

**REUSABLE MODELLING AND SIMULATION OF
FLEXIBLE MANUFACTURING FOR NEXT
GENERATION SEMICONDUCTOR MANUFACTURING
FACILITIES**

By

Khaled S. El-Kilany, B.Sc., M.Sc.

This thesis is submitted as the fulfilment of the requirement for the award of the degree
of Doctor of Philosophy (PhD) by research to:

Dublin City University

Prof. Mohie El Baradie

Dr. Paul Young

Prof. Saleem Hashmi

The School of Mechanical and Manufacturing Engineering

NOVEMBER 2003

I hereby certify that this material, which I now submit for assessment on the programme of study leading to the award of the degree of Doctor of Philosophy (PhD) is entirely my own work and has not been taken from the work of others save and to the extent that such work has been cited and acknowledged within the text of my work.

Signed:  (Candidate)

ID No.: 99145693

Date: 16 November 2003.

To the most wonderful man I have ever known...

My Father

May God Rest His Soul

ACKNOWLEDGEMENT

I thank God for granting me patience and perseverance to get through all the difficult times I faced, and for making some of the finest people I've ever met a part of life during my research period.

I thank Prof. M. H. Elwany (Faculty of Engineering – Alexandria University) for offering me this research opportunity. Prof. M. A. El Baradie tireless efforts, in setting up a research scheme with Intel, resulted in this opportunity. My deepest thanks and gratitude go to these two gentlemen; as if it weren't for them I would have not had this opportunity at the first place. I would like to further extend my thanks to Prof. El Baradie and Prof. Saleem Hashmi for their fatherly guidance, care, and concern; I'm honoured to have you, Sirs, as my supervisors.

It gives me a great pleasure and delight to thank Dr. Paul Young from all my heart and soul. Words just fall short in trying to express my deepest gratitude to him. His continues supervision; original ideas and brotherly guidance were the cornerstones that made this work come to light. Not only has he helped with my work, but also he helped with almost every single problem I faced during my stay in Ireland. Thank you dear sir for your tireless efforts and patience in guiding me through this work.

Many thanks to all the people I worked with at Intel Ireland for their friendliness and cooperation in providing all the required information needed for the research, especially Ger P Ryan and Kieran O'Connell. My thanks extend to Prof. James P Ignizio for all of his advice and helpful recommendations throughout the work.

I would also like to thank Prof. E. H. Roushdy, Prof. O. M. Gaafar, and Prof. O. Abdel Aziz (College of Engineering – Arab Academy for Science and Technology) for all their support, understanding, and patience.

To all my friends for their care and support, especially Amr Arisha for continuously offering help in every possible way. I would also like to thank everyone who has contributed, by any way, to my fulfilment of this work.

My deepest love and appreciation goes to my lovely wife, Asia, who supported me and encouraged me all the way showing how much she believes in me and in my work, with her unconditional love, limitless kindness, constant care, sincere prayers... and in every other possible way, simply just by being the nice person that she is. I just can't imagine how I could have accomplished this work without having her beside me. It is really difficult to part with such a wonderful person, and just for that I thank all her family for their trust and support.

To my dear son, Mohamed, for having the sweetest smile that is full of hope and happiness no matter what happens, it kept me going on for many times.

At last, but not in any measure least, to my family, my great mother and my lovely sister, for their love, trust, understanding, and for seeing me through this journey and for always standing by me providing every possible help even when we're far away from each other. Thank you for your endurance, serenity and for barring with me.

TABLE OF CONTENTS

LIST OF FIGURES.....	IV
LIST OF TABLES.....	VIII
ABSTRACT.....	IX
NOMENCLATURE.....	X
1 INTRODUCTION.....	1
1.1 Mission and Scope.....	2
1.2 Outline of the Thesis.....	4
2 LITERATURE REVIEW.....	5
2.1 Introduction.....	5
2.2 Semiconductor Manufacturing.....	9
2.2.1 History and Background.....	9
2.2.2 The Semiconductor Manufacturing Process.....	13
2.2.3 Conversion to 300mm Wafer Size.....	19
2.2.4 300mm Wafer Fabrication Facility.....	21
2.2.5 Key Differences and Major Concerns.....	25
2.3 Studying The Behaviour of 300mm Wafer Fabrication Facilities.....	26
2.3.1 General Approaches of System Studying.....	26
2.3.2 Systems Modelling.....	28
2.3.3 Modelling and Simulation.....	29
2.3.4 Issues in Modelling and Simulation.....	41
2.3.5 Simulation Software.....	47
2.3.6 Modelling and Simulation of Semiconductor Manufacturing Systems.....	49
2.4 Conclusions.....	56
3 AUTOMATED MATERIAL HANDLING SYSTEMS FOR 300 MM WAFER MANUFACTURING FACILITIES.....	57
3.1 Front Opening Unified Pod.....	58
3.2 300mm Stockers.....	59
3.3 300mm Production Equipment.....	61

3.4	Interbay Automated Material Handling System.....	62
3.5	Intrabay Automated Material Handling System	63
3.5.1	Effect of Increasing FOUPs Weight.....	64
3.5.2	Components of the Intrabay AMHS.....	64
3.5.3	Lots Delivery/Collection To/From Production Tools/Stockers	65
3.5.4	Different Visions for the Intrabay AMHS	66
3.6	AMHS Delivery Models of the Future	67
3.7	Definition of System Under Study	68
3.7.1	Rules for Lot Movements	68
3.7.2	Routing Rules.....	69
3.7.3	Pickup/Drop-off at Tools and Stockers.....	71
3.7.4	Moves Allowed at Production Tools.....	72
3.8	Background	74
3.8.1	Motivation	75
3.8.2	List of Issues	76

4 DEVELOPMENT OF A GENERIC TOOL FOR INTRABAY AMHS MODELLING AND SIMULATION (GTIA-M&S)..... 79

4.1	Introduction	79
4.1.1	The Extend™ Simulation Environment.....	79
4.1.2	Background Material	83
4.1.3	Attributes Definition	85
4.2	Problem Formulation	86
4.2.1	Objectives.....	87
4.2.2	Bounds of the Problem and Tool Complexity	87
4.2.3	Assumptions	90
4.2.4	Selection of Performance Measures	91
4.2.5	Tool Time Frame	91
4.2.6	End-User.....	92
4.3	Collection and Processing of Real System Data	92
4.3.1	Data Requirements.....	92
4.3.2	Identifying the Sources of Data	95
4.3.3	Collection of Information and Data.....	96
4.3.4	Handling the Different Data Problems.....	99
4.3.5	Probability Distributions for Random Variables	100
4.3.6	Choosing the User Input Interface.....	101
4.3.7	Building the Database	104
4.3.8	System Performance Data.....	105
4.4	Tool Conceptualisation.....	106
4.4.1	Intrabay AMHS.....	106
4.4.2	Re-entrant Material Flow	106
4.4.3	Generation of Different Entities	108
4.4.4	Decision Making Points.....	110
4.4.5	Interaction of Components	121
4.5	Translation	134
4.5.1	Challenges	134
4.5.2	Component Based Modelling.....	139
4.5.3	Incorporation of Special Features.....	141
4.5.4	Tool Refinement	143

4.6	Verification.....	143
4.7	Validation	144
4.8	Experimental Design.....	145
4.9	Production Runs and Analysis	145
4.10	Documentation	146
4.11	Implementation	146
4.12	Conclusions	147
5	SIMULATION EXPERIMENTS, RESULTS, AND ANALYSIS	148
5.1	Description of Experiments.....	148
5.1.1	Purpose	148
5.1.2	Model Developed.....	149
5.1.3	Performance Measures.....	149
5.1.4	Simulation Parameters	153
5.2	Experiment One: Intrabay AMHS Without Merge/Diverge.....	154
5.2.1	Determining the Number of OHVs Required	155
5.2.2	Determining the Achievable Moves Per Hour	158
5.3	Experiment Two: Intrabay AMHS With Merge/Diverge.....	159
5.3.1	Determining the Number of OHVs Required	160
5.3.2	Determining the Achievable Moves Per Hour.....	160
5.4	Comparing Intrabay AMHS Performance.....	161
5.4.1	Analysing the Performance of the Two Technologies.....	163
5.4.2	Comparing the Two Technologies for the Same Number of OHVs	163
5.4.3	Confirmation of the Results.....	168
5.4.4	Summary of Analysis and Results.....	169
6	CONCLUSIONS AND RECOMMENDATIONS FOR FUTURE WORK	171
6.1	Conclusions	171
6.2	Recommendations for Future Work.....	172
7	REFERENCES	173
	APPENDICES.....	182

LIST OF FIGURES

CHAPTER 1	1
Figure 1.1: Mission and scope of research.....	3
CHAPTER 2	5
Figure 2.1: Forces shaping the 21st century manufacturing environment.	7
Figure 2.2: Original graph by Gordon Moore [12].	10
Figure 2.3: Application of Moore's law at Intel [12].	11
Figure 2.4: 90 nm generation interconnects [15].	12
Figure 2.5. Basic steps of semiconductor manufacturing	14
Figure 2.6: Wafer fabrication process [20].	15
Figure 2.7. Re-entrant material flow [17].	18
Figure 2.8. Typical 300mm fab layout.	23
Figure 2.9. Difference between 200mm and 300mm fabs [31].	24
Figure 2.10: Different methods of approaching a system under study.	28
Figure 2.11: Steps in a simulation study [49].	37
Figure 2.12: Simulation takes too long to produce results [58].	43
Figure 2.13: Model reuse spectrum.	45
Figure 2.14: Modelling and simulation in semiconductor manufacturing, application area and purpose.	50
Figure 2.15: Two distinct model parts communicating with each other [74].	51
Figure 2.16: Relationship between effort to build the model and quality of results [76].	52
CHAPTER 3	57
Figure 3.1: The 300mm Automated Material Handling System (AMHS) concept.....	57
Figure 3.2: Front Opening Unified Pod.	58
Figure 3.3: Stockers linking the different material handling systems [31].	59
Figure 3.4. Multi-level input and output ports of a stocker.	60
Figure 3.5: Inside a stocker showing FOUPs stored on shelves and the stocker input/output ports (Courtesy of Daifuku Co., Ltd. [87]).	60
Figure 3.6: Mini-environment and loading ports [21].	61

Figure 3.7: Different parts inside a production tool (Courtesy of JMAR Technologies, Inc. [88]).	62
Figure 3.8: Monorail Overhead Shuttle (Courtesy of Shinko Technologies, Inc. [89]).	63
Figure 3.9: Overhead Hoist Vehicle (Courtesy of Shinko Technologies, Inc. [89]).	64
Figure 3.10: Bay width savings realised when using OHVs [90].	65
Figure 3.11: OHVs dropping off and picking up FOUPs at loading ports of production tools [91].	65
Figure 3.12: Linking a number of bays using merge/diverge links.	66
Figure 3.13: Moving lots between bays with and without merge/diverge.	67
Figure 3.14: Stocker-Tool-Stocker lot delivery in a bay.	69
Figure 3.15: Different routes an OHV can loop.	69
Figure 3.16: Decision tree for selection of different routes.	71
Figure 3.17: Illustration of the different statuses of OHVs and their respective pickup/drop-off case.	72
Figure 3.18: Locations of a lot ready for pickup and its effect on the pickup/drop-off process.	73

CHAPTER 4 79

Figure 4.1: Different animation types offered by Extend TM .	80
Figure 4.2: Creation of hierarchical blocks using Extend TM .	81
Figure 4.3: Dialog parameters cloning.	82
Figure 4.4: Component-based modelling of reusable components in Extend TM .	83
Figure 4.5: Sources of data.	95
Figure 4.6: Created AutoCAD layer and imported by Proof Animation TM .	100
Figure 4.7: Different data categories defined in the database.	104
Figure 4.8: Referencing data in one table to data in another table by indexing.	105
Figure 4.9: Material flow within GTIA-M&S.	107
Figure 4.10: Lots generation and Routing.	109
Figure 4.11: OHVs generation and routing.	110
Figure 4.12: Part of a model constructed using GTIA-M&S showing different decision points.	111
Figure 4.13: Distribution of lots sent to a bay over its stockers.	112
Figure 4.14: Deciding whether or not to store lots in a stocker.	113
Figure 4.15: Lot pickup/drop-off at stockers.	114
Figure 4.16: Selection of bay for lot collection.	115
Figure 4.17: Routing logic at merge/diverge link 1.	116

Figure 4.18: Lot pickup/drop-off at tools.....	117
Figure 4.19: Selection of bay for next move.....	118
Figure 4.20: Routing logic at merge/diverge link 2.....	120
Figure 4.21: Selecting stocker for next OHV move.....	121
Figure 4.22: IDEF0 node diagram.....	122
Figure 4.23: Emulation of pick up and drop-off of lots.....	135
Figure 4.24: Preventing OHVs overflow.....	136
Figure 4.25: Using a gate block for accurate decision-making.....	137
Figure 4.26: Using DB-aware attributes for changing the number of free loading ports.....	139
Figure 4.27: Configuring a merge/diverge link.....	140
Figure 4.28: Model-building template.....	142
CHAPTER 5.....	148
Figure 5.1: Developed model.....	149
Figure 5.2: Delivery time analysis – Bay 01 – 16 OHVs – merge/diverge ON.....	151
Figure 5.3: Delivery time distribution.....	152
Figure 5.4: Moves per hour analysis – Bay 01 – 16 OHVs – merge/diverge ON.....	152
Figure 5.5: Using the MPH for the three bays to determine the wamup period.....	153
Figure 5.6: Different replications for a model running under same operating conditions.....	154
Figure 5.7: Delivery time versus number of OHVs without merge/diverge.....	156
Figure 5.8: Different combinations of 7 OHVs.....	157
Figure 5.9: Different combinations of 8 OHVs.....	157
Figure 5.10: Delivery time versus number of OHVs without merge/diverge, after modification.....	158
Figure 5.11: Moves per hour achieved without merge/diverge.....	159
Figure 5.12: Delivery time versus number of OHVs with merge/diverge.....	160
Figure 5.13: Moves per hour achieved with merge/diverge.....	161
Figure 5.14: Bay 01 results when running with 7 OHVs.....	164
Figure 5.15: Bay 02 results when running with 7 OHVs.....	165
Figure 5.16: Bay 03 results when running with 7 OHVs.....	165
Figure 5.17: Improvement gained in each bay by having the merge/diverge switched ON.....	166
Figure 5.18: Average results across the three bays.....	167
Figure 5.19: Overall improvement gained by having the merge/diverge switched ON.....	168

Figure 5.20: Mean and 2-sigma values of the percentage improvement gained with merge/diverge..... 169

CHAPTER 6 171

LIST OF TABLES

CHAPTER 1	1
CHAPTER 2	5
Table 2-1: Challenges of manufacturing companies.	6
CHAPTER 3	57
CHAPTER 4	79
Table 4-1: List of attributes used in GTIA-M&S.	86
Table 4-2: Sample of the modified process flow used in GTIA-M&S.	94
Table 4-3: Different variable types used in GTIA-M&S.....	101
Table 4-4: List of the different decision points.....	111
Table 4-5: The different functions modelled using the IDEF0.	122
CHAPTER 5	148
Table 5-1: Number of OHVs set in each bay without merge/diverge.....	155
Table 5-2: Comparing the results of the two experiments.....	162
Table 5-3: Percentage improvement gained with merge/diverge ON for the different runs.	168
CHAPTER 6	171

ABSTRACT

Automated material handling systems (AMHS) in 300 mm semiconductor manufacturing facilities may need to evolve faster than expected considering the high performance demands on these facilities. Reusable simulation models are needed to cope with the demands of this dynamic environment and to deliver answers to the industry much faster. One vision for intrabay AMHS is to link a small group of intrabay AMHS systems, within a full manufacturing facility, together using what is called a Merge/Diverge link. This promises better operational performance of the AMHS when compared to operating two dedicated AMHS systems, one for interbay transport and the other for intrabay handling. A generic tool for modelling and simulation of an intrabay AMHS (GTIA-M&S) is built, which utilises a library of different blocks representing the different components of any intrabay material handling system. GTIA-M&S provides a means for rapid building and analysis of an intrabay AMHS under different operating conditions. The ease of use of the tool means that inexperienced users have the ability to generate good models. Models developed by the tool can be executed with the merge/diverge capability enabled or disabled to provide comparable solutions to production demands and to compare these two different configurations of intrabay AMHS using a single simulation model. Finally, results from simulation experiments on a model developed using the tool were very informative in that they include useful decision making data, which can now be used to further enhance and update the design and operational characteristics of the intrabay AMHS.

NOMENCLATURE

AMHS	Automated Material Handling System
AGV	Automated Guided Vehicles
AI	Artificial Intelligence
AITE	Automated Inspecting and Testing Equipment
AMT	Advanced Manufacturing Technology
APL/U	Automated Parts Loading/ Unloading
AS/RS	Automatic Storage and Retrieval System
ATC	Automated Tool Changes
CAD	Computer-Aided Design
CAE	Computer-Aided Engineering
CAM	Computer-Aided Manufacturing
CAPP	Computer-Aided Process Planning
CE	Concurrent Engineering
CIM	Computer Integrated Manufacturing
CM	Cellular Manufacturing
CNC	Computer Numerical Control
CRN	Common Random Numbers
DB	Data Base
DD	Direct Delivery
DLFT	Die-Level Functional Testing
DNC	Direct Numerical Control
ES	Expert System
FA	Factory Automation
Fab	Wafer Fabrication Facility
FIFO	First-In-First-Out
FLC	Fuzzy Logic Controller
FMC	Flexible Manufacturing Cell
FMS	Flexible Manufacturing System
FOUP	Front Opening Unified Pod
GCAPP	Generative CAPP
GT	Group Technology
GTIA-M&S	Generic Tool for Intrabay AMHS Modelling and Simulation

GUI	Graphical User Interface
I300I	International 300mm Initiative
IC	Integrated Circuit
IDEF	Integration DEFinition language
IDEF0	Intergration DEFinition language 0
IHS	Intelligent Hybrid Systems
IMS	Intelligent Manufacturing Systems
IPC	Inter-Process Communication
JIT	Just-In- Time
LAN	Local Area Network
LP	Loading Port
MC	Machine Cell
MD	Merge/Diverge
MPH	Moves Per Hour
MRP	Material Requirements Planning
MRPII	Manufacturing Resources Planning
MTBF	Mean Time Between Failures
NC	Numerical Control
NGM	Next Generation Manufacturing
NN	Neural Networks
ODBC	Open Database Connectivity
OHT	Overhead Hoist Transport
OHV	Overhead Hoist Vehicles
OR	Operations Research
PFA	Production Flow Analysis
PGV	Personnel Guided Vehicles
S ^G CAPP	Semi-Generative
SKR	Stocker
SME	Society Of Manufacturing Engineers
TTR	Time To Repair
VCAPP	Variant CAPP
VRT	Variance-Reduction Technique
WAN	Wide Area Network
WIP	Work-In-Progress
WSPW	Wafer Starts Per Week

1 INTRODUCTION

Next Generation Manufacturing (NGM) is a sweeping vision of the integrated, globally competitive manufacturing enterprise in the 21st Century. As businesses enter the twenty-first century, they have to cope with an increasingly competitive environment. Among other things, the competitive environment may be characterised by product proliferation with short and uncertain life cycles, innovative process technologies, customers who simultaneously demand quick response, lower costs, and greater customisation. Manufacturing systems have changed beyond recognition due to this new business environment and consequently these systems are typically huge and complex. To survive and prosper in this dynamic and uncertain competitive world, manufacturing systems must be able to react quickly and effectively to unanticipated changes. The automation and integration of almost every aspect of a manufacturing system can realise the required level of responsiveness in such a system.

Advanced manufacturing technology (AMT) is rapidly evolving to provide automation and integration capabilities for the manufacturing enterprise, offering hardware and software automation solutions for a multitude of manufacturing activities. With the increasing number of AMT solutions, coupled with the growing size and complexity of global manufacturing enterprises, it is not unusual to have massive amounts of information being transacted across the enterprise. Computer integrated manufacturing (CIM) automates the information transactions effectively and ties the different components of AMT in a manufacturing enterprise. Furthermore, artificial intelligence (AI) is offering more and more solutions to almost every function of a manufacturing system to automate decision-making and to help manufacturers maintain and further expand their competitive advantage. Indeed, the current advancement in AI techniques has led to the development of intelligent manufacturing systems (IMS).

Semiconductor manufacturing industry is characterised by extremely short product life cycle, constantly decreasing profit margins, and increasingly intense competition. Investments in developing process technologies and manufacturing facilities are well

over a billion dollars. Manufacturing of a semiconductor requires highly complex and sensitive manufacturing processes of features that are well in the nanometres range using highly customised manufacturing equipment. This, arguably, qualifies the process of manufacturing semiconductors on silicon wafer substrates to be one of the most expensive complex manufacturing processes in existence.

300mm wafer fabrication is the most recent development in the semiconductor manufacturing, where semiconductors are manufactured on a silicon wafer substrate of 300mm diameter wafers replacing the older 200mm ones. A new automation component, the intrabay automated material handling system AMHS, is needed in the 300mm manufacturing facilities to support high volume manufacturing and to solve safety and ergonomics issues. The design, operation, and configuration of the intrabay AMHS is, as still, ambiguous to the industry. The complexity of these systems coupled with their relative novelty has provided researchers a green field for study and analysis. Predictive discrete event simulation models seem to be the only paramount method of evaluating the different designs and configurations, proposed for the intrabay AMHS, and understanding the behaviour of the non-existing system under different operating conditions.

1.1 MISSION AND SCOPE

Since the research work started, it was known that the final target is the next generation semiconductor manufacturing facilities, which are the 300mm manufacturing facilities. However, two main lines of work are identified as shown in Figure 1.1.

An intelligent group technology approach to the design of the semiconductor facility machine cells, which is based on fuzzy logic for arranging the machines with a group technology layout. This line of work has stopped at establishing the concepts of the approach as the industry showed little interest in implementing different layout for their facility. Appendix (B) presents the outcome of work done in that area.

The industry is concerned to know the behaviour and performance to expect from the automated material handling system used in 300mm wafer fabrication facilities. Analytical solution for this system is not available, making simulation a reasonable methodology for studying the AMHS.

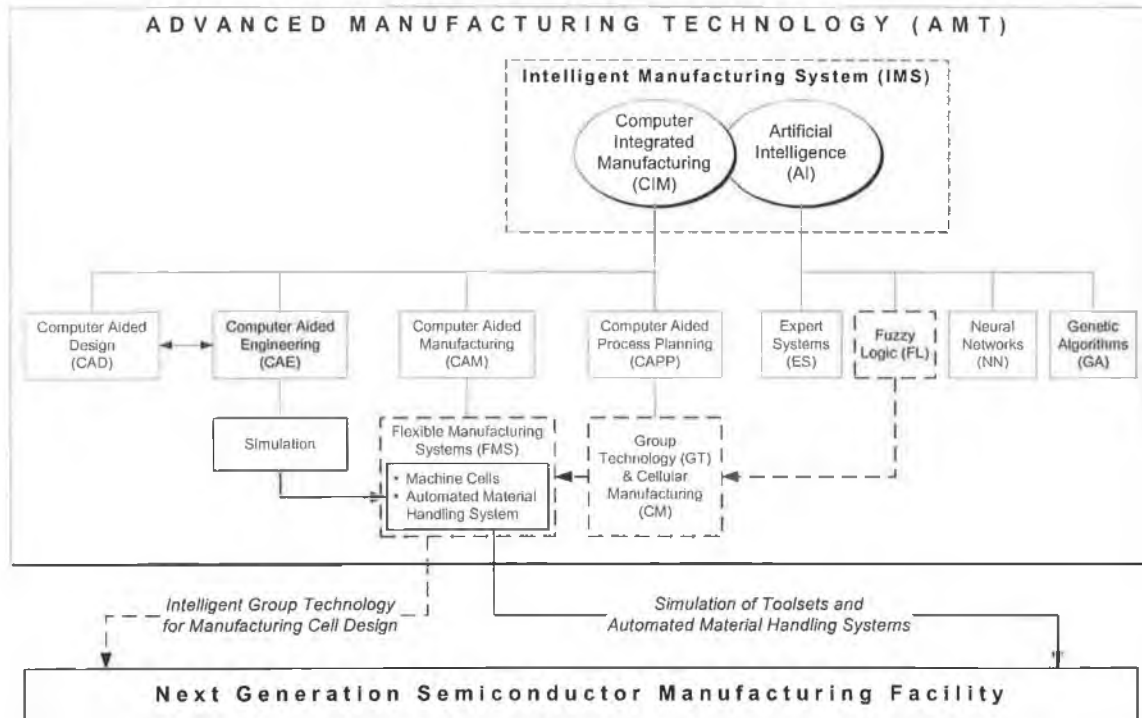


Figure 1.1: Mission and scope of research.

This line of work evaluates the intrabay AMHS running under different operating conditions and with different configurations. Essentially there are two visions proposed for the way that the intrabay AMHS will be implemented, and simulation is the only good method available to predict and compare the performance of the intrabay AMHS running with those two different technologies. The design of the intrabay AMHS is highly iterative, and modelling it is not a one-off process, hence a tool has been created specifically to address this problem, a generic tool that allows fast modelling and simulation of new designs of the system and can compare the two visions by merely switching between them. The tool comprises a number of reusable components that can be connected together, forming a network of components, to rapidly model an intrabay AMHS from scratch. In this way, rather than offering the industry one model with long development time and the need for rebuilding whenever the real system changes, the tool allows building several models with different configurations much faster. One of the key characteristics of this tool is that it includes both the production equipment and the intrabay AMHS in a single model showing the interaction between these two primary parts of a wafer fabrication facility. This helps in understanding the behaviour of the system and the integration between the production equipment and the AMHS. In addition it clearly shows the impact that these two parts have on one another. Results

from simulation experiments on a model developed using the tool show their merit in that they include many useful decision making data, which can now be used to further enhance and update the design and operational characteristics of the intrabay AMHS.

1.2 OUTLINE OF THE THESIS

The thesis consists of six main chapters and five appendices.

Chapter two covers a survey of literature offering background and related work to this research.

Chapter three covers a detailed description of the system under study and defines the scope of the work.

Chapter four details the development of the generic tool for modelling and simulation of intrabay AMHS.

Chapter five presents simulation experiments conducted using the developed tool, as well as the results and analysis of the simulation output.

Chapter six covers the conclusions and recommendations for future work.

Finally, the appendices comprise the following:

- Papers arising from this research.
- Outcome of work undertaken in the early period of this research on the fuzzy approach to group technology.
- Comprehensive survey of the artificial intelligence tools.
- Description of initial simulation efforts conducted for the system.
- Brief introduction to the IDEF0 modelling technique and IDEF0 charts used to describe the developed tool.

2 LITERATURE REVIEW

2.1 INTRODUCTION

Globalization of enterprises, and the means by which these enterprises stay competitive are considered, by some, the hottest issues in industry. Globalization includes two aspects, markets and manufacturing. Products are produced for a global market and trade barriers have been torn down, in a way that competitive companies today regard the world as their market. Enterprises are moving in a multinational direction with their production units spread all over the world [1].

Growing global competition, coupled with technology advancements, has greatly complicated the ongoing manufacturing operations of companies. Flexibility, speed, and the ability to serve the customer are the three important aspects of operational capability, to stay competitive in future global markets. These capabilities are achieved by the manufacturer's ability to excel in product development, production operations, customer order processing and delivery, and the development of supply networks [2].

Today's Dilemmas

The leaders of today's manufacturing companies are facing a wide range of challenges [3]. The most important challenges, listed in Table 2-1, are all due to the geographically distributed manufacturing units of next generation manufacturing (NGM) enterprises. Consequently, NGM enterprises are growing in size and complexity. Effective use and management of these worldwide enterprises to ultimately achieve cost and cycle time reduction is critical to excel in manufacturing.

Table 2-1: Challenges of manufacturing companies.

Most Important Challenges	Other Important Challenges
<ul style="list-style-type: none"> • Globalization • Improving Knowledge Management • Cost and Cycle Time Reduction • Improving Supply Chains Globally • Manufacturing at Multiple Locations in Many Countries • Managing the Use of More Part-Time, Temporary, and Contract Workers 	<ul style="list-style-type: none"> • Developing Employee Relationships Based on Performance • Improving Human Resources Management • Improving the Execution of Strategic Plans • Ongoing Measurement and Analysis of Organizational Processes • Developing a Consistent Global Corporate Culture • Outsourcing of Manufacturing • Creating a Learning Organization

Drivers Shaping the Future

Not only do companies have to deal with many dilemmas and challenges, but they also have to prepare for a future that is a juggernaut. The following eight closely coupled forces, shown in Figure 2.1, are shaping the future:

1. **Globalization.** Markets and business competition are undergoing globalization.
2. **Customers.** Customer expectations are rapidly rising. Customized products with high functionality and higher quality need to be delivered faster to customers.
3. **Information.** Information about ideas, technologies, and business ventures is available globally. Internet access is eliminating artificial restrictions on the flow of information. Companies that can convert this information into useful knowledge faster than their competitors will be at an advantage.
4. **Technology changes.** New technologies are coming to maturity and supplanting old ones on an ever-faster cycle.
5. **Technologies access.** Globally, everyone has access to advanced technology, leading to more competitors even in traditionally poorly developed countries.
6. **Upgrading of skills.** All countries are upgrading the skills of their citizens, resulting in more intense competition at all skill levels.

7. **Countries industrialization.** Countries are industrializing more and more, leading to an increase in the environmental responsibilities and resource constraints.
8. **Workforce diversity.** The diversity of workforce has led to an increasing need for coordinating the activities of the people within a manufacturing enterprise.

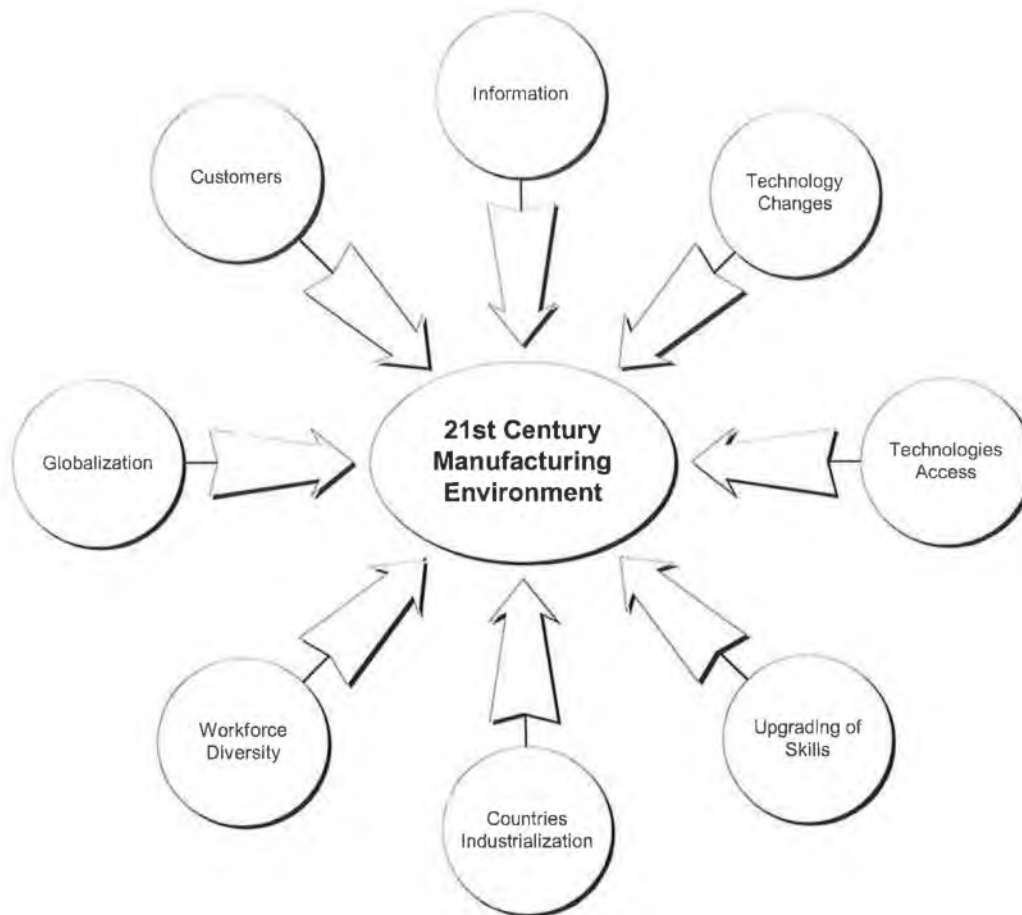


Figure 2.1: Forces shaping the 21st century manufacturing environment.

Business Drives for Manufacturing

There are a number of different drives that have led to the current developments in manufacturing technology and approaches to manufacturing problems. These can essentially be divided into two groups: economic pressures and more general commercial issues.

Economic pressures have led to a particular drive to reduce work in progress and inventory and the size of manufacturing facilities. These stocks represent large amounts of idle capital that could be better employed.

Commercial market pressures have had a number of effects too - customers now demand increased product variety and expect shorter delivery times. These have resulted in the reduction of batch sizes, which are traditionally associated with economic production, placing increased emphasis on the construction of more flexible and programmable facilities that are easier to change from the manufacture of one product to another. Another significant market pressure has been consumer demand for increased quality. One of the major sources of variability in product quality is the human operator, who tires and gets bored. Most programmable automation is introduced to give consistent quality and also significant improvement in quality by operating in constant conditions [4]. In fact, automation of the different processes and functions in manufacturing is a major advance in manufacturing practice, and its impact can hardly be overestimated [5].

To that end, flexible manufacturing systems (FMS) have been developed with the hope that they would provide a means to tackle a threefold challenge: better quality, lower cost, and shorter lead-time. These systems were developed out of the need for producing diverse products in smaller batches while maximizing utilization of machines [6]. An FMS is an integrated production system composed of a set of independent machining centres linked by an automated material handling system for part handling and tool handling. The group of related machines that perform a particular process or step in a larger manufacturing process form a flexible manufacturing cell (FMC) [7].

The nature of flexible manufacturing systems means that the systems design function and sometimes the product design function are also subject to change [8]. Specification of the type and capacity of the material-handling system is one of the major FMS design problems [9] addressed by this work.

In addition to automation, opportunities to apply intelligence, in different functions of manufacturing, are always being pursued. Thus NGM systems can respond and react to unexpected situations without human intervention. Artificial intelligence techniques are presented in Appendix (C).

The purpose of this literature review is to build a solid base of characteristics and current issues related to the automation in semiconductor manufacturing industry and the use of modelling and simulation to study the behaviour of these semiconductor manufacturing systems.

2.2 SEMICONDUCTOR MANUFACTURING

The semiconductor industry is a typical example for manufacturing system applying different advanced manufacturing technologies. Semiconductor manufacturing enterprises compete with each other on a global level and are usually spread all over the world.

2.2.1 History and Background

Although semiconductors have been used in electronics since the early decades of this century, it was the invention of the transistor in 1948 that set the stage for what would become one of the greatest technological advancements in all of history. Since then integrated circuit (IC) technology can be found in an incredible range of products.

The major advantage of ICs is their reduced size and cost. As fabrication technology becomes more advanced, the size of devices decreases. Consequently, more components can be put onto a chip (a small slice of semiconductor on which the circuit is fabricated), thereby reducing cost.

Chips fabrication takes place over the surface of a circular disk called the wafer, and many identical circuits are generated at the same time. More chips can be fabricated at the same time by decreasing device sizes and increasing wafers diameter. Finally, from a manufacturing perspective, batch processing and process automation have helped to reduce the cost of each completed circuit [10].

Moore's Law

Gordon Moore is a founder of Intel and the formulator of Moore's Law. Moore recognized that the reduced cost advantage of integrated circuits continues to increase as the technology evolves toward the production of larger and larger circuit functions on a single semiconductor substrate.

For simple circuits, the cost per component is almost inversely proportional to the number of components, the result of the equivalent piece of semiconductor in the equivalent package containing more components. However, as components are added, there is a tendency to raise the cost per component to compensate for the increased complexity. Thus there is a minimum cost at any given time in the evolution of the technology. However, the minimum is rising rapidly while the entire cost curve is falling [11] (see Figure 2.2).

Figure 2.2 shows Moore's original graph, which confirms the decrease in relative manufacturing cost per component as the number of components per integrated circuit increases.

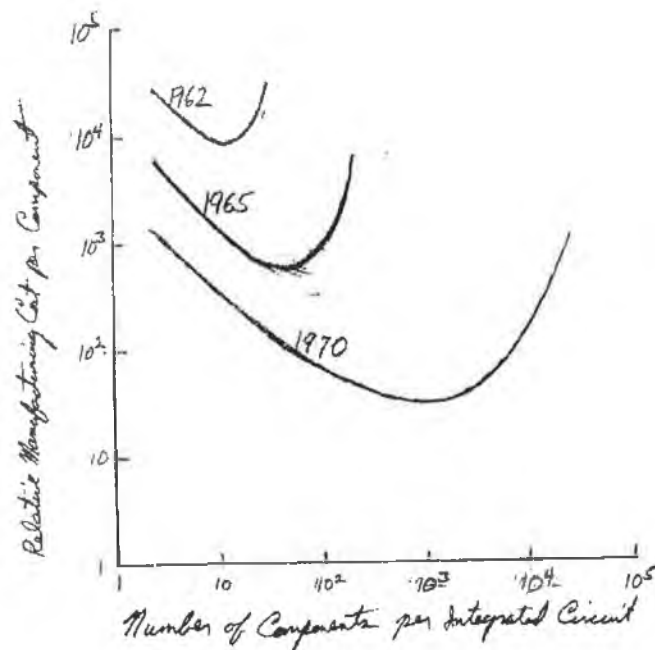


Figure 2.2: Original graph by Gordon Moore [12].

Moore's law is the major driver of the semiconductor industry towards increasing the number of transistors (components) in a chip. Intel has been applying this law effectively over the years and the number of transistors doubles every 18 months on the average, as shown in Figure 2.3.

MOORE'S LAW

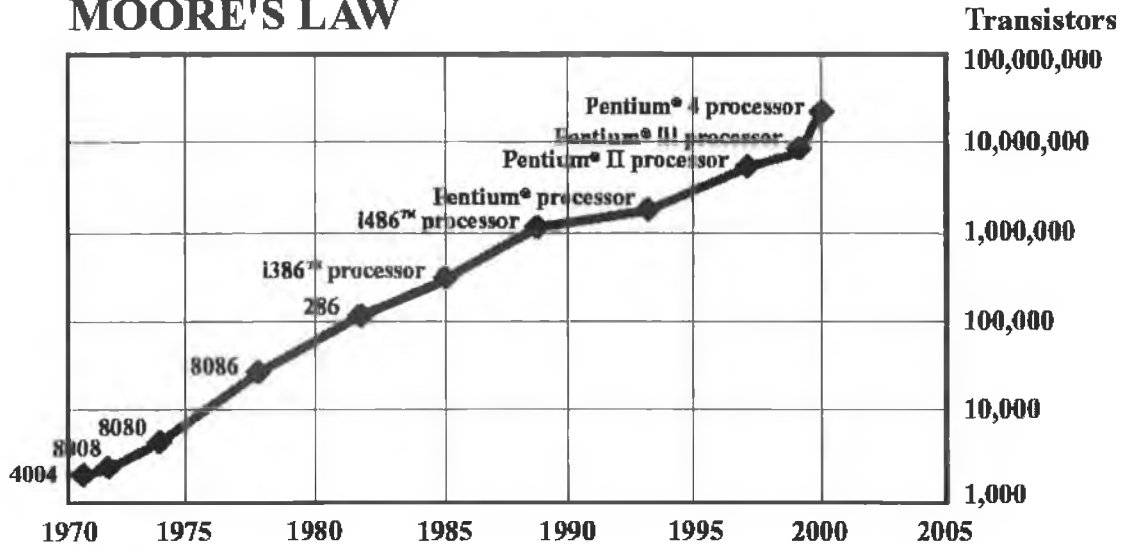


Figure 2.3: Application of Moore's law at Intel [12].

The driving force that has allowed for the above to be consistently valid is the shrinkage of the dimensions of transistors. A new technology generation every 2 years is realised, which effectively leads to:

- Doubling the number of chips per wafer through shrinking the line-width, which is the width of the smallest feature, by a factor of 0.7X.
- Reduction in the transistor cost by half.
- Doubling the microprocessor speed.

Currently, the line width that can be achieved in production is 90 nanometres as shown in Figure 2.4. With new technology generations introduced every 2 years, it is expected by the year 2007 that the line width will be in the order of 45nm. This line width is the minimum achievable [13], which means that the technology barrier for the semiconductor industry will be reached within 4 to 6 years. Therefore, cost effective manufacturing must be pursued to reduce the cost of chips, because further feature size reduction will be unattainable after few years [14].

To conclude, the industry has been on a 2-year cycle for the past decade with no fundamental barriers to extending Moore's Law. Scaling down of the feature size will

continue till the minimum achievable size is reached. At that point of time, the greatest challenge will be to drive costs down through cost effective manufacturing.

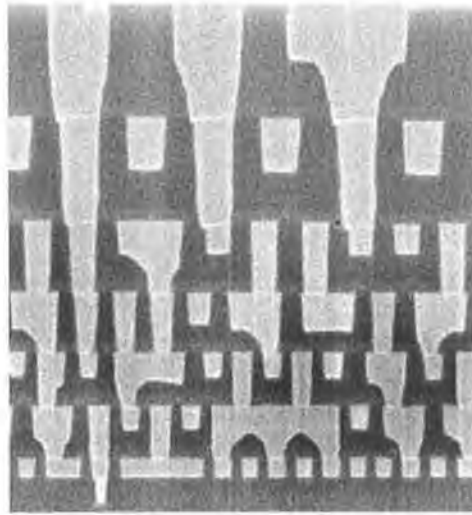


Figure 2.4: 90 nm generation interconnects [15].

Characteristics of the Semiconductor Industry

A very dynamic environment characterises semiconductor-manufacturing industry. The technology to produce integrated circuits is always changing in response to the demand for faster and more powerful devices, and the elapsed time between the development of a new profitable technique and its transfer to actual production is very short. This requires an industry that must be very flexible to adapt to constant changes in operating procedures and challenges to produce new devices [16].

In addition, there is clear recognition that the dynamics of change are progressing so fast that a competitive advantage may quickly shift from company to company or from nation to nation. Hence, the product and market conditions of the semiconductor industry can be described as follows [17]:

1. World market: Supply and demand are not regionally limited. Because consumers worldwide use PCs and entertainment electronics products. Moreover, homogeneous semiconductor products, such as memories, have uniform prices. Thus the semiconductor market is already a world market.

2. **Innovation competition:** The semiconductor market is characterized by a high rate of innovation, short product life cycles, and a low degree of process differentiation. Therefore, there is a strong competition in terms of both price and innovation.
3. **Fast and strongly fluctuating market growth.**
4. **Manufacturing in networks:** Because enormous investment is necessary, the above-mentioned characteristics lead to high investment risk. Semiconductor manufacturers react to this risk by establishing network-wide cooperation.

Continuous productivity improvement is needed in order to maintain the industry's historical 30% per-year per-function reduction in cost.

2.2.2 The Semiconductor Manufacturing Process

Semiconductor manufacturing is among the most complex manufacturing processes. Every semiconductor manufacturing process starts with raw wafers, a thin disc made of silicon or gallium arsenide. Depending on the diameter of the wafer, up to a few hundreds of identical chips can be made on each wafer, building up the electronic circuits layer by layer [18].

The manufacturing process, starting from raw material to assembled chips, is comprised of four main steps as follows [19]:

1. **Wafer Fabrication:** A typical wafer is fabricated through a series of more than 500 process steps. Virtually all of the processing is automated, but there are still many potential sources of failure that lead to defective wafers. Therefore, parametric tests are performed throughout the fabrication process. Parametric tests measure physical and electrical parameters of the wafer such as electrical conductivity and behaviour of individual sample components.
2. **Wafer Probe:** Die-level functional testing (DLFT) typically occurs once the wafers are completely fabricated and the dice are completely formed and functional. This test measures the operational quality of the every individual dice on the wafer. A robotic machine presses electrical probes onto the die contacts

(hence, the term “wafer probe”), input signals are fed to the circuit, and output signals are measured. Functional testing simulates normal and abnormal operating conditions (e.g., high, normal, and low voltage tests). If a die fails the functional test, an ink dot is placed on it (inking decision), so that it will not be packaged later. After inking, wafers are shipped to a separate location for assembly.

3. **Chip Assembly:** To convert the wafers into assembled chips, the wafers are cut into individual dice by a high-precision diamond saw. The resulting chips are mounted into packages, electrical contacts are bonded in place, and then a protective covering is added.
4. **Final Test:** Once the chips are assembled, they are tested again to ensure that the assembly process was successful. Assembly tests usually repeat many of the functional tests that were performed during DLFT. The test results from package testing are used to decide which chips to sell.

Most of the semiconductor manufacturers do not perform all the four steps in one facility. However, the manufacturing process is divided into a front end and a back end, as shown in Figure 2.5, each is carried out in a separate facility. Wafer fabrication is characteristic of the front end of the process, having the transport unit lots of 25 wafers. On the other hand, integrated circuits sawed out of wafers are processed individually at the back end.

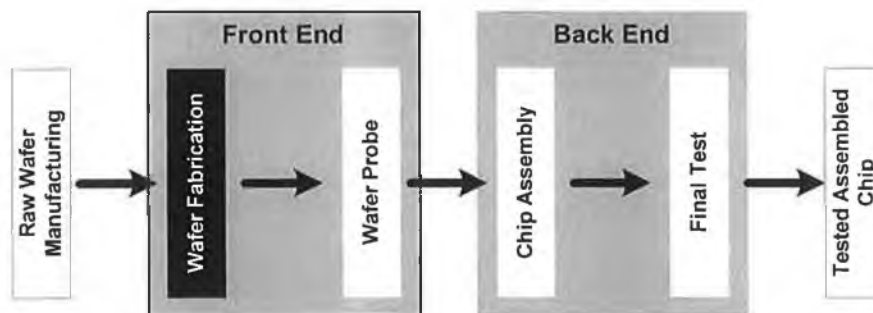


Figure 2.5. Basic steps of semiconductor manufacturing.

A semiconductor manufacturing facility performing the front end of the process is called a “fab”, which refers to the wafer fabrication. The work presented here focuses on the wafer fabrication part of the process, as highlighted in Figure 2.5.

Detailed Description of the Wafer Fabrication Process

The wafer fabrication process is a complex process requiring several steps by special machine tools. Figure 2.6 shows in details the steps of the wafer fabrication, which are as follow:

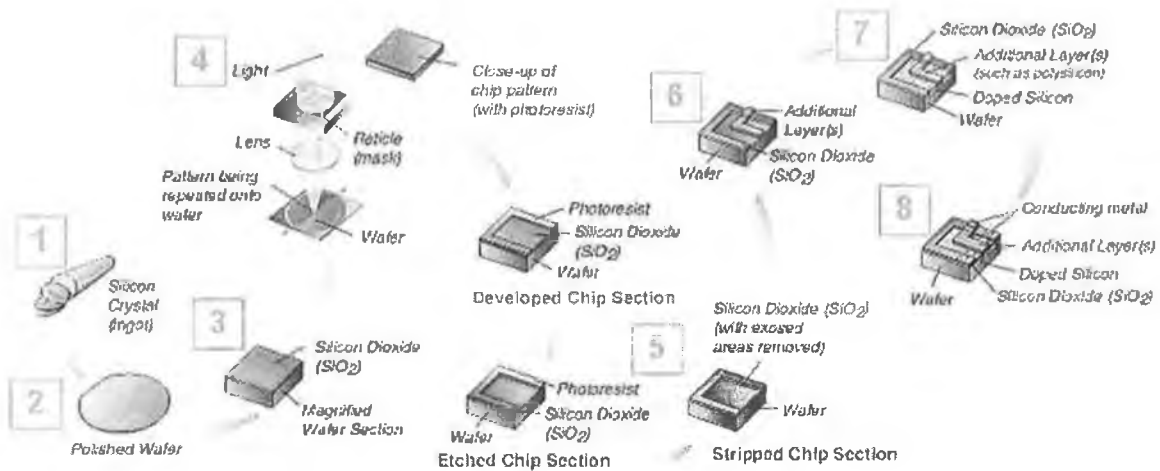


Figure 2.6: Wafer fabrication process [20].

1. To make wafers, polycrystalline silicon, containing elements that can modify its conductivity, is melted. The melted silicon is used to grow silicon crystals (or ingots) that are sliced into wafers.
2. To remove even the tiniest scratches and impurities, one side of each wafer is polished to a mirror-like finish. Chips are built on this surface.
3. As the first step to making a chip, a layer of silicon dioxide (SiO_2) glass is grown or *deposited* on the wafer, which is called dielectric (because it does not conduct electricity). Later channels will be etched or otherwise formed into the dielectric for conducting materials. SiO_2 may also be grown or deposited in later steps in the process as layers of circuits are built into the chips.
4. *Photolithography*, lithography for short, is a process used to create multiple layers of circuit patterns on a chip. First, the wafer is coated with a light-sensitive chemical called photoresist. Then light is shone through a patterned plate called a mask or reticle to expose the resist.

5. Following the lithography process, the wafer goes to the *etch* area where materials are removed in a series of steps. Exposure to light in lithography causes portions of the resist to “harden” (or become resistant to certain chemicals). The “non-hardened” resist is washed away in the development process. Then the material below it, for example SiO₂, is etched away. Finally, the hardened resist is stripped off so that the material underneath forms a three-dimensional pattern on the wafer. The first lithography and etch process will result in a pattern of SiO₂.
6. Through several lithography and etch steps, subsequent layers of various patterned materials are built up on the wafer to form the multiple layers of circuit patterns on a single chip.
7. To control the flow of electricity through a chip, certain areas of the wafer are exposed to chemicals that change its ability to conduct electricity.
 - Atoms from the chemicals, called doping materials, can be *diffused*, or forced into areas of the silicon wafer through chemical exposure and heating.
 - Another doping process, called ion *implantation*, bombards or shoots sections with charged atoms called ions to displace silicon atoms.

Dopant atoms displace some of the wafer’s original silicon atoms to make the wafer either more or less conductive.

8. Finally, all the millions of individual conductive pathways must be connected in order for the chip to function. This includes vertical interconnections between the layers as well as horizontal interconnections across each layer of the chip. The portions of the chip that conduct electricity form the chip’s interconnections. A conducting metal (usually a form of aluminium) is deposited on the entire wafer surface. Unwanted metal is removed using lithography and etch processes to leave microscopically thin lines of metal interconnects.

Although the process might seem to be similar to other manufacturing processes, yet it is characterized by its high technological complexity.

Process Complexity

Manufacturing conditions in the semiconductor industry is far from being conventional. The semiconductor manufacturing process can be characterized as follows:

1. Yield variation and changes in productivity: Compared with other industrial products, process and product yields are relatively low and vary greatly. Learning effects, product reruns (ramp ups), and new processes or technologies, which depend on the quantities sold, cause major forecasting problems.
2. Equipment availability: Semiconductors are manufactured close to the technological limits, requiring highly sensitive manufacturing processes. As a result, equipment availabilities fluctuate greatly due to regular maintenance and adjustments applied to the equipment.
3. Low operation flexibility: In comparison with conventional lot wise production of individual goods according to the shop floor principle, operation flexibility is extremely low. Process sequences are fixed in terms of process technology, and it is impossible to alter the processing sequence. In addition, a load change at alternate machines is hardly possible because of the highly specialized equipment.
4. Low capacity flexibility: Chip factories are intensive in capital or operational costs and for technological reasons cannot be easily extended. Hence area optimization precedes material flow optimization. High investment demands usually causes chip factories to be run 24 hours a day, 365 days a year (i.e., capacity flexibility is almost zero).
5. Shared facilities: Continually high product and process development necessary for innovation competition, as well as high investments demands for clean rooms and manufacturing devices. This can sometimes lead to the use of equipment by both production and development (shared facilities).
6. Data availability and maintenance: The enormous data volume results from the information necessary for each operation, as well as a large number of process steps. Currently, for one wafer fab, transaction volumes are approximately

200,000 movements per day. The constantly high share of new products and short product life cycles requires the constant updating of data.

7. Long lead time and complex material flow: In wafer manufacturing semiconductor products go through a large number of processes. Loop processes (wafers often pass the equipment several times) necessary for reasons of process technology. In combination with the multitude of work steps, lead to a highly complex material flow (see Figure 2.7).

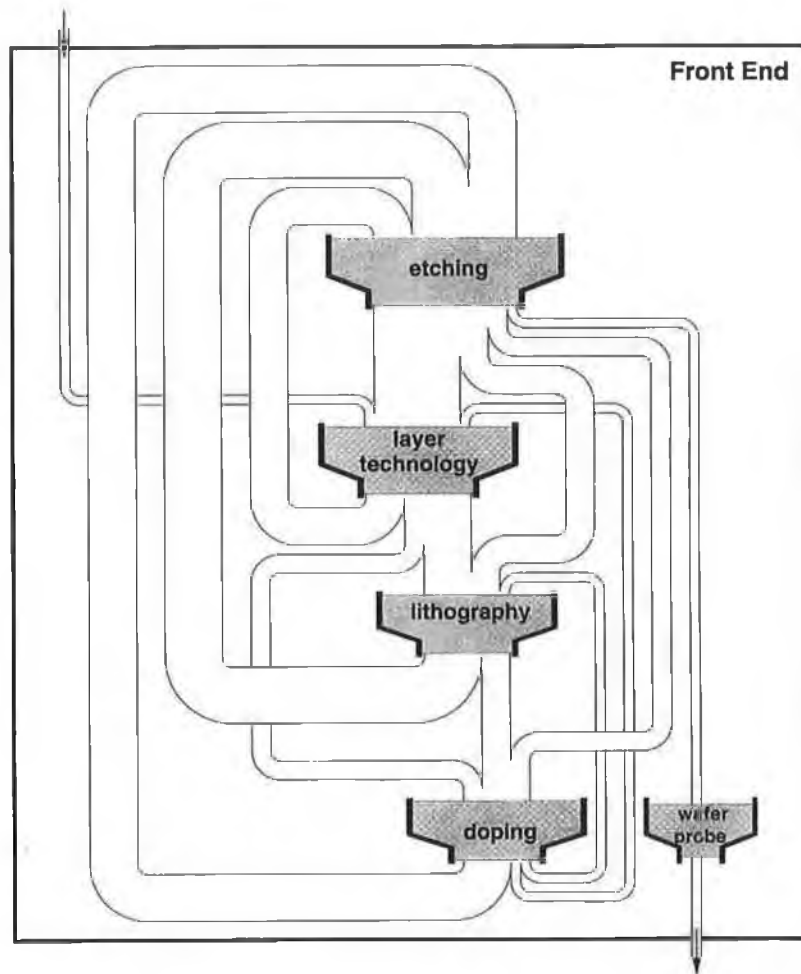


Figure 2.7. Re-entrant material flow [17].

It is quite clear now that new ideas for cost effective manufacturing must be aggressively pursued to maintain industry profitability and growth. These ideas must go beyond simply enlarging the scale of production to achieve efficiency, as this cannot be achieved due to the complexity of process discussed earlier. Developing and adopting a

new manufacturing technology has become the only way for a semiconductor manufacturing facility to remain competitive and to maintain manufacturing cost effectiveness. Therefore, semiconductor manufacturers have been stimulated to increase the wafer diameter size from 200mm to 300mm.

2.2.3 Conversion to 300mm Wafer Size

Beginning in 1993 a small group of people at Intel began thinking seriously about a transition from 200mm manufacturing to the next wafer size. By early 1994, the industry reached consensus that the right wafer size was 300mm. The industry realised these two key points [21]:

1. Output of the 200mm factories (in terms of chips) was low and didn't match the growth in the microprocessor business. This led semiconductor manufacturers to conclude that before the end of the decade, factories would need to be added at the rate of two per year. The complexities of construction, site selection, staffing and management development required to support that growth suggested that it was time to begin thinking seriously about the next wafer size.
2. Many of the costs of manufacturing are proportional to the number of wafers moved, and not to the size of those wafers.

Main Factors Behind the Conversion to the 300mm Technology

Semiconductor manufacturers' decision to convert to 300mm wafers was largely driven by cost. Some of the cost benefits associated with the transfer to 300mm are [22]:

- The primary cost benefit is the ratio of die yield per 300 mm wafer to die yield per 200 mm wafer. The yield improvement ranges from 2.25X to 2.5X.
- The next cost benefit is quantity of output. The output of a 300 mm factory is 1.6-2.2X that of a 200 mm factory producing the same product. The output ratio also reduces the impact of the facility cost on product.
- Finally, A 300 mm factory will cost close to the same as a 200 mm factory but the product output is higher.

A combined 30% reduction in cost is expected with the transition to the 300mm technology.

Previous Wafer Size Conversions

Single large companies have led past wafer size conversions. These companies incurred significant conversion costs while paving the way for other companies [22].

- Intel's groups in Albuquerque, New Mexico led the transition from 100mm to 150mm in 1983 and 1984.
- IBM in Burlington, Vermont led the transition from 150mm to 200mm.

Both companies did not realise any real benefit from leading the previous conversions as the unbalanced investment necessary to support development of each piece of equipment, necessary to equip a full pilot line, did not lead to any practical business advantage for the two companies. This was due to the following reasons [23]:

1. It took many months from the development of equipment prototypes to the delivery of high volume manufacturing tools, thus eliminating any time advantage in converting to a larger wafer size.
2. Any other semiconductor company was able to purchase, with practically no time lag, the same equipment developed by the conversion leader without having invested in a costly equipment development.

Based on these negative experiences, it became clear in 1995 that no company was planning on coming forward and leading the 300 mm wafer size conversion and some form of consortium was the only alternative approach left for the industry in order to make progress in this important cost reduction program [21-23].

The International 300mm Initiative and Technology Transfer

A group of countries formed the International 300mm Initiative (I300I) in 1996 to pave the road for a technology transfer from 200mm to 300mm wafers. I300I began with developing a factory vision for 300mm fabricators. The approach began at the wafer and drove for development of supported/endorsed standards for 300mm wafers, wafer carriers, the equipment interface, and the overall factory layout [24].

It is important to define and use standards for the 300mm wafer fabrication technology for the following reasons [25]:

- Standards provide a mechanism for integrating the different elements of a factory cost-effectively, and in a faster, better and cheaper method.
- Standardization provides opportunities for all firms, having products that meet that standard, to compete in that business.
- There are significant opportunities for 300 mm, which will ultimately lead to significant cost reduction opportunities. This was not the case for the 200mm because many factory elements were not well integrated.

In 2000 the semiconductor industry, through its International SEMATECH website, has definitively confirmed 300mm as the new wafer diameter for mass-production in manufacturing wafer fabrication to implement in the next 2 to 3 years. This wafer diameter is already being implemented in several pilot lines around the world (Intel being the first in the World [26, 27]). I300I provided the foundation necessary to get the semiconductor industry on the road to 300mm from 200mm wafers.

2.2.4 300mm Wafer Fabrication Facility

Production Equipment

In a typical fab, hundreds of various machines exist. Few machines are used for only one dedicated processing step. Most machines are designed to carry out several very similar processing steps during the whole processing sequence and for multiple products. Machines of the same type are usually grouped into work centres (referred to as toolsets) for a number of reasons including reduction of setup time, redundancy in case of breakdowns, efficient utilization of operators, and having backup when maintenance work is done [18].

The 300mm toolsets are much larger in both footprint and height. Every tool has a number of loading ports (for loading and unloading of lots) and a mini-environment to isolate wafers from the open fab air where wafer boxes can be opened and the wafers are accessed. Because wafers are isolated from the open fab environment, a slightly lower class of cleanroom is allowed in the bays.

Material Handling System

The 300mm material handling system is a fully automated system for moving material around the entire facility. Lot deliveries are of two distinct types:

- Delivery of lots from a specific bay to another through an Automated Material Handling System (AMHS), called *interbay* AMHS.
- Delivery of lots from a stocker to a specific toolset within a single bay achieved by another AMHS called *inrabay* AMHS.

It should be noted that delivery from a stocker to a specific toolset was originally performed manually, and the use of the intrabay AMHS is due to the introduction of the new technology.

Layout of the 300mm Fab

In 300mm wafer fabrication facilities, functional area layouts where similar groups of processing equipment (from a function perspective) are located together in a single area will be adopted [28]. These functional areas can typically be lithography, etching, planning, diffusion, thin films, and metrology.

Functional areas are further divided into groups of equipment called bays. A bay is typically a number of machines facing each other, having a storage area at its front end, and a path between the machines and the storage area for delivery of lots.

The storage area is composed of one or more storage slots called stockers, having sizes that vary from one bay to another (depending on the number of lots that could be processed within a specific bay). In that sense, a bay is considered as an FMS within the facility for processing batches of wafers, however, the product variety in a semiconductor manufacturing facility is limited.

The 300mm factory layout is illustrated shown in Figure 2.8, which shows the different functional areas found in a factory, the bays within a functional area, and the internal structure of a bay. Also shown in the figure are the two types of AMHS for moving material around the factory.

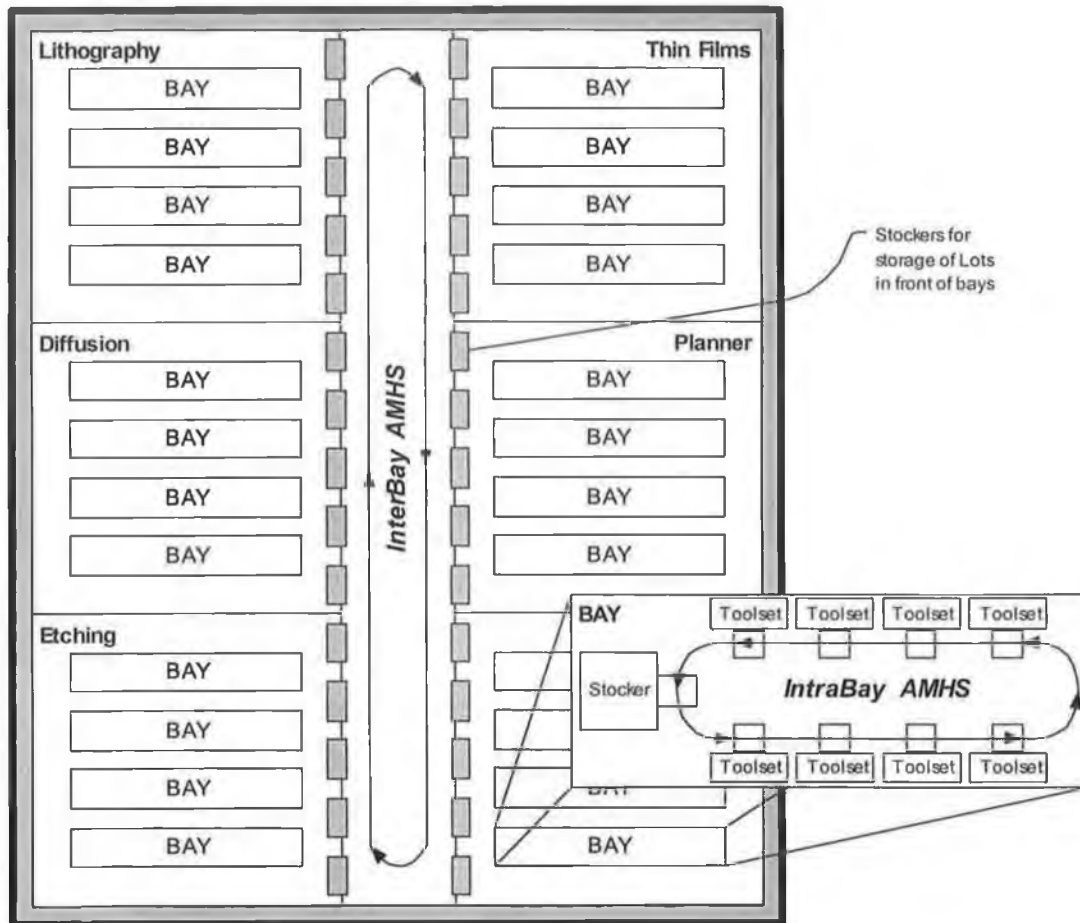


Figure 2.8. Typical 300mm fab layout.

Group technology layout for semiconductor manufacturing facilities promises better cycle times and lower work-in-process inventory levels [29]. Adopting the idea of GT layout, a layer-based approach is used to solve the fab layout problem that groups the machines required to manufacture continuous process layers in the same bay, or area. In fact, the layer-based layout is expected to be a good and feasible alternative especially for a 300mm wafer fab [30].

However, the semiconductor industry is reluctant to adopt other strategies for layouts because the functional layout offers a number of advantages in terms of maintenance and operation of the equipment.

Differences Between 200mm and 300mm Fab

As mentioned before, the 300mm toolsets are much larger in both footprint and height, which enable these toolsets to produce higher number of wafers. A single 300mm tool

consumes more floor space than a 200mm tool would consume. However, a 300mm fab can still hold almost the same number of toolsets as a 200mm fab using the same floor space.

Figure 2.9 shows the main differences between a 200mm and a 300mm fab. The narrower bay width in case of the 300mm fab provides effective use of the valuable clean room floor space. Furthermore, the toolsets are closely packed to each other and operators can access the toolsets from the chase area (back of the tool) for maintenance reasons for example. This explains why a 300mm fab can still hold the same number of toolsets.

Also noted from the figure that the other major difference is the higher level of automation. Most of the manual material handling on the shop floor is eliminated, and the intrabay AMHS is used to transfer material from stockers to machines. However, in case of intrabay AMHS breakdown, manual carts can be used to deliver the lots to the tools.

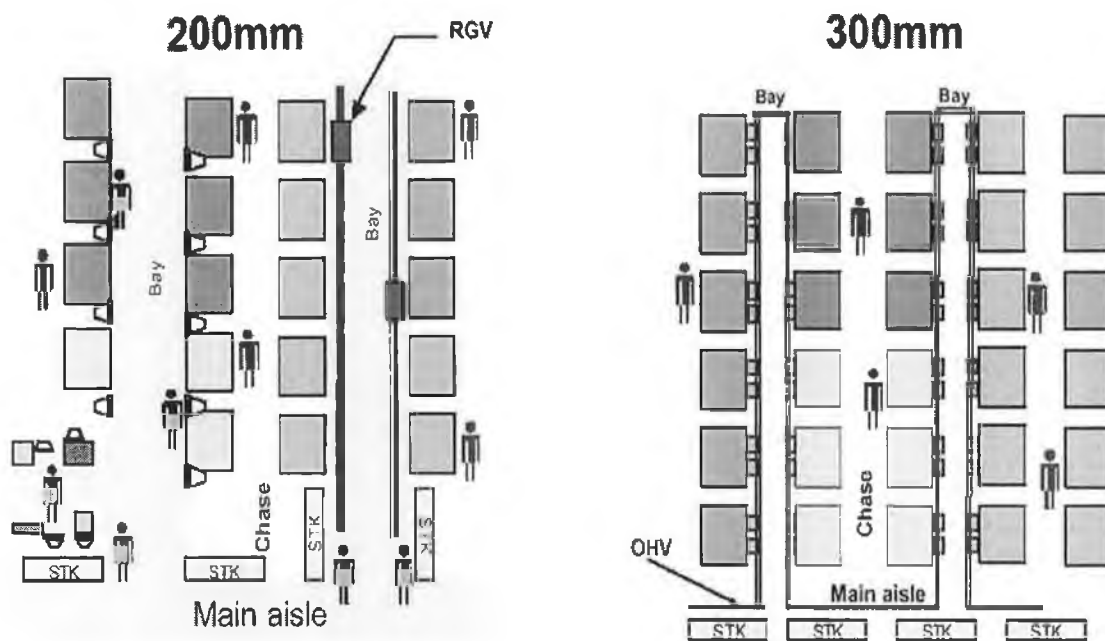


Figure 2.9. Difference between 200mm and 300mm fabs [31].

2.2.5 Key Differences and Major Concerns

This section summarises the key differences between the 300mm and the 200mm technologies, and highlights the benefits associated with each point. Furthermore, the major concerns of the industry imposed by this technology conversion are presented.

- 300mm wafers are handled in sealed boxes that can only be opened by the tool, and thus the risk of contaminating the wafer is reduced. Sealing of boxes coupled with production equipment mini-environment allow for lower level of cleanroom to be used.
- 300mm production equipments are larger in size and can have higher capacity to produce a larger number of wafers. Loading ports and a mini-environment for opening wafer carriers are integrated to the equipment.
- Higher level of automation in 300mm fabs as result of introducing the intrabay AMHS component, which promises faster and more effective lots handling inside the bays. The transportation system is composed of both an interbay and intrabay transport devices.
- Narrower bay width results in better utilisation of the 300mm fab space, and compensates the increase of the production equipment footprint.
- As a result of all the above differences, cost effectiveness of up to 30% reduction in manufacturing cost can be achieved with the 300mm wafer manufacturing technology.

Although the new tools for 300mm appear to be based on better designs than any previous wafer-size generation, yet only time will tell as to how well the tools perform. This totally new equipment set has provided a new set of challenges for engineers. In addition, the new mini-environments and robots add new failure modes in the tools.

Interoperability issues between the materials transferred, the hundreds of equipment loading ports, and the intrabay AMHS presents a major concern for the industry. Consequently, the industry is searching for the appropriate methods for studying the behaviour of these new systems.

2.3 STUDYING THE BEHAVIOUR OF 300MM WAFER FABRICATION FACILITIES

The previous section introduced the 300mm wafer fabrication technology and highlighted the complexity of the semiconductor manufacturing system and the competitive market conditions. Furthermore, these two factors are coupled with the expensive and increasing cost of fabs. Typical fab cost figures are as follows [32]:

- Production equipment: \$1,000,000 - \$7,000,000 per tool.
- Complete AMHS (interbay and intrabay): \$50,000,000 - \$100,000,000.
- Total 300mm fab cost: \$1.3 - \$2.0 billion.

However, it is not uncommon to have a single 25-wafer carrier with a combined retail value of over \$1,000,000. Therefore, relatively small improvements (or problems) result in huge cost impacts. This necessitates an exact understanding of how to design and operate these facilities efficiently.

2.3.1 General Approaches of System Studying

The different methods of analysing and understanding a system under study are as follows [33, 34]:

Experiment with Actual System vs. Experiment with a Model of System

The model and system must be distinguished clearly. A model is a representation of a group of objects or ideas in some form other than the entity itself, and a system is a group or collection of interrelated elements that cooperate to achieve a stated objective [35].

If it is possible (and cost-effective) to alter the system physically and then let it operate under the new conditions, it is probably desirable to do so. However, it is infeasible to do this in semiconductor manufacturing facilities, because such an experiment would often be too costly to the system.

The “system” might not even exist, such as the intrabay AMHS studied in this research; nevertheless there is a need to study it in its various proposed alternative configurations to see how it should be built in the first place. For these reasons, it is necessary to build a model as a representation of the semiconductor manufacturing system and study it as a replacement for the actual system.

Physical Model vs. Mathematical Model

A model can be a physical, mathematical, or otherwise logical representation of a system, entity, phenomenon, or process [36]. Physical models (also called iconic models) are not typical of the kinds of models that are usually of interest in systems analysis. Occasionally, however, it has been found useful to build physical models to study engineering or management systems; examples include tabletop scale models of material-handling systems.

The vast majority of models built for manufacturing systems are mathematical. This model represents a system in terms of logical and quantitative relationships. These relationships are then manipulated and changed to see how the model reacts, and thus how the system would operate under new conditions.

Analytical Solution vs. Simulation

Once a mathematical model is built, it must then be examined to see how it can be used to answer the questions of interest about the system it is supposed to represent.

If the model is simple enough, it may be possible to work with its relationships and quantities to get an exact, analytical solution. However, semiconductor manufacturing systems are highly complex, so that valid mathematical models of them are themselves complex, precluding any possibility of an analytical solution. Therefore, the model must be studied by means of simulation, i.e., numerically exercising the model for the inputs in question to see how they affect the output measures of performance.

Figure 2.10 shows the general ways in which a system might be studied. The figure also shows that the selected approach to study the semiconductor manufacturing facility is through modelling and simulation.

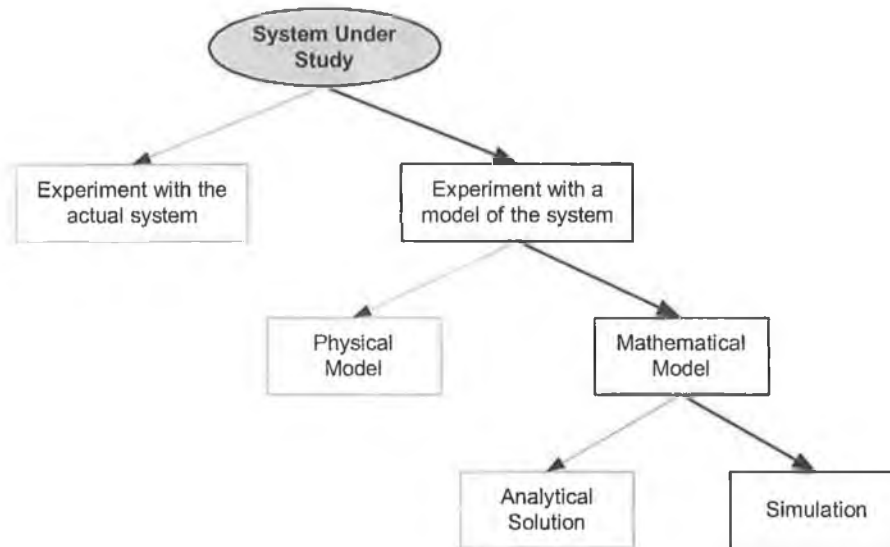


Figure 2.10: Different methods of approaching a system under study.

2.3.2 Systems Modelling

System modelling is a logical representation of the important parts of the system. Modelling is the process of producing a model. A model is a representation of the construction and working of some system of interest, which is similar to but simpler than the system it represents [37]. Models of manufacturing systems are generally represented as a computer program, which represents the most important system components and the way in which they interact [38].

The usefulness of models has been demonstrated in describing, designing, and analyzing manufacturing systems. The advantages of a model are three-fold [39].

1. A model is easier and cheaper to operate than the real system.
2. Irrelevant details, with respect to the study objective, can be ignored.
3. Easier analysis and decision-making are realised with the use of models.

However, because of the complex large-scale nature of manufacturing systems; model building is a complex process and involves both inductive and deductive reasoning. In addition, there are no established, published principles for modelling, as modelling is considered an art and a creative activity [40].

Pidd [41] presented what appears to be a number of modelling principles, nevertheless these are only guidelines that can be considered in a modelling activity and are not rules that must be followed, because modelling activities differ from one to the other.

2.3.3 Modelling and Simulation

Simulation is the imitation of the operation of a real-world process or system over time [42]. When modelling systems, their dynamic behaviour is usually of interest. Typically, the model is placed at some initial point in phase space and watch it mark out a path through the future. Simulation embodies this concept because it involves running a model of a system by starting with the system status at an initial point in time and evaluating the variables in a model over time to ascertain the dynamic performance of the model of the system [40]. In addition, one of the model validation techniques is simulating the model under known input conditions and comparing model output with system output [37].

Manufacturing simulation has been one of the primary application areas of simulation technology. It has been widely used to improve and validate the designs of a wide range of manufacturing systems. Typically, manufacturing simulation models are usually used either to predict system performance or to compare two or more system designs or scenarios [43].

Different Types of Simulation Model

Static vs. Dynamic Simulation Models

A static simulation model is a representation of a system at a particular time, or one that may be used to represent a system in which time simply plays no role; examples of static simulations are Monte Carlo models. On the other hand, a dynamic simulation model represents a system as it evolves over time, such as a conveyor system in a factory.

Deterministic vs. Stochastic Simulation Models

If a simulation model does not contain any probabilistic (i.e., random) components, it is called deterministic. In deterministic models, the output is “determined” once the set of

input quantities and relationships in the model have been specified. Simulations carried out using spreadsheets are a typical example of this type.

Many systems, however, must be modelled as having at least some random input components, and these give rise to stochastic simulation models. Most queuing and inventory systems are modelled stochastically. Stochastic simulation models produce output that is itself random, and must therefore be treated as only an estimate of the true characteristics of the model; this is one of the main disadvantages of simulation.

Continuous vs. Discrete Simulation Models

Loosely speaking, discrete and continuous simulation models are defined analogously to the way discrete and continuous systems were defined above. It should be mentioned that a discrete model is not always used to model a discrete system and vice versa. The decision, whether to use a discrete or a continuous model for a particular system, depends on the specific objectives of the study.

For example, a model of vehicle flow on an AMHS loop would be discrete if the characteristics and movement of individual vehicles are important. Alternatively, if the vehicles can be treated “in the aggregate,” the flow of vehicles can be described by differential equations in a continuous model.

Discrete-event simulation is one of the most widely used methods to study, analyse, design, and improve manufacturing systems [44]; hence, it is used in this research to study and analyse the performance of the semiconductor manufacturing system.

Structural Components of Discrete Event Simulation

A discrete-event simulation is one in which the state of a model changes at only a discrete, but possibly random, set of simulated time points. During a simulation run an internally managed stored data value tracks the passage of simulated time, which advances in discrete steps (typically of unequal size) during the run. After all possible actions have been taken at a given simulated time; the time is advanced to the start of the next earliest event [45].

Time is advanced using a time advance mechanism, which is done by ordering all known events into a chronological order of occurrence, and letting the simulation time

advance from one event to the next in the ordered sequence. The state of the model between events remains unchanged, thus skipping from one event to the next, without considering the time in between those two events, loses no information [46].

The execution of a run thus takes the form of a two-phase loop: “carry out all possible actions at the current simulated time,” followed by “advance the simulated time,” repeated over and over again until a run-ending condition is reached.

A number of modelling concepts have to be defined so that a discrete-event simulation model can be well understood [42]. These are briefly introduced in the following sections.

Entities and Attributes

A system is composed of objects or *entities*, which can cause changes in the state of the simulation. The collection of entities that compose a system rely primarily of the objective of a particular study. Consequently, different study objectives for the same system result in including different entities in the model, and the collection of entities that compose a model for one study might be only a subset of the overall model for another [33].

For example, in the study of the semiconductor manufacturing system, if one wants to study the impact of the different breakdown rates of the toolsets on the overall output of the system, then the system can be defined as that portion that contains the toolsets. If, on the other hand, the time to repair the toolsets and different policies of maintenance are considered in the study as factors that would affect the breakdown rate and factory output, then the system is expanded in an obvious way; entities like labour and equipment needed to repair the toolsets are added to the system.

Entities are defined using *attributes*, which represent the properties of an entity. Attributes are characteristics of a given entity that are unique to that entity. Attributes are critical to the understanding of the performance and function of entities in the simulation.

Semiconductor fabrication facilities contain thousands of simultaneous production lots, resulting in a simulation model that may also contain thousands of entity records, each entity having a number of attributes. That is why the speed and space complexity of these simulations must be at least on the order of some polynomial of the number of lots in the factory. That is why simulation models of such a system have been shown to increase exponentially in their execution time [47].

Activities and Delays

An *activity* represents a time period of a specified length. For example, vehicles are one of the AMHS components (entities) each vehicle has a certain speed, capacity, breakdown rate... etc (attributes); pickup, drop-off are the activities carried out by the vehicles.

A *delay* is an indefinite duration that is caused by some combination of system conditions. When an entity joins a queue for a resource, the time that it will remain in the queue may be unknown initially since that time, which is the delay, may depend on other events that may occur.

System State Variables

The *state* of the system is defined to be that collection of variables needed to describe the system at any point of time; again, this is relative to the objective of the study. In the semiconductor manufacturing system example, the state variables are the different status of the toolsets (idle, processing, or down), and/or vehicles (empty, full, or down).

Events

An *event* is defined as an instantaneous occurrence that may change the state of the system. Events can be *endogenous*, occurring within the system, or *exogenous*, occurring in the environment affecting the system. In semiconductor manufacturing, an endogenous event can be start or end of processing or breakdown of toolsets, start or end of delivery of lots by the AMHS...etc; an exogenous event can be lots arrival to the system, or even customer orders.

Resources

A *resource* is a static entity that provides service to dynamic entities. The resource can serve one or more than one dynamic entity at the same time, i.e., operate as a parallel server. A dynamic entity can request one or more units of a resource. If denied, the requesting entity joins a queue, or takes some other action (i.e., diverted to another resource, ejected from the system). If permitted to capture the resource, the entity performs an activity, then releases the resource. There are many possible states of a resource. Minimally, these states are idle and busy. But other possibilities exist including failed, blocked, or starved.

List Processing

Entities are managed by allocating them to resources that provide service, by attaching them to event notices thereby suspending entities activity into the future, or by placing them into an ordered list. *Lists* are used to represent queues. Lists are often processed according to FIFO (first-in-first-out), but there are many other possibilities.

Steps of a Simulation Study

General methodologies for carrying out a simulation study are found in several sources [33-35, 37]. The specific methodology selected and presented here, while containing many points from these, is mainly based on a “client – system analyst” relationship [48, 49], which suits the nature of this work.

1. Problem Formulation

Every simulation study begins with a statement of the problem. Problem statement is provided by a client or prepared by the simulation analyst. The simulation analyst must ensure that the problem is clearly understood.

Furthermore, the client must understand and agree with the formulation of the problem and the set of assumptions upon which it was based. It is possible that the problem may need to be reformulated as the simulation study progresses. Finally, all possible alternative designs for the real system must be known by the system-analyst.

2. Setting of Objectives and Overall Project Plan

The objectives indicate the questions that are to be answered by the simulation study. The project plan should include a statement of the various scenarios that will be investigated.

By defining the objective the bounds of the problem can be defined in terms of scope and level of detail required to analyse it. Performance measures for evaluating and comparing different system configurations are identified. Sometimes it is necessary to know at this stage the time frame of the model (how frequently it will be used and for how long), and the targeted end-user.

3. Model Conceptualisation

The conceptual model, a series of mathematical (formulas, equations...etc.) and logical relationships concerning the components and the structure of the system (block diagrams, flowcharts...etc.), abstracts the real world system under investigation. It is recommended that modelling begin simply and that the model grow until a model of appropriate complexity has been developed. Then, special features can be added (animation, user interface...etc.) and final refinement of model is done.

For example, consider the model of a manufacturing and material handling system. The basic model with the arrivals, queues, and servers is constructed. Next, the material handling capabilities are added. Finally, animation for showing material flow can be added.

4. Data Collection

Shortly after the proposal is accepted, schedule of data requirements should be submitted to the client. In the best of circumstances, the data needed in the format required can be submitted to the simulation analyst in an electronic format. However, the required data can be either in available or the data are delivered in different formats than anticipated. Data-related problems should never be ignored or underestimated, as they often cause significant delays in model development [50]. Usually the simulation analyst constructs the model while the data collection is progressing.

The required data format must be accurately defined, to facilitate introducing the data to the developed model. Furthermore, the probability distributions for any random variables must be defined at this stage. Finally, data on performance of the real system (if the system exists), which can be used for validation purposes, must be collected.

5. Model Translation

The conceptual model constructed in step3 is coded into a computer-recognizable form, an operational model.

6. Verified?

Verification concerns the operational model, which makes sure that the model is operating as intended by the system-analyst. It is highly advisable that verification takes place as a continuing process and not to wait until the entire model is completed to begin the verification process.

7. Validated?

Validation is the determination that the conceptual model is an accurate representation of the real system, and that the model can be substituted for the real system for the purposes of experimentation. An ideal way to validate the model is to compare its output to that of the real system.

Unfortunately, there is not always a real system (such as in the design of a new system). In that case the client confirms that the model is operating in a manner similar to the real system. In that case the client checks the face validity of the model. This is achieved through animation checks, sensitivity of the model to the input data, and whether or not the model results are reasonable.

8. Experimental Design

For each scenario that is to be simulated, decisions need to be made concerning the length of the simulation run, the number of runs, and the manner of initialization (discussed later in more detail).

9. Production Runs and Analysis

Production runs, and their subsequent analysis, are used to estimate measures of performance for the scenarios that are being simulated.

10. More Runs?

Based on the analysis of runs that have been completed, the simulation analyst determines if additional runs are needed and if any additional scenarios need to be simulated.

11. Documentation and Reporting

Documentation is necessary for numerous reasons. If the simulation model is going to be used again by the same or different analysts, it may be necessary to understand how the simulation model operates. This will stimulate confidence in the simulation model so that the client can make decisions based on the analysis. In addition, if the model is to be modified, this can be greatly facilitated by adequate documentation. The result of all the analysis should be reported clearly and concisely. This will enable the client to review the final formulation, the alternatives that were addressed, the criterion by which the alternative systems were compared, the results of the experiments, and analyst recommendations, if any.

12. Implementation

Implementation of the model is the last stage in a simulation project. A simulation study whose results are never implemented is most likely a failure. However, the results of a simulation study can be understanding of a system, performance comparison of two systems, or the number of some entity required for efficient system performance. Thus, a simulation study is successful when the results of the study, whatever that may be, is used by the client.

The steps of a simulation study are shown in Figure 2.11. It must be noted that although the figure shows the steps to be carried independently, most of the time several steps are performed concurrently (e.g. model conceptualisation, data collection, model translation, verification, and validation).

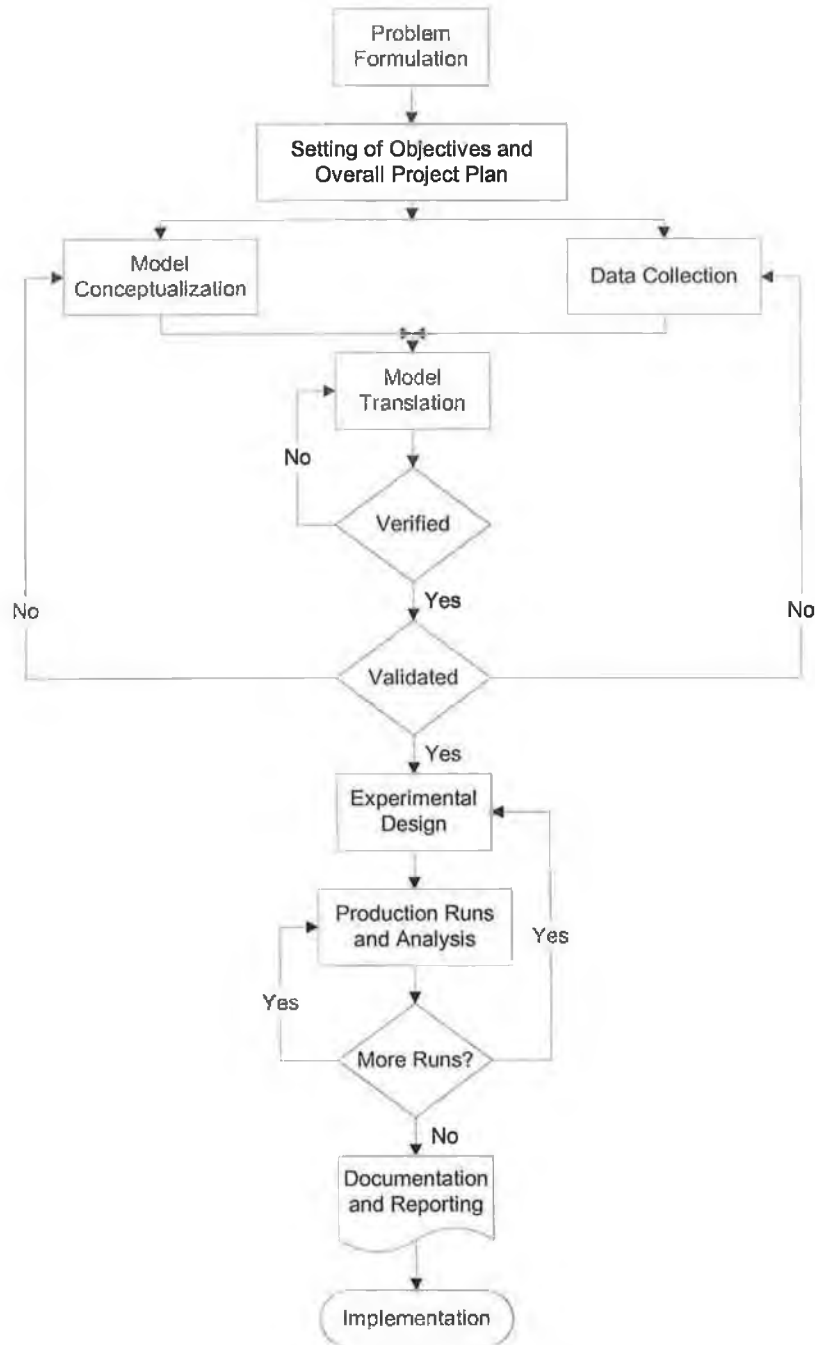


Figure 2.11: Steps in a simulation study [49].

Design and Analysis of Simulation Experiments

Simulation input is usually of a random nature, thus a simulation run produces a statistical estimate of the (true) performance measure not the measure itself [51]. In order for an estimate to be statistically precise (have a small variance) and free of bias, the analyst must specify for each system design of interest appropriate choices for the following:

Length of Each Simulation Run

Length of a simulation run depends on whether the simulation model is built for a terminating or a non-terminating system [49].

TERMINATING SYSTEM

By definition, a terminating system is one that has a fixed starting condition and an event definition that marks the end of the simulation. The duration of the simulation can be fixed by specifying a finite length of time to simulate or by limiting the number of entities created or disposed. Therefore, the length of simulation run in this case is known.

NON-TERMINATING SYSTEM

In a non-terminating system, the duration is not finite, such as semiconductor facilities running 24 hours a day, 7 days a week. The objective in simulating a non-terminating system is to understand its steady state behaviour.

Length of the Warm-Up Period

Most simulations of manufacturing systems generally begin with the system in an empty and idle state. This results in the output data from the beginning of the simulation run not being representative of the desired steady state behaviour of the system.

To study the steady state behaviour accurately, the effects of the initial conditions or transient phase must be removed from the simulation results. Therefore, simulations are often run for a certain amount of time, the warm-up period, before the output data are actually used to estimate the desired performance measure.

Number of Replications

Replications are independent simulation runs using different random numbers for each run. Number of replications required for analysing the performance of a system depends on whether a single system is analysed or two systems are compared to each other.

REPLICATIONS NEEDED FOR A SINGLE SYSTEM

Generally, when analysing the performance of a single system, it is recommended to make at least three to five replications for each system design, and using the average of the

estimated performance measures from the individual runs as the overall estimate of the performance measure [51].

A more accurate method of calculating the required number of replications is by using the t distribution, which is based on the standard deviation of the performance measure (calculated from test runs of the model) and level of confidence required [49].

REPLICATIONS NEEDED FOR COMPARING TWO SYSTEMS

When comparing two systems together it is usually recommended to assign the same random number seeds to all data points, using a technique called common random numbers (CRN). The intuition behind CRN is that a fairer comparison among systems is achieved if both systems are subjected to the same experimental conditions, and specifically the same source of randomness [48, 52].

CRN is a variance-reduction technique (VRT), which reduces the variance of an output random variable of interest (the performance measure) without disturbing its expectation. This results in greater precision of the simulation output, and thus smaller confidence intervals can be achieved with much fewer replications [33].

Reproducing the exact same sequence of random numbers requires the following capabilities [53]:

- **Reproducibility.** Starting at the same initial “seed” can reproduce a sequence of random numbers.
- **Synchronisation.** Forcing the random numbers to be used for the same purpose in each system.

Advantages and Disadvantages of Simulation

Listing the advantages and disadvantages of simulation concludes the modelling and simulation section of this literature review. These are found in most literature [33, 37, 48, 49], however only the ones of interest are listed here.

Advantages of Simulation

1. Most complex, real-world systems with stochastic elements cannot be accurately described by mathematical models that can be evaluated analytically. Thus, simulation is often the only type of investigation possible.
2. Simulation allows the estimation of performance of existing and non-existing systems.
3. New hardware designs, physical layouts, transportation systems...etc. can be tested.
4. Time can be compressed or expanded allowing for speed up or slow down of the phenomena under investigation.
5. Insight can be obtained about the interaction of variables and the importance of variables to the performance of the system.
6. A simulation study can help in understanding how the system operates rather than how individuals think the system operates.
7. "What-if" questions can be answered, which is particularly useful in the design of new systems.
8. Alternative proposed system designs can be compared via simulation to see which best meets a specified requirement.

Disadvantages of Simulation

1. Each run of a stochastic simulation model produces only estimates of a model's true characteristic for a particular set of stochastic input parameters.
2. Simulation models are often expensive and time-consuming to develop.
3. Simulation can sometimes be used when an analytical solution is possible, or even preferable.

2.3.4 Issues in Modelling and Simulation

Due to the increasing complexity of manufacturing systems and the continuous development of new technologies, several issues have emerged related to modelling and simulation studies. Consequently, simulation studies have acquired a number of new characteristics over the past years. Among these new characteristics, the following is of particular interest to this research [54]:

- Simulation studies normally take a couple of months to a year, due to the complexity of the environment, the number of processes, and the required detail.
- Simulation studies have a repetitive characteristic and are used several times in decision-making processes not one-shot-models.

These two characteristics are directly related to the issues of model complexity and reusability of models, presented next.

Complexity of Simulation Models

In the simulation of manufacturing systems, simplification is often not attempted. A common practice is to build and use the most complex model that can be built in the time available, based the misguided assumption that more detailed and complex models are necessarily more accurate and therefore better. However, projects often fail because too much detail is included in the model [55].

Chwif *et al.* [56] presented an excellent paper on simulation model complexity in which they confirmed the lack of a general single definition of model complexity. However, they presented a number of definitions that contribute to the overall characterisation of a model complexity.

Level of Understanding

This definition relates the model complexity with the cognitive aspect. The difficulty of understanding a model directly increases the complexity of that model. On the other hand, model simplicity can be achieved with higher levels of “transparency” related to understanding of the model.

Level of Detail

The level of detail of a model determines the amount of information that is contained in that model. Again, model complexity increases with the increase of the level of detail. It should be noted that both the level of abstraction and level of detail represent the quantity of information contained in a model. However, the higher the level of abstraction the smaller the amount of information included in a model [57].

Scope of Model

A model's scope (or perspective [57]) refers to the mechanism that determines what set of model information is relevant to a given modelling goal and, therefore, needs to become an explicit part of a model. In other words, model's scope defines the boundaries of the system under study.

A model level of detail and scope are usually used in connection to each other. Although there is a strong relationship between them, yet, each one is a different component of complexity. Thus, abstractions determine detail (quantity of information) and scope determines relevance. As an example when simulating manufacturing systems, then either the entire facility (large scope) or just one work centre (small scope) can be modelled. In both cases, the work centre can be modelled with an aggregated processing time or it can be modelled with sub-processing times, breakdowns, shift patterns, material handling equipment, etc., increasing the amount of information. In general large scope models will have less details than small scope models. Often, the aggregate information used in large scope models is taken from related small scope models.

Size of Model

The number of parts and elements found in a model represents its size, which is directly related to the model complexity. The more parts and elements of a model, and the more amount of information used to define those parts and elements, the larger the model's size, and the more complex it is. Number of elements increase as the scope of the model increase, and the amount of information defining each element increase by increasing the level of detail.

Establishing the size of the model is an important step in simulation modelling and is often done early in the model development life cycle. A good modelling practice is to select the smallest size that will adequately address modelling objectives.

Reusable Simulation Models

In a rapid growth high technology industry it is infeasible to continually employ sufficient staff to create unique models of all of the possible scenarios that require examination.

Typically, the model cannot be completed before the project requirements have been modified. Moreover, if simulation does not give credible answers in the time frame the customer wants, a loss of interest in using simulation can occur, as the model won't be of any value to the customer (see Figure 2.12). Therefore, the simulation cycle time must be reduced such that it provides reasonable results to the customer.

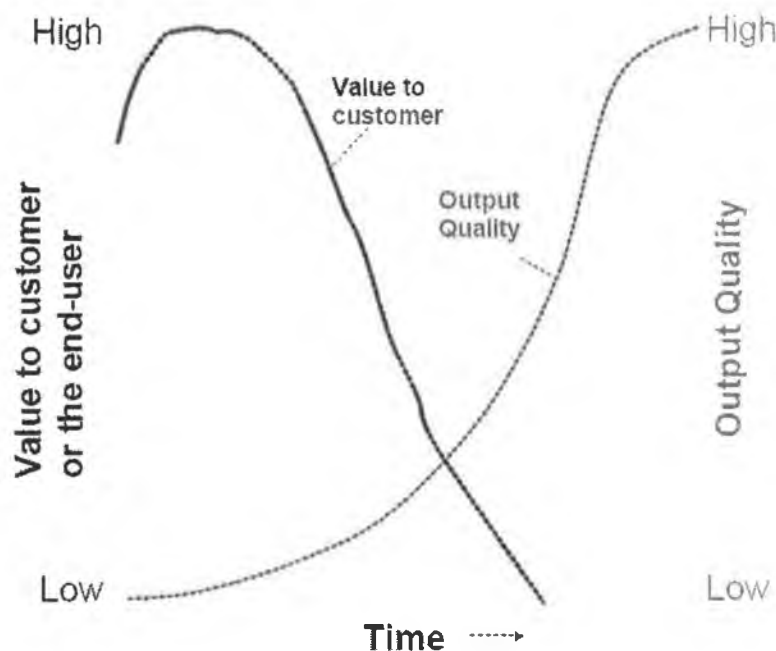


Figure 2.12: Simulation takes too long to produce results [58].

An alternative to unique model creation is the reuse of an existing generic model. Generic models are similar to a group of software tools called simulators; software packages that contain a pre-programmed model. Investigation has indicated that a special purpose reusable generic model, designed to address the set of issues faced by a

specific commercial entity, is efficient and necessary for fast model turnaround. If correctly developed, the generic model can be reused, thereby reducing model-building time as well as increasing simulation accuracy. The generic/specific concept investigates the ability to define models that would be applicable to a wide range of situations, yet are only detailed enough to accurately represent the system being studied. The generic/specific concept is of particular value to situations where the same basic model type is used to evaluate many diverse yet fundamentally identical systems [59].

Through model reuse, the modeller has several apparent opportunities to save time on building the model. These can include the following [60]:

- **Reuse of basic modelling components.** The modeller reuses the basic modelling components (workstations, resources, etc.) that come with the commercial-off-the-shelf modelling package.
- **Reuse of subsystem models.** The modeller has models of various “generic” factory parts that he or she has previously developed or has access to through a model library (a conveyor subsystem is often a good example of this) that can be adapted and used in a new model representing the factory. Alternatively, the factory owner might have models of factory parts that were previously developed and makes these available to the modeller.
- **Reuse of a similar model.** The modeller has previously developed a model that has similar features to the factory being studied. This model is adapted appropriately.

Figure 2.13 shows a spectrum of different types of model reuse [61]. It shows types (or levels) of model reuse with two different axes. The first, frequency, indicates that reuse is much more frequent at the right-hand end of the spectrum, where all modellers engage in code reuse provided with a simulator. The second axis, complexity, runs in the vertical direction, making the point that code reuse is relatively easy, whereas successful reuse of entire simulation models can be very difficult indeed.

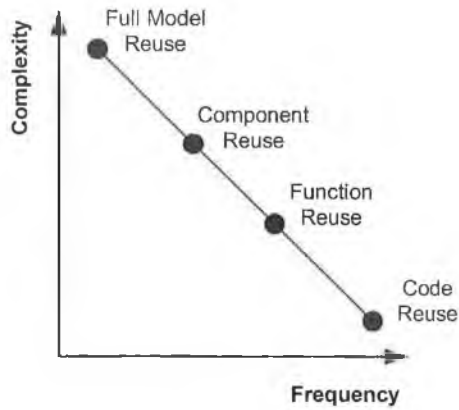


Figure 2.13: Model reuse spectrum.

It should be noted that if the full model is reused in application domains with widely ranging objectives, making it what is called an “all-purpose” model, the complexity of such will increase dramatically. Hence, such type of reuse is impossible to achieve [62].

Function reuse refers to the reuse of functions provided with a simulator like random number generators. Furthermore, component reuse is of particular to the research.

Component-Based Modelling

The component-oriented worldview sees a simulation as being composed of a set of components. It takes a divide-and-conquer approach in which the whole simulation is partitioned into a number of smaller simulation tasks, which are modelled individually as components [41].

The immediate benefit of doing so is that the complexity is significantly reduced. Each component is now a smaller task whose internal logic is much simpler than that the whole simulation. With this approach, reusability can also be achieved if the components are designed in such a way that context-relevant information is not embedded into the code of the components [63].

In a panel discussion exploring the future of simulation [64], Pegden quoted the following “...one of the promising ideas is the concept of having pre-built models or model components that can be plugged together to form a model of our system. The idea is that we simply select these components from a library and use them directly... build each model component once, verify its operation, and then make it available in a

library to be used in many different applications”. Component libraries are being developed to allow rapid prototyping of new designs, which provides a means of creating a rich depiction of systems from reusable models rather than starting from scratch each time [65].

Component-based modelling allows users to quickly and efficiently create high fidelity simulation models by linking independent model objects. The result of linking these models is a model network that can be used to evaluate the aggregate performance of the system as well as investigate the interactions and performance of the individual component models. Model users should be able to assemble the model component parts in a plug-in manner, thus minimizing the time, cost and expertise required to construct comprehensive models within the context of their organization [66].

Definition of Component

The term component has been used in computer science often without elaboration [63].

A component is defined as an object with an interface that enables it to be combined with other objects.

The interface, either explicit or implicit, prescribes in what kind of interaction the components would be involved with other components. All interactions between components must be reflected in the interface. Components refer to individual models, such as work cells, machine tools, or material handling subsystems, which represent portions of a larger system aggregate.

Modularisation as a Prerequisite to Components

Modularity is based on locality and encapsulation [67]. Locality is the notion that all information relevant to a design decision should be kept in one place, i.e. within a module. Encapsulation, or information hiding, separates internal, hidden aspects of a module from the external.

A modular model should satisfy two conditions [68]:

1. The model or component must not directly access the state of any other model or component.

2. The model must have recognized input and output ports through which all interaction with the exterior is mediated.

Given the properties of a modular model, it should be clear that modularity ought to make a model easier to understand and safer to modify. This is because changes in any design decisions are local and are independent of the rest of the model, since they are encapsulated within a single module.

A further advantage is that this modularity supports the re-use of model components, since modular models are defined with no direct reference to the state of their potential co components.

This improves the likelihood that modular models may be built, at least in part, from existing components. It also provides an attractive way of introducing hierarchical components into simulation models.

Benefits of Component-Based Modelling

Two important benefits can be derived from component-based modelling; reduction in model complexity, and reusability of components to develop models.

- A reduction in the complexity of the model, from a construction perspective, is attained because the model developer is able to reuse appropriate components without having to re-invent them [67].
- Ability of simulation novices (who have knowledge of the problem domain, but only limited knowledge of simulation) to perform the simulation study. If it proves possible to train a domain expert to develop models of their problem situations, no knowledge is lost in the communication process between domain expert and model developer [54].

2.3.5 Simulation Software

Development of Simulation Models

Three broad perspectives on how to approach the development of simulation models are easily identifiable from the literature [69]:

- The first perspective focuses on using a graphical user interface (GUI) that allows the user to build the model on the screen, connect the components by arcs to represent the model logic, and run the simulation.
- The second perspective is underpinned by the belief that no simulation program is able to model all types of systems behaviour without making some simplifications or modifications. This suggests that models should be developed from scratch by using a simulation modelling language.
- The third perspective concentrates on using a GUI that is able to automatically generate code, with the modeller making changes to the generated code to match the system needs. This approach is popular because it can reduce the model production time and provides the user with access to low level programming structures where needed.

For every perspective mentioned above, there exists an equivalent simulation software type.

Types of Simulation Software

Generally, simulation packages have a wide variety of features and characteristics. For example, some packages require coding effort, while some others need little or no programming effort. Simulation software tools can be classified into three categories as follows [70]:

1. **Data-driven simulators:** which are computer packages that allow the modeller to model systems with little or no programming.
2. **General-purpose simulation languages:** which are simulation packages used for modelling different types of systems with different characteristics.
3. **Program generators:** which are used as another way of making simulation more accessible to non-computer specialists. A program generator is a computer program which itself generates another program. The program generator can be changed to meet the system needs.

Current Trends in Simulation Software

Matt Rohrer [71] quoted the following “The number of those using models built by others will continue to climb. Also, we will see more expert users creating templates as a business venture”.

These trends will drive the simulation vendors to create more open products and tools to assist the model builders in creating templates. The interface must be easily customized, allowing for spreadsheet data input and mouse-based interaction with the model. Simulation products must also provide for customization through a structured language.

Template models can reduce model-building effort. While no two companies conduct business in exactly the same way, it is very common for companies within the same industry to perform the same tasks. A flexible template model can provide a framework to which detail can be added to accurately depict specific processes. Template models, developed by experts, cannot only help reduce model development time, but can provide valuable insight into the appropriate inputs to be considered when analyzing a given process.

2.3.6 Modelling and Simulation of Semiconductor Manufacturing Systems

Discrete event simulation is an important tool for analysing the characteristics of complex semiconductor manufacturing systems. Simulation is now recognised as a way of impacting decision-making through system performance analysis.

Figure 2.14 shows the different levels of abstraction for a semiconductor manufacturing system. Each level of abstraction presents an application area for modelling and simulation. Different purposes of the modelling and simulation activity can be identified, depending on the application area. The figure also shows the increasing level of complexity with lower abstraction levels. Finally, the virtual factory and production tools (detailed modelling of internal operations) are *not* the focus of this research.

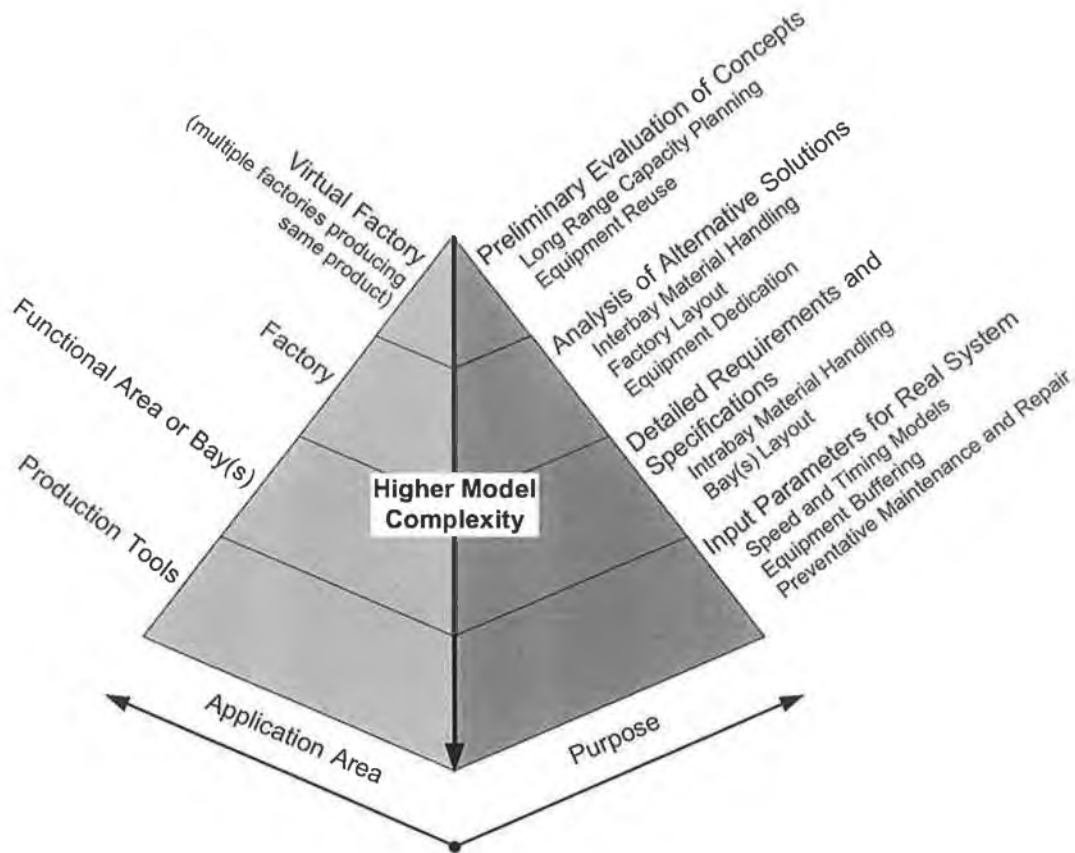


Figure 2.14: Modelling and simulation in semiconductor manufacturing, application area and purpose.

Simulation of the Entire Factory

The current “standard practice” for modelling wafer fabrication operations is to develop a single model that encompasses the entire scope of fab operations [72]. This effort and scope is necessary to capture the constraining interactions of material handling and storage systems, process equipment, human operators and product flow throughout wafer fabrication.

Due to the size of these facilities and the enormous complexity of the wafer fabrication process, these models are typically very large and sophisticated. Different approaches to modelling and simulating an entire factory are evident from literature.

1. Splitting up the model into two distinct parts that communicate with each other [72-74]. The two parts, shown in Figure 2.15, are the production tools processing the wafers and the material handling system that delivers the wafers to the tools for processing. The drawback of this will be the loss of an overall

picture on interactions that occur between different components outside the boundaries of these sub models.

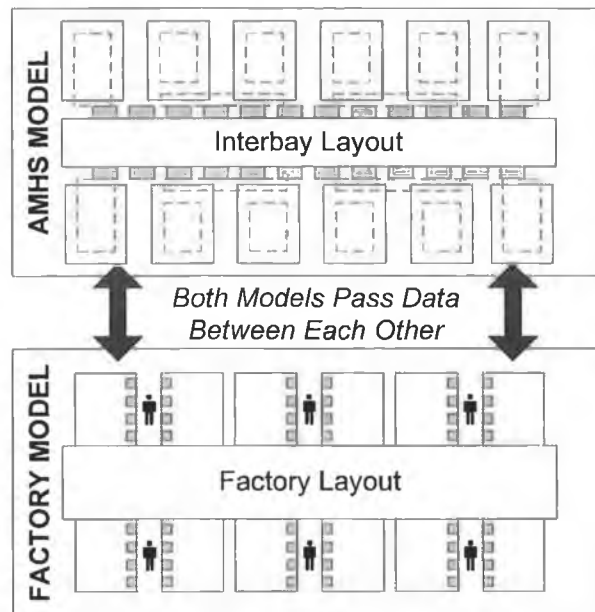


Figure 2.15: Two distinct model parts communicating with each other [74].

2. Modelling and simulation of the whole system with an increased level of detail as a single integrated model [75]. The single model, consisting of both the production equipment and the material handling system, can offer a better opportunity to understand interactions between the different areas and components of the system. However, a model with that complexity can take days to complete a single simulation run.
3. Simulating a single process area out of the whole fabrication complex [76]. This permits a more focused model to be built in the same time required to build a larger, less detailed model. The focused model is easier to update, and allows a quicker reaction to specific factory problems (Figure 2.16). In fact, it is believed that properties cannot be detected if a full fab model is constructed.

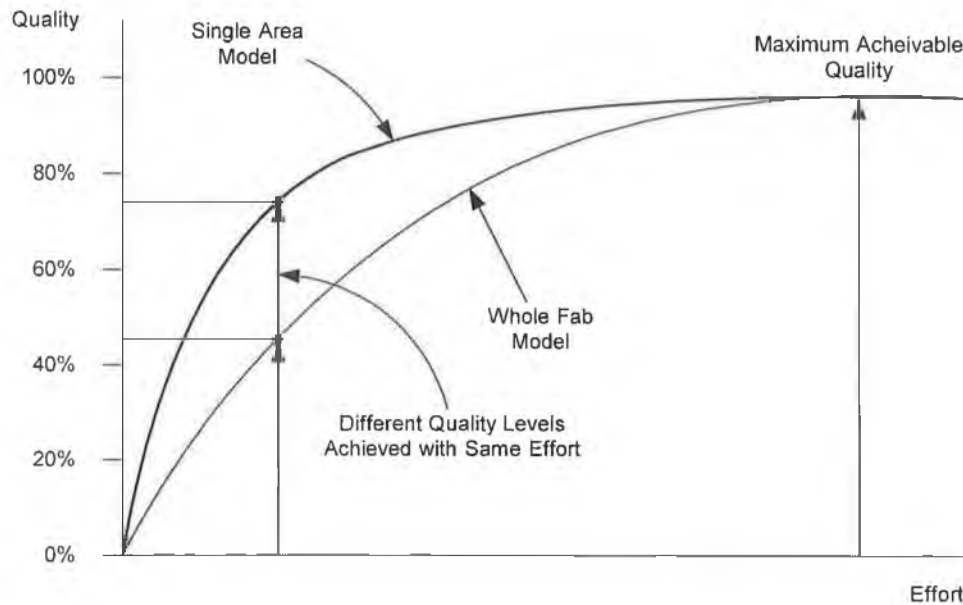


Figure 2.16: Relationship between effort to build the model and quality of results [76].

This third approach is more feasible than the first two approaches because it shows the overall picture of interactions occurring between the different parts of a fab. In addition, it allows development of high quality models in a shorter time. Finally, execution of a simulation model of a part of the fab is faster than execution of a simulation model of the entire fab.

Automated Material Handling System

The iterative nature of the AMHS design process – initial design, customer review, modified design, customer review, etc. – demands that the modelling tools used must be highly flexible, so that customers' changes can be quickly and easily incorporated in model revisions. The competitive pressures for quick turnarounds of high volumes of AMHS design projects mandate the use of special purpose reusable simulation models to reduce model-building time and increase model accuracy [59].

Simulation is a popular tool for accurately estimating the performance of an automated material handling system (AMHS). However, accuracy of the model is normally dependent on a detailed description of the AMHS physical system components and their coordinate positions. Automatically generating the physical system components, based on data extraction from a CAD layout file of the system, reduces simulation model

building time and increases model accuracy [77]. Four conditions must be fulfilled in order for the methodology to be applicable.

1. A CAD layout of the designed system must exist.
2. The simulation model has to be based on a dimensional accurate representation of the physical system components.
3. Models created must be similar in logic and general structure so that the same CAD transfer template can be re-used.
4. The methodology has to be applicable to a number of subsequent projects in order to justify the implementation effort.

If all these conditions hold, automatic model creation will produce significant reductions in the model building effort.

Modelling and Simulation of Interbay AMHS

The use of simulation in the design of interbay AMHS have been documented in the past [78], unlike intrabay AMHS. Interbay AMHS can be modelled by developing a From:To matrix that shows the hourly demand of bay-to-bay movements. The matrix is then coupled with a second From:To matrix that represents the mean transport time of automated material handling [79].

Modelling and Simulation of Intrabay AMHS

Simulation can play a key role during the design of intrabay AMHS by verifying that the system will not negatively affect throughput of an area. One of the modelling methodologies proposed for the intrabay AMHS is to develop a core model (representing the material handling part) with add-on components (representing the types and numbers of production equipment, or the type of outputs required) [78].

Such an approach can be useful when the material handling system is well developed, and thus the core model will not require modifications; this is not the case with the intrabay AMHS addressed by this research work, because different visions for this system are still evolving.

The literature shows that modelling and simulation has been extensively used to study intrabay AMHS. However, this literature does not show the methodology applied or even the assumptions made when developing these models. For example, Mackulak and Savory [80] performed a simulation-based experiment for comparing two different types of intrabay AMHS using a two-bay model. Their results show that an intrabay AMHS similar to the merge/diverge AMHS, presented in section 3.5.4 below, improves the performance of the system. Furthermore, they mentioned that the simulation analysis was performed using a sophisticated and detailed AMHS simulation model, however, no details pertaining to model development or systems being modelled were given. Other literature [81] has also addressed the merge/diverge system and compared it to another intrabay AMHS running without merge/diverge using simulation models. The results show that the merge/diverge system is more economical and offers better delivery time.

Simulation models were also used to predict the improvement of fab performance when implementing an intrabay AMHS [82]. While a general overview of the system under study was provided and an overall increase in the fab throughput was reported, no information regarding the detail of the model was given. This result gave no information about the interaction between the AMHS and the tools.

The overhead intrabay AMHS was also compared to other transport solutions, such as the continuous flow transport [32], using simulation models. The developed models did not explicitly contain production tools as components, however production tools were included through the creation of movement request data. Results of the comparison showed that the overhead intrabay AMHS is a better solution for transporting lots within a bay.

Efforts in reducing the simulation model development time have been reported as well, of which an integrated modelling methodology that combines a simulator and an expert system [83]. The expert system helps in preliminary analysis and selection among several alternatives. This methodology has been tested in interbay AMHS, and is expected to be applied to the intrabay AMHS.

Simulators Used in Semiconductor Manufacturing Companies

Looking at the simulation software survey carried out by OR/MS Today [84] it shows that there isn't much simulation software designed and developed specifically for a certain area of application in a semiconductor manufacturing system. Even if, such software exists, it won't be developed to solve a specific problem within its area of application.

This is simply because companies in the same field are in an ongoing state of competition, thus it is expected that very few competitors would be applying the same design, planning, and operational activities. Consequently, simulation software vendors are inclined to develop general application software rather than a very specialized one, in order to capture as much as it can of the market.

Software developed for simulating semiconductor-manufacturing devices (such as production and material handling equipment) are supplied by a small number of vendors (like Wright Williams and Kelly, FlexSim Software Products, and Brooks Automation) [85]. When it comes to using these packages for solving a specific problem within a factory, it would still require a great deal of time to tailor the software for this problem. This "tailoring" activity consumes valuable planning and design time to get a simulation model up and running.

This problem has been sensed by a number of simulation software vendors, represented in the drive toward building more and more "templates" that can be used by modellers, similar to that described by Collins *et. al.* [86]. However, developing one's "own template" that suits the needs of the client is inevitable in most cases. This has the advantage of developing a reusable model that could be tweaked-up easily to fit a problem within the application domain, rather than starting the whole modelling activity from scratch. Such an advantage is extremely important for models developed for non-existing systems and for addressing problems in the design and planning phase, as the data and operational requirements are constantly changing. Without a reusable model, every time system and the data associated with it changes; there will be a need for developing a new model to cope with these changes, which is impractical.

2.4 CONCLUSIONS

There are several findings that can be drawn from the literature survey presented above:

- The semiconductor industry must introduce technology revolutions at least every two years, requiring the introduction of new “cost effective” manufacturing solutions to the industry at almost the same rate.
- The most recent technological change that the semiconductor industry faces is the introduction and development of 300mm wafer fabrication technology.
- 300mm-wafer carriers require fully automated handling system and the best means of implementing this is still not known.
- As the physical AMHS system does not exist it must be studied using simulation as it is too complex for pure analytical modelling.
- Existing reported results report improvements in overall factory performance, but do not show the interaction between the AMHS and the production tools.
- Reusable simulation models are needed for the intrabay AMHS as the system design process is a highly iterative. Developing a model from scratch whenever a design change occurs is infeasible.
- Component-based modelling can reduce the complexity of simulation models and can allow reusability of these components for building an aggregate simulation model.
- Developing focused simulation models of sub-sections of a semiconductor manufacturing facility is a better approach for studying AMHS/Tool interaction than using a single model of the facility.

3 AUTOMATED MATERIAL HANDLING SYSTEMS FOR 300 MM WAFER MANUFACTURING FACILITIES

This chapter introduces the different components required for material handling in the 300mm wafer manufacturing facilities. The 300mm AMHS is fully automated and serves the purpose of storing and transporting material around the fab. Materials transported are lots of wafers, which are held special boxes called FOUPs. There are a number of components that contribute to realisation of the 300mm AMHS concept shown in Figure 3.1, which are described in more detail throughout the chapter.

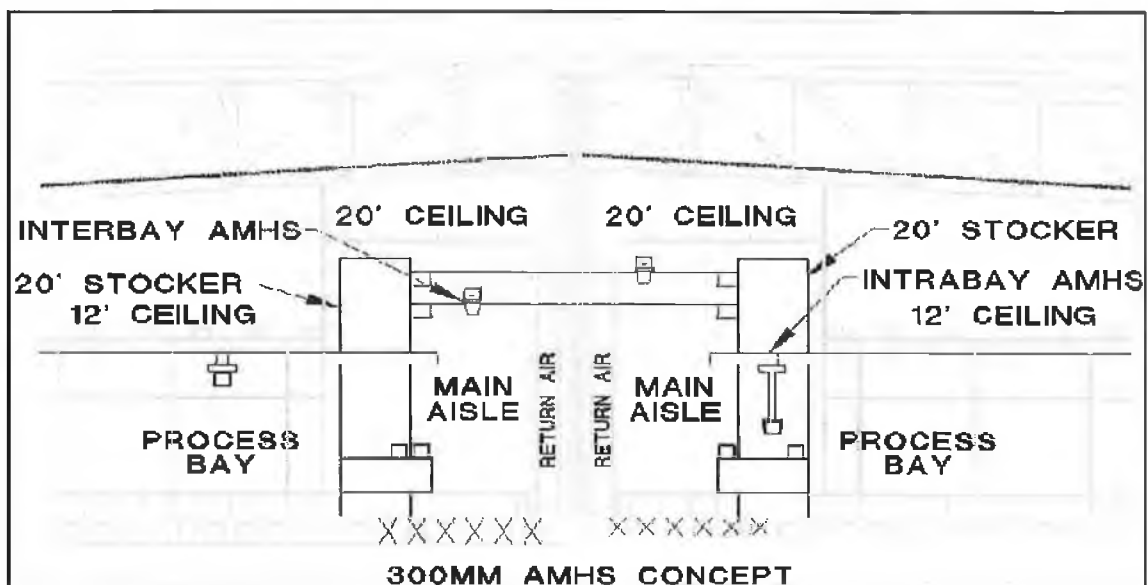


Figure 3.1: The 300mm Automated Material Handling System (AMHS) concept.

These components are:

- FOUPs for holding wafers.
- Stockers for storage of FOUPs.
- Interbay AMHS for moving lots between bays.

- Intrabay AMHS for moving lots between stockers and production tools in a bay.
- Personnel Guided Vehicles (PGV) for manual transport of FOUPs.

3.1 FRONT OPENING UNIFIED POD

The front opening unified pod (FOUP) is used in the 300mm environment for holding the wafers lot. A FOUP can carry up to 25 wafers, which is the dominant lot size handled in the factory. Furthermore, a bar code is given for every FOUP to track the lot it is carrying and identify the manufacturing process required for that lot at any time. The database, which defines the processing stages of the lot in the FOUP can be accessed at both stocker and production tool, but can only be updated at a stocker.

A production tool can only open the front opening of a FOUP, as shown in Figure 3.2. The wafers are completely sealed inside the FOUP, which reduces the risk of lot contamination. Compared to a regular 200mm wafers box, a FOUP is twice the weight and 325% the volume. Consequently, introduction of the intrabay AMHS and vertical stockers are needed to overcome these weight and size issues. These issues directly affect the transportation and storage of FOUPs in a fab.



Figure 3.2: Front Opening Unified Pod.

3.2 300MM STOCKERS

As mentioned earlier, material storage space became a key concern due to the increased volume of the FOUP. It has been estimated that stocker floor space would nearly double and consume a significant amount of the total factory floor space, if the 200mm stockers were used.

This problem is overcome by using vertical stockers that are over 20 feet tall (their base start under floor level). The stockers are all aligned and concentrated along the centre of the factory, where each stocker is associated with a particular bay. Stockers link the different types of material handling systems within the factory as shown in Figure 3.3.

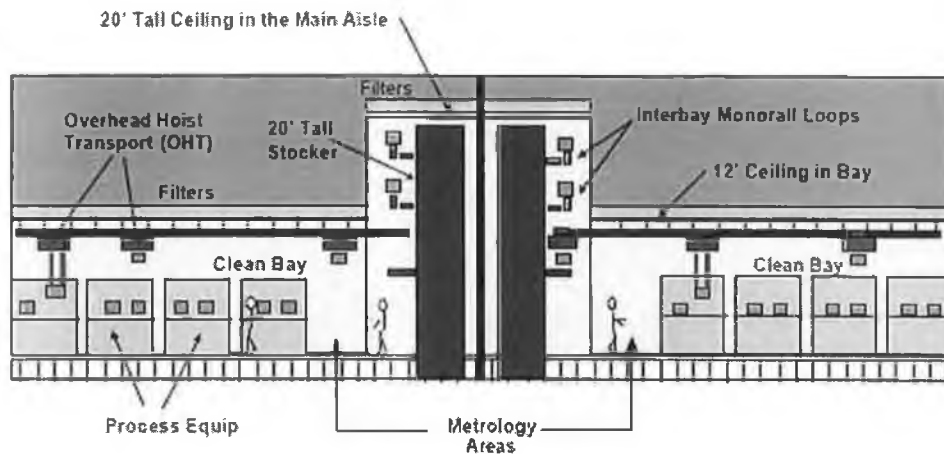


Figure 3.3: Stockers linking the different material handling systems [31].

Figure 3.4 shows that stockers are interconnected via 2 levels of interbay monorail to speed up material transport and save space. Stockers are also connected at a third level, which are the intrabay transport loops, on which the Overhead Hoist Vehicles (OHV) operate and deliver FOUPs to production equipment in each bay. Finally, at a fourth level input and output ports are present for PGVs in case of failure of any of the OHVs.

Thus, a stocker typically has a total of eight input/output ports, which are:

- 4 interbay input/output ports for interfacing with the two interbay loops.
- 2 intrabay input/output ports for interfacing with the intrabay loop.
- 2 operator ports for interfacing with PGVs.

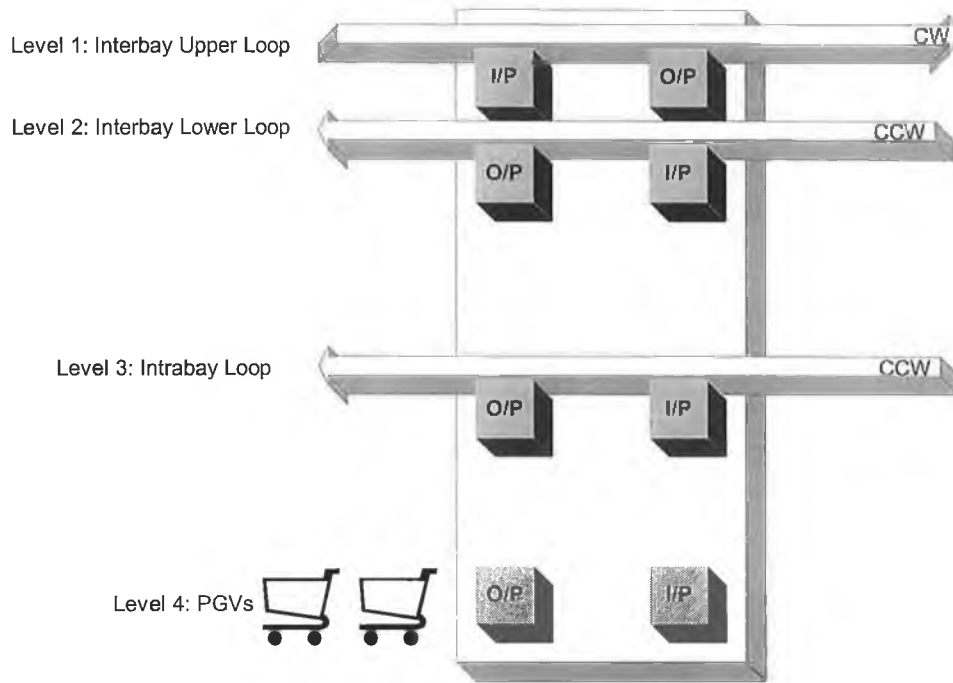


Figure 3.4. Multi-level input and output ports of a stocker.

Figure 3.5 shows the internal structure of a stocker. Several shelves are found in a stocker for storing the FOUPs, received from the different input ports, and accurately identifying the location of each FOUP. A robot is responsible for moving the FOUPs, inside the stocker, between the different input/output ports.

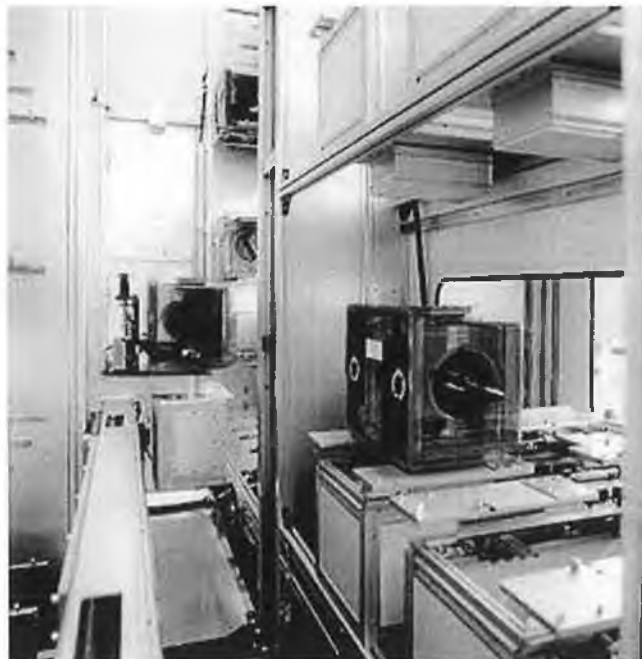


Figure 3.5: Inside a stocker showing FOUPs stored on shelves and the stocker input/output ports (Courtesy of Daifuku Co., Ltd. [87]).

The moves done by a stocker robot can be as follows:

- Movement of a FOUP from an interbay input port (level 1 or 2) to the stocker shelf.
- Movement from the stocker shelf to an output port (level 3 or 4).
- Movement from an input port (level 3 or 4) to the stocker shelf.
- Movement from the stocker shelf to an output port (level 1 or 2).

The time needed by a stocker for receiving a FOUP at an input port and releasing that FOUP at an output port is referred to as the stocker cycle time, which can become a system-limiting factor in case of high volume manufacturing.

Finally, FOUPs cannot be released from a stocker if the production tool required for processing the lot is not ready (full, or down). The production tool needed for the lot next manufacturing step is identified through the bar code on the FOUP.

3.3 300MM PRODUCTION EQUIPMENT

The 300mm production equipment, either process tools or metrology tools, are all based around the same template. Every tool has a mini-environment to isolate wafers from the open fab air where FOUPs can be opened and the wafers are accessed.

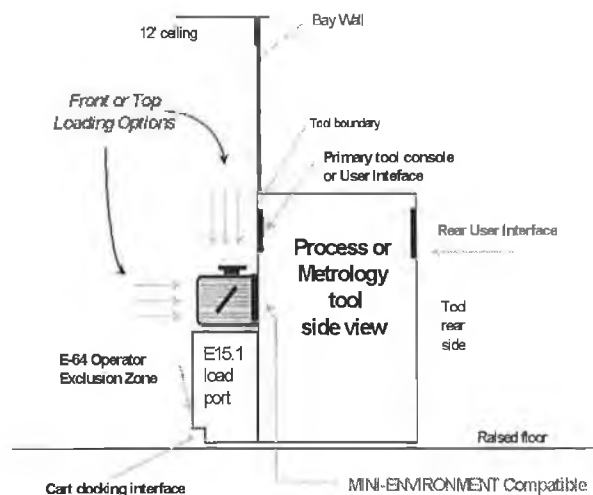


Figure 3.6: Mini-environment and loading ports [21].

Furthermore, common load-port interfaces (loading ports) exist for all tools, which act as an input port and an output port at the same time, and facilitates FOUP loading and unloading by an OHV (top loading) or a PGV (front loading) as shown in Figure 3.6. Because wafers are isolated from the open fab environment by the FOUP and mini-environment, a slightly lower class of cleanroom can be used in the bays.

In a single production tool each wafer passes through several stages, as shown in Figure 3.7. At each stage the tool performs the same operation on every wafer till all the wafers in a lot are finished and are returned back to the FOUP. The time required by the production tool to complete processing of a lot is the processing time of that tool.

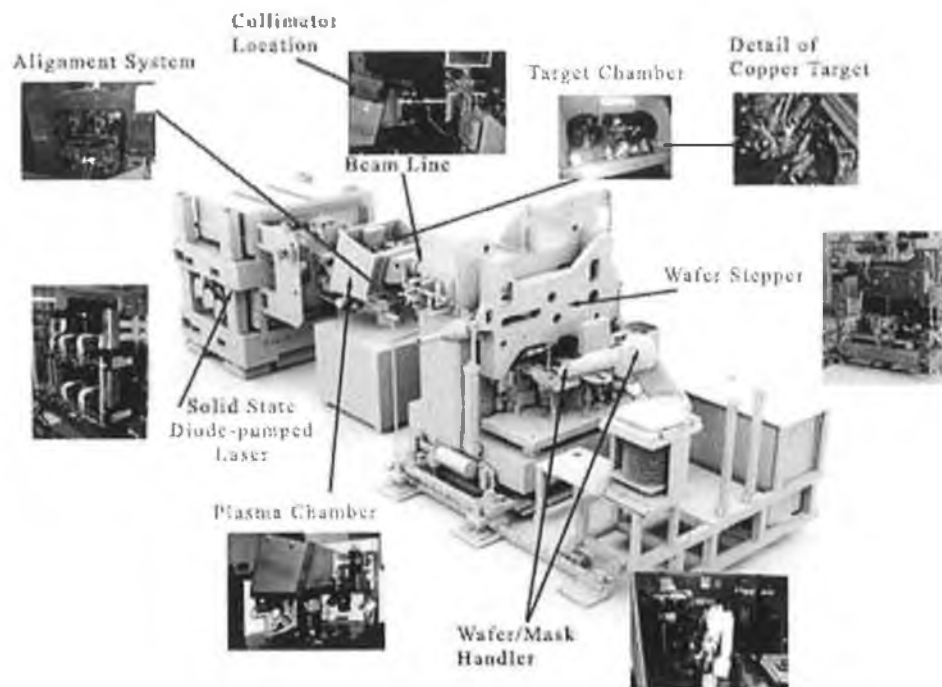


Figure 3.7: Different parts inside a production tool (Courtesy of JMAR Technologies, Inc. [88]).

3.4 INTERBAY AUTOMATED MATERIAL HANDLING SYSTEM

This is one of the two major mechanical components responsible for material handling in a 300mm semiconductor manufacturing facility. The interbay AMHS is responsible for transferring lots from one bay to another.

The transport system is an Overhead Shuttle for transferring lots between stockers by running several vehicles along a monorail, as shown in Figure 3.8. The way or path

where the interbay is placed is called a cleanway. Lot transfer from one bay to another is made by moving lots from stocker to stocker along this route.



Figure 3.8: Monorail Overhead Shuttle (Courtesy of Shinko Technologies, Inc. [89]).

Furthermore, the interbay AMHS is developed in such a way that ensures rapid transfer of lots between bays, guaranteed by using two loops of interbay AMHS running in different directions.

This means that if a vehicle passes a lot waiting for pickup, it will never have to travel over a full loop to pickup that lot, as another vehicle running on the other loop (travelling in the different direction) can pickup the lot.

Finally, the interbay AMHS has long been in use in the 200mm environment and the best means of operation is well established.

3.5 INTRABAY AUTOMATED MATERIAL HANDLING SYSTEM

This is the second major mechanical component responsible for material handling. The intrabay AMHS is used *only* in the 300mm environment because of the increase of FOUPs weight. This *new* automation component is a major concern for the semiconductor industry in general, and Intel in particular, because the best method of configuring it is still unknown. Therefore, this research work focuses on investigating the operation of this automation component.

3.5.1 Effect of Increasing FOUPs Weight

In a high volume manufacturing environment, lots have to be moved between stockers and production tools at a high rate. The weight of these lots (18 lb. – 8 Kg.) makes manual transfer impossible. This is due to ergonomic and safety reasons, the lots cannot be carried manually on regular basis.

Furthermore, the 300mm bay size is too narrow to allow for easy manoeuvring and parking of manual carts (PGVs), even if a PGV can carry two lots at time it would still be inappropriate to use in high volume manufacturing.

Thus, the direct effect of this increase in weight is the necessity of implementing an intrabay AMHS for moving lots between stockers and production tools inside a bay.

3.5.2 Components of the Intrabay AMHS

An intrabay AMHS is an Overhead Hoist Transport (OHT) for transferring lots between the production equipment and stockers using Overhead Hoist Vehicles (OHVs). The OHVs, shown in Figure 3.9, are responsible for picking up and dropping off lots at the stockers and at the production tools.



Figure 3.9: Overhead Hoist Vehicle (Courtesy of Shinko Technologies, Inc. [89]).

With the advent of the OHV system for movement of wafers between stockers and tools, the lots not only move faster but also enable the use of narrower bays. Because floor running rail-guided intrabay vehicles no longer run down the bay at floor level

(common in 200mm), the extra clearance space is not needed and significant floor space can be saved (as illustrated in Figure 3.10).

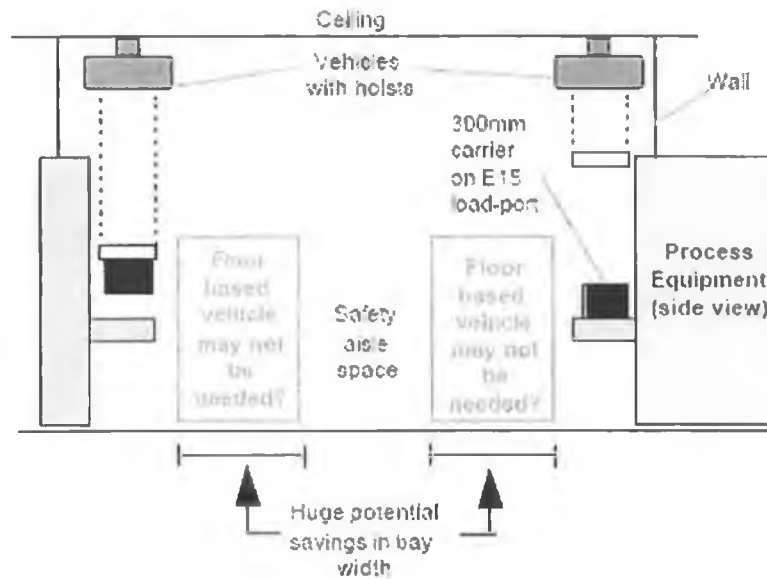


Figure 3.10: Bay width savings realised when using OHVs [90].

3.5.3 Lots Delivery/Collection To/From Production Tools/Stockers

OHVs have front and side sensors; for sensing other vehicles ahead and for accurate positioning over production tools/stockers loading ports. Thus, whenever an OHV reaches the required production tool/stocker, the side sensor receives a signal telling the OHV to stop and perform the lot delivery or collection as shown in Figure 3.11.

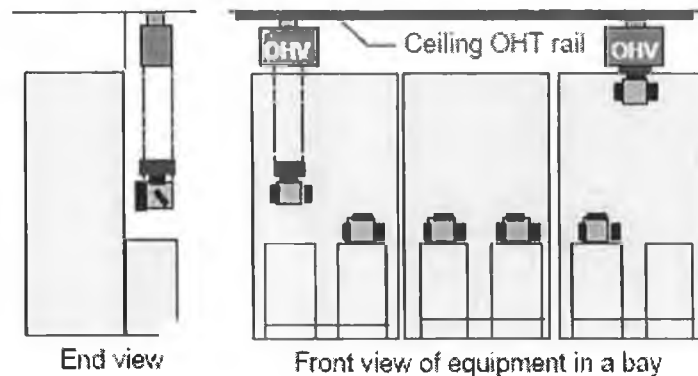


Figure 3.11: OHVs dropping off and picking up FOUPs at loading ports of production tools [91].

3.5.4 Different Visions for the Intrabay AMHS

The previous discussion assumes that the interbay AMHS is dedicated to moving lots between bays and the intrabay AMHS is dedicated for moving lots within a bay, which is one of the visions of operating the intrabay AMHS.

Another vision for the intrabay AMHS is to link a small group of intrabay AMHS systems, within a full manufacturing facility, together using what is called a Merge/Diverge link. This promises better operational performance of the AMHS when compared to operating two dedicated AMHS systems.

It must be noted that in a merge/diverge AMHS, a limited number of bays can be connected together using the same intrabay AMHS through merge/diverge links (because of the material handling systems technology requirements dictated by the international technology roadmap for semiconductors [92]). If a lot is to be transferred to a bay that lies outside the merge/diverge link, then the interbay AMHS will be used (as shown in Figure 3.12).

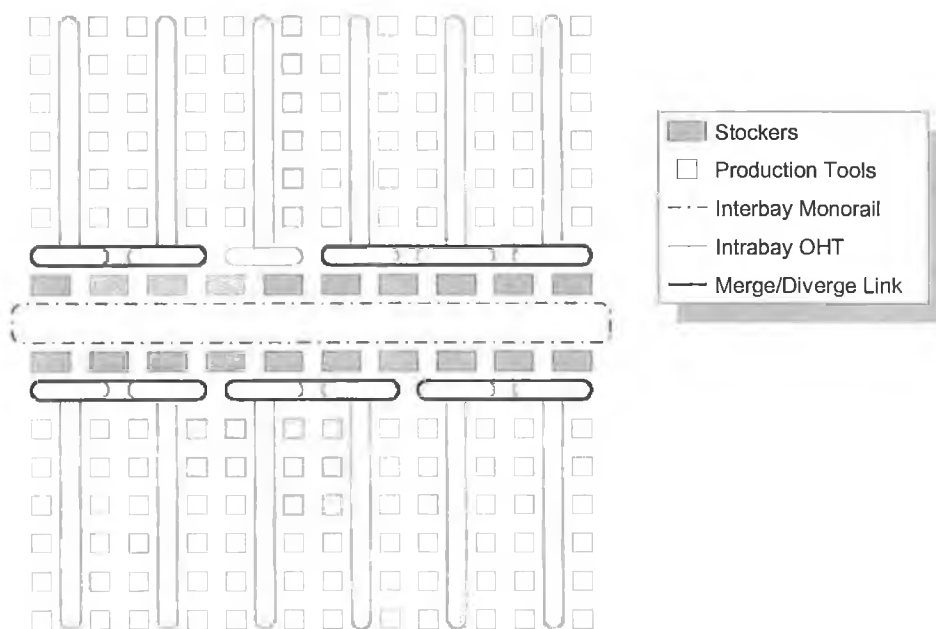


Figure 3.12: Linking a number of bays using merge/diverge links.

The merge/diverge AMHS enables the handling of lots between a specified number of bays using the intrabay AMHS and a number of merge/diverge links. Accordingly, the necessary interbay AMHS moves will be reduced. With merge/diverge, OHVs can

diverge into a bay, merge with other OHVs, or simply move between bays through the different merge/diverge links.

Figure 3.13 shows that without merge/diverge three moves are needed to transfer a lot, and the use of the interbay AMHS as well. On the other hand, with merge/diverge this can be done in two moves only, without the interbay AMHS. This results in reduction of time needed to transfer a lot between two bays (given that the two bays are linked).

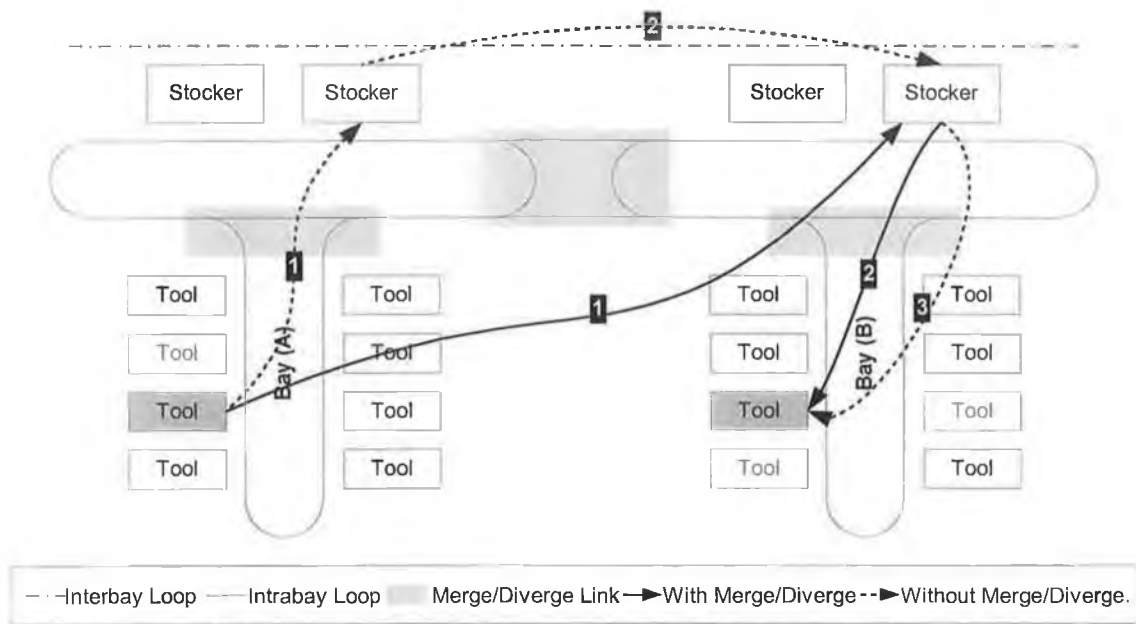


Figure 3.13: Moving lots between bays with and without merge/diverge.

Merge/diverge intrabay AMHS presents another challenge for the semiconductor manufacturers, because of the several possibilities of configuring it and particularly the method of routing free vehicles.

3.6 AMHS DELIVERY MODELS OF THE FUTURE

Today, the merge/diverge technology is possibly the latest technology for the intrabay AMHS [92]. However, chipmakers have begun to focus their attention on the AMHS delivery model looking for one that can optimize fab productivity.

One such model is the Direct Delivery (DD) model, which would replace the traditional interbay and intrabay distinctions in the future. The direct delivery model is a seamless

delivery system using one track and one type of vehicle, which can be thought of as an extension of the merge/diverge vision.

3.7 DEFINITION OF SYSTEM UNDER STUDY

The system under study is defined as a bay (or bays) containing stockers for storing lots, production tools for processing of lots, and intrabay AMHS for moving these lots between stockers and tools. Furthermore, the intrabay AMHS encompasses OHVs, bay loop(s), and merge/diverge links (if applicable).

Before detailing the rules that govern the operation of the intrabay AMHS, it should be noted that lot delivery refers to moving a lot from a stocker to a production tool, lot return refers to moving a lot from a production tool to a stocker, and lot collection refers to lot pickup at stockers or production tools.

3.7.1 Rules for Lot Movements

There are two main types of moves that can be made in an intrabay AMHS, Stocker-Tool-Stocker moves, and Tool-to-Tool moves.

- Stocker-Tool-Stocker moves mean that the lot leaves from a stocker and goes to a tool, and is returned back to a stocker, even if the next process step is to be carried out by a tool that the OHV passes by.
- Tool-To-Tool moves mean that a lot can be transferred from a tool to another tool directly, given that the next process step is to be carried out by a tool that the OHV passes by. However, the final destination of the lot is a stocker.

Figure 3.14 shows a Stocker-Tool-Stocker move to transfer a lot from tool (1) to tool (2); after being processed at tool (1), the lot has to go to the stocker first then it goes from the stocker to tool (2).

Due to the current technology that applies to intrabay AMHS, Tool-To-Tool moves cannot be done and only Stocker-Tool-Stocker moves are allowed [92].

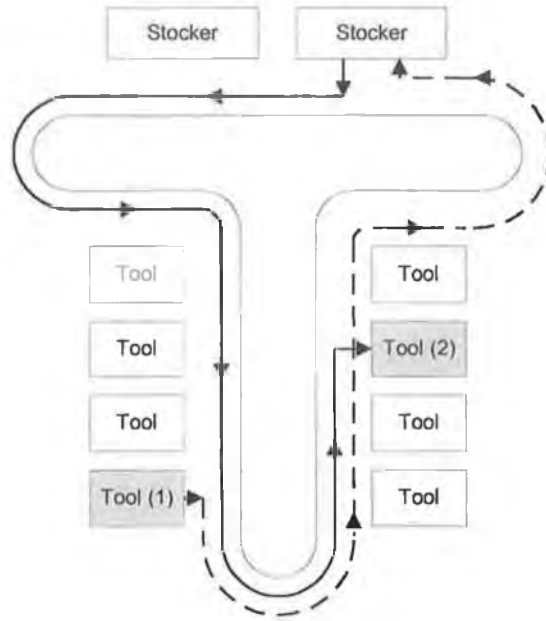


Figure 3.14: Stacker-Tool-Stacker lot delivery in a bay.

3.7.2 Routing Rules

Figure 3.15 shows a typical two bay layout with the track broken down into segments of routes, each segment is given a number and the start/end points of these segments are indicated by the thick vertical lines. The thicker track shows the outer loop that connects the two bays together. The grey shaded sections are the merge/diverge links.

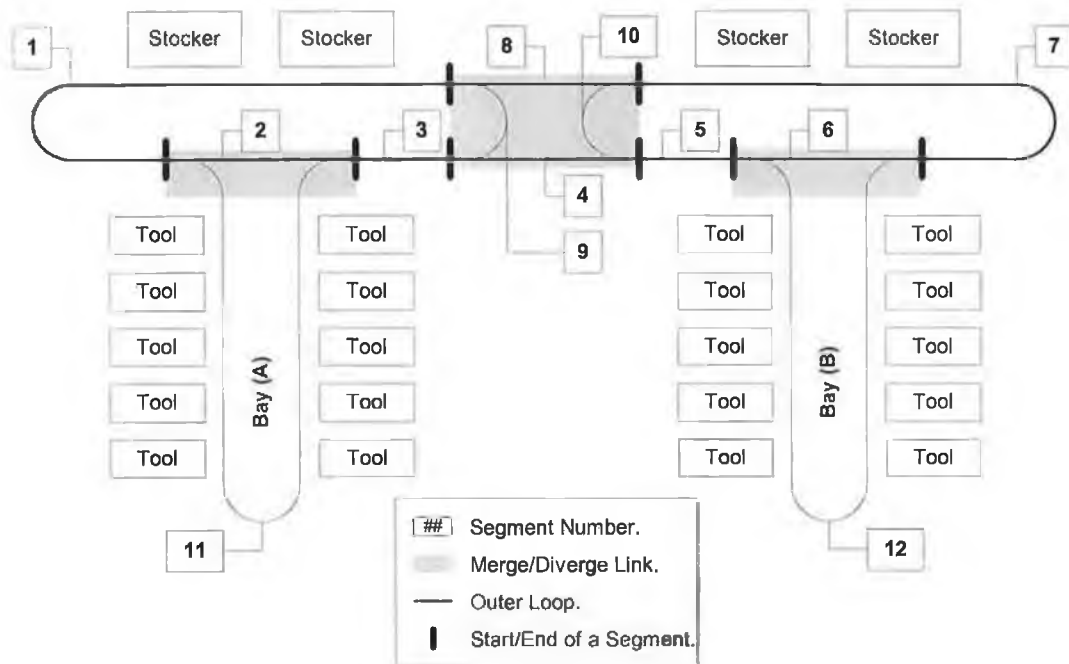


Figure 3.15: Different routes an OHV can loop.

Using this layout, the outer loop is the segments 1-2-3-4-5-6-7-8, bay (A) loop is segments 1-11-3-9, and bay (B) loop is segments 7-10-5-12.

For a single bay layout (without merge/diverge), OHVs are not allowed to pass through segments 2, 4, 6, and 8. This means they are restricted to their bay loop regardless of the transfer required for manufacture, and thus routing rules are not applied in that case.

When the two bays are linked, the route that an OHV takes for collecting or returning a lot depends on:

- Number of lots waiting in the two bays.
- Current location of OHV on the intrabay loop.
- Bay for next processing step.

The decision tree in Figure 3.16 illustrates the route that should be selected based on those parameters.

The default route for empty OHVs is the outer loop. If there is a lot awaiting pickup in one of the bays, then the OHV will diverge into the bay to collect the waiting lot. Furthermore, if there are waiting lots in both bays, the OHV goes to the bay with the higher number of lots waiting collection. Finally, if there is a tie between the numbers of waiting lots in the two bays, the OHV collects the lot from the bay closer to it.

On returning the lot, the OHV returns the lot to the stocker that belongs to the bay in which the next process step is to take place. If the lot is to be returned to a bay that is not included in the merge/diverge link, bay (C) for example, the lot will be returned to the stockers of the same bay it was collected from.

This approach to OHV route selection for collection and return of lots, keeps lot deliveries within the same bay where possible. This means that a lot leaving a stocker in bay (A) will always be delivered to a tool in bay (A) and the OHV will travel through (1-11); similarly for delivering a lot from a stocker in bay (B), the OHV will travel through (7-10-5-12). Consequently, whether the merge/diverge links exist or not, OHVs will always travel on the bay loop when delivering lots.

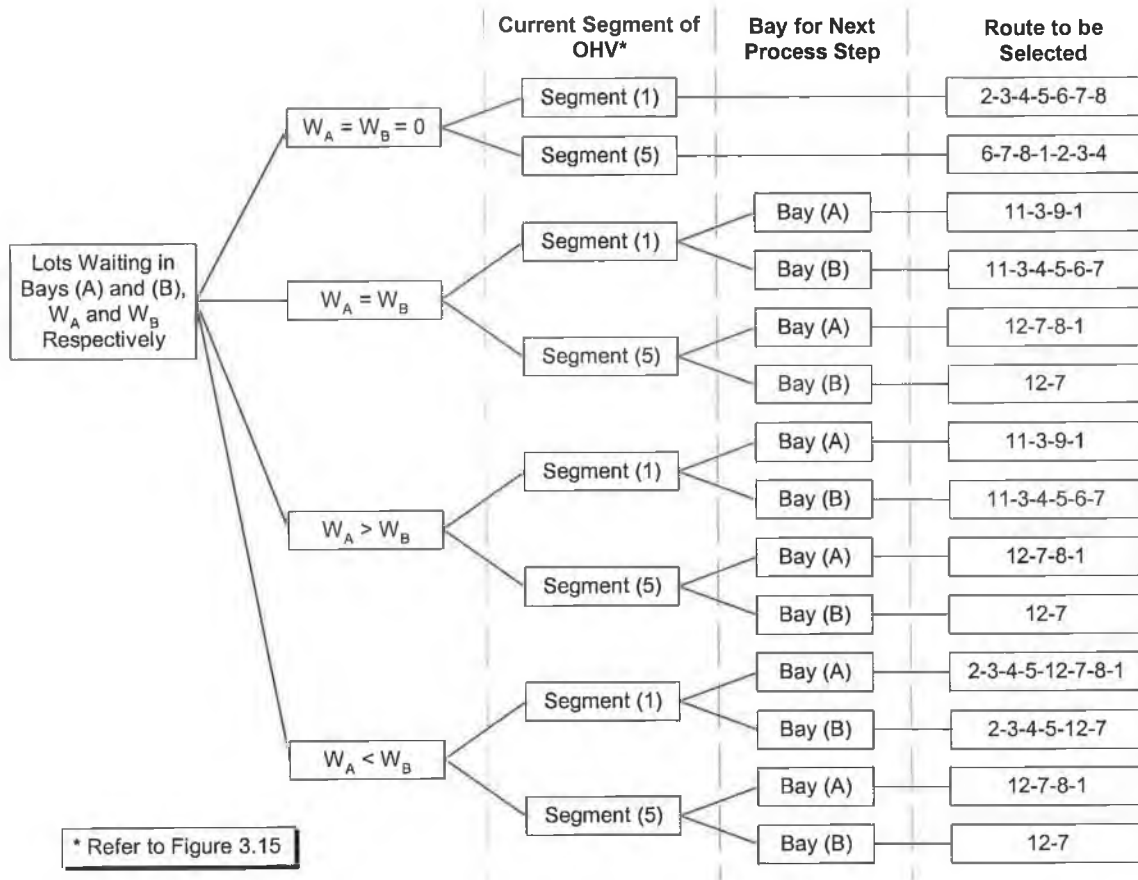


Figure 3.16: Decision tree for selection of different routes.

3.7.3 Pickup/Drop-off at Tools and Stockers

Whenever a vehicle arrives at a tool or a stocker, certain checks are made to determine whether the vehicle is to pickup (collect) or drop-off (deliver) a lot at this tool/stocker. The action to be taken, whether pickup or drop-off depends on the status of the OHV.

The OHV can be:

1. Empty.
2. Carrying a lot for processing.
3. Carrying a processed lot.

As shown in Figure 3.17, each status represents a different case for pickup/drop-off at tools/stockers as follows:

If an empty OHV arrives at a tool or a stocker, a check is made whether there is a lot ready to be picked up or not. The OHV picks up the lot unconditionally.

A full OHV carrying a processed lot arriving at a stocker, and the stocker satisfies a number for conditions qualifying it for accepting the lot (refer to 4.4.4 for explanation); the OHV drops-off the lot at the stocker input port. But if that OHV, with the processed lot, arrives at a tool loading port, it takes no action because tool-to-tool delivery is not allowed.

A full OHV carrying a non-processed lot arrives at a stocker, the OHV takes no action. This prevents lots from being returned to stockers without being processed. However, if that OHV arrives at a tool, and this tool qualifies for processing the lot (refer to 4.4.4 for explanation) the OHV drops-off the lot.

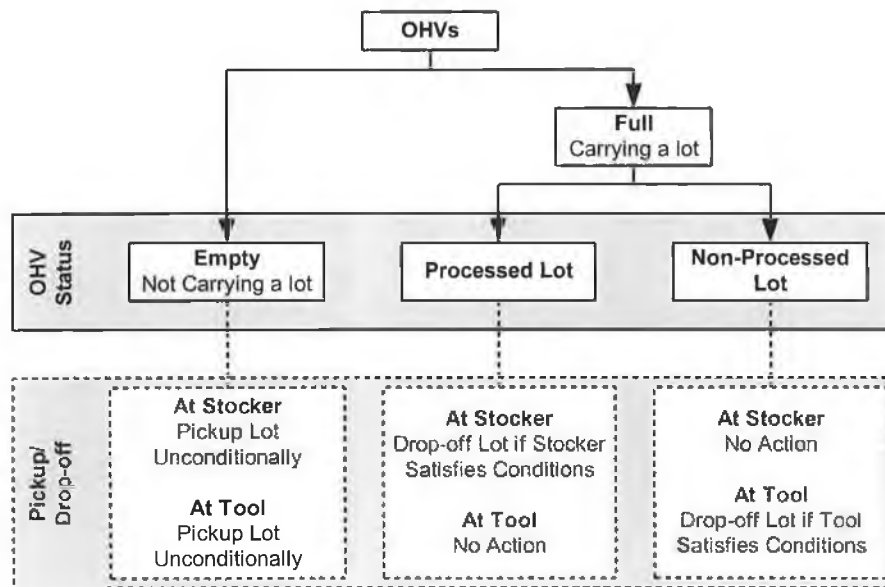


Figure 3.17: Illustration of the different statuses of OHVs and their respective pickup/drop-off case.

3.7.4 Moves Allowed at Production Tools

OHVs run on the intrabay loops in a counter-clock wise direction only, and thus for an OHV to go to a certain point it passed, it has to travel over the whole intrabay loop to reach the point it passed again rather than simply going back to that point. This has a major effect on the pickup/drop-off moves at tools, as production tools have a number of loading ports each acting as an input and output port at the same time.

Consider the following situation of a full OHV arriving at a tool to make a drop-off, and the tool already has a waiting lot ready for pickup; the ability of the OHV to pickup the waiting lot depends on which loading port the lot lies, as illustrated in Figure 3.18.

If the lot waiting for pickup lies on a loading port following an empty one, as shown in Figure 3.18 (A); the OHV will be able to do the two moves, dropping-off the lot it carries, and picking up the waiting lot.

However, if the waiting lot lies on a loading port preceding all empty ones, as shown in Figure 3.18 (B); the OHV will not be able to do the two moves and it will only drop-off the lot it carries, simply because the OHV cannot go back.

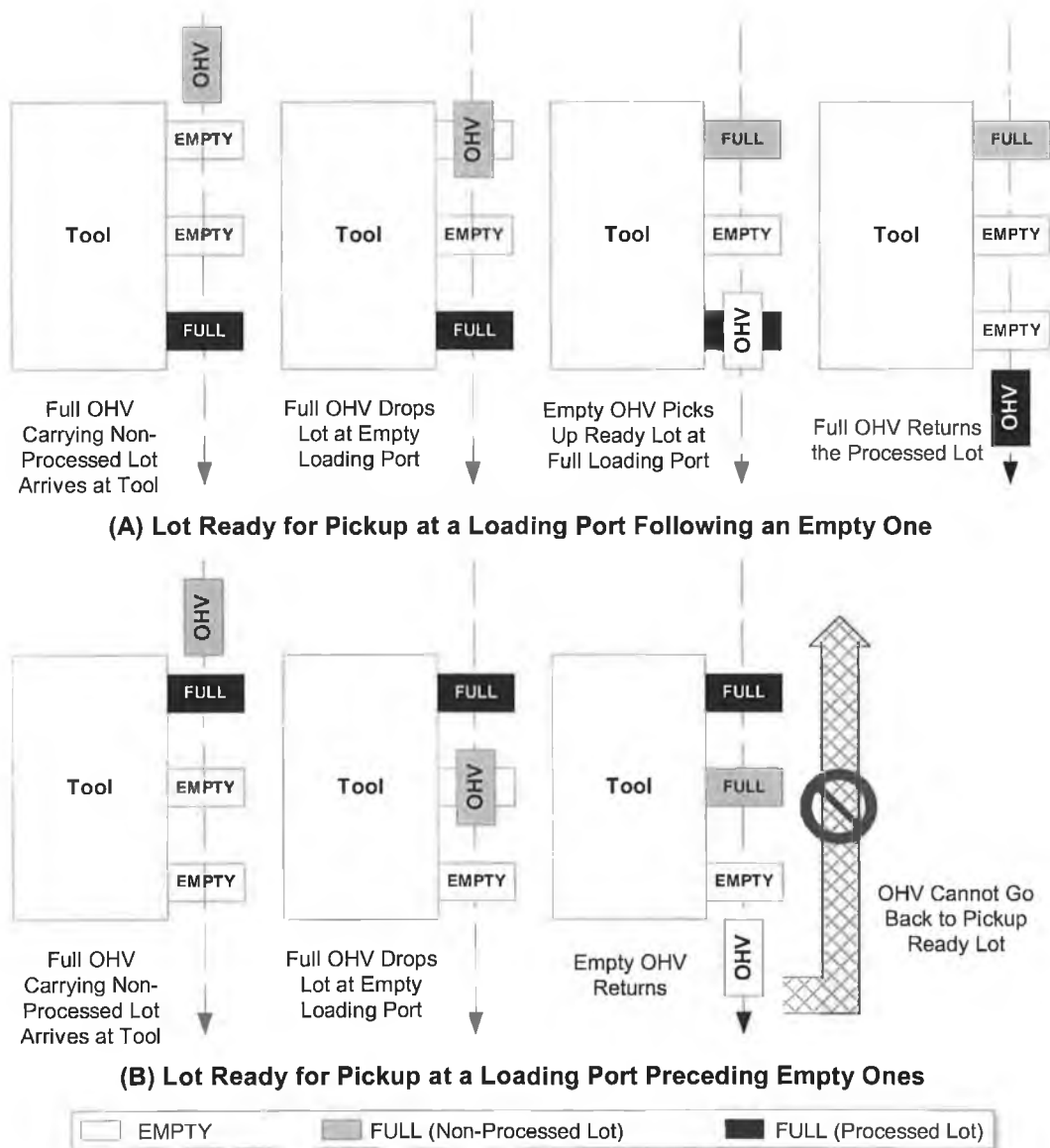


Figure 3.18: Locations of a lot ready for pickup and its effect on the pickup/drop-off process.

It is assumed that the empty loading port, where the lot is to be dropped-off, will always be the first available port that the OHV comes across. Moreover, the waiting lot (if it exists) will always lie on the last loading port the OHV passes by, and thus the OHV will always be capable of performing the two moves.

It should be noted that the issue presented in this section doesn't apply to stockers, as they have a dedicated input port and a dedicated output port.

3.8 BACKGROUND

During the research period in industry, simulation capabilities for the intrabay AMHS were developed. The industry partner is a worldwide corporation and relies on a number of simulation tools for solving different design and planning problems. Another of their manufacturing sites, similar to the plant in Ireland, has been using Extend™ simulation environment and recommended using it for this project.

The simulation project aims at finding the interaction between a *new* intrabay AMHS and the tools, in other words to understand:

- How the new material handling system will affect the performance of tools.
- How long will a lot be waiting at a tool's loading port to be picked up.
- How fast will the AMHS deliver lots to tools?
- How long will the machine stay idle, waiting for a lot to arrive?

It must be stressed that the system under study is the next generation of planned manufacturing at the facility, and that the real system being modelled does not yet exist.

The scope of the project is mainly of the intrabay AMHS. Interbay AMHS and labour are out of this project's scope. The toolsets are part of the project due to the nature of the objective, which is the interaction between tools and the AMHS. Thus, the toolsets must be considered, yet with lower level of detail.

Furthermore, due to the nature of this research project the simulation model developed must target non-simulation experts. Therefore, easy to use and user-friendly model is most favourable.

3.8.1 Motivation

The simulation study is motivated by the semiconductor industry transfer to the 300mm wafer technology leading to the increase of wafers weight. Lots of 25 wafers (predominantly) are handled on the shop floor, and thus lots cannot be handled manually due to ergonomic and safety reasons. Furthermore, the use of manual carts for handling the lots cannot be done, because the bays are too narrow to allow for easy manoeuvring of these carts.

This consequently led to the introduction of a new automation component for handling lots inside the bays, represented by the intrabay AMHS, as the lots can no longer be transferred from a stocker to a toolset manually.

As mentioned before, Intel is already using an AMHS for delivery of lots between the functional areas in their current fab. This case is different for the new fab, in which an Intrabay AMHS is to be used for delivery of lots to the toolsets within the bays.

The main concern of the industry is that the impact of the intrabay AMHS on the performance of the toolsets is unknown, as this material handling system hasn't been implemented before. Therefore, the industry is still searching for the best configuration of the Intrabay AMHS, which will consequently lead to a higher performance from a toolset perspective.

This briefly means that the transfer to 300mm wafer technology resulted in:

- Increase of wafers weight (18 lbs compared to 8 lbs).
- Wafers cannot be handled manually (safety and ergonomics constraints).
- Implementing intrabay AMHS is a must.
- Intrabay AMHS impact on performance of production tools is *unknown*.

- Therefore, simulation is needed to predict the performance of AMHS and production tools in the new 300 mm environment.

3.8.2 List of Issues

The main concern is how to predict the performance of a new non-existing system under different operating conditions and using different configurations. There are other issues that must also be considered in this simulation study, which are:

- The nature of the design process of the intrabay AMHS is highly iterative, which leads to continuous change in data and system configuration.
- Furthermore, a general misconception and unawareness of the mode of operation of the system exist.
- There exists two different visions for the intrabay AMHS and Intel is unable to determine which is more suitable for implementation.

The models required to simulate these systems are predictive discrete event simulation models. At the beginning, a simulation model (referred to in text as the first simulation model or the old model) was built to test the capabilities of ExtendTM in modelling an intrabay AMHS. The plan was to build a model for two bays, using a hierarchical approach, focusing on the intrabay AMHS component. Afterwards, the two bays are to be duplicated until the full factory has been modelled (refer to Appendix D for further information about this model).

Due to lack of data and information pertained to the performance and operation of the intrabay AMHS, the focus has shifted towards modelling the production tools in more detail. Afterwards, information related to the intrabay AMHS gradually appeared. Due to the scope of the simulation project, focusing on the intrabay AMHS, these pieces of information needed to be incorporated into the model. Consequently, model complexity, in terms of size, will be enormous and would ultimately be infeasible to simulate.

Although the first model was not implemented, valuable insights related to the nature of the system under study as well as the capability of ExtendTM as a simulation tool to model intrabay AMHS, have been gained. These insights paved the way for better development of simulation models for the system under study.

Throughout the modelling process the following major problems have been identified:

- The design process related to all the devices of the facility under development, specifically of the AMHS, is of a highly iterative nature.
- Data gathering is difficult, and maintenance of these data is even more difficult.
- ExtendTM simulation environment is equipped standard blocks for modelling common material handling systems, however these are not capable of accurately modelling the different components of a complex AMHS.
- Modelling both the production tools and AMHS with high level of details will ultimately lead to an infeasible model, specially if one model is built for the entire facility.
- General confusion and misconception, has been sensed, about how the intrabay AMHS will operate and interact with the production tool.

Based on these problems, a tool has been developed that considers and effectively overcomes those problems. The tool provides a user-friendly interface for building models, from scratch, for users with minimal simulation experience. The approach towards building that tool is as follows:

- Identify and model the main components of any intrabay AMHS.
- Compile all components as building blocks in a library.
- Build into the model the capability to switch between different types of intrabay AMHS technologies known to Intel at the date of model building.
- Each block must be generic in nature and can be re-used with minimum amount of change.
- Keep the data controlling the various blocks in a central database.
- Allow for versatile addition and/or removal of any component.
- Model only the most important details of production tools and intrabay AMHS and for a limited number of bays.

- Add simple animation capabilities that helps in showing the interaction between the production tools and the intrabay AMHS.

The newly developed tool is called GTIA-M&S, Generic Tool for Intrabay AMHS Modelling and Simulation. The steps of building GTIA-M&S are discussed in details in the next chapter.

It must be noted that GTIA-M&S is a tool containing a number of sub-models representing different parts of an intrabay AMHS. GTIA-M&S has been developed using the component-based modelling approach (discussed previously in 2.3.4), making these sub-models the “*components*” needed for developing larger aggregate models of a bay or number of bays.

The main objective behind building the GTIA-M&S is to provide a generic and user-friendly modelling and simulation capability to a non-simulation expert for modelling and simulation of the next generation of intrabay AMHS of a semiconductor fabrication facility. This means that the tool not only provides the end-user with ready-made models that could be simulated easily, but it also provides a tool for modelling intrabay systems from scratch.

The tool developed was built using the ExtendTM simulation environment. In the development of the generic tool for modelling and simulation of intrabay AMHS, the feature of creating hierarchical blocks and adding the blocks to a library was well employed. A library of different blocks representing the different components of any intrabay material handling system was put together, providing a tool that allows rapid building and analysis of an AMHS under different operating conditions (see section 4.1.1 for details).

In the upcoming text, GTIA-M&S will be referred to as “*the tool*”, and production equipment will be referred to as “*production tools*”.

4 DEVELOPMENT OF A GENERIC TOOL FOR INTRABAY AMHS MODELLING AND SIMULATION (GTIA-M&S)

4.1 INTRODUCTION

This section aims at familiarising the reader with the ExtendTM simulation environment, data/planning documents, and the definition of some of the terms that will follow in the text.

4.1.1 The ExtendTM Simulation Environment

ExtendTM, from Imagine That, Inc., is a visual, interactive simulation tool. ExtendTM contains a built-in development system that allows the user to construct components and build custom user interfaces. Models are constructed graphically by dragging and dropping blocks (high-level model components) from library windows onto the model worksheet. Data can be entered directly into block dialogs, interactively using controls, or read from files as the simulation runs. Output is in the form of plots, histograms, tables, and customizable reports [93]. The random numbers generated by blocks supplied with ExtendTM are reproducible and synchronised (refer to 2.3.3 above), which allows use of the technique of common random numbers [53].

Customising ExtendTM

Users are keen to find better compatibility, more flexibility, and ease of use in simulation software. Furthermore, users are searching for more simulation capabilities dedicated to a specific application domain [94]. This is achieved by customising a commercial-off-the-shelf product to suit the needs and the application domain of the user; ExtendTM, in that regard, is very powerful.

The graphical interface of Extend™ can reflect the appearance of the model application. Interfaces, components, and graphics can be used which tailor the model to a specific application area. The most visible aspect of a custom model is the user interface. By modifying an existing interface or creating a new one, the simulation modeller is able to create a model that can be exercised by someone more familiar with the system than with the simulation tool. Models can be built that fit naturally into the conceptual framework of the person using the model. The following sections will describe some of the tools provided in Extend™ for customising a model [95].

Animation

Animation is a powerful presentation and debugging tool that can greatly increase model clarity. In Extend™, animation icons moving from block to block represent the flow of items through the system. Users can choose from a number of icons provided with Extend™ or create their own in an external drawing package. In addition, custom animation can be added to display pictures and text, level indicators, and pixel maps.

Figure 4.1 shows a part of a model developed for the research, illustrating the different types of animation that can be added to a model.

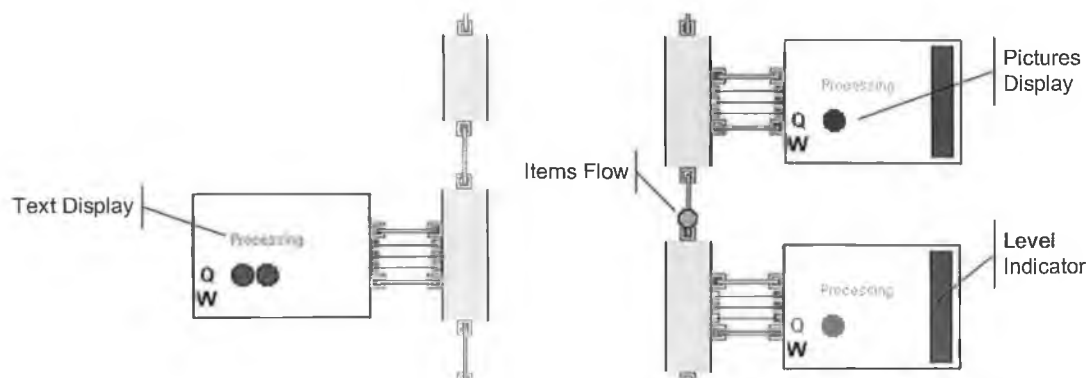


Figure 4.1: Different animation types offered by Extend™.

Hierarchical Modelling

Hierarchy allows models to be subdivided into logical components or sub-models. A single descriptive icon, similar in shape to a part of the real system, can represent each sub-model. This can greatly simplify the representation of a model and allow the

modeller to hide and show model details as appropriate for the target user, as well as creating a model with a more familiar look to the user (as shown in Figure 4.2).

Creation of hierarchical blocks is straightforward, simply selecting a group of blocks and choosing 'Make Hierarchical' from a drop menu, a section of the model can be encapsulated within a hierarchical block. The icon for the hierarchical block can be modified by using the built-in icon editor or by importing an existing picture. Finally, the number of hierarchical layers allowed in Extend™ is unlimited.

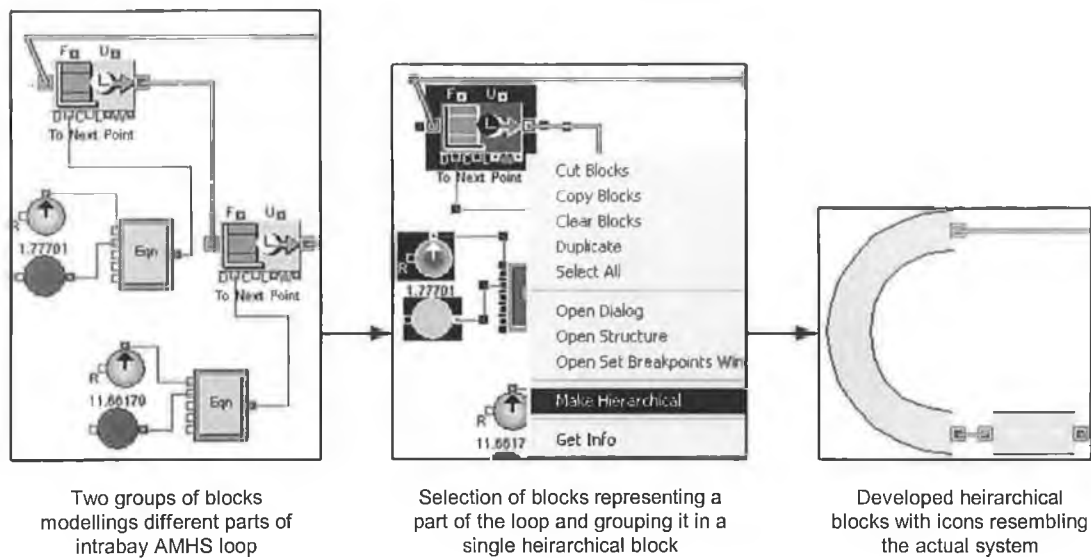


Figure 4.2: Creation of hierarchical blocks using Extend™.

Dialog Cloning and the Notebook

Input and output parameters associated with the model can be found in the dialogs of the appropriate blocks. This can make searching for specific data more difficult when working with large models containing many layers of hierarchy. An effective way of dealing with this is to use the notebook and cloning feature, as shown in Figure 4.3.

The notebook is a separate window associated with each model. Initially the notebook is a blank worksheet to which text, pictures, and clones can be added. With the notebook, a single custom interface is created for the model consolidating critical parameters and results to a central location.

Clones are direct links to dialog parameters and can be created by selecting the Cloning Tool from the tool bar and dragging a dialog parameter from a block dialog to the

notebook or model worksheet. Once a clone is created, any changes to the clone are immediately reflected in the block and visa versa.

Therefore it is no longer necessary to access the block's dialog directly to change an input parameter or view updated results. Creative use of the notebook can result in a simple yet effective interface for a large, complex model.

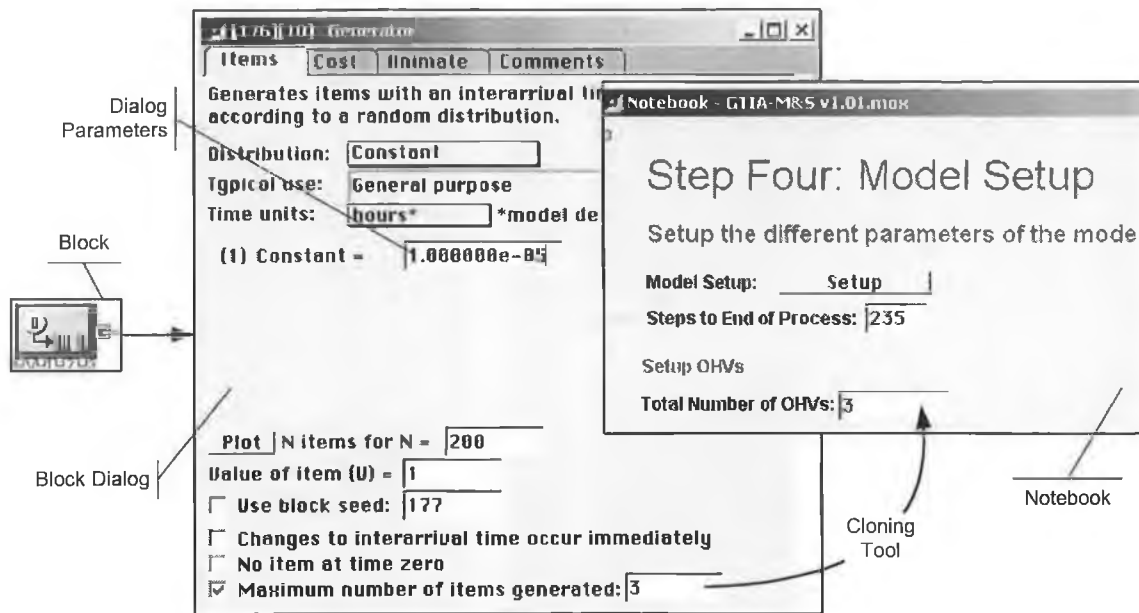


Figure 4.3: Dialog parameters cloning.

Block Development

The Extend™ block development environment provides a way for users to expand the modelling capabilities to perform unusual or highly specialized tasks. Extend™ uses a programming language called “ModL”, which is used to define the behaviour of the block. This language provides high-level functions and features while having a familiar look and feel for users with experience programming in C.

ModL language allows editing the icon, dialog, help text, and functionality of any block. In addition, the language can be used to create a new *custom block* from scratch.

Integrated Relational Database

Extend™ features an integrated relational database that provides a complete data management system for model input and output. The database is built directly into the

model to hold product data, process information, and experiment with scenarios. By separating the data from the model, the database enables fast scenario implementation, flexible analysis and improved project management.

This database is relational, and parent-child relationships can be used to better organize the information in the model. For example, each entry in a table of part types can reference its own unique routing table. This is an extremely powerful feature for organizing the information used in complex simulation models. This allows the modeller to easily modify model parameters in a central location without having to change various values that are distributed throughout the model [96].

Component-based Modelling of Reusable Models Using Extend™

Model builders can quickly build and analyze systems, with little or no programming, using Extend™ pre-built modelling components. However, Extend™ can also be used to develop new reusable modelling components [96].

The components that comprise a model (component-based modelling) can be created, reused, and even distributed to others. Component-based modelling is either by grouping existing blocks to save them as one hierarchical block, or by developing custom blocks [97]. Newly created components are then stored in a library, as shown in Figure 4.4.

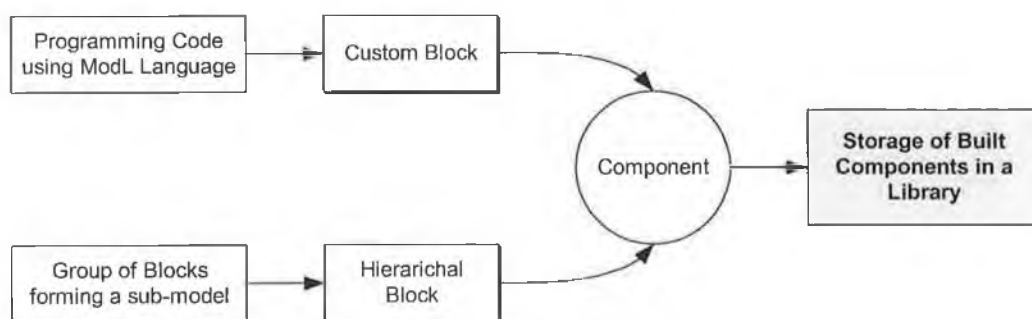


Figure 4.4: Component-based modelling of reusable components in Extend™.

4.1.2 Background Material

Most of the data used in the firm is stored in excel spreadsheets, which are sometimes used as deterministic simulation models as well.

Ramp Profile

Introduction of wafers to the fab is defined by a “ramp profile”. The ramp profile indicates the number of wafer starts per week (WSPW) for the expected product life (2-3 years).

Capacity Model

A deterministic capacity simulation model is used to determine the number of single production tools for every toolset needed to process the number of wafer starts. This model also contains detailed information about the production tools:

- Run Rate: number of wafers that can produced in one hour. This value is an averaged value for all the operations performed by a production tool.
- Number of operations performed: this indicates the number of times the same lot visits a production tool for processing (because of the re-entrant flow).
- Sampling rates: some of the production tools are metrology tools for performing parametric tests during the wafer fabrication process. Hence, sampling rates apply to these metrology tools to decide whether to sample the lot or not. Sampling rates are defined for every operation performed by a metrology tool, thus this operation can be either performed by that tool or skipped.

Process Flow

Process flow is created for every product type, which depends on the technology used for manufacturing the chip. Typically, the process flow is composed of a number of steps identified by a *step sequence number*, which are the number of steps needed to complete the manufacturing of a wafer. For each step the following is defined:

- A brief description of the operation done.
- The production tool needed for carrying out the operation.
- Whether the operation is needed for processing a wafer or testing a wafer.

Factory Layout

Layouts drawing for the fab, generated by AutoCAD, should be based on the capacity model and the process flow. These drawings show the following:

- Positioning of production tools in the different bays.
- The number of loading ports for each production tool.
- The interbay and intrabay AMHS tracks.

The production tools, number of single tools in a toolset, and their bay locations are extracted from these drawings and summarised in an excel spreadsheet.

Other

Other deterministic spreadsheet-based models include a model that predicts the amount of time needed to complete the manufacture of one lot, and another model, the ‘moves per hour’ model that predicts the expected number of lot moves in a bay per hour. Both models use data from the capacity models for defining the production tools.

4.1.3 Attributes Definition

GTIA-M&S generates two main types of entities, which are the lot and the OHV. These entities are defined throughout any model by a number of attributes, which are listed in Table 4-1. There are two main types of attributes used in GTIA-M&S:

- Regular attributes: values of these attributes can be assigned, changed, or simply read.
- Database-aware attributes (shown in *italic*): these are used to read values from the database related to the attribute defined, and it can also change parameter values in the database.

Table 4-1: List of attributes used in GTIA-M&S.

Attribute	Function
<i>Attributes Defined for Lots</i>	
AlternateBay	Defines an alternate bay in which there exists a similar tool.
Avail_LPs	Defines the total number of empty (free) loading ports for a specific production tool type in a specific bay. The number of available loading ports is incremented or decremented in the appropriate database cell defined by the production tool and bay numbers.
Bay	Defines the bay number where the lot should exist or go to.
LotProcessed	Defines the status of lot, whether it is processed or not.
ProcessingTime	Defines the processing time needed to complete a specific operation by a production tool.
SamplingRate	States whether a lot should be sampled or not. The attribute value can be either 1 (sample lot) or 0 (don't sample lot, or skip operation). An empirical table generates (0, 1) values according to a sampling rate. For process tools the attribute value is always 1, however it varies for metrology tools.
StepSeqNumber	This is a database-aware attribute, which returns different values in the 'Process Flow' table based on the step sequence number.
Stocker	Defines the stocker number.
Tool	Defines the tool number.
Waiting_Lots	Defines the number of waiting lots in each bay. It increments or decrements the number of lots awaiting collection in the appropriate database cell defined by bay number.
<i>Attributes Defined for OHVs</i>	
Collect	Defines the bay number that an OHV should collect a lot from.
OHVBay	Defines the bay for which the OHV belongs to.
OHVFull	Defines the status of an OHV, whether it is empty or carrying a lot.
OHVNumber	Identifies each OHV generated by a number.

4.2 PROBLEM FORMULATION

The problem formulation is an important step in the tool building process, in which the client and the system analyst agree on the objectives of the tool, scope and level of detail required, list of assumptions, and select the performance measures that tool must report. In addition, the time frame of using the tool is defined, as well as the tool end-user. These are all crucial points that ultimately contribute to the success of the project and are therefore discussed in details.

4.2.1 Objectives

The following objectives were agreed in discussion with production staff:

- Predict the performance of the next generation intrabay AMHS.
- Clearly identify the interaction between the intrabay AMHS and the production tools within a bay.
- Provide a tool that allows rapid model building and analysis of an AMHS under different operating conditions.
- Decide whether or not to use the merge/diverge intrabay AMHS technology.
- Find the impact of implementing different intrabay AMHS technologies on toolsets.
- Provide planning based recommendations (OHV numbers, merge/diverge...).

4.2.2 Bounds of the Problem and Tool Complexity

The size of the problem of the facility under development cannot be captured with only one model; as such model will probably end up being infeasible. Therefore, bounds must be defined for the problem under study. Furthermore, details or even components that can be removed, without losing the essence of the real system and/or losing the integrity of the model, must be overlooked.

This means that the tool developed must be complex enough to represent an abstract of the real system, but not too complex. Complexity must be considered from all different aspects. Tool scope, level of details, level of understanding, and size; are all important to accurately define the complexity of a model.

Scope of the Tool

The scope of the tool is defined by the boundaries of the problem. The tool targets the intrabay AMHS, which links production tools to stockers within a bay. In addition, the tool is developed to evaluate the merge/diverge intrabay AMHS, which connects two or more bays together. Finally, the stocker to tool delivery times and moves per hour in a

bay are the two major performance measures required to evaluate the performance of the different intrabay AMHS visions.

Therefore, it is reasonable enough to narrow down the scope of the tool to a bay or a number of linked bays, and not the entire facility. Furthermore, the production tools and stockers must be included to show clearly the handling of lots between those two components.

Level of Details

Increasing the level of details that the tool is capable of modelling greatly increases the complexity of the models developed by it.

The tool provides a level of detail that is sufficient to emulate the system under study. Some details have been overlooked to reduce the developed models complexity. However, being able to model the production tools and the material handling system at the same time elevates this.

A number of general assumptions (presented in 4.2.3 below) and other specific assumptions (presented in 3.7 above) are made to improve the level of detail the tool is capable of. Furthermore, the levels of detail of the different tool components can be categorized as level one, two, and three.

Level One

This is the lowest level of detail, in which a major component lies outside the boundaries of the problem and, hence it is not considered in the study. These include the stocker robots, labour, PGVs...etc.

Level Two

This is the next level of detail, in which a major component lies outside the boundaries of the problem, however it is still presented by a delay or a probability distribution. Examples for this are the interbay AMHS, other part of the facility not modelled (as the tool models a number of bays at a time), the internal processes of a production tool...etc.

Level Three

This presents the highest level of detail attainable by Extend™, which used for the components of absolute importance to the problem formulation. These include the flow of lots and OHVs, routing of OHVs, queuing and routing points, different delivery/collection moves...etc.

Level of Understanding

The tool presents the user with a friendly-user interface and a library of components that resembles the real system components in appearance. Because of the high user-friendly environment provided by the tool, the end-user can easily develop models *blindly* and in little time. The models that are developed, looking like the real system and coupled with animation to show the different moves of the AMHS, will offer an extremely high level of understanding to the end-user.

On the other hand, if a tool component needs reconfiguration, most of the connectors and blocks are clearly labelled to show its function. Furthermore, hierarchal blocks are used, whenever possible, to combine a number of blocks into a single block, which improves the understanding of the model.

It must be noted that using hierarchal blocks usually mislead the developer about the size of the model, because they can contain a large number of blocks and still appear as a single block. Furthermore, at several instances using multi-level hierarchal blocks can necessitate that the user opens these blocks, to the lower most level, to configure it and leading to great deal of time spent to configure a single block. Although, all components of the tool are developed using hierarchal blocks, yet they are defined with absolute care to insure that these shortcomings are overcome.

Size of Model

GTIA-M&S develops models of bay (or bays) using a number of components. The size of a model developed by GTIA-M&S depends on the size of the components, and the number of components used in the model.

The size of the components depends on the tool scope and the level of detail. The tool scope and level of detail are chosen carefully to arrive at a reasonable size of

components, and ultimately taking less time to simulate. Whereas, the number of components used in the model depends on the size of the problem addressed (number of bays, number of production tools...etc.), which the user has to define carefully before modelling it.

Finally, ExtendTM ModL language, which requires a skilled computer programmer, can be used to hard code the tool components. In that case, the components will be much smaller in size and will simulate faster, yet a great deal of flexibility and ease of understanding of these components are sacrificed.

4.2.3 Assumptions

A number of assumptions are listed here, which are mainly due to operational data and information unavailability. These are general assumptions; other more specific assumptions and rules applied to the tool are presented in 3.7 above.

- No downtimes for tools or OHVs (refer to 4.3.2 below).
- OHVs speed is constant and no acceleration or deceleration is allowed, due to the reasons discussed in section 4.5.1 below.
- Processing time on a production tool is constant. However, the processing time for the same production tool changes according to the manufacturing step.
- Interbay AMHS is factored as a uniformly distributed delay time extracted from other models developed by Intel. Also, lots pickup and drop-off at production tools and stockers are factored using uniformly distributed delay time obtained from other simulation models.
- Only one product type is allowed, therefore no prioritisation of lots. However, higher priority is given to lots with higher manufacturing step sequence number and thus are closer towards the end of manufacturing. This helps in pushing lots towards completion and exit of the system.

4.2.4 Selection of Performance Measures

Performance measures selected (in consultation with fab engineers) for this study are all specific to the intrabay AMHS. They are, typically, delivery times and the number of moves the AMHS is capable of per hour. Although the tool is capable of presenting several performance measures that provide better understanding of the behaviour of the system, yet the industry is interested only in delivery times and moves per hour.

Measures expressed as time, are reported as a mean and standard deviation. The mean and standard deviation are used to calculate the 3-sigma peak value for every measure. Presenting the performance measures as these peak values allows the evaluation of a worst case within the 99% probability level.

It should be noted that based on the International Technology Roadmap for Semiconductors [92], these measures should be expressed as 2-sigma (within the 96% probability level), rather than 3-sigma peak value. However, the system dealt with is relatively new and high levels of risk and uncertainty are associated with its operational performance, that is why the industry still prefers to use the 3-sigma as a factor of safety and to be more confident in building decisions based on the results obtained from the simulation model.

4.2.5 Tool Time Frame

The tool developed is to be used and reused regularly over substantial periods of time. Initially the design of the intrabay AMHS changes frequently, however as time passes the design of the intrabay AMHS will become more fixed, and the frequency of changes in it will reduce. When this happens the usage of this tool is expected to become occasional. However, it is recommended that the tool be used whenever design decisions are undertaken to predict the impact of those decisions on the performance of the system.

That is why the tool is built around a number of easily configurable blocks, which can be reused several times to reflect most of the design changes that occur.

4.2.6 End-User

The automation group within the firm is the major beneficiary of this tool. The group can and will use the tool for modelling any bay of interest to test the impact of design alterations on the intrabay AMHS performance. This group of people are not trained to using simulation models in their studies; instead they are well experienced with using MS Excel as a deterministic simulation tool.

Therefore, the tool targets non-simulation experts and presents them with a simple easy to use solution for testing different intrabay AMHS configurations. In addition, the tool uses Excel spreadsheets as an optional data input/output interface to suite the end-user preference.

4.3 COLLECTION AND PROCESSING OF REAL SYSTEM DATA

This step presents the major constraint to the progress of the project. The amount of time and effort that go into identifying, specifying, collecting, synthesizing, and maintaining the data used in modelling efforts is tremendous. In addition, the lack of perceived value of these efforts by semiconductor management doesn't promise improvement in that regard (for example, by making data available in the right format needed for simulation efforts), as confirmed by the statement:

“data quality and availability is a tremendous problem, much of that we own internally” [98].

4.3.1 Data Requirements

GTIA-M&S requires a considerable amount of data that are found in a number of sources, and in different formats. As mentioned before, the corporation has other simulation capabilities in different sites. These capabilities are well established and most of the time are a good source of data. In fact, some of these capabilities are built for the same process that GTIA-M&S is modelling. Therefore, most of the data related to the process are extracted from these models and used in the tool.

Ramp Profile

The ramp profile is a spreadsheet that shows the weekly WSPW, and can be easily retrieved. It is used in GTIA-M&S to calculate the lot inter-arrival rate, which is the time needed to introduce a lot to a model. The mean value of the inter-arrival rate can be calculated based on the WSPW as follows:

$$\text{Lot Inter-Arrival Rate} = 25 * 168 * (1 / \text{WSPW}) \quad \text{hr/lot.}$$

Stockers

Number, capacity, and ID of stockers belonging to every bay in the factory need to be identified to determine where the lots are allowed to queue. These can be retrieved from the layout and the moves per hour model.

Lots sent to a bay having a number of stockers are distributed over the stockers with a specified proportion between the stockers in the bay. This proportion, which is extracted from the “moves per hour” model, is used to determine which stocker should receive the lot.

Bays

A list of all bays in the fab is required as well as the stockers belonging to each bay. The bay play an important in identifying the location of GTIA-M&S components, even most data collected from models developed are per bay. The data related to the bays was found in the layout drawings, and from the other simulation models developed by the industry partner.

Production Tools

This represents all the data needed to define the production tools used in the fab, and it includes:

- Production tool list showing all the types of toolsets used in the fab. This has been retrieved from another simulation model.
- Bay location of every production tool, which is retrieved form the layout drawings.

- Number of single tools in a toolset, which can be found in more than one source as the capacity models, layout drawings, moves per hour model...
- Number of loading ports per production tool, which can also be found in more than one source such as the layout drawings, moves per hour model...

Processing time for production tools and sampling rate for metrology tools are also data related to production tools, however they are defined in the process flow.

Process Flow

The process flow is one of the most important requirements needed by the tool, as it determines the flow of lots in a developed model. The process flow used in GTIA-M&S is a modified one, as illustrated in Table 4-2. For every step sequence number the following additional information is defined:

- Bay number where the production tool belongs.
- Alternate bay number(s) if the production tool has duplicates in other bays.
- Processing time in hours, which the time needed to process one lot. This value is derived from the run rate per layer and not from the average run rate.
- Sampling rate for production tools for every process step, meaning that for metrology tools a process step can be either performed or skipped.

Table 4-2: Sample of the modified process flow used in GTIA-M&S.

Step Seq. No.	Step Name	Tool	Type	Additional Information			
				Bay	Alternate Bay	Processing Time	Sampling Rate
10			Process/ Metrology				

The first part of the process flow (step sequence number, step name, tool, and tool type) can be easily retrieved as a spreadsheet. Bay and alternate bays are responsible for the routing of lots, which determines where the lot should be located at any stage of the process, and are based on analysis of the process flow, layout, and intuition. Run rate per layer (wafer/hour) is retrieved from other simulation models and is converted into

the processing time (hour/lot), Processing Time = 25/Run Rate. Finally, the sampling rate is retrieved from other simulation models as well.

Data Related to the AMHS

GTIA-M&S builds models of a part of the fab resembling the actual system by laying out the AMHS track, production tools, and stockers. Thus, the most important data required to define the AMHS is the distances between the different components. This can be retrieved from the layout drawings, which is the only source for these distances.

Other data related to the intrabay AMHS are needed to defined the performance of the AMHS, these are typically the OHV speeds on straight and curved tracks, and the amount of time needed to pickup and drop-off lots at production tools ad stockers. These values are retrieved from the output of simulation models within the cooperation.

4.3.2 Identifying the Sources of Data

The sources of data and their expected format were identified during development of the first simulation model, and then used for GTIA-M&S, as shown in Figure 4.5.

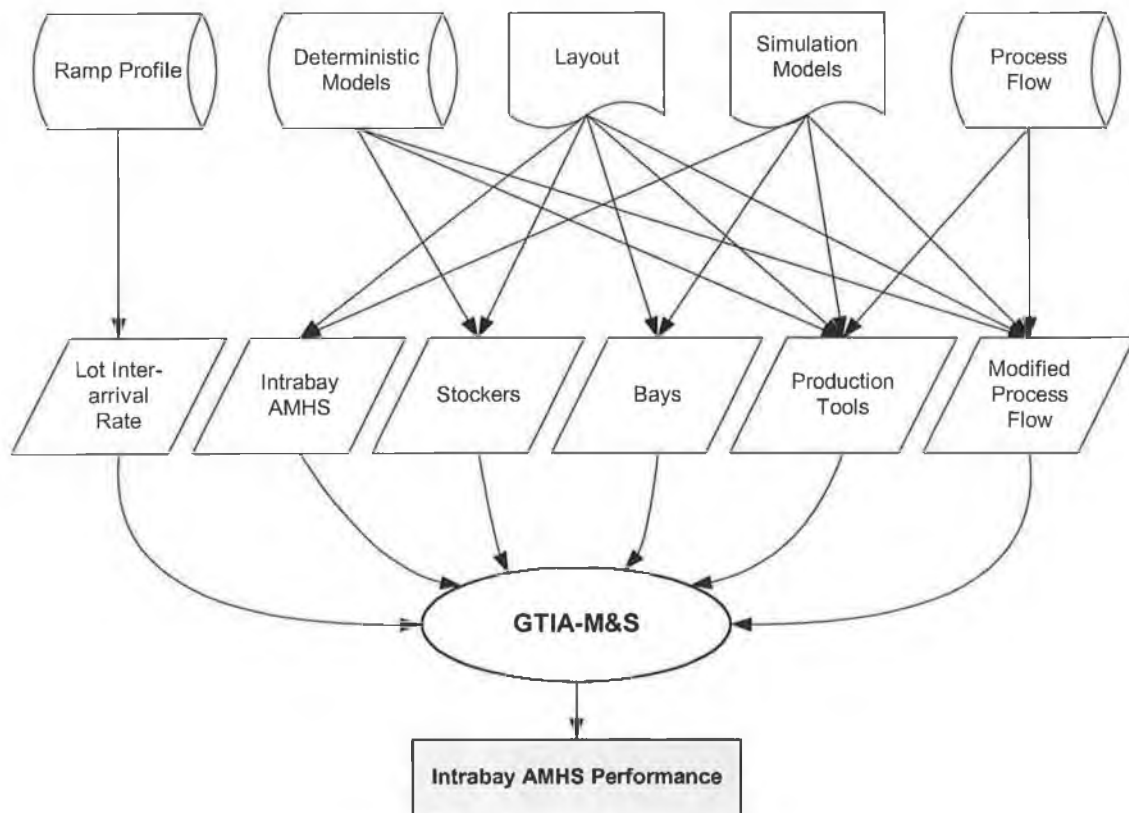


Figure 4.5: Sources of data.

The above figure clearly shows that in most cases, there are several sources for a single data item, therefore, the data are first validated, and then used in the tool. Validation of data is mostly carried out with the aid of excel spreadsheets. Data from different sources are entered in a single spreadsheet, put in the same format, and are then compared against each other. This comparison can sometimes be with the aid of some formula that automatically identifies any discrepancies between the data values, or it can be manually checking every cell in the spreadsheet to point out these differences. Any inconsistency in data is reported back to the source in order to understand the reason, valid values are then identified and used in the tool.

Inconsistency of data values is not the only identifiable data problem, however other problems have been faced and are reported in the next section.

4.3.3 Collection of Information and Data

Information about the operating procedures of the system under study was collected, even before collection of data. Meetings with people from the firm were held weekly to gather all the possible information. This was one of the great hurdles because the system is still under development and there remains a general confusion and misconception about how the system actually operates. A piece of information collected one week could be overturned by new information within a couple of weeks.

For example, the intention to use the merge/diverge capabilities for the intrabay AMHS was unknown for almost nine months. Furthermore, after being informed of the merge/diverge technology, there was as yet no firm decision about whether or not to implement it. Hence, the GTIA-M&S is developed with the capability to switch between the two technologies.

Updating of Data

The first major problem experienced with data is the need for continuous updating. Data collected, checked, processed, and used in the tool, can sometimes remain valid for a couple of weeks only. This is because the system is still under development. Major updates to all the data take place every three months, and minor changes are common during these three months.

Data Inconsistencies

Similar data are found in more than one source, which means that there is duplication of data. The major concern is data inconsistencies that occur whenever an update of data takes place, as data changed in one document is not reflected in the other.

Data about production tools is one of the data that suffer from great inconsistency. Production tools are found in the layout, in the process flow, and in the production tools capacity model. The production tool list, which was never found on its own, changes every now and then. A production tool may be removed, added, or simply renamed and only parts of these changes are captured by the different documents.

For example, a production toolset appears on the layout containing four tools but appears in the capacity model containing three tools, and doesn't appear at all in the process flow file. When this conflict in data was reported, the problem was rooted to the update of data. The layout document was updated first, and then the number of tools changed and only later the capacity model changed. Finally, the process flow file was the most up-to-date document and didn't appear to show the production tool simply because the production tool ID had been changed. This dramatic example is common and unless the system analyst is used to such inconsistencies and how to deal with them, data collection becomes impossible.

Data Requirement Confusion

This is another problem experienced during the research work, which actually delayed model development at one stage for a couple of months. Confusion occurs when the system analysts requires data about a specific process, and the client supplies this data in a *processed* form. The cause of this problem is that the client is not sure about how the data is used in the tool development. Furthermore, the system analyst, being inexperienced with the manufacturing process, is unable to determine if the data values are reasonable or not.

For example, the analyst asks for the process flow showing all the manufacturing steps and the production tool needed at each step in addition to the processing time at each step. The client supplies a document with all the required data; however, the processing times were already synthesised to predict the amount of time needed to completely

manufacture a lot (to predict the cycle time). This data, being processed, is supposed to be the output of a simulation model and not the input. However, the analyst feeds the data to the model, which shows that lots cannot be finished even in two years. At first, the problem is thought to be with the model developed. Only after other documents were supplied showing different values for the processing times was the problem identified and the problem traced back to original document.

Data Unavailability

Data related to several aspects of the facility is either missing, or cannot be retrieved. Examples of these are:

- Metrology tools processing times and run rates are missing, and data from other models are used.
- Process flow shows the production tool required for each manufacturing step, however, it does not show the location of that tool. Most of the time there exists duplicate production tools in more than one bay. Choosing the bay for the manufacturing step is based on intuition and analysis of process flow and facility layout.
- OHV breakdown and maintenance times are not available, because the system is still under development. A development site that tests the new intrabay AMHS could be a good source of data, yet they are reluctant to release much information as the system is still considered as a “work-in-progress”. In addition, even if the data is released, it will not be representative of a high volume manufacturing facility.
- Production tools breakdown and maintenance *raw data* were missing. Production tools for the 300mm technology are present in development sites, and because this is not a typical high volume manufacturing scenario, data extracted from these production tools was sometimes inaccurate.
- Processing time of a lot on the same production tool differs depending on the manufacturing step. Processing times were available as an average for all the manufacturing steps performed by a production tool. However, processing time

per tool per manufacturing step is needed for the purpose of this study, which was unavailable and again must be extracted from another model.

Data Format

Most of the data collected are not ready to be used directly in a simulation model. Some of this data needs processing prior to being used. For example, run rates of production tools are converted to processing time, and wafer starts per week are converted into lot inter-arrival times. Other data may exist only in hard copy and is not available electronically. This is evident in the distances travelled by OHVs, which is a key input needed for modelling any AMHS. There is no single document that lists the distances between the different production tools, lengths of bays and so on. However, every major part of the factory, for example a bay or a functional area, is available in a separate AutoCAD factory layout electronic document.

4.3.4 Handling the Different Data Problems

In time the methodology applied in creating the GTIA-M&S overcame most of the data problems presented above.

Specifically, the GTIA-M&S uses a central database that helps in remotely changing almost every block in a model developed by the tool. Furthermore, the generic nature and flexibility of the tool adapts to these changes easily and quickly. With the first model, every time the data changed, a great deal of time was needed to incorporate the changes into the model.

GTIA-M&S handles the issue of routing lots, where duplicate production tools exist, by introducing an extra parameter, which defines alternate bays for each production tool, indicating the presence of a duplicate tool in other bays.

To overcome the format of factory layout problem, the different files, for the set of bays modelled by GTIA-M&S are merged into a single drawing, which shows the stockers, production tools, and intrabay AMHS track. A different layer is defined in the AutoCAD file that is superimposed over the intrabay AMHS track. Circles are used as placeholders at the input and output ports of the stockers and the loading ports of the

production tools, as shown in Figure 4.6. All the layers are then deleted, except for the created one, which is saved in a DXF format.

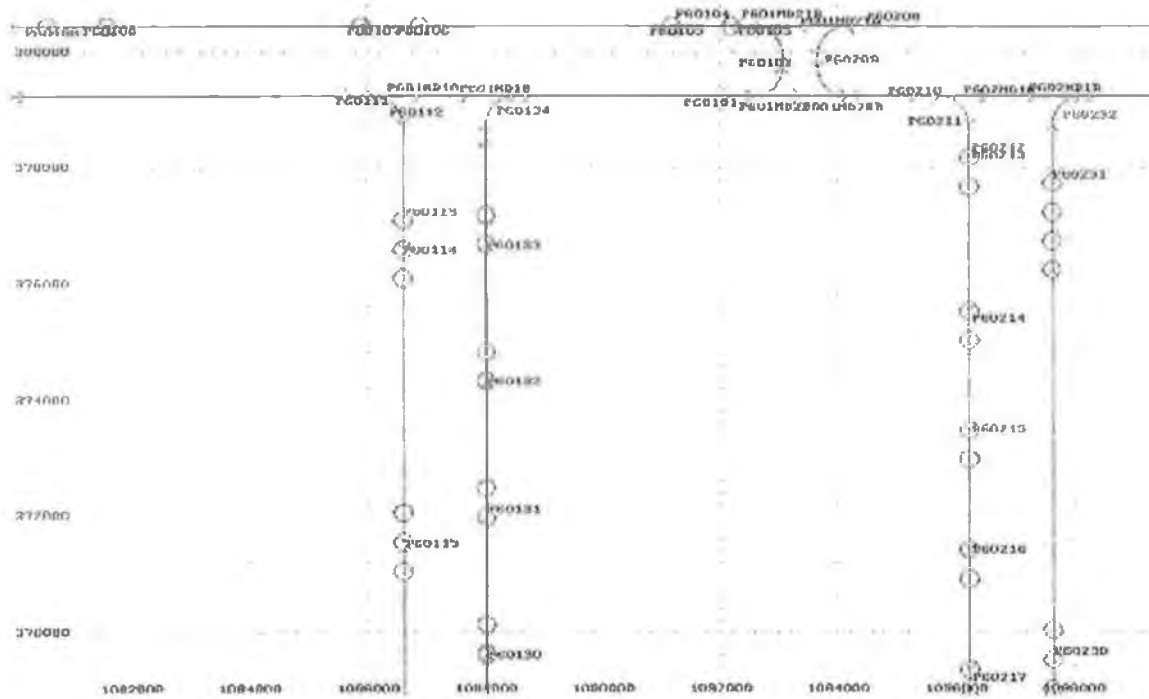


Figure 4.6: Created AutoCAD layer and imported by Proof Animation™.

Proof Animation™, included in the Extend™ simulation package, is a 2.5D animator that provides simultaneous animation during the simulation run [99]. Proof Animation™ can import DXF files for animation purposes, however, it can be used for reporting the absolute distances between any two points accurately by simply connecting these two points. Therefore, Proof Animation™ is used to import the DXF file of the created layer, and the absolute distances between the different placeholders are reported by connecting them. Finally, the distances are fed to an excel file to perform any required calculations (changing the units for example).

4.3.5 Probability Distributions for Random Variables

GTIA-M&S uses different variable types when running developed models, as listed in Table 4-3. Fixed variables, which have a value that remains constant over the whole simulation run, such as distances and OHVs speed.

Other variables value change as the simulation run progresses, these variables can be either dependent or independent. The value of a dependent variable, such as processing time, changes based on the value of another variable, which is usually an independent one, such as the step sequence number.

Finally, a number of stochastic variables are used in the tool to represent the randomness of these variables, for example:

- On the basis of common usage with available models and to produce comparable results, an exponential distribution is used for the arrival rate of entities. This is a standard method used when arrivals occur “at random” [100].
- Uniform distributions are used for interbay delays and pickup and drop-off times at production tools and stockers, which are extracted from another simulation model used by the corporation.
- Variables with limited possible values and known probability for each value, empirical distributions are used such as the sampling rate and preferred stocker (discussed in section 4.4.5 below).

Table 4-3: Different variable types used in GTIA-M&S.

Variable	Description	Example
Fixed	Value of variable does not change over the whole simulation run.	Distances. Speed of OHVs.
Dynamic – Independent	Value of variable changes during the simulation run, and does not depend on other variables.	Step Sequence Number.
Dynamic – Dependant	Value of variable changes during the simulation, and depends on the value of other variables.	Processing Time.
Stochastic	Variable value exhibits a random nature.	Inter-arrival Time

4.3.6 Choosing the User Input Interface

Direct data input cannot be used because of the large number of blocks used, and manual configuration of every single block is not practical. Furthermore, a large amount of data is necessary to run the tool, for example the process flow is composed of around 500 steps each comprising a number of data about several parameters.

ExtendTM offers four different methods for data input in addition to direct data input in a block and these have been critically evaluated through tests carried out on the first simulation model. These methods are:

1. The notebook of ExtendTM.
2. The inter-process communication (IPC) library.
3. Global arrays.
4. The SDI database.

ExtendTM Notebook

ExtendTM notebook can be used for controlling model parameters, reporting simulation results, and documenting the model. This feature offers great flexibility as notebook dialogs replicate local block dialogs. As such they are not suitable for entering large quantities of data. Dialog items control the values of the different parameters defining a block. Any dialog item can be *cloned* and inserted in the notebook.

The notebook has been used for a small amount of data entry, mainly for setting the simulation parameters.

IPC Library

The IPC library is a special library that comes with ExtendTM for reading/writing from/to excel files in real time. The major problem with these blocks is that it remains connected to the excel file throughout the whole simulation run, and increases the simulation runtime drastically.

Any advantages offered by this library are abolished by the slow execution of models, and thus it is not suitable for models as large as those developed using the tool.

Global Arrays

Global arrays are general-purpose arrays available anywhere in the model. In addition, these blocks can be configured to access excel files using the Windows ODBC (Open Database Connectivity).

Lack of user-friendliness is the major drawback of this method. The arrays rows and columns cannot be renamed, which is a problem if such an array is large in size, and the user will not be able to identify the data required in this array. Furthermore, configuring the global arrays to access excel files is not straightforward. For example, the excel file name and location must be written manually. The user is prompted that an error occurred (and doesn't mention that the error is simply in the file name), if any mistake in writing the file name and location exists.

Another drawback of this method, as well as the IPC method, is that developed models will not operate without the excel file making it difficult to transfer these models.

SDI Database

This is the most suitable way of data management, provided by ExtendTM, which can be achieved in GTIA-M&S. The SDI database handles the different problems found with the other data input/output interfaces, and provides several other advantages as well.

The SDI database is completely integrated with ExtendTM such that the database is embedded in the tool. Although, the database management system includes data manipulation features in both ExtendTM and Excel, yet it doesn't need the excel file for running the models developed by the tool.

The database is supplied with a number of blocks for writing/extracting data to/from the different components of the tool. Thus, almost any part of any component can read data from the database and can report results as well to the database.

Any change in a section of the database means a change in all the components using that section, which means the effect is global. This offers great flexibility in configuring the different components of the tool, and provides the end-user with a straightforward method of experimenting with the models developed under different operating conditions. For example, changing the OHV speed on straight and curved parts of the AMHS track is achieved by changing the speed parameters once in the database, without the need to change every track segment modelled.

A random number generator is built into the database, which is used by different components of the tool, to generate parameter values according to their various

distributions. For example, sampling rate, pickup and drop-off times, and interbay AMHS delays.

Finally, the database can categorise the different data into groups (Figure 4.7), which helps the end-user in finding the data that he/she is looking for easily. This is especially important because of the large amount of data that the tool is dealing with. For example, grouping the different outputs of the model under 'Simulation Results' helps the user to locate the output of a simulation run quickly.

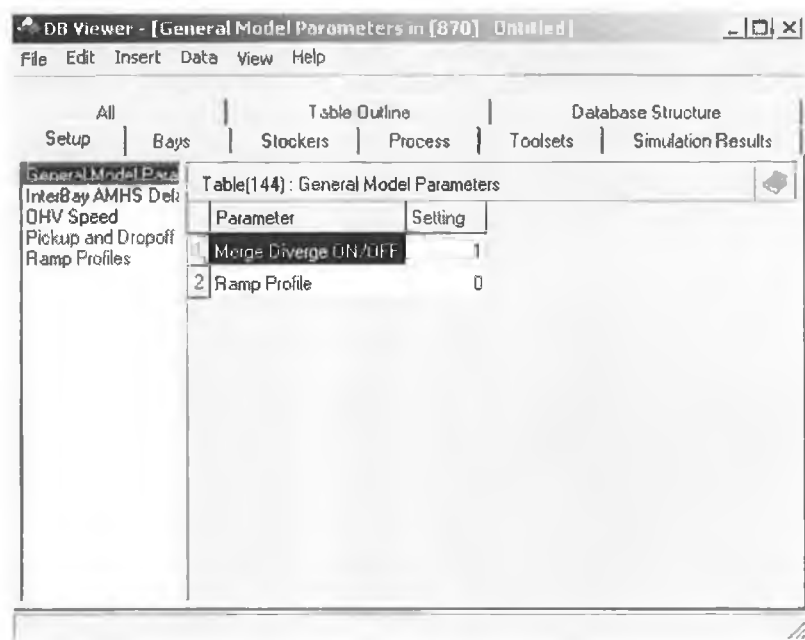


Figure 4.7: Different data categories defined in the database.

4.3.7 Building the Database

The database is composed of several tables; each table has a specific number of rows and columns. Every major subset of data is collected in a table, and the number of rows and columns corresponds to the amount of data in that subset.

First, the basic data, such as the bays and production tools lists, are entered in the database each in a table. Then the repeated data from different tables are cross-referenced, as shown in Figure 4.8. For example, the process flow table has a production tool and a bay column, which are referenced from the production tools list and bays list. Referencing data in one table to data in another table is called indexing, which allows the end-user to deal with 'string' rather than 'integer' values when configuring the tool

components. For example, when the end-user defines a stocker it is better to select the name of the stocker, rather than entering a number representing that stocker ID number.

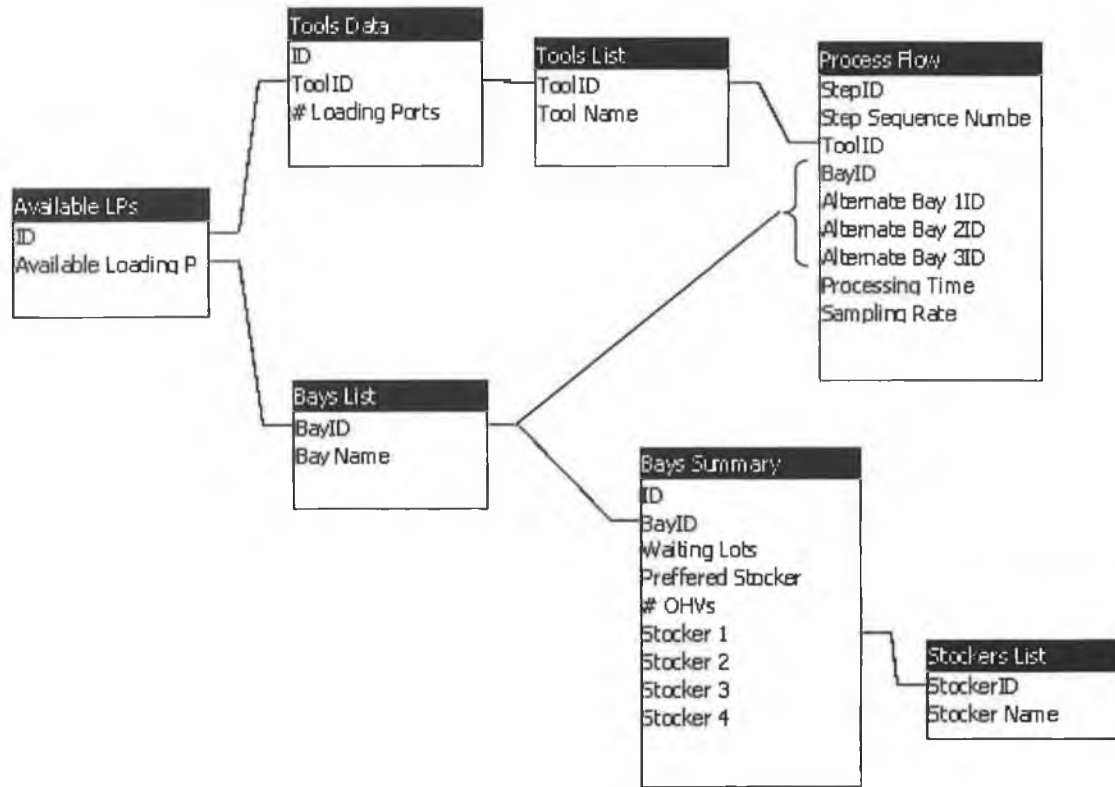


Figure 4.8: Referencing data in one table to data in another table by indexing.

Data, such as bay and stocker lists, can be directly imported from excel spreadsheets. In addition, manual data entry into the database is possible either by entering data directly into a database cell or by selecting the value from a list of possible entries.

Finally, after all the collected data are fed into the database, any redundant data is identified and deleted at the end of the project in the tool refinement step (refer to 4.5.4 below).

4.3.8 System Performance Data

System performance data is live data extracted from the real system, after running with a particular configuration, and used for validation purposes. Unfortunately, this type of data is not available because the real system is still under development, and there is, as yet, no similar system implemented elsewhere. However, other techniques are used to validate the GTIA-M&S and are discussed in 4.7 below.

4.4 TOOL CONCEPTUALISATION

The tool conceptualisation started by the basic functions required for modelling the AMHS. Different approaches to modelling the intrabay AMHS are evaluated, then the flow of material (lots being handled) are modelled. Afterwards, controlling the flow of material and vehicles is modelled. The different components needed by the tool to represent a part of the real system are then modelled, and the final step is to define the interaction between those components. The interaction between the components is identified through IDEF0 modelling, which is presented at the end of this section.

4.4.1 Intrabay AMHS

As mentioned before, the intrabay AMHS is composed of a track and a number of vehicles. The track is modelled as short route segments connected together, each segment with a specific length and capacity (maximum number of OHVs it can accommodate). The start and end of these routes are usually input or output ports of production tools or stockers.

The vehicles are represented as entities (rather than resources). Joining and splitting an OHV entity and a lot entity realises the lot pickup and drop-off, respectively.

4.4.2 Re-entrant Material Flow

A common characteristic of the semiconductor manufacturing process is re-entrant material flow, which means that the same product goes through the same production tool several times finishing a manufacturing step. Each product requires a number of manufacturing steps, dictated by the process flow, for completion.

For each lot, defining an attribute called the *step sequence number* captures this characteristic. The step sequence number plays a key role in defining all processing requirements. Using that attribute the production tool required for processing the lot is identified, the bay it is found in, the processing time, and the sampling rate. Furthermore, the step sequence number attribute value is incremented by one, whenever the lot finishes a manufacturing step, indicating the completion of that step. Finally, the

lot exits when the attribute value reaches the value of the last manufacturing step, which is when the lot is completed.

As mentioned before, GTIA-M&S is used for developing models of a bay or a number of bays and not a single model for the entire facility. However, the lots require manufacturing steps in all bays and not only the bays modelled. This issue is overcome by building a delay module responsible for delaying lots for a specific amount of time, which is equal to the amount of processing time spent by the lot in the remaining part of the facility. A delay is applicable whenever a lot leaves the modelled bays and returns, which may occur several times for a lot.

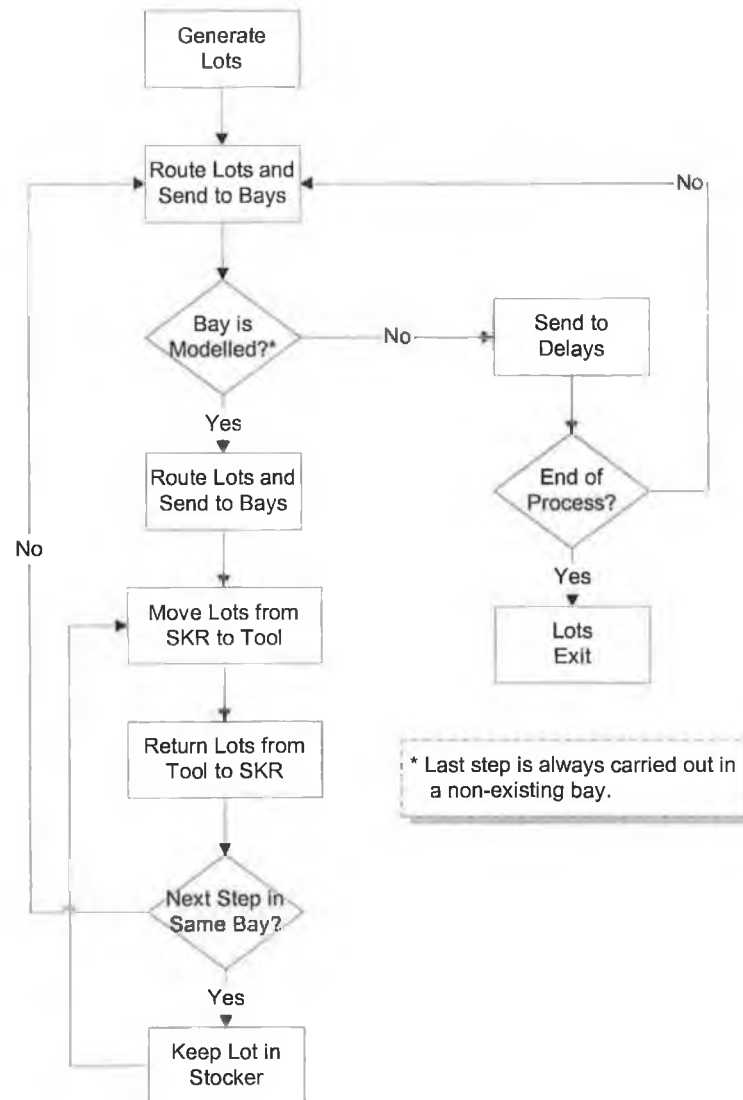


Figure 4.9: Material flow within GTIA-M&S.

These two concepts, re-entrant flow and delaying of lots, are illustrated by Figure 4.9. After generation of lots and routing these lots to bays, a check is made to determine whether the bay is modelled or not. If the bay is not modelled, the lot is sent to the delays module. After delaying the lot, the step sequence number is incremented and if it is equal to the last manufacturing step, the lot exits, otherwise it is routed again to another bay.

An additional last step is added to the process flow, which is always carried out in a non-existing bay. This guarantees that the lot will always go the delays module to check if it has completed all the required steps or not, and that at the end of the manufacturing process the lot can be released.

If the bay is modelled, the lot is sent to a stocker in that bay, moved from stocker to production tool, processed, and returned to stocker. If the next manufacturing step is in the same bay, the lot remains in the stocker, otherwise it is routed to another bay.

4.4.3 Generation of Different Entities

In general, there are two types of entities generated and fed into the models developed these are the lots and the OHVs. Both entities require routing to the right location in a model. Lots are routed to bays then stockers in these bays, and OHVs are routed to a number of OHV entry points along the intrabay AMHS track.

Lot Generation and Routing

Generation of lots, routing, and sending to the different bays depend on several conditions as shown in Figure 4.10. Inter-arrival rate, hours per lot, is calculated based on the ramp profile for timely lot generation. For every generated lot the step sequence number is defined, which is first used to retrieve the sampling rate from the database. If the lot will not be sampled, the step sequence number is incremented by one, and the sampling check is performed again. Otherwise, routing to bays continues.

When a production tool exists in more than one bay, it is said to have alternate bays. Alternate bays are retrieved from the database based on the step sequence number. If alternate bays exist, the production tool (in the original bay) and the number of available loading ports of that tool are retrieved. If the production tool has no available loading

ports, the rest of the production tools (duplicate ones in alternate bays) are checked for loading ports availability. The bay that collectively has the maximum number of available loading ports of the tool, is the bay selected for the next step and the lot is sent to that bay.

If such bays do not exist, the bay number is retrieved from the database, and the lots are sent to that bay.

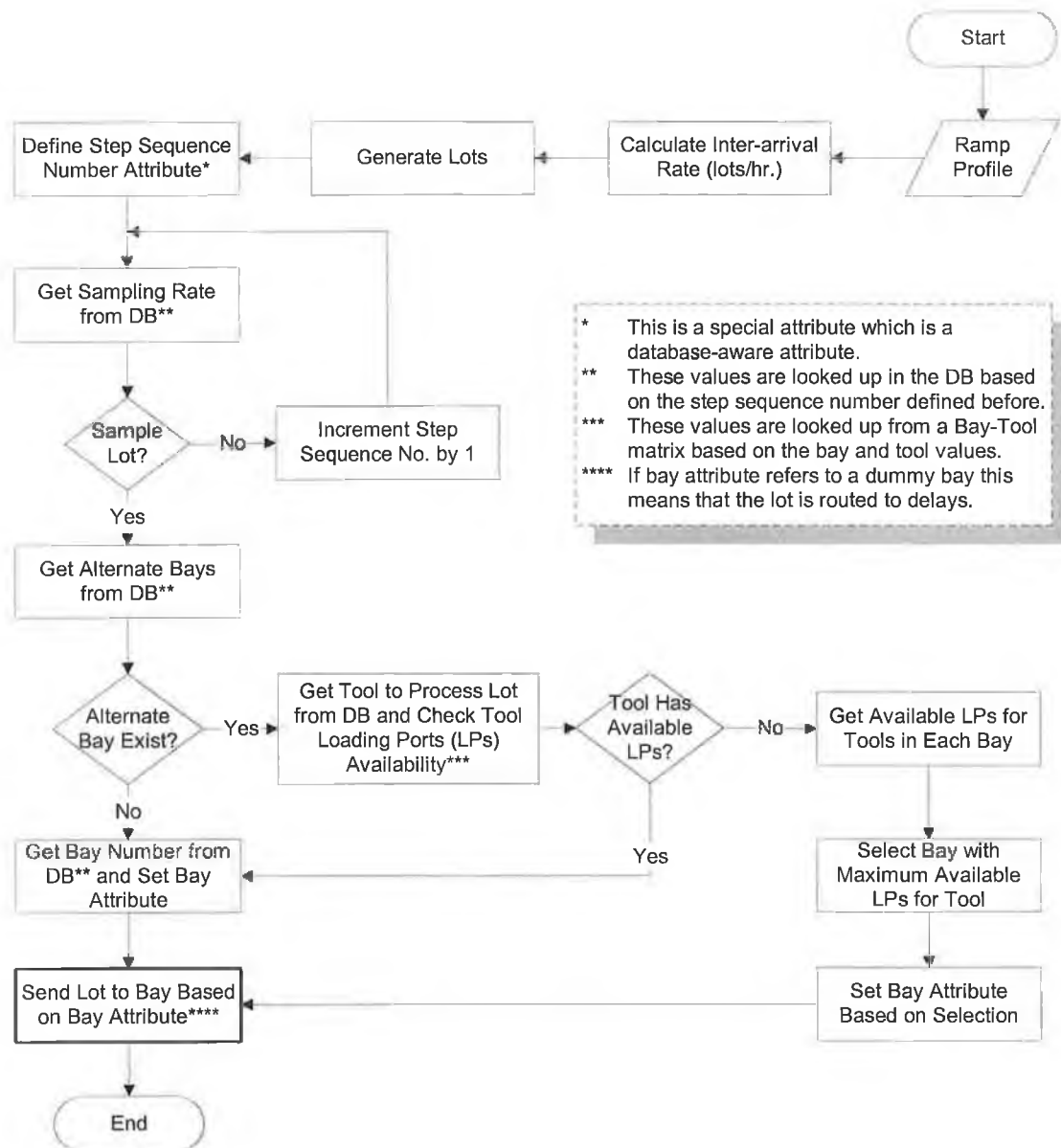


Figure 4.10: Lots generation and Routing.

NB: The bold outline around the “Send Lot to Bay Based on Bay Attribute****” box indicates that this is the final outcome of the decision process. The same is true in all subsequent flowcharts.

OHVs Generation and Routing

Timely generation of OHV depends on the number of OHVs required as well as the OHVs inter-arrival rate. Introduction of OHVs occurs at different entry points, which depends on the merge/diverge status, as shown in Figure 4.11.

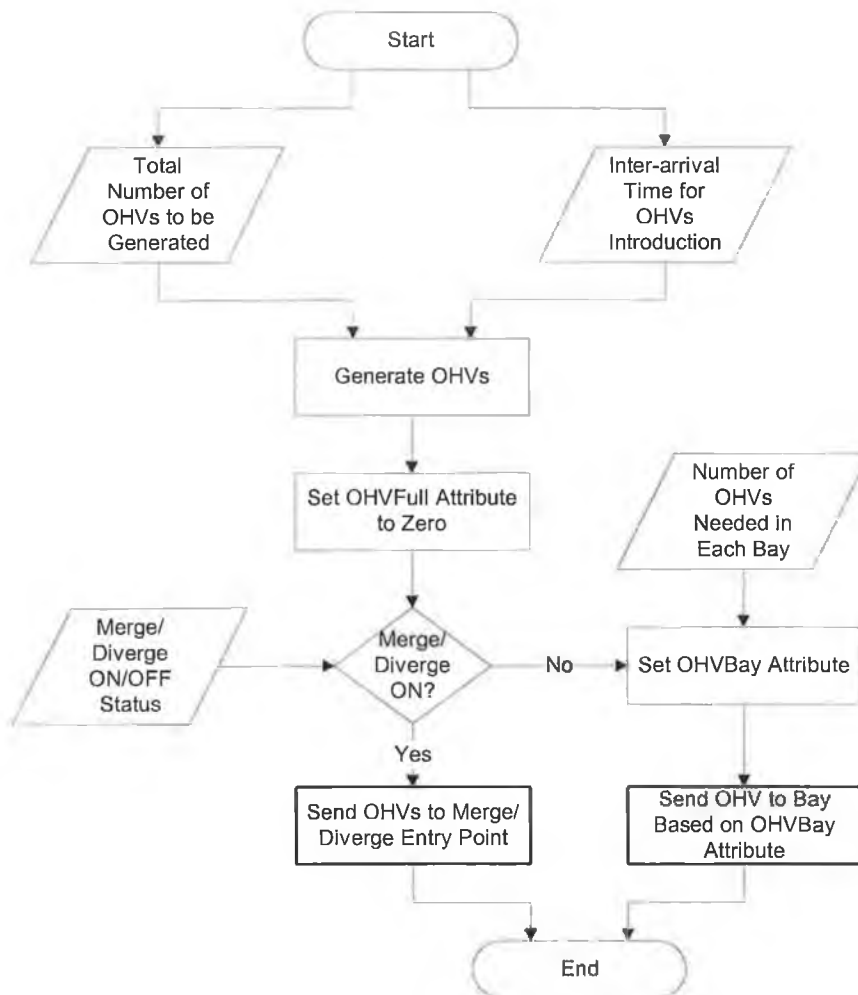


Figure 4.11: OHVs generation and routing.

If the merge/diverge is ON, all the OHVs are sent to a single entry point for all bays in the merge/diverge link. Otherwise, the number of OHVs assigned in each bay is determined and these OHVs are introduced through entry points, each located in a bay.

4.4.4 Decision Making Points

To control the flow of entities in the model, whether it is a lot or an OHV, several decision-making points are chosen at different places. These points, shown in Figure

4.12, either set a specific attribute that is used later in decision-making, or route the entity to a required position.

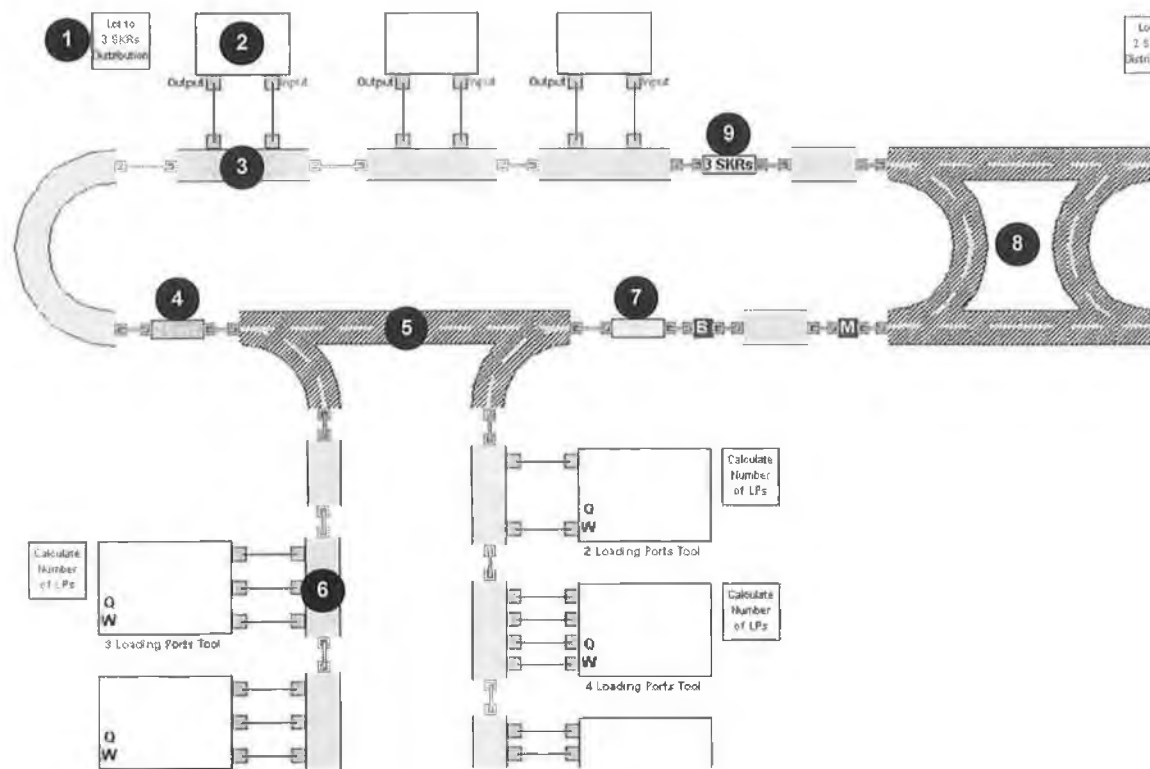


Figure 4.12: Part of a model constructed using GTIA-M&S showing different decision points.

Table 4-4 lists the decision-making points location, purpose, and outcome, which are presented next in details.

Table 4-4: List of the different decision points.

No.	Location	Purpose	Outcome
1	Before Stockers	Sending Lots to Stockers	Stocker to Receive Lot
2	At Stockers	Storing Lots in Stockers	Store or Release Lot
3	Route Facing Stocker	Drop-off/Pickup Decisions	Lot Drop-off/Pickup
4	Before MD Link 1	Setting Collect Attribute	Bay to Collect Lot From
5	MD Link 1	OHVs Routing Decisions	Diverge Into Bay or Not
6	Route Facing Tool	Drop-off/Pickup Decisions	Lot Drop-off/Pickup
7	Before MD Link 2	Setting Bay Attribute	Bay to Send OHV to
8	MD Link 2	OHVs Routing Decisions	Move on Outer Loop or Not
9	After MD Link 2	Select SKR for Next Move	Stocker Attribute

Decision Point 1: Sending Lots to Stockers

This decision point is responsible for distributing the lots sent to the bay over the stockers in that bay. The process starts by generating a random stocker number based on an empirical distribution, as illustrated by the flowchart in Figure 4.13. The stocker number is called the *preferred* stocker number, and the next step is to check the capacity of the preferred stocker. If it has enough capacity to accept the lot, the lot is assigned a stocker attribute value equal to the preferred stocker number. Otherwise, the stocker holding the minimum number of lots is identified and the stocker attribute value is equal to the identified stocker number. The lot is then sent to the stocker based on the stocker attribute value.

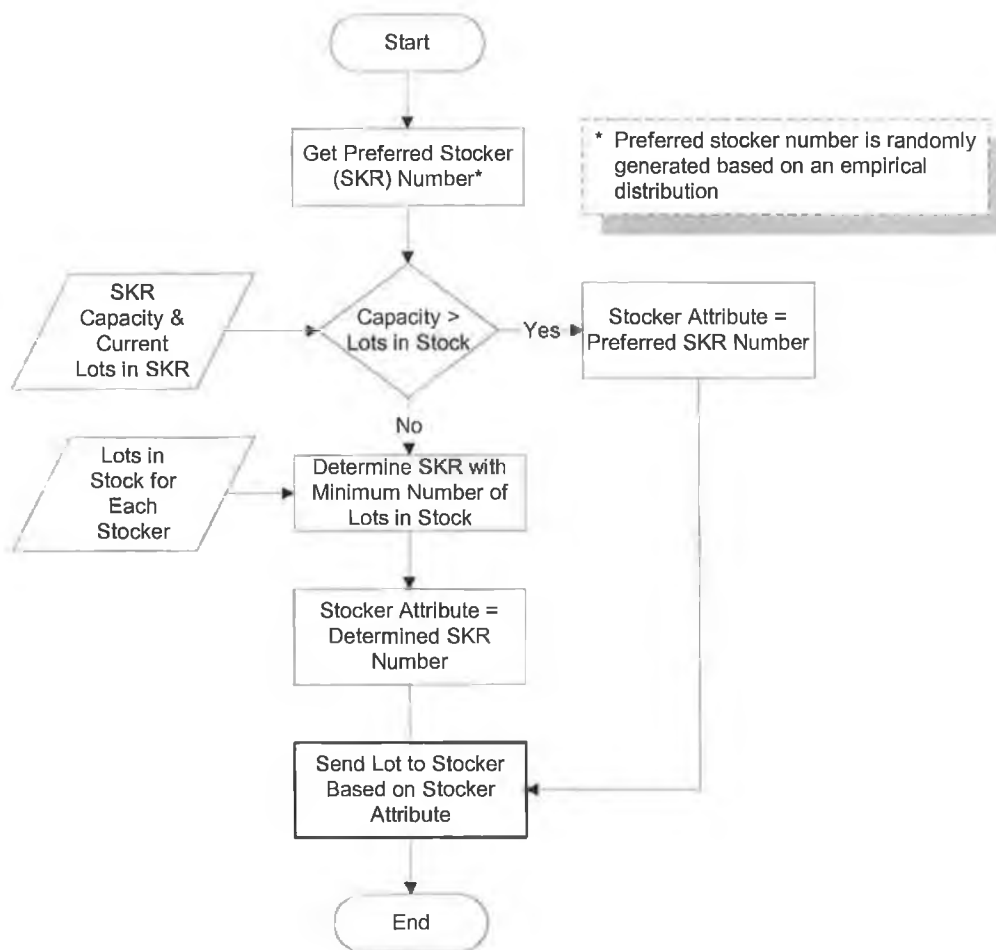


Figure 4.13: Distribution of lots sent to a bay over its stockers.

Decision Point 2: Storing Lots in Stockers

Figure 4.14 shows that lots sent to a stocker, in the previous decision point, are either prepared for pickup by OHVs immediately, or stored in that stocker. Lots are stored (queued) in stockers when the production tool, required to process these lots, is not available.

Lots are queued in designated queues. For every production tool in the bay, a queue exists in which lots are allowed to wait for the production tool to become free. Whenever a loading port of a production tool is free, the queue corresponding to that production tool releases a lot, which is prepared for pickup.

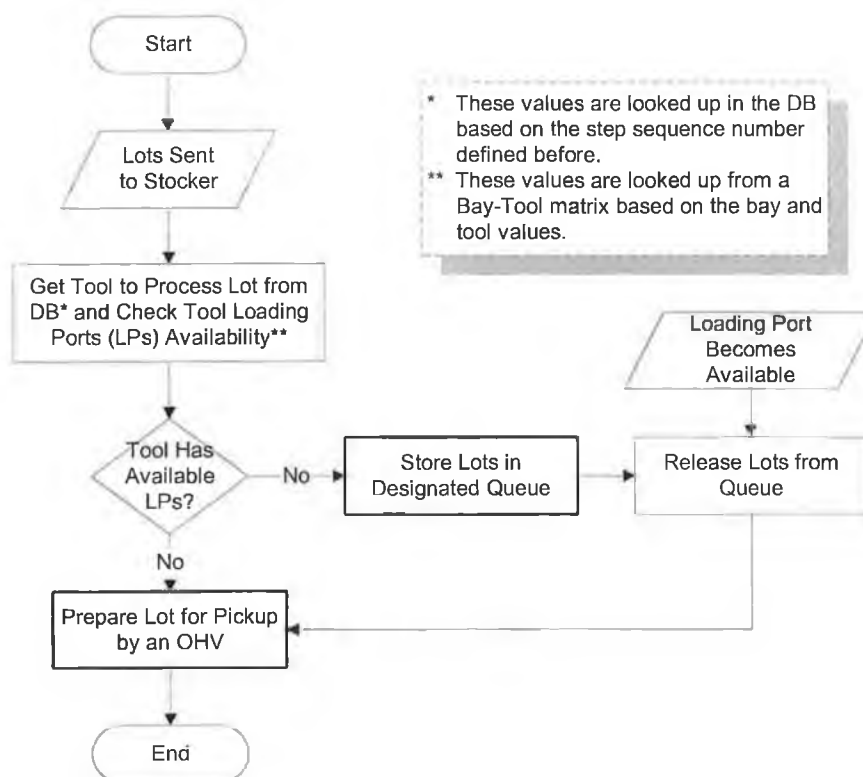


Figure 4.14: Deciding whether or not to store lots in a stocker.

Decision Point 3: Drop-off/Pickup at Stockers

Every stocker has an input port and an output port. The OHV always passes by the input port first then the output port making the OHV capable of two moves at all stockers.

Consequently, at each stocker two decisions are made; whether or not to drop-off of a lot, and then whether or not to pickup another lot, as shown in Figure 4.15.

The OHV drops off a lot given that it is full, the stocker is not full, the lot is processed, and the stocker attribute value (determined at decision point 9) equals the stocker number; otherwise, the OHV takes no action at the stocker input port and moves to the stocker output port.

The OHV picks up a lot given that it is empty, there is a lot waiting collection, and the stocker attribute value equals the stocker number (stocker attribute value is assigned to the OHV at decision point 9); otherwise, the OHV takes no action at the stocker output port and moves to the next stocker.

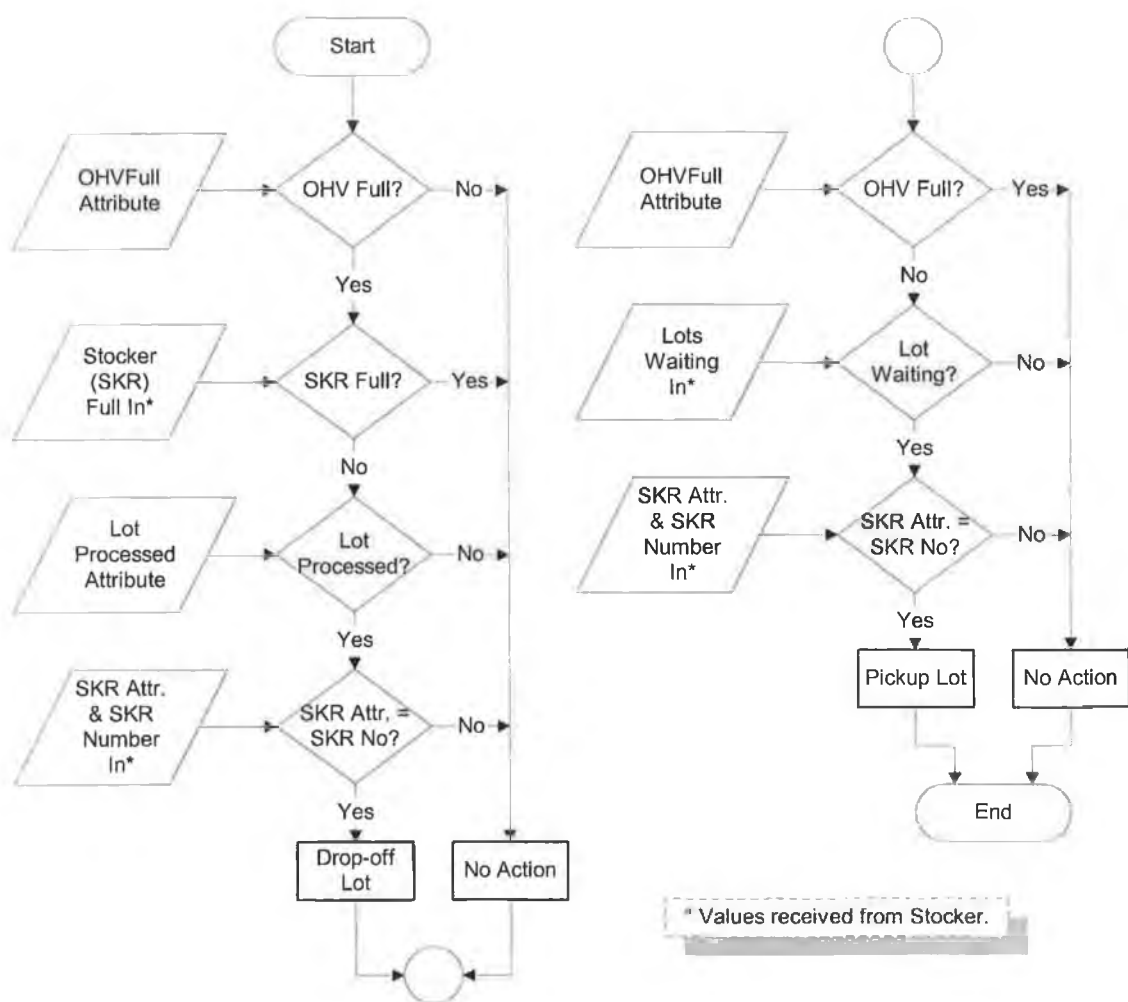


Figure 4.15: Lot pickup/drop-off at stockers.

Decision Point 4: Setting Collect Attribute

The simple, yet important, decision process at this point sets the collect attribute. This attribute helps in identifying the bay where the maximum number of waiting lots are

located. Collection of lots is from that bay in order to balance the number of waiting lots in all bays.

When the maximum number of waiting lots is zero, then there are no lots ready for collection and the attribute is set to zero preventing the OHV from diverging into any bay; otherwise, the attribute value equals the number of the bay holding the maximum number of waiting lots.

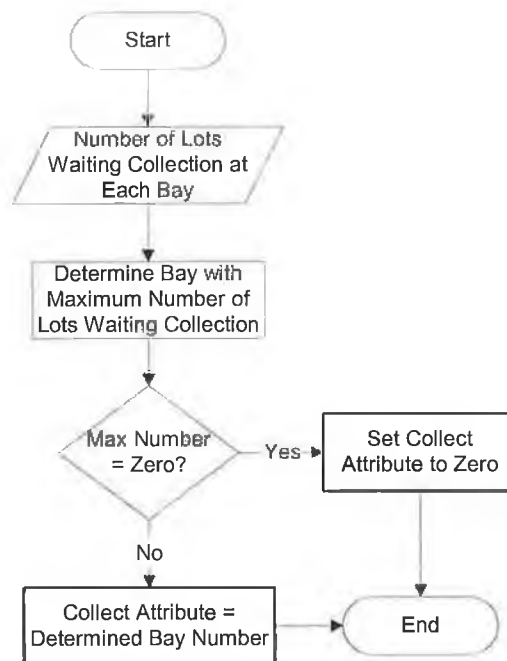


Figure 4.16: Selection of bay for lot collection.

Decision Point 5: OHVs Routing Decisions at Merge/Diverge Link 1

At merge/diverge link 1, a decision is made to determine whether the OHV will diverge into the bay or continue on the outer loop. This primarily depends on the merge/diverge status. When the merge/diverge is OFF, the OHV at all times will diverge and go into bay. On the other hand, if the merge/diverge is ON other factors are needed to make the decision, as shown in Figure 4.17.

First if the OHV is empty, meaning that it will not deliver a lot, a check is made to determine if the OHV will collect a lot from that bay or not (depending on the collect attribute set in the previous decision point). If a lot requires collection, the OHV diverges into the bay; otherwise the OHV remains on the outer loop.

Second, if the OHV is carrying a processed lot, it remains on the outer loop because the lot has to be returned to a stocker.

Finally, if the OHV is carrying an unprocessed lot, a check is made to determine whether to deliver the lot at the current bay or not. If the bay holding the production tool required for processing the lot is the same bay that the OHV is situated, then the OHV diverges into that bay; otherwise the OHV continues on the outer loop.

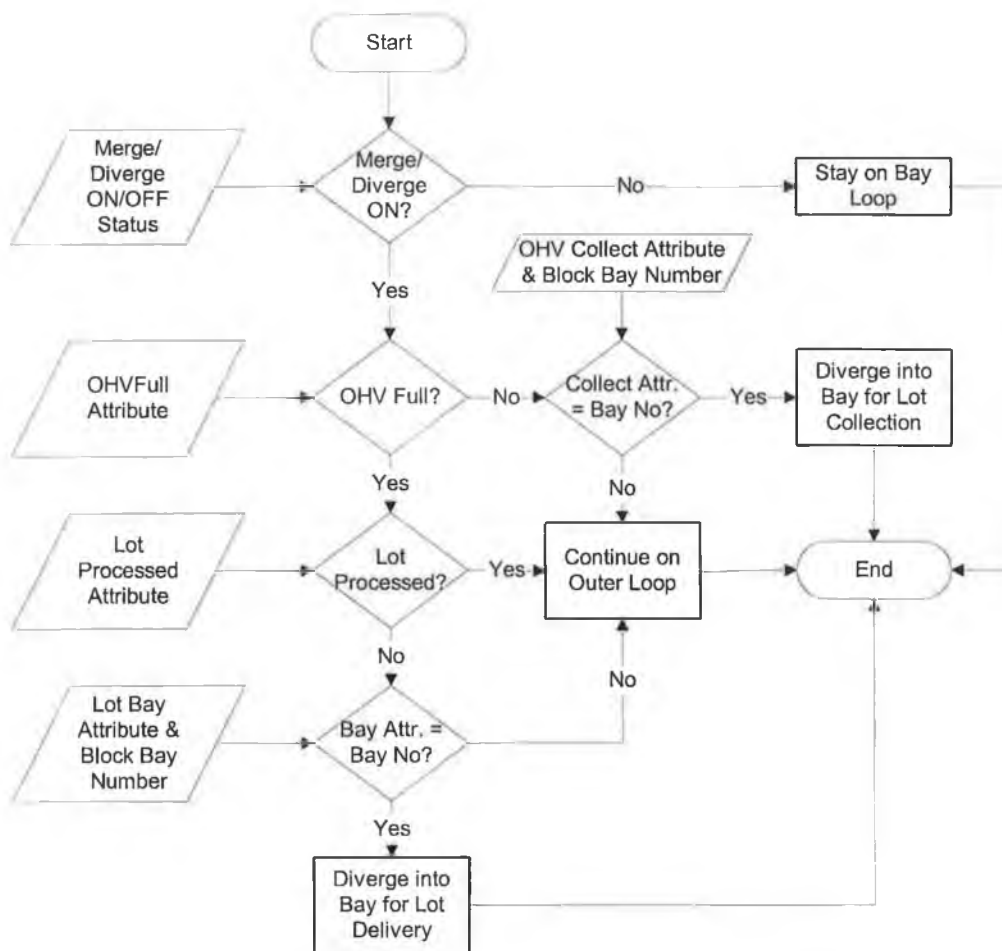


Figure 4.17: Routing logic at merge/diverge link 1.

Decision Point 6: Drop-off/Pickup at Production Tools

This decision point, located in the route in front the production tool block, is responsible for determining whether the OHV picks up or drops off a lot at that tool. Based on the assumptions presented in section 3.7.4 above, an OHV can perform two moves at a production tool, drop-off of a lot (given that the OHV is full) then pickup of another lot (given that a lot is waiting collection). Therefore, the decision process is for the same

production tool and starts by deciding on the lot drop-off at the production tool, as shown in Figure 4.18.

If the OHV is carrying a processed lot, tool attribute of the lot equals the tool number, and a free tool loading port exists, then the OHV drops off the lot; otherwise the OHV takes no action and the next decision is made, which is lot pickup. If the OHV is empty and a lot is waiting for collection, the OHV picks the lot; otherwise, no action is taken and the OHV proceeds to the following production tool.

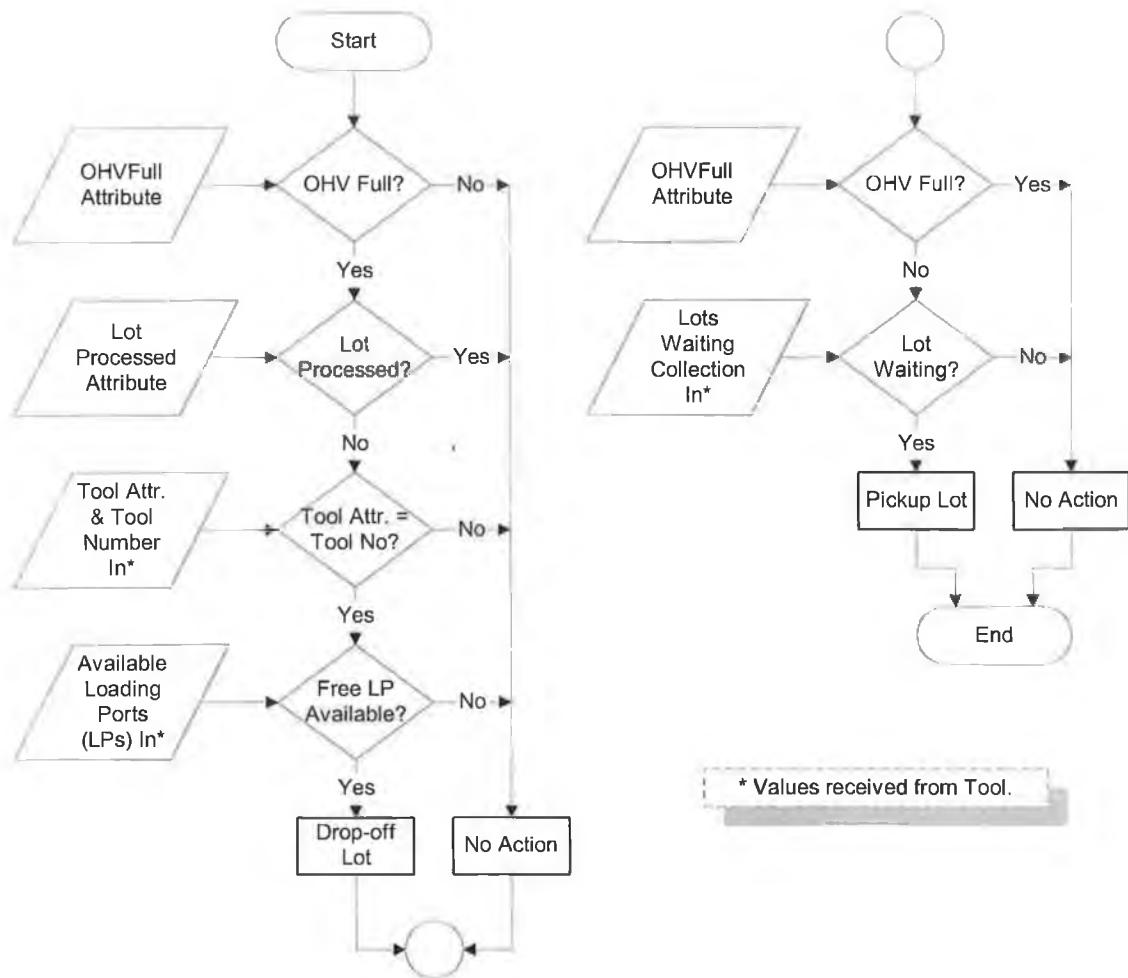


Figure 4.18: Lot pickup/drop-off at tools.

Decision Point 7: Setting Bay Attribute After Lot Collection

A lot collected from a production tool is returned to a stocker belonging to a specific bay. The decision process at this point is required to determine the bay that the stocker belongs to. A bay attribute is assigned to every lot routed, as discussed before in 4.4.3

above, which is needed to identify the bay in which the lot is processed. This attribute remains assigned to the lot even after pickup, as it guides the OHV to the bay to which the lot is returned. If this bay attribute remains unchanged, then the OHV returns the lot to a stocker in the same bay.

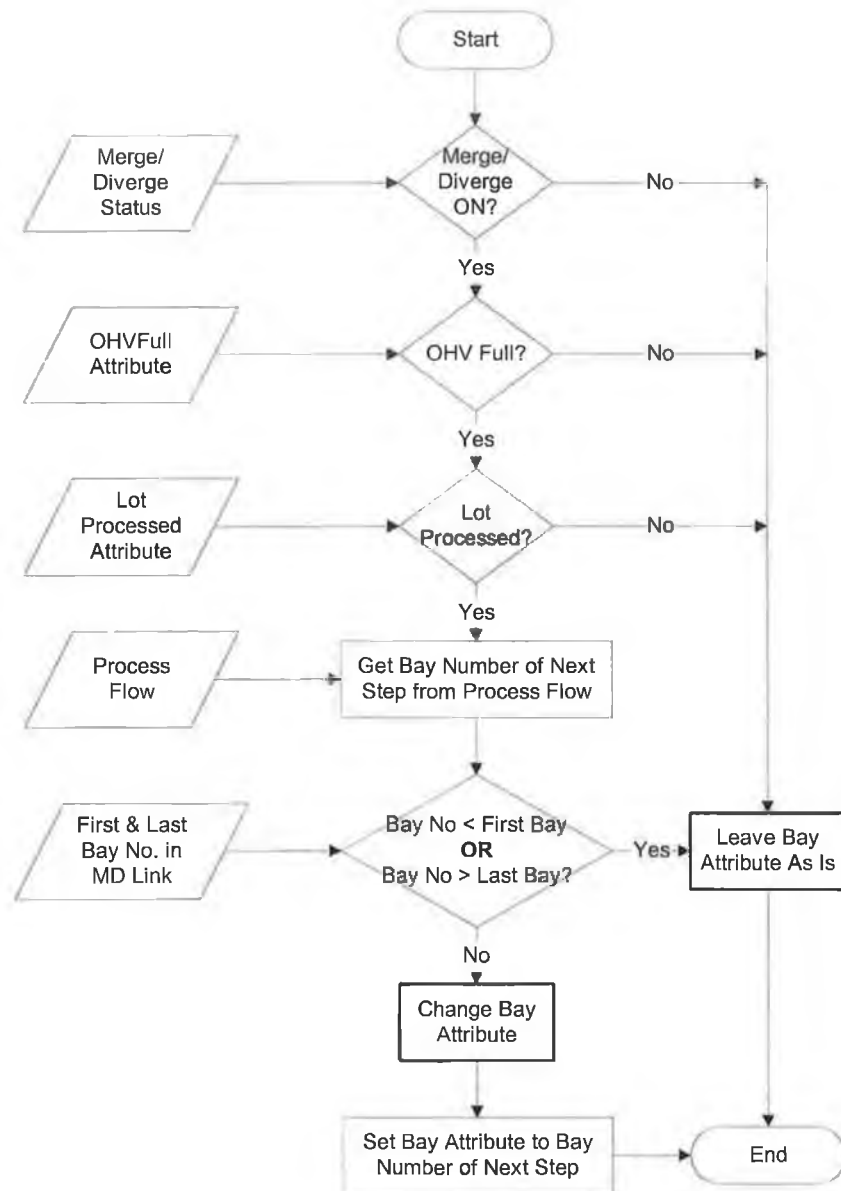


Figure 4.19: Selection of bay for next move.

On the other hand, if the next manufacturing step of the lot is in another bay, then the lot must be returned to a stocker belonging to that bay, meaning that the bay attribute must be changed. However, the attribute is not changed in all cases, because the decision depends on several other factors, as shown in Figure 4.19.

The bay attribute remains unchanged if the following conditions are satisfied:

- Merge/diverge is OFF, as the OHV must remain on the bay loop.
- OHV is not carrying a lot, as there isn't a lot to return and change is not needed in that case.
- OHV is carrying an unprocessed lot, as the bay attribute assigned to the lot determines the location of the production tool required for processing the lot.
- Bay for next manufacturing step doesn't belong to the linked bays. In that case the lot requires moving using the interbay AMHS, and thus it is logical to return it to a stocker belonging to the same bay.

If any of these conditions is not satisfied, then the bay attribute value is changed to the value of the next manufacturing step bay number.

Decision Point 8: OHVs Routing Decisions at Merge/Diverge Link 2

At the merge/diverge link 2 routing decision (illustrated in Figure 4.20) is made to determine whether the OHV will continue travelling on the outer loop, or diverge and continue on the bay loop.

If the merge/diverge is OFF, this means that the link is non-existent and that the OHV will continue on the bay loop. On the other hand, if the merge/diverge is ON this means that the link exists, and decision must be made to determine the route to take.

If the OHV is not carrying a lot, then it will continue on the outer loop, which the default route for free OHVs. Otherwise, the bay attribute value (set at the previous decision point) is compared with the bay number (reported by the merge/diverge link block). If the two values are equal, which means that the lot is returned to a stocker in the same bay, the OHV diverges to the bay loop; otherwise, the OHV will continue travelling on the outer loop.

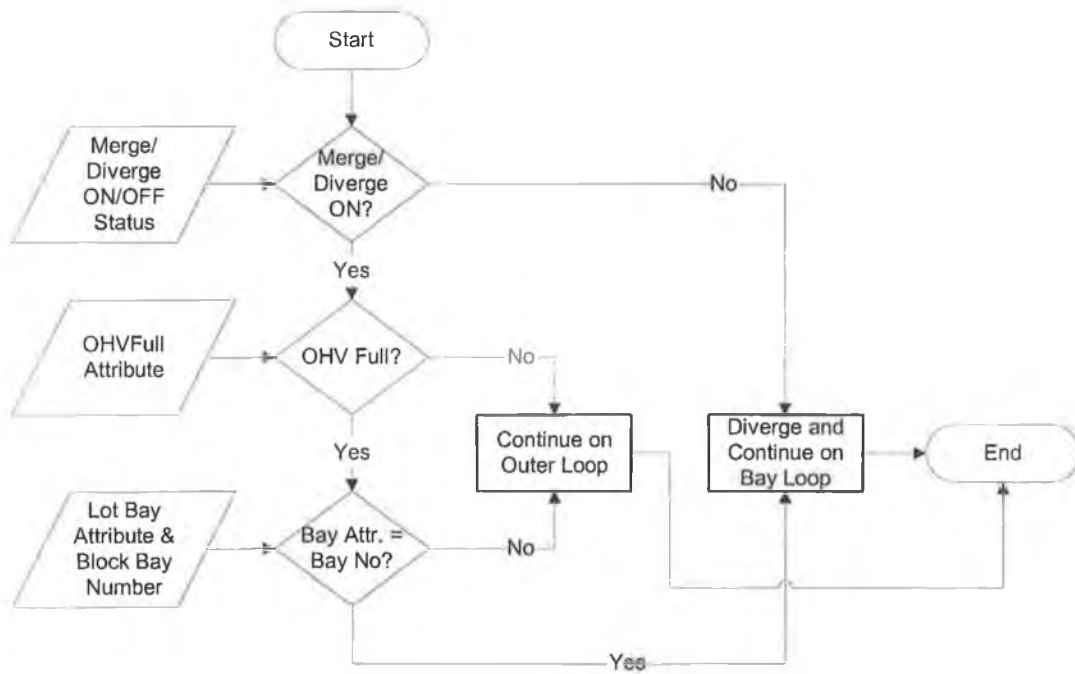


Figure 4.20: Routing logic at merge/diverge link 2.

Decision Point 9: Selection of Stocker for Next Move

The selection of a stocker for the next move, which is lot pickup or drop-off (decision point 3), performed by an OHV is shown in Figure 4.21. The outcome of the decision making process here is the stocker attribute, which will guide the OHV to the required stocker.

If the OHV is not carrying a lot, then it can pickup one at a stocker it passes by. The stocker to pickup the lot from, in that case, is the one with the maximum number of lots awaiting collection.

If the OHV is carrying a lot, the bay attribute (assigned at decision point 7) is checked to determine whether the OHV is located in the correct bay (where the lot is returned to) or not. The stocker attribute is set to zero, when the OHV is not at the correct bay, preventing lot dropping off at any stocker in that bay.

If the OHV is at the correct bay, the preferred stocker is retrieved and the capacity of that stocker is checked. The stocker attribute value is set to the preferred stocker number, given that it has the capacity to accommodate the lot; otherwise, the stocker attribute value is set to the stocker holding the minimum number of lots.

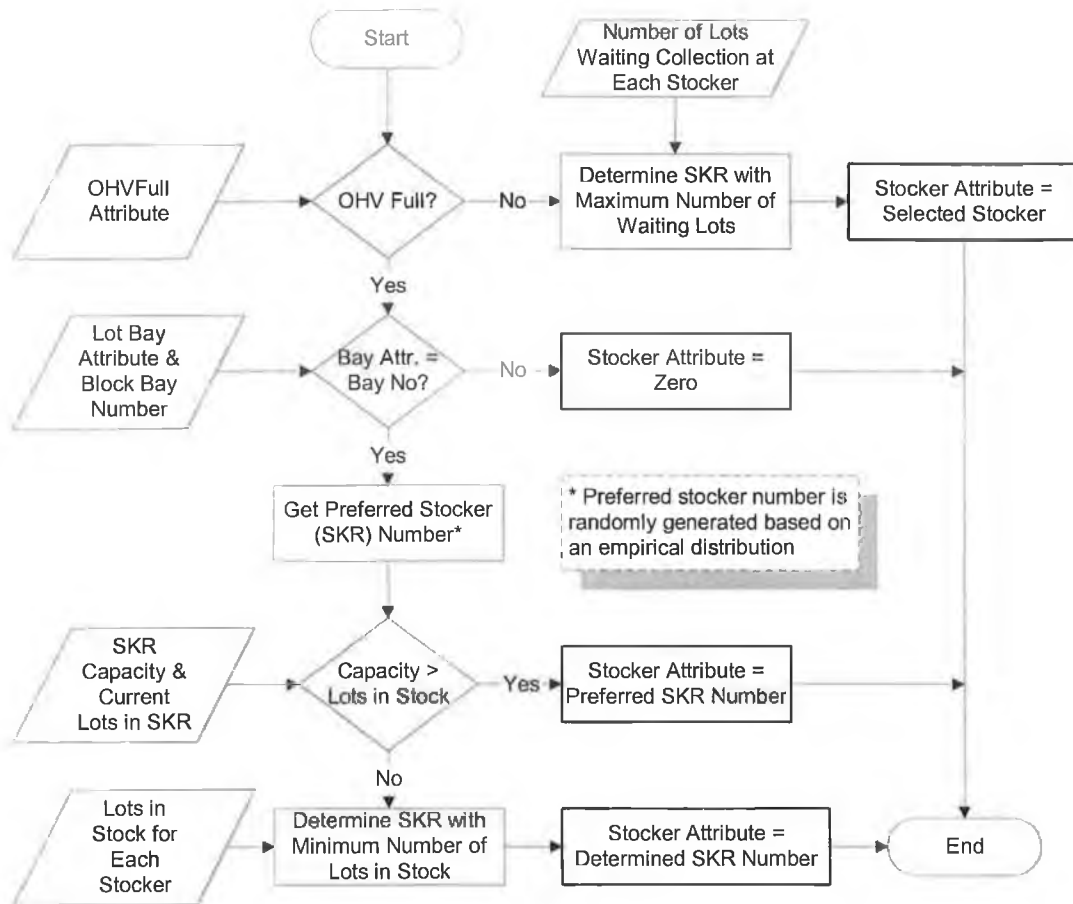


Figure 4.21: Selecting stocker for next OHV move.

4.4.5 Interaction of Components

This section provides detailed description of the structure of GTIA-M&S using IDEF0*. The purpose IDEF0 modelling is to fully describe how the different components of the tool operate and interact together, which is useful if a different system analyst reconfigures the tool. Furthermore, this section doesn't describe *how to use* GTIA-M&S, as this is described in section 4.10 below. IDEF0 is used here to describe every function performed by the tool, and to whatever detail needed for each function as illustrated by the IDEF0 node diagram in Figure 4.22. Furthermore, the different functions that have been modelled and decomposed using IDEF0 are listed in Table 4-5. Finally, The flowcharts presented in the previous section can be referred to for understanding the different decision making processes.

* Appendix E offers a brief description of IDEF0 as well as the created charts, which can be unfolded and referred to while reading through the detailed nodes description.

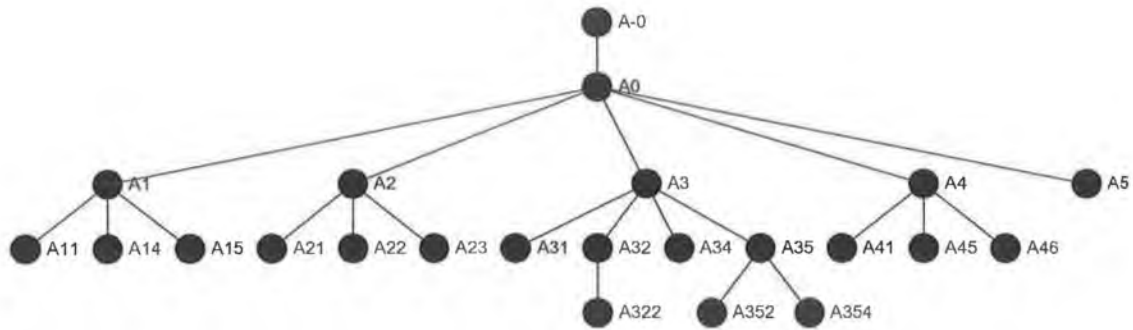


Figure 4.22: IDEF0 node diagram.

Table 4-5: The different functions modelled using the IDEF0.

A-0	Studying Different Types of Intrabay AMHS Using GTIA-M&S
A0	Main Activities modelled using GTIA-M&S
A1	Entities Generation and Routing
	A11 Generation of Lots
	A14 Routing Lots to Bays
	A15 Routing of OHVs
A2	Different Activities Performed in a Stocker
	A21 Distribution of Lots Over Stockers in a Bay
	A22 Storing Lots in Stockers
	A23 Queuing Lots Ready for Pickup by OHVs
A3	Moving Lots Using Intrabay AMHS
	A31 Collection of Lots from Stockers
	A32 Delivering Lots from Stocker to Tool
	A322 Dropping-off Lot at Desired Tool
	A34 Collection of Lots from Tools
	A35 Returning Lot to Stocker
	A352 Determining the Bay for Next Move
	A354 Dropping-off Lot at Stocker
A4	Processing of Lots Using Production Tools
	A41 Setting the Stocker to Tool Transportation Time
	A45 Changing the Values of Attributes Before and After Releasing Lots from a Tool
	A46 Queuing Processed Lots Ready for Collection at Tools
A5	Collecting, Calculating, and Writing of Performance Related Data

NODE A-0 – DIAGRAM 1

This diagram is a special case of a one-box IDEF0 context diagram, containing the top-level function being modelled along with statements of model purpose and viewpoint.

The top-level function being modelled is the studying of intrabay AMHS using simulation, which represents the upper-most level in the IDEF0 hierarchy. The purpose of the IDEF0 model is to provide a system analyst with a means of comprehension and documentation of the GTIA-M&S underlying structure and model design.

ExtendTM simulation environment has been used as the discrete event simulator in this study. Based on the different user inputs, selected performance measures are evaluated. The performance of the intrabay AMHS relies heavily on the routing logic of the vehicles moving the material as well as the mode of operation. There are two main modes of operations, with merge/diverge and without merge/diverge. GTIA-M&S has been developed using ExtendTM to study the performance of the intrabay AMHS according to the user input and under the different operating conditions.

NODE A0 – DIAGRAM 2

GTIA-M&S is capable of emulating a number of essential activities and functions required to represent an intrabay AMHS. These activities start by the production and routing of lots and OHVs to the different bays, then storing those lots in stockers, moving lots from stockers to the production tools, processing lots, and returning lots back again to stockers using the intrabay AMHS. Finally, the various data, stored as attributes in the flowing entities, are collected to evaluate a set of performance measures. These are the five main activities described in this diagram.

The user inputs the wafer starts per week along with the number of OHVs required. The first activity generates these two types of entities using a number of ExtendTM blocks. Lots are sent to the bays, based on the process flow, and ultimately stored in the stockers. OHVs are sent to different OHV entry points, depending on the mode of intrabay AMHS operation. Lots and OHVs are emulated entities produced by ExtendTM.

Storage of lots generated is the second activity. Lots generated and sent to a bay are stored in the stockers found in that bay. In addition, lots that have finished processing on production tools are sent back to the stockers. Lots are released from a stocker either to be moved to a production tool for processing, or to be sent back to the first activity for routing. Stocker storage capacity is the main constraint for this activity. Again the functions of the stocker are emulated using different ExtendTM blocks.

Moving of lots using the intrabay AMHS is the third and most important activity (from a model scope perspective), and thus this activity is the most detailed one. Non-processed lots released from the stockers are moved by the intrabay AMHS using an *emulated* OHV and delivered to a designated production tool (dictated by the process flow). The intrabay AMHS then moves the processed lot from the production tool back again to the stockers. During the processing time, the OHV rotates on a default route depending on the merge/diverge status. The sequence and logic of moves is dependant on the routing logic. Furthermore, one or more OHVs must be available to carry out these moves.

Lots processing is the fourth activity that models the actual processing of lots on the different production tools. Again production tools in this case are emulated tools using a combination of Extend™ blocks. This processing activity depends on the production tool capacity.

Finally, the OHVs carrying a processed lot are passed through a data collection point (at the bay exit) where all the attributes assigned to the lots and OHVs are extracted and used for estimating the different performance measures.

NODE A1 – DIAGRAM 3

Described in this diagram are the details of generating and routing lots and OHVs. Using a generator block, a number of lots are fed into the model every simulated week depending on the wafer starts per week.

Similarly, OHVs are fed into the model based on the user input stating the number of OHVs needed in a simulation run. The lots generated are assigned a number of attributes that are needed throughout the whole model, such as the tool attribute, sampling rate, alternate bays...etc. These lots are then routed to bays, based on the process flow, using a number of routing blocks. OHVs are routed similarly, depending on the merge/diverge status, to the different OHV entry points in bays.

NODE A11 – DIAGRAM 4

At the start of the simulation a time advance mechanism is activated. Every 168 hours is equivalent to one week, and a number representing the simulated week is generated and used as a row number in the ramp profile table (found in the built-in

database). The database then returns the wafer starts per week (WSPW) based on the simulated week (row number). WSPW is then fed into an equation block to calculate the inter-arrival rate of lots based on the equation shown in diagram. Accordingly, the generator block generates lots based on the calculated inter-arrival rate value.

It should be noted that no lots are allowed at time zero (phantom lots) and that lots are generated one at a time. Any lots generated by error are discarded and are not allowed to enter the model.

Finally, the tool allows for setting an initial WIP (for verification purposes), and hence the final activity in generation of lots is to add the amount of that initial WIP.

NODE A14 – DIAGRAM 5

After generation of lots and assigning the different attributes, the lots are routed and sent to bays. The bay for the next manufacturing step is identified from the process flow, and is stored as an attribute in lots. These lots are either newly generated lots, or lots needing further processing and thus are feeding back from a bay.

The manufacturing step denotes both processing and inspection activities. Inspection activities, carried out by metrology tools, are based on sampling rate. The tool would skip a manufacturing step if the sampling rate shows that the lot won't be sampled, and hence, it will feed back for setting the bay attribute again.

In some cases the production tool qualified for the manufacturing step can have alternate bays (refer to 4.4.4 above). The availability of these bays is only checked if the original production tool is not available (full). If the duplicate tool in the alternate bay is available, the bay attribute is changed to the alternate bay, which will ultimately lead to sending the lot to the alternate bay.

Finally, lots are sent to bays by comparing the bay attribute stored in the lot with a list of bays numbers and throws the lots to these bays accordingly.

NODE A15 – DIAGRAM 6

The generated OHVs are sent to the different OHVs entry points depending on whether the merge/diverge is set ON or OFF. If the merge/diverge is ON, the number of OHVs required per merge/diverge link is sent directly to the merge/diverge entry

point. On the other hand, with merge/diverge OFF, the numbers of OHVs required in a bay are assigned an OHVBay attribute equivalent to that bay; the throw block sends the required number of OHVs to each bay using that attribute.

NODE A2 – DIAGRAM 7

The lots sent to a bay are distributed over the different stockers in that bays depending on the stocker capacity and number of lots present in each stocker. The stocker can store the lots sent by the previous activity (lots distribution) as well as lots returned by the intrabay AMHS from production tools. In addition, lots ready for collection can queue at the output port of the stocker, and are released upon the arrival of an empty OHV.

As previously mentioned, lots returning from a bay are received by the stocker. Depending on the bay for the next manufacturing step dictated by the process flow, these lots are either stored in the stocker for further processing within the same bay, or are sent to another bay. Lots sent to another bay are the lots feeding back, which is the output of activity 5 (node A2 – diagram 7) as well as the input to activity 1 (node A14 – diagram 5).

NODE A21 – DIAGRAM 8

Before distribution of lots over the different stockers in a bay occurs, a catch block catches the lots sent to that bay. The bay attribute stored in the lot entity is compared against the catch block number of each bay, if both numbers match, catching of lots takes place.

Distribution of lots is required when a bay has one or more stocker, which is the situation of most bays. Depending on the number of stockers in a bay, each stocker will receive a specified proportion of the total number of lots sent to a bay. For example, consider a bay having two stockers, and the total number of lots sent to that bay is 100 lots. The proportion in this case will be 45% for the first stocker and 55% for the second stocker. Thus, 45 lots will be sent to the first stocker and 55 will be sent to the second one.

This proportion has been modelled in GTIA-M&S using a probability distribution. An empirical table returns a value representing the stocker number based on the probability of that value occurring. The probabilities entered in the empirical table

are equal to the proportions for each stocker. The stocker number returned is the “preferred stocker” for receiving a lot, and this is how the preferred stocker is identified.

If the preferred stocker is not full, its number is set as the stocker attribute for the lot. On the other hand, if the preferred stocker is full, the stocker holding the least number of lots is set as the stocker attribute (as an alternative to the preferred stocker when full). Finally, based on the stocker attribute assigned to lots, lots are sent to stockers using a throw block.

NODE A22 – DIAGRAM 9

A catch block catches the lots sent to the stocker, by the previous activity, given that the stocker has space to accommodate those lots. A gate is present at the start of the stocker, which acts as a safety precaution for preventing lots from passing through if the stocker is full. This gate counts the number of lots passing through, the ones that are sent to stocker and the ones that are returning back from bay after processing, as well as senses the lots that are leaving the stocker. Thus, the number of lots present in the stocker is known, and lots are allowed to pass only if this number is less than the stocker capacity.

Before sending the lot to a production tool, stated in the process flow, the availability of the toolset is checked. This availability depends on the total number of empty loading ports of production tools forming the toolset. This number is read directly of the database, and if the number of empty loading ports is greater than zero the tool is said to be free, otherwise it is full.

If the toolset is free, the lots are sent to the stocker output port and waits for collection by an OHV. In contrast, if it is full, lots queue in the stocker until the toolset becomes available again, and are then released to join the other lots waiting for collection.

Lots sent to output port, either sent directly or released from a queue, pass by a change attribute block to decrement the number of available loading ports by one (of the tool that the lots will be delivered to) in the database to overcome the narrow time window effect (refer to 4.5.1 below).

NODE A23 – DIAGRAM 10

The activity of queuing lots ready for pickup by OHVs, described here, begins by setting the start delivery time attribute. This activity is merely tagging the lot passing through by the current time value in the form of an attribute. This attribute will be used later for calculating the stocker to tool delivery time (refer to node A41).

The lots are then queued for collection and wait for pickup by an OHV. The number of lots waiting collection is reported by the queue block and stored in the database, which can be used when an OHV drops-off a lot at the stocker.

On the arrival of an OHV that is qualified for picking up the lot, the lot is released from the queue. Again, the queuing block reports the queuing time. Mean and standard deviation of the waiting time at each stocker are stored in the database. In addition, the waiting time is assigned as a waiting time at stocker attribute, which will be used for calculating the mean and standard deviation of waiting times at all stockers in a bay (refer to node A5).

NODE A3 – DIAGRAM 11

This node describes the key activity that GTIA-M&S was built for, which is the moving of lots using an intrabay AMHS operating with and without merge/diverge. The collection and return cycle starts by an empty OHV collecting a lot from a stocker. After that, the full OHV delivers the lot to a production tool for processing, and the OHV becomes empty again. When there are no lots to collect, the empty OHV travels on a default route depending on the merge/diverge status (whether it is ON or OFF). As soon as the production tool finishes processing the lot, a signal notifies the OHV that a lot is ready for collection, thus the OHV travels to the production tool and collects this lot. Finally, the full OHV returns the lot back to the stocker and once again it becomes empty.

NODE A31 – DIAGRAM 12

A gate block is used to allow only one OHV to pass for making the pickup from the stocker. Any other OHVs following that OHV will be blocked and allowed to queue until the OHV finishes the pickup (refer to 4.5.1 below for explanation).

After the OHV is allowed to pass through the gate, it picks up the lot at the stocker, given that the OHV is empty and that the lot is ready for collection and waiting at the stocker output port. Lot ready for collection, which is a constraint to this activity, is the same output of activity 5 in node A22 – diagram 9. The pickup time duration in minutes is extracted from the database, converted to hours, and used as the pickup delay time.

After delaying the OHV, the lot coming from the stocker and the empty OHV are joined together into a single item representing an OHV carrying a lot. An attribute is then assigned to the new item to define it as a full OHV. Also, the start of transportation time attribute is set, which is equal to the current time. This attribute is used in the following activity for calculating the stocker to tool transportation time.

NODE A32 – DIAGRAM 13

Delivering lots from a stocker to a production tool is described here. The OHV carries the lot and moves it on the intrabay loop, with or without merge/diverge, till the required production tool is reached and the lot is dropped off at that tool.

To identify the required production tool, the tool number (identifier sent by the production tool block) of every production tool the OHV passes by, and the tool number attribute stored in the entity are compared. The OHV recognises that the required production tool is reached when those two values match. At that point, the OHV drops off the unprocessed lot given that the tool has an empty loading port.

Afterwards, the OHV carrying the lot is split into two items (one lot and one OHV) and the OHV is redefined as an empty OHV. The stocker to tool transportation time is calculated and stored as an attribute, which will be used later for estimating the mean and standard deviation of transportation time (refer to Node A5 – Diagram 23).

NODE A322 – DIAGRAM 14

This provides further details about how a lot is dropped off at a production tool. As mentioned before, the tool attribute is compared with the tool number reported by the production tool block. If the two values are not equal, the OHV moves to the next production and compares the two values again. This continues till the OHV arrives at tool where the tool number is equal to the tool attribute. If the two values

are equal, loading port availability is checked and the lot is dropped if an empty loading port exists at the tool. If no empty loading ports were found, the OHV moves to the next tool.

It should be noted that what is meant by the lot dropping off, discussed here, is only the amount of time the OHV is delayed and not the actual dropping off of the lot. The lot is actually dropped off when the OHV and lot are split into separate items (activity 3 – node A32 – diagram 13).

NODE A34 – DIAGRAM 15

Collection of lots from production tools starts by the selection of a bay for lot collection. The selection of bay depends on the number of waiting lots in that bay. The bay with the maximum number of lots waiting collection is chosen, to maintain a balanced number of waiting lots in all bays at all times.

One OHV is allowed to pass for picking up the lot and a decision is made whether or not to pickup the lot (reasons for that are explained in 4.5.1 below). The OHV is delayed for a certain amount of time equivalent to the pickup time, given that the OHV is empty and that the lot is ready for collection. The lot and OHV are then joined together using a batching block, producing a single item that signifies an OHV carrying a lot. Finally, the single item produced is assigned an attribute that defines the OHV as a full OHV.

NODE A35 – DIAGRAM 16

On returning the lot to stocker, the route selected depends heavily on the merge/diverge status. When the merge/diverge off, the OHV will move on the intrabay loop and drops the lot at a desired stocker belonging to the same bay from which the lot was collected. However, when the merge/diverge is on, the desired stocker can belong to any bay in a merge/diverge link. Therefore, the bay for the next manufacturing step must be determined first, send the OHV to the determined bay, and drop-off the lot at a stocker belonging to that bay. The actual drop-off of lot is modelled by splitting the single entity representing a full OHV in two entities, the processed lot and the OHV. Finally, the OHV must be redefined as an empty OHV again.

A352 – DIAGRAM 17

This node provides further details related to the selection of the bay where the OHV will be sent to, for returning the lot to a stocker. A new bay attribute other than the originally assigned one in node A14 – diagram 5 is assigned to the OHV if and only if the OHV is full, carrying a processed lot, and the merge/diverge is on, otherwise, the bay attribute is left as it is. Keeping the bay attribute unchanged means that the lot will be delivered to a stocker in the same bay.

In order to assign a new bay attribute to the OHV, the bay number for the next manufacturing step is retrieved from the database, depending on the process flow. The new bay number is checked to determine if it is one of the linked bays in the merge/diverge link or not. On verifying that the new bay number belongs to the merge/diverge link, the number is assigned to the OHV. If the new bay number doesn't belong to the merge/diverge link, the bay attribute is left unchanged and the OHV continues using the original bay attribute.

NODE A354 – DIAGRAM 18

After determining the bay to send the OHV to and setting the bay attribute, this attribute is compared to the bay number reported by stockers that the OHV passes by. If the two values don't match, the OHV moves to the next stocker and compares the two values again, and so on till a stocker returns a bay value that is equal to the bay attribute. When the two values match, the OHV recognizes that it has reached a stocker in the required bay. If the stocker doesn't have available space to accommodate the lot, the OHV goes to the following one. On the other hand, if the stocker has available space and the lot carried by the OHV is a processed one, the OHV drops off the lot.

It should be noted that this comparison is not needed if the merge/diverge is off, because the lot will be dropped off at one of the stocker belonging to the same bay. In this case the OHV is sent to the delay activity right away, given that the carried lot is processed.

NODE A4 – DIAGRAM 19

This describes lots processing using production tools. On the arrival of a lot to a production tool, the stocker to tool delivery time is calculated and set as an attribute.

A gate block is used to count the number of lots entering the production tool block, as well as to sense the number of lots leaving the production tool. At any moment of time, the gate block is aware of the number of lots present in the production tool block, and will not allow the number of lots to exceed the number of loading ports for that tool.

The lots are allowed to queue at the production tool buffer, which is equal to the number of loading ports, if the production tool is busy processing another lot. Furthermore, processing of lot takes place immediately, given that the tool is free and ready to process the lot.

A number of attributes are changed after processing the lot, and then the lot is queued for collection by an OHV. On the arrival of the OHV, the lot is released from the stocker and a number of attributes are changed again before it is picked up by the OHV.

NODE A41 – DIAGRAM 20

The start of delivery time attribute, which was set at the stocker (Activity 1 – Node A23 – Diagram 10), is retrieved. The value of this attribute and the current time are used to calculate the stocker to tool delivery time. The calculated time is then assigned as an attribute to the unprocessed lot, which will later be used to estimate the mean and standard deviation of the delivery time over the simulation run.

NODE A45 – DIAGRAM 21

There are a number of attributes that must be changed (value of attribute) before queuing a lot ready for collection, and after releasing it. These attributes are vital in representing the state of the system being modelled, as they represent finishing of a manufacturing step, existence of lots waiting for collection, and availability of loading ports. Furthermore, these attributes are database aware, meaning that the changes in their values are reflected in the database altering the state of all blocks using these attributes. Finally, the use of this type of attributes is the only way to overcome the time window effect (discussed in 4.5.1 below) that is experienced on modelling the intrabay AMHS.

Before a lot is allowed to queue for pickup by an OHV, the step sequence number is incremented by one, to represent that the current manufacturing step has finished. Also, the number of waiting lots in the bay is incremented by one and reported to database. When the lot is released from the queue, the number of waiting lots in the bay is decremented meaning that the lot has been picked up. In addition, to indicate that a loading port has been freed up, the number of available loading ports is incremented by one, which is again reported to the database.

NODE A46 – DIAGRAM 22

The processed lots are held in a queue waiting for an OHV qualified for pickup to arrive. The queue block returns the queue length value, to notify passing OHVs with the presence of waiting lots. The processed lot is released from the queue when an OHV arrives and is ready to collect the lot. The queuing time of the lot is reported by the queue block and stored as an attribute. This attribute value will be used later to compute the mean and standard deviation of lots waiting time at production tools. Moreover, the lot released is defined as a processed lot by changing the processed lot attribute value to 1.

NODE A5 – DIAGRAM 23

All the OHVs pass by a data collection point located at the bay exit. At this point, performance related data are collected, calculated, and written to the appropriate place in the database. This function describes the various data-related activities performed GTIA-M&S. All OHVs pass by this point; however, only OHVs that are carrying processed lots are selected for data collection.

The moves per hour are first calculated by counting the number of items passing through every hour. These values are written in an expandable table in the database, with each row representing an hour, the number of rows expands according to the simulation run time.

All the performance specific attributes that have been set by the different activities in the model are retrieved. These attributes are the waiting time at tools, waiting time at stockers, stocker to tool delivery time, and stocker to tool transportation time.

The values of these attributes go into a statistics block for calculating the mean and standard deviation of these values. Finally, the calculated mean and standard deviation is written to the database in a designated place, identified by row number, column number, and table name.

4.5 TRANSLATION

Translation is concerned with transforming the tool conceptualisation, presented in the previous step, into a computer recognizable form. This form is the operational components forming GTIA-M&S and constructed using ExtendTM.

4.5.1 Challenges

Experience gained from constructing the first model, showed that ExtendTM is not equipped with blocks capable of modelling the complex structure of the intrabay AMHS. Presented here are some of the challenges of modelling different components of the intrabay AMHS using ExtendTM.

Modelling of OHVs and Intrabay Track

ExtendTM provides three types of blocks for modelling material handling systems these are: route for empty vehicles, route for full vehicles, and the vehicles.

The first drawback is the two types of routes, which necessitates building two tracks, one for empty OHVs and another for full OHVs. Modelling the track in that manner will increase the complexity of the model twofold. In addition, the level of understanding the model will decrease, as it will no longer offer the end user or the analyst with a replica of the real system.

The second drawback is related to the standard block to represent vehicles. These blocks do not allow attribute assignment, and attributes play a key role at all the decision-making points. In addition, due to the re-entrant flow of lots, the block loses track of actual number of OHVs specified by the user.

Those two major drawbacks lead to search for other possible solutions for modelling the intrabay track and the OHVs.

Modelling of OHVs

The vehicles or OHVs are represented as regular entities, rather than a special resource, that are transferred between the activities. An OHV performs three basic moves; moving from one point to another, picking up a lot, and dropping off a lot. The OHV entity must be capable of performing these moves.

Emulation of an OHV picking up a lot is by batching the OHV and lot to form a single item representing an OHV carrying a lot, as shown in Figure 4.23 (A). To emulate an OHV dropping-off a lot, the single item is split into two items, the OHV and the lot, by un-batching as shown in Figure 4.23 (B).

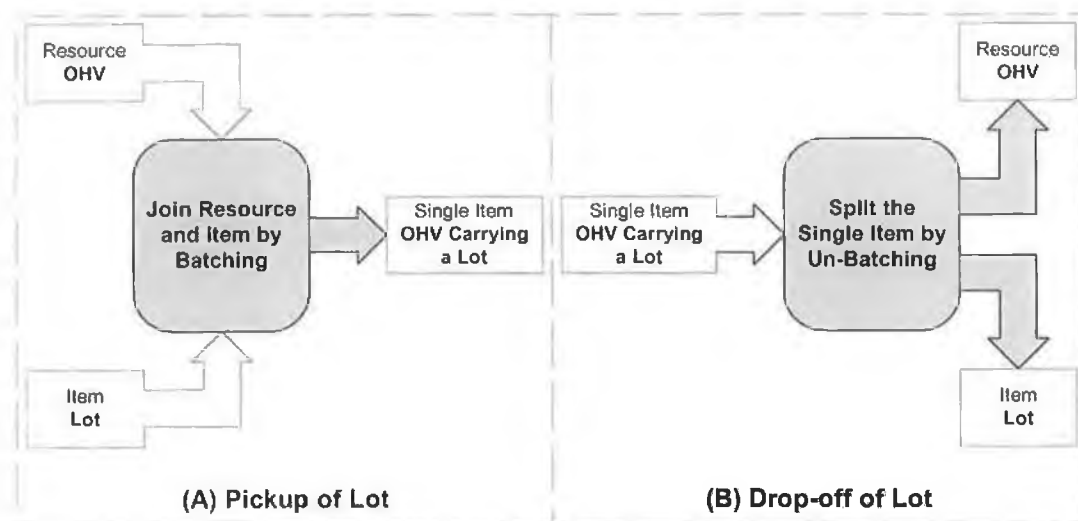


Figure 4.23: Emulation of pick up and drop-off of lots.

Modelling of the Intrabay Loop

The intrabay loop is broken up into small segments; each segment has a specified length (entered by the user). This length is equal to the distance to be travelled by the OHV to get from one end to the other. For example, inside a bay the starting/ending points are typically loading ports of different production tools. The time needed to travel between these two points is determined based on the OHV speed (extracted from the database) and the segment length, which is computed using an equation block ($\text{Time} = \text{Distance} / \text{Speed}$).

Each track segment must be represented by a multiple activity block. The block delays one or more OHVs according to the travel time as calculated above and the capacity of the track segment in terms of the number of vehicles that can be simultaneously on the segment.

Multiple activity blocks release the entity requiring the least remaining delay first. If OHVs entering these blocks have different speeds, then it can lead to changing the sequence of OHVs because one OHV may need less time, to travel between two points, than another. Therefore, although the OHVs accelerate and decelerate at different points on the loop, it is assumed that OHVs run at a constant speed to maintain their sequence.

OHVs Overflow

When an OHV picks up or drops off a lot, at a stocker or at a production tool, the remaining OHVs must be held back or else OHVs will overflow as shown in Figure 4.24. OHVs overflow leads to alterations in the sequence of OHVs, and erroneous representation of the real system.

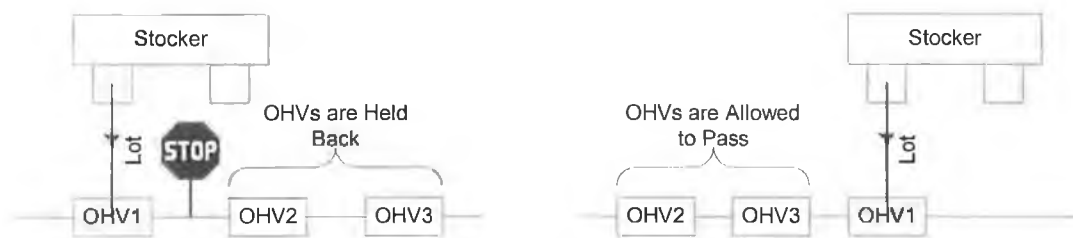


Figure 4.24: Preventing OHVs overflow.

This overflow of OHVs problem originates because OHVs are modelled as a regular entity, and not as a special entity developed specifically for material handling. A feature, using a gate block, is introduced to prevent overflow of OHVs. The gate block passes only one OHV, to pickup or drop-off a lot, and holds the remaining OHVs at the previous loop segment; thus, OHV overflow is avoided.

Decision Making

Gate blocks also play an important role in the decision-making processes found at different points. As, it has been found that ExtendTM is unable to accurately perform

decision-making if items are not effectively restricted from passing through, especially when the decision requires the retrieval of attributes.

Figure 4.25 shows that, even if a stop block exists and the animation shows no items passing by the ‘get attribute’ blocks, the attributes are still retrieved and when the item actually passes by these blocks, the item is lost. Gate blocks on the other hand effectively prevent the item from passing through as well as retrieval of attributes. This implies that relying fully on animation for verification purposes can be misleading, and that such a method of verification must be used carefully.

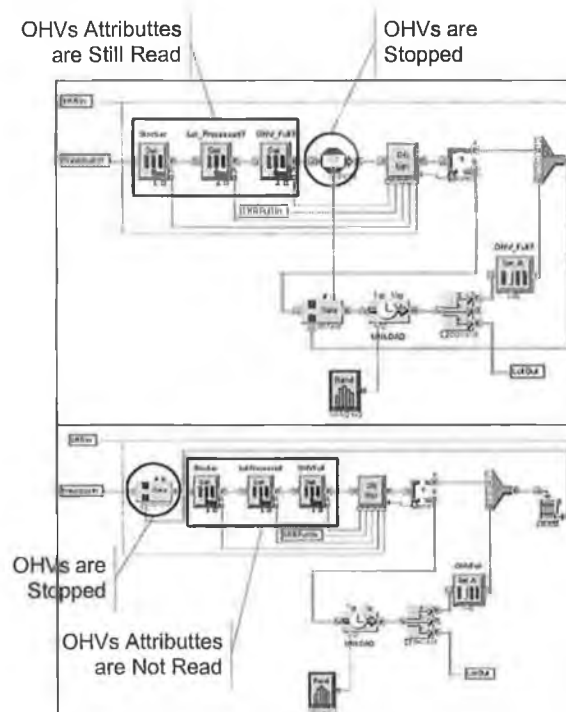


Figure 4.25: Using a gate block for accurate decision-making.

Number of Available Loading Ports

Availability of production tool loading ports plays an important role in GTIA-M&S, as lots are not released from the stocker and sent to a production tool, unless that tool has free loading ports. Determining the number of available loading ports of a production toolset in a bay, and changing this number whenever a lot is released from a stocker or production tool are all computed in the database.

A component called ‘calculate loading ports available’ calculates the number of loading ports available for a production toolset. This number is calculated based on the number

of loading ports of a single tool and the number of tools in the production toolset. The database reports the number of loading ports for each production toolset type, and the user specifies the number of tools in a toolset depending on the number of tools in that toolset located in the bay.

The number of available loading ports is stored in a tool-bay matrix, with the production tool type as rows, and the bays as columns. For every production toolset in a bay, the total number of available loading ports is written to the database.

Given a specific toolset in a specific bay, every time a lot is qualified and sent to this toolset the available loading ports decrease by one. Furthermore, every time a lot is picked up by an OHV at the production tool this number increases by one.

Narrow Time Window Effect

The time window between the different events being modelled in an intrabay AMHS is very short, due to the high travelling speed of OHVs, which results in a large number of events occurring within seconds making it very difficult to control the flow of entities in the model.

For example, when delivering a lot from a specified stocker to a specified production tool, several events can occur and decisions can be made during the delivery time. An OHV only delivers a lot if the production tool, which will receive the lot, has an empty loading port. The number of free loading ports must be decreased by one, as soon as the decision is made to deliver the lot, at the stockers and not at the production tool where the lot is actually delivered. Otherwise, other lots will be sent to the same production tool, because the delivery time is very long compared to the time window, and other decisions will send more lots.

This issue has been effectively handled using the database aware attributes, which is an attribute capable of changing a value in a specific cell in the database whenever an entity passes by. Looking at Figure 4.26, whenever a decision is made at a stocker to deliver the lot it passes by a database aware attribute block, and decrement the number of available loading ports of the specified production tool by one. This change in value is directly reported to the database. On the other hand, as soon as the lot is picked up by

an OHV, it passes by another DB aware block, available loading ports of the specified production tool is incremented by one in appropriate cell of the database.

The same problem occurs with the number of waiting lots in bays, and has been overcome in the same way.

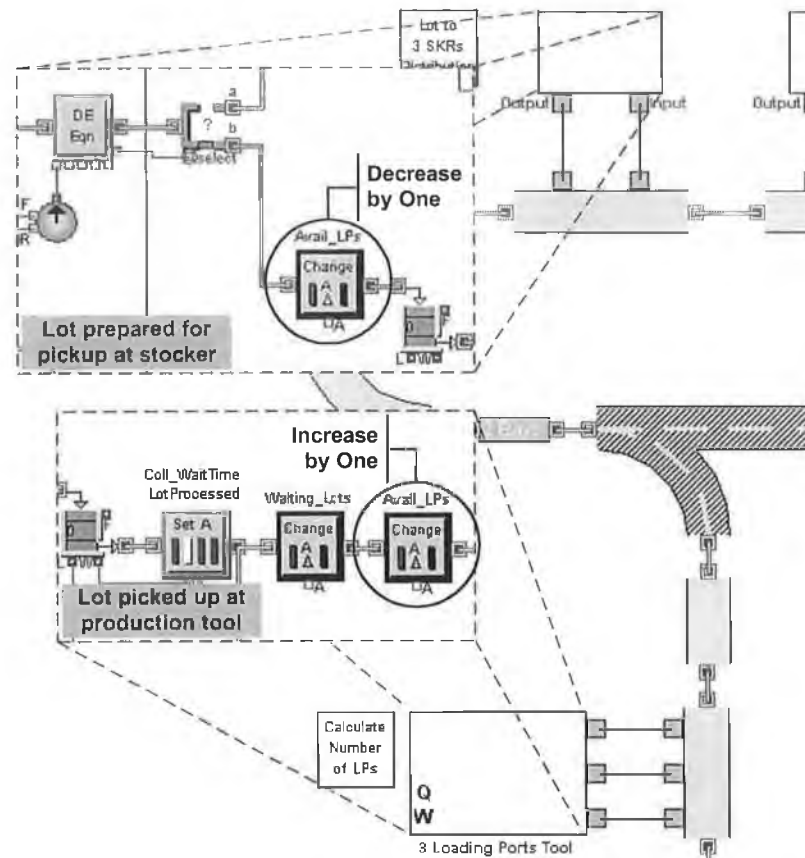


Figure 4.26: Using DB-aware attributes for changing the number of free loading ports.

4.5.2 Component Based Modelling

After the first modelling experience, it is clear that creating blocks for modelling the different components found in an intrabay AMHS is inevitable. Component based modelling starts by identifying the major components required to model the system under study. Then, each of these components is modelled on its own, which is done by assembling a number of building blocks in a single component. Finally, the interaction and communication between the different components is established.

The major types of components that are modelled for GTIA-M&S are:

- Stockers.
- Production tools with different loading ports (from one to four loading ports per production tool) and on the different sides of a bay (east and west).
- A number of straight and curved routes needed to model all the different possibilities for routes directions.
- Merge/diverge links.
- Different types of components required for decision-making.

While modelling the different components, the generic nature and flexibility of all components must be maintained. For example, a merge/diverge link component that links two bays together (merge/diverge link 2, refer to 4.4.4 above) can link any two bays together. As shown in Figure 4.27, from the top-left side simply selecting the names of the two bays, from a list of bays, configures the component to act as the merge/diverge link between these bays.

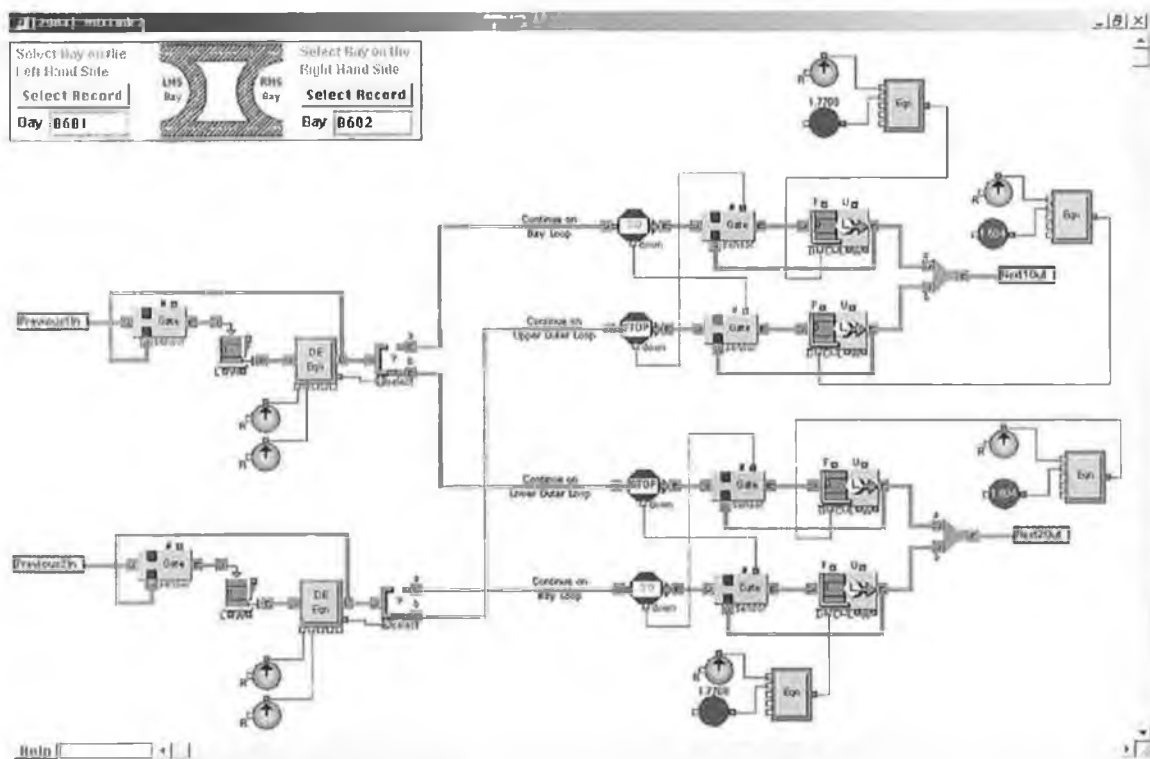


Figure 4.27: Configuring a merge/diverge link.

After a component is built, its operation and interaction with other components is verified (discussed in 4.6 below). Finally, if the component functions as intended, it is added to a library that holds all the components built specially for GTIA-M&S.

4.5.3 Incorporation of Special Features

This step comes after finishing the bulk of the modelling process presented in the previous sections. It is concerned with presenting the end-user a number of special features that can increase the friendliness and usability of the tool. The tool is presented to a user, who has never used it before, and the user interaction with the tool is monitored. Any difficulties in using the tool as well as the user needs are documented. After that, remedies to these difficulties are incorporated in the tool, which effectively increases the user acceptance of the tool.

One major difficulty encountered was getting started with using the tool. Although, the user finds a library comprising a number of components that are capable of modelling the system under study; yet, the user is confused with respect to which component to use, where to place it, and how to connect it to other components. This is outlined in the following sections.

Tool User-Interface

This feature provides the user with a six-step guide to model building and simulation using GTIA-M&S, these steps are:

- Step One: Check Your Data
- Step Two: Open Library
- Step Three: Build Your Model
- Step Four: Model Setup
- Step Five: Simulation Setup and Run
- Step Six: View Results

Through the interface the user can check/edit the data in the database, load and open the GTIA-M&S library, then builds his/her model with the aid of the model-building template (discussed below). After the model is developed, the different operating conditions are set in step four, such as the merge/diverge status. The simulation parameters are then set and model run is executed. Finally, at step six the results of the run are obtained and can be exported to Excel.

Model-Building Template

The model-building template, shown in Figure 4.28, guides the user while modelling a bay. The template clearly shows the sequence of laying out the components, the required components, the recommended location of each component, and how to configure these components.

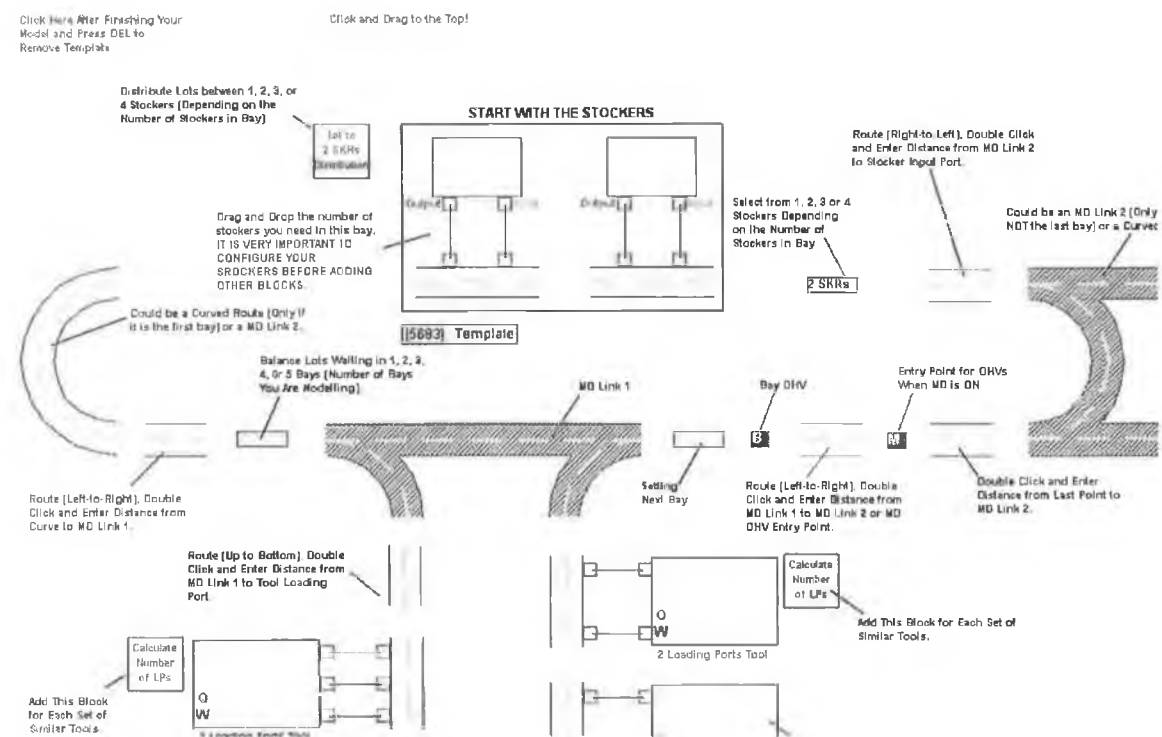


Figure 4.28: Model-building template.

Built-in Help

Although configuring of components appears easy and straightforward the user may still be confused about how to configure a component, simple help is built-in every component that guides the user to configuring that component.

Animation

Simple animation is built in GTIA-M&S to clearly show the different moves of the intrabay AMHS and its interaction with the stockers and production tools.

Furthermore, the system analyst uses animation for verification (refer to 4.6 below) and validation (refer to 4.7 below) purposes, with caution. Animation is used to insure that components are operating as intended. Also, it is used to insure that the models operation (developed by the tool), as a whole, confirms to the information gathered about the system under study.

4.5.4 Tool Refinement

This is the final step in the tool development, in which all the components of the tool and the database are refined and any redundant parts are deleted. These are some of the tool refinements:

- Some parts of the tool are simplified, by combining a number of blocks in one block for example, which can ultimately decrease the size of the models developed.
- Component input interfaces are refined and assume a standard look, which helps the user become familiar with the tool easily.
- Centralisation of data that is common in a number of components.
- Similar data in the database are grouped together.

Verification of every refinement is always carried out, to ensure that the component is still operating as intended after any modifications.

4.6 VERIFICATION

As mentioned in the previous step, animation has been built into the tool to provide the user with an understanding about the system behaviour. Animation also plays a key role in the verification and validation process for models of non-existing systems.

Verification or debugging of a tool is a continuous process and is carried out in parallel to the translation process. Every component modelled, and even part of a component in some cases, is verified. Verification in this case is by testing a variety of settings of the input parameters and checking that the output is reasonable. Readout blocks, provided by ExtendTM, can be used to read the output value of every block in the component to insure that those blocks are functioning in the intended manner.

Afterwards, a number of verified components are connected together to produce a model of reasonable size, for one bay for example. This model is run with limited inputs and assumptions, for example with one lot, one OHV, and without random variables. At this stage, animation's key role in the verification process is obvious, as it clearly shows if this simplified model runs as intended or not, for example routing of OHVs is accurate and the time needed to travel from one point to the next is correct.

Finally, the simplified model is expanded to a larger model with all inputs. Verification is carried out again, with the help of animation and the readout blocks, to check if the model will operate as intended or not. Initial WIP (presented in 4.4.5 above) can be used to speed up lots introduction to the model and thus the steady state can be reached faster. However, initial WIP must only be used for verification purposes.

In addition, verification of the input parameters, especially the randomly generated ones, is needed to ensure the model is using valid values for the different input parameters. For example, the variables generated by an empirical table are collected, and the proportion of each value generated is determined to check if it matches the required probability of occurrence of each of these values.

4.7 VALIDATION

Total validation of the tool developed cannot be achieved, because neither the real nor a similar system has been implemented to date.

Each component of the tool is validated against the information provided about the operation of the underlying system by ensuring that the tool responds in an acceptable manner to changes in inputs and replicates the information gathered about the system

operation accurately. The outputs, especially delivery and waiting times, show the characteristics expected by the industry partner in relation to the input scenarios used.

These points increase the confidence in the validity of the tool. Finally, the tool can be used to build models of the real system once it is operational, and validation of these models can then be achieved by comparing the output of the real system to that of the models.

4.8 EXPERIMENTAL DESIGN

This section provides a guide to setting the different simulation parameters for running models developed by GTIA-M&S. Stress is made to the fact that the values presented here are just recommendations, and that the end-user can choose any values for these parameters that suit his/her needs.

The ramp profile determines the number of lots to be generated every week, therefore a warm-up period of one week is set after which the model is expected to reach steady state. If the warm-up period is to be extended for any reason, this is directly done through the user interface.

Assume that a model developed by GTIA-M&S is run for one week. During this period the tool will collect a huge number of data about the performance of the intrabay AMHS. This is due to the high travelling speed of OHVs, which means short travel times (minutes or even seconds) and also means a great number of events occurring during that period. In addition, whenever an event occurs calculations for every block in the model are performed, which exponentially increases the simulation run execution time. Therefore, it is sufficient to set the simulation run time (including the warm-up period) at two weeks.

4.9 PRODUCTION RUNS AND ANALYSIS

After modelling a bay or group of bays using GTIA-M&S, the model is simulated to predict the performance of the intrabay AMHS under different operating conditions.

The production run objective must be clearly identified, as well as the number of runs needed.

GTIA-M&S reports the results to the database. Then, the results can be exported to an excel file for further analysis, if required.

A detailed description of this step is presented in the next chapter, which presents simulation experiments for analysing the performance of the intrabay AMHS running with different configurations.

4.10 DOCUMENTATION

Extensive additional documentation of the tool is not required, as almost every part of the tool is self-documenting. The six-step user guide to building models is straightforward, which promises fast model development with little effort. Furthermore, every component is accompanied with simple help about how to configure that component. In addition, the model-building template will get the user started with modelling right away, without being confused about which block to use and where to use it.

Finally, it is highly recommended that the user prepares the data required such as the length of the different loop segments, for modelling the part of the facility that he/she is interested in studying, before using the tool.

4.11 IMPLEMENTATION

The tool was used to generate the models for the following section of this work. These are based on actual plans for a new facility currently 80% complete. In addition, the GTIA-M&S has been reviewed by a number of production personnel, and a general acceptance of the tool has been reported. Indeed, the company intend to use it to generate several models of other AMHS/tool interaction so that greater understanding of them can be used in future design.

4.12 CONCLUSIONS

The following key points are observed during the tool development:

- Data collection and formatting is one of the greatest hurdles to any simulation project.
- The GTIA-M&S central database is the best solution for maintaining the different input/output data of a complex large model.
- Proof Animation™ is an effective method, which is carried out by the modeller, for determining the distances required to simulate any AMHS using CAD drawings only.
- Standard blocks for material handling systems supplied with Extend™ are not capable of modelling the intrabay AMHS.
- Developing a library of reusable components can be accomplished with the Extend™ simulation environment.
- Modelling of an intrabay AMHS is, not by any measure, an easy task because of the following:
 - Different types of entities are in motion in the model.
 - A large number of attributes are required to define each entity.
 - Several decision-making points are needed to control the flow of entities.
 - The narrow time window imposed by the high velocity of requires tight control of decisions.
- Combining flowcharts with IDEF0 shows the logic of decision-making and describes how the model works at the required level of detail.

5 SIMULATION EXPERIMENTS, RESULTS, AND ANALYSIS

To show the capabilities of the tool, two simulation experiments for comparing the different performance parameters of an intrabay AMHS operating with and without merge/diverge are undertaken.

5.1 DESCRIPTION OF EXPERIMENTS

The number of OHVs needed for running a bay or a group of bays is a major concern for the industry, as there are several factors that require consideration when specifying the number of OHVs. These factors can include the number of production tools in each bay, the type of the production tool (process, or metrology), the arrival rate of lots (from outside and inside the fab)...etc. As a result, simulation experiments are needed to factor all these parameters and predict the effect of changing the number of OHVs on the major performance indices.

In addition, due to the presence of two technologies applied to the intrabay AMHS, the experiments must explain the difference between the two technologies, showing the pros and cons of each technology.

5.1.1 Purpose

The purpose of the experiments conducted is twofold:

- First, to predict the number of OHVs required to attain specified target delivery times and report the number of moves per hour achievable when using this number of OHVs.
- Second, to analyse the performance of an intrabay AMHS running with and without merge/diverge, by comparing additional performance measures.

5.1.2 Model Developed

A model of 3 bays with different production tools, number of stockers, and track lengths is built as illustrated in Figure 5.1. It must be noted that each of the three bays comprises a different number and types of production tools having different run rates. Moreover, some of these tools are metrology tools, which means that sampling of lots takes place and production steps can be skipped for such tools. The first bay, bay 01, has the largest number of production tools and the longest track. Moreover, bay 01 has the largest number of process production tools, while bay 02 has the largest number of metrology production tools.

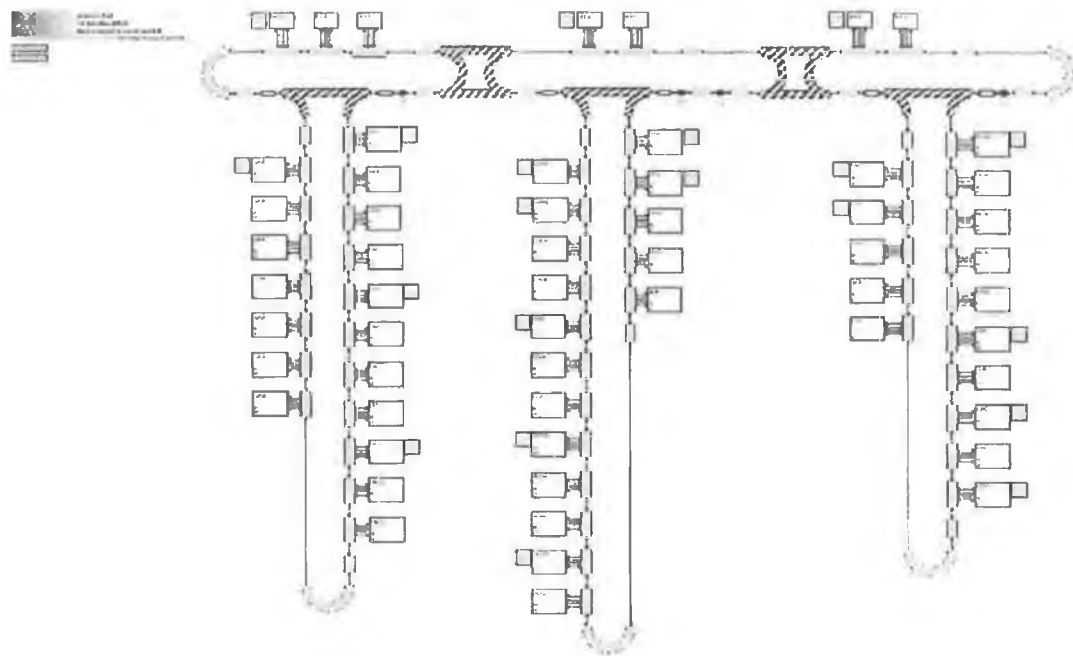


Figure 5.1: Developed model.

5.1.3 Performance Measures

There are two viewpoints from which the performance of the intrabay AMHS is evaluated:

- From a production tool perspective (how fast will it serve the tool)
- From an intrabay AMHS operation perspective (delivery times, waiting times, congestion...).

The industry is always concerned about the delivery times and the number of physical lot moves that can be achieved. However, for detailed analysis purposes those two measures will not suffice to evaluate and understand the behaviour of the system, hence, other measures that are useful in analyzing the intrabay AMHS must be used. The performance parameters selected are:

- **Stocker-to-Tool Delivery Time:** which comprises waiting time of lot at stocker to be picked up, pickup time at stocker, drop-off time at tool, and delays due to track traffic. This performance measure is typically set at a target value of 4 – 8 minutes.
- **Stocker-to-Tool Transportation Time:** which is the transportation time from stocker to tool and it only includes delays due to traffic on the intrabay track. Large variation in the transportation time value is a good indicator of any congestion that can occur at any point of an intrabay loop.
- **Lot Waiting Time at Stockers:** which is the queuing time of lots ready for pickup at stockers.
- **Lot Waiting Time at Production Tools:** which is the queuing time of lots ready for pickup at tools.
- **Moves Per Hour (MPH):** a move is the physical transport of a lot from stocker to tool or from tool to stocker, and moves per hour is the total number of moves in a bay that occur in one hour. Moves per hour essentially depend on the number and type of production tools in a bay.

The model produces a mean and a standard deviation for each of the time parameters mentioned before. The 3-sigma peak value of time is used to represent the expected worst case. However, the model reports the cumulative MPH that occurs in a bay every hour. From the cumulative value, the number of moves in each hour of simulated time in that bay is computed. After that, using these calculated MPH values, the mean and standard deviation are estimated and the 2-sigma peak value is used to speculate the moves per hour achievable.

Distribution of Time Parameters

Figure 5.2 shows a plot of approximately 4,500 data points for delivery time recorded in a three bay model running with 16 OHVs and merge/diverge set to ON. The plot shows a number of points that are above the 3-sigma limit, and the remaining points lying between the upper and lower limits.

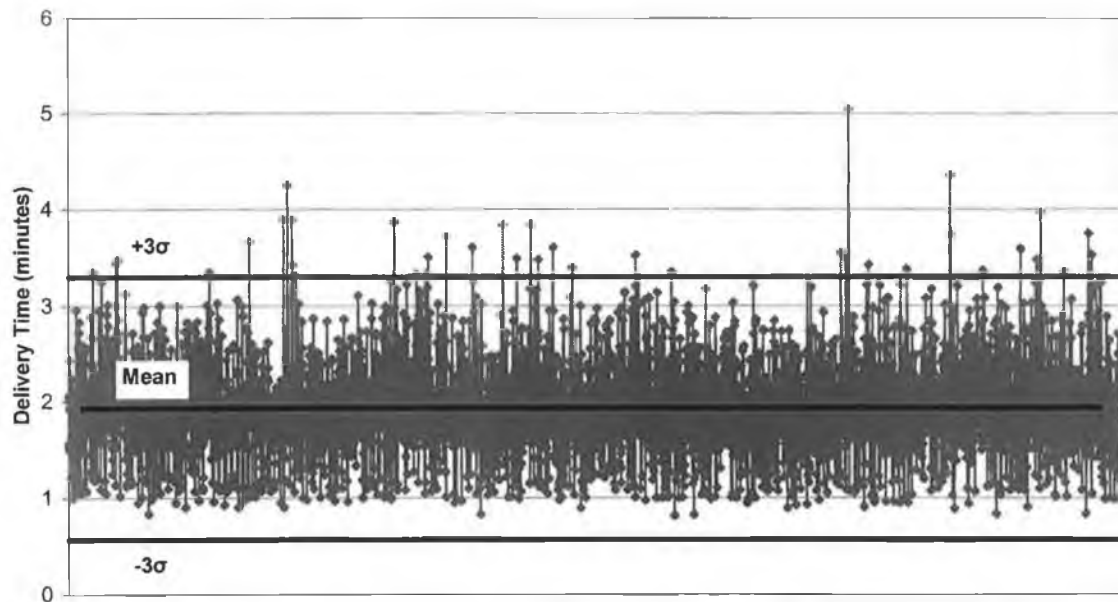


Figure 5.2: Delivery time analysis – Bay 01 – 16 OHVs – merge/diverge ON.

Consequently, further analysis of the delivery is carried out; distribution of the delivery time, shown by the histogram in Figure 5.3, reveals that the proportion of the points lying above the upper limit (39 points) is negligible compared to the rest of the data. In that case, it can be assumed that the 3-sigma value represents the data (since $39/4500 < 1\%$, therefore 99% probability level represents the data).

It should be noted that the problem lies in the fact that the firm is assuming a normal distribution of all the data, while most of the data exhibit a certain amount of skewness, due to the fact that there are absolute minimum times for a particular event, but no limitations on maximum times.

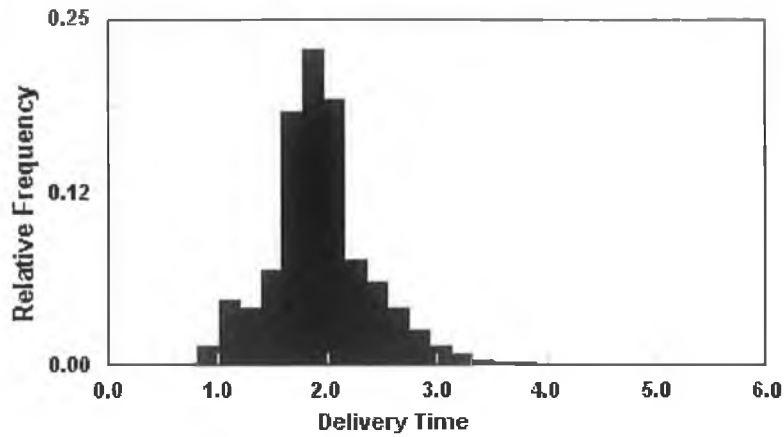


Figure 5.3: Delivery time distribution.

Distribution of Moves per Hour

The same issue is found with the moves per hour distribution. As shown in Figure 5.4, the 3-sigma value doesn't represent the majority of the data, or even the indicative maximum, and therefore the 2-sigma value is used for evaluating the moves per hour.

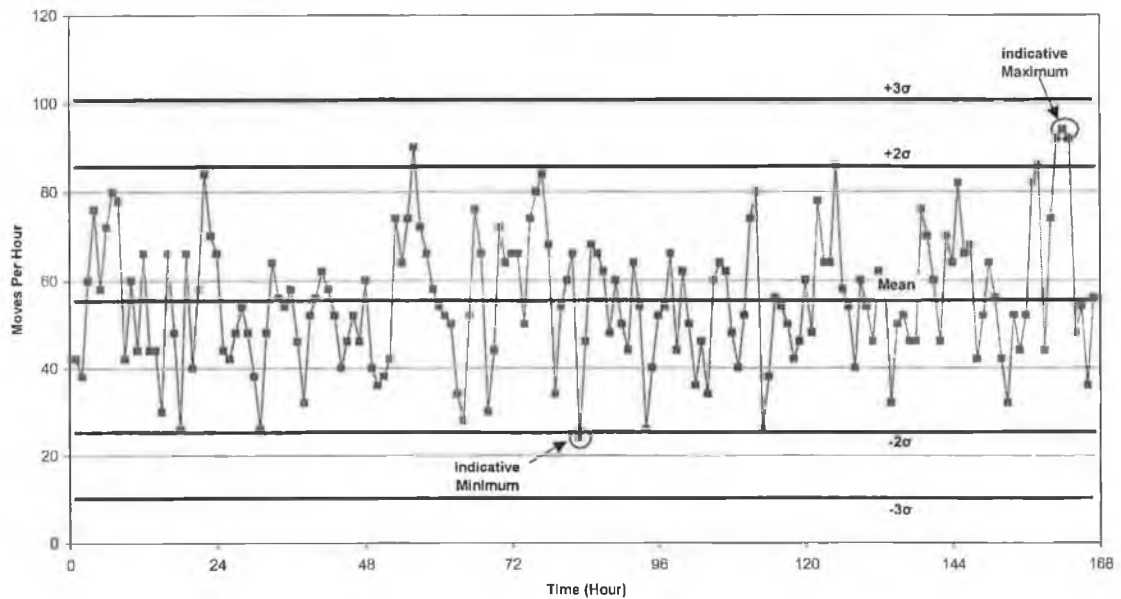


Figure 5.4: Moves per hour analysis – Bay 01 – 16 OHVs – merge/diverge ON.

5.1.4 Simulation Parameters

The length of the simulation runs have been set to two weeks, with one-week warm-up period, and lot arrival rate of 0.84 lots/hour. Lot arrival rate corresponds to 5,000 wafer starts per week (WSPW).

A check is made to ensure that the selected warm-up period guarantees collection of data under steady state conditions. Figure 5.5 shows the change in the moves per hour value over a two-week simulation run and confirms that a steady state is reached after one week (168 hours).

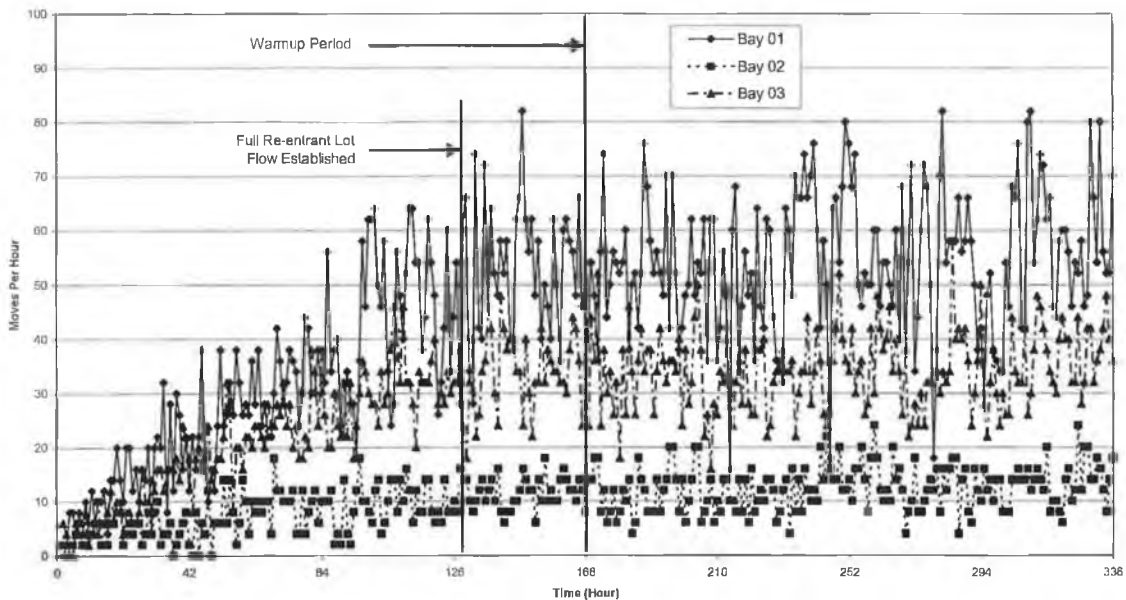


Figure 5.5: Using the MPH for the three bays to determine the warmup period.

It must be noted that the model is set in a way to end the manufacturing process after all the processing steps in the three bays are completed, which means lots exit the model after approximately 128 hours (lot cycle time). This assures that the selected warm-up period is sufficient for the simulation experiments.

Setting the length of the simulation run to two weeks would require a range of half an hour to 4 hours of computer time to complete one replication depending on the number of OHVs. The greater the number of OHVs the longer it takes to complete a simulation run. Furthermore, it must be noted that these values are reported for the model

developed here and for OHVs number varying from 3 to 16. Changing the size of the model and number of OHVs will have a direct effect on the simulation run time.

Finally, CRN is used for all the experiments to compare the intrabay AMHS running with merge (ON/OFF), and thus one replication for every data point is used. However, if a single system is studied it is recommended to perform independent runs of that system, under same operating conditions, to check the variability of the output value (see Figure 5.6).

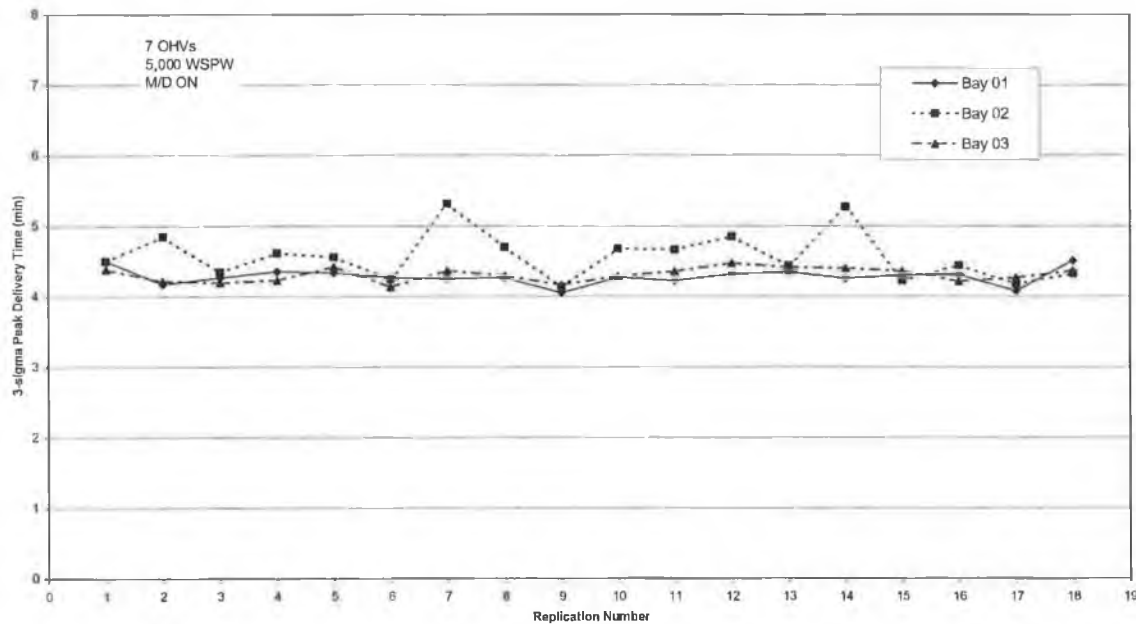


Figure 5.6: Different replications for a model running under same operating conditions.

5.2 EXPERIMENT ONE: INTRABAY AMHS WITHOUT MERGE/DIVERGE

The first experiment ran the developed model without the merge/diverge, and varied the number of OHVs from 3 – 16. Because the model in this case is running without merge/diverge, the number of OHVs in each bay is first determined based on values obtained from the existing “moves per hour” deterministic simulation model. The ratio of expected moves per hour for Bay 01 : Bay 02 : Bay 03 is 3 : 1 : 2. Consequently the number of OHVs in each bay is as illustrated in Table 5-1.

Table 5-1: Number of OHVs set in each bay without merge/diverge.

Total Number of OHVs	Based on 3:1:2 Proportion			Number of OHVs Used		
	Bay 01 OHVs	Bay 02 OHVs	Bay 03 OHVs	Bay 01 OHVs	Bay 02 OHVs	Bay 03 OHVs
3	1.5	0.5	1.0	1	1	1
4	2.0	0.7	1.3	2	1	1
5	2.5	0.8	1.7	2	1	2
6	3.0	1.0	2.0	3	1	2
7	3.5	1.2	2.3	4	1	2
8	4.0	1.3	2.7	4	1	3
9	4.5	1.5	3.0	4	2	3
10	5.0	1.7	3.3	5	2	3
11	5.5	1.8	3.7	5	2	4
12	6.0	2.0	4.0	6	2	4
13	6.5	2.2	4.3	7	2	4
14	7.0	2.3	4.7	7	2	5
15	7.5	2.5	5.0	7	3	5
16	8.0	2.7	5.3	8	3	5

5.2.1 Determining the Number of OHVs Required

Figure 5.7 shows the effect of changing the number of OHVs on the 3-sigma peak stocker-to-tool delivery time. The chart shows that running the three bays with 9 OHVs can guarantee a stocker-to-tool delivery time between 4 and 8 minutes. However, the number of OHVs selected is 10 to ensure that the delivery time remains within limits should one vehicle go down.

Further examination of the chart (Figure 5.7) indicates unbalanced delivery time performance at lower number of OHVs, which is due to restricting a certain number of OHVs to service a certain bay. For example, setting the number of OHVs to service a bay with high output, like bay 01, to one or two will not be enough to service this bay, leading to the high values of delivery time ‘spikes’ displayed in the chart. Having bay 01 suffering from the low number of OHVs makes that bay a bottleneck (Bay 01 bottleneck) for the rest of the bays and can lead to interrupts in the material flow of other bays. The same example is evident in Bay 03 bottleneck, in which bay 03 doesn’t have enough OHVs to service it, and thus it interrupts material flow to bay 01.

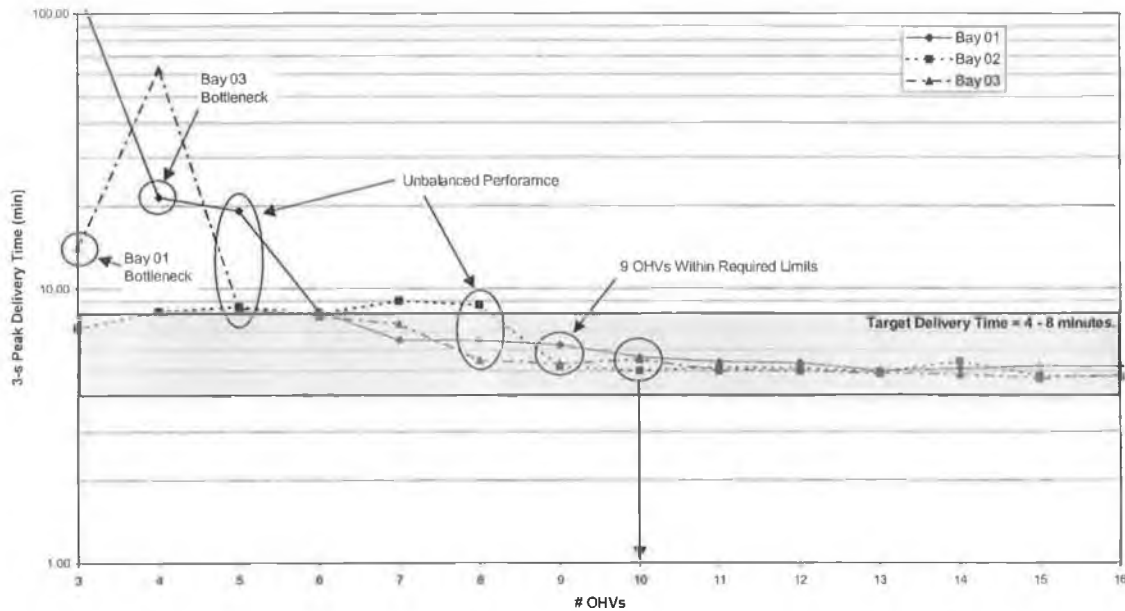


Figure 5.7: Delivery time versus number of OHVs without merge/diverge.

Finally, if it assumed that 6 OHVs are sufficient to satisfy the delivery performance target, and one OHV breaks down forcing the system to run with only 5 OHVs very high delivery times can be expected due to the unbalanced performance between the bays.

Testing Other Combinations of OHV Number in Bays

The values presented in Table 5-1 are used as a guideline only, and other combinations of OHVs number in each bay, for the critical points, should be tested.

Figure 5.8 presents different combinations of 7 OHVs in the three bays. Looking at this figure, it is clear that bay 01 requires at least 4 OHVs to achieve the target delivery time. Furthermore, bay 02 requires 2 OHVs because assigning 1 OHV only to this bay results in delivery times above target level. Finally, bay 03 should have at least 2 OHVs, as only 1 OHV in bay 03 results in drastic decrease in delivery time performance.

This shows that 7 OHVs, with all the tested combinations, are not enough to achieve the delivery time target.

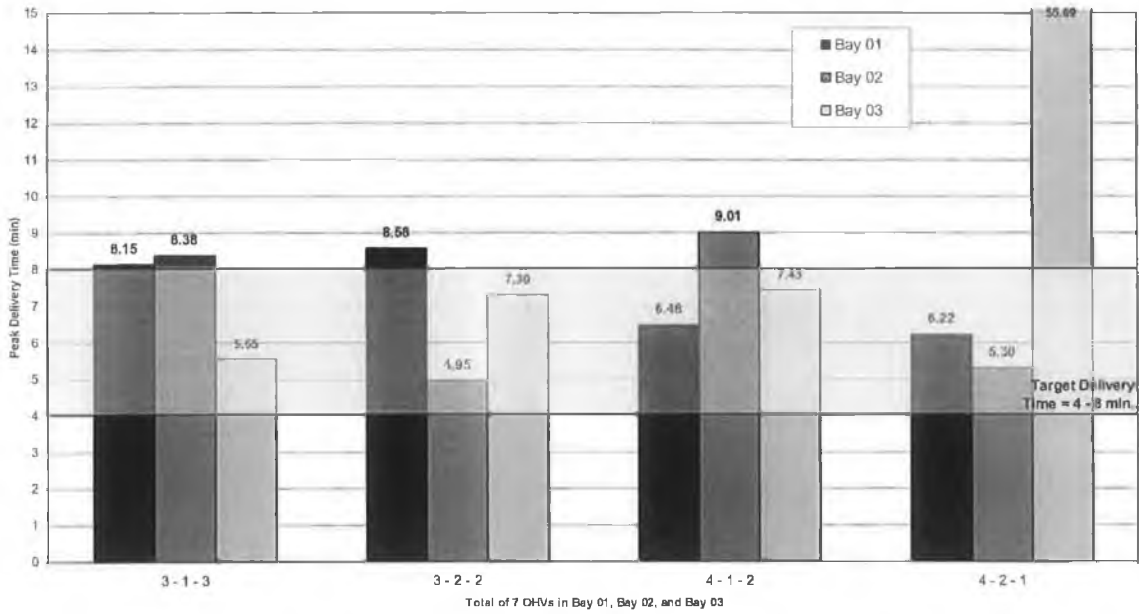


Figure 5.8: Different combinations of 7 OHVs.

The same test is carried out for different combinations of 8 OHVs. Figure 5.9 shows that by using 4 : 2 : 2 OHVs in bay 01 : bay 02 : bay 03 can accomplish the target delivery time using 8 OHVs, while proportional distribution based on the supplied data does not.

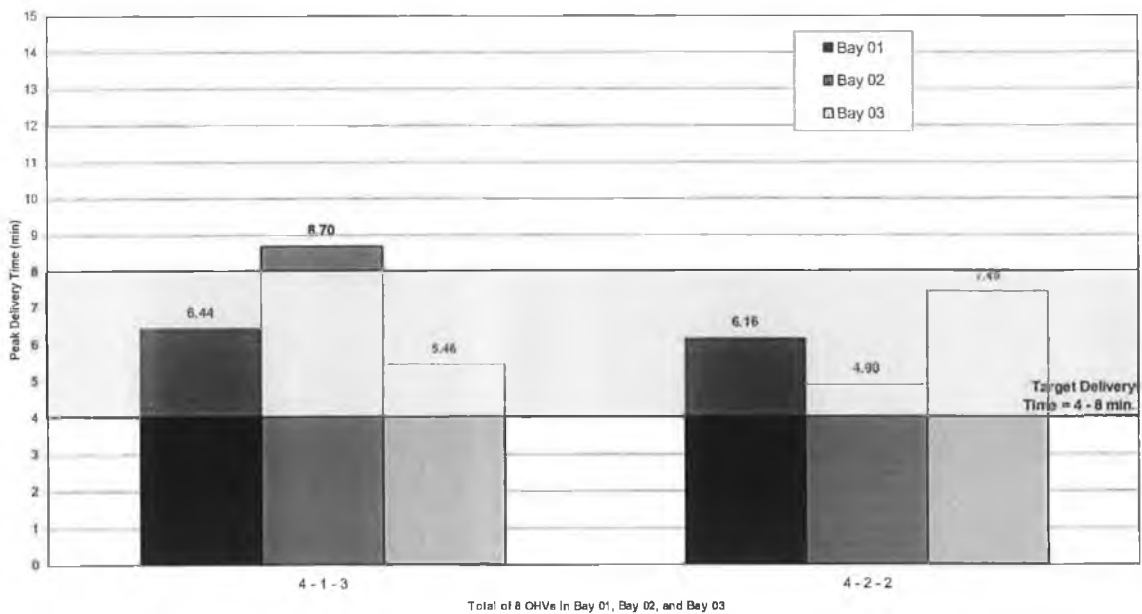


Figure 5.9: Different combinations of 8 OHVs.

The result of the previous test is implemented for the delivery time values. Figure 5.10 shows a modified chart of the effect of changing the number of OHVs on the 3-sigma peak delivery time. The chart now shows that running the three bays with 8 OHVs can guarantee a stocker-to-tool delivery time between 4 and 8 minutes. However, the number of OHVs selected is 9 to guarantee safe operating conditions.

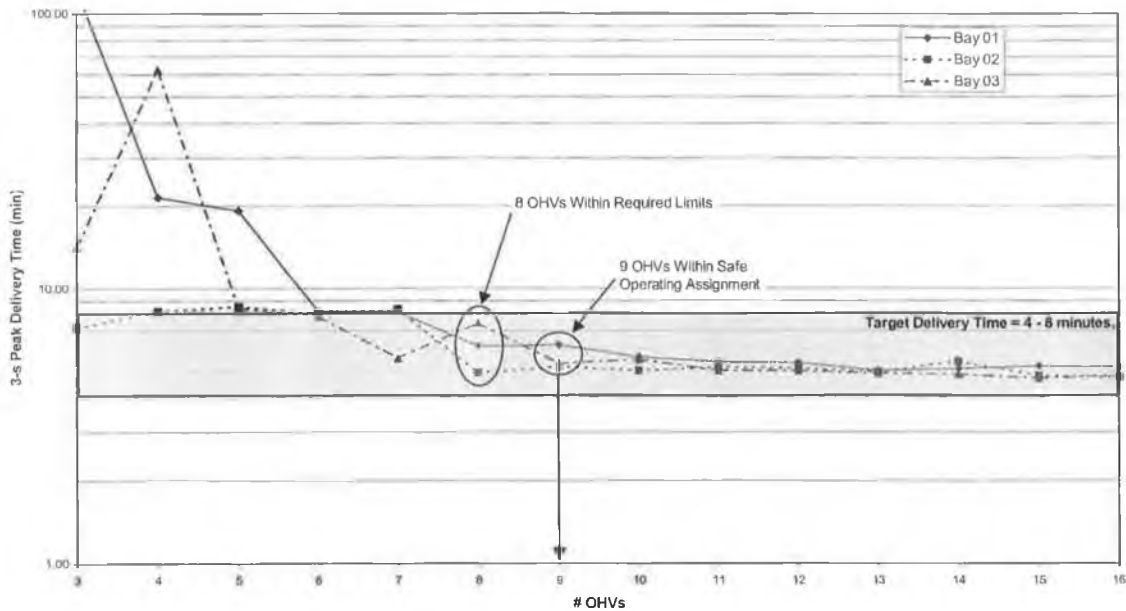


Figure 5.10: Delivery time versus number of OHVs without merge/diverge, after modification.

5.2.2 Determining the Achievable Moves Per Hour

The numbers of moves per hour (MPH) against the number of OHVs simulated without merge/diverge are illustrated in Figure 5.11. The moves per hour, for all bays, start at low values when the number of OHVs is low. Then the MPH increases till it reaches a near steady state (small variations at higher number of OHVs). This indicates that the maximum achievable MPH level is reached, which corresponds to the limitation by production tools.

The chart clearly shows that since bay 01 holds the largest number of production tools, most of which are processing tools, therefore the moves per hour in that bay is the highest. On the other hand, since bay 02 holds the largest number of metrology tools, for which production steps can be skipped, that bay has the lowest number of moves per hour.

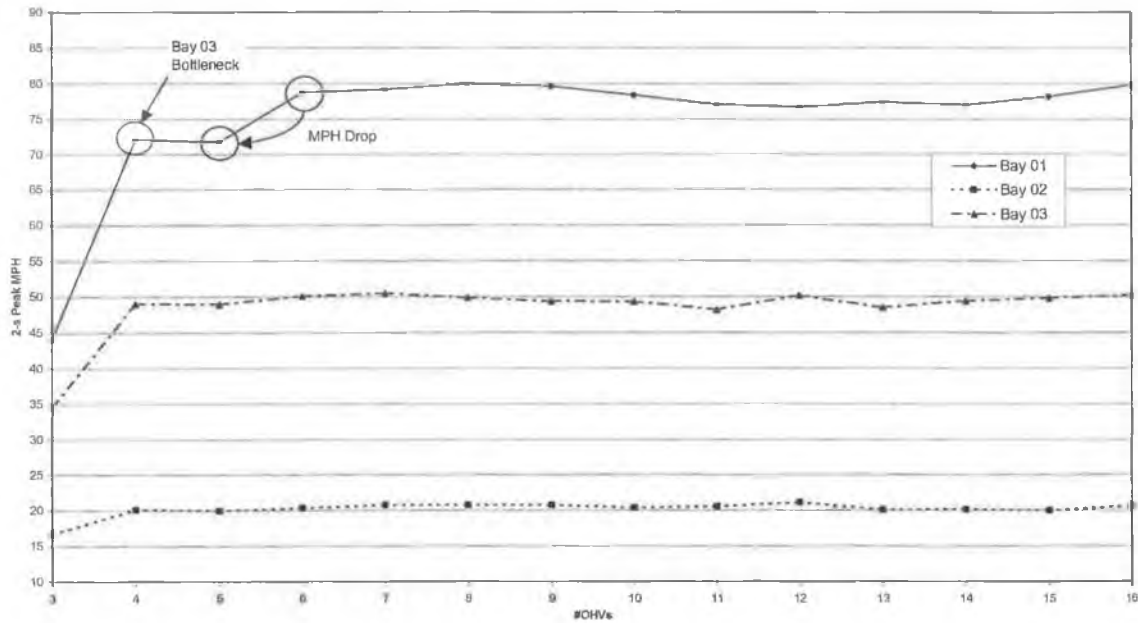


Figure 5.11: Moves per hour achieved without merge/diverge.

Furthermore, reducing the number of OHVs from 6 to 5 shows a drop in the MPH, because the number of OHVs in bay 01 is no longer sufficient to serve it. Also, at 4 OHVs, MPH for bay 01 is almost the same as for 5 OHVs because bay 03 in that case is unable to move the lots out, as the number of OHVs serving bay 03 is insufficient. It must be noted that those two deductions comply with the delivery time performance, discussed previously.

5.3 EXPERIMENT TWO: INTRABAY AMHS WITH MERGE/DIVERGE

The second experiment ran the developed model with the merge/diverge turned ON, and varied the number of OHVs from 3 – 16. Because the model in this case is running with merge/diverge, there is no need to determine the number of OHVs in each bay and hence, the number of OHVs varied is per merge/diverge loop (linking the three bays together) rather than per bay.

5.3.1 Determining the Number of OHVs Required

Figure 5.12 shows the effect of changing the number of OHVs on the 3-sigma peak stocker-to-tool delivery time. The chart shows that running the merge/diverge loop with 6 OHVs can guarantee a stocker-to-tool delivery time between 4 and 8 minutes. However, the number of OHVs selected is 7 to ensure that the delivery time remains within limits should one vehicle breakdown.

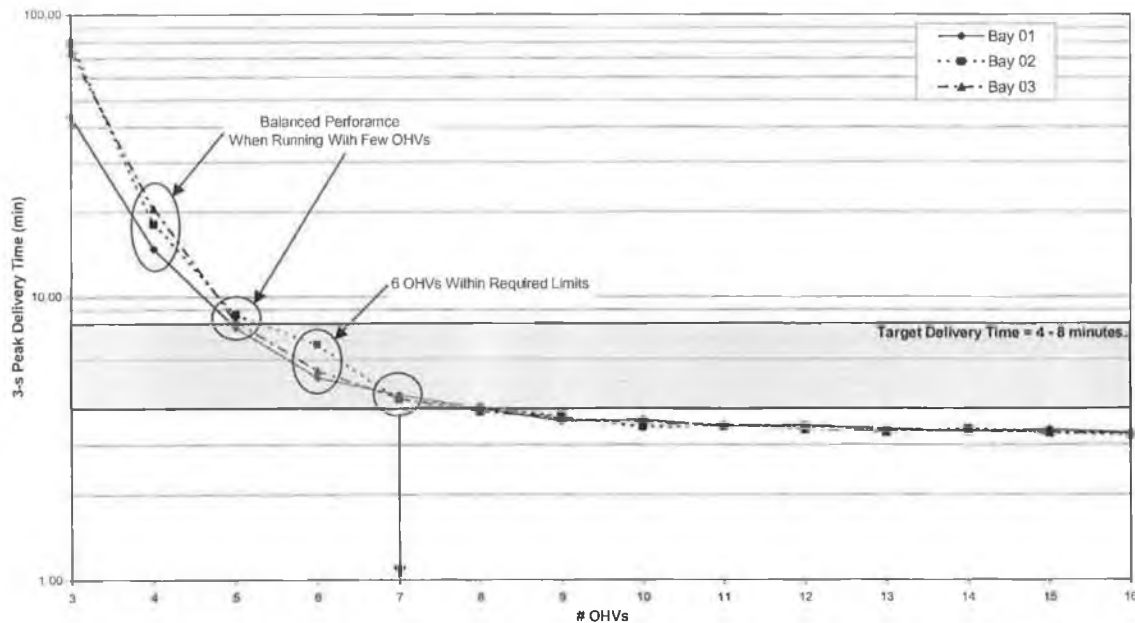


Figure 5.12: Delivery time versus number of OHVs with merge/diverge.

Examination of the chart shows that the merge/diverge intrabay AMHS can achieve delivery times that are lower than the stocker-to-tool delivery time target. In addition, the balanced performance for all bays means that when an OHV breaks down, the breakdown effect is distributed over the rest of the bays.

Finally, with the merge/diverge ON no bottlenecks have been identified for any of the three bays, which is again due to the balanced performance of all bays.

5.3.2 Determining the Achievable Moves Per Hour

The number of moves per hour achieved with varying the total number of OHVs in the merge/diverge loop is illustrated in Figure 5.13. The moves per hour stabilises quickly

at 4 OHVs, with little variation in the MPH value with increasing the OHVs number, indicating that the achievable MPH in bays is reached.

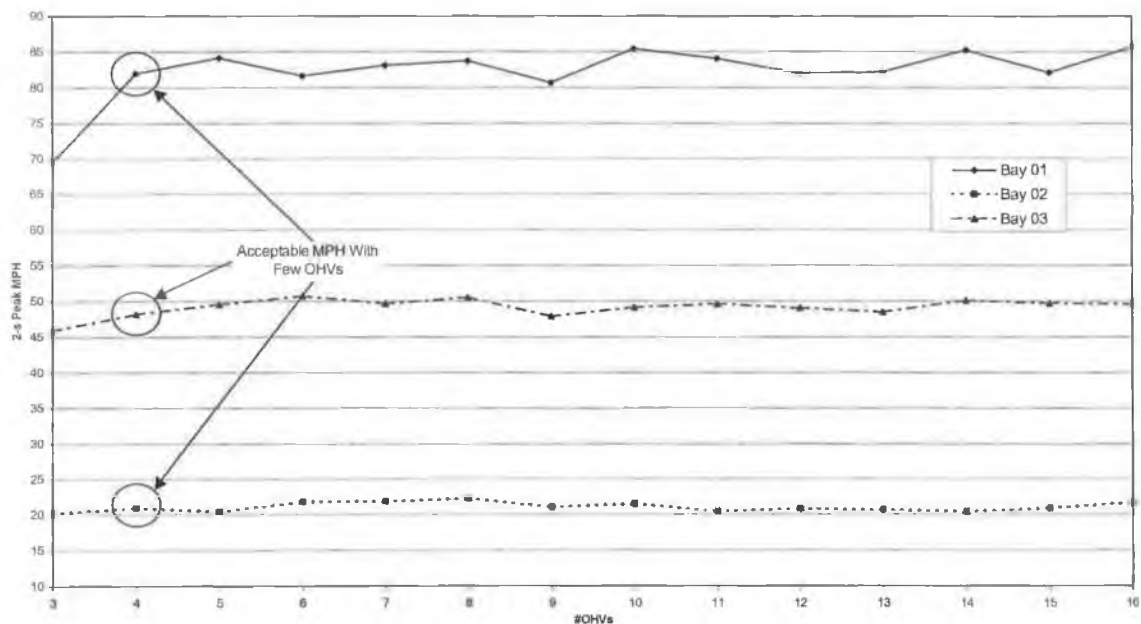


Figure 5.13: Moves per hour achieved with merge/diverge.

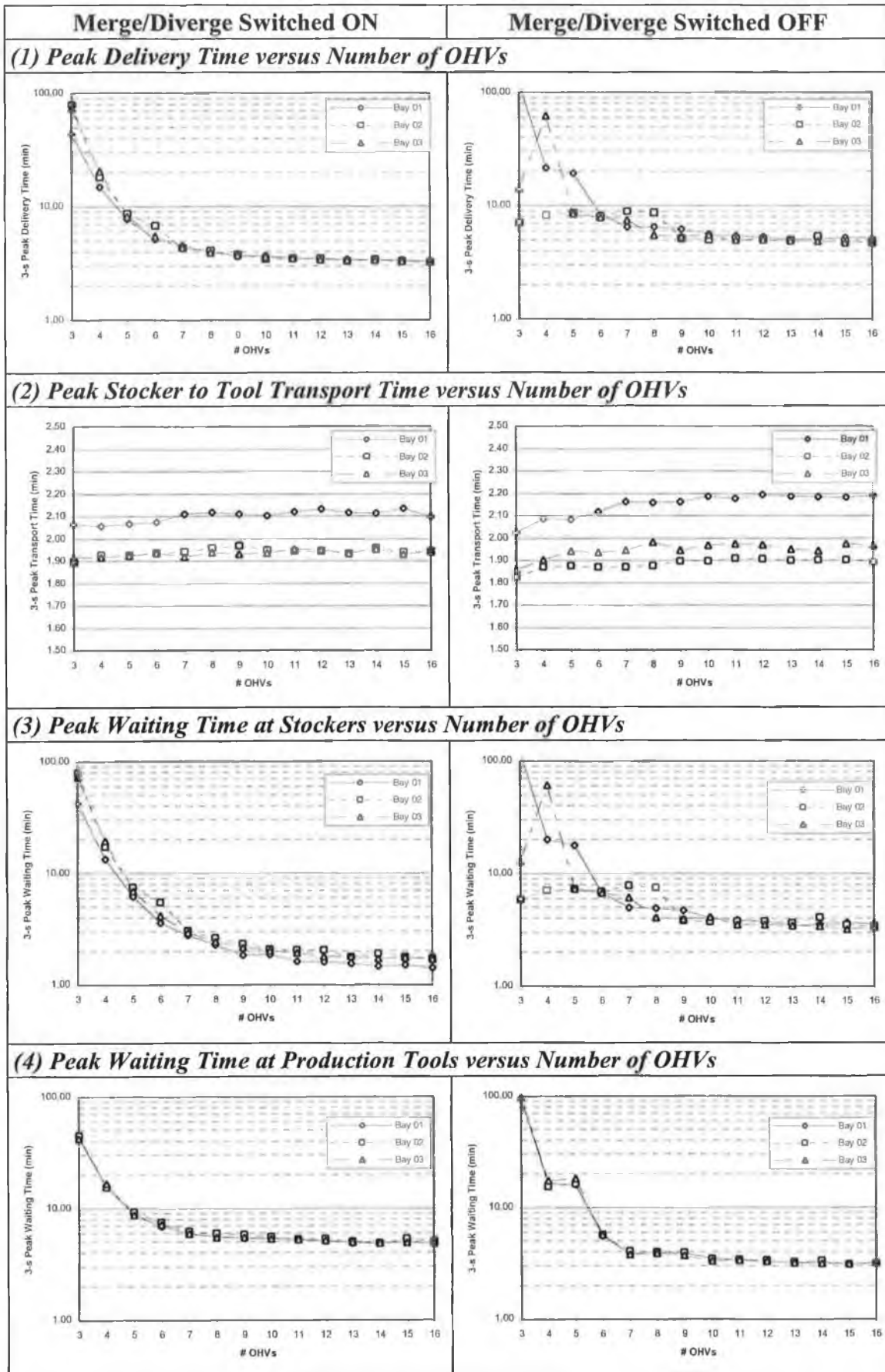
Once again, no signs of any sudden drops in the MPH value, indicating that no material interrupts have occurred.

5.4 COMPARING INTRABAY AMHS PERFORMANCE

After determining the number of OHVs required and the achievable MPH in each bay, with and without merge/diverge, critical comparison of the two technologies is the next step. The comparison presented here is also based on the output results of the previous experiments; however, the remaining performance measures are considered as well.

Variations of all measures with respect to changing the number of OHVs are plotted and presented in Table 5-2. The table shows every performance measure obtained from the simulation experiments for merge/diverge ON and OFF.

Table 5-2: Comparing the results of the two experiments.



5.4.1 Analysing the Performance of the Two Technologies

The previous discussion, related to the stocker-to-tool delivery time, clearly shows that in case of merge/diverge better delivery performance can be achieved. What is not clear is that OHVs travel on the same track for lot delivery in both cases (refer to the routing rules in 3.7.2 above), yet the lots reach the tools faster in case of merge/diverge ON.

To determine the cause of this phenomenon, the delivery time is broken down to its different components, which are waiting time at stockers, pickup time at stocker, stocker to tool transportation time, and drop-off time at tool. Pickup and drop-off times are very similar in both cases and won't be considered for further analysis.

Changing the number of OHVs has a similar effect on the peak stocker to tool transport time with and without merge/diverge, respectively. This is understandable because the comparable distances travelled in both cases, and any delays due to traffic would probably be an OHV picking up or dropping off a lot; again, these are very slight and lead to similar delays in both cases.

Looking at waiting times at stockers versus the number of OHVs with and without merge/diverge respectively (Table 5-2), it is quite clear that in case of merge/diverge the waiting time of lots ready for pickup is much lower. This is due to the fact that free OHVs run on the outer loop linking the three bays together, thus free OHVs are always close to stockers and can pick up ready lots much faster than is the case without merge/diverge where free OHVs are looping inside the bays away from the stockers.

On the other hand, without merge/diverge OHVs are looping inside the bays near the tools, which means that pickup performance at the tools should be better without merge/diverge. The charts in Table 5-2 confirm the better pickup performance at tools by OHVs without merge/diverge.

5.4.2 Comparing the Two Technologies for the Same Number of OHVs

The total number of OHVs allocated for the three bays is fixed and the values of the performance measures are compared. The number of OHVs selected is 7, which has been determined in 5.3.1 above for the intrabay AMHS running with merge/diverge.

Comparing Results for Each Bay

The outcomes of the simulation runs for bay 01, bay 02, and bay 03 are shown in Figure 5.14, Figure 5.15, and Figure 5.16, respectively. Having merge/diverge switched ON, all the figures indicate the decrease in the stocker-to-tool delivery time due to the decrease in waiting time at stockers. On the other hand, having merge/diverge switched OFF reduces the waiting time at tools significantly.

The stocker-to-tool transportation time in bays 01 and 03 is lower with merge/diverge; however, it is higher in bay 02. The improvement in this measure experienced in bay 01 and 03 is due to the fact that free OHVs are running on the outer loop leaving the inside of these bays almost traffic free, which is true for bay 02 as well.

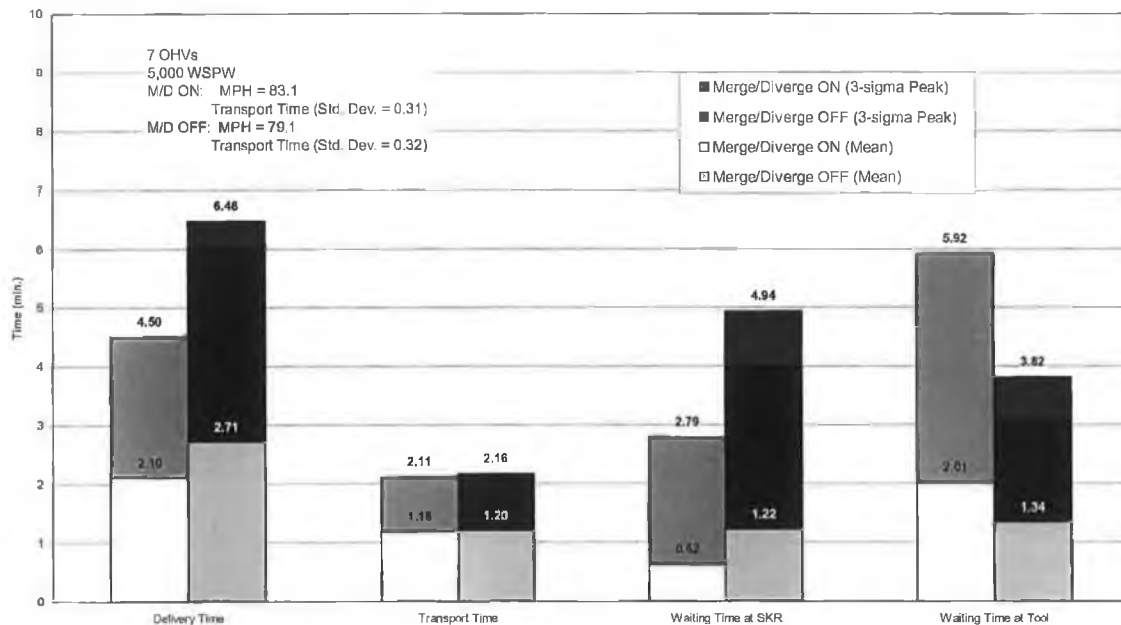


Figure 5.14: Bay 01 results when running with 7 OHVs.

Yet the increase in bay 02 transport time (both peak value and standard deviation) with the merge/diverge switched ON, depicted in Figure 5.15, reflects variations in transportation time due to congestion that occur at some point on the intrabay loop.

This can only mean that the inconsistency in the delivery time is not due to traffic inside the bay itself (as the case with the other bays), but rather due to traffic on the outer loop specifically at the bay entry point or at the merge/diverge link. Thus, the bay(s) that lie in the middle of a merge/diverge link are the most likely to suffer, and the first bay in a

merge/diverge link is the least likely to suffer, from any traffic congestion on the outer loop.

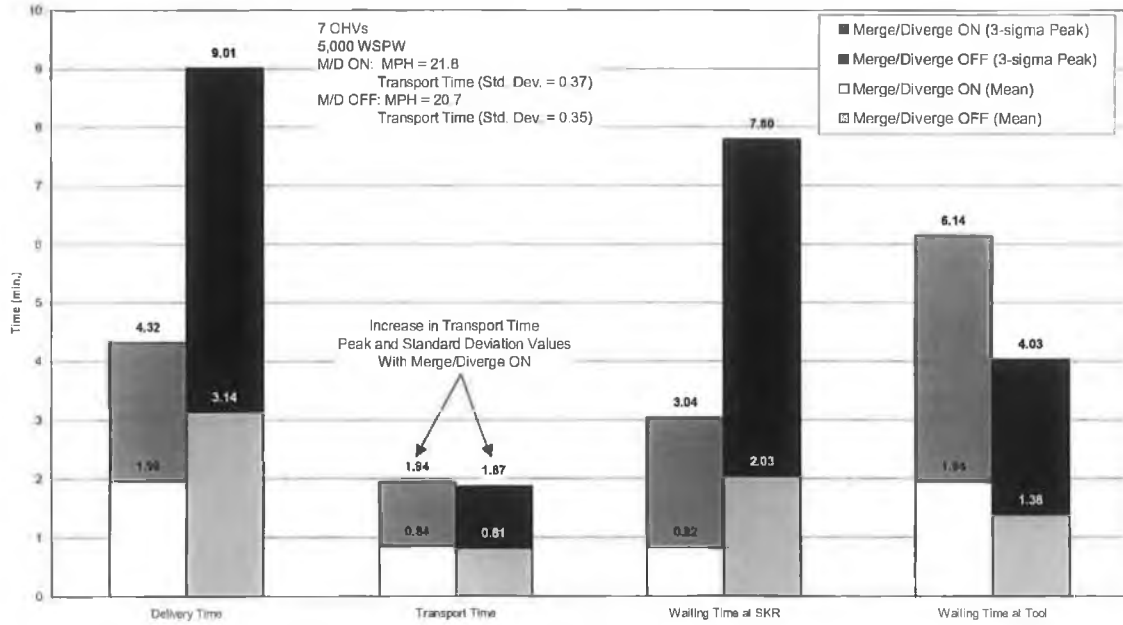


Figure 5.15: Bay 02 results when running with 7 OHVs.

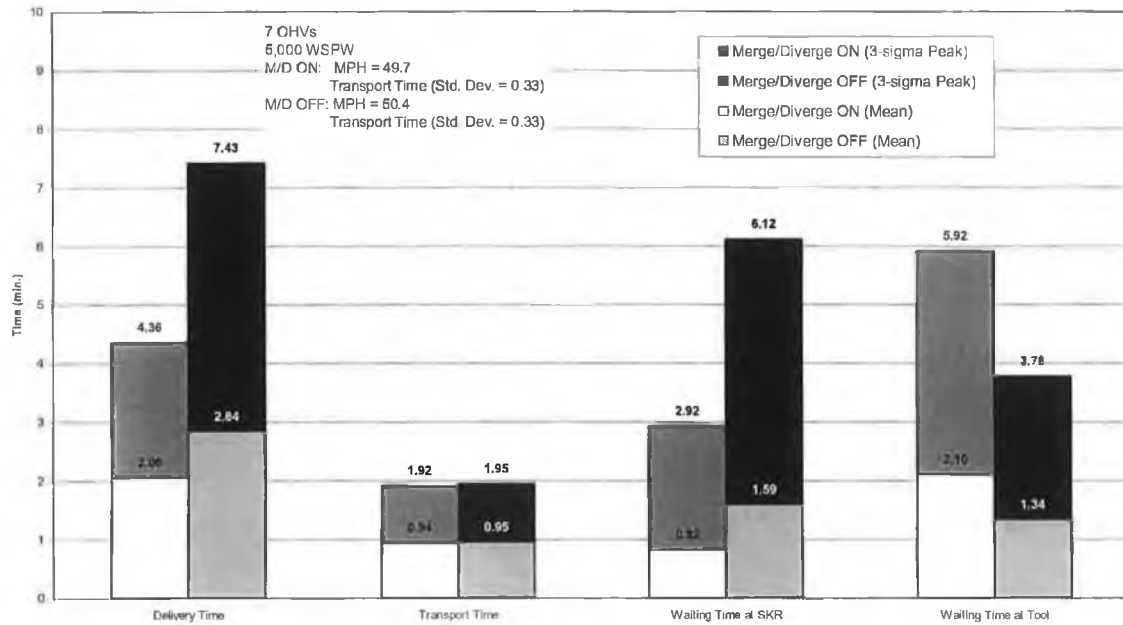


Figure 5.16: Bay 03 results when running with 7 OHVs.

Improvements with Merge/Diverge in Each Bay

The chart in Figure 5.17 shows the percentage improvement attained by adopting the merge/diverge intrabay AMHS. The negative percentage values for some of the measures indicate that those measures are better without the merge/diverge.

The chart shows that improvement of the waiting time at stockers and deterioration of the waiting time at production tools, almost cancel the effect of each other and result in slight increase (and even decrease in bay 03) in the achievable MPH.

Finally, the chart clearly shows that the middle bay in a merge/diverge loop suffers from an increase in the transportation time due to traffic at the bay entrance (merge/diverge link).

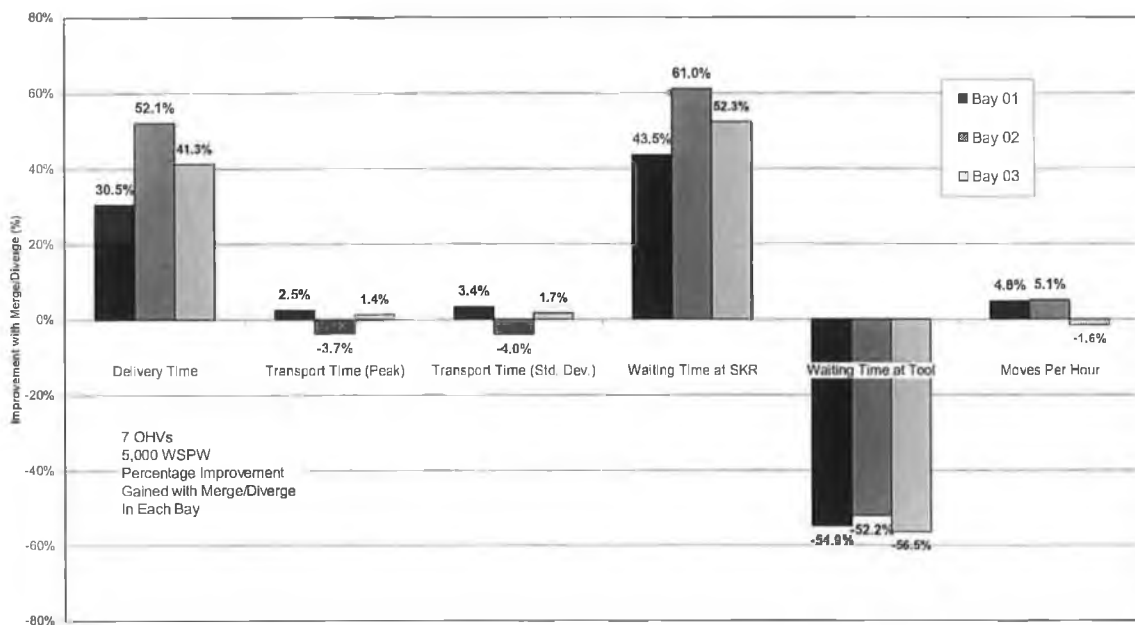


Figure 5.17: Improvement gained in each bay by having the merge/diverge switched ON.

Comparing Averaged Results Across Bays for Each Technology

After presenting the results of all bays separately, it is interesting to see the average performance of the three bays with and without merge/diverge. The average results across the three bays are shown in Figure 5.18.

Decrease of the stocker-to-tool delivery time, due to a decrease of the lot waiting time at stockers, occurs with merge/diverge switched ON. This decrease in delivery time is not due to an improvement in the lot transportation time, as the time in both cases is almost the same. Waiting time at production tools increased with the merge/diverge, which resulted in cancelling the effect of delivery time improvement. Consequently, the MPH increased slightly from 50 MPH to 51.5 MPH.

The 1.5 moves increase in the MPH translates into an extra 252 moves per week, which can have huge impacts on the factory. For example, an average process flow has 500 steps, meaning that completing manufacturing of a lot requires approximately 1,000 moves. Therefore, the extra 252 moves per week translates to an extra lot pushed out of the fab approximately every 4 weeks, keeping in mind that this lot has a combined retail value of well over a million dollars.

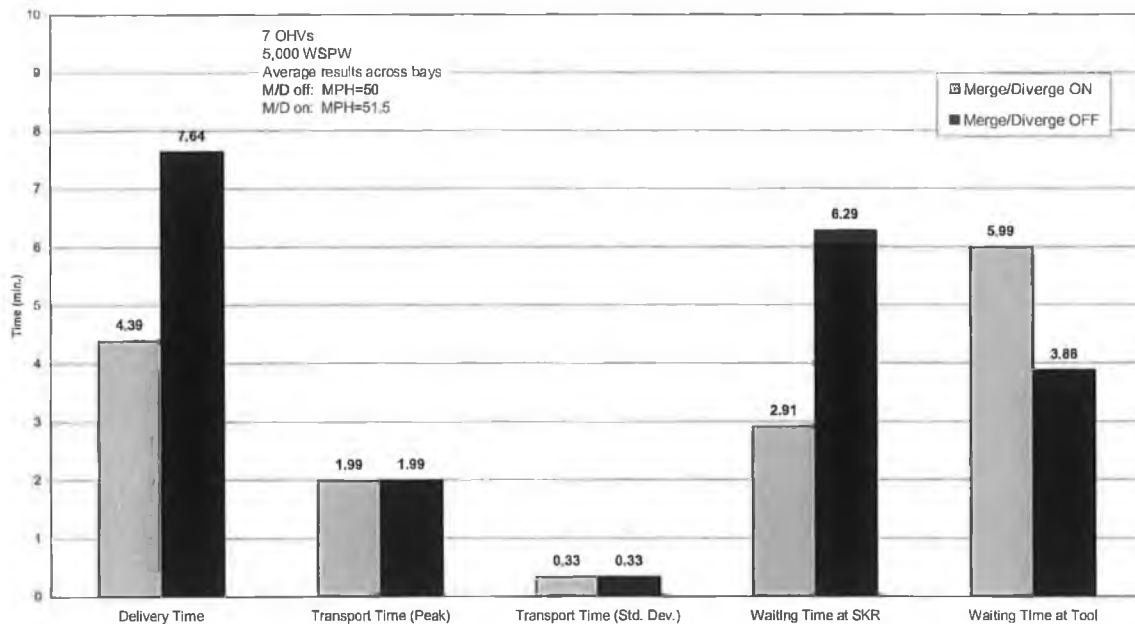


Figure 5.18: Average results across the three bays.

Figure 5.19 shows the percentage improvement gained with merge/diverge ON. The results shown in the figure are averaged across the three bays. The negative percentage value of the 'waiting time at tool' parameter means that this parameter is better with merge/diverge OFF.

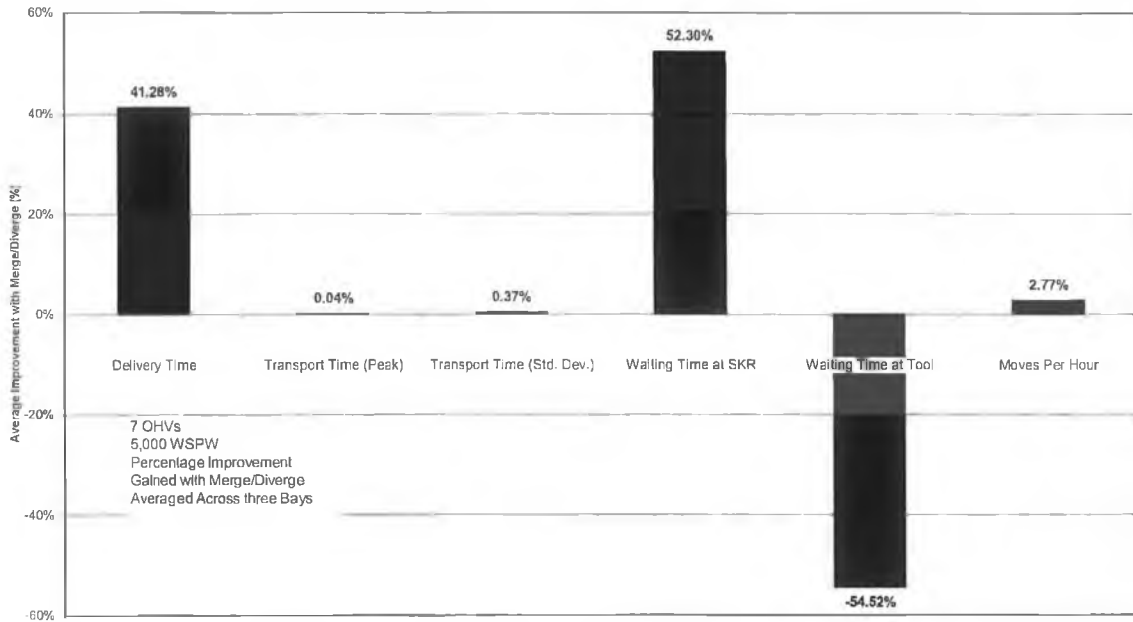


Figure 5.19: Overall improvement gained by having the merge/diverge switched ON.

5.4.3 Confirmation of the Results

In order to verify whether the improvement values reported from the single experiment in the previous section actually reflect the general behaviour of the model, ten different replications are conducted comparing merge/diverge ON/OFF using the constant random number approach (a total of 20 model runs). Examining the values in Table 5-3, which reports the result of these replications (averaged across the three bays), it is noticed that there is a consistency in the values of percentage improvement gained with the merge/diverge switched ON.

Table 5-3: Percentage improvement gained with merge/diverge ON for the different runs.

Run	Delivery Time			Peak Transport Time			Waiting Time at Stocker			Waiting Time at Tool			Moves Per Hour		
	M/D ON	M/D OFF	% Imprvmt	M/D ON	M/D OFF	% Imprvmt	M/D ON	M/D OFF	% Imprvmt	M/D ON	M/D OFF	% Imprvmt	M/D ON	M/D OFF	% Imprvmt
1	4.11	7.26	43.41	1.99	2.00	0.14	2.64	5.95	55.56	5.88	3.76	-56.22	49.63	47.65	4.00
2	4.26	7.05	39.62	1.98	1.99	0.18	2.79	5.71	51.13	6.20	3.75	-65.05	49.09	48.38	1.44
3	4.30	7.96	45.98	1.99	1.98	-0.22	2.81	6.69	58.00	6.17	3.81	-61.78	48.78	48.54	0.50
4	4.34	7.70	43.64	1.99	1.98	-0.38	2.90	6.40	54.76	6.16	3.84	-60.61	51.19	49.09	4.10
5	4.31	7.00	38.40	1.99	1.98	-0.27	2.83	5.67	50.18	6.03	3.84	-57.08	50.76	47.05	7.31
6	4.44	7.15	37.89	1.98	1.99	0.50	3.00	5.81	48.34	6.10	3.70	-64.89	50.40	48.26	4.25
7	4.37	6.97	37.33	2.00	1.99	-0.02	2.88	5.61	48.78	5.79	3.79	-52.87	49.50	46.81	5.44
8	4.47	7.50	40.35	1.98	1.98	-0.01	2.99	6.20	51.70	6.05	3.77	-60.26	49.99	48.29	3.40
9	4.24	7.37	42.56	1.99	1.99	0.09	2.73	6.06	54.90	5.99	3.78	-58.42	49.49	47.90	3.22
10	4.60	7.18	35.83	2.00	1.99	-0.40	3.12	5.85	46.68	6.01	3.94	-52.66	50.99	48.48	4.92
Mean	40.50%			-0.04%			52.00%			-56.97%			3.86%		
Std. Dev.	3.27%			0.28%			3.67%			4.41%			1.93%		

Further analysis of the improvement in the moves per hour value shows that based on a 2-sigma range (96% probability level), the value of the MPH improvement varies from

0.002% to 7.716%. However, using a 3-sigma range (99% probability level), the value of the MPH improvement varies from -1.927% to 9.645%. Therefore, at a 95% confidence level, merge/diverge ON results in a better performance.

Figure 5.20 presents the mean improvement values gained with merge/diverge switched ON, and shows the 2-sigma values for all the parameters as well.

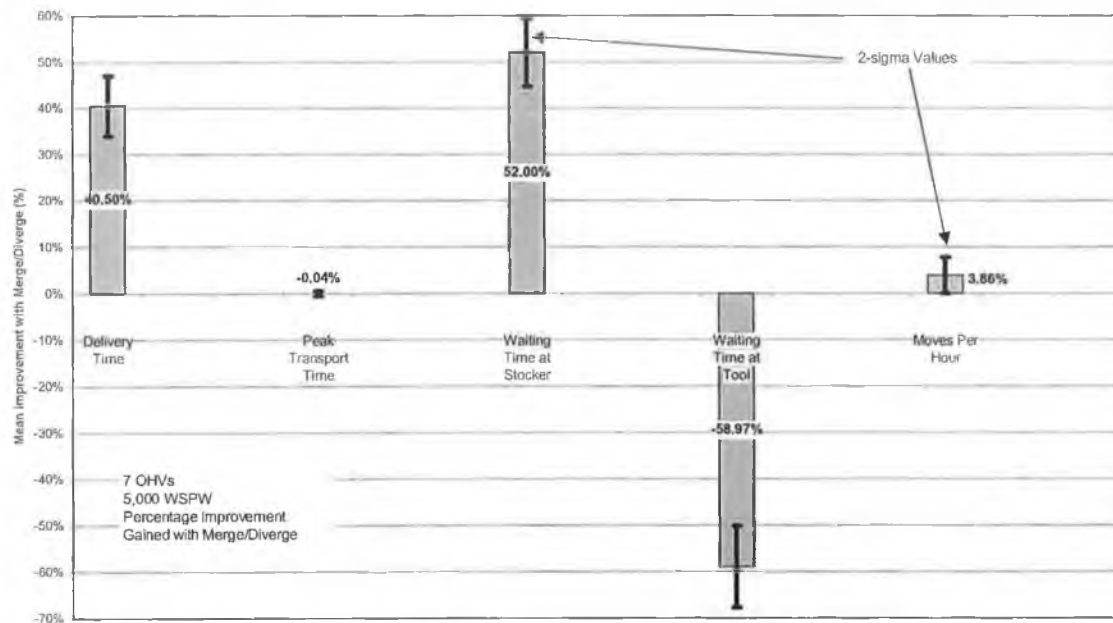


Figure 5.20: Mean and 2-sigma values of the percentage improvement gained with merge/diverge.

5.4.4 Summary of Analysis and Results

Merge/diverge technology results in improved delivery time because free OHVs are able to travel around the outer loop connecting the three bays, rather than going in and out of a bay and travelling long distances for no reason. Lots waiting at stockers have a better chance to be picked up faster with merge/diverge (as the OHVs are running on the outer loop close to the stockers) this directly lowers the lot waiting time at stockers (major constituent of the delivery time) and ultimately reduces the delivery time.

When determining the number of OHVs required in the three bays for achieving a specified delivery time target with and without merge/diverge, it is found that fewer number of OHVs are required to attain the same target with merge/diverge switched on. This results in huge savings of investments in the intrabay AMHS OHVs, making the merge/diverge technology more economical.

With merge/diverge balanced performance of all bays is noticed, which allows for easier planning for the merge/diverge intrabay AMHS. In addition, the balanced performance means predictable behaviour of the system, from an operational point of view. Furthermore, in case of OHV breakdown, any drop in the performance of the intrabay AMHS is distributed on all bays belonging to the merge/diverge link, leading to a reduced OHV breakdown effect. Finally, because of the balanced performance and reduced breakdown effect, the risk of running the system with fewer numbers of OHVs is reduced.

On the other hand, having all the vehicles running on the outer loop takes the OHVs away from any lot waiting collection at the production tools. That is why the waiting time at tools increased as OHVs travel longer distances to reach the lot ready for pickup. On the contrary, running the intrabay AMHS with the merge/diverge switched OFF, the OHVs are continuously looping inside the bay and have a better chance to collect lots from production tools.

Finally, the increase in the achievable moves per hour noticed with the use of merge/diverge is limited by the increase of the lot waiting time at production tools. MPH can be improved further if the lot waiting time at production tools is reduced, and this can be achieved through better routing rules for the free OHVs. The OHV supplier has come to the same conclusion, with separate studies of the system, and is looking into possible solutions for alternative routing rules of free OHVs.

Running the intrabay AMHS with the merge/diverge technology lead to:

- Improved delivery time.
- Fewer OHVs required.
- Balanced performance.
- Easier planning for intrabay AMHS.
- Reduced OHV breakdown effect.
- Waiting time at tools increased.
- Small but significant increase in achievable moves per hour.

6 CONCLUSIONS AND RECOMMENDATIONS FOR FUTURE WORK

6.1 CONCLUSIONS

Conclusions drawn from this work are:

1. Modelling in a rapidly changing environment is not a simple task.
 - Design changes are highly frequent.
 - The data is continuously changing.
 - Rapid model building is needed.
2. The developed tool, GTIA-M&S, has the following features:
 - Allows rapid configuration of new models.
 - Central generalised database holds key data for use in any model.
 - Components developed form a library, which is easily distributed to many users.
 - Capable of switching merge/diverge technology ON and OFF.
 - Models production tools explicitly as components.
3. Comparison of merge/diverge ON and OFF shows:
 - Stability of merge/diverge ON in balancing OHVs between bays.
 - Improvement of delivery time with fewer vehicles when merge/diverge is ON.

- Overall performance could be further improved with better routing rules.

The simulation project clearly shows that the development of the GTIA-M&S was successful. Production personnel are eager to use the tool as part of their decision making process for evaluating different intrabay AMHS designs. Furthermore, analysis of results obtained from the simulation experiments has increased their confidence in using the tool, as it confirms the current issues being addressed by the intrabay AMHS supplier.

6.2 RECOMMENDATIONS FOR FUTURE WORK

Development of models for the remaining bays in the facility using the tool.

Developing a neural network meta-model that can data mine different input/output configurations, which can be extracted from different models created by GTIA-M&S.

Validation of models developed by the tool with real system data whenever the real system is implemented.

Programming custom blocks for the library components, when the system design is fixed, to execute simulation runs faster.

7 REFERENCES

- [1] Rolstadas, A. "Enterprise Performance Measurement." International Journal of Operations & Production Management 18.9/10 (1998): 989-999.
- [2] Willis, T. H. "Operational Competitive Requirements for the Twenty-First Century." Industrial Management & Data Systems 98.2 (1998): 83-86.
- [3] Jordan, J. and Michel, F. Next Generation Manufacturing (NGM). Society of Manufacturing Engineers, 1999.
- [4] Williams, D. J. Manufacturing Systems: An Introduction to the Technologies. Halsted Press, 1988.
- [5] Zijm, W. H. M. "Towards Intelligent Manufacturing Planning and Control Systems." OR Spektrum 22 (2000): 313-345.
- [6] Shulka, C. S. and Chen, F. F. "The State of the Art in Intelligent Real-Time FMS Control: A Comprehensive Survey." Journal of Intelligent Manufacturing 7 (1996): 441-455.
- [7] Rehg, J. A. and Kraebber, H. W. Computer-Integrated Manufacturing. Second ed: Prentice-Hall, Inc., 2001.
- [8] O'Sullivan, D. "Development and Operation of FMS Using Structured Techniques." International Journal of Production Planning & Control 2.2 (1992): 122-134.
- [9] Kouvelis, P. "Design and Planning Problems in Flexible Manufacturing Systems: A Critical Review." Journal of Intelligent Manufacturing 3 (1992): 75-99.
- [10] Kalpakjian, K. M. "Fabrication of Microelectric Devices." Manufacturing Engineering and Technology. Ed. S. Kalpakjian. Third ed: Addison-Wesley Publishing Company, Inc., 1995. 1012-1031.
- [11] Moore, G. E. "Cramming More Components onto Integrated Circuits." Electronics 38.8 (1965): 114-117.
- [12] Moore, G. Original and Current Graph. Online. Intel Corporation. Available: ftp://download.intel.com/intel/intelis/museum/research/arc_collect/history_docs/pdf/original+graph.pdf. 25 September 2003.
- [13] Awtar, S. "Automated Material Handling Systems for the Back-End of Semiconductor Fabrication." Ph.D. Thesis: Initial Chapters. MIT.

- [14] Silverman, P. Advanced Technology Challenges. 2001. Online. Intel Corporation. Available: <ftp://download.intel.com/research/silicon/AMATSemicon.pdf>. 30 March 2003.
- [15] Moore, G. No Exponential Is Forever ... but We Can Delay 'Forever'. 2003. Online. Intel Research. Available: [ftp://download.intel.com/research/silicon/Gordon Moore ISSCC 021003.pdf](ftp://download.intel.com/research/silicon/Gordon%20Moore%20ISSCC%20021003.pdf). 21 March 2003.
- [16] Edgar, T. F., Butler, S. W., Campbell, W. J., Pfeiffer, C., Bode, C., Hwang, S. B., Balakrishnan, K. S., and Hahn, J. "Automatic Control in Microelectronics Manufacturing: Practices, Challenges, and Possibilities." Automatica 36.11 (2000): 1567-1603.
- [17] Wesrkämper, E., Schmidt, T., and Wiendahl, H.-H. "Production Planning and Control with Learning Technologies: Simulation and Optimization of Complex Production Processes." Knowledge-Based Systems: Techniques and Applications. Ed. C. T. Leondes. Vol. 3: Academic Press, 2000. 840-887.
- [18] Schoemig, A. K. "On the Corrupting Influence of Variability in Semiconductor Manufacturing." Proceedings of the 1999 Winter Simulation Conference. Ed. G. W. Evans. Phoenix, AZ, 1999. 837-842.
- [19] Fountain, T., Dietterich, T. G., and Sudyka, B. "Mining IC Test Data to Optimize VLSI Testing." Proceedings of the sixth ACM SIGKDD international conference on Knowledge discovery and data mining. Boston, MA, USA: ACM, 2000. 18-25.
- [20] Semiconductor Manufacturing Process. 2003. Online. International SEMATECH. Available: <http://www.sematech.org/public/news/mfgproc/mfgproc.htm>. 23 September 2003.
- [21] Seligson, D. Planning for the 300mm Transition. Q4. 1998. Online. Intel Technology Journal. Available: <http://developer.intel.com/technology/iti/q41998/pdf/300mm.pdf>. 24 September 2003.
- [22] Garrett, T. M. "Converting to 300 Mm." Proceedings of 1999 ISSM (International Symposium on Semiconductor Manufacturing), 1999.
- [23] Gargini, P. and Garrett, T. "Intel's 300 Mm Conversion." Proceedings of Third IFST (International Forum on Semiconductor Technology), 2000.
- [24] Quinn, T. and Bass, E. 300 Mm Factory Layout and Material Handling Modeling: Phase I Report. 1999. Online. International SEMATECH. Available: <http://www.sematech.org/public/docubase/document/3688beng.pdf>. 20 March 2003.
- [25] Gargini, P. and Pillai, D. Factory Considerations for High Volume Manufacturing Using 300 Mm Wafers. 1997. Online. Intel Research. Available:

- <ftp://download.intel.com/research/silicon/factoryconsiderations.pdf>. 18 March 2003.
- [26] Intel Activates Its 300 Millimetre Wafer Program. 9 June 1999. Online. Intel Press Room. Available: <http://www.intel.com/pressroom/archive/releases/Cn060999.htm>. 24 April 2003.
- [27] Intel First to Ship Microprocessors Using 300 Mm Wafers on 0.13-Micron Technology. 25 February 2002. Online. Intel Press Room. Available: <http://www.intel.com/pressroom/archive/releases/20020225manf.htm>. 24 April 2003.
- [28] Gartland, K. Automated Material Handling System (AMHS) Framework User Requirements Document: Version 1.0. 1999. Online. International SEMATECH, Inc. Available: <http://www.sematech.org/public/docubase/document/3793atr.pdf>. 24 April 2003.
- [29] Davis, T. and Selep, E. "Group Technology for High-Mix Printed Circuit Assembly." 8th IEEE/CHMT International Electronic Manufacturing Technology Symposium, IEMT Conference, 1990. 264-269.
- [30] Chang, C.-F. and Chang, S.-K. "A Layer-Based Approach for Semiconductor Fabrication Facilities." IEEE/SEMI Advanced Semiconductor Manufacturing Conference: IEEE, 1998. 385-390.
- [31] Garrett, T. Intel'S 300mm Program Is on-Line. 2001. Online. Electronic Journal (Japanese). Available: <http://www.intel.com/research/silicon/wafers.htm>. 23 September 2003.
- [32] Paprotny, I., Shiau, J.-Y., Huh, Y., and Mackulak, G. T. "Simulation Based Comparison of Semiconductor AMHS Alternatives: Continuous Flow Vs. Overhead Monorail." Proceedings of the 2000 Winter Simulation Conference. Eds. J. A. Joines, R. R. Barton, K. Kang, and P. A. Fishwick. Orlando, FL, 2000. 1333-1338.
- [33] Law, A. M. and Kelton, W. D. Simulation Modelling and Analysis. Second Edition ed: McGraw-Hill Inc., 1991.
- [34] Pidd, M. "An Introduction to Computer Simulation." Proceeding of the 1994 Winter Simulation Conference. Eds. J. D. Tew, S. Manivannan, D. A. Sadowski, and A. F. Seila: IEEE, 1994. 7-14.
- [35] Shannon, R. E. "Introduction to the Art and Science of Simulation." Proceedings of the 1998 Winter Simulation Conference. Eds. D. J. Medeiros, E. F. Watson, J. S. Carson, and M. S. Manivannan. Washington, DC, 1998. 7-14.
- [36] Gansler, J. S. DoD Modeling and Simulation (M&S) Glossary. 1998. Online. DoD Modeling and Simulation (M&S) Management. Available: <https://www.dmsomil/public/library/policy/guidance/p500059m.pdf>. 4 May 2003.

- [37] Maria, A. "Introduction to Modeling and Simulation." Proceedings of the 1997 Winter Simulation Conference. Eds. S. Andradóttir, K. J. Healy, D. H. Withers, and B. L. Nelson. ATLANTA, GEORGIA, 1997. 7-13.
- [38] Seila, A. F. "Introduction to Simulation." Proceedings of the 1995 Winter Simulation Conference. Eds. C. Alexopoulos, K. Kang, W. R. Lilegdon, and D. Goldsman: IEEE, 1995. 7-15.
- [39] Ben-Arieh, D. "Concurrent Modelling Language (CML) for Discrete Process Modeling, Simulation and Control." Journal of Intelligent Manufacturing 3 (1992): 31-41.
- [40] PRIKSTER, A. A. B. "Principles of Simulation Modelling." Handbook of Simulation: Principles, Methodology, Advances, Applications, and Practice. Ed. J. Banks: John Wiley & Sons, Inc., 1998. 31-51.
- [41] Pidd, M. "Five Simple Principles of Modelling." Proceedings of the 1996 Winter Simulation Conference. Eds. J. M. Charnes, D. M. Morrice, D. T. Brunner, and J. J. Swain. Coronado, California: Institute of Electrical and Electronics Engineers, 1996. 721-728.
- [42] Banks, J. "Introduction to Simulation." Proceedings of the 2000 Winter Simulation Conference. Eds. J. A. Joines, R. R. Barton, K. Kang, and P. A. Fishwick. Orlando, FL, 2000. 9-16.
- [43] Miller, S. and Pegden, D. "Introduction to Manufacturing Simulation." Proceedings of the 2000 Winter Simulation Conference. Ed. P. A. Fishwick. Orlando, FL, 2000. 63-66.
- [44] Bodner, D. A. and McGinnis, L. F. "A Structured Approach to Simulation Modelling of Manufacturing Systems." Proceedings of the 2002 Industrial Engineering Research Conference, 2002.
- [45] Schriber, T. J. and Brunner, D. T. "Inside Discrete-Event Simulation Software: How It Works and Why It Matters." Proceedings of the 1999 Winter Simulation Conference. Eds. P. A. Farrington, H. B. Nembhard, D. T. Sturrock, and G. W. Evans. Phoenix, AZ, 1999. 72-80.
- [46] Bennett, B. S. Simulation Fundamentals. Prentice Hall International, 1995.
- [47] Banks, J. "Panel Session: The Future of Simulation." Proceedings of the 2001 Winter Simulation Conference. Eds. B. A. Peters, J. S. Smith, D. J. Medeiros, and M. W. Rohrer. Arlington, VA, 2001.
- [48] Banks, J., John S. Carson, I., and Nelson, B. L. Discrete-Event System Simulation. Second Edition ed: Prentice-Hall, Inc., 1996.
- [49] Banks, J. "Principles of Simulation." Handbook of Simulation: Principles, Methodology, Advances, Applications, and Practice. Ed. J. Banks: John Wiley & Sons, Inc., 1998. 3-29.

- [50] Chance, F., Robinson, J., and Fowler, J. "Supporting Manufacturing with Simulation: Model Design, Development, and Deployment." Proceedings of the 1996 Winter Simulation Conference. Ed. J. J. Swain. Coronado, California: Institute of Electrical and Electronics Engineers, 1996. 114-121.
- [51] Law, A. M. and McComas, M. G. "Simulation of Manufacturing Systems." Proceedings of the 1999 Winter Simulation Conference. Eds. P. A. Farrington, H. B. Nembhard, D. T. Sturrock, and G. W. Evans. Phoenix, AZ, 1999. 56-59.
- [52] Goldsman, D. and Nelson, B. L. "Comparing Systems Via Simulation." Handbook of Simulation: Principles, Methodology, Advances, Applications, and Practice. Ed. J. Banks: John Wiley & Sons, Inc., 1998. 273-306.
- [53] Simulation Dynamics Sdi Industry Version 5.0: Reference Manual. 2001. Simulation Dynamics, Inc.
- [54] Valentin, E. C. and Verbraeck, A. "Guidelines for Designing Simulation Building Blocks." Proceedings of the 2002 Winter Simulation Conference. Ed. J. M. Charnes. San Diego, CA, 2002. 563-571.
- [55] Brooks, R. J. and Tobias, A. M. "Simplification in the Simulation of Manufacturing Systems." International Journal of Production Research 38.5 (2000): 1009-1027.
- [56] Chwif, L., Barretto, M. R. P., and Paul, R. J. "On Simulation Model Complexity." Proceedings of the 2000 Winter Simulation Conference. Ed. R. R. B. J. A. Joines, K. Kang, and P. A. Fishwick. Orlando, FL, 2000. 449-455.
- [57] Benjamin, P., Erraguntla, M., Delen, D., and Mayer, R. "Simulation Modeling at Multiple Levels of Abstraction." Proceedings of the 1998 Winter Simulation Conference. Ed. M. S. Manivannan. Washington, DC, 1998. 391-398.
- [58] Pillai, D. Integrated Factory Designs Using Modeling and Simulation. 1999. Online. Intel Research. Available: <ftp://download.intel.com/research/silicon/integratedfactorydesignsusing.pdf>. 30 March 2003.
- [59] Mackulak, G. T., Lawrence, F. P., and Colvin, T. "Effective Simulation Model Reuse: A Case Study for AMHS Modeling." Proceedings of the 1998 Winter Simulation Conference. Eds. D. J. Medeiros, E. F. Watson, J. S. Carson, and M. S. Manivannan. Washington, DC, 1998. 979-984.
- [60] Paul, R. J. and Taylor, S. J. E. "What Use Is Model Reuse: Is There a Crook at the End of the Rainbow?" Proceedings of the 2002 Winter Simulation Conference. Ed. J. M. Charnes. San Diego, CA, 2002. 648-652.
- [61] Pidd, M. "Simulation Software and Model Reuse: A Polemic." Proceedings of the 2002 Winter Simulation Conference. Eds. E. Yücesan, C.-H. Chen, J. L. Snowdon, and J. M. Charnes. San Diego, CA, 2002. 772-775.

- [62] Rioux, G. P. and Nance, R. E. "Generalizing: Is It Possible to Create All-Purpose Simulations?" Proceedings of the 2002 Winter Simulation Conference. Ed. J. M. Charnes. San Diego, CA, 2002. 783-790.
- [63] Chen, G. and Szymanski, B. K. "COST: A Component-Oriented Discrete Event Simulator." Proceedings of the 2002 Winter Simulation Conference. Eds. E. Yücesan, C.-H. Chen, J. L. Snowdon, and J. M. Charnes. San Diego, CA, 2002. 776-782.
- [64] Kachitvichyanukul, V., Henriksen, J. O., Pegden, C. D., Ingalls, R. G., and Schmeiser, B. W. "Simulation Environment for the New Millennium (Panel)." Proceedings of the 2001 Winter Simulation Conference. Ed. M. W. Rohrer. Arlington, VA, 2001. 541-547.
- [65] Kantner, M. "Hierarchical Modeling and Multiresolution Simulation." Proceedings of the 1999 Winter Simulation Conference. Ed. G. W. Evans. Phoenix, AZ, 1999. 548-552.
- [66] Heim, J. A. "Integrating Distributed Simulation Objects." Proceedings of the 1997 Winter Simulation Conference. Eds. S. Andradóttir, K. J. Healy, D. H. Withers, and B. L. Nelson. ATLANTA, GEORGIA, 1997. 532-538.
- [67] Randell, L. G., Holst, L. G., and Bolmsjö, G. S. "Incremental System Development of Large Discrete-Event Simulation Models." Proceedings of the 1999 Winter Simulation Conference. Ed. G. W. Evans. Phoenix, AZ, 1999. 561-568.
- [68] Pidd, M. and Castro, R. B. "Hierarchical Modular Modelling in Discrete Simulation." Proceedings of the 1998 Winter Simulation Conference. Eds. D. J. Medeiros, E. F. Watson, J. S. Carson, and M. S. Manivannan. Washington, DC, 1998. 383-389.
- [69] Odhabi, H. I., Paul, R. J., and Macredie, R. D. "The Four Phase Method for Modelling Complex Systems." Proceedings of the 1997 Winter Simulation Conference. Ed. B. L. Nelson. ATLANTA, GEORGIA, 1997. 510-517.
- [70] Eldabi, T. and Paul, R. J. "Flexible Modeling of Manufacturing Systems with Variable Levels of Detail." Proceedings of the 1997 Winter Simulation Conference. Eds. S. Andradóttir, K. J. Healy, D. H. Withers, and B. L. Nelson. ATLANTA, GEORGIA, 1997. 801-808.
- [71] Banks, J. "The Future of Simulation Software: A Panel Discussion." Proceedings of the 1997 Winter Simulation Conference. Ed. B. L. Nelson. ATLANTA, GEORGIA, 1997. 166-173.
- [72] Norman, M., Tinsley, D., Barksdale, J., Wiersholm, O., Campbell, P., and MacNair, E. "Process and Material Handling Models Integration." Proceedings of the 1999 Winter Simulation Conference. Eds. P. A. Farrington, H. B. Nembhard, D. T. Sturrock, and G. W. Evans. Phoenix, AZ, 1999. 1262-1267.

- [73] Campbell, P. L., Rohan, D., and MacNair, E. A. "A Model of a 300mm Wafer Fabrication Line." Proceedings of the 1999 Winter Simulation Conference. Ed. G. W. Evans. Phoenix, AZ, 1999. 909-911.
- [74] DeJong, C. D. and Fischbein, S. A. "Integrating Dynamic Fab Capacity and Automation Models for 300mm Semiconductor Manufacturing." Proceedings of the 2000 Winter Simulation Conference. Eds. J. A. Joines, R. R. Barton, K. Kang, and P. A. Fishwick. Orlando, FL, 2000. 1505-1509.
- [75] Schulz, M., Stanley, T. D., and Renelt, B. "Simulation Based Decision Support for Future 300mm Automated Material Handling." Proceedings of the 2000 Winter Simulation Conference. Ed. R. R. B. J. A. Joines, K. Kang, and P. A. Fishwick. Orlando, FL, 2000. 1518-1522.
- [76] Peikert, A., Thoma, J., and Brown, S. "A Rapid Modeling Technique for Measurable Improvements in Factory Performance." Proceedings of the 1998 Winter Simulation Conference. Eds. D. J. Medeiros, E. F. Watson, J. S. Carson, and M. S. Manivannan. Washington, DC, 1998. 1011-1015.
- [77] Paprotny, I., Zhao, W., and Mackulak, G. T. "Reducing Model Creation Cycle Time by Automated Conversion of a CAD AMHS Layout Design." Proceedings of the 1999 Winter Simulation Conference. Ed. G. W. Evans. Phoenix, AZ, 1999. 779-783.
- [78] Jefferson, T., Rangaswami, M., and Stoner, G. "Simulation in the Design of Ground-Based Intrabay Automation Systems." Proceedings of the 1996 Winter Simulation Conference. Eds. J. M. Charnes, D. M. Morrice, D. T. Brunner, and J. J. Swain. Coronado, California: Institute of Electrical and Electronics Engineers, 1996. 1008-1013.
- [79] Pierce, N. G. and Stafford, R. "Simulating AMHS Performance for Semiconductor Wafer Fabrication." International Symposium on Semiconductor Manufacturing: IEEE, 1995. 165-170.
- [80] Mackulak, G. T. and Savory, P. "A Simulation-Based Experiment for Comparing AMHS Performance in a Semiconductor Fabrication Facility." IEEE Transactions on Semiconductor Manufacturing 14.3 (2001): 273-280.
- [81] Kurosaki, R., Shimura, T., Komada, H., Kojima, T., and Watanabe, Y. "Low Cost and Short Lead Time AMHS Design Using Interbay/Intrabay Diverging and Converging Method for 300mm Fab." The Ninth International Symposium on Semiconductor Manufacturing, 2000. 48-51.
- [82] Kaempf, U. "Automated Wafer Transport in the Wafer Fab." IEEE/SEMI Advanced Semiconductor Manufacturing Conference: IEEE, 1997. 356-361.
- [83] Nadoli, G. and Rangaswami, M. "An Integrated Modelling Methodology for Material Handling Systems Design." Proceedings of the 1993 Winter Simulation Conference. Eds. G. W. Evans, M. Mollaghasemi, E. C. Russel, and W. E. Biles: IEEE, 1993. 785-789.

- [84] Simulation Software Survey. February 2001. Online. OR-MS Today. Available: <http://lionhrtpub.com/orms/surveys/Simulation/Simulation.html>. 11 May 2003.
- [85] Mason, S. J. and Jensen, P. A. "A Comparison Study of the Logic of Four Wafer Fabrication Simulators." Proceedings of the 1996 Winter Simulation Conference. Eds. J. M. Charnes, D. M. Morrice, D. T. Brunner, and J. J. Swain. Coronado, California: Institute of Electrical and Electronics Engineers, 1996. 1031-1038.
- [86] Collins, D. W., Collins, L. D. A., and Franklin, B. Simulation Model Analysis of AMHS W/Area Elevators for an 8 Inch Wafer Fab. 2002. Online. ACADZ inc. Available: <http://www.acadz.com/Simulation%20Model%20Analysis%20of%20AMHS.pdf>. 14 November 2003.
- [87] Next-Generation Transport Systems for 300 Mm Plants. 2003. Online. Daifuku Co., Ltd. Available: <http://www.daifuku.co.jp/english/products/cfa/300.html>. 23 September 2003.
- [88] Cpl™ Technology. 2003. Online. JMAR Technologies, Inc. Available: <http://www.nanolitho.com/gallery.html>. 23 September 2003.
- [89] Automatic Handling Systems. 2003. Online. Shinko Technologies, Inc. Available: <http://www.shinkotech.com/handle.htm>. 23 September 2003.
- [90] Pillai, D. D. and Srinivasan, S. R. "Material Handling Automation - Trends, Vision, and Future Plans." Proceedings from the International Symposium on Semiconductor Manufacturing '96. Tokyo, Japan, 1996. 1-4.
- [91] Sohn, B., Pillai, D., and Acker, N. "300mm Factory Design for Operational Effectiveness." Proceedings of European IEEE/SEMI Semiconductor Manufacturing Conference, 2000.
- [92] International Technology Roadmap for Semiconductors 2002 Update. 2002. Online. Available: <http://www.itrs.net/ITRS/Files/2002Update/Home.pdf>. 14 June 2003.
- [93] Banks, J. "Software for Simulation." Handbook of Simulation: Principles, Methodology, Advances, Applications, and Practice. Ed. J. Banks: John Wiley & Sons, Inc., 1998. 813-835.
- [94] Baldwin, L. P., Eldabi, T., Hlupic, V., and Irani, Z. "Enhancing Simulation Software for Use in Manufacturing." Logistics Information Management 13.5 (2000): 263-270.
- [95] Rivera, J. "Modeling with Extend™." Proceedings of the 1998 Winter Simulation Conference. Eds. D. J. Medeiros, E. F. Watson, J. S. Carson, and M. S. Manivannan. Washington, DC, 1998. 257-262.
- [96] Krahl, D. "The Extend Simulation Environment." Proceedings of the 2002 Winter Simulation Conference. Eds. E. Yücesan, C.-H. Chen, J. L. Snowdon, and J. M. Charnes. San Diego, CA, 2002. 205-213.

- [97] Extend Advantages. 2003. Online. Imagine That Inc. Available: http://www.imaginethatinc.com/sols_advantage.html. 23 September 2003.
- [98] Fowler, J. W., Fu, M. C., Schruben, L. W., Brown, S., Chance, F., Cunningham, S., Hilton, C., Janakiram, M., Stafford, R., and Hutchby, J. "Operational Modeling & Simulation in Semiconductor Manufacturing." Proceedings of the 1998 Winter Simulation Conference. Ed. M. S. Manivannan. Washington, DC, 1998. 383-389.
- [99] ExtendTM Version 5 User'S Guide. 2000. Imagine That, Inc. Available: <http://www.imaginethatinc.com>.
- [100] Biller, B. and Nelson, B. L. "Answers to the Top Ten Input Modeling Questions." Proceedings of the 2002 Winter Simulation Conference. Eds. E. Yücesan, C.-H. Chen, J. L. Snowdon, and J. M. Charnes. San Diego, CA, 2002. 35-40.

APPENDICES

Appendix A: Publications Arising from this Work.

Appendix B: Group Technology.

Appendix C: Artificial Intelligence Techniques.

Appendix D: Initial AMHS Simulation Model.

Appendix E: The IDEF0 Function Modelling Method.

Appendix A

A PUBLICATIONS ARISING FROM THIS WORK

- [A.1] El-Kilany, K. S., Young, P., and El-Baradie, M. A. "Intelligent Planning and Control of Integrated Manufacturing Systems: An Overview." 11th International Conference on Flexible Automation and Intelligent Manufacturing, FAIM 2001. Ed. T. Szécsi. Dublin, Ireland: Dublin City University, 2001. 229-240. Vol. 1. 2 vols.
- [A.2] El-Kilany, K. S., Young, P., and El-Baradie, M. A. "A Fuzzy Approach to the Group Technology Problem." The Sixth International Conference on Production Engineering and design for Development (PEDD6). Cairo, Egypt: Ain-Shams University, 2002. 201-212. Vol. 1. 2 vols.
- [A.3] El-Kilany, K. S., Young, P., and El-Baradie, M. A. "Comparing Different Types of Semiconductor Material Handling Systems Using Simulation." 32nd International Conference on Computers and Industrial Engineering. Ed. C. Heavey. Limerick, Ireland, 2003. 647-652. Vol. 2. 2 vols.
- [A.4] El-Kilany, K. S., Young, P., and El-Baradie, M. A. "Generic Tool for Modelling and Simulation of Semiconductor Intrabay Material Handling System." International Conference on Advances in Materials and Processing Technology. Ed. S. J. Hashmi. Dublin, Ireland: Dublin City University, 2003. 547-550. Vol. 1. 2 vols.
- [A.5] El-Kilany, K. S., Young, P., and El-Baradie, M. A. "Generic Tool for Modelling and Simulation of Semiconductor Intrabay Material Handling System." Submitted to the Journal of Materials Processing Technology (2003).

Journal paper [A.5] is an extended version of the conference paper [A.4].

B GROUP TECHNOLOGY

There is a growing need to make batch manufacturing more efficient and productive. Also, there is an increasing trend to achieve a higher level of integration of design and manufacturing functions in a firm. One of the approaches that are directed towards both of these objectives is group technology [B.1].

The term group technology (GT) was first used by Professor Mitrovanof of Leningrad University as the title for his research into the relationship between component-shape and processing methods. Among other findings, he showed that lathes could be setup to make a number of different but similar parts, one after the other at the same setup. It was then found that by adding other machines, groups could be formed which completed parts [B.2].

Group technology evolved in this way into a form of product organization, which can be used for component processing in any factory. Thus, GT is a form of product organization in which there is a total division of machines into groups, which complete all the components they make. With GT groups, different parts may use different combinations of the machines in a group, in different sequences.

B.1 DEFINING GROUP TECHNOLOGY

Gallagher and Knight [B.3] defined group technology (GT) as a manufacturing philosophy or principle with far reaching implications. The basic concept is relatively simple: identify and bring together related or similar parts and processes, to take advantage of the similarities that exist during all stages of design and manufacture. Similarly, Snead [B.4] stated that GT aims at classifying of parts and/or the grouping of machines in order to increase the efficiency of part design and production.

In context of batch production and flexible manufacturing systems, Durie and Crocker [B.5] defined GT from a production-specific perspective as the replacing of traditional

jobbing shop manufacture by the analysis and grouping of work into families and the formation of groups of machines to manufacture these families on a flow-line principle with the object of minimizing setting times and throughput times.

Another definition of GT, which is similar to Durie and Crocker's, is [B.6]: group technology is a technique, which allows the production of components required in small batches to achieve similar economic advantages to those associated with continuous flow-line production. The approach is to analyse a company's total product and to identify within this product, families of components that are related by a similarity in shape and/or the type of production facilities required to manufacture them.

Group technology can also be defined as a theory of management [B.7]. This is a broader definition, which is based on the simple principle that similar *things* should be done similarly. For manufacturers, these *things* could be products, processes, machines...etc.

From these definitions it is clear that identifying similar parts and grouping them together is the one of the main issues in GT.

B.1.1 Part Families

A part family is a group of parts having some similar attributes. Part similarities are of two types [B.1]:

Design attributes, such as geometric shape and size.

Manufacturing attributes, which can be the sequence and type of processing steps required for making the part.

When forming a part family, both design and manufacturing attributes must be considered. Figure B.1 and Figure B.2 indicate that the consideration of only one set of attributes, either design or manufacturing, may be misleading.

The two parts shown in Figure B.1 are similar from a design viewpoint but quite different in terms of manufacturing. Furthermore, the thirteen parts shown in Figure B.2 have similar manufacturing requirements, but their geometry characteristics do not permit them to be grouped as a design parts family.

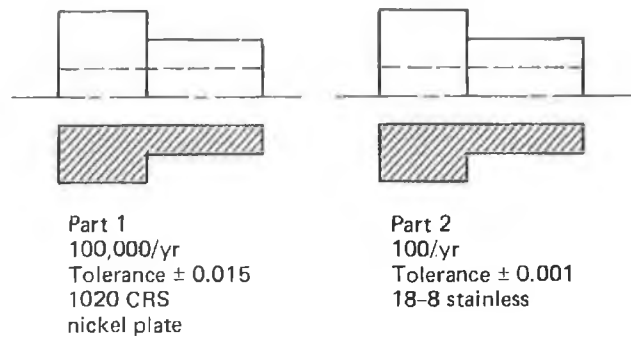


Figure B.1: Two parts of identical shape and size but different manufacturing requirements [B.1].

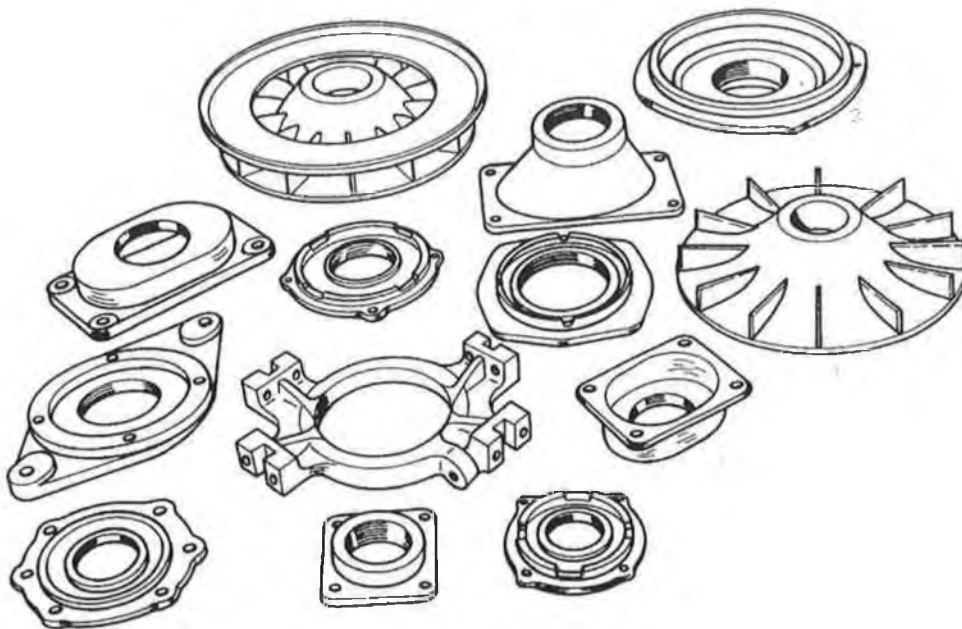


Figure B.2: Thirteen parts with similar manufacturing process requirements but different design attributes [B.1].

B.1.2 Machine Groups

Machine groups are groups of machines, arranged in a cell, to produce similar families of parts. Identifying the group of machines is the second issue addressed by GT.

B.1.3 Cellular Manufacturing

Cellular manufacturing can be defined as an application of GT that involves grouping machines or process based on parts or part families they process. CM entails dividing

the manufacturing facility into small groups or cells of machines, each cell being dedicated to a specified set of part types [B.8].

The third issue of GT is that part families and machine cells should be identified in such a way that parts with similar size, design features, functions and processing requirements are produced *completely* in a cell sharing common resources, e.g. machines, tools and labour.

Before proceeding with further discussion of GT, it is crucial at this point to show the relationship between GT and CIM. The role of GT in integrating and automating various parts of CIM should be clearly understood.

B.1.4 Role of Group Technology in CIM

Kamrani, *et al* [B.9] recognised that GT is an essential element of the foundation for the successful development and implementation of computer integrated manufacturing, through the application of the part family formation and its analysis. GT acts as an information and knowledge system organizer within manufacturing environments. A part classification system, which is an essential part of the grouping task, can be used as the means of describing parts and can be readily integrated with the database, making it a key factor in the integrated CAD and CAPP as shown in Figure B.3.

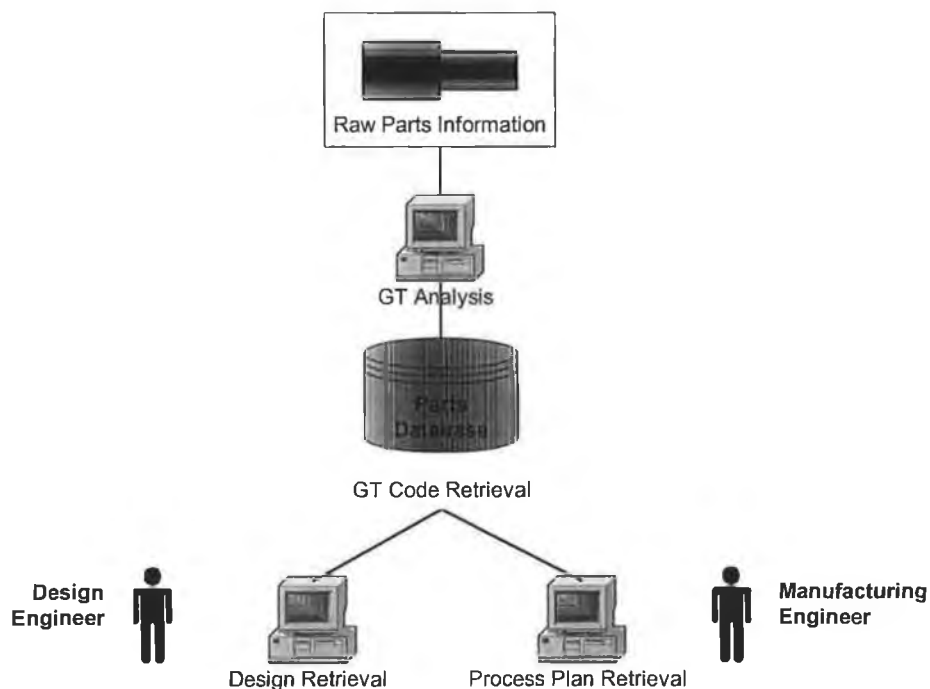


Figure B.3: Importance of parts classification and coding for design and manufacture.

Furthermore, they stress that all too often GT is misunderstood as an another term for classification and coding technique to group parts into families to support the creation of dedicated manufacturing cells. The group technology systems are much more extensive and sophisticated in their capabilities. A major part of manufacturing is a series of data-processing operations. All manufacturing tasks require creating, storing, transmitting, analyzing and modifying data. There is a need to automate and integrate these data manipulating functions and the total solution could be achieved through GT applications.

B.2 APPROACHES TO THE GROUP TECHNOLOGY PROBLEM

B.2.1 General Methods of Solving the GT Problem

There are three general methods for solving the GT problem. All three methods are time consuming, and involve the analysis of much data [B.1].

Visual Inspection

The visual inspection method is the least sophisticated and least expensive method. It involves the classification of parts into families by looking at either the physical parts or their photographs and arranging them into similar groupings. This method is considered to be the least accurate of the three [B.1], it leads to very inconsistent results because seldom will two people group a set of parts into the same families [B.10].

Classification and Coding

This is the most time consuming and complicated of the three methods. Classification of parts is the process of categorizing parts into groups, sometimes called families, according to a set of rules or principles. The objectives are to group together similar parts and to differentiate among dissimilar parts. Coding of a part is the process of assigning symbols to the part. These symbols should have meanings that reflect the attributes of the part, thereby facilitating analysis [B.10].

Parts classification systems fall into one of three categories:

1. Systems based on part design attributes.

2. Systems based on part manufacturing attributes.
3. System based on both design and manufacturing attributes.

The types of design and manufacturing attributes that are typically included in classification schemes are listed in Table B-1.

Table B-1: Design and manufacturing attributes typically included in a group technology classification system [B.1].

<i>Part Design Attributes</i>	
Basic external shape	Major dimensions
Basic internal shape	Minor dimensions
Length/diameter ratio	Tolerances
Material type	Surface finish
Part function	
<i>Part Manufacturing Attributes</i>	
Major process	Operation sequence
Minor operations	Production time
Major dimension	Batch size
Length/diameter ratio	Annual production
Surface finish	Fixtures needed
Machine tool	Cutting tools

Although well over 100 classification and coding systems have been developed for group technology applications, all of them can be grouped into three basic types: hierarchical, attribute, and hybrid.

Hierarchical Code

In this type of code, the meaning of each character is dependent on the meaning of the previous character. A hierarchical code provides a large amount of information in a relatively small number of digits. Defining the meanings for each digit in a hierarchical system can be difficult, although application of the defined system is relatively simple.

Attribute Code

An attribute code is also called a chain code, or a fixed-digit code. The meaning of each character in an attribute code is independent of any other character; thus, each attribute of a part can be assigned a specific position in an attribute code.

Hybrid Code

In reality, most coding systems use a hybrid (mixed) code so that the advantages of each type of system can be utilized. The first digit, for example, might be used to denote the type of part. The next five positions might be reserved for a short attribute code that would describe the attributes of that part.

Production Flow Analysis

The production flow analysis (PFA) method is a production-based approach. The purpose of this approach is to group parts that share common processing requirements. It utilises a machine-part incidence matrices consisting of binary (0,1) identifier used to record the manufacturing sequence and processing for part and machine components. The matrix is rearranged to identify machine-part families. The process of rearranging the machine-part incidence matrix is difficult and sometimes subjective [B.11].

Figure B.4 shows the part-machine incidence matrix before and after rearrangement into part families and machine groups.

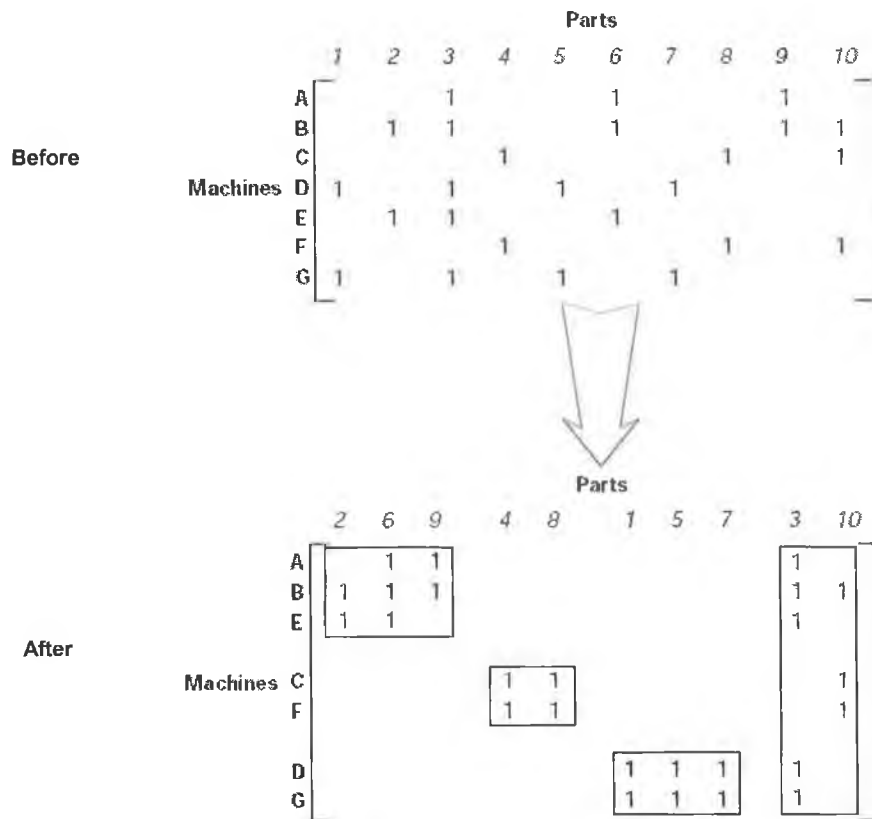


Figure B.4: Part-machine incidence matrix before and after rearrangement.

B.2.2 Intelligent Approaches to Solving the Group Technology Problem

Due to the inherent problems related to the conventional methods of solving the GT problem, researches were forced to search for other ways and approaches to solving this problem, or at least overcoming the weaknesses of the conventional methods. Artificial intelligence techniques are an efficient way to solve the various problems of GT and an effective way for overcoming these weaknesses. In that connection, a fuzzy model has been developed to approach the group technology problem in a new way. The model description and outcome are reported in the paper, [B.12], included below.

B.3 REFERENCES

- [B.1] Groover, M. P. Automation, Production Systems, and Computer Integrated Manufacturing. Prentice-Hall International, Inc., 1987.
- [B.2] Burbidge, J. L. Period Batch Control. Oxford University Press, Inc., 1996.
- [B.3] Gallagher, C. C. and Knight, W. A. Group Technology Production Methods in Manufacture. Ellis Horwood Limited, 1986.
- [B.4] Snead, C. S. Group Technology: Foundation for Competitive Manufacturing. Van Nostrand Reinhold, 1989.
- [B.5] Durie, F. R. E. and Crocker, A. C. "Group Technology." Group Technology. British Productivity Council, 1970. 1-13.
- [B.6] Bennett, D. "A Systematic Approach to Group Technology." The Principles and Practice of Group Technology. The Group Technology Centre, 1972.
- [B.7] Askin, R. G. and Standridge, C. R. Modeling and Analysis of Manufacturing Systems. John Wiley & Sons, Inc., 1993.
- [B.8] HERAGU, S. S. "Group Technology and Cellular Manufacturing." IEEE Transactions on Systems, Man, and Cybernetics 24.2 (1994): 203-215.
- [B.9] Kamrani, A. K., Sferro, P., and Handelman, J. "Critical Issues in Design and Evaluation of Computer Aided Process Planning Systems." Computers & Industrial Engineering 29.1-4 (1995): 619-623.
- [B.10] Bedworth, D. D., Henderson, M. R., and Wolfe, P. M. Computer-Integrated Design and Manufacturing. McGraw-Hill, Inc., 1991.
- [B.11] Ang, D. S. "Identifications of Part Families and Bottleneck Parts in Cellular Manufacturing." Industrial Management & Data Systems 98.1 (1998): 3-7.

- [B.12] El-Kilany, K. S., Young, P., and El-Baradie, M. A. "A Fuzzy Approach to the Group Technology Problem." The Sixth International Conference on Production Engineering and design for Development (PEDD6). Cairo, Egypt: Ain-Shams University, 2002. 201-212. Vol. 1. 2 vols.

A FUZZY APPROACH TO THE GROUP TECHNOLOGY PROBLEM

K. S. El-Kilany, P. Young, and M. A. El Baradie

School of Mechanical and Manufacturing Engineering,
Dublin City University,
Glasnevin, Dublin 9,
Ireland.

ABSTRACT

This paper presents a new approach to the group technology problem based on fuzzy logic. The approach uses fuzzy logic to quantify various design and manufacturing attributes of parts, as well as to quantify machine cell attributes. Both parts and machine cells (MC) attributes are related together through fuzzy rules, and a fuzzy output representing the strength of the relationship between the part and MC is obtained. Thus, part families are identified from the existing machine cells; furthermore, the families are allocated to the cells in a single step. The approach reduces time of forming the part families and the time for assigning the families to their most appropriate machine cells. Moreover, the approach presents other advantages in relation to handling the fuzziness of the group technology problem.

KEYWORDS: Group technology, cellular manufacturing, fuzzy logic, part families formation.

1 INTRODUCTION

In the current business environment, competitiveness of manufacturing companies is determined by their ability to respond quickly to the changing environment and to produce high-quality products at lower costs. Manufacturing companies are striving to achieve these capabilities through automation and innovative concepts.

Application of artificial intelligence (AI) and computer integrated manufacturing (CIM), as one of the automation concepts, has become a necessity to achieve high automation within a manufacturing facility. On the other hand, one of the innovative concepts that is recognised by the manufacturing industries as a major driver to achieve world-class capabilities, is the concept of group technology (GT) and cellular manufacturing (CM). Many large and medium-size manufacturing firms have adopted GT/CM concepts, and experienced reduction in manufacturing lead time, material handling cost, and improvement in quality among other benefits.

Group technology is a manufacturing philosophy or principle with far reaching implications. The basic concept is relatively simple: identify and bring together related or similar parts and processes, to take advantage of the similarities that exist during all stages of design and manufacture [1]. GT aims at classifying parts and/or grouping machines in order to increase the efficiency of part design and production [2]. GT could also be thought of as a theory of management that is based on the simple principle that similar things should be done similarly. One tenet of GT entails dividing the manufacturing facility into small groups or cells of machines, each cell being dedicated to a specified set of part types. The term cellular manufacturing is often used in this regard [3]. Thus, CM can be defined as an application of GT that involves grouping machines or process based on parts or part families they process [4]. Benefits of applying GT are numerous, such as reduced work-in-process inventory, reduced set-up times, higher productivity, etc.

The main issue in GT is how to identify part families and machine cells. A part family is a group of parts having some similar attributes, while a machine cell is a group of machines arranged to produce similar families of parts [5]. Part families and machine cells should be identified in such a way that parts with similar size, design features, functions and processing requirements are produced completely in a cell sharing common resources, e.g. machines, tools and labor [6]. This problem is quite complex and challenging with a number of researchers developing techniques to solve it. The various classifications of these GT techniques have already been reported [4].

Contemporary GT methods apply either coding and classification systems (CCS) or production flow analysis (PFA). In CCS, a coding system assigns codes to parts according to part features and production requirements. Based on these codes, parts are grouped into part families. The disadvantage of the system is that a significant effort is required to code parts. In addition, parts of similar size, shape, and function may not necessarily use the same set of machine tools and other resources [7]. The PFA method is a production-based approach. The purpose of this approach is to group parts that share common processing requirements. It utilizes machine-part incidence matrices consisting of binary (0,1) identifier used to record the manufacturing sequence and processing for part and machine components. The matrix is rearranged to identify machine-part families. The process of rearranging the machine-part incidence matrix is difficult and sometimes subjective [8].

Venugopal [6], reviewed several approaches to the GT problem and concluded that one of the common drawbacks of these contemporary methods, beside those mentioned before, is that the introduction of new parts (or machines) in the system required the algorithm to be rerun with the whole set of data. In addition, these procedures have assumed that a given part (or machine) can be a member of only one part family (or machine cell). Furthermore, he presented AI tools, e.g. neural networks and fuzzy models, as a set of new tools to overcome the above limitations along with other advantages. Finally, he clearly indicated the trend and scope for further research with respect to applications of fuzzy models to the part-machine grouping problem.

The objective of this paper is to present a new approach to the GT problem based on fuzzy logic. The paper is outlined as follows: section 2 presents briefly the fuzzy set theory and its application to the GT problem, section 3 explains the proposed approach, section 4 details the fuzzy model development, finally, section 5 concludes the paper.

2 RELATED WORK AND BACKGROUND

2.1 Fuzzy Set Theory

The basic idea in the fuzzy set theory is the notion of a fuzzy set, which is the mathematical concept of a set that allows partial membership for elements in its domain. Thus, an element (x) may partially belong to a fuzzy set (A). Each fuzzy set is characterized by its membership function $f_A(x)$

that states, for any given (x) in the domain (A), “how much of (x) belongs to (A)” [9]. Thus, a membership function is the relationship between the values of an element and its degree of membership in a set. A fuzzy model, which refers to fuzzy sets and fuzzy analysis, is suitable for dealing with impreciseness that may exist in the parameters of any system.

As mentioned before, GT is used to group parts (or machines) into parts families (or machine cells). The similarity in a part family is the basis for grouping. However, the similarity of properties may not be precise. Therefore, fuzzy models can be used, in conjunction with GT, for defining such similarities [10].

To illustrate this imprecision in parts grouping, suppose that the cylindrical parts shown in figure 1 are grouped based on a certain attribute, dimensions for example, using the OPITZ CCS coding system [11]. The dimensions attribute is expressed by the first digit in the supplementary code. The parts will be coded, based on its diameter (D) in millimeters, as follows: $D \leq 20$ (code value 0), $20 < D \leq 50$ (code value 1), $50 < D \leq 100$ (code value 2), and so on. Such a coding procedure raises a very important question, which is why parts A and B having very different diameters, 26 mm and 50 mm, would be given the same code value (1)? On the other hand, parts B and C having similar diameters, 50 mm and 52 mm, would be given different code values (1 and 2 respectively) and consequently grouped into two different part families. This problem is clearly due to the crisp boundaries that are set between the different code values. Such a problem could be overcome by using the FST in the part family formation problem.

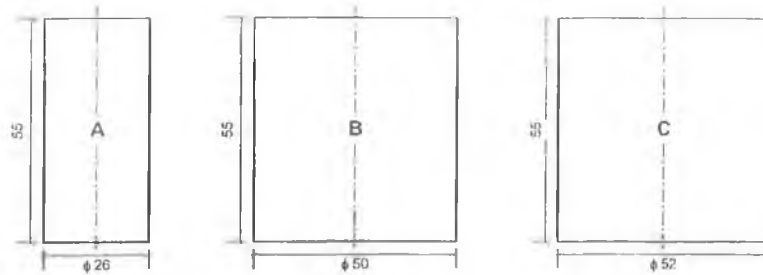


Figure 1. Parts with different diameters.

Likewise, the problem arises in PFA, as these procedures assume that a part feature is developed on a single machine. However, alternate machines can be suitable for developing the same part feature. This is dependent on the uncertainties in the machine attributes, such as tolerances and surface finishes that a given machine could produce. In other words, uncertainty arises in the routing information when alternative routes are possible for processing the components. Under this circumstance also, the fuzzy set can provide a more realistic environment for decision-making.

Many researchers have sensed the need for fuzzy models in order to account for the fuzziness inherent in the GT problem, and approached the problem through the fuzzy point of view.

2.2 Fuzzy Models for the GT Problem

Several fuzzy approaches have been contributed by researchers. The first of these approaches is fuzzy mathematical programming, which is a fuzzified version of conventional methods as presented by Xu and Wang [12]. To incorporate the uncertainty, which is inherent in the measurement of similarities between parts fuzzy mathematics are employed. They present fuzzy cluster analysis and fuzzy pattern recognition as useful tools in GT implementations. These tools generate results that are more realistic as similarity is measured based on common sense. Furthermore, the selection of features is based on characteristics of the machining process, which

facilitates manufacturing-oriented applications. Another fuzzified version of conventional methods is the fuzzy clustering approach [13-14].

The second approach is developing part codes using fuzzy sets. Liao [15], identifies CCS as a major approach to solve the part family formation problem in GT applications. He also affirms that traditionally, all features in a classification and coding system are considered crisp, whilst it is more realistic that some features should be fuzzy. FST has been used to define fuzzy features, such as length-to-diameter ratio and surface roughness, and to code parts using various coding structures under a fuzzy environment. In addition, several measures that could be used to compute the similarity between the fuzzy features have been identified.

Ben-Arieh, Lee, and Chang [16], agree that the main disadvantage of current coding systems is that it has been done in a very crisp and deterministic manner, which lack informative representation of the parts. They present a methodology for coding parts using fuzzy codes, which aims at developing a coding scheme that captures the degree of agreement of the attribute values with the code digits. This is achieved using fuzzy code numbers, which represent the degree of membership of an attribute value in a code. Furthermore, three types of part attributes have been identified; attributes that have a crisp value (e.g., length and aspect ratio), an interval value (e.g. tolerance and surface roughness) or a fuzzy value (e.g. primary shape).

The third approach is quantifying parts features using fuzzy logic. Ben-Arieh and Triantaphyllou [17], stress that no matter how the coding system is designed, there are always similar parts that are forced to have different codes. A methodology for quantifying part features for grouping has been presented. They consider three types of features used for part grouping: quantitative, qualitative, and quantitative with subjective meaning. The analytical hierarchy process (AHP) has been used to quantify the two fuzzy features identified (qualitative, and quantitative with subjective meaning). The methodology compares the discrete values of the fuzzy features to each other. The main disadvantage of their methodology arises from the need to estimate $K(K - 1)/2$ pairwise comparisons, where K is the number of qualitative descriptors.

Gill and Bector [18], state that part features are too ill defined and too complex to analyze using traditional discrete methods. However, these features could be best described using phrases in natural language. Thus, an approach based on fuzzy linguistics to quantify data for part family formation problem has been suggested. Their approach differs from Ben-Arieh and Triantaphyllou in the sense that it uses linguistic terms to describe parts features and the weights, and uses natural language computation technique to aggregate the features to form composite attribute fuzzy sets for each part. Furthermore, Narayanaswamy, Bector, and Rajamani [19], present methods to quantify the fuzziness in part features as well as machining process. This has been achieved through the of fuzzy logic concepts to the machine-part incidence matrix.

The final approach is fuzzified versions of other AI techniques, e.g. fuzzy expert systems and fuzzy neural networks. These include the application of a fuzzy decision making system in part-machine grouping [20] and using a fuzzy neural network for grouping parts based on images captured using a vision sensor [21].

This brief review clearly indicates the weakness of traditional approaches to the GT problem and indicates the trend toward the application and development of fuzzy models.

3 THE PROPOSED APPROACH

The approach uses fuzzy logic to quantify (fuzzify) various design and manufacturing attributes of parts, as well as to quantify machine cell attributes. Both fuzzified parts and machine cell (MC)

attributes, which are considered the inputs to the fuzzy model, are related together through fuzzy rules. Using a set of a given part features and a set of a given machine cell features a fuzzy output, representing the strength of the relationship between the part and MC, is obtained. This output indicates if it is suitable to machine the part within the machine cell or not, and thus will be termed the suitability index.

This process is repeated for the same part and the remaining machine cells. Whenever the suitability index reaches a certain value, predetermined by the user, the part is allocated to the machine cell. Alternatively, the machine cell with the highest value of suitability index is identified, and the part is allocated to it. Afterwards, another part is selected and the same process is repeated for it, and the process continues for the rest of the parts. After all the parts have been allocated to a certain machine cell, the final output would be families of parts that are readily assigned to their most suitable machine cell.

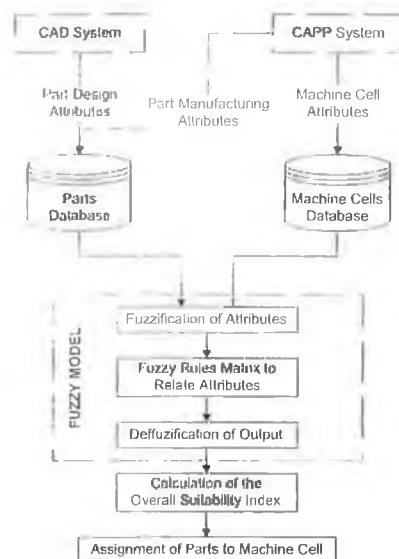


Figure 2. The proposed approach.

Fig. 2 illustrates the system's operation. It is clear from the figure that databases for machine cells and each part within a manufacturing facility are required. The data needed could be extracted directly from the CAD and CAPP systems (assuming that it is integrated within the CIMS of the company), or it could be simply built by the user.

The databases should include all the attributes needed to define both the cells and parts. As mentioned before, these attributes are considered as the inputs to the fuzzy model. Table 1 presents the design and manufacturing attributes that should be defined for parts, while Table 2 presents the cell attributes that are to be defined and its relation to the part attributes.

Table 1. Design and manufacturing attributes of parts.

Design Attributes	Manufacturing Attributes
<ul style="list-style-type: none"> ▪ Basic internal and external shape. ▪ Major and minor dimension. ▪ Length/diameter ratio. ▪ Tolerances. ▪ Material type. ▪ Surface finish. ▪ Part function. 	<ul style="list-style-type: none"> ▪ Primary processes. ▪ Secondary and finishing processes. ▪ Tolerances. ▪ Surface finish. ▪ Sequence of operations. ▪ Tools, dies, and fixtures. ▪ Production quantity and production rate.

Table 2. Cell attributes and their relation to part attributes.

Cell Attributes	Relation to Part Attributes
<ul style="list-style-type: none"> ▪ Shape of parts that could be produced. ▪ Maximum and minimum overall size of parts that could be produced. ▪ Accuracy of machines. ▪ Finishing machines. 	<ul style="list-style-type: none"> ▪ Part design attributes.
<ul style="list-style-type: none"> ▪ Machines included within the cell. ▪ Other equipment. 	<ul style="list-style-type: none"> ▪ Primary, secondary, and finishing processes. ▪ Tools, dies, and fixtures.
<ul style="list-style-type: none"> ▪ Machines layout. 	<ul style="list-style-type: none"> ▪ Sequence of operations.
<ul style="list-style-type: none"> ▪ Machines capacity. 	<ul style="list-style-type: none"> ▪ Production quantity and production rate.

After building the databases, the attributes of parts along with the attributes of machine cells are fuzzified, as mentioned before; this is detailed in the next section.

4 FUZZY MODEL DEVELOPMENT

Development of the fuzzy model for the GT problem is carried out as follows:

- Defining input and output variables, and identifying linguistic terms that describe these variables; hence, fuzzification of variables.
- Developing a fuzzy rules matrix to relate inputs to output.
- Defuzzification of output.

These steps are presented in details in this section. Surface finishing of parts is used as an example to illustrate the steps of the model development.

4.1 Defining Inputs

The required surface finish of a part is considered as the first input (IN1), while the MC finishing capability is considered as the second input (IN2). The linguistic terms used to define each input, along with its corresponding value is given in Table 3. It should be noted that these values are guided by the Design For Manufacturing Handbook [22].

Table 3. The fuzzy model inputs, linguistic terms and values.

Required Surface Finish of Part (IN1)					Machine Cell Finishing Capability (IN2)				
Linguistic Term		Corresponding Value in μm .			Linguistic Term		Corresponding Value in μm .		
Abbr	Name	Min	Avg	Max	Abbr	Name	Min	Avg	Max
SU	Super		0.05	0.40	S	Super	0.05	0.20	0.40
VF	Very Fine	0.20	0.80	1.40	O	Ordinary	0.20	0.80	1.60
F	Fine	0.80	1.60	2.40	P	Poor	0.80	3.20	8.00
ST	Standard	1.60	3.20	4.80					
R	Rough	3.20	6.40	8.00					

Figures 3 and 4 illustrate the set shapes of input 1 and input 2, respectively. The thick lines shown in the figures represent crisp values of inputs for a sample part and machine cell. IN1 has a value of 4.0 μm ., which is interpreted into *standard* surface finish with a membership of 0.5 and *rough* surface finish with a membership of 0.25. On the other hand, IN2 has a value of 4.0 μm ., which is interpreted into *poor* finishing capability with a membership of 1.

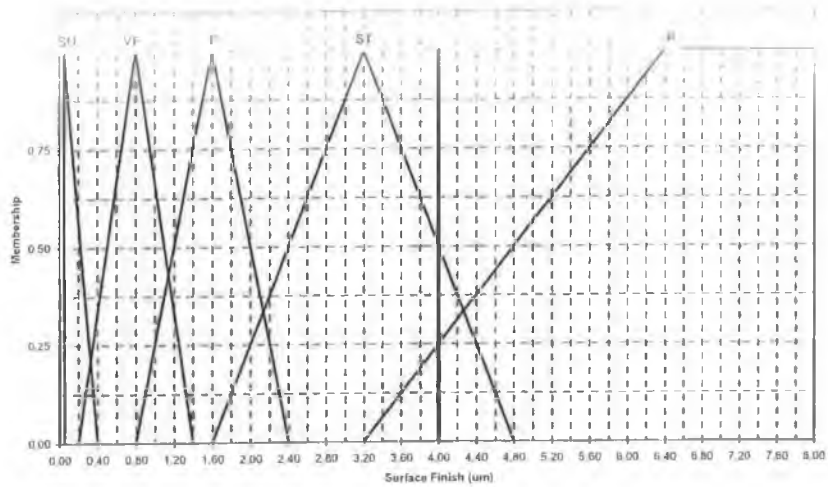


Figure 3. Input 1 set shape.

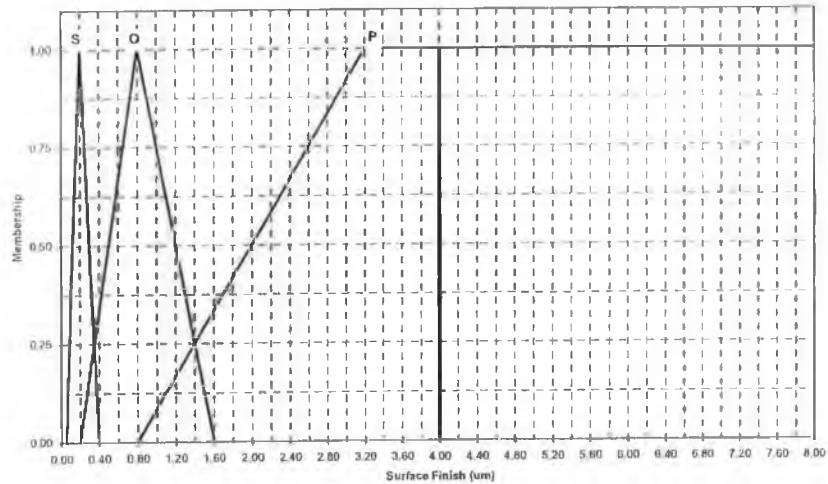


Figure 4. Input 2 set shape.

4.2 Defining Output

The output represents how suitable the MC is to produce the required part based on a certain feature. The output (OUT) is fuzzified, and the linguistic terms used to describe its values are given in Table 4.

Table 4. The fuzzy model output, linguistic terms and values.

Suitability Index (OUT)				
Linguistic Term		Corresponding Value in %		
Abbr.	Name	Min.	Avg.	Max.
VH	Very High	85	100	–
H	High	60	75	90
M	Moderate	35	50	65
L	Low	10	25	40
VL	Very Low	–	0	15

4.3 Fuzzy Rules Matrix

The fuzzy rules matrix is used to combine the two inputs of the model to arrive at the output, which is the suitability index. The matrix is shown in figure 5, and the notations used in the figure are the ones defined in Tables 3 and 4. Using the same input values, mentioned before, IN1 has a linguistic term of either *standard* or *rough*, and IN2 has a linguistic term of *poor*. Combining these two inputs gives the shaded part (upper right) of the matrix. This means that in some situations more than one rule may fire, and the output value is obtained based on all the rules. Hence, rules are working fuzzy as well.

2. IN2					
P	VL	VL	M	VH	VH
O	VL	H	VH	L	VL
S	VH	VH	L	VL	VL
1.IN1	SU	VF	F	ST	R

Figure 5. The fuzzy rules matrix.

4.4 Defuzzification of Output

After the output has been obtained, it is defuzzified into a crisp value as shown in figure 6. The thick vertical line represents the crisp output value. The output in figure 6 is *very high* with membership of 0.54, which is equivalent to a crisp value of 93.

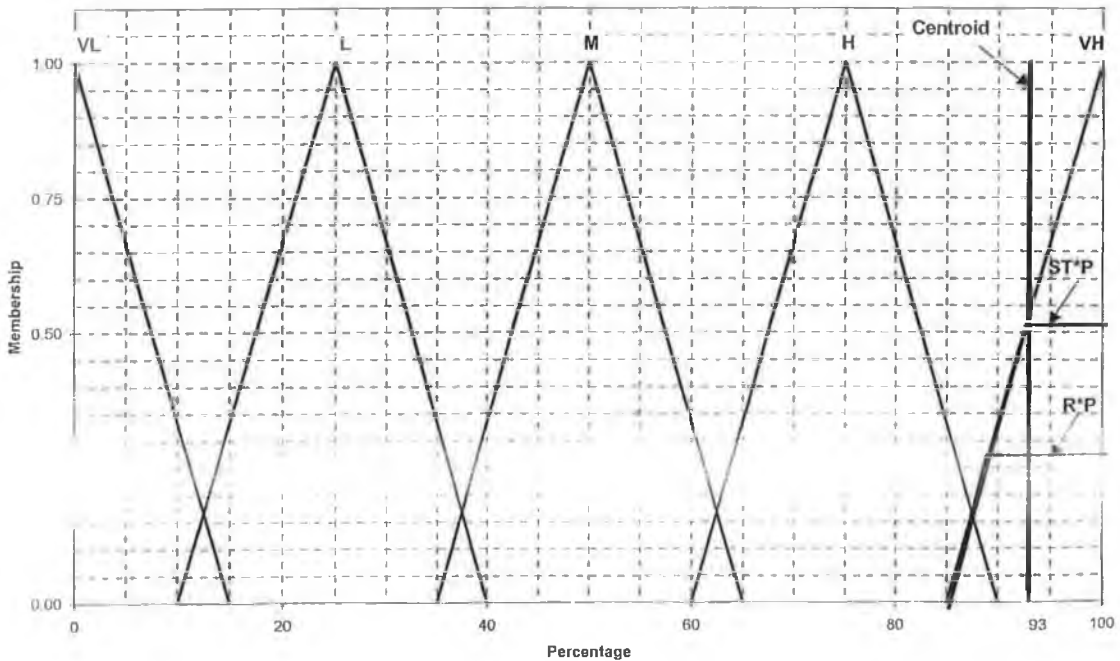


Figure 6. Defuzzification of output.

The need for defuzzifying the output is that the suitability index for each individual feature of a part (F) is used to calculate the overall suitability index (SI). The overall suitability index is calculated by using weighted average, as follows:

$$SI = \frac{W_1 F_1 + W_2 F_2 + \dots W_n F_n}{W_1 + W_2 + \dots W_n}$$

Where,

SI : Overall suitability index (%).

F : Suitability index for an individual feature (%).

W : Weight given for each feature.

n : Number of features.

The weight given for each feature should be flexible enough to be modified, which allows the user to set a value for it based on his/her needs. Finally, based on the value of SI , the part is either allocated to the machine cell or not.

5 CONCLUSIONS

A new approach to the GT part family formation problem has been presented. A review of literature has been conducted and clearly indicated the weakness of traditional GT methods. Moreover, the review showed a strong trend towards the application of fuzzy models to the GT problem. Hence, this is due to the fuzziness inherent in grouping similar parts together, which is the concept upon which GT is based.

The proposed approach is effective in part family formation. Different design and manufacturing attributes of parts have been identified. In addition, machine cells' attributes and their relation to parts' attributes have been presented. The approach quantifies parts' features using linguistic terms, as well as machine cells attributes. The fuzzy model presented uses these quantified features as fuzzy inputs, and by using a fuzzy rules matrix, a fuzzy output is obtained which is the suitability index. The model is simple and rules could be modified easily.

Compared to other approaches, the proposed approach reduces the time needed to form part families and assigning them to their suitable machine cells. This is due to following reasons:

- It performs a smaller number of comparisons, as comparisons are between parts and machine cells; rather than parts and each other.
- Groups parts into families, and assign them to the most suitable machine cell in one step; rather than forming part families and then assigning the families to the machine cells

Finally, the approach is expected to expand its range of applications, by defining more parts and machine cells attributes. Furthermore, it could be applied to handle other problems, for example creating machine cells based on pre-defined part families.

REFERENCES

- [1] GALLAGHER, C.C. and KNIGHT, W.A., "Group Technology Production Methods in Manufacture", John Wiley & Sons, Inc., (1986).
- [2] SNEAD, C.S., "Group Technology: Foundation for Competitive Manufacturing", Van Nostrand Reinhold, (1989).

- [3] ASKIN, R.G., and STANDRIDGE, C.R., "Modeling and Analysis of Manufacturing Systems", John Wiley & Sons, Inc., (1993)
- [4] HERAGU, S.S., "Group Technology and Cellular Manufacturing", IEEE Transactions on Systems, Man, and Cybernetics, Vol. 24, No. 2, pp. 203-215, (1994).
- [5] BEDWORTH, D.D., HENDERSON, M.P., and WOLFE, P.M., "Computer Integrated Design and Manufacturing", McGraw-Hill, Inc., (1991).
- [6] VENUGOPAL, V., "Soft-Computing-Based Approaches to the Group Technology Problem: A State-of-the-art Review", International Journal of Production Research, Vol. 37, No. 14, pp. 3335-3357, (1999).
- [7] CHENG, C., GOH, C., and LEE, A., "A Two-Stage Procedure for Designing a Group Technology System", International Journal of Operations & Production Management, Vol. 15, No. 6, pp. 41-50, (1995).
- [8] Aug, D.S., "Identifications of Part Families and Bottleneck Parts in Cellular Manufacturing", Industrial Management & Data Systems, Vol. 98, No. 1, pp. 3-7, (1998).
- [9] DRAKOPOULOS, J.A., "Probabilities, Possibilities, and Fuzzy Sets", Fuzzy Sets and Systems, Vol. 75, pp. 1-15, (1995).
- [10] DU, T.C., and WOLFE, P.M., "Implementation of Fuzzy Logic Systems and Neural Networks in Industry", Computers in Industry, Vol. 32, pp. 261-272, (1997).
- [11] OPITZ, H., "A Classification System to Describe Workpieces", Pergamon Press, (1970).
- [12] XU, H., and WANG, H., "Part Family Formation for GT Applications Based on Fuzzy Mathematics", International Journal of Production Research, Vol. 27, No. 9, pp. 1637-1651, (1989).
- [13] GINDY, N.N.Z., RATCHEV, T.M., and CASE, K., "Component Grouping for GT Applications—A Fuzzy Clustering Approach with Validity Measure", International Journal of Production Research, Vol. 33, No. 9, pp. 2493-2509, (1995).
- [14] MASNATA, A. and SETTINER L., "An Application of Fuzzy Clustering to Cellular Manufacturing", International Journal of Production Research, Vol. 35 No. 4, p. 1077-1094, (1997).
- [15] LIAO, T.W., "Classification and Coding Approaches to Part Family Formation Under a Fuzzy Environment", Fuzzy Sets and Systems, Vol. 122, pp. 425-441, (2001).
- [16] BEN-ARIEH, D., LEE, S.E., and CHANG, P.T., "Fuzzy Part Coding for Group Technology", European Journal of Operational Research, Vol. 92, pp. 637-648, (1996).
- [17] BEN-ARIEH, D., and TRIANTAPHYLLOU, E., "Quantifying Data for Group Technology with Weighted Fuzzy Features", International Journal of Production Research, Vol. 30, No. 6, pp. 1285-1299, (1992).
- [18] GILL, A. and BECTOR, C.R., "A Fuzzy Linguistic Approach to Data Quantification and Construction of Distance Measures for the Part Family Formation Problem", International Journal of Production Research, Vol. 35, No. 9, pp. 2565-2578, (1997).
- [19] NARYANASWAMY, P., BECTOR, C.R., and RAJAMANI, D., "Fuzzy Logic Concepts Applied to Machine-Component Matrix Formation in Cellular Manufacturing", European Journal of Operational Research, Vol. 93, pp. 88-97, (1996).
- [20] GUNGOR, Z. and ARIKAN, F., "Application of Fuzzy Decision Making in Part-Machine Grouping", International Journal of Production Economics, Vol. 63, pp. 181-193, (2000).
- [21] KUO, R.J., CHI, S.C., and TENG, P.W., "Generalized Part Family Formation Through Fuzzy Self-Organizing Feature Map Neural Network", Computers & Industrial Engineering, Vol. 40, pp. 79-100, (2001).
- [22] BRALLA, J.G., "Design For Manufacturing Handbook", McGraw-Hill Inc., (1999).

C ARTIFICIAL INTELLIGENCE TECHNIQUES

C.1 EXPERT SYSTEMS

An Expert System is a program that attempts to mimic human expertise by applying inference methods to a specific body of knowledge. This body of knowledge is called the domain [C.1]. Expert systems derive their name from the fact that the system contains the expertise required to solve specific domain related problems.

To help humans advance the accuracy of their diagnoses, reduce the time needed to diagnose a situation and thus, decrease the cost of their assessments, expert systems can play a great role [C.2].

C.1.1 Structure of Expert Systems

An expert system typically consists of two main components: the knowledge base and the inference engine. The rules and facts form the knowledge base. These include definitions of the objects and variables related to the application domain and the relationships among them [C.3].

The inference engine, on the other hand, contains the inference strategies used by the expert system. It processes the elements in the knowledge base to solve a particular problem within the application domain.

A user interface program or protocol is used to interact with the user. During such interaction, the user's responses are processed by the inference engine and the system queries through the facts and rules stored in the knowledge base. The structure of an expert system is shown in Figure C.1.

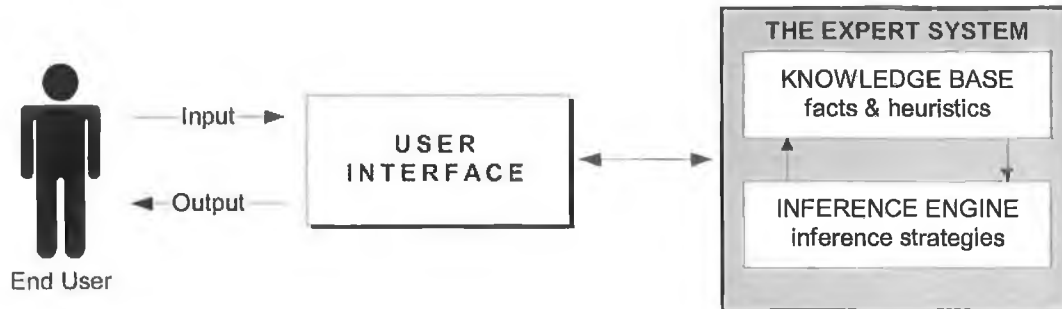


Figure C.1: Structure of an expert system.

C.1.2 Problem Solving in Expert Systems

There are three main activities involved in problem solving within expert systems [C.4]:

1. **Searching.** The purpose of an ES is to develop and recommend a proposed solution or a set of alternatives to a given problem. To accomplish this task, the ES must conduct a search for the solution. The two fundamental search strategies employed by an ES are forward and backward chaining:

Forward chaining proceeds from a premise to some conclusion and is said to be data-driven.

Backward chaining is a goal-driven approach and proceeds from a tentative conclusion to a premise to determine if the data support that conclusion.

2. **Satisfying constraints.** Constraint satisfaction works by process of elimination. It starts with the entire state space and narrows down the original state space to a required subset.
3. **Reasoning.** Reasoning is the process of accumulating information by inference until the solution to the problem is deduced.

C.1.3 Uncertainty in Expert Systems

Since the real world is characterised by uncertainty and fuzziness, various sources of them may be encountered in an expert system. Actually, there are two primary sources of uncertainty [C.5]:

Uncertainty with regard to the validity of the knowledge base content.

Uncertainty with regard to the validity of a user response.

C.1.4 Expert System Learning

Researchers reported that expert system learning techniques were still much specialised. Very few existing programs had real learning abilities [C.6]. Learning is defined as the ability of the expert system (ES) to improve system recommendations based on experience, which can be achieved in three different ways:

The range of the knowledge domain is expanded: the ES can hence say more.

The accuracy of recommendations is enhanced: the ES can hence be more reliable.

The ES speed is improved: recommendations are hence generated in a faster manner.

Uncertainty and lack of learning capabilities, experienced in expert systems, lead researchers to direct their focus to other AI techniques like fuzzy logic and neural networks.

C.2 FUZZY LOGIC

Most knowledge is heuristic and, hence, is usually imprecise and uncertain [C.7]. Consequently, uncertainty is present in expert systems because it relies on knowledge bases as mentioned before. This prompts researchers to devise new ways that are capable of considering uncertainty during inference processing. Fuzzy logic is now recognised as an excellent tool for performing reasoning upon common sense and vague knowledge-bases [C.8].

Fuzzy logic allows the representation and processing of uncertain or vague information such as linguistic statements. Judging the degree of membership of classes provides a way of quantifying information that otherwise could not be used in mathematical modelling. Fuzzy logic has the potential to represent complex situations, and can be

implemented on localised control systems, which makes this approach very appealing for application in future generations of automated manufacturing systems[C.9].

In 1965, Lotfi Zadeh introduced the concept of a fuzzy set, that is, a class with no sharp boundaries, in which he provided a basis for a qualitative approach to the analysis of complex systems in which linguistic rather than numerical variables are employed to describe system behaviour and performance. In this way, a much better understanding of how to deal with uncertainty may be achieved, and better models of human reasoning may be constructed [C.10].

C.2.1 Fuzzy Set Theory

The basic idea in the fuzzy set theory is the notion of a fuzzy set, which is the mathematical concept of a set that allows partial membership for elements in its domain. An element (x) may partially belong to a fuzzy set (A). Each fuzzy set is characterized by its membership function $f_A(x)$ that states, for any given (x) in the domain (A), “how much of (x) belongs to (A)”. Thus, a membership function is the relationship between the values of an element and its degree of membership in a set [C.11]. A fuzzy model, which refers to fuzzy sets and fuzzy analysis, is suitable for dealing with impreciseness that may exist in the parameters of any system.

Fuzzy logic deals with fuzzy sets and logical connectives for modelling human-like reasoning problems of the real world. A fuzzy set includes all elements of the universal set of the domain but with varying membership values in the interval (0,1) [C.12].

C.2.2 Fuzzy Systems

Fuzzy sets form the building blocks for fuzzy IF – THEN rules. A fuzzy system is a set of fuzzy rules that converts inputs to outputs [C.13]. Building fuzzy systems is carried out as follows [C.14]:

Select the input and output variables.

Define the fuzzy sets.

Construct the relationship between input and output (fuzzy rules).

Run a simulation of the system model.

Interpret the output (defuzzify to get expected values).

C.2.3 Fuzzy Logic Control

Fuzzy logic has been efficiently used in systems control. Systems control in manufacturing are responsible for sequencing and scheduling orders, monitoring and executing detailed plans, and monitoring system status in an environment that is prone to disruptions (e.g. machine failures, inaccurate information). Appropriately designed fuzzy-logic controller (FLC) can support the requirements of system adaptability and reconfiguration, which is capable of flexible and responsive recovery from disruptions in manufacturing systems [C.15].

Fuzzy control is founded on fuzzy logic, which is attractive due to its similarity to the human way of thinking. The purpose of a fuzzy logic controller (FLC) is to achieve and maintain some desired state of a process or a system. This is performed by monitoring several process variables and modifying selected control parameters [C.16].

A fuzzy controller is a fuzzy system model. It employs fuzzy sets to represent the properties of each control and solution variable. Then the controller processes its input and output by using a set of production IF – THEN rules that associate an input value, through a collection of fuzzy sets, into a new output representation.

The structure of a basic FLC is shown in Figure C.2. The system consists of four principal components: the condition interface, a knowledge base that consists of fuzzy control rules and fuzzy set definitions database, the fuzzy controller itself, and the action interface.

The condition interface is responsible for three main activities, which are; measurement of input variables, scaling the input values into a corresponding universe of discourse, and mapping the input variables into suitable linguistic values, or fuzzification of input variables.

The knowledge base contains the knowledge of the application domain and the control goals. It consists of a set of control rules and a set of definitions. These definitions

represent the descriptions of the fuzzy sets that correspond to the various input parameters.

The fuzzy controller is the unit that combines all the elements of the system into a decision-making unit operating in real time or in timely intervals.

The action interface is the unit responsible to communicate the decision made by the FLC back into the controlled system. This unit translates the decision variable into a corresponding universe of discourse and converts the fuzzy values into a crisp control action or decision, or defuzzification of the decision made.

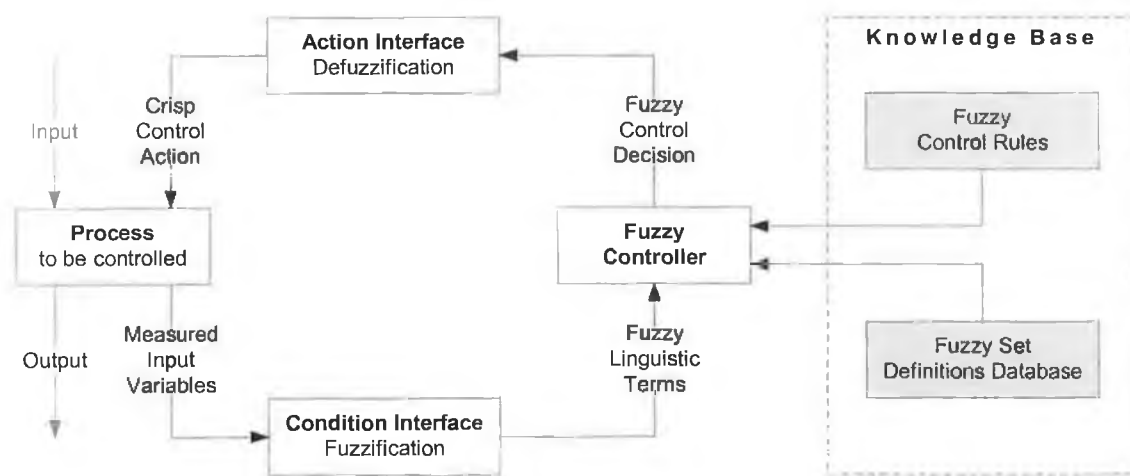


Figure C.2: Structure of a fuzzy logic controller.

C.3 NEURAL NETWORKS

Expert systems suffer from the fact that they rely solely upon expert's knowledge in the knowledge base, and cannot make decisions based on new situations, which are not in the knowledge base. In other words, expert systems are not self-learning systems [C.17].

Neural networks, on the other hand, can learn from experience and generalise from previous examples, but lack the ability of explaining decisions like the expert systems. Moreover, they can modify their behaviour in response to the environment. These systems are large networks of processing elements or nodes which process information dynamically in response to external inputs. The nodes are simplified models of biological neurones. The knowledge in a neural network is distributed throughout the

network in the form of inter-node connections and weighted links, which form the inputs to the nodes.

Basically, learning is a process by which the free parameters (i.e., the weighted links) of a neural network are adapted through a continuing process of stimulation by the environment in which the network is embedded [C.18]. A typical network model has a set of input patterns and a set of output patterns. The role of the network is to perform a function that associates each input pattern with an output pattern. A learning algorithm uses the statistical properties of a set of input-output pairs, called the training set, to generalise. With this model, statistical inferences can be developed. Statistical inference allows for exceptions and randomness in the association between two variables [C.19].

C.3.1 Neural Networks and Modelling

Neural networks behave as model free estimators, they can capture and model complex input-output relationships even without the help of a mathematical model. This property is extremely useful in situations where it is hard to derive a mathematical model that links the various parameters [C.20]. From the modelling perspective, interest in neural networks is motivated by the following important properties:

They learn by experience (not by modelling or programming).

They can generalise; i.e. they can also work in situations not taught to them.

They have inherently parallel and distributed structures.

They can form arbitrary continuous non-linear mappings.

C.3.2 Back-propagation Neural Networks

There are two classes of neural networks, supervised and unsupervised. The back-propagation network is the most popular example of a supervised network. It is very powerful for constructing non-linear transfer functions between several continuous valued inputs and one or more continuously valued output. The network basically uses a multi-layer perceptron structure; with one input layer, one output layer, and some hidden layers between the input and output layers. Back-propagation networks get its

name from the manner in which it processes errors during training [C.21]. The structure of a typical back-propagation network is shown in Figure C.3.

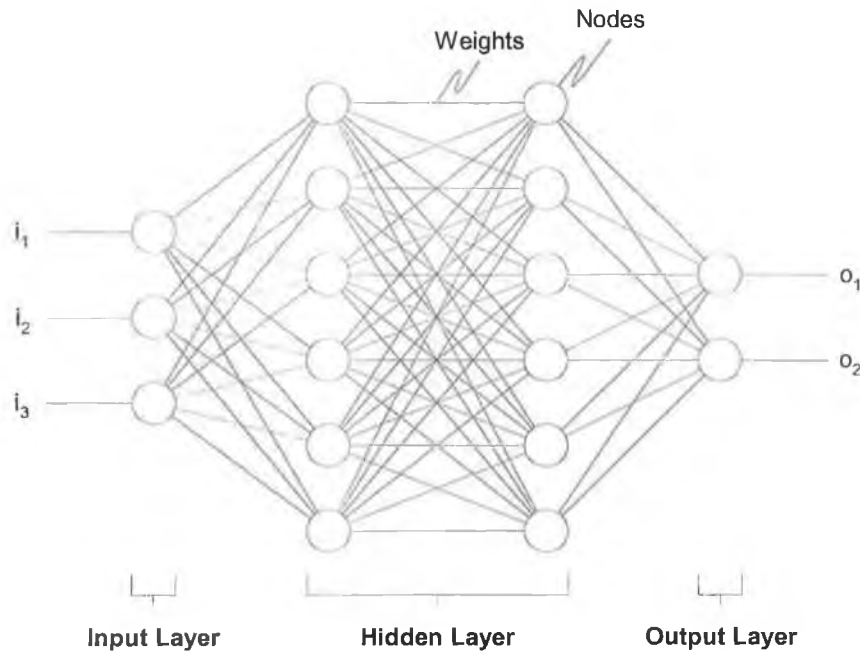


Figure C.3: Back-propagation neural network.

After the initial weights have been randomly specified and the input has been presented to the neural network, each neurone computes weighted sum of inputs from all neurones in the preceding layer and outputs the sum calculated to succeeding layers.

The sums and activation (output) values for each neurone in each layer are propagated forward through to the entire network to compute an actual output and error for each neurone in the output layer. The error for each neurone is computed as the difference between actual output and its corresponding target output. The partial derivative of sum-squared errors of all the neurones in the output layer is propagated back through the entire network and the weights are updated.

C.4 INTELLIGENT HYBRID SYSTEMS

The designing of hybrid architectures for intelligent systems is a new field of artificial intelligence research concerned with the development of the next generation of intelligent systems. Intelligent hybrid systems (IHS) represent a combination of

different intelligent techniques such as fuzzy logic, neural networks and expert systems. The limitations of intelligent techniques are a main reason for the creation of these systems in which two or more techniques are combined to overcome the limitations of each individual technique [C.22-24]. Figure C.4 shows the combination of AI techniques to develop new hybrid systems.

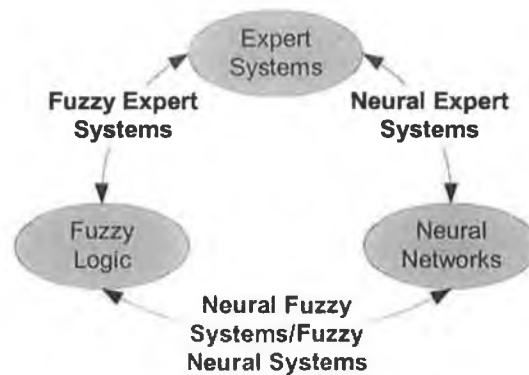


Figure C.4: Hybrid systems developed by combining different AI techniques.

C.4.1 Fuzzy Expert Systems

Fuzzy expert systems use fuzzy data, fuzzy propositions, and fuzzy logic. Fuzzy rules and the membership functions make up the knowledge base of the system. It is typically an expert system, which incorporates fuzzy sets and/or fuzzy logic into its reasoning process and/or knowledge representation scheme. Fuzzy expert systems are proposed to behave more human like because they handle concepts and reasoning in the same fuzzy and natural way as the human being does [C.25].

Application of the well-established theory of fuzzy sets and fuzzy logic to the problem of modelling imprecision and uncertainty in expert systems; provides a theoretical foundation for this crucial portion of an expert system [C.26].

Fuzzy expert systems can be used for real-time process condition monitoring, using dynamic membership functions. Based on dynamic membership functions, this expert system can perform fuzzy inference to deal with time-varying situations in process operation [C.27].

C.4.2 Neural Expert Systems

The development of integrated technologies of neural networks and expert systems has shown some advancement. The complementary features of neural networks and expert systems allow the combination of these two technologies to make more powerful systems than can be built with either one by itself. This integration emerges as new systems called neural expert systems. Neural expert systems are expert systems that have neural networks for their knowledge bases [C.22].

Expert systems techniques are best when hard and fast, reasonable sized, stable sets of rules can be derived. Neural networks can address the knowledge acquisition bottleneck by creating knowledge from training data and storing the information as connection weights.

C.4.3 Neural Fuzzy Systems/Fuzzy Neural Systems

Neural networks and fuzzy logic are the two key technologies that have recently received growing attention in solving real world, non-linear, and time variant problems. These techniques, because of their learning and reasoning capabilities, do not need a mathematical model of the system, which may be difficult to obtain for complex systems. Intelligent systems based on neural fuzzy techniques have shown good potential to solve many complex real world problems [C.28]. Neural fuzzy systems (neuro-fuzzy) combine the capabilities of neural networks (e.g. ability to learn) and the capabilities of fuzzy systems (e.g. processing of vague information) [C.29].

Neural fuzzy systems can generate fuzzy logic rules and membership functions for complex systems for which a conventional fuzzy approach may fail. Neural net technology can be used to learn system behaviour based on system input-output data. This learned knowledge is used to generate fuzzy logic rules and membership functions, significantly reducing the development time.

In return, this combination, neural networks and fuzzy logic, can also help in expressing the weights of the neural nets using fuzzy rules, forming a fuzzy neural system. Fuzzy neural networks are an expansion of classical neural networks. These networks can formally represent and process both the qualitative (linguistic) and quantitative information, which usually describe a complex, multidimensional systems or decision-

making processes [C.30]. The fuzzy rules linguistically describe how to upgrade the weights as well as to state the desired output of the neurons of the hidden layers [C.31]. This provides greater insights into the neural nets, thus leading to a design of better neural nets.

C.5 REFERENCES

- [C.1] Darlington, K. The Essence of Expert Systems. Prentice Hall, Pearson Education Limited, 2000.
- [C.2] Vouros, G. "Representing, Adapting and Reasoning with Uncertain, Imprecise and Vague Information." Expert Systems with Applications 19.3 (2000): 167-192.
- [C.3] Cavusgil, S. T. and Evirgen, C. "Use of Expert Systems in International Marketing: An Application for Co-Operative Venture Partner Selection." European Journal of Marketing 31.1 (1997): 73-86.
- [C.4] Jayaraman, V. and Srivastava, R. "Expert Systems in Production and Operations Management: Current Applications and Future Prospects." International Journal of Operations and Production Management 16.12 (1996): 27-44.
- [C.5] Metaxiotis, K. S. and Samouillidis, J.-E. "Expert Systems in Medicine: Academic Illusion or Real Power?" Information Management and Computer Security 8.2 (2000): 75-79.
- [C.6] Raggad, B. G. and Gargano, M. L. "Expert System: Defection and Perfection." Logistics Information Management 12.5 (1999): 395-406.
- [C.7] D'Aquillaa, R. O., Crespoa, C., Matec, J. L., and Pazos, J. "An Inference Engine Based on Fuzzy Logic for Uncertain and Imprecise Expert Reasoning." Fuzzy Sets and Systems 129 (2002): 187-202.
- [C.8] Yen, V. C. "Rule Selections in Fuzzy Expert Systems." Expert Systems with Applications 16 (1999): 79-84.
- [C.9] Monfared, M. A. S. and Steiner, S. J. "Fuzzy Adaptive Scheduling and Control Systems." Fuzzy Sets and Systems 115 (2000): 231-246.
- [C.10] Zadeh, L. A. Achievements and Principal Contributions. Online. The Berkeley Initiative in Soft Computing. Available: <http://www.cs.berkeley.edu/~zadeh/acprco.html>. 14 July 2003.
- [C.11] Drakopoulos, J. A. "Probabilities, Possibilities, and Fuzzy Sets." Fuzzy Sets and Systems 75.1 (1995): 1-15.
- [C.12] Konar, A. Artificial Intelligence and Soft Computing. CRC Press, 2000.

- [C.13] Kosko, B. Fuzzy Engineering. Prentice-Hall Inc., 1997.
- [C.14] Cox, E. The Fuzzy Systems Handbook. Academic Press, Inc., 1994.
- [C.15] Brennan, R. W. and Foroughi, B. "A Control Framework to Support Responsive Manufacturing." International Journal of Agile Management Systems 1.3 (1999): 159-168.
- [C.16] Arieh, D. B. and Lee, E. S. "Fuzzy Logic Controller for Part Routing." Design and Implementation of Intelligent Manufacturing Systems. Ed. M. Jamshidi: Prentice-Hall Inc., 1995. 81-106.
- [C.17] Ali, A. F., Vassiliadis, C. A., Gunasekera, J. S., Anbajane, R., and McDonald, S. "An Intelligent Hybrid System for Synthesis and Control of Metal Forming Processes." Design and Implementation of Intelligent Manufacturing Systems. Eds. H. R. Parsaei and M. Jamshidi: Prentice-Hall Inc., 1995. 39-63.
- [C.18] Haykin, S. "Neural Networks - a Guided Tour." Soft Computing and Intelligent Systems: Theory and Applications. Ed. M. M. Gupta: Academic Press, 2000. 71-80.
- [C.19] Gupta, P. and Sinha, N. K. "Neural Networks for Identification of Non-Linear Systems: An Overview." Soft Computing and Intelligent Systems: Theory and Applications. Eds. N. K. Sinha and M. M. Gupta: Academic Press, 2000. 337-356.
- [C.20] Raj, K. H., Sharma, R. S., Srivastava, S., and Patvardhan, C. "Modelling of Manufacturing Processes with ANNs for Intelligent Manufacturing." International Journal of Machine Tools and Manufacture 40 (2000): 851-868.
- [C.21] Martin, N. M. and Jain, L. C. "Introduction to Neural Networks Fuzzy Sets Genetic Algorithms and Their Fusion." Fusion of Neural Networks Fuzzy Sets and Genetic Algorithms. Eds. L. C. Jain and N. M. Martin: CRC Press LLC, 1999. 1-12.
- [C.22] Yahia, M. E., Mahmood, R., Sulaiman, N., and Ahmad, F. "Rough Neural Expert Systems." Expert Systems with Applications 18 (2000): 87-99.
- [C.23] Su, D. and Wakelam, M. "Intelligent Hybrid System for Integration in Design and Manufacture." Journal of Materials Processing Technology 76.1-3 (1998): 23-28.
- [C.24] Rao, M., Sun, X., and Feng, J. "Intelligent System Architecture for Process Operation Support." Expert Systems with Applications 19.4 (2000): 279-288.
- [C.25] Rocha, A. F., Giorno, F., Leão, B., and Theoto, M. "The Physiology of the Expert System." Fuzzy Expert System. Ed. A. Kandel: CRC Press Inc., 1992. 81-98.
- [C.26] Hall, L. O. and Kandel, A. "Fess: A Reusable Fuzzy Expert System." Fuzzy Expert System. Ed. A. Kandel: CRC Press Inc., 1992. 181-193.

- [C.27] Feng, E., Yang, H., and Rao, M. "Fuzzy Expert System for Real-Time Process Condition Monitoring and Incident Prevention." Expert Systems with Applications 15 (1998): 383-390.
- [C.28] Khan, E. "Neural Fuzzy Based Intelligent Systems and Applications." Fusion of Neural Networks Fuzzy Sets and Genetic Algorithms. Eds. L. C. Jain and N. M. Martin: CRC Press LLC, 1999. 107-139.
- [C.29] Nürnberger, A., Radetzky, A., and Kruse, R. "Using Recurrent Neuro-Fuzzy Techniques for the Identification and Simulation of Dynamic Systems." Neurocomputing 36.1-4 (2001): 123-147.
- [C.30] Satlak, G. "Fuzzy Neural Networks in Intelligent Manufacturing Systems." International Workshop on Intelligent Data Acquisition and Advanced Computing Systems: Technology and Applications, 2001. 203-206.
- [C.31] Castro, J. L., Delgado, M., and Mantas, C. J. "A Fuzzy Rule-Based Algorithm to Train Perceptrons." Fuzzy Sets and Systems 118.2 (2001): 359-367.

D INITIAL AMHS SIMULATION MODEL

D.1 OVERVIEW

The model is based on the hierarchical modelling approach, which allows decomposing of the simulation models into modules and sub-modules. This approach has the advantage of building different modules separately and ultimately linking them together. Moreover, it allows integrating additional features (as add-ons) to the model whenever needed. Finally, decomposition of modules into sub-modules could be achieved to any level of detail needed.

Another feature of the model is the flexibility upon which it is based. The process flow, layout of tools, and even number of production tools required for the manufacturing process are not fixed yet; thus the model should be able to accommodate to changes and updates of this information.

Lots flowing through the model are identified by their layer number (step sequence number), which tells the model where the lot should be going and where is it coming from.

Moreover, the model aims at presenting the end-user a friendly interface for data input and retrieval coupled with a variety of charts and reports. This interface is typically an excel file that is either imported into the model, reflecting changes in inputs; or exported from the model, reporting the different simulation outputs.

D.2 MAIN MODULES

The model has three main modules and an auxiliary data collection and analysis module. The modules, shown in Figure D.1, are as follows:

1. *Ramping and Hot Lots*. This is used to generate the various types of lots into the model on timely basis as dictated by the ramp profile.
2. *Lots Processing*. Simulates the actual processing of lots within the two bays considered in this model.
3. *Lots Delays*. Simulates the delay caused by other bays that are not considered in this model.

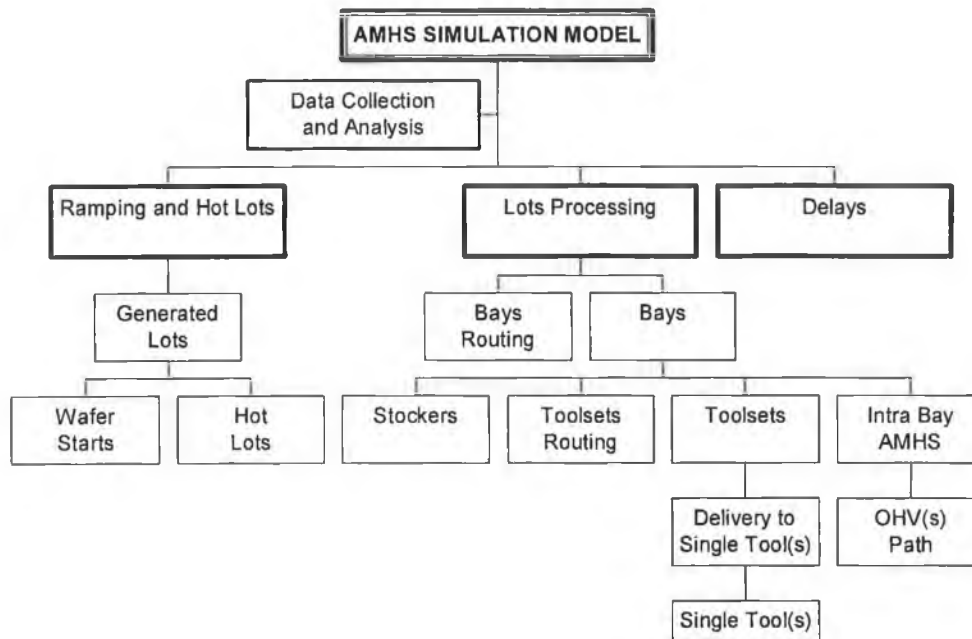


Figure D.1: Hierarchical structure of the model and its main modules.

D.3 ASSUMPTIONS

A number of assumptions for the operation of the model have been made. Presented here are the essential assumptions required to define the system being modelled.

D.3.1 Re-routing of Lots

Lots may have to be re-routed to another production tool of the same type but in another bay. If the production tool is not available and the lot has waited for a certain amount of time for it, the lot is re-routed to the other bay (this amount of time is not available yet).

D.3.2 AMHS Deliveries

The track is the loop on which OHVs continuously circle starting from stockers and back to it again.

OHVs circle the loop anti-clock wise.

An OHV can only carry one Lot.

OHV deliveries are Stocker – Tool – Stocker.

Logic behind delivery to tools is assumed to be loading the first available tool.

D.3.3 Delays

These are used to emulate the processing of lots in other bays that have not been modelled. This delay time is equivalent to the processing time that would be needed by production tools in the remaining bays to process a lot, however it does not consider the number available from each production tool, AMHS delays...etc.

D.4 BASIC CAPABILITIES

The model developed has a number of capabilities that can be identified by identifying the factors that have been modelled, representing the model level of detail, and by the outputs reported from the model.

D.4.1 Level of Detail

Products

The model handles only one type of product. In addition, it has the capability to account for hot lots, which have higher priority than ordinary lots. Lot size is constant, for both types of lot, and is equal to 25 wafers.

Ramp Profiles

There are two ramp profiles used in the model, a 2.6K WSPW and 5K WSPW. Also, there is a separate one for hot lots.

Process Flow

The model is using a compressed version of the process flow having all the processing times in bays, other than the modelled ones, summed up.

Learning Rate

The model uses learning rates to adjust the tools processing time according to the simulated workweek. The learning rate increases the processing time at the start of production and then decreases it by time when production stabilises.

Run Rate

Production tools run rate is extracted from a simulation model used in the firm, and the run rate unit is wafers/hour. However, the value that the model is using for presenting the time needed to process a lot in hours, is the processing time which is calculated using the following formula:

$$\text{Processing Time} = \frac{25}{\text{Run Rate}} * \frac{1}{\text{Learning Rate}} \quad \text{Hours / Lot.}$$

Sampling Rates

These were also extracted from the simulation model used in the firm. The model deals with it as a value that is outputted according to a certain probability (sampling rate). Depending on this value the lot is either skipped or processed on a specific toolset.

Production Tools Downtime

The mean time between failures (MTBF) and time to repair (TTR) are used to simulate the breakdown of production tools. MTBF and TTR have a lognormal distribution, and they represent both scheduled and unscheduled downtimes. Raw data extracted from the production tools are not available.

Production Toolsets Ramping and Number of Single Tools

Wafer fabrication starts in the fab in low production volumes depending on the ramp profile, and therefore production can start with lower number of single tools forming a

production toolset. However, the number of single tools increases incrementally as the production volume increases, till the fab is running at full capacity and with all the production tools installed.

Number of single tools forming a production toolset is defined in a production toolsets ramp table. Each week shows the number of single tools in a production toolset needed depending on the production volume in that week. The values used to represent the number of single tools, which are extracted from the layout, change depending on the simulated week.

AMHS Parameters

The AMHS is modelled by two main parameters, distance travelled by OHVs and the speed of OHVs. The distance is extracted from the layout, which is the distance from the input/output port of a stocker to a mid-distance between loading ports of a production tool. The speed is calculated as a weighted average of the OHV speed on straight and bent tracks using the following formula:

$$V = \frac{aX + bY}{a + b}$$

Where,

V : the speed of the OHV.

a, X : Length of straight track, and OHV speed on straight respectively.

b, Y : Length of bent track, and OHV speed on bent tracks respectively.

It should be noted that the model represents distances in ft, and speed in ft/hr, because the standard blocks used for material handling that are provided with ExtendTM use these units. Moreover, the number of OHVs in each bay can be changed depending on the user input. Finally, the model has the capability to shutdown a whole intrabay AMHS according to MTBF and TTR, however the data related to this capability have not been acquired, and therefore the AMHS is assumed to run without any breakdowns.

D.4.2 Outputs

Toolsets Availability and Utilisation

Availability and utilisation reported by ExtendTM are averaged over the whole simulation run; however, the firm is concerned with their weekly value, which cannot

be retrieved directly using standard Extend™ machine blocks. Other blocks are needed to accumulate the amount of time a production tool is processing a lot and the breakdown time.

Therefore, availability is calculated per week for every single tool in a production toolset using the following formula:

$$Availability = 1 - \frac{Sum\ of\ Downtime}{168}$$

Where, ‘sum of downtime’ is the total amount of time the tool is down in a specific week due to scheduled and unscheduled downtime.

Similarly, utilisation is calculated per week for every single tool in a production toolset, using the following formula:

$$Utilization = \frac{Total\ Time\ In\ Use}{168}$$

Where, ‘total time in use’ is the amount of time the tool is used to process lots in a specific week.

Figure D.2 shows an example of calculating the weekly availability and utilisation of a production tool. Using the equations given above, the availability in this example is 90% and the utilisation is 75%.

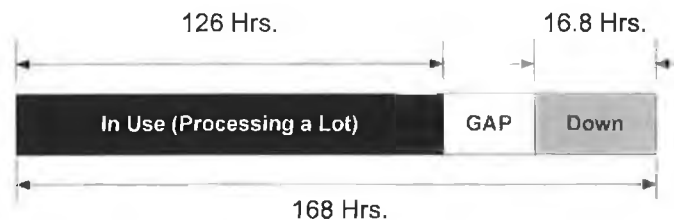


Figure D.2: Availability and utilisation example.

AMHS Availability and Utilisation

These are calculated for the OHVs and are based on the percentage of time the OHVs is down or in use over the course of the simulation run.

AMHS Moves Per Hour (MPH)

These are the total number of lots that are output from all the tools in a specific bay every hour, which are averaged per week.

Outs per Week

These are the number of wafers out per week, which can also be obtained separately for hot lots.

Throughput Time

The factory throughput time is calculated in days. The model tags the lots that are introduced into the model to know the time it has started manufacturing, and when it is out it checks the current time of the system; the difference between the two values represent the total amount of time a specific lot spends in the model. Again, this is obtained separately for hot lots.

Amount of WIP

Represents the total number of wafers that have not completed manufacturing and are still being processed. This is calculated on a weekly basis as follows:

$$WIP = (Lots\ IN - Lots\ OUT) * 25 \quad Wafers$$

E THE IDEF0 FUNCTION MODELLING METHOD

E.1 IDEF HISTORY

IDEF is an Integration computer-aided manufacturing DEFinition language that consists of a set of re-engineering techniques developed by the Air Force to facilitate manufacturing automation [E.1].

The IDEF techniques have been developed to describe, specify, and model manufacturing systems in a structured graphical form. Following structured system analysis methodology, these methods supply a powerful means of analysis and development of systems [E.2]. The interest of the work is in the IDEF0 method within the IDEF family.

E.2 THE IDEF0 FUNCTION MODELLING METHOD

IDEF0 (Integration DEFinition language 0) is used for functional or activity modelling of a wide variety of automated and non-automated systems for existing and non-existing systems [E.3].

The two primary modelling components used in IDEF0 are:

Functions (represented on a diagram by boxes)

Data and objects that inter-relate those functions (represented by arrows).

IDEF0 describes any process as a series of linked activities, each with inputs and outputs. External or internal factors control each activity, and each activity requires one or more mechanisms or resources [E.4]. Inputs are data or objects that are consumed or transformed by an activity. Outputs are data or objects that are the direct result of an activity. Controls are data or objects that specify conditions that must exist for an

activity to produce correct outputs. Finally, mechanisms (or resources) support the successful completion of an activity, but are not changed in any way by the activity. Figure E.1 illustrates generically how IDEF0 is used to depict activities, inputs, outputs, controls, and mechanisms.

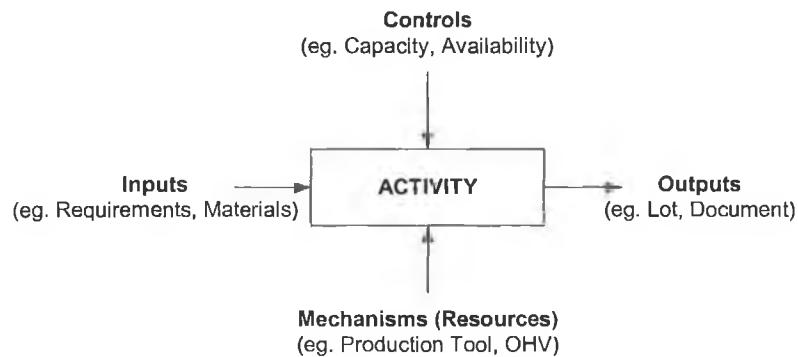


Figure E.1: IDEF0 building block.

The essence of IDEF0 is its hierarchical approach, in which a basic, single-activity description of the process is decomposed systematically into its constituent activities. This decomposition can be to whatever level of detail appropriate for the purposes at hand. IDEF0 can help in describing exactly what is happening in a system and in as complete a level of detail as desired. The result of applying IDEF0 to a system is a model that consists of a hierarchical series of diagrams, text, and glossary cross-referenced to each other [E.5].

As a function modelling language, IDEF0 has the following characteristics [E.6]:

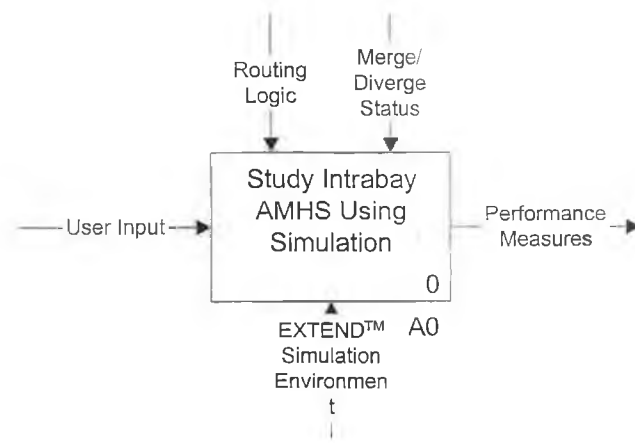
1. It is comprehensive and expressive, capable of graphically representing a wide variety of manufacturing operations to any level of detail.
2. It enhances communication between systems analysts, developers and users through ease of learning and its emphasis on hierarchical exposition of detail.
3. It is well tested and proven, through many years of use in Air Force and other government development projects, and by private industry.
4. Numerous commercial products support development and analysis of IDEF0 diagrams and models.

On evaluating the IDEF0 modelling technique for describing simulation models, IDEF0 shows a number of strengths and weaknesses in its use. IDEF0 is well suited for documenting the flows in a simulation model. It does not cover all the aspects of the model documentation and do not say much about low-level logic and therefore in most cases it should be coupled with flowcharts. It cannot handle parallel processes well, but it plays an important role in giving an overall understanding of the simulation model. The quality of being comprehensive and broadly used among engineers makes the IDEF0 model documentation a good choice [E.7].

MS Visio has been used to model GTIA-M&S, in full, using the IDEF0 function modelling method; the result of this modelling activity is the charts that follow.

E.3 REFERENCES

- [E.1] Kappes, S. "Putting Your IDEF0 Model to Work." Business Process Management Journal 3.2 (1997): 151-161.
- [E.2] Plaia, A. and Carrie, A. "Application and Assessment of IDEF3 - Process Flow Description Capture Method." International Journal of Operations & Production Management 15.1 (1995): 63-73.
- [E.3] Whitman, L., Huff, B., and Presley, A. "Structured Models and Dynamic Systems Analysis: The Integration of the IDEF0/IDEF3 Modeling Methods and Discrete Event Simulation." Proceedings of the 1997 Winter Simulation Conference. Eds. S. Andradóttir, K. J. Healy, D. H. Withers, and B. L. Nelson. ATLANTA, GEORGIA, 1997. 518-524.
- [E.4] Fülischer, J. and Powell, S. G. "Anatomy of a Process Mapping Workshop." Business Process Management Journal 5.3 (1999): 208-237.
- [E.5] Perera, T. and Liyanage, K. "Idef Based Methodology for Rapid Data Collection." Integrated Manufacturing Systems 12.3 (2001): 187-194.
- [E.6] Integration Definition for Function Modeling (IDEF0). 21 December 1993. Online. National Institute of Standards and Technology. Available: <http://www.idef.com/Downloads/pdf/idef0.pdf>. 15 May 2003.
- [E.7] Oscarsson, J. and Moris, M. U. "Documentation of Discrete Event Simulation Models for Manufacturing System Life Cycle Simulation." Proceedings of the 2002 Winter Simulation Conference. Eds. E. Yücesan, C.-H. Chen, J. L. Snowdon, and J. M. Charnes. San Diego, CA, 2002. 1073-1078.



PURPOSE: TO UNDERSTAND AND DOCUMENT THE MODEL DESIGN AND STRUCTURE OF THE GTIA-M&S
VIEWPOINT: SYSTEM ANALYST

