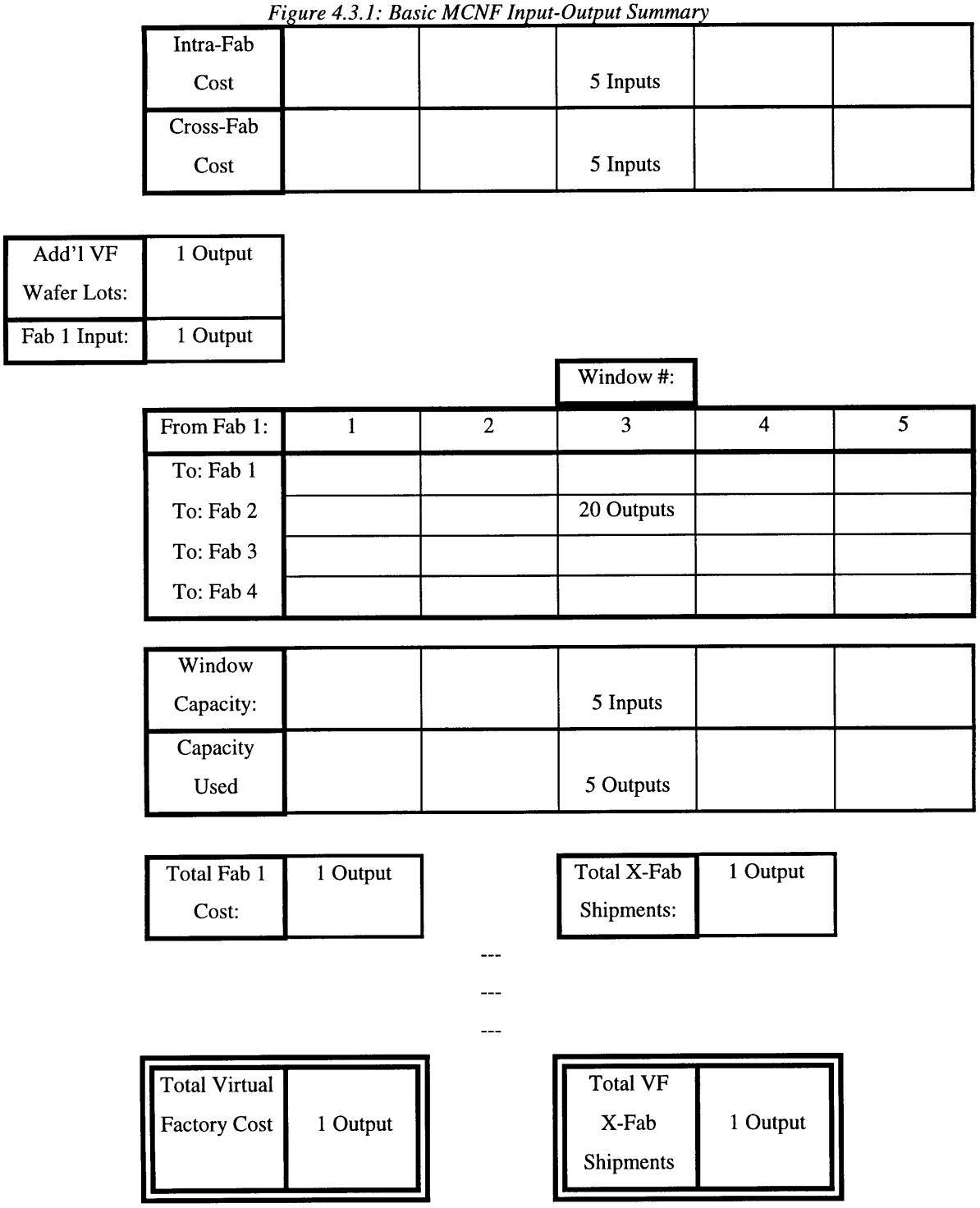
The objective function of the semiconductor fabrication MCNF model is in essence the same as that of the generic MCNF model. In the context of the fabrication process, however, that objective simply measures the number of additional wafer lots (i.e., additional lots in excess of the standard, overall virtual factory WSC) to pass from a designated input node, through every process window (i.e., every *column* of nodes in Figures 4.1.1 and 4.2.1) of the virtual factory node network, and end up at a designated output node with as little shipping between the individual fabs (i.e., the *rows* of nodes in Figures 4.1.1 and 4.2.1) as possible. Thus, what this particular MCNF model is optimizing is not the flow of *all* wafer starts through the virtual factory, but only those *additional* wafer starts intended to utilize the fabs' excess, idle capacity.

4.3. Semiconductor Fabrication MCNF Model Operation.

The model developed to achieve the objective stated in Section 4.2 is a linear program which we solve utilizing Frontline Systems, Inc's Large-Scale LP Solver for Microsoft Excel. More efficient, specialized algorithms are available for solving minimum cost network flow problems. However, we chose to use an Excel-based linear programming model and solver because our problems are relatively small by network flow standards (i.e., 1500 - 2500 decision variables and 5000 - 6000 constraints for a typical virtual factory scenario) and because Frontline's add-on Excel solver (Excel's standard solver cannot solve for more than 200 variables) was readily available. Figure 4.3.1 displays the basic single fab input-output summary format that would be replicated in Excel to solve

the semiconductor fabrication MCNF model for a four fab, five window process flow. (Note: The figure below would be repeated 4 times, one for each fab, in a four fab virtual factory.)

Figure 4.3.1: Basic MCNF Input-Output Summary



4.3.1. Model Inputs.

The model has two primary inputs: The cost of shipping wafers between fabrication sites for each of the process windows and the node capacity for every process window at each individual fab.

Input 1: Cross/Intra-Facility Cost Weightings for Shipping Wafer Lots.

Since a key component of this model is to find cross-facility processing paths with a minimal number of wafer lot shipments between fabs, a crucial input to the model is the cost weighting factor, C_{ij}, assigned to each wafer lot shipped from one node to another. Assuming that wafer lots transported from one window to a same-fab node in the next window would not incur any additional real cost or risk above that which is normally seen in any individual fab, it is logical to assign all same-fab C_{ij}'s a value of zero. We would typically, however, assign cross-fab C_{ij}'s nonzero values based on factors such as distance between the two fabs in question and/or any special monetary or technical risks associated with cross-facility shipments between two particular process windows. For example, we might assign cross-facility wafer lot shipments between fabs that are less than 1000 miles apart for low technical risk process windows a cost weighting of one, while assigning cross-facility wafer shipments between fabs that are more than 1000 miles apart for high technical risk process windows a cost weighting of five. The fine tuning capability of possible cost weighting scales is nearly limitless.

Input 2: Node Capacities.

The final primary model input is each virtual factory network node's processing capacity (i.e., the processing capability of each fab for every process window). As stated previously in Section 4.2.2, these node capacities are essential for restricting the model from shipping more wafer lots to a particular node than it can handle.

4.3.2. Model Constraints.

The results of any optimization study run on this model are based on four primary constraints: flow balance, node capacity, tool capacity, and nonnegative integer flows.

The first two constraints are common to all MCNF models, as indicated previously. The last two constraints, however, are specific to this particular model.

Constraint 1: Flow Balance.

As described earlier in Section 4.1, the concept of flow balance in any MCNF model is described by: $\sum_j X_{ij} - \sum_k X_{ki} = B_i$ for every node in the network. In this equation, B_i is the net supply (i.e., outflow minus inflow) at node i. Since every wafer lot that enters the node network must pass all the way through it, B_i must always be set equal to zero. In Excel, we model this zero net supply at every node by setting the sum of outputs from all virtual factory fabs going to one particular fab at a given process window equal to the output of that particular fab at the next process window (i.e., the 'Capacity Used' cell for that fab). For example, in the four fab, five window MCNF model described earlier, a zero net supply of wafers at a given node, defined by shipments from Window 2 in Fabs 1-4 into Window 3 at Fab 3, would be ensured by setting the sum of cells marked below in Figure 4.3.2 equal to the single cell marked in Figure 4.3.3. (i.e., Cell 1 + Cell 2 + Cell 3 + Cell 4 = Cell 5)

Figure 4.3.2: Example Input-Output Summary for Lots Shipped at Window 2

			Window #:		
Fab 1	1	2	3	4	5
To: Fab 1					
To: Fab 2					
To: Fab 3		Cell 1			
To: Fab 4					
Window Capacity:					
Capacity Used					

Fab 2	1	2	3	4	5
To: Fab 2					
To: Fab 1					
To: Fab 3		Cell 2			
To: Fab 4					
Window Capacity:					
Capacity Used					

Fab 3	1	2	3	4	5
To: Fab 3					
To: Fab 1					
To: Fab 2		Cell 3			
To: Fab 4					
Window Capacity:					
Capacity Used					

Fab 4	1	2	3	4	5
To: Fab 4					
To: Fab 1					
To: Fab 2		Cell 4			
To: Fab 3					
Window Capacity:					
Capacity Used					

Figure 4.3.3: Example Input-Output Summary for Lots Shipped at Window 3

			Window #:		
From Fab 3:	1	2	3	4	5
To: Fab 1					
To: Fab 2					
To: Fab 3					
To: Fab 4					
Window Capacity:					
Capacity Used			Cell 5		

Constraint 2: Node Capacity.

As stated before in Sections 4.2.2 and 4.3.1, in this model, no more wafer lots should be shipped to a specific node than there is idle capacity there to process them. To ensure this constraint, we restrict the value of each fab's 'Capacity Used' cell for each window to be less than or equal to its corresponding 'Window Capacity' cell directly above it in Frontline's Large-Scale LP Solver set-up routine.

Figure 4.3.4: Example Node Capacity Constraint Set-up

Window Capacity:	Cell 1	Cell 2	Cell 3	Cell 4	Cell 5
Capacity Used	Cell 6	Cell 7	Cell 8	Cell 9	Cell 10

In this single fab example with five process windows, the model would be constrained so that (Cell X) ≤ (Cell X-5).

Constraint 3: Tool Capacity.

The node capacity constraint detailed above ensures that no more wafer lots are sent to a node than can be processed by the tool with the least excess capacity in the window and fab that defines it. Additionally, the node capacity constraint ensures an equal flow of added wafer lots into and out of each node. Given the previously discussed re-entrant flow nature of the fabrication process, however, it is also necessary to monitor and limit the cumulative amount of wafer lots processed on every individual tool in each individual fab under steady state conditions, a constraint not provided for in a standard MCNF model. This constraint is different from the simple node capacity constraint in that it limits the usage of individual tool groups *across multiple nodes*. For example, consider two nodes: X and Y. Both nodes contain processes that are run on tools 1, 2, and 3, with the constraint step being processed on tool 2. The node capacity constraint will restrict the flow of wafers through X and Y to be less than the excess capacity available on tool 2. That is:

EQUATION 4.3.1: EXAMPLE NODE CAPACITY CONSTRAINTS

$$\text{(Flow through Node X)} \leq \text{(Excess Tool 2 Capacity)}$$
$$\text{(Flow through Node Y)} \leq \text{(Excess Tool 2 Capacity)}$$

Conversely, the tool capacity constraint will restrict the *sum* of the flows through X and Y to be less than the excess capacity available for each individual tool. That is:

EQUATION 4.3.2: EXAMPLE TOOL CAPACITY CONSTRAINTS

$$\text{(Flow through Node X)} + \text{(Flow through Node Y)} \leq \text{(Excess Tool 1 Capacity)}$$
$$\text{(Flow through Node X)} + \text{(Flow through Node Y)} \leq \text{(Excess Tool 2 Capacity)}$$
$$\text{(Flow through Node X)} + \text{(Flow through Node Y)} \leq \text{(Excess Tool 3 Capacity)}$$

This additional constraint to the standard MCNF model can be accomplished in Excel by creating another input-output summary table similar to the one below for each fab in the virtual factory.

Table 4.3.1: Example Single-Fab Tool Constraint Table

Fab 1 Tool:	Excess Capacity Available:	Excess Capacity Used:
Tool "A"	Input Cell 1	Output Cell 1
Tool "B"	Input Cell 2	Output Cell 2
Tool "C"	Input Cell 3	Output Cell 3
Tool "D"	Input Cell 4	Output Cell 4
Tool "E"	Input Cell 5	Output Cell 5

The way in which this table operates is simple. Each time a particular process window in the given fab is utilized for processing wafers, every tool in that window is incremented in its respective output cell (e.g., Output Cells 1-5 in the example above) by the number of wafer lots that were processed. To meet tool capacity limitations, then, the model's user needs simply to input each tool's excess capacity available in its respective 'Excess Capacity Available' input cell and then constrain its corresponding 'Excess Capacity Used' output cell to be less than or equal to it in Frontline's Large-Scale LP Solver set-up routine (i.e., in Solver, constrain (Output Cells 1-5) $\leq$ (Input Cells 1-5)).

Constraint 4: Nonnegative, Integer Flows.

A final constraint that is not typical of a standard MCNF model, but is necessary for this particular model, is the nonnegative, integer flow of wafer lots from node to node. The reasons for this limitation should be intuitive. First, nonnegativity is required simply because it is impossible, and hence useless to consider, transportation of negative wafer lots within and between fabs. Next, given the previously mentioned assertion that industry experts would consider shipping only whole lots of wafers, it is also useless to consider fractional lots, which might also at times even imply fractional wafers.

This combination of constraints can be easily achieved in Frontline's Large-Scale LP Solver set-up routine by restricting all of the cells in the basic input-output summary table that indicate the volume of wafers transported over every arc in the network (e.g., the large grouping of twenty output cells in Figure 4.3.1) to be both nonnegative and integer.

4.3.3. Model Outputs.

This model has six primary outputs: additional virtual factory wafer lots, optimal cross-facility processing paths, quantitative node utilization data, excess tool capacity utilization data, total virtual factory cost, and total number of virtual factory cross-site shipments.

Outputs 1 and 2: Additional Virtual Factory Wafer Lots and their Optimized Cross-Site Processing Path(s).

Because we want to maximize the number of extra wafer lots processed across the virtual factory *and* simultaneously minimize the amount of shipping in their corresponding optimized process path(s), this model presents what is known as a 'Multi-object Optimization Problem.' There are two methods for solving a problem of this nature. The first method, known as the 'Weighted Function' approach, involves solving a single objective function in which the multiple objectives have each been assigned weighting factors that indicate their level of importance in determining the overall solution to the problem. For example, to solve the multi-object optimization problem presented in this thesis, we would solve an objective function of the form:

EQUATION 4.3.3: WEIGHTED FUNCTION APPROACH OBJECTIVE EQUATION

$$\text{Objective} = \text{Maximize}[\theta * \text{Additional Wafer Lots} - \phi * \text{Total Shipping Costs}]$$

where θ and ϕ are the variable importance weighting factors. If the total number of additional wafer lots is the more important objective for a given scenario, then the value of θ will be set higher than the value of ϕ. If the total shipping cost is of paramount importance, however, then ϕ will be assigned a value higher than θ.

The other method used in solving multi-object optimization problems like this involves optimizing only one of the objectives while allowing the model to vary the other objective within constrained limits. For example, to solve the multi-object optimization problem presented in this thesis, we would either 1) Minimize the total cost of cross-site processing while allowing the model to vary the number of additional wafers starts added

(subject to a lower-bound constraint) or 2) Maximize the total number of additional wafer starts while allowing the model to vary the total cost of cross-site processing (subject to an upper-bound constraint).

The basic tenets of the Theory of Constraints (TOC) would suggest that the maximum number of additional wafer lots capable of being processed would be equal to lowest cumulative process window capacity across the virtual factory. However, given that these tools are utilized for several process windows simultaneously in a steady state condition (i.e., re-entrant flow), the actual number of lots that can be feasibly processed is often much less.

Regardless of the multi-objective optimization problem solving method used, the various process path(s) by which the additional wafer starts are fabricated are often equally important as the total number of additional lots processed. This output can be seen via the numerous output cells that indicate the volume of wafer lots transported over the various network arcs in the Excel model input-output summary (e.g., again, the large grouping of 20 output cells in Figure 4.3.1).

Output 3: Quantitative Node Utilization Data.

For reasons that will be discussed later, it is also often useful to know the utilization level of the various nodes in the network being modeled. This information is contained in the 'Capacity Used' output cells for each fab and process window. It is also possible to present this output information graphically in Excel, which is often more useful in recognizing cross-facility processing patterns, by selecting all of the 'Capacity Used' output cells for a particular fab and processing them through Excel's Chart Wizard.

Output 4: Excess Tool Capacity Utilization Data.

Again, for reasons that will be discussed later, it is also often useful to know the utilization level of excess tool capacity at each fab. This information is contained in the

'Excess Capacity Used' output cells of the Single-Fab Tool Constraint Tables (e.g. Figure 4.3.5).

Outputs 5 and 6: Total Virtual Factory Cost and Number of Cross-Site Shipments.
The MCNF model objective function being minimized specifies the total virtual factory cost of cross-facility processing a given number of wafer lots. This output appears in the basic input-output summary's 'Total Virtual Factory Cost' output cell for any given cost weighting scale.

Additionally, the basic input-output summary of every virtual factory scenario solution specifies the total number of cross-facility shipments required to process a given number of additional wafer lots. Excel calculates this output by incrementing the 'Total VF X-Fab Shipments' output cell by one every time a cross-site path output cell becomes nonzero. The model includes this particular metric as a standard output is because numerous interviews with industry experts revealed that most feel it is a crucial component of any go/no-go decision making process for a project such as this.

4.4. Important Considerations Regarding the Operation of This Model.

When using this particular model for the purpose of planning cross-facility processing routes for a network of semiconductor fabrication facilities, it is important to keep two considerations in mind:

Consideration 1: This model does not give "the" solution.
The model presented in this thesis provides the user with an optimal process path solution that best serves as a *baseline* result for a particular number of additional virtual factory wafer starts, a given virtual factory set-up, and a specific cost weighting scale. In most cases, however, there is an appreciable degree of flexibility and/or ambiguity in one or all of these metrics. Therefore, since multiple optimal solutions are possible through manipulations of uncertain inputs, solution refinement becomes a crucial component of the model's use. Optimal solutions should always be tailored as much as possible to the

specific needs of individual fabs with constant comparisons made to each individual input metric's effect on total virtual factory output.

Consideration 2: This model does not factor in the effects of current work-in-progress (WIP) or overall throughput times.

Due to the relatively long cycle times often required for the fabrication of semiconductor products, the number of wafers started at any given time in any given facility will have a significant effect on the available tool capacities for that fab several months into the future. Therefore, before any decision to add additional wafer starts to multiple fabs could be made, the company would have to carefully consider existing WIP levels in those fabs.

5. Example Case Study.

This chapter presents a case study using a simplified virtual factory scenario consisting of three production facilities fabricating a single, identical product using six different equipment types on a ten-step process flow. The purpose of this simplified study is to make the reader more comfortable with the concepts introduced thus far by way of a single, yet comprehensive example.

5.1. Overview.

Consider a semiconductor manufacturing company with three geographically disparate fabrication facilities capable of producing a single, identical product. The facilities use identical equipment types and an identical process technology to produce the product. The only differences between the facilities of this virtual factory network are minor process recipe differences, necessitated by differed local operating conditions, and varied overall wafer starts capacity (WSC) of each fab, intended for company-wide manufacturing strategy reasons.

Section 5.2 presents the initial situation of this virtual factory with detailed information concerning the overall wafer starts capacity, installed equipment set, and amount and location of excess capacity at each facility. In addition, this section will identify and quantify the capacity of each MCNF model representation node, which collectively represent this particular network of facilities. Section 5.3 presents a solution to the given initial conditions, using a general form of the optimization model discussed in Chapter 4, and Section 5.4 explains the results in detail. The purpose of this last section is to develop a more intuitive understanding of the model and the potential implications of its results.

5.2. Initial Conditions.

For this case, we will consider three fabrication facilities, denoted simply as Fab 1, Fab 2, and Fab 3. Section 5.2.1 presents the wafer starts capacity (WSC) data for each facility overall and for each tool type within them. Section 5.2.2 then calculates the necessary

installed equipment base for each facility. In this example, the equipment sets at each facility are identical with the exception of the number of installed units as the facilities differ in size and overall capacity.

5.2.1. Wafer Starts Capacity Data.

Each tool utilized in this particular process flow has a WSC metric that represents the capacity available from a single unit of that equipment type per unit of time. Normally, this metric is product specific, but because we will be considering only a single product, we will require only a single WSC per tool. Additionally, all WSC metrics given or derived in this example will be assumed to be in units of wafer starts per week (WSPW). Table 5.2.1 lists the desired overall wafer starts capacity of each virtual factory component facility. Table 5.2.2 lists the rated WSC of each equipment type within each component facility.

Table 5.2.1: Desired Facility WSC Data (in WSPW)

Fabrication Facility:	Desired Wafer Starts Capacity:
Fab 1	5000
Fab 2	7500
Fab 3	3000

Table 5.2.2: Tool WSC Data (in WSPW)

Equipment Type:	Rated WSC:
Tool #1	350
Tool #2	5000
Tool #3	1500
Tool #4	400
Tool #5	2500
Tool #6	700

5.2.2. Installed Equipment Sets.

Before optimizing this situation, we need to determine what the installed equipment set at each facility will be, based on the production targets of Table 5.1. We determine the results, calculated using Equation 2.5.1, by dividing the desired production capacity of a

given fab by the respective WSC metric for each tool and rounding up to the nearest integer. Tables 5.2.3-5.2.5 summarize the results for each fab and tool type.

Table 5.2.3: Installed Equipment Set (Fab 1)

Equipment Type:	Desired Fab Capacity:	Fractional Tools Required	Integer Tools Installed:
Tool #1	5000 WSPW	5000/350 = 14.3	15
Tool #2	5000 WSPW	5000/5000 = 1	1
Tool #3	5000 WSPW	5000/1500 = 3.3	4
Tool #4	5000 WSPW	5000/400 = 12.5	13
Tool #5	5000 WSPW	5000/2500 = 2	2
Tool #6	5000 WSPW	5000/700 = 7.1	8

Table 5.2.4: Installed Equipment Set (Fab 2)

Equipment Type:	Desired Fab Capacity:	Fractional Tools Required	Integer Tools Installed:
Tool #1	7500 WSPW	7500/350 = 21.4	22
Tool #2	7500 WSPW	7500/5000 = 1.5	2
Tool #3	7500 WSPW	7500/1500 = 5	5
Tool #4	7500 WSPW	7500/400 = 18.8	19
Tool #5	7500 WSPW	7500/2500 = 3	3
Tool #6	7500 WSPW	7500/700 = 10.7	11

Table 5.2.5: Installed Equipment Set (Fab 3)

Equipment Type:	Desired Fab Capacity:	Fractional Tools Required	Integer Tools Installed:
Tool #1	3000 WSPW	3000/350 = 8.6	9
Tool #2	3000 WSPW	3000/5000 = 0.6	1
Tool #3	3000 WSPW	3000/1500 = 2	2
Tool #4	3000 WSPW	3000/400 = 7.5	8
Tool #5	3000 WSPW	3000/2500 = 1.2	2
Tool #6	3000 WSPW	3000/700 = 4.3	5

Although this method results in an equipment set for each fab that is roughly balanced (in terms of capacity), one or more tools in each facility stand out as constraints because of the exact integer quantities of tools that are required to meet the desired production level

of that fab. For Fab 1, the constraint tools are Tool #2 and Tool #5, which require exactly one and two tools respectively to meet a production goal of 5000 WSPW. For Fab 2, the constraint tools are Tool #3 and Tool #5, which require exactly five and three tools respectively to meet a production goal of 7500 WSPW. For Fab 3, the constraint tool is Tool #3, which requires exactly two tools to meet a production goal of 3000 WSPW.

5.2.3. Amount and Location of Excess Capacity.

The remaining, non-constraint tools in each facility have excess capacity above the fractional number of tools required. This excess capacity on some tools in one or more of the facilities provides the opportunity to produce additional wafer starts across the virtual factory by processing wafers across multiple facilities. Before we can test this particular optimization model, however, we must ascertain the exact amount and location of this disparate excess capacity. Equations 4.2.1 & 4.2.2 permit us to easily calculate the excess capacity that is available for each equipment set at each fab. Tables 5.2.6 - 5.2.8 summarize the results of these calculations.

Table 5.2.6: Excess Capacity by Tool Type (Fab 1)

Equipment Type:	Units Required to Meet Desired Capacity:	Total Available Capacity:	Excess Capacity Above Fab's Rated WSC:
Tool #1	15	15*350 = 5250	5250 - 5000 = 250
Tool #2	1	1*5000 = 5000	5000 - 5000 = 0
Tool #3	4	4*1500 = 6000	6000 - 5000 = 1000
Tool #4	13	13*400 = 5200	5200 - 5000 = 200
Tool #5	2	2*2500 = 5000	5000 - 5000 = 0
Tool #6	8	8*700 = 5600	5600 - 5000 = 600

Table 5.2.7: Excess Capacity by Tool Type (Fab 2)

Equipment Type:	Units Required to Meet Desired Capacity:	Total Available Capacity:	Excess Capacity Above Fab's Rated WSC:
Tool #1	22	22*350 = 7700	7700 - 7500 = 200
Tool #2	2	2*5000 = 10000	10000 - 7500 = 2500
Tool #3	5	5*1500 = 7500	7500 - 7500 = 0
Tool #4	19	19*400 = 7600	7600 - 7500 = 100
Tool #5	3	3*2500 = 7500	7500 - 7500 = 0
Tool #6	11	11*700 = 7700	7700 - 7500 = 200

Table 5.2.8: Excess Capacity by Tool Type (Fab 3)

Equipment Type:	Units Required to Meet Desired Capacity:	Total Available Capacity:	Excess Capacity Above Fab's Rated WSC:
Tool #1	9	9*350 = 3150	3150 - 3000 = 150
Tool #2	1	1*5000 = 5000	5000 - 3000 = 2000
Tool #3	2	2*1500 = 3000	3000 - 3000 = 0
Tool #4	8	8*400 = 3200	3200 - 3000 = 200
Tool #5	2	2*2500 = 5000	5000 - 3000 = 2000
Tool #6	5	5*700 = 3500	3500 - 3000 = 500

5.2.4. Defining the Virtual Factory MCNF Nodes.

Although a typical semiconductor manufacturing process flow might consist of hundreds of steps which could be broken into tens or even hundreds of process windows (as defined in Section 4.2.1), for this particular case example, we will be examining a simple, ten-step process flow. The six equipment types utilized for this simplified, re-entrant flow are allocated as follows:

Table 5.2.9: Equipment Types Utilized for Each Process Step.

Process Step Number:	Equipment Type:
1	1
2	2
3	3
4	1
5	4
6	5
7	4
8	6
9	3
10	5

For the sake of additional model simplicity, we will also assume that each one of these ten steps is fully cross-facility process capable (i.e., each step is a process window in and of itself). Given, then, that a node for this particular MCNF model is defined by both fab and process window (Figure 4.2.1), we now have all of the necessary components to define our MCNF nodal network:

Table 5.3.0: Example Virtual Factory Node Assignments

Process Step Number:	Fab 1	Fab 2	Fab 3
1	Node 1	Node 2	Node 3
2	Node 4	Node 5	Node 6
3	Node 7	Node 8	Node 9
4	Node 10	Node 11	Node 12
5	Node 13	Node 14	Node 15
6	Node 16	Node 17	Node 18
7	Node 19	Node 20	Node 21
8	Node 22	Node 23	Node 24
9	Node 25	Node 26	Node 27
10	Node 28	Node 29	Node 30

Finally, using Equations 4.2.3 & 4.2.4 and our results from Tables 5.2.6 - 5.2.8, we need to quantify the capacity of each of the nodes in our network in terms of rounded 25 wafer lots for both the model's Node Capacity Constraint (Section 4.3.2, Constraint 2) and the

Tool Capacity Constraint (Section 4.3.2, Constraint 3). Table 5.3.1 summarizes these calculations.

Table 5.3.1: Example Virtual Factory Node Capacities.

Node Number:	Excess Node Capacity Available:	Excess Node Capacity in Lots:	Final MCNF Model Node Capacity:
1	250	250/25 = 10	10
2	200	200/25 = 8	8
3	150	150/25 = 6	6
4	0	0	0
5	2500	2500/25 = 100	100
6	2000	2000/25 = 80	80
7	1000	1000/25 = 40	40
8	0	0	0
9	0	0	0
10	250	250/25 = 10	10
11	200	200/25 = 8	8
12	150	150/25 = 6	6
13	200	200/25 = 8	8
14	100	100/25 = 4	4
15	200	200/25 = 8	8
16	0	0	0
17	0	0	0
18	2000	2000/25 = 80	80
19	200	200/25 = 8	8
20	100	100/25 = 4	4
21	200	200/25 = 8	8
22	600	600/25 = 24	24
23	200	200/25 = 8	8
24	500	500/25 = 20	20
25	1000	1000/25 = 40	40
26	0	0	0
27	0	0	0
28	0	0	0
29	0	0	0
30	2000	2000/25 = 80	80

Problem Statement:

Now that we qualitatively and quantitatively understand the existing virtual factory scenario, we ask: Can additional wafer starts capacity be realized by simply allowing wafers to be cross-processed between multiple fabrication facilities? More specifically, can a goal of 250 extra wafers per week (10 lots) be processed across this hypothetical virtual factory without the addition of any new equipment?

This goal represents a modest weekly production increase of 1.6% and would be obtained without *any* capital investment in new processing equipment.

5.3. Optimization Model.

A general form of the optimization program presented in Section 4.3 can be used to determine an optimal cross-facility process path that maximizes the wafer starts capacity of this virtual factory. Recall that the resulting process path solution to this model will be for *additional* wafer starts added to the virtual factory. We assume that each individual facility will continue to process its normal allotted weekly wafer starts, up to its designated WSC, concurrently.

5.3.1. Case Example Model Inputs.

As previously described in Section 4.3.1, the inputs to this model are: 1) The desired number of total additional virtual factory wafer starts (in lots), 2) The cross/intra-facility cost weighting factors, and 3) The amount of excess capacity available for each equipment type at each facility (i.e., node capacities). Figure 5.3.1 indicates the representation of these inputs in the actual MS Excel implementation of this model.

Figure 5.3.1: Model Inputs.

Initial Additional VF Input:
10

Cost Data:										
Cost Xfer Inside:	0	0	0	0	0	0	0	0	0	0
Cost Xfer Outside:	1	1	1	1	1	1	1	1	1	1

Fab 1 Tool Capacity:		
Tool Number:	Tool Cap:	Cap. Used
1	10	0
2	0	0
3	40	0
4	8	0
5	0	0
6	24	0

Fab 2 Tool Capacity:		
Tool Number:	Tool Cap:	Cap. Used
1	8	0
2	100	0
3	0	0
4	4	0
5	0	0
6	8	0

Fab 3 Tool Capacity:		
Tool Number:	Tool Cap:	Cap. Used
1	6	0
2	80	0
3	0	0
4	8	0
5	80	0
6	20	0

As indicated above, this model employs a simplified cross/intra-facility process weighting system. All wafer lot transfers that occur within a single facility are assigned a cost weighting, C_{ij}, of zero, while all transfers between distinct facilities are given a cost weighting of one.

5.3.2. Case Example Model Outputs.

As described in Section 4.3.3, the outputs of this model are: 1) The optimized cross-facility process paths of all additional wafer lots added to the virtual factory, 2) The MCNF node utilization data, 3) The excess tool capacity utilization data, 4) The total cost of processing the additional wafer lots (given the input cost weighting factors assigned), and 5) The total number of combined cross-facility shipments required to process the additional wafer lots. For this particular scenario, Figure 5.3.2 specifies the representation of these outputs in the actual MS Excel implementation of this model.

Figure 5.3.2: Model Outputs.

Xfer from Fab 1:										
Process Window #:	1	2	3	4	5	6	7	8	9	10
Step Tool Number:	1	2	3	1	4	5	4	6	3	5
To Fab 1	0	0	10	0	0	0	8	10	0	0
To Fab 2	0	0	0	4	0	0	0	0	0	0
To Fab 3	0	0	0	6	0	0	0	0	10	0
Node Out Total:	0	0	10	10	0	0	8	10	10	0
Node In Total:	0	0	10	10	0	0	8	10	10	0

Total Cost Fab 1 =	20

Xfer from Fab 2:										
Process Window #:	1	2	3	4	5	6	7	8	9	10
Step Tool Number:	1	2	3	1	4	5	4	6	3	5
To Fab 2	8	0	0	0	0	0	0	0	0	0
To Fab 1	0	8	0	0	0	0	0	0	0	0
To Fab 3	0	0	0	0	4	0	0	0	0	0
Node Out Total:	8	8	0	0	4	0	0	0	0	0
Node In Total:	8	8	0	0	4	0	0	0	0	0

Total Cost Fab 2 =	12

Xfer From Fab 3										
Process Window #:	1	2	3	4	5	6	7	8	9	10
Step Tool Number:	1	2	3	1	4	5	4	6	3	5
To Fab 3	2	0	0	0	6	2	0	0	0	10
To Fab 1	0	2	0	0	0	8	2	0	0	0
To Fab 2	0	0	0	0	0	0	0	0	0	0
Node Out Total:	2	2	0	0	6	10	2	0	0	10
Node In Total:	2	2	0	0	6	10	2	0	0	10

Total Cost Fab 3 =	12

Total VF Cost =	44

Total # X-Ships:	8

Fab 1 Tool Capacity:		
Tool Number:	Tool Cap:	Cap. Used
1	10	10
2	0	0
3	40	20
4	8	8
5	0	0
6	24	10

Fab 2 Tool Capacity:		
Tool Number:	Tool Cap:	Cap. Used
1	8	8
2	100	8
3	0	0
4	4	4
5	0	0
6	8	0

Fab 3 Tool Capacity:		
Tool Number:	Tool Cap:	Cap. Used
1	6	2
2	80	2
3	0	0
4	8	8
5	80	20
6	20	0

As can be seen from these outputs, it is possible to process an additional 250 wafer starts within this particular virtual factory scenario by utilizing cross-facility processing. The total cost factor of the project, given the simplified weighting factors used, is 44, while the total number of cross-facility shipments required to process all of the wafer lots is eight.

5.4. Detailed Summary of Results.

The purpose of this section is to explain in detail the optimized cross-facility process paths generated by the model and depicted in Figure 5.3.2., as they may be more difficult for a first time user of the model to follow. In addition, this section contains a brief constraint analysis of the scenario results in order to foster an even better understanding of the model and this particular virtual factory's limitations.

5.4.1. Summary of Optimized Cross-Facility Process Paths.

A verbal, step by step summary of the optimized process paths depicted in Figure 5.3.2 is as follows:

- Initially, eight wafer lots are loaded into Fab 2 and processed for step 1, and two lots are loaded into Fab 3 and processed for step 1.
- For step 2 of the process, the eight lots in Fab 2 are again processed in Fab 2 and then shipped to Fab 1 for step 3. The two lots in Fab 3 are again processed in Fab 3 for step 2 and then shipped to Fab 1 for step 3.
- All ten lots are then processed in Fab 1 for steps 3 and 4. After they have been processed for these steps, four lots are shipped to Fab 2 and six to Fab 3 for step 5.
- The four lots in Fab 2 are then processed for step five and then sent to Fab 3 for step 6, while the six lots in Fab 3 are processed for step 5 and retained in Fab 3 for step 6.

- All ten lots are then processed in Fab 3 for step six. After that step, two lots remain in Fab 3 for step 7, while the other eight are shipped to Fab 1 for step 7.
- The two lots in Fab 3 are then processed for step 7 and then sent to Fab 1 for step 8, while the eight lots in Fab 1 are processed for step7 and retained in Fab 1 for step 8.
- All ten lots are then processed in Fab 1 for steps 8 and 9.
- Finally, for the last process step, all ten lots are shipped to and processed at Fab 3.

It might not be intuitive that the paths just described are indeed the lowest cost/least shipping paths possible for this scenario. After all, for every equipment set in the virtual factory except tool four, at least one or more of the facilities has at least ten lots of extra capacity. This would intuitively point to more simplified processing paths than the ones just detailed. The reason for this rather counter-intuitive complexity is, again, the re-entrant flow nature of semiconductor fabrication discussed in Section 2.3. Thus, when allotting the excess capacity available at a particular tool group at a particular fab, one must be cognoscente of *all* steps that will be performed concurrently on that equipment when the process is in steady state.

5.4.2. Constraint Analysis.

By comparing the total excess tool capacity used to the total excess tool capacity available for this example scenario, we can determine what the overall constraint to the virtual factory is, and hence, what limits increasing the cross-facility fabrication capacity.

Figure 5.4.1: Virtual Factory Constraint Analysis.

Total VF Tool Capacity:		
Tool Number:	**Tool Cap:**	**Cap. Used**
1	24	20
2	180	10
3	40	20
4	20	20
5	80	20
6	52	10

As can be clearly seen in the shaded area of Figure 5.4.1, the overall virtual factory constraint in this scenario is tool group four. This result is somewhat surprising in that tool group four is not the constraint equipment group in any of the individual facilities. In fact, with tool groups two and five acting as constraints in Fab 1, tool groups three and

five acting as constraints in Fab 2, and tool group three acting as the constraint in Fab 3, this scenario demonstrates that a virtual factory of manufacturing facilities *can* have as its overall constraint a tool group that is *not* a constraint in any of the component facilities.

Clearly as well, if the cross-facility fabrication capability of this particular scenario were to be enhanced, we would need to focus on adding capacity to tool group four. We will address this issue again, in detail, in Chapter 7.

Although the virtual factory network presented in this chapter was simplified for demonstration purposes, the reader should be aware of the extent to which actual networks can be larger. While the scenario presented here contained only 90 decision variables (i.e., three possible shipping choices for each of three fabs across ten nodes) and 228 constraints (i.e., 18 for tool capacity, 90 for flow balance, 30 for node capacity, and 90 for nonnegative, integer flow), an actual virtual factory scenario might contain upwards of 2500 decision variables and 6000 constraints. For this reason, solving these complex problems in practice typically requires larger-scale, commercial-grade linear programming systems.

6. Assessment of Technical Risks and Other Concerns.

Though the potential fabrication capacity, and hence revenue, gains from cross-facility processing merit serious consideration, they must also be cautiously measured against the significant technological risks and additional organizational effort needed to make such an operational policy possible. In addition, the basic numerical assumptions required to develop the requisite MCNF model of such a plan provide an added element of risk and might limit the extent to which cross-processing gains reach fruition. This chapter will address these concerns.

The first section of this chapter discusses the numerous technical risks inherent in any cross-facility fabrication policy. The second section addresses the issue of organizational implications of cross-facility processing. The third section explores the uncertainties and risks intrinsic to the basic numerical assumptions required for such a model to be developed, and the forth section addresses realistic limitations to the model's upside capacity gain potential.

6.1. Technical Risks.

Although there are literally hundreds of technical risks and concerns associated with the cross-facility processing of wafers, all can be summarily grouped into one of the two primary risk/concern categories listed in Table 6.1 and detailed thereafter.

Table 6.1: Technical Risk Groups Inherent to Cross-Facility Fabrication.

Shipping Environment:
• Particulate Contamination and Wafer Breakage.
• UV Light Exposure.
• Temperature and Humidity Deviations
Wafer Misprocessing:
• Loss of Wafers in Misrouted Lots Only.
• Loss of All Lots on the Tool.

6.1.1. Shipping Environment Risks and Concerns.

In any decision to transport wafer lots that have already undergone some level of processing, the first concern is the integrity of the environment in which the wafers will be shipped. The ultra-clean, delicate environment in which the wafers' valuable devices were fabricated (or partially fabricated) must be painstakingly maintained in their journey from one virtual factory facility to the next. Amongst the more prominent risks/concerns in this area are:

- **Particulate Contamination and Wafer Breakage.** No plan for cross-facility processing would be acceptable without sufficient safeguards against both particulate contamination of the wafers and wafer loss due to physical breakage. Contaminates of less than one micron in size on a wafer can be just as critical to wafer integrity and result in the same revenue losses as physically breaking the wafer. Risk in the former category can be minimized through vigilant extension of environmental standards and handling/exposure procedures, already in use within each facility, throughout the shipping process (i.e., at packaging, shipping, unpackaging, flow re-introduction, etc.). Risk management in the latter category, however, must also focus on the training and validation of third party product handlers (i.e., contracted shippers). They too must apply acceptable wafer handling techniques, analogues to those used for the initial delivery of clean wafers to each facility, at each of the steps in the shipping process they own.
- **UV Light Exposure.** Although strongly discouraged (see Section 4.2.1), a company might ship wafer lots from one facility to another with either patterned or unpatterned photo-resist intact on them. If this decision is made, the wafers being shipped run the risk of partial patterning/development defects in any areas that might have been accidentally exposed to UV light during the shipping process. Thus, it is recommended for all wafer shipments, and compulsory for all shipments which container wafers with photo-resist intact, that adequate protection against UV light exposure to afforded at all phases of the wafer transportation process.

- **Temperature and Humidity Deviations.** Similar to the UV light concern, the temperature and humidity of the shipping environment must also be kept at fabrication facility levels at all phases of the shipping process. Elevated temperature levels in the shipping environment might cause inadvertent wafer processing at many points in the process. For example, high shipping environment temperatures might accelerate incidental oxide growth beyond acceptable levels at certain points in the process flow or result in unwanted resist development at others. In addition, unacceptably high levels of humidity in the shipping environment might 'lift' certain thin films deposited on the wafer at various steps in the process. Given that the fabrication of semiconductor devices tends to be concentrated in the warmer and/or more humid climates of the world (i.e., the southwest U.S., southeast Asia, etc.), any cross-facility transportation plan must procedurally assure that wafer lots are not allowed to sit in hot, humid shipping and receiving facilities without adequate environmental safeguards.

6.1.2. Wafer Misprocessing Risks and Concerns.

Wafer lots would most commonly be misprocessed in shipping them from one fabrication facility to another by being introduced into the wrong processing 'short loop' upon entering the new fab's process flow. In semiconductor manufacturing, a 'full loop' represents the entire fabrication process flow for a given product, while a short loop represents only a part of that full loop. The number of process steps in short loops can vary from one or two (e.g., the deposition of a film and its corresponding dimensional check step) to several dozen. The only limiters to a particular short loop's origin and termination steps are the process window considerations mentioned in Section 4.2.1 and the judgment of the personnel at each participating facility.

There are essentially two consequences of introducing a wafer lot from one facility into the wrong processing short loop at another. Both are costly, but one can potentially be devastating to product *beyond* the misprocessed lot.

Consequence 1: Loss of Wafers in Misrouted Lots Only.

If the facilities involved in this particular cross-facility short loop misrouting are fortunate, the erroneous process steps that the transferred lot(s) have undergone will result in the functional loss of the wafers in those designated lots only. This could easily be the case with a repeated short loop at any number of points in the process. For example, suppose a lot of wafers has an oxide layer grown on them in Fab A and are then to be sent to Fab B to have that oxide patterned and etched in a predetermined short loop. Due to a routing error, however, these wafers are sent to a short loop in Fab B that designates that the oxide just grown in Fab A be grown again in Fab B. The result, of course, would be that these wafers are now functionally worthless, given they have roughly twice the specification thickness of a key oxide layer. The other wafers in the oxide growth furnace, however, are not damaged by the misprocessing.

Consequence 2: Loss of All Lots on the Tool.

Unlike the example above, many short loop misroutings between separate facilities can result in the loss of *all* wafers being processed on the tool with the misrouted lot. In addition, if the error is not immediately detected, the misrouting can result in the continued loss of all wafers processed on the tool even after the misrouted lot has moved on. For example, suppose a metal layer has been deposited on a lot of wafers in Fab A, and these wafers are to be shipped to a short loop in Fab B, where that particular metal layer will be patterned and etched. Due to a routing error, however, these wafers are sent to a short loop in Fab B that designates a high temperature annealing of a recent ion implantation step. The outcome of this error would likely be the complete contamination of the diffusion furnace and loss of all wafer lots in it through a process known as 'outgassing,' where certain solids release gases and/or vapors under a specific range of temperatures and pressures. In addition, all subsequent lots loaded into that particular furnace will be lost until the tool is adequately cleaned, or more likely, replaced.

While the overall risk of misprocessing wafer lots might increase with the introduction of a cross-facility fabrication plan, the risk already exists without it. The current situation in

many fabrication facilities today requires the processing of many different products, utilizing a variety of different process flows, on the same equipment on a daily basis. If safeguards are already in place to contend with process flow misroutings for this kind of environment, it is unlikely that wafer lots from other facilities would pose much of an additional problem.

6.2. Organizational Implications of Cross-Facility Processing.

As cross-facility processing capability is already in place for more case-specific instances at Intel and other semiconductor manufacturers, the additional physical infrastructure required to implement an ongoing cross-facility processing plan would be minimal. In fact, if the overall volume of cross-processed wafer lots is low enough in relation to overall virtual factory capacity (as in the example of Chapter 5), it is likely that the existing physical and human infrastructure in production and operations control would be sufficient to handle the increased traffic of wafer lots entering and leaving the facility.

Where additional infrastructure will definitely be required, however, is the area of organizational and/or procedural alterations to production and operations control. More specifically, the following are seen as necessary to sustaining any organization's plan for continuous, frequent cross-facility processing:

- **Key Contacts:** All functional area supervisors, shift managers, and key production control personnel from each virtual factory member facility need to know and be in constant contact with one another regarding operational conditions and concerns in their areas/fabs. Monthly or quarterly face-to-face meetings are also advisable.
- **Priority Systems:** Most fabs already have systems in place that prioritize re-entrant work in progress (WIP) within their facility. These systems must be expanded to comprehend cross-processed WIP from other facilities as well.
- **Lot ID's and Monitoring Systems:** The ID code assigned to any cross-facility processed lot must remain constant throughout the entire process flow. In addition, all virtual factory member facilities should be afforded continuous, real-time tracking

capability of all cross-facility processed lots for which they have at least partial ownership.

- **Reward and Ownership Systems:** Fabs in many companies are typically rewarded based on their total number of wafers produced, or 'outs.' Clearly, a system must be put in place that will also reward individual fabs for their processing contributions to wafers that were processed across multiple facilities. These systems should also clearly spell out who has ownership of the cross-processed wafers at all points in the complete process flow.
- **Engineering Awareness:** Engineering teams that typically conduct tests and/or experiments in the fab by continuously introducing 'test wafers' into fab short loops need to be aware of the strategic importance of cross-facility processing and the implications of their actions on its operation. Large numbers of engineering test wafers introduced into particular short loops at particular fabs can severely limit, if not totally eliminate, cross-facility processing capability for a given virtual factory network.

Implementation of the organizational systems and procedures identified above will also serve to further reduce many of the technical risks identified in Section 6.1.

6.3. Assessment of Numerical Assumptions.

Excess capacity availability calculations made for each facility's tool groups, and subsequently MCNF model nodes, rely on two important assumptions 1) The processing capability of each tool group in each fab does not change over time, and 2) Each virtual factory member facility will always run at optimal production levels. While these assumptions are necessary for the model to function at a given snapshot in time, they are very unrealistic in a real world production environment. Not only will a typical virtual factory of production facilities see quarterly (if not monthly or even weekly) changes in the production capacity of some tool groups through continuous improvement projects and operational enhancements, but it will also experience wide, almost instantaneous

decreases (or even temporary elimination) of processing capacity in some areas due to tool outages or unscheduled preventative maintenance.

Clearly any operationally viable implementation of this model would have to be continuously updated on at least a weekly basis to comprehend the ongoing changes in the processing capability of all individual tool groups throughout the virtual factory. Thus, every week, additional wafer starts newly introduced to the virtual factory would be allocated across member facilities according to the optimized paths generated by the MCNF model for that week. Additional wafer starts already in the virtual factory process flow would then either be kept on the optimized path determined for them by the MCNF model on their week of introduction or reallocated according to capacity data generated by the new week's model, whichever the virtual factory team suggested in Section 6.2 deems most appropriate.

6.4. Limitations to Capacity Gains.

In studies utilizing this model and methodology already conducted, and in the example virtual factory design of Chapter 5, weekly production volume gains discovered possible to date have been on the order of 0.5% to 2.75% of total virtual factory network volume. Given that it is not at all uncommon for weekly production volume in this kind virtual factory network to fluctuate as much as 5% to 10% per week (given the factors addressed in Section 6.3), it is entirely possible that the capacity gain benefits of such a processing program may often be completely overshadowed by the typical variation in weekly production volume.

It should be noted, however, that given the number of die typically contained on each wafer and the levels of revenue that those die can conceivably generate, a 0.5% to 2.75% production volume increase can often result in increased gross revenues of one to several hundred million dollars.

Additionally, it should be noted that the greatest capacity gains observed occurred in facilities whose overall WSC was not well planned against the processing capabilities of the individual tool groups that comprise them (i.e., the overall WSC was always a small fractional multiple of all the associated tool groups). Thus, the greatest potential benefit for a model such as this appears to lie with virtual factories whose component facility capacities were not well planned initially.

7. Additional Model Uses and Benefits.

In spite of the associated risks and concerns just discussed, the cross-facility fabrication planning model and methodology presented in this thesis does have additional uses that can significantly benefit a network of virtual factory facilities. This chapter addresses these benefits. The first two sections discuss the model's usefulness in conducting cost/benefit analyses for both incremental tool additions and/or reductions to the virtual factory, while the third section addresses the model's usefulness in strategically designing future virtual factory networks with cross-facility processing intent.

7.1. Analyses of Incremental Tool Additions.

The Theory of Constraints (TOC) teaches us that the most effective way to expand capacity in a production facility is to increase the processing capability of the facility's constraint or 'bottleneck' operation. This principle is in fact no less valid for a virtual factory of facilities viewed collectively than it is of any single facility viewed alone. The only difference, of course, is in identifying and exploiting the true constraint of the virtual factory overall through comprehensive analysis of total virtual factory capacity in each functional area and tool group (See Section 5.4.2). Additionally, it is also important to recall that the overall constraint of the virtual factory *does not* have to be a constraint of any of the member facilities (The constraint analysis of our example virtual factory scenario in Section 5.4.2 illustrates this possibility).

Clearly, the most simple way to expand productive capacity across any virtual factory tool group (i.e., beyond simple continuous improvement projects) is to add incremental tools of that particular group to one or more of the member facilities. Given these tools can often cost millions of dollars to merely purchase and additional hundreds of thousands of dollars to then install, it becomes very important to know a) What capacity, and hence revenue, gains are enabled by the tool addition(s)? and b) Where in the virtual factory should the tool(s) be allocated (i.e., in which facility)? Without this information, any attempt to conduct a cost/ benefit analysis of the project would be futile.

Consider as an example cost/benefit analysis the same virtual factory network described in Chapter 5. As demonstrated in Figure 5.4.1, tool four, which experienced full utilization of its 20 additional wafer lots of capacity, was the overall constraint tool group of this particular virtual factory network. Recall from the initial information given in that example that one additional tool of the tool four group would yield 400/25 = 16 lots of excess capacity. Remember also that two process steps were run on tool four. Thus, given its re-entrant flow nature and the fact the tool one operates near its capacity with 83% utilization of its excess capacity available, we would not necessarily expect that the 16 extra lots of capacity afforded by the addition of one tool four will translate into 16 extra wafer lots processed overall. How many total additional wafer lots will be realized? And where is the optimal facility location for the new tool? These questions can be easily answered with three quick variation tests of the original virtual factory scenario of Chapter 5. More specifically, holding all other inputs constant, the excess capacity available for tool four will be increased by 16 lots individually for each of Fabs 1, 2, and 3 in a separate model scenario. For each scenario, then, the new additional wafer lots capacity will be measured along with the total cost and total number of shipments required to process the wafers. Table 7.1 indicates the results of this study.

Table 7.1: Incremental Tool Four Addition Study Results.

Tool Four Added to:	Total Additional Wafer Starts Possible:	Total Virtual Factory Cost:	Total Number of Cross-Facility Shipments:
Base Case: No Addition	10	44	8
Fab 1	12	48	6
Fab 2	12	56	10
Fab 3	12	48	7

As can be seen in these results, the addition of one tool four in this particular virtual factory network will enable an additional two wafer lots (50 wafers) per week over Chapter 5's base case scenario at a total virtual factory cost of 48. Given that locating the tool in Fab 1 yields the lowest overall cost and number of cross-facility shipments, it is

clearly the best facility in which to install the tool. In addition, it is also interesting to note that while the overall cost of processing the 12 wafer lots was more than the base case example of 10 lots (48 vs. 44 in the base case), the overall number of shipments required to process 12 wafer lots with the additional tool is actually *less* than with the base case of 10 lots (6 vs. 8 in the base case).

From this point in the project investigation, the decision to go forward would have to center upon the results of a net present value (NPV) analysis of the tool addition, which would weigh the discounted sum of potential future revenues enabled by the tool's addition against its associated purchase, installation, and organizational/shipping costs.

7.2. Analyses of Incremental Tool Reductions.

A company's return on net assets (RONA) is a key ratio by which the quality of its operations is judged on Wall Street and throughout the financial community. If a company were thus enabled to generate the same amount of revenue on a decreased asset base, its RONA would improve, and its operations would be viewed more favorably by those third parties who routinely perform valuations of the company for the world at-large. The important questions, however, are 1) Which assets can be liquidated? And 2) Once liquidated, how will product be processed without the removed equipment? The answer to these questions is the subject of this section.

Consider again the basic virtual factory scenario presented in Chapter 5. Which capital tool assets can be eliminated from this particular virtual factory network without affecting the overall processing capability of all three facilities when viewed collectively? The answer can readily found by applying Equation 7.1 to each tool group in the virtual factory.

EQUATION 7.1: CALCULATION OF NONESSENTIAL TOOLS FOR EACH VF TOOL GROUP.

Number of Nonessential Tools = ROUND DOWN[(Total Excess Capacity Available)/(Tool WSC)]

When this equation is applied to the initial data provided for the example network of facilities, the resultant nonessential tools are identified as those whose Equation 7.1 result is greater than one.

Table 7.2: Nonessential Process Equipment for Example VF Network.

VF Tool Group:	Excess Capacity Available:	Tool WSC:	Nonessential Tool Calculation:	Number of VF Nonessential Tools:
Tool 1	24 Lots	14 Lots	24/14 = 1.71	1
Tool 2	180 Lots	200 Lots	180/200 = 0.90	0
Tool 3	40 Lots	60 Lots	40/60 = 0.67	0
Tool 4	20 Lots	16 Lots	20/16 = 1.25	1
Tool 5	80 Lots	100 Lots	80/100 = 0.80	0
Tool 6	52 Lots	28 Lots	52/28 = 1.86	1

Thus, for this particular network example, we can eliminate one tool of tool groups one, four, and six from the virtual factory without impacting its total output. Viewed from a RONA perspective, this outcome would result in an equivalent level of production capacity obtained by utilizing millions (if not tens of millions) of dollars less in capital equipment assets. Determining *how* the wafers would be processed across all facilities is a simple reformulation of the Chapter 5 base model scenario, which would comprehend the new WSC's for each virtual factory tool group.

Again, however, any decision to move forward with this particular project would need to be based on a sound NPV analysis of the tool eliminations. Such an analysis would not include the one-time, up front revenue gain of selling each tool in the open market, but also the on-going, added costs associated with the new level of cross-facility shipping necessitated by the project.

7.3. Strategic Virtual Factory Design with Cross-Processing Intent.

Today, most semiconductor fabrication facilities are designed without a single thought given to their future cross-facility processing capability. Typically, a high-level decision is made for new fabrication facility to have a specific wafer starts capacity. This WSC determination is usually a function of both projected market demand and the company's existing WSC in other facilities. Once a WSC has been determined for a fab, it is outfitted for equipment types in much the same way as was demonstrated in Section 5.2.2. That is, tools are allotted to each facility until the rated WSC of the tool group meets or exceeds the desired WSC of the fab.

Given this background information, the question arises as to the extent of enhanced cross-facility process capability that is possible from a virtual factory network that has been strategically designed to implement it. Studies utilizing the model presented in this thesis conducted to date have shown that although cross-facility processing might enhance weekly production capacity by only 0.5%-2.75% for existing virtual factory networks, capacity gains on the order of 5% - 10% are possible for virtual factory networks whose capacity and tool allocations were poorly planned.

Two methods exist for developing comprehensive, optimized virtual factory designs. The first method utilizes a more time-consuming, iterative approach, while the second implements a combination of minimum cost network flows and integer programming.

7.3.1. Iterative Virtual Factory Design Method.

The method used for the studies conducted to date was an iterative process which involved constant model reformulations according to the following procedures:

- Divide the desired overall WSC of the virtual factory by the number of desired member facilities. This number will be each individual facility's initial WSC.

- Allocate tools to each facility until the tool group's combined WSC meets or exceeds the fab's desired WSC *plus* a quantity equal to the number of desired cross-processed lots per week multiplied by the number of process steps which are run on that particular tool group. That is:

EQUATION 7.2: INITIAL TOOL GROUP WSC FOR STRATEGICALLY DESIGNED VF FACILITIES.

Tool Group WSC = [Fab WSC + (Number of Desired Cross-processed Lots) * (Number of Process Steps Run on that Tool)]

Then iteratively:

- Use the model to determine the cross-facility processing capability and cost of the virtual factory at that point.

- Exploit the virtual factory constraint according to the principles and methods outlined in Section 7.1.

- Eliminated nonessential tools from the virtual factory according to the principles and methods outlined in Section 7.2.

Although this process might seem unduly time consuming, it is possible for an experienced user of the model to conduct a complete a theoretical virtual factory design in this manner in only a day or less, a small time allocation for a project of multibillion dollar implication.

7.3.2. MCNF/Integer Programming Virtual Factory Design Method.

The other method that could be used to strategically design a complete virtual factory network employs a combination of minimum cost network flows and integer programming. Using this method, the quantity of installed tool capacity for each tool group at each virtual factory member facility becomes a decision variable instead of a constraint. These installed tool capacity decision variables can then be appropriately constrained to meet certain strategic goals of the virtual factory system overall. Example constraints could include: individual fab WSC goals, individual fab capital expenditure limitations, or both. After the new model has been developed, this method of design could prove more exacting (depending on the level of detail of the decision variable constraints) and less time consuming than the iterative process just discussed.

8. Conclusion.

This thesis has presented a new approach for increasing semiconductor processing capacity for a virtual factory network of independent facilities. In a July 1997 on-line article, Murphy [14] emphasizes the importance of Intel's need to find additional fabrication capacity quickly when he states: "Because Intel's out of capacity, they're sold out through September on the new stuff. It leaves a window for both AMD and Cyrix because Intel cannot possibly build enough chips." This situation points to the very essence of the cross-facility fabrication planning strategy: Instant additional processing capacity for only the cost of shipping the wafers.

In a September 1994 article, Uzsoy, Lee and Martin-Vega [19] maintain that "overall, the semiconductor industry provides a host of very difficult and challenging problems in production planning and scheduling. The complex nature of semiconductor manufacturing provides an area where the use of more advanced techniques may yield considerable benefits." Clearly, the linear optimization program and process methodologies presented in Chapters 4,5 and 7 of this thesis are examples of these 'advanced techniques.'

As addressed in detail in Chapter 6, however, implementation of a cross-facility processing model like the one presented is not without its many risks and concerns. More specifically, the many technical, financial, and organizational risks and concerns posed by the adoption of this model could easily outweigh any gains in production capacity, and hence revenue, for many companies and virtual factory design scenarios. Clearly, then, the level of success experienced in utilizing a model such as this is case-specific.

Depending on the nature of the original virtual factory design, thus, the most appreciable benefits from a cross-facility fabrication planning model like the one presented might be those which were detailed in Chapter 7. That is to say, if the initial virtual factory network design in question does not lend itself well to cross-facility processing as is (in terms of

% capacity gained), true benefit may lie in the analyses of adding or eliminating tools from the network and/or developing comprehensive 'next-generation' network redesign plans.

References

[1] Babikian, Richard and Marijke Sanders, "Application of a TOC Based Wafer Starts Modulation Methodology for Fab 10 Production Line Management," IMEC Publication, Intel Corporation, 1997.

[2] Bailey, Troy A., "Semiconductor Cluster Tool Availability," MSM/MSEE Thesis, MIT, June 1994.

[3] Barrett, Todd, "Change Management: Leveraging the Role of Data as a Semiconductor Manufacturing Process Evolves," M.S. Thesis, MIT, June 1996.

[4] Fong, Douglas, "An Overview of CMP," Graduate School of Engineering Report, Department of Materials Science, MIT, May 1997.

[5] Frontline Systems, Inc., "Microsoft Excel Enhanced Solver User's Guide Version 2.0."

[6] Goldratt, Eliyahu M., and Jeff Cox, *The Goal: A Process of Ongoing Improvement*, North River Press, Inc., Second Revised Edition, 1992.

[7] Goldratt, Eliyahu M., *It's Not Luck*, North River Press, Inc., 1994.

[8] Han, Byungho, "Development, Implementation, and Application of Project Prioritization System in Wafer Fabrication Manufacturing," MSM/MSCEE Thesis, MIT, June 1996.

[9] Intel 1996 Annual Report.

[10] Keith, Christopher K., "Accelerating Time-to-Money for a Semiconductor Wafer Fab Capacity Ramp," MSM/MSMSE Thesis, MIT, June 1996.

[11] McBride, Eric and Karl Kempf, "Metering Fab Starts by Application of Control Charts," IMEC Publication, Intel Corporation, 1997.

[12] Mlynarczyk, Michele M., "Improving Synergy in Multi-Site Microprocessor Manufacturing: An Analysis of a Copy Exactly Approach," MSM/MSMSE Thesis, MIT, June 1995.

[13] Muir, Steven H., "Semiconductor Fabrication Capacity Gains Through Optimal Allocation of Product Loadings and Equipment Sharing Across Process Technologies," MSM/MSEE Thesis, MIT, May 1997.

[14] Murphy, Michael, "Editor's Note," *California Technology Letter*, July 1997.

[15] Nahmias, Steven, *Production and Operations Analysis*, Irwin, Second Edition, 1993.

[16] Nickel, Randall T., "Improving Productivity of Semiconductor Manufacturing Cluster Tools," MSM/MSEE Thesis, MIT, May 1994.

[17] Rodgers, Bob, Bill Schofield, Dennis Danielson, Brian Verwer, and Karl Kempf, "A WIP Management Policy to Maximize Fab Velocity," IMEC Publication, Intel Corporation, 1997.

[18] Uzsoy, Reha, Chung-Yee Lee, and Louis A. Martin-Vega, "A Review of Production Planning and Scheduling Models in the Semiconductor Industry – Part I: System Characteristics, Performance Evaluation and Production Planning," *IIE Transactions*, Volume 24, Number 4, September 1992.

[19] Uzsoy, Reha, Chung-Yee Lee, and Louis A. Martin-Vega, "A Review of Production Planning and Scheduling Models in the Semiconductor Industry— Part II: Shop-Floor Control," *IIE Transactions*, Volume 26, Number 5, September 1994.

[20] Winston, Wayne L., *Introduction to Mathematical Programming: Applications and Algorithms*, Duxbury Press, Second Edition, 1995.

[21] Wolf, Stanley and Richard N. Tauber, *Silicon Processing for the VLSI Era, Volume 1: Process Technology*, Lattice Press, 1986.