ABSTRACT

Modeling and Analysis of Semiconductor Manufacturing Processes Using Petri Nets

by

Xinyao Wang

This thesis addresses the issues in modeling and analysis of multichip module (MCM) manufacturing processes using Petri nets. Building such graphical and mathematical models is a crucial step to understand MCM technologies and to enhance their application scope.

In this thesis, the application of Petri nets is presented with top-down and bottom-up approaches. The theory of Petri nets is summarized with its basic notations and properties at first. After that, the capability of calculating and analyzing Petri nets with deterministic timing information is extended to meet the requirements of the MCM models. Then, using top-down refining and system decomposition, MCM models are built from an abstract point to concrete systems with timing information. In this process, reduction theory based on a multiple-input-single-output modules for deterministic Petri nets is applied to analyze the cycle time of Petri net models. Besides, this thesis is of significance in its use of the reduction theory which is derived for timed marked graphs - an important class of Petri nets.

MODELING AND ANALYSIS OF SEMICONDUCTOR MANUFACTURING PROCESSES USING PETRI NETS

by
Xinyao Wang

A Thesis
Submitted to the Faculty of
New Jersey Institute of Technology
in Partial Fulfillment of the Requirements for the Degree of
Master of Science in Electrical Engineering
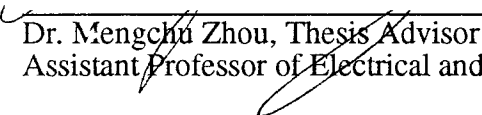
Department of Electrical and Computer Engineering

May 1993
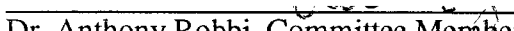
APPROVAL PAGE

Modeling and Analysis of Semiconductor Manufacturing Processes using Petri Nets

Xinyao Wang

5/13/93

Dr. Mengchu Zhou, Thesis Advisor   (date)
Assistant Professor of Electrical and Computer Engineering, NJIT

5/12/93

Dr. Anthony Robbi, Committee Member   (date)
Associate Professor of Electrical and Computer Engineering, NJIT

5/12/93

Dr. David Wang, Committee Member   (date)
Assistant Professor of Computer and Information Science, NJIT

BIOGRAPHICAL SKETCH

Author: Xinyao Wang

Degree: Master of Science in Electrical Engineering

Date: May 1993

Date of Birth:

Place of Birth:

Undergraduate and Graduate Education:

- Master of Science in Electrical Engineering,
  New Jersey Institute of Technology, Newark, NJ, 1993

- Bachelor of Science in Electrical Engineering,
  Jiaotong University, Shanghai, P. R. China, 1991

Major: Electrical Engineering

To my family

ACKNOWLEDGMENT

I take this opportunity to express my deep gratitude to Dr. Mengchu Zhou, my thesis advisor for his invaluable support and guidance throughout the research and development of this thesis. I thank him for his patience and for all of his suggestions to refine and improve this research work as well as manuscript.

My sincere appreciation goes to Drs. Anthony Robbi and David Wang, who agree to work on the thesis committee, given the very short notice.

I would like to thank to all my friends for their support and encouragement during my studies.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

CHAPTER 1

INTRODUCTION

Being the case with the most advanced technologies, the drive towards high performance multichip packaging has given rise to a number of engineering challenges. These challenges can be grouped into three major categories [13]:

1. Design and fabrication of the thin film interconnections

2. Integrated circuited substrate compatibility

3. Module assembly and reliability

This thesis will discuss the basic processes required for fabricating thin film multilayer interconnection and multichip mounting, with the application of Petri nets to modeling and analysis of such manufacturing processes. Petri nets are a powerful tool for studying systems involving concurrency, synchronization and choices [44]. They can be used to analyze and evaluate important industry management indices, such as cycle time, firing rate, and sequence.

1.1 Background

The term "multichip module" (MCM) may yield some confusion for those who are not intimately familiar with that technology. Simply stated, it is a new technology for integrated-circuit (IC) processing, which emerged in the late 1980s [18], and is considered as the most challenging technology in semiconductor industries in the 1990s [12, 26]. The distinguishing feature of multichip modules is that one or more pre-tested chips are mounted onto the circuited silicon wafer board, instead of the traditional IC fabricating

methods. This new technology addresses packaging problems in a manner that emulates a wafer scale while relaxing the limits imposed by having to repair defective cells [10]. It is being aggressively developed by several commercial firms and is also likely to find applications within the next couple of years [5]. In addition, the silicon wafer substrate provides a solid foundation for introduction of the advanced interconnection technologies.

A MCM manufacturing process can be described as a discrete event system (DES). As defined by Ramdgae and Wonham [28], "A DES is a dynamic system with a discrete state space and piecewise constant state trajectories, the time instant at which state transition occur, as well as the actual transitions, will in general be unpredictable."

The state transitions of a DES are called events and may be associated with the physical phenomenon that causes the change in state. For example, in a manufacturing environment, a typical event could be "machine A is working on part A". The purpose of modeling such a system is the design of a control structure to drive its state space trajectories, and a model should then provide a guide to constructing a controller. Once a model of the system has been constructed, the analysis and computation of the model turn ahead. Petri nets may be one of the efficient solutions to such a problem [7].

Petri nets are a graphical and mathematical modeling tool applicable to many DES's, such as manufacturing, control and information processes. As a graphical tool, Petri net can be used as a visual-communication aid similar to flow charts, block diagrams and networks. As a mathematical tool, it is possible to set up state equations, algebraic equations, and other mathematical models governing the behavior of systems. In a word, Petri nets can be used by both practitioners and theoreticians [26].

1.2 Motivation

The real-time control of manufacturing systems for processing integrated circuit (IC) is a complex task including the real-time management of factory operation as well as its integration with the direct control of machines or robots. Generally, such systems are decomposed into various levels, each in charge of a certain class of decisions [24]. Basically, these levels are: local control of machines, coordinate level control, and global level control.

gateway to the upper manager

global level controller

network

coordinate level controller

local controller

equipment
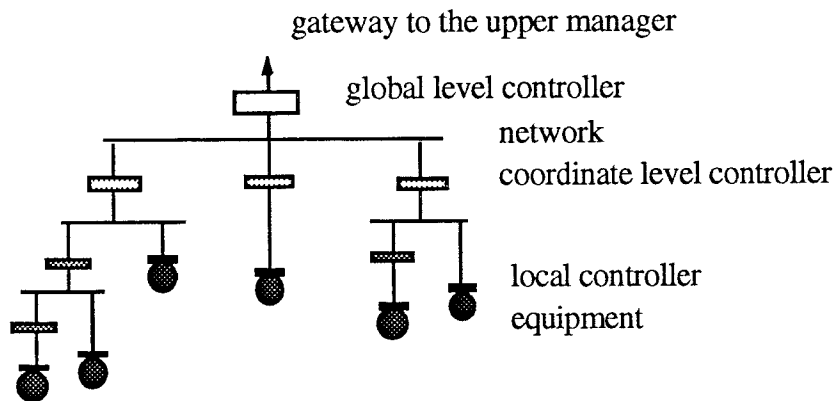
Fig. 1.1 A control system for manufacturing processes

To set up a control system like the example in Figure 1.1 is the real purpose of computer aid manufacturing processes. It realizes the real-time control in a manufacturing process. To a complex system, it is hard to expound the IC manufacturing processes in a word. Hence, modeling is a base for setting up control systems. For a large system, the modeling processes usually deal with several discrete sub-systems, and then mingle them together. Petri nets are an ideal tool for such work. Simulation is the critical step toward control systems, as it provides some useful parameters and detailed performance information. Programmed control is the way to convert the simulation model into a real-time system.

1.3 Methodology

There are several requirements for good methodologies: the need to model from a global viewpoint through a top-down approach and in details through a bottom-up approach; the need to study not only the static but also the dynamic behavior, particularly when disruption occurs (this places an emphasis on simulation); and the need for a compatible model for the transition between construction, simulation, and implementation [41, 42]. In short, all of them bring to light the most important problem which is encountered with FMS (flexible manufacturing system) design: modeling.

We need models to describe and formalize manufacturing systems in the design process. These models can be used as a reference and a guide all through the design. To describe reality and to find specification errors during the design (rather than during implementation), we need powerful design tools. Such tools must be supported by computers, and so the requirement of formality. Finally, some methods are needed in order to organize the data acquisition in a structured way during the analysis and design process. In all, a methodology should be less sensitive to specification anomalies, rigorous enough to verify designs, and flexible enough to be implemented on a computer. Such a methodology will impart many desirable properties in a design such as ease of understanding and management, efficient implementation, and ease of debugging and maintaining the system. At last, this division will result in prototypes and in a system which has "distributed control" [11, 31].

Petri nets are a graphical and mathematical tool developed dozens of years, and verified that they are a powerful tool in modeling and simulation processes. Application of Petri nets to various fields in manufacturing systems have resulted in many successes [42,44]. In Petri nets, places represent availibility of resources, operation processes, or

conditions, transitions model the events, start, or termination of operations, arcs indicate the relationship between places and transitions, and tokens in places present the dynamics to Petri nets with their flowing graphs. Thus, marked Petri nets can be used to study the dynamic behavior of the modeled DES.

1.4 Objective

This research will apply Petri nets standard to the problem of modeling a control system for integrated circuit (IC) fabrication for multichip modules (MCMs). The literature, however, does not pay sufficient attention to the reduction theory for deterministic timed Petri nets (TPNs) so that setting up the foundation to this area is necessary before modeling, and more, it gives significant sight to the TPN systems. Modeling the IC manufacturing processes yields an architectural framework for the control system, which can be directly implemented via their conversion to computer code later. An analysis of the resultant designs will give insight to the system on the different levels. Along with a greater understanding of the particular automation needs of this industry, possible extensions and further avenues of exploration will be offered to the designers.

Chapter 1 presents an overview of the thesis while skimming through the MCM technology and fabrication, Petri net analysis, performance and challenges, and finally focusing on the purpose of the research.

Chapter 2 contains a summary of MCM technology, such as its challenges, processes and the comparison with other relative technologies - PCB and VLSI.

Chapter 3 is concerned with a Petri net methodology. An overview of its development, definitions and characteristics is given while focusing on the fundamental theories of Petri nets.

Chapter 4 provides the reduction theory of Petri nets. It is really mathematical and logical analysis. Importantly, it can ease up the problem of net-complexity in analyzing DPN systems.

Chapter 5 deals with the highlight in the whole report. Petri nets are applied to a fabricating system, which concerns the MCM technology. A critical evaluation will be offered in an attempt to set up the model of MCM manufacturing processes and data acquisition into different levels. This evaluation will be based on a variety of perspectives, from the theoreticians' view to operators'.

Finally, Chapter 6 summarizes the results and conclusions of the research. Along with the analysis and results in this research, avenues of the future research are suggested and discussed.

CHAPTER 2

MULTICHIP MODULE TECHNOLOGY

With the rapid advances in wafer fabrication technology, IC designers always attempt to increase wiring density and executing speed [37, 39]. Among the alternative new technologies,wafer scale integration (WSI) has the best opportunity to become a mainstream electronics technology, as it has quite high densities, large area monolithic circuits and hybrid substrate. Being a branch of WSI, multichip module (MCM) is rapidly being developed as an advanced packaging system expected to relieve the chip-to-chip interconnection limits imposed by the conventional packaging of individual ICs [3, 32, 33]. It appears that a significant opportunity now exists for development of MCM commercial product — large scale integrated circuits [21]. Hence, MCM technologies are considered as an important and longer term evolutionary direction for silicon ICs.

2.1 Introduction

WSI has endured several infant and notorious commercial failures [1, 15]. But with the researches to develop the fundamental principles, which are needed as the underlying foundation for WSI components, WSI has received increasing attention as a major evolutionary direction in IC fabrication when the advantages of WSI become compelling over other advanced technologies [5, 16]. Multichip module (MCM), sometimes called as "hybrid WSI" or "pseudo-WSI" [15, 18], is the development of WSI, i.e., mounting the ready-made chips onto the silicon wafer circuited board. MCMs relieve the limitation of chip-to-chip interconnection imposed by the traditional IC packaging. Such an advanced

technique is essential to match the performance requirements (density and speed) of today's IC industry and more importantly the emerging VLSI (very large scale integration) technologies of silicon IC fabrications.

By using thin film technologies on large area substrates, the fabrication of MCMs draws out the techniques typically used for IC fabrication, as MCMs implement the passive interconnections in a MCM substrate. The similarity to IC technologies and resulting tight spacing of unpacked ICs promises higher density and higher speed "circuit boards". While an extensive set of references is provided to avoid repeating excessively detailed discussions available in the cited literature, this thesis provides a board overview of the objectives and motivations of the considerable work on the wafer level system components.
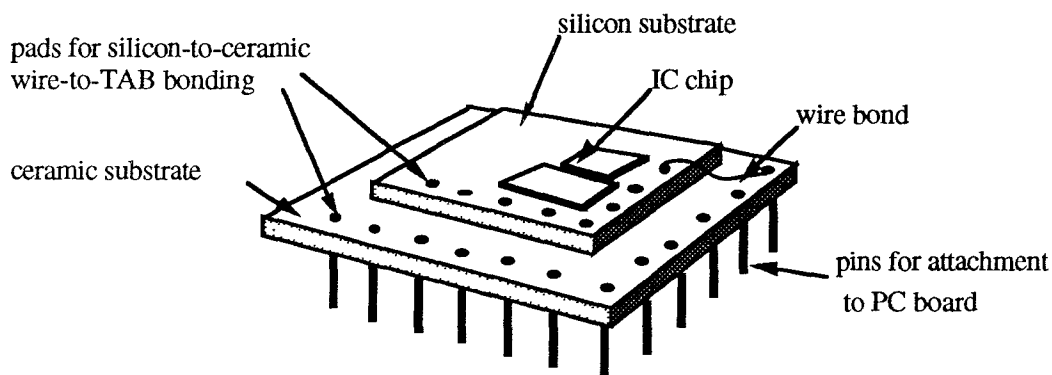
Fig. 2.1 "Silicon circuit board" packaging for high density multichip module

MCM technology provides the advantages of implementing conventional chip-to-chip interconnections with the more benign environment of "on-chip" interconnections [2, 39]. However, the use of thin film technologies does not eliminate several of traditional capabilities provided by conventional printed circuit board (PCB). Providing these PCB capabilities within the thin film technologies of MCMs is a major challenge. As a part of wafer scale integration, MCMs have had to confront similar requirements and have been unable to achieve commercial success. MCMs significantly relax the difficulty of achieving

higher yields through selective detection and repair of fault-producing defects since MCMs use passive interconnections (avoiding the need to consider defects within active devices) and have larger feature sizes than that of wafer scale ICs [29].

IC fabrication requires the use of mechanical and optical equipment and materials capable of precisely maintaining close tolerances and small geometry. As any high-technology field, a large amount of technical jargon has involved. It should be mastered by anyone wishing to be conversant with those working in the area [6, 8, 9, 30].

An *integrated circuit* (IC) is a combination of interconnected circuit elements inseparably associated on or within a continuous substrate.

A *monolithic IC* is an IC whose elements are formed in place upon or within a semiconductor substrate with at least one of the elements formed within the substrate.

A *hybrid IC* is a combination of two or more IC types or an IC with some discrete elements.

A *wafer* is a basic physical unit used in the process. It generally contains a large number of identical ICs.

A *substrate* is the supporting material upon or within which an IC is fabricated or attached.

A *test plug* is a special chip that is repeated only a few times on each wafer. It is used to monitor the process parameters of the technology. If the measurements of key parameters at the plug level are not acceptable, the wafer will be discarded.

A *test cell* is a special cell on each wafer. It differs from the test plug in that circuit designers include this cell specifically to monitor the performance of elementary subcircuits or subcomponents.

Considerable effort has been expended in the WSI field, toward using the entire wafer as a single IC, but this approach is challenged by defects in processing and associated decline in yield and by inherent delay in signals that must transverse the wafer. However, it is still a big challenge [36].

2.2 Process Overview

The fabrication of ICs requires the execution of a large number of individual and general complex interactive operations. To provide an overview of the IC fabrication process, this section outlines these steps in approximately the same order in which they occur [13, 14, 26]. As mentioned in many papers, MCMs avoid repeating excessively detailed logic designs, but provide the interconnection system on the silicon circuited board, so that it saves much detailed work in the design.

(a) Provide polished silicon wafer

(b) Plant CMOS components in wafer

(c) Coat the GND level on the wafer
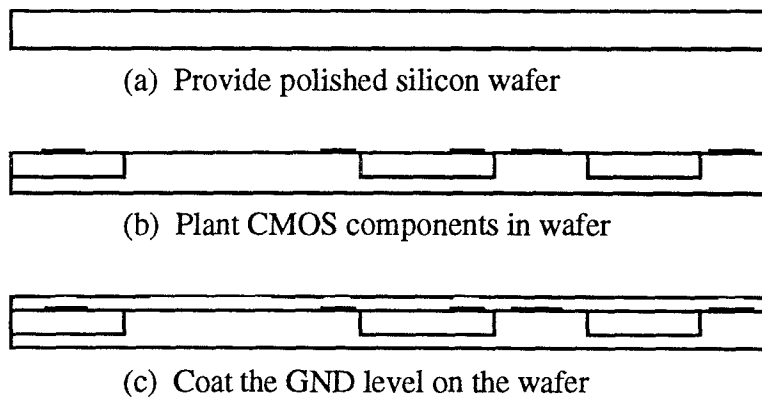
Fig. 2.2 Processes of IC fabrication (to be continued)

(d) Circuit the lower level board

(e) Coat the GND level on the wafer
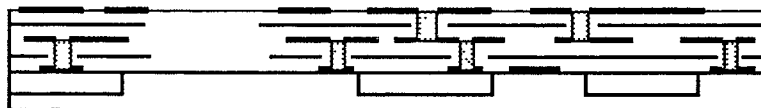
(f) Circuit the upper level board
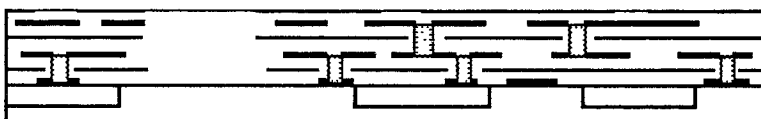
(g) Coat the GND level on the wafer
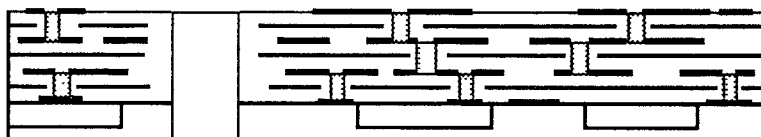
(h) Coat solder on pads and drill
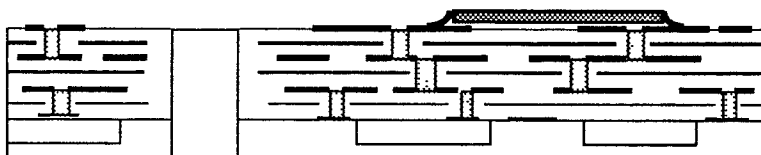
(i) Mount chips on circuited substrate

Fig. 2.2 Processes of IC fabrication (continued)

Figure 2.2(a), it shows a polished silicon wafer to the fabrication work-stations.

Figure 2.2(b) shows the process of planting CMOS switches and other simple transistors in the polished substrate. In most papers on multichip modules [6, 10, 26], this step is ignored, as the authors always emphasize on the advantages of silicon circuited

board. But in some cases, this step is valuable, as it can reduce the errors of assembly in chip-to-chip interconnections. Needless to say, the reliability of VLSI CMOS technologies is higher than that of chip-to-chip interconnections [40], so that planting CMOS devices on the raw substrate is the first step in such a system.

In Figure 2.2(c), there are two steps: plating dielectric and metallizing the first GND level. First, silicon dioxide is planted as the interlevel dielectric. Metallization is the next step to the dielectric. GND layer is used to avoid the electrical-magnetic impact between two circuited layers.

Figure 2.2(d) includes two steps: dielectric and metallization. Polyimide is applied as interlevel dielectric. The goal of metallization is to settle the lower level circuits and interconnecting with CMOS devices. It is done in the same way as that of Figure 2.2(c). Obviously, the metallization process here is much more difficult than that of GND level.

The processes from Figure 2.2(e) ~ 2.2(g) are similar to that of Figure 2.2(c) or Figure 2.2(d). In these processes, the upper level of the circuit is patterned and there are GND levels between two neighboring circuited levels. Polyimide is planted as dielectric between any different metal levels.

As shown in Figure 2.2(h), the upper level circuit is provided to the silicon circuited board, and I/O pads for chip-to-chip interconnections and bonding pads for packaging are planted in this process. Solder is supplied on the I/O pads finally.

Finally in Figure 2.2(i), the joining process for chip-to-substrate interconnections is realized. The ready-made chips are mounted onto the substrate under the C-4 (controlled collapse chip connection) technology.

2.3 Technical Terminologies

The IC fabrication processes have many special terminologies, which relate to the detailed work in fabrication of MCMs. It should be made clear during the research [6, 8, 17, 22, 30, 35, 36, 38].

2.3.1 Preparation

Once wafers have properly prepared, they should be pretreated before any fabrication processes.

The wafers are chemically and mechanically cleaned to remove dust and other forms of airbone and human contamination. They are then forced dried and given a short bake to remove residual surface moisture permitting good resist adhesion. The wafer is coated immediately after prime.

2.3.2 Patterning

Patterning is the essential process in the IC fabrication processes which is an imaging process for metallization and interconnections. This complicated process can be divided into several steps.

Coating: coating is always performed in the environment with solvent exhaust. The overall objective of this step is to provide a uniform and defect-free layer of a photo- or electro- sensitive masking material. Spin coating is a widely used technique for applying photoresist to wafers.

Softbaking: softbaking is used to remove the solvents present in the spin-coating film of photoresist and is usually performed immediately down stream from the coater in a track-type wafer handling system.

Exposure: the goal of exposure is to transmit a latent image of a desired pattern into the photoresist film. The role of photomask is obviously paramount, since the mask is a passive tool that gives the photoresist its latent image. It must therefore be as clean and defect-free as the coated and softbaked wafer is hopefully.

Development: in-line wafer developing is widely used now. It carries each wafer alone a track, where it is placed under a stream or spray of developer for a predetermined time. Rinsing and drying follow the developer step, the parts are agitated, to provide uniform developing action. The development process is one in which the developer selectively attacks and removes either exposed regions (positive resist) or unexposed region (negative resist) leaving behind image to serve as the mask for etching, or, in some case, metallization.

Postbaking: postbaking provides "insurance" that the resist is bonded well to the underlying oxide. But it is not always used. And the process is the same as that of softbaking.

2.3.3 Etching

Wet- and dry-etching techniques are widely used in developing, batch, and in-line processing. The goal is to precisely remove the semiconductor layer exposed by the developing process. Complete removal of developed resists essential and any residues left on the semiconductor layer may prevent or inhibit an etching action, a phenomenon called "blocking".

Silicon dioxide etching: The most common etching application in IC fabrication is etching undoped SiO_2. Removing selective areas of the initial undoped oxide is done by imaging the oxide with photoresist to protect those areas on which the oxide will not be etched. Since most resists bond ready to undoped SiO_2 and the etchant is not corrosive to the resists commonly used, the etching application is relatively easy to perform and simple to control. Etching doped silicon dioxide is more difficult than that of the undoped. Phosphorous-doped SiO_2 is the most common application in this category, where openings are made in the doped layer to provide for aluminum ohmic contacts. After etching the phosphorus-doped contact holes, aluminum is evaporated over the wafer and flows into these holes to make electric contact with the silicon.

Polyimide etching: Polyimide layers are etched in aqueous solvent solutions or reactive-ion etched using very thick coatings of resist as a sacrificial mask layer. "Sacrificial" etching involves the application of the photoresist coating that is thicker than the layer to be etched. After imaging the resist, the wafer is placed in the reactive-ion etch chamber and both polyimide and photoresist are removed or etched at the approximately same rate. Since the photoresist is thicker than the underlying polyimide, the polyimide layer is etched through first and the etching is terminated, leaving a thin layer of photoresist that was not sacrificed in the etching to protect the polyimide layer needed for insulation or protection of the ICs.

Aluminum etching: Since aluminum metallization serves as the interconnection layer for all the devices on the chip. Aluminum is imaged with positive photoresist mostly. Aluminum etching is accomplished with both wet and dry etchants, and both approaches are capable of meeting IC fabrication.

Removal: Photoresist removal is typically done in a batch mode, by immersing wafers in a heated resist-stripping solution or placing them in a batch plasma-stripping chamber, when an oxygen plasma removes the resist. The goal is to leave behind a surface defects in doping or metallization processes.

2.3.4 Metallization

Metallization is used in the interconnection layer on the ICs, making ohmic contact to the devices formed in the silicon and connecting these to the bonding pads on the chip's edge. The choice of metal in a multilevel interconnection structure is based on its resistivity, cost, adhesion properties to the dielectric layer, deposition technique, and resistance to long-term corrosion. The three metals of choice are gold, aluminum and copper for their superior conductivity and ease of deposition. Aluminum is used here because it adheres well to both silicon and dielectric, it can be easily vacuum deposited, and it has high conductivity. In addition, small amount of copper is added to reduce the potential for electromagnet effects, where current applied to the device induces mass transport of the metal. Small amount of silicon is also added to aluminum metallization to reduce the formation of metal "spikes" that occur over the contact holes. In some cases, titanium is added for the same reason.

Aluminum sputtering is commonly used in IC metallization processes, and is popular because the adhesion of deposited metal is excellent. It is done by ionizing inert-gas particles in an electric field (producing a gas plasma) and then directing them toward the source or target, where the energy of these gas particles physically dislodges, or "sputters off", atoms of the aluminum or other source materials.

2.3.5 Dielectrics

Dielectric materials are planted to isolate any metal layers. The purpose of coating dielectric is not only to isolate any two metal layers, but to protect the ready-made layers. Oxidation

of silicon is perhaps the most fundamental of all processes used in IC fabrication. Silicon dioxide (SiO_2) is a very stable material and is almost universally used as the surface for resist-imaging operation. Silicon interacts with oxygen. High-pressure oxidation, operating at up to 25 atmosphere (atm), permits better product.

Polyimide is also used as insulating layers, most notably between two metal layers. Polyimide tends to smooth abrupt underlying irregularities, thus reducing the effects of sharp boundaries of underlying metal layers. Spin coating is widely used in this process.

Diffusion: "Diffusion", in a classical sense, is the uniform distribution of particles within a fixed volume of space according to a physical mechanism that begins with the same particles in a concentrated state in the same fixed volume. The IC analog of this physical behavior is the thermally induced distribution of impurity atoms throughout the silicon crystal-lattice structure.

2.3.6 Joining processes

The controlled collapse chip connection (C-4) technology [32, 36] is applied for chip-to-substrate interconnections. As a two-dimensional interconnection by nature, C-4 has the small inductance and capacitance of the solder joints compared with bonding wires or TAB (tape automated bonding) leads. The key feature of the C-4 technology is the deposition of a thin film structure and solder on the I/O pads. The thin film structure, usually termed ball-limiting metallurgy (BLM) or under-bump metallurgy (UBM), serves the purpose of an adhesion layer, a diffusion barrier, and a solderable base of controlled wetting area. The solder pads are aligned and tacked onto the matching footprint on the substrate, and connection to the substrate is made during the reflow of the deposited solder. The surface tension of the molten solder prevents the chip from collapsing during reflowing.

Evaporation through metal masks and electroplating is known as the method of solder deposition. In a solder evaporation process, mechanical fixtures are necessary to align and clamp the metal mask to the wafer or to the substrate.

Reflow of deposited solder pads in N_2 is successful because it avoids reflowing the bumped dies before they are tacked for joining.

Implementation of this process could be simpler if the chips are specially designed for the C-4 process, namely having small circular via openings of uniform size in an area array format located near the center of the die.

2.3.7 Testing and packaging

After being processed, ICs are tested and packaged. The first step in a testing process generally involves a process verification to make certain that the process parameters are within the tolerances acceptable for the product, and test plugs are specially designed for this purpose. Then a wafer prober is used to make mechanical contact with the test plugs so that electrical measurements can be made. After being probed, the wafer is scribed both horizontally and vertically between adjacent dies, and the dies are separated. Following the separation, the individual dies are die attached, or die bonded, to a carrier or to the IC package itself. Wire bonds, typically gold, are subsequently made from the pins of the package to the appropriate locations on the die. After wire bonding, the packages are closed, and a final electrical test is completed.

2.4 MCMs vs PCBs and VLSIs

Multichip modules reflect physical characteristics which are shown partially in two distinct technologies — printed circuit boards (PCBs) and VLSI integrated circuits. Experiences within these technologies provide a foundation for development of MCM techniques, and are briefly summarized below.

PCBs, developed over many years, provide a very flexible platform on which complex systems comprised of large number of components can be assembled. The advantages of PCBs are:

        a. Usage of pretested components

        b. Inspectability

        c. Repairability and adaptability

        d. Ease of assembly

The disadvantages of PCBs are:

        a. Large components

        b. Large size

        c. Exposure to the air

Compared with PCBs, MCMs are also assembled with a lot of components, and inspectable and repairable, but more difficult than PCBs, as MCMs are of mini-size. MCMs integrate the circuits and components. Furthermore, they are sealed in a LCC (leadless chip carrier) package [29].

The VLSI technology has been developed for more than ten years. And it is still the mainstream in IC fabrication industries. The advantages are:

        a. Small size but large scale integration

        b. Inspectability (but more difficult than that of PCBs)

        c. Isolation of the air

The disadvantages are:

        a. Unrepairability

        b. Repeatedly detailed work

Compared with VLSI, MCM technology provides an environment for designers to use ready-made components without repeating the detail, and for operators to repair the defect components or the defect areas on substrates. It provides the flexibility of design and operation in IC fabrications. Furthermore, MCMs preserve the advantages of VLSI technology [40].

Multichip modules have qualitative similarities to the technologies discussed above, and it is the technology in relation to both PCB and VLSI technologies. But it is also clear that MCMs have unique characteristics which will make direct use of manufacturability techniques difficult from those other technologies.

Multichip modules represent the extension of the microfabrication and thin film techniques from the arena of the device structures into the arena of the system-level interconnections. MCMs address the high density, chip-to-chip interconnections using VLSI technologies. For such reasons, MCMs are presently evolving rapidly towards a commercial, generic packaging of IC components.

CHAPTER 3

PETRI NET THEORY

Modeling and analyzing large scale discrete event systems (DES's) present a big challenge. As a graphical and mathematical modeling tool, Petri nets provide strong reliability and computability for the study of DES. As a graphical tool, they can be used as a visual-communication aid, similar to flow charts, block diagrams and networks. In addition, tokens are used in these nets to simulate the dynamic and concurrent activities of systems. As a mathematical tool, they make it possible to set up state equations, algebraic equations, and other mathematical models governing the behavior of systems. Shortly, Petri nets are flexible and powerful so that they could be used not only by practitioners, but also by theoreticians [25, 31]. To some extent, for large scale systems, Petri net could be so complicated that they would lead to the problem of state space explosion. Out of question, simplification and reduction methods for the Petri net system become more important than that of any previous time [19, 20]. This research also faces to this complexity problem, and thus the reduction theories are applied to the analysis of MCM manufacturing processes.

3.1 Historical Review

Historically speaking, the concept of "Petri net" has its origin in Carl Adam Petri's dissertation, University of Darmstadt, West Germany, 1962 — "Kommunikation mit Automaten" (Communication with Automata). In the early development of Petri nets (in the late 1960's and early 1970's), the pioneers used it in some control systems. Because of

the original Petri nets lack some conditions (such as firing rate, firing time), the applications were confined to a narrow area. Since 1979, the Europeans have been very active in organizing workshops and publish conference proceedings on Petri nets. More than four tutorials were brought out. Year by year, Petri nets became more and more noticeable and popular for describing and studying the information processing systems that are characterized as being concurrent, asynchronous, distributed and/or parallel. From 1985, another series of international workshops were initiated. These series place emphasis on time and stochastic nets and their applications to performance evaluations. They are called timed-Petri net and stochastic-Petri net, respectively [25]. Petri nets become a practical tool for real-time control systems. Nowadays, Petri nets are widely applied in manufacturing control systems, discrete event systems, data-flow computing systems, fault-tolerant systems, asynchronous circuits and structures, compiler and operating systems, and so on. There are many kinds of Petri nets. Generally, they are classified into ordinary Petri nets, deterministic timed Petri nets, stochastic Petri nets and colored Petri nets, and they can also be trans-sorted to two or more classes. For different systems, different Petri nets are created and applied. The theories of Petri nets are developing so quickly that they become one of the most powerful tools to study discrete event systems [28]. This chapter indicates the essential elements of Petri nets.

3.2 Fundamentals

A Petri net is a directed graph consisting of two types of nodes, places and transitions. Places always follow transitions and vice versa. Places and transitions are connected by directed arcs.

Formally, an ordinary Petri net is defined as

$$Z = (P, T, I, O, m)$$

where P: a set of places represented by circles,

T: a set of transitions represented by bars,

I: $P \times T \to N$; the number of arcs from p to t, where $N=\{0, 1, 2, \cdots\}$

O: $P \times T \to N$; the number of arcs from t to p, and

m: a vector whose i^{th} component represents the number of tokens in i^{th} place.

I, O are two functions connecting transitions with places: I is the input function, and O, the output. They are both represented as matrices. These four items: set of places, set of transitions, input function and output function, define the structure of the Petri net.
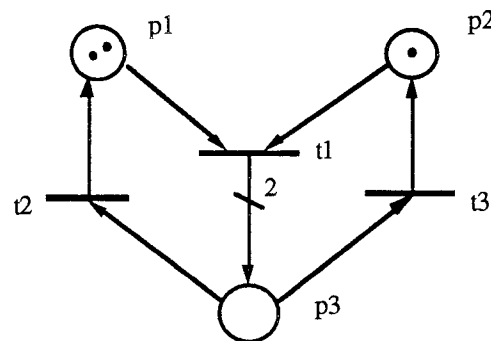
Fig. 3.1 An example of Petri net

A Petri net graphically consists of

1. *Circles* (called places) represent conditions or availability of resources, e.g.,"machine 1 available", "parts on machine 1".

2. *Bars* (called transitions) represent the initiation or termination of an event. There are two kinds of bars, vacant bars stand for deterministic timed transitions, i.e., there is time-delay when firing such a transition, and solid bars stand for immediate transitions i.e., zero time-delay.

3. *Black dots* (called tokens) in resource places, represent the availability of the corresponding resources.

4. *Arcs* represent the direction of the operating resource flow.

A transition $t \in T$ is enabled by a marking m if $\forall\, p \in P$

$$m(p) \geq I(p,t)$$

The firing of transition t generates a new marking m' with:

$$m'(p) = m(p) + O(p,t) - I(p,t).$$

3.3 Basic Definitions

First, denote preset and postset as follows:

preset of place p

$$\cdot p = \{t\colon t \in T\ \&\ O(p,t) \neq 0\},$$

postset of place p

$$p \cdot = \{t\colon t \in T\ \&\ I(p,t) \neq 0\},$$

preset of transition t

$$\cdot t = \{p\colon p \in P\ \&\ I(p,\, t) \neq 0\},$$

postset of transition t

$$t \cdot = \{p\colon p \in P\ \&\ O(p,\, t) \neq 0\},$$

Reachability set is:

A marking m' is immediately reachable from m if firing an enabled transition in m yields m'. Given the initial marking m_0 of the Petri net, it is called reachability set or forward marking class (denoted by $R(m_0)$) — the set of all possible reachable markings. In

other words, a marking m belongs to $R(m_0)$, if there exists a firing sequence q leading from m_0 to m, i.e.,

$$R(m_0)=\{m| \exists q \in T^*, q[m_0>m]\}$$

where T^* represents the set of all possible orders of the elements in T, and

$q[m_0>m]$ denotes that firing q at m_0 leads to m.

Ordinary Petri nets (OPN)

$I(p,t) \leq 1$; $O(p,t) \leq 1$ for any places and transitions.

i.e., there are not any multiple arcs in the graph.
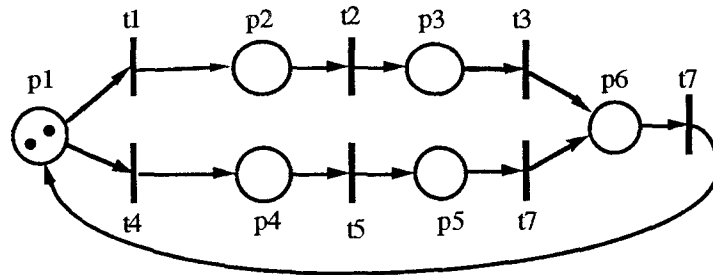
Fig. 3.2 An example of State Machine

State machine Petri net (SM)

An ordinary Petri net with

$|\cdot t|=1$; $|t\cdot|=1$ for any transitions in a graph

i.e., tokens will not increase or decrease during firing any transitions.
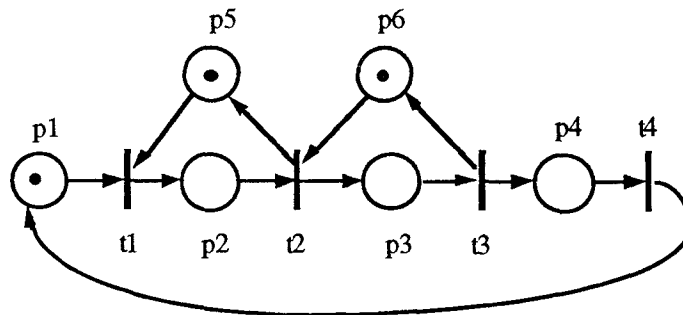
Fig. 3.3 An example of Marked Graph

Marked graph Petri net (MG)

An ordinary Petri net with

$$|\cdot p|=1; \quad |p\cdot|=1 \quad \text{for any places in a graph}$$

i.e., tokens in any cycle is constant, no matter which transition fires.

Free choice Petri net (FC)

An ordinary Petri net with

either $|p_i\cdot| \leq 1$ or $\cdot(p_i\cdot) = \{p\}$ for any place in graph

i.e., each arc from a place is either single outgoing or unique to a transition.
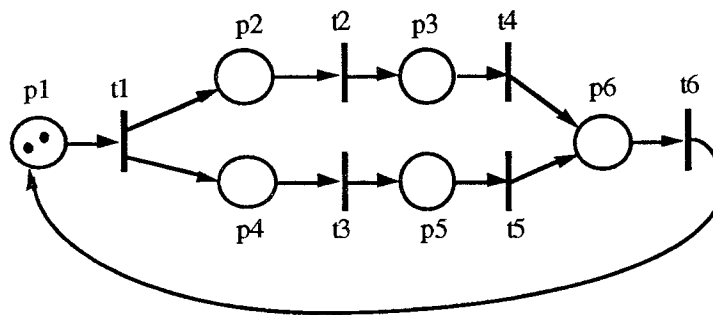
Fig. 3.4 An example of Free Choice

Extended Petri net (EPN)

A general Petri net with inhibitor arcs, probabilistic arcs or random switches, priority functions and so on.

The relationship among these graphs are shown in Figure 3.5.

Once a model of Petri net is set up, analysis is desirable to determine which properties the net possesses.
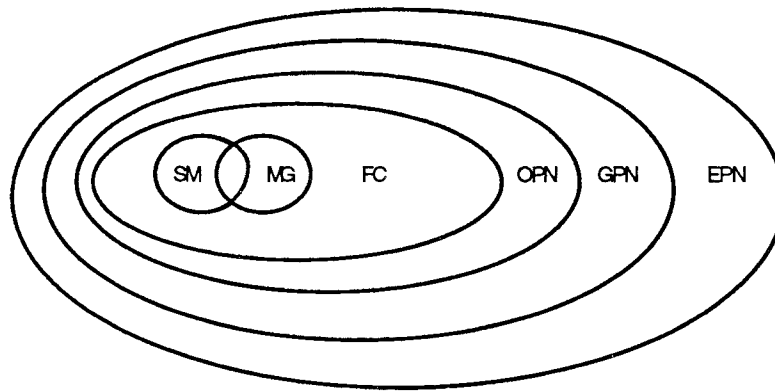
Fig. 3.5 The relations of Petri nets

3.4 Properties of Petri Nets

1) *Boundedness*

A place is K-bounded if there exists an upper bound to the number of tokens that can be simultaneously held in the place. A Petri net is K-bounded, if each place in the net is bounded and the maximum bound is K. A 1-bounded net is said to be safe.

In manufacturing environments, the boundedness of a system indicates the absence of overflows. In other words, if a system is not bounded, i.e., at least a place is not bounded, it will become a trap to absorb tokens, and then the further steps become meaningless. Therefore, a model of real systems should be bounded for any initial markings.

2) *Liveness*

A transition is live if there exists a sequence g for any marking m in the system, firing g at m will lead to a marking that enables t. A Petri net is live if all of its transitions are live.

Liveness implies no deadlock. If a transition cannot fire, it is a redundant transition and can be eliminated from the net. However, it needs to be identified since it may represent an error in the model or an inconsistency in the modeling system.

In manufacturing environments, the liveness of a system indicates the system can run well, without causing the system deadlocked. Thus, the liveness is a most important property in the system analysis. In a live net, not only a transition is firable in any given marking at least, but also it keeps the firing potent in all marking reachable from any initial marking.

3) *Reversibility*

A Petri net is reversible if for any marking, it can go forward and reach the initial marking, i.e., the initial marking is reachable from any reachable markings.

In manufacturing environments, the reversibility means the system will be re-started. Generally, when a part is finished in the process, it will input a new part automatically.

4) *Conservation*

A Petri net is conservative if the number of tokens in the net keeps constant. It implies that each transition in such a net is conservative, in the sense that the number of inputs of each firable transition is equal to that of outputs. More generally, weights can be defined for each place allowing the number of tokens to change as long as the weighted sum is constant. It guarantees the system resources are neither destroyed nor created.

5) *Persistency*

A Petri net is persistent if for any two enabled transitions firing one does not disable the other. In other word, a transition in a persistent net, once it is enabled, will stay enabled until it fires.

3.5 Timed Petri Nets

The concept of time is not given in the original definitions of Petri nets. However, for performance evaluation and scheduling problems of dynamic systems, it is at present necessary to introduce time delays associated with transitions and/or places.

It is interesting in finding how fast each transition can initiate firing in a periodically operated timed Petri net. An important parameter is cycle time, which is defined as the time to complete a firing sequence leading back to the starting marking after firing each transition at least once.

In a deterministic timed marked graph, one can assign the delay-time τ_i to each transition. The total delay in a directed cycle is the sum of the delay-time of all transitions in that cycle. Compared with that of other cycles, a maximum delay time can be obtained and is the cycle time of such a net, i.e.,

$$\pi_{min} = \max\{\text{the total delay in } C_k / m_0(C_k)\}$$

where $m_0(C_k)$ denotes the number of tokens in cycle C_k at M_0.

As defined, the number of tokens in a cycle remains unchanged in an MG during firing sequences.

3.6 Petri Net Design Method

Since in a moderately sized manufacturing system, the complement level of detail may be unreasonable, there exist two paths — top-down method and bottom-up method. Top-down method sets up an aggregate model and neglects the low-level details initially, then refines the system by expanding places and/or transitions to incorporate details into the models. In contrast, bottom-up method first classifies them into subnets,models each subnet, then merges them together by sharing nodes in the system [41, 42, 44].

Top-down techniques

Transitions and places in a net can be replaced by a more detailed sub-net step by step. This method is very attractive from the design point of view because details are introduced step by step, thereby, reducing the complexity to be dealt with at any phase in the process. Using this method, it is easy to analyze the properties of the net, such as liveness, boundedness, reversibility and others.

A Petri net model can be derived for a manufacturing system given a list of operations and their relations. The procedure is outlined as follows:

Step 1. Set up an aggregated model of the system with properties of boundedness, liveness and reversibility.

Step 2. Perform decomposition and refinement

Step 3. Appropriately add the places to represent the availability of resources.

This top-down approach method is a natural decomposition of a manufacturing system.

Bottom-up techniques

In the bottom-up method, synthesis can start with a very simple separated net, then at each synthesis step, these subnets can be merged into new nets, with sharing places and/or transitions. The procedure is:

Step 1. Separate the objective into some separate operations and set up simple subnets corresponding to the detailed operations.

Step 2. Add the resource places to the subnets

Step 3. Merge the subnets by sharing the common places, transitions or paths.

The bottom-up method is very natural from the view of the operators, as it goes from the detailed activities to the top-level outlines.

Being a graphic and mathematic tool, Petri net applies a strict theory for modeling and analysis. Facing a big system, we can separate it into several small-sized subsystems, using the top-down or bottom-up method. Then the system will be made clear and analyzed through the subnets, and finally synthesized.

CHAPTER 4

REDUCTION OF DETERMINISTIC TIMED PETRI NETS

Development of Petri nets brings light to the discrete event system (DES) analysis. However, one faces the exponential growth of the state space with net size and the number of jobs and resources. The net becomes more and more complicated and unreadable. To challenge such a problem, we resort to a general methodology — reduction. This chapter investigates reduction theory for the deterministic timed Petri net (TPN).

4.1 Fundamentals

A timed Petri net is a six-tuple set

$$Z=(P, T, I, O, m_0, \tau)$$

where P, T, I, O, m_0 are defined in the last chapter, and

τ: the set of firing time functions and τ_i is used to denote the time delay associated with the ith transition.

All delays are deterministic. Denote these timed Petri nets by TPNs in this thesis.

An *elementary path* is a sequence of nodes: $x_1 x_2...x_n$, $n \geq 1$, such that $\exists$ an arc (x_i, x_{i+1}), $1 \leq i < n$ if $n > 1$, and $x_i = x_j$ implies that $i = j$, $1 \leq i$, $j \leq n$. An *elementary circuit* is $x_1 x_2...x_n$, $n > 1$ such that $x_i = x_j$, $1 \leq i < j \leq n$, implies that $i = 1$ and $j = n$. A PN is strongly connected if for every pair of vertices v_i, $v_j \in P \cup T$, there exists an elementary path from v_i to v_j, i.e., vertices in the net can definitely connected with directed paths from each other.

32

As discussed in Chapter 1, the analysis of deterministic timed Petri nets faces the complexity problem. Therefore, reduction techniques need to be used to mitigate the effort of checking Petri net properties. The reduction steps should not lose important properties about the net.

First, introduce the following some definitions about subnets [19].

Definition 1. $Z'=(P', T', I', O')$ is a subnet of $Z=(P, T, I, O)$ if $P' \in P$, $T' \in T$, and if there exist $I'(p, t)=I(p, t)$, $O'(p, t)=O(p, t)$ for any $p \in P'$ and $t \in T'$. It is also written as

$$Z \supset Z'.$$

Then the input place set of Z' is
$$P_{in} = \{p \in P': \exists t \in {}^{\cdot}p \ \& \ t \notin T'\}$$
the output place set of Z' is
$$P_{out} = \{p \in P': \exists t \in p^{\cdot} \ \& \ t \notin T'\}$$
the input transition set of Z' is
$$T_{in} = \{t \in T': \exists p \in {}^{\cdot}t \ \& \ p \notin P'\}$$
the output transition set of Z' is
$$T_{out} = \{t \in T': \exists p \in t^{\cdot} \ \& \ p \notin P'\}$$

Definition 2. Given $Z \supset Z'$, Z' is a place subnet/block iff $P_{in} \neq 0$, $P_{out} \neq 0$, $T_{in}=0$ and $T_{out}=0$; Z' is a transition subnet/block iff $P_{in}=0$, $P_{out}=0$, $T_{in} \neq 0$ and $T_{out} \neq 0$; Z' is a p-t subnet/block iff $P_{in}=0$, $P_{out} \neq 0$, $T_{in} \neq 0$ and $T_{out}=0$ and Z' is a t-p subnet/block iff $P_{in} \neq 0$, $P_{out}=0$, $T_{in}=0$ and $T_{out} \neq 0$.

Definition 3. A TPN Z'' is called an equivalent subnet of a marked graph Z', iff under the same markings, 1) the delay time from any input node to any output node in Z'' equals to that of Z'; and 2) the number of nodes in Z'' is less than that of Z'.

4.2 Calculations of Cycle Time

We are always interested in how fast each transition can initiate firing in a periodically operated TPN. As stated earlier, cycle time is the time to complete a firing sequence leading back to the starting marking after firing each transition at least once. On reducing deterministic Petri nets the delay time calculation is critical.

4.2.1 A simple circuit

The cycle time for the simple circuit as shown in Fig. 4.2, is

$$\pi = \frac{\Sigma \tau_i}{m_c}$$

where τ_i is the delay time of ith transitions associated to the circuit,

m_c is the total number of tokens in the cycle
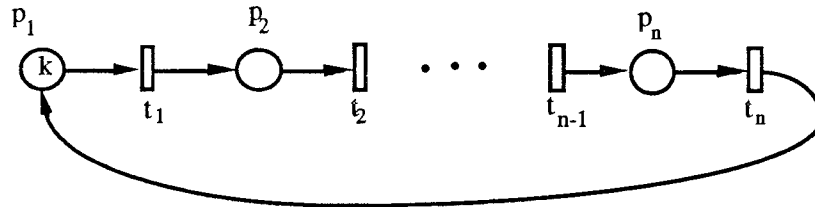
Fig. 4.1 A single circuit

One should notice that the cycle is associated with the number of tokens in the cycle as well as the total delay time of the cycle. In Fig. 4.1, transitions t_1, t_2, t_3, $\cdots$, t_n can be concurrent iff they are all enabled, and furthermore, transition t_i can pass any number of tokens simultaneously. Under such conditions, the formula above is obtained.

4.2.2 Marked graphs

In a marked graph, its cycle time is

$$\pi_c = \max(\pi_1, \pi_2, ..., \pi_n)$$

where π_i $(1 \leq i \leq n)$: the cycle time of ith circuit.

Furthermore, this formula can be applied to many other deterministic timed Petri nets to compute the deterministic time delay and search critical cycles which bound the system performance.
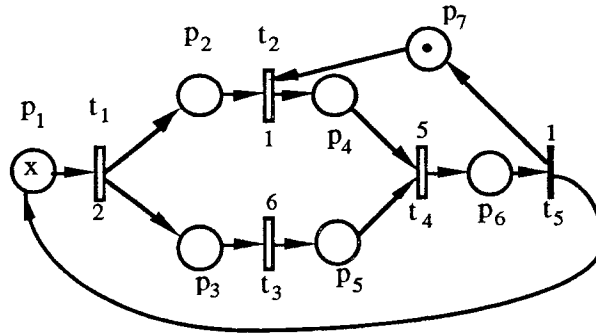
Fig. 4.2 A marked graph with parallel paths

There are three loops in Fig. 4.2, and

$$\pi = \max(\frac{14}{x}, \frac{9}{x}, 7)$$

Definitely, $\frac{14}{x} > \frac{9}{x}$ for x is a positive integer. If x=1, the cycle time π depends on the big loop $p_1 t_1 p_3 t_3 p_5 t_4 p_6 t_5 p_1$, i.e., π=14; and the idle time in loop p_7 is 14-7=7, i.e., the efficiency of control loop $p_7 t_2 p_4 t_4 p_6 t_5 p_7$ is 50%. But if x>2, π=7, which is independent of the big loop and the initial tokens in place p_1. When x=2, both delay times in these loops are 7. It means there is no waiting time either loop, and at this time the efficiency of the system and the utility is the highest. In conclusion, the assignment of tokens in resource p_1 and control place p_7 can affect the system optimum, and here $m(p_1)=2 \times m(p_7)$ is the best assignment of tokens for the net in Fig. 4.2.

The calculation of cycle time for marked graph is the foundation of deterministic timed Petri nets. In other words, the calculation of other types of Petri nets follows after their conversion to marked graphs.

4.2.3 Nets with multiple-weighted arcs

First of all, the net should be bounded, live and reversible. Here is a model of a circuit with multiple-weighted arcs, as shown in Fig. 4.3.
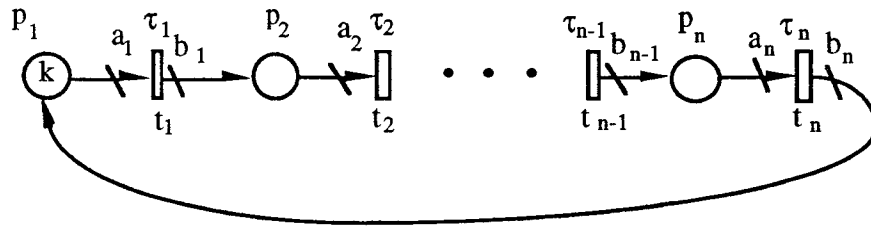
where a_j stands for the weight of input arc $I(p_j, t_j)$ and b_j, output arc $O(p_{j+1}, t_j)$ $(1 \leq j \leq n)$.

Fig. 4.3 A circuit with multiple-weighted arcs

There exist two conclusions:

1. $\Pi a_j = \Pi b_j$ $(j=1,2,..., n)$ [4]

2. $m_0(p_1) \geq \max(\prod_{j=1}^{k} \frac{a_j}{b_{j-1}}, \quad k=1, 2, \cdots, n)$

where $b_0 = 1$.

The first point guarantees that tokens in net are bounded during processing. The second is a necessary condition for the live net, but not sufficient. For instance, $n=4$ with $a_1=3$; $b_1=2$; $a_2=5$; $b_2=3$; $a_3=2$; $b_3=5$; $a_4=4$ and $b_4=4$, the value from the equation in the second condition is 7.5, actually the minimum tokens in p_1 without deadlock is 13. Only if the factor of a_j^i/b_{j-1}^i or b_{j-1}^i/a_j^i is a positive integer for applicable 1 and j in the net and all of

its reduced nets, Condition is necessary and sufficient for the liveness of the net. An example follows.

Fig. 4.4 introduces a reduced circuit of a multiple-weighted one with $n=4$. In the original circuit, there are 4 pairs of a_j and a_{j-1}, i.e., a_2 and b_1; a_3 and b_2; a_4 and b_3; and, a_1

and b4. Reduction can be done in any subnets of the original circuit, and here, a unit transition subnet is reduced in the example of Fig. 4.4. First, mark all nodes in the original circuit with the superscript of "0" which means 0-step reduction, and after reduction, mark the reduced circuit with 1 which means 1-step reduction. In this example, $a_2^1 = \max(a_2^0,$ $a_2^0 a_3^0 / b_2^0$) and $b_2^1 = \max(b_3^0, b_2^0 b_3^0 / a_3^0$). The ith reduction can be done only under the condition that the factor of $\frac{a_j}{b_{j-1}}$ or $\frac{b_{j-1}}{a_j}$ in $(i-1)^{th}$ circuit is an integer. Otherwise, the reduction cannot be done, which means there may be remaining tokens in the reducing subnet, then the equation in Condition 2 in page 36 is not sufficient for this case. We can see the original circuit can be reduced to a single-place-single-transition circuit with weights of input and output arcs equal to that in Condition 2 if in any step of reduction satisfies that the factor of a_j^i / b_{j-1}^i or b_{j-1}^i / a_j^i in i^{th} reduced circuit is an integer for all applicable i and j.
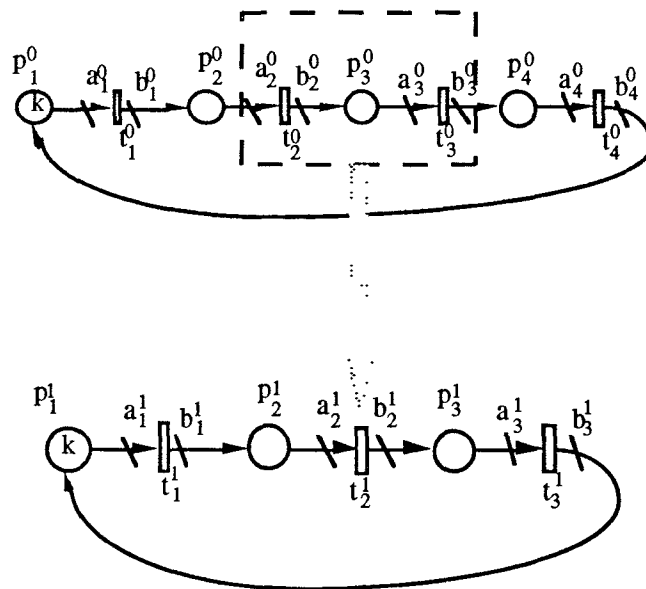
Fig. 4.4 A multiple-weighted circuit with n=4 and one of its reduced circuit

In most manufacturing processes, this limitation can be met. The following calculations are limited under this condition - there is no remaining token in places except p_1 in a firing loop.

Now, mark the transition t_i in Fig.4.3 with time delay τ_i $(1 \leq i \leq n)$, ~~respectively~~ and

denote $m = \max(\displaystyle\prod_{j=1}^{k} \frac{a_j}{b_{j-1}}$, $k=1, 2, \cdots, n)$.

The minimum number of active tokens in a circuit is an integer which comes from

the factor $\dfrac{m_0(p_1)}{m}$, denoted as m'. Then, the cycle time is

$$\pi_c = \text{(total delay in the circuit/minimum active tokens)}$$

$$= \frac{\Sigma \tau_i}{m'}$$

Fig. 4.5 is a model of multiple-weighted maked graph (WMG) with two circuits

$p_1 t_1 \cdots t_4 p_1$ and $p_5 t_2 p_3 t_3 p_6$. That the net is bounded, live and reversible is prerequestive to

the calculation. These two circuits should satisfied these three conditions:

1. $\Pi a_j = \Pi b_j$ $(j=1, 2, 3, 4)$

2. $m_0(p_1) \geq \max(\displaystyle\prod_{j=1}^{k} \frac{a_j}{b_{j-1}}$, $k=1, 2, 3, 4)$

3. Either $\dfrac{a_j}{b_{j-1}}$ or $\dfrac{b_{j-1}}{a_j}$ is a positive integer.
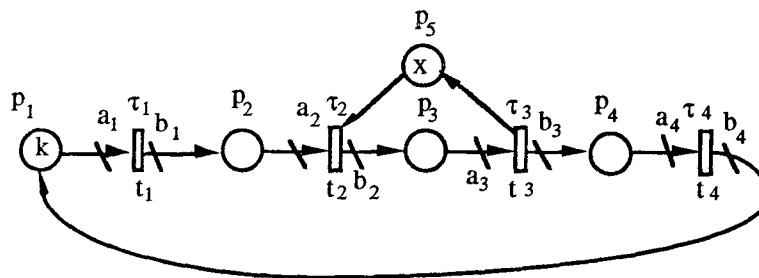
Fig. 4.5 A model of multiple-weighted marked graph

Now it is clear that $b_2 = a_3$ due to the loop $p_5 t_2 p_3 t_3 p_6$ and the net can be reduced to

that in Fig. 4.6. $m = \max(a_1, \dfrac{a_1 a_2}{b_1}, \dfrac{a_1 a_2 a_4}{b_1 b_3})$ and m' is the integer part of the factor $\dfrac{k}{m}$.
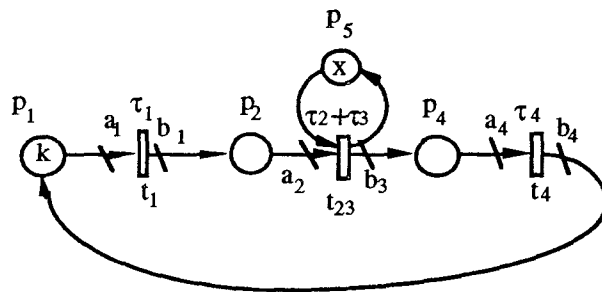
Fig. 4.6 One-step reduction of a WMG model

If $m=1$, it means that $a_1=1$, $b_1 \geq a_1 a_2 = a_2$ and $b_1 = \beta a_1$ (β is a natural number) due to the conditions and formula listed above. Then if $x \geq \beta$, $\pi_c = \max(\dfrac{\tau_1+\tau_2+\tau_3+\tau_4}{m'}, \dfrac{\tau_2+\tau_3}{x})$.

Otherwise, $x<\beta$, $\pi_c = \max(\dfrac{\tau_1+\tau_4+\frac{(\tau_2+\tau_3)\beta}{x}}{m'}, \dfrac{(\tau_2+\tau_3)\beta}{x})$.

If $m \geq 2$, then it is more difficult. From the formula associated to m, we can find that either a_1, $\dfrac{a_1 a_2}{b_1}$ or $\dfrac{a_1 a_2 a_4}{b_1 b_3}$ is equal to m.

First, $a_1=m$. In this case, $b_1 \geq \dfrac{a_1 a_2}{m} = a_2$ and we get $b_1 = \beta a_2$ (β is a natural number).

If $x \geq \beta$, $\pi_c = \max(\dfrac{\tau_1+\tau_2+\tau_3+\tau_4}{m'}, \dfrac{\tau_2+\tau_3}{x})$. Otherwise, $\pi_c = \max(\dfrac{\tau_1+\tau_4+\frac{(\tau_2+\tau_3)\beta}{x}}{m'}, \dfrac{(\tau_2+\tau_3)\beta}{x})$.

If $a_1 < m$ and $\dfrac{a_1 a_2}{b_1} = m$, it means that $\dfrac{a_2}{b_1} = \dfrac{m}{a_1} > 1$. we get that $a_2 = \beta b_1$ where β is a natural number and $\beta \geq 2$. Then the cycle time $\pi_c = \max(\dfrac{\tau_1+\tau_2+\tau_3+\tau_4}{m'}, \dfrac{\tau_2+\tau_3}{x})$.

The last case is that $a_1 < m$, $\dfrac{a_1 a_2}{b_1} < m$ and $\dfrac{a_1 a_2 a_4}{b_1 b_3} = m$. Similarly, we get $a_4 = \beta b_3$, i.e., the number of firing t_{23} is n times as much as that of t_4. We also get: if $x \geq \beta$, then the cycle time $\pi_c = \max(\dfrac{\tau_1+\tau_2+\tau_3+\tau_4}{m'}, \dfrac{\tau_2+\tau_3}{x})$ and if $x<\beta$, $\pi_c = \max(\dfrac{\tau_1+\tau_4+\frac{(\tau_2+\tau_3)\beta}{x}}{m'}, \dfrac{(\tau_2+\tau_3)\beta}{x})$.

4.3 Reduction of Timed Petri Nets

The purpose of reduction is to simplify the nets while preserving interesting properties and thus, save the time of computation and analysis.

4.3.1 Reduction approaches in marked graphs
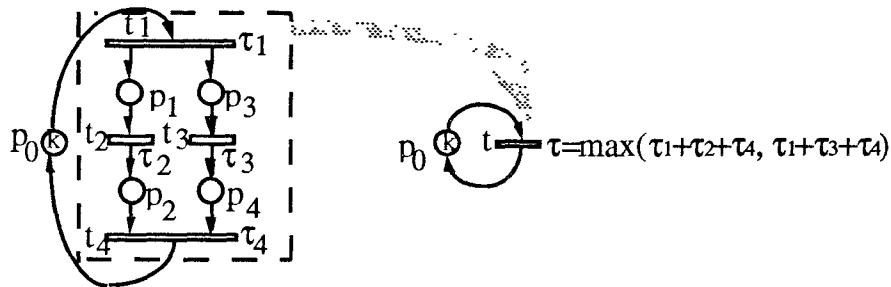
Fig. 4.7 Reduction of a net with concurrent paths

To illustrate the significance of reduction of marked graphs, sequential combination of three subnets of Fig. 4.7 is evaluated using reduction as shown in Fig. 4.8. The reduction is separated into three steps. First, the upper concurrent subnet is reduced to a macro transition subnet, whose delay time is $max(\tau_1+\tau_2+\tau_4, \tau_1+\tau_3+\tau_4)$, according to the formula above. The following steps are in the same way. The reduction processes are illustrated in Fig. 4.8. In this case, the number of computation is decreased from 2^3 to $2 \cdot 3$. As a conclusion, in a concurrent sequence, the reduction methodology saves the number of computation from Πn_i to Σn_i, where n_i represents the number of concurrent paths in i^{th} order.

Another example shown in Fig. 4.9 is used to show the reduction of a subnet containing tokens. In this case, we need to keep the record of the maximum cycle time of the loops in the subnet, denoted by π_s, as the cycle time is the maximum value of total local

circuits, it saves the calculation space and time to keep the maximum cycle time instead of the whole loops in the subnet.

$$\pi_s(i) = \max(\text{total local cycle time of the } i^{th} \text{ subnet}, \pi_s(i-1))$$

$\tau' = \max(\tau_1 + \tau_2 + \tau_4, \tau_1 + \tau_3 + \tau_4)$

$\tau'' = \max(\tau' + \tau_5 + \tau_7, \tau' + \tau_6 + \tau_7)$

$\tau''' = \max(\tau'' + \tau_8 + \tau_{10}, \tau'' + \tau_9 + \tau_{10})$
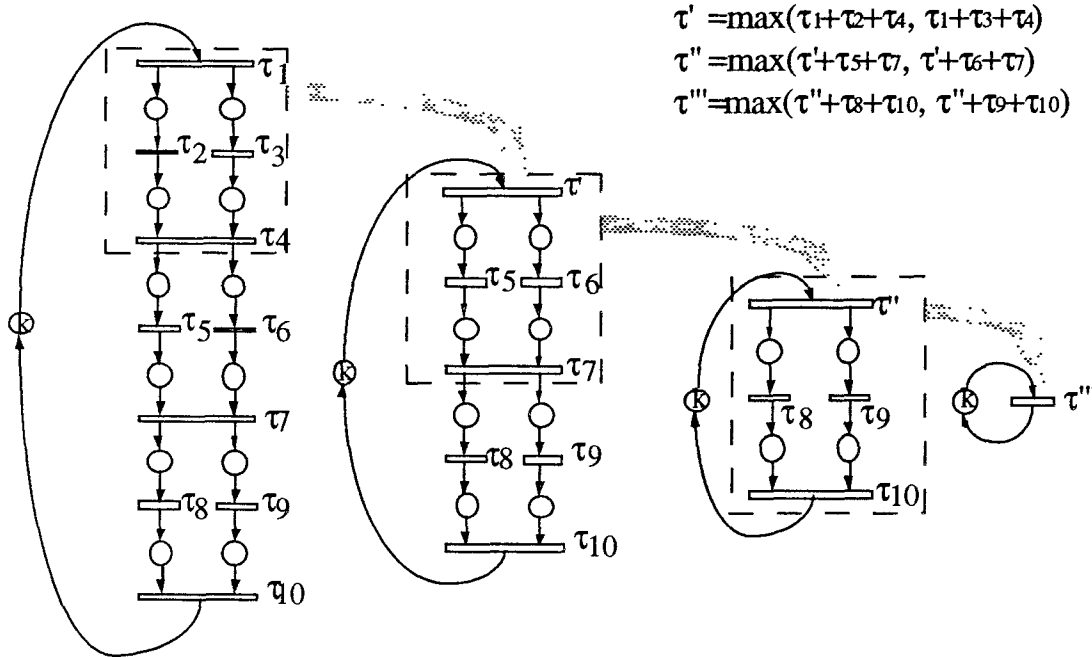
Fig. 4.8 Three-step reduction of subnets with concurrent paths

$\tau' = \max(\tau_4 + \tau_5 + \tau_7, \tau_4 + \tau_6 + \tau_7)$

$\pi_{s1} = \max(\tau_4 + \tau_5, \tau_5 + \tau_7, \tau_4 + \tau_6, \tau_6 + \tau_7)$

$\alpha = \tau' - \tau_4$

$\tau'' = \max(\tau_1 + \tau_2 + \tau', \tau_1 + \tau_3 + \tau')$

$\pi_{s2} = \max(\tau_1 + \tau_2, \tau_2 + \tau' - \alpha, \tau_1 + \tau_3, \tau_3 + \tau' - \alpha, \pi_{s2})$
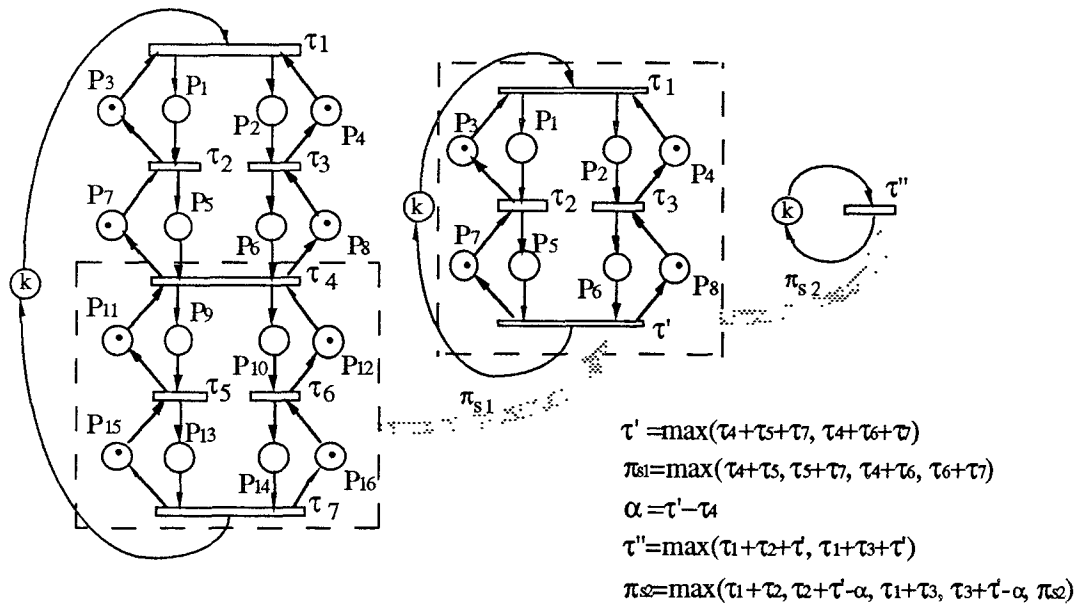
Fig. 4.9 Reduction of a net with loops in subnets

In addition, we need to amend the equivalent net. Subnet reduction affects the circuits involving the incoming and/or outgoing arcs, as it changes the time delay information on the *macro node*, which represents the reduced subnet. The transitions t' and t" in the example are marco transitions. An *incoming arc* is an arc which goes to the output node in the subnet, and an *outgoing arc* comes from the input node in the subnet. In the example, arcs (p_7, t_4) and (p_8, t_4) are the outgoing arcs of the subnet t_4-t_7, and there are no incoming arcs. After reduction, they both move to the macro node t'. Now these arcs are marked with *artificial delay* α, and $\alpha=\tau_{new}$ -τ_{old}, where τ_{new} is the delay time of the macro node, τ_{old} is that of the input or output transition. Here, the incoming arcs have the same artificial delay as each other, so do the outgoing arcs. Then, the cycle time of a circuit is

$$\pi=\frac{\sum\tau_i\text{-}\alpha_i}{\sum m_i}$$

where α_i is the sum of all artifitial delays assotiated with i^{th} circuit.

The last example in Fig. 4.10 is used to illustrate the case that elementary paths contain tokens. An *elementary path* is a path from the input node to the output node in a subnet. In this case, we need to consider the token flow. Each of concurrent paths in a subnet may contain a number of tokens. Their minimum number should be added to all of the input places to the macro transition of the subnet, where it satisfies {p: p∈ $^\bullet$t$_{macro}$}. The other conditions remain as the preceding cases. Hence, the cycle time of ith circuit is

$$\pi=\frac{\sum\tau_i\text{-}\alpha_i}{\sum m_i\text{-}\mu_i}$$

where μ_i represents the sum of artificial tokens in the ith loop.

$\tau' = \max(\tau_4 + \tau_5 + \tau_7, \ \tau_4 + \tau_6 + \tau_7)$

$\pi_{s1} = \max(\tau_4 + \tau_5, \ \tau_5 + \tau_7, \ \tau_4 + \tau_6, \ \tau_6 + \tau_7)$

$\alpha = \tau' - \tau_4$

$\tau'' = \max(\tau_1 + \tau_2 + \tau', \ \tau_1 + \tau_3 + \tau')$

$\pi_{s2} = \max(\tau_1 + \tau_2, \ \tau_2 + \tau' - \alpha, \ \tau_1 + \tau_3, \ \tau_3 + \tau' - \alpha, \ \pi_{s1})$
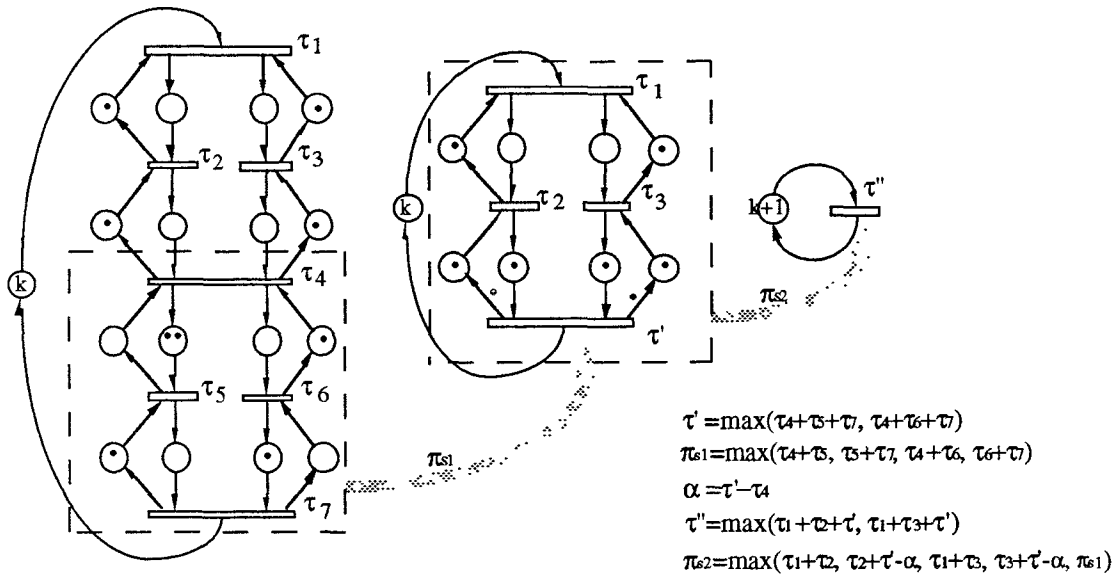
Fig. 4.10 Reduction of subnets with tokens in elementary paths

4.3.2 Reduction example for an FMS cell

The reduction method is applied to the model of an FMS cell which is operating in the factory floor of Information Technology Center, New Jersey Institute of Technology.
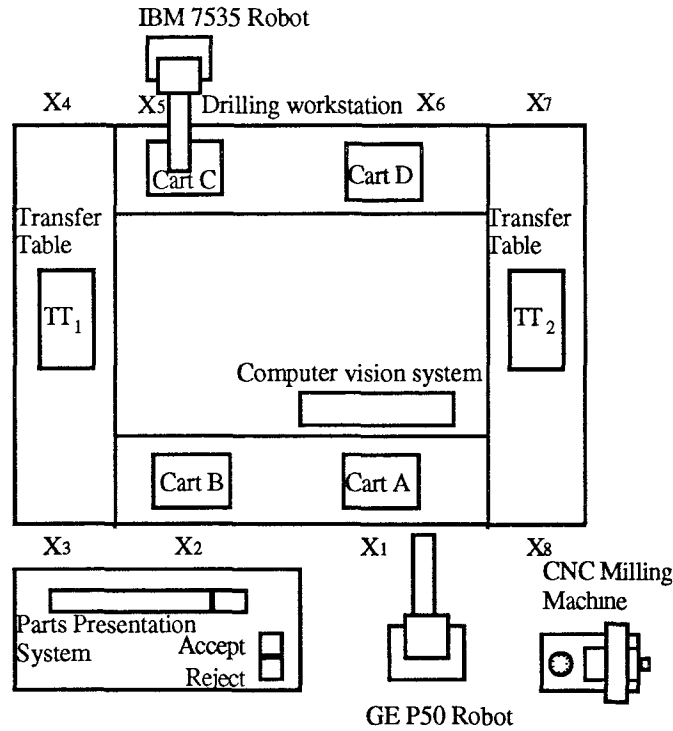
Fig. 4.11 Flexible manufacturing system cell [42]

It is observed that all four carts will move from their original positions and then back to the same one during a complete cycle. Hence the conveyor system is repetitive. Since every cart can move towards only one direction and transfer tables carry a cart from one to the other and go back once it is done, the system is decision-free. The basic operations with carts and transfer tables are moving and waiting. Positions are important and can be modeled as single resources since every position cannot be occupied by two carts at any time. Four carts will be modeled then as four tokens since they are moving from a position to another [42].

Based on the above analysis, a model of Petri net can be set up, as each moving between two positions as a transition, and each waiting as a place. Picture transitions t_1-t_{10} and places px_1-px_8. Then the availability of the position x_1-x_8 is modeled as places py_1-py_8. Arcs represent the operation flows and the work-situations, referred to Fig. 4.12(a) and Table 4.1 [42].

Table 4.1 Description and time table for FMS cell [42]

Node	Description	τ_i	Node	Description	τ_i
px_1	Cart available at loading station at X_1	74	py_1	Availability of cart halting at X_1	0
px_2	Cart halting position at X_2	5	py_2	Availability of cart halting at X_2	0
px_3	Cart loaded on TT_1 at X_3	2	py_3	Availability of TT_1 at X_3	0
px_4	Cart transported to X_4 by TT_1	3	py_4	Availability of TT_1 at X_4	0
px_5	Cart available at drilling-station at X_5	73	py_5	Availability of cart halting at X_5	0
px_6	Cart halting position at X_6	75	py_6	Availability of cart halting at X_6	0
px_7	Cart loaded on TT_2 at X_7	2	py_7	Availability of TT_2 at X_7	0
px_8	Cart transported to X_8 by TT_2	68	py_8	Availability of TT_2 at X_8	0
t_1	Cart moves from X_1 to X_2	3	t_2	Cart loaded on TT_1	3
t_3	TT_1 & loaded cart move to X_4	3	t_4	Cart unloads from TT_1 to X_5	4
t_5	Cart moves from X_5 to X_6	3	t_6	Cart loaded on TT_2	3
t_7	TT_1 & loaded cart move to X_8	3	t_8	Cart unloads from TT_2 to X_1	5
t_9	TT_2 moves back from X_8 to X_7	3	t_{10}	TT_1 moves back from X_4 to X_3	3
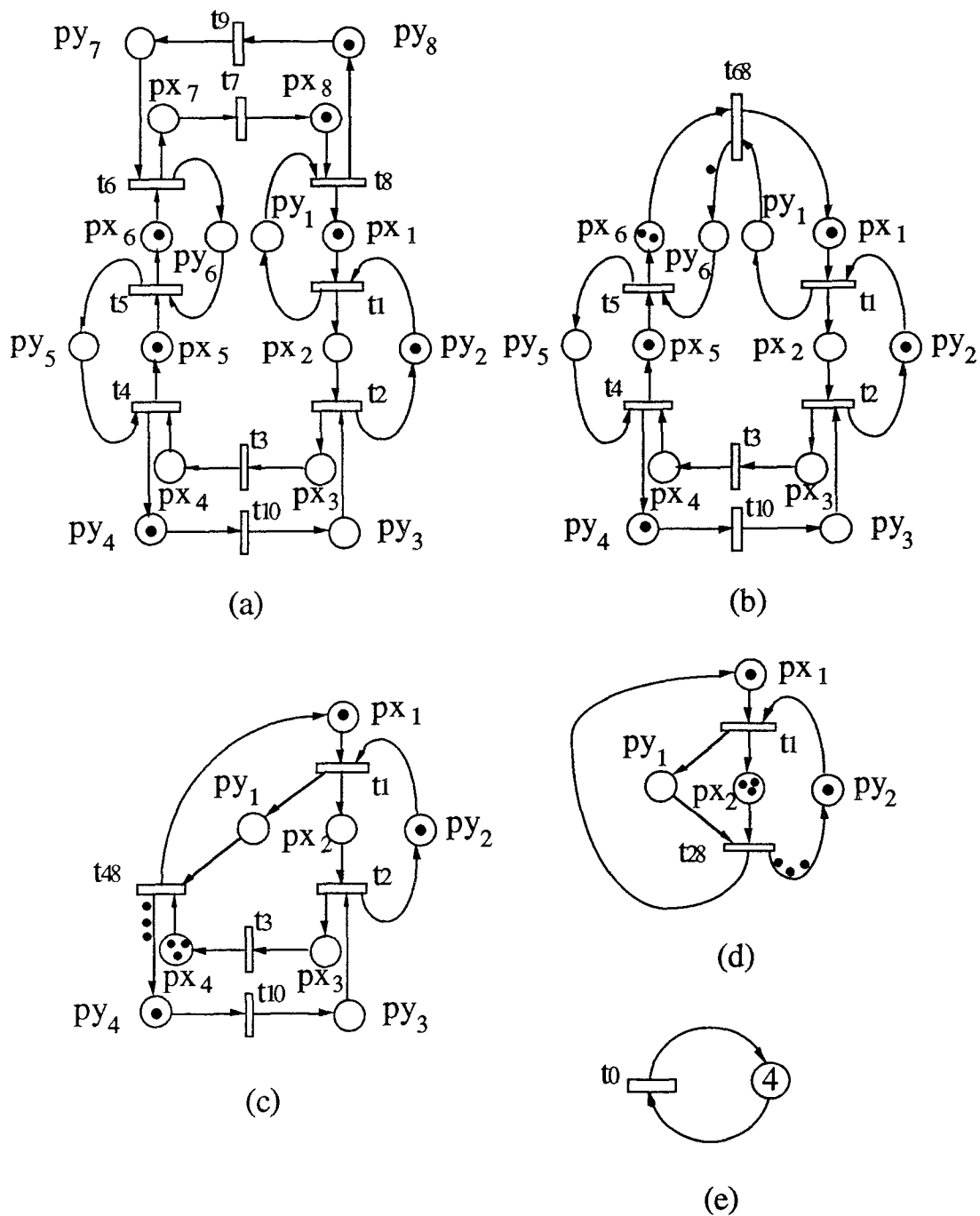
Fig. 4.12 The model of material handling system and the reduction processes

The first reduction is based on the transition subnet with px_7, t_6, t_7, px_8, t_8, py_7, t_9, and py_8. The transition t_6 is the input transition and t_8 the output transition. The elementary path is $t_6px_7t_7px_8t_8$ and there is a loop in the subnet. This subnet can be

reduced to a macro transition, denoted as t_{68}. First, the delay time of t_{68} is the sum of delays on the elementary path, i.e., $\tau(t_{68})=\tau(t_6)+\tau(px_7)+\tau(t_7)+\tau(px_8)+\tau(t_8)=81$. π_{s1}=(sum of delays in the loop/sum of tokens in the loop) $=\frac{82}{2}=42$. Then find the accessaries which are associated to the reduction. Here, we have an outgoing arc (t_6, py_6), an incoming arc (py_1, t_8), and one token on the elementary path in the subnet. Store the artificial delays on the outgoing and incoming arcs, $\alpha_{in1}=\tau_{68}-\tau_8 =76$ and $\alpha_{out1}=\tau_{68}-\tau_6=78$. Then add a token to the pre-place of t_6, px_6, $m(px_6)=1+1=2$ and mark the artificial token on the outgoing arc (py_1, t_8), $\mu_1=1$. Now the first step reduction is finished. The following steps are similar to the first step with considering the previous information, such as π_{i-1} and τ_{si-1} and applying the formular in page 43. They are listed in Table 4.2.

Table 4.2 Results of stepwise reduction for an FMS cell

Step	Subnet	Macro	τ_s	π_s	arc$_{out}$	α_{out}	μ	$\cdot t$	m($\cdot t$)	arc$_{in}$	α_{in}	Reference
1	t_6-t_8	t_{68}	81	42	(t_6, py_6)	78	1	px_6	2	(py_1, t_8)	76	Fig. 4.12(b)
2	t_5-t_{68}	t_{58}	159	81	(t_5, py_5)	156	2	px_5	3	(py_1, t_{68})	154	
3	t_4-t_{58}	t_{48}	238	82	(t_4, py_4)	234	3	px_4	3	(py_1, t_{58})	233	Fig. 4.12(b)
4	t_2-t_{48}	t_{28}	249	82	(t_2, py_2)	246	3	px_2	3	(py_1, t_{48})	244	Fig. 4.12(c)
5	t_1-t_{28}	t_{18}	255	82	(t_1, py_1)	254	3	px_1	4	(py_1, t_{28})	250	
6	---	t_0	329	82	–	--	–	–	4	–	–	Fig. 4.12(d)

$$\pi= \max(\frac{329}{4}, 82)= 82.25$$

We obtain the correct result. During the stepwise reduction, the amount of computation in this model is not released but the graph is simplified and can stop at any step if needed.

Reduction theory is the response to the problem of net complexity. This chapter discusses the condition of reductions and the methods of reduction. It is clear that we should keep

the right size of the subnets, as any big subnets may be quite complicated and difficult to evaluate the nets themselves. A marked graph net or subnet can be reduced to any degree required according to the formula developed in this chapter.

CHAPTER 5

MODELING AND ANALYSIS OF MCM PROCESSES

Petri nets, as a graphical and mathematical tool, shine significant efficiency in the modeling, analysis, simulation and control of manufacturing systems. Applications of Petri nets to the various fields in flexible manufacturing processes have resulted in great successes. It is highly desirable for the theoreticians, system analysts and practical operators to set up models associated to the multichip module fabricating processes. However, the application to such a sophiscated system is a big challenge.

5.1 Overview

In this particular work, an advanced MPU (microprocessor unit) is combined with a coprocessor in a multilayer ceramic PGA package using C-4 technique as shown in Fig. 5.1.

a cell

a wafer
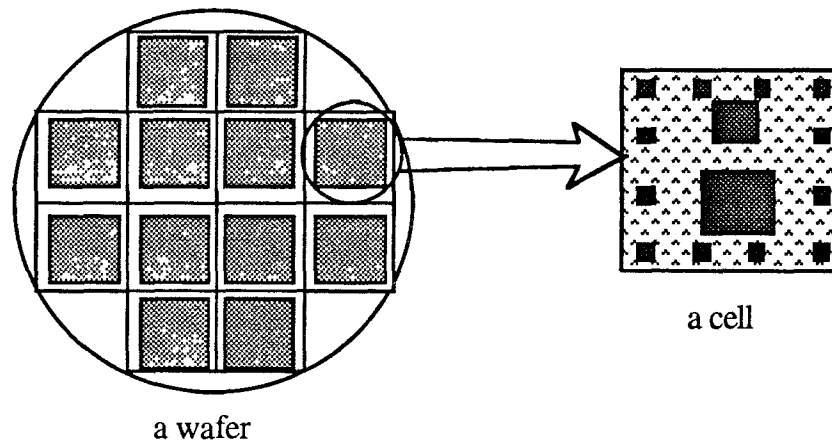
Fig. 5.1 An layout of a multichip module

48

As reviewed in Chapter 2, multichip module (MCM) technology is a fresh activity in semiconductor manufacturing. Since they eliminate a lot of repeating detailed constructions in the IC design and fabrication, they receive more and more attention by the IC manufacturers. Significant commercial contributions have been made by some large IC manufacturers.

MCMs mingle the PCB technology with VLSI technology. In other words, they import the mini-sized PCB to the IC fabrications. However, the fabrication processes of MCMs constitute a complicated concurrent manufacturing system, in which the circuited wafers and related chips are prepared and then assembled together, tested and packaged finally. With such a concept, we outline the MCM fabrication processes in Fig. 5.2 and Table 5.1.
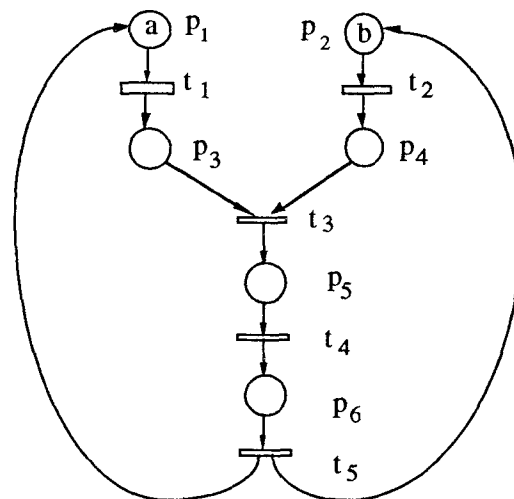
Fig. 5.2 Overview of MCM processes (1)

Table 5.1 Overview of MCM processes (1)

p_1: raw wafers available	p_2: raw chips available
p_3: circuit on wafers	p_4: prepare chips
p_5: mount chips on substrate	p_6: test, package, output

The model is a marked graph (MG), which is bounded, live and reversible. There are no conflict transitions in the net. The paths $p_1t_1p_3$ and $p_2t_2p_4$ form a parallel synchronization structure, i.e., t_3 is enabled only when there are tokens in both p_2 and p_4. If the delay times of t_1 and t_2 are the same, there is no idle time in either of those two places. Such a simple graph partially shows characteristics of MCM manufacturing processes.

5.2 System Description

The manufacturing processes of MCMs are much more complicated than that shown in Fig. 5.2, which is just an outline of the MCM concept. There is much work to do during the processes, such as a lot of different qualification tests in processes. Furthermore, the tests may get two results — qualified and unqualified. The qualified samples still keep in the process, and the unqualified are picked out of the process and either discarded or reworked. Now we can derive a model with more details introduced in Fig. 5.3 and Table 5.2. The net is still bounded, live and reversible, but it is not a marked graph yet.

5.3 Jobshops in Processes

The model obtained for MCM fabrication processes may be fine for system analysis, but it cannot be used in the practical control systems yet, as it is too abstract to understand. They are quite complicated processes, and roughly, they can be decomposited into four subnets: substrate process, chip-inspection process, C-4 mounting process and packaging and testing process. And the subnets may include several sub-subnets, in the light of specific conditions.
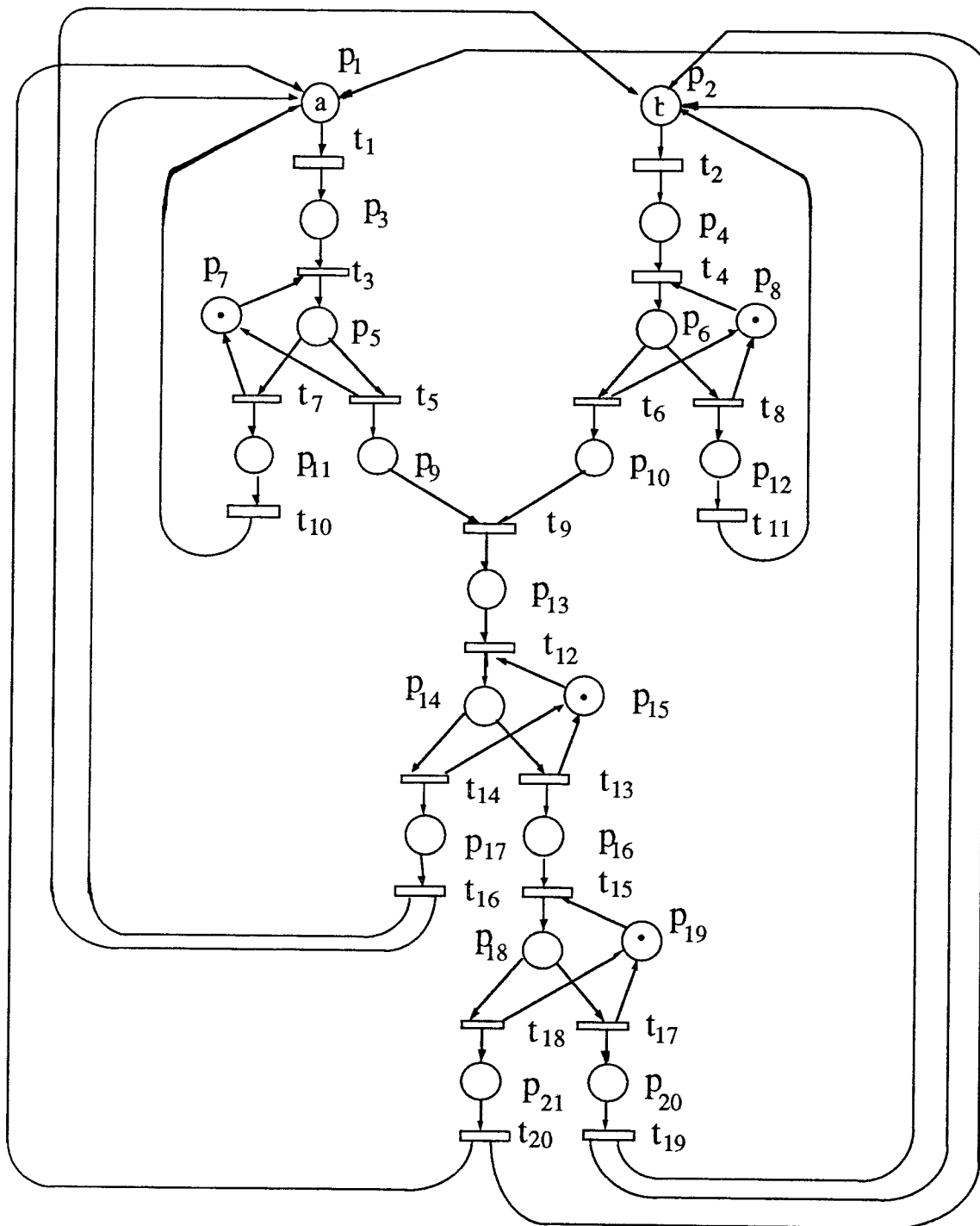
Fig. 5.3 Overview of MCM processing (2)

Table 5.2 Overview of MCM processes (2)

p_1: raw wafers available	p_2: raw chips available
p_3: circuit on wafers	p_4: prepare chips
p_5: test circuit substrate	p_6: test chips
p_7: test device available	p_8: test device available
p_9: accept the qualified	p_{10}: accept the qualified
p_{11}: discard the unqualified	p_{12}: discard the unqualified
p_{13}: mount ICs on substrate	p_{14}: test samples
p_{15}: test device available	p_{16}: package the qualified
p_{17}: discard the unqualified	p_{18}: test product
p_{19}: test device available	p_{20}: output the qualified
p_{21}: discard the unqualified	

Chapter 2 discussed the processes of MCM fabrication. Here are their illustrations with Petri nets. First, let us look back at Fig. 2.2 again. A raw wafer is supplied to the system, then the circuit and ground layers are coated onto the wafer several times, as it is a multilayer unit. This process is done in the sealed container. A robot loads or unloads a part between the operations shown in Fig. 5.4 and Table 5.3.

This subnet is a model with a sequential mutual exclusion, i.e., $(p_{a20}, (t_{a1}, t_{a2}) \cup (t_{a3}, t_{a4}) \cup (t_{a5}, t_{a6}) \cup (t_{a7}, t_{a8}) \cup (t_{a9}, t_{a10}) \cup (t_{a11}, t_{a12}) \cup (t_{a13}, t_{a14}))$. Three distinguishing features are observed:

1) No place can hold two or more tokens except place p_{a1}

2) No transition can fire concurrently.

3) If the next station is not available, the part would not be carried by the robot.

The first point emphasizes the capacity of places in the net — one token in a place at most, except p_{a1}, and the second point expresses that transitions fire asynchronously in the net, and the last, guarantees the liveness of the net. These features are very important.
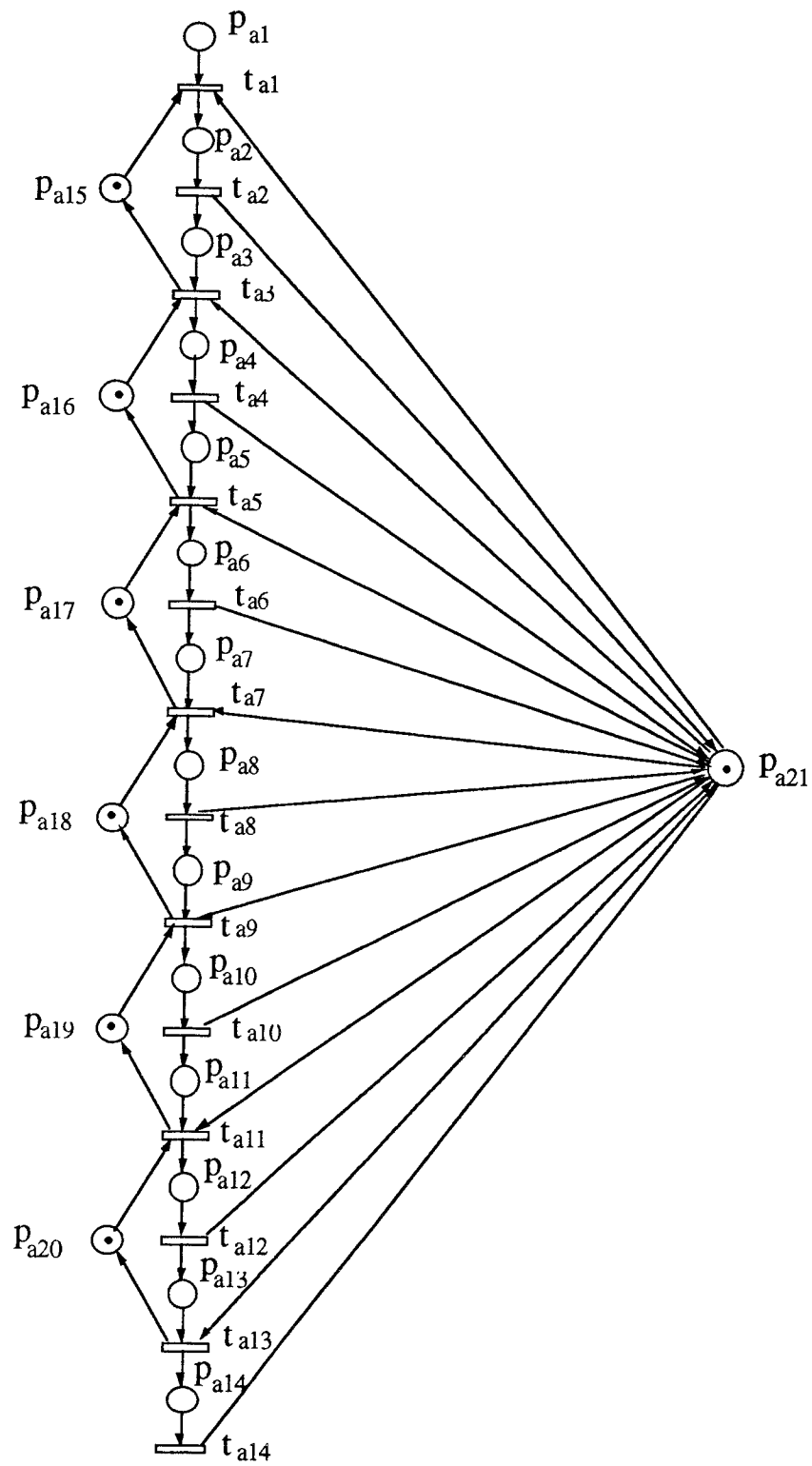
Fig. 5.4 Substrate process

Table 5.3 Substrate process

p_{a1}: raw wafer available	p_{a2}: load by robot
p_{a3}: preliminary clean	p_{a4}: unload/load by robot
p_{a5}: plant CMOS elements	p_{a6}: unload/load by robot
p_{a7}: ground & circuit 1	p_{a8}: unload/load by robot
p_{a9}: ground & circuit 2	p_{a10}: unload/load by robot
p_{a11}: ground & circuit 3	p_{a12}: unload/load by robot
p_{a13}: drill & coat solder	p_{a14}: unload by robot
p_{a15}: workstation 1 available	p_{a16}: workstation 2 available
p_{a17}: workstation 3 available	p_{a18}: workstation 4 available
p_{a19}: workstation 5 available	p_{a20}: workstation 6 available
p_{a21}: robot available	

The CMOS process (replacing p_{a5}) is shown in Fig. 5.5 and Table 5.4, and grounding and circuiting, p_{a7}, p_{a9} and p_{a11}, have almost the same process as that of Fig. 5.6 and Table 5.5. The difference is the delay times of transitions in the net. The place p_{a13} presents the drilling and coating solder process which is shown in Fig. 5.7 and Table 5.6. The jargon in the processes has been introduced in Chapter 2.

These graphs have the same function as that of flow charts, which express the process flows. Furthermore, they indicate the related stations, specific machines and other utilities in the nets.

Fig. 5.8 represents a Two-chip test process. Since there are two kinds of chips involved and their operations are presented by two parallel equally likely paths ($t_{b1}p_{b3}$ $t_{b3}p_{b5}t_{b5}p_{b9}t_{b8}p_{b13}t_{b12}p_{b15}t_{b14}$ and $t_{b1}p_{b2}t_{b2}p_{b4}t_{b4}p_{b8}t_{b6}p_{b12}t_{b10}p_{b16}t_{b14}$). An automatic pick-&-place machine is placed as a sharing resource which leads to a parallel mutual extension, (p_{b6}, (t_{b3}, $t_{b5})\cup(t_{b2}$, t_{b4})). As known, the ready-made chips are needed as much as twelve pieces
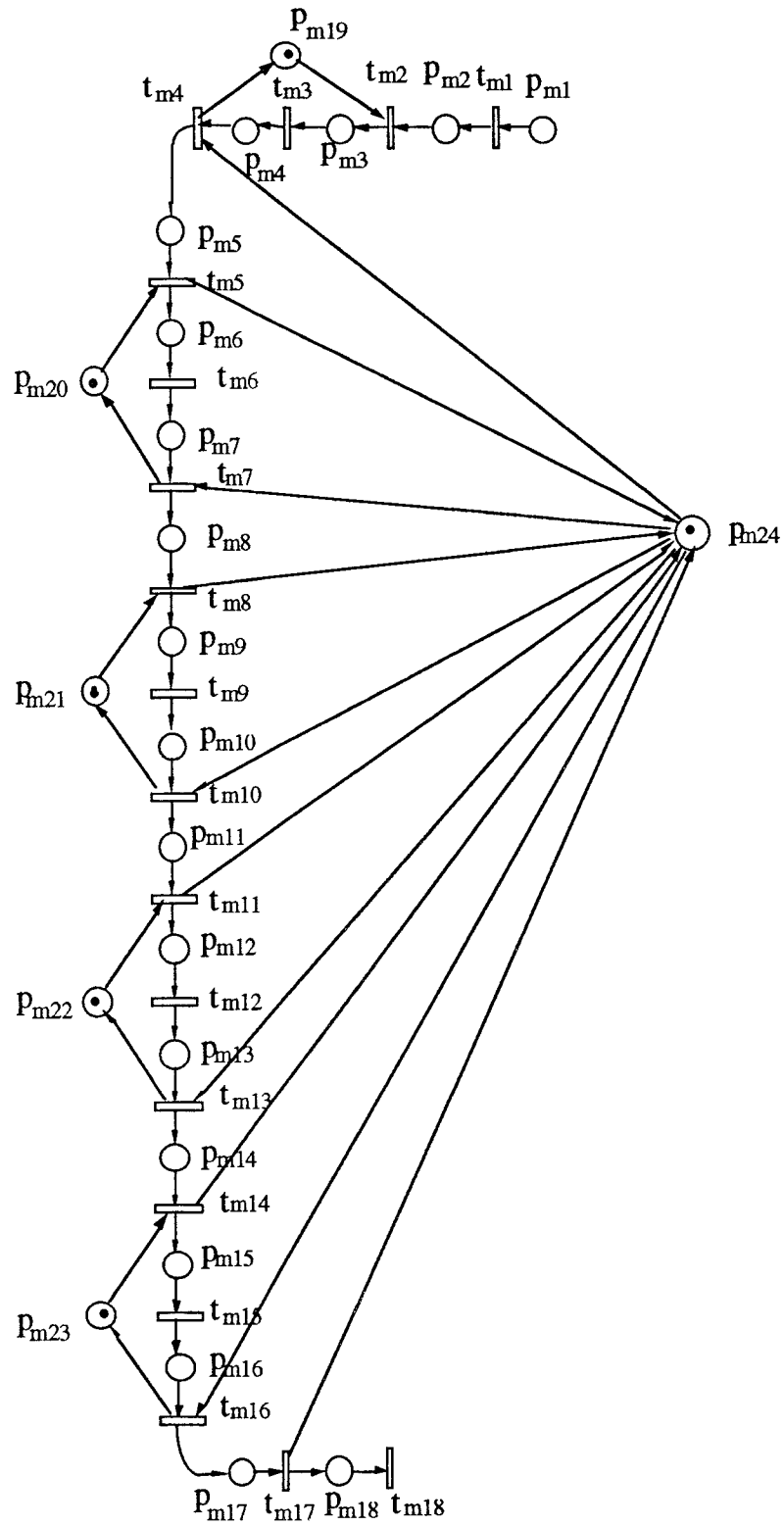
Fig. 5.5 CMOS process
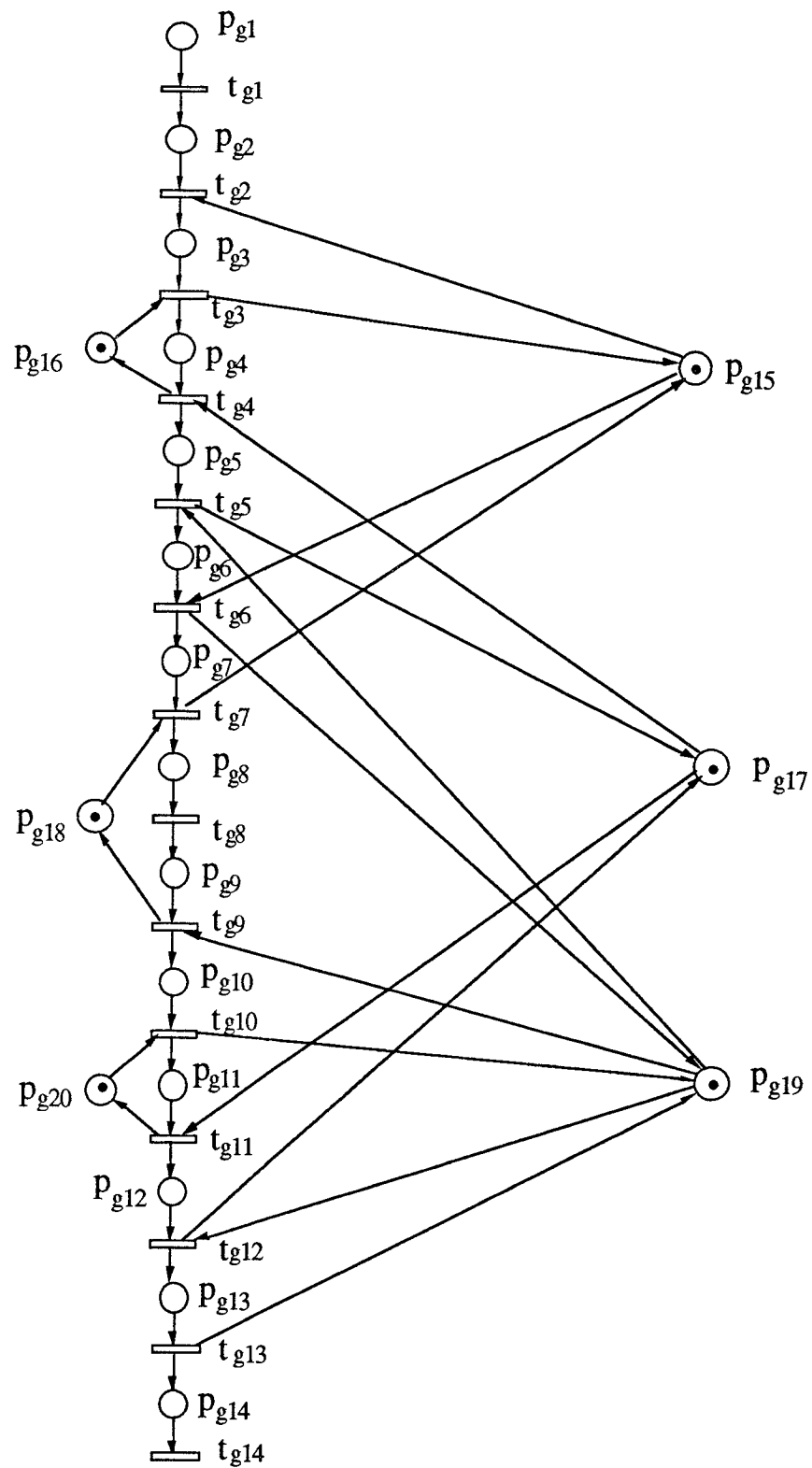
Fig. 5.6 Ground and circuit process

Table 5.4 CMOS process

P_{m1}: wafer available	P_{m2}: clean
P_{m3}: pattern p-well	P_{m4}: deposit/diffuse p-type impurities
P_{m5}: strip photoresist	P_{m6}: pattern dielectric
P_{m7}: etch dielectric	P_{m8}: strip photoresist
P_{m9}: pattern polysilicon	P_{m10}: etch polysilicon
P_{m11}: strip photoresist	P_{m12}: pattern p-channel & p+ guard rings
P_{m13}: p+ implant	P_{m14}: strip photoresist
P_{m15}: pattern n-channel & n+ guard rings	P_{m16}: n+ implant
P_{m17}: strip photoresist	P_{m18}: grow thin oxide
P_{m19}: workstation 1 available	P_{m20}: workstation 2 available
P_{m21}: workstation 3 available	P_{m22}: workstation 4 available
P_{m23}: workstation 5 available	P_{m24}: anti-resist available

Table 5.5 Ground and circuit process

P_{g1}: wafer available	P_{g2}: clean
P_{g3}: coat dielectric	P_{g4}: pattern ground
P_{g5}: coat metal	P_{g6}: strip photoresist
P_{g7}: coat dielectric	P_{g8}: pattern photo-via-holes
P_{g9}: etch dielectric	P_{g10}: strip photoresist
P_{g11}: pattern circuit board	P_{g12}: coat metal
P_{g13}: strip photoresist	P_{g14}: grow thin oxide
P_{g15}: dielectric available	P_{g16}: workstation 1 available
P_{g17}: metal available	P_{g18}: workstation 2 available
P_{g19}: anti-resist available	P_{g20}: workstation 3 available

per wafer each. Hence, it repeats testing at least 12 times each for a set, as there may exist the unqualified chips. After testing, if a chip is unqualified, it is discarded, and otherwise, accepted. The qualified chips are then stored in the buffer, P_{b19}. Table 5.7 is related to Fig. 5.8.
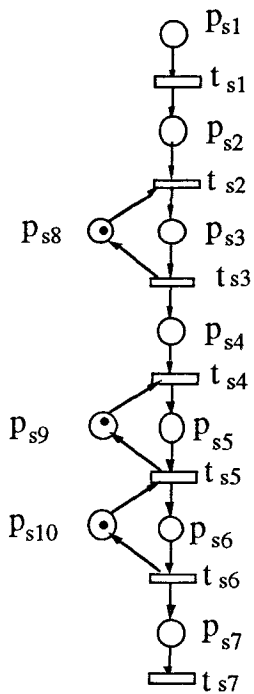
Fig. 5.7 Drill & coat solder process
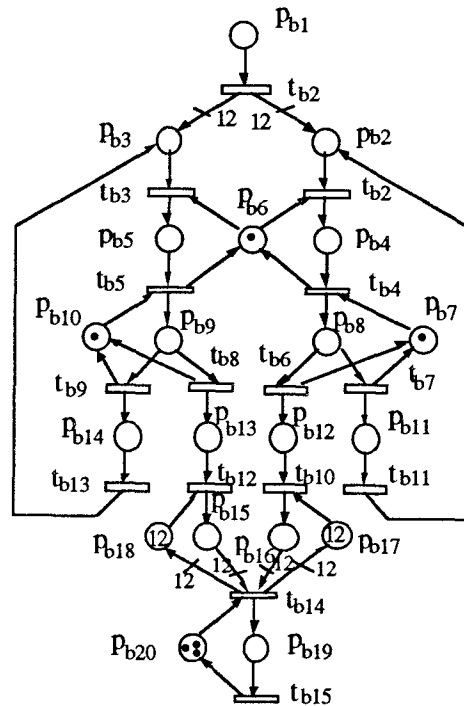
Fig. 5.8 Two-chip testing process

Table 5.6 Drill and coat solder process

p_{s1}: wafer available	p_{s2}: clean
p_{s3}: drill	p_{s4}: clean
p_{s5}: supply solder-aid	p_{s6}: coat solder
p_{s7}: clean	p_{s8}: drill available
p_{s9}: solder-aid available	p_{s10}: solder available

Table 5.7 Two-chip test process

p_{b1}: chips available	p_{b2}: chip B available
p_{b3}: chip A available	p_{b4}: load on test-table
p_{b5}: load on test-table	p_{b6}: p&p machine available
p_{b7}: test-machine B available	p_{b8}: test B
p_{b9}: test A	p_{b10}: test-machine A available
p_{b11}: discard B	p_{b12}: accept B
p_{b13}: accept A	p_{b14}: discard A
p_{b15}: unload A from testing	p_{b16}: unload B from testing
p_{b17}: 12 pieces a set of B	p_{b18}: 12 pieces a set of A
p_{b19}: store in buffer	p_{b20}: buffer available

C-4 packaging process is represented in Fig. 5.9 and Table 5.8, where pre-tested chips are mounted onto the circuited wafer. Among the operations, the surface mounting of components is one of the most important and critical operations in the process. An automatic pick-and-place machine picks chips from the stack, and then places them onto the corresponding positions. It is programmed that it placed all 12 A-chips, then does B-chip. This sub-process is simulated as Fig. 5.10 and Table 5.9, where p_{e17} presents the preparation of the circuited wafer, p_{e10} and p_{e11}, stand for the pick-and-place machine whose operations follow the instruction above. This is a concurrent process.
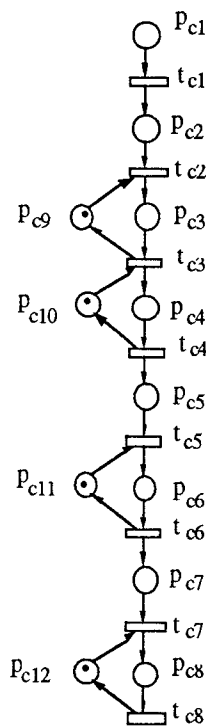
Fig. 5.9 C-4 process
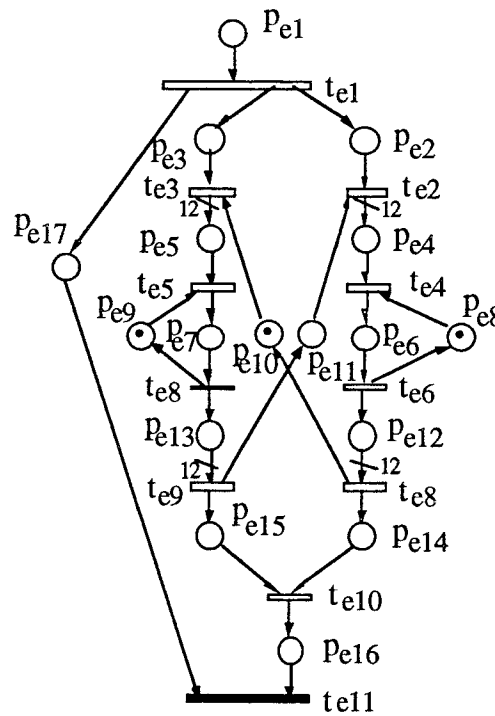
Fig. 5.10 Place chips on board process

Table 5.8 C-4 process

p_{c1}: materials are ready	p_{c2}: clean board
p_{c3}: place ICs on substrate	p_{c4}: solder reflow
p_{c5}: clean	p_{c6}: wave remind solder
p_{c7}: final clean	p_{c8}: coat glass
p_{c9}: p&p machine available	p_{c10}: infrared device available
p_{c11}: wave machine available	p_{c12}: SiO$_2$ available

Table 5.9 Place chips on board process

p_{e1}: materials available	p_{e2}: chip B gets ready
p_{e3}: chip A gets ready	p_{e4}: 12 pieces a unit
p_{e5}: 12 pieces a unit	p_{e6}: place chip B on board
p_{e7}: place chips on board	p_{e8}: robot available
p_{e9}: robot available	p_{e10}: p&p machine for chip A
p_{e11}: p&p machine for chip B	p_{e12}: finish chip B process
p_{e13}: finish chip A process	p_{e14}: chip B all on board
p_{e15}: chip A all on board	p_{e16}: finish mounting processes
p_{e17}: wafer gets ready	

Fig. 5.11 illustrates the final jobs in the semiconductor manufacturing system. There are two tests: electrical test and identification test. If both are passed, the products will be identified and output after packaged. This process was discussed in Section 2.10.
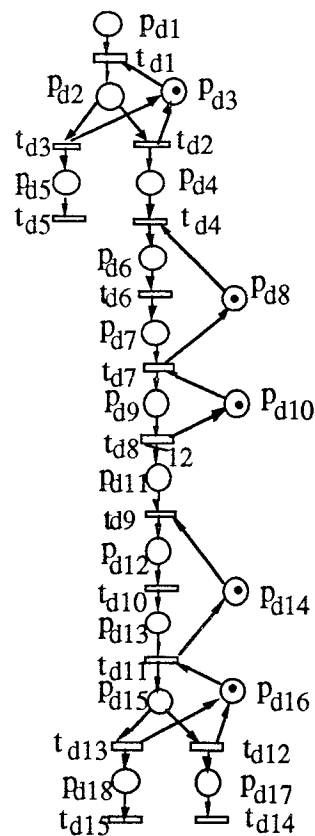
Fig. 5.11 Final package and test process

Table 5.10 Final package and test process	
p_{d1}: products available	p_{d2}: electrical test
p_{d3}: electrical test available	p_{d4}: accept the qualified
p_{d5}: discard the unqualified	p_{d6}: probe the wafer
p_{d7}: scribe on wafer	p_{d8}: prober available
p_{d9}: dice the wafer	p_{d10}: dice machine available
p_{d11}: split into 12 pieces	p_{d12}: wire bonding
p_{d13}: package the chip	p_{d14}: package machine available
p_{d15}: identification test	p_{d16}: identification test available
p_{d17}: output the identified	p_{d18}: discard the unqualified chips

5.4 Augmentation with Time

Time information is critical in quantitative analysis of event-driven systems. With the Petri nets excluding the time information, we can comprehend the whole processes of the MCM manufacture as above, but we cannot perform the desired performance analysis, unfortunately. Hence, augmenting with time information is very important and necessary for manufacturing systems. As addressed before, the TTPN (transition timed Petri net) is applied in this work. However, one faces some problems, for instance, $|\cdot t| \geq 2$. The time information on a transition always follows the preceding place in an elementary path, rather than alternate paths.

On enhancing the models of MCM fabrication systems with time information, keeping the original features is necessary. The time information is listed in Tables 5.11 – 5.20, which relate to Figs. 5.4 - 5.11.

Table 5.11 Time information for COMS process (Fig. 5.5)

t_{m1}	0.1	t_{m2}	1	t_{m3}	4.3	t_{m4}	7
t_{m5}	2	t_{m6}	4	t_{m7}	7	t_{m8}	2
t_{m9}	6	t_{m10}	7	t_{m11}	2	t_{m12}	8
t_{m13}	11	t_{m14}	2	t_{m15}	7	t_{m18}	11
t_{m17}	2	t_{m18}	2				

Table 5.12 Time information for ground & circuit level 1 process (Fig. 5.6)

$t1_{g1}$	0.1	$t1_{g2}$	1	$t1_{g3}$	5	$t1_{g4}$	10
$t1_{g5}$	9	$t1_{g6}$	2	$t1_{g7}$	5	$t1_{g8}$	7
$t1_{g9}$	14	$t1_{g10}$	2	$t1_{g11}$	15	$t1_{g12}$	10
$t1_{g13}$	2	$t1_{g14}$	2				

Table 5.13 Time information for ground & circuit level 2 process (Fig. 5.6)

$t2_{g1}$	0.1	$t2_{g2}$	1	$t2_{g3}$	5	$t2_{g4}$	14
$t2_{g5}$	9	$t2_{g6}$	2.4	$t2_{g7}$	5	$t2_{g8}$	11
$t2_{g9}$	14	$t2_{g10}$	2	$t2_{g11}$	19	$t2_{g12}$	10
$t2_{g13}$	2	$t2_{g14}$	2				

Table 5.14 Time information for ground & circuit level 3 process (Fig. 5.6)

$t3_{g1}$	0.1	$t3_{g2}$	1	$t3_{g3}$	4	$t3_{g4}$	6
$t3_{g5}$	9	$t3_{g6}$	2	$t3_{g7}$	8	$t3_{g8}$	7
$t3_{g9}$	16	$t3_{g10}$	2	$t3_{g11}$	12	$t3_{g12}$	10
$t3_{g13}$	2	$t3_{g14}$	3				

Table 5.15 Time information for drill and coat solder process (Fig. 5.7)

t_{s1}	0.1	t_{s2}	1	t_{s3}	11	t_{s4}	1
t_{s5}	10	t_{s6}	8	t_{s7}	2		

Table 5.16 Time information for place chips on board process (Fig. 5.10)

t_{e1}	0.5	t_{e2}	1	t_{e3}	1	t_{e4}	0.1
t_{e5}	0.1	t_{e6}	0.4	t_{e7}	0.4	t_{e8}	0.2
t_{e9}	0.2	t_{e10}	1.2	t_{e11}	0		

Table 5.17 Time information for substrate process (Fig. 5.4)

t_{a1}	0.5	t_{a2}	1	t_{a3}	4	t_{a4}	2
t_{a5}	π_{coms}	t_{a6}	2	t_{a7}	$\pi 1_g$	t_{a8}	2
t_{a9}	$\pi 2_g$	t_{a10}	2	t_{a11}	$\pi 3_g$	t_{a12}	2
t_{a13}	π_s	t_{a14}	1				

Table 5.18 Time information for two-chip test process (Fig. 5.8)

t_{b1}	0	t_{b2}	0.1	t_{b3}	0.1	t_{b4}	0.3
t_{b5}	0.3	t_{b6}	2	t_{b7}	2	t_{b8}	2
t_{b9}	2	t_{b10}	3	t_{b11}	1	t_{b12}	3
t_{b13}	1	t_{b14}	0	t_{b15}	1		

where t_{b6} and t_{b7} are mutually exclusive, $p(t_{b6})$=0.98 and $p(t_{b7})$=0.02

t_{b8} and t_{b9} are mutually exclusive, $p(t_{b8})$=0.98 and $p(t_{b9})$=0.02

Table 5.19 Time information for C-4 process (Fig. 5.9)

t_{c1}	0.1	t_{c2}	2.2	t_{c3}	t_p	t_{c4}	10
t_{c5}	2	t_{c6}	10	t_{c7}	4	t_{c8}	4.5

Table 5.20 Time information for final process (Fig. 5.11)

t_{d1}	0.1	t_{d2}	5	t_{d3}	5.5	t_{d4}	3
t_{d5}	1	t_{d6}	5	t_{d7}	3.5	t_{d8}	5
t_{d9}	0.1	t_{d10}	3	t_{d11}	3.5	t_{d12}	3
t_{d13}	3.5	t_{d14}	6	t_{d15}	1		

where t_{d2} and t_{d3} are mutually exclusive, $p(t_{p2})$=0.97 and $p(t_{p3})$=0.03

t_{d12} and t_{d13} are mutually exclusive, $p(t_{p12})$=0.99 and $p(t_{p13})$=0.01

5.5 Reduction Approaches

The model of MCM manufacturing systems is so big and complicated that reduction needs

for the analysis purpose. This reduction is done in several steps.

5.5.1 Models of substrate processes

Reduction procedure starts from the basic level of the processes. First, let us look back at Fig. 5.5, which is the model of CMOS process. According to the Feature 1 of the substrate model, there is at most one token which can enter the process at one time. Thus no transitions are in conflict, so none can be enabled simultaneously. Thus, it can be considered as a sequential graph, ignoring the loops in this subnet. Similarly, the subnets of ground-and-circuit, drill-and-coat processes are analyzed in the same way. The cycle time of these subnets, π_{cmos}, $\pi1_g$, $\pi2_g$, $\pi3_g$ and π_s, are listed in Table 5.21.

These subnets can be reduced to that in Fig. 5.12 and represent the marco nodes in the substrate processes — $p_{a5}t_{a5}$, $p_{a7}t_{a7}$, $p_{a9}t_{a9}$, $p_{a11}t_{a11}$ and $p_{a13}t_{a13}$, respectively.

Table 5.21 Delay time of subnets in the substrate process

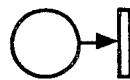| π_{cmos}	$\pi1_g$	$\pi2_g$	$\pi3_g$	π_s
83.4	84.1	96.5	82.1	33.1

Fig. 5.12 A p-t subnet

Now look back at the substrate model in Fig. 5.4. The time information is fixed by the constant values in Table 5.21. It is bounded, but not safe usually. It allows only one part in the workstation at most, and a robot to load and unload parts to or from the workstations once a time. If the following workstation is not available, the part cannot be loaded via the robot. Under such circumstances, we can draw an equivalent graph instead of that in Fig. 5.4. The equivalent graph is based on Fig. 5.4 without the utilities of the robot. Obviously, it is a marked graph and keeps the properties of the original graph. An assumption is made that parts may transfer between stations simultaneously, except the

contiguous ones. Originally, it is defined that the transference of parts are under the control of a robot, and it does once a time. But as seen, the delay time of load/unload by the robot is much less than that of working in the critical workstations, i.e., compared with the working time of the processes, the conflict waiting time is so short that it can be ignored. Under such circumstances, the equivalent simplification of the model is tenable. Now it is easy to calculate and reduce to a simple graph.

According to the reduction theory, the model of substrate processes can be reduced to a p-t subnet with the time-delay of $\Sigma\tau_i$=397.7 units, and critical cycle time in this subnet is π_s=100.5 units.

Test follows the substrate processes, shown in Fig. 5.13 with delay times of transitions, τ_1=1, τ_a=397.7, τ_2=2, τ_3=2.5, τ_4=5 and τ_5=2. The probability of the qualified is 97%, which flow through path $t_2p_5t_4$ and that of the unqualified is 3%, flowing through $t_3p_4t_5$ and then discarded.
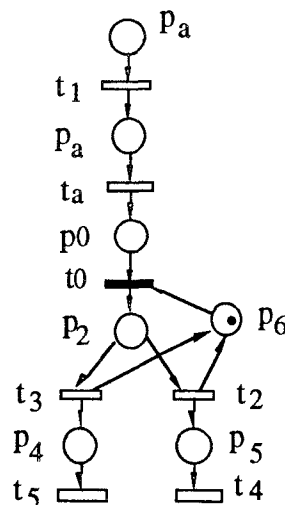

Fig. 5.13 Substrate fabrication and test process

Table 5.22 Results of substrate fabrication and test

τ_a(ideal)	probability	critical π(B)
405.7	97%	100.5

5.5.2 Models of 2-chip test process

2-chip test is modeled in Fig. 5.8, where an automatic pick-&place machine behaviors as a parallel mutual exclusion, i.e., $(p_{b6}, (t_{b3}, t_{b5}) \cup (t_{b2}, t_{b4}))$, and each chip testing process has an equally likely path, delay time distribution and probability distribution.

The path from p_{b3} to t_{b12} or t_{b13} is the process of testing chip A, and the number of tokens in the resource place is k, which is greater than or equal to 12. First discuss the path from t_{b3} to t_{b12}, compared with the loops in the subnet, $\pi'=\max(\pi_1, \pi_2, \pi_3)= \max(\frac{5.36}{k}, 2, 0.4)= 2.0$ units. Hence, the delay time in this sub-subnet is dominated by a critical loop in it. Again, the graph in Fig. 5.8 is a synchronous model with two equally likely paths, furthermore, the delay time of parallel mutually exclusive loops are shorter than π'. Such that, the delay time of the 2-chip test subnet can be considered to be equal to the value of delay time in a path, i.e., 2.0 units with 98% qualified.
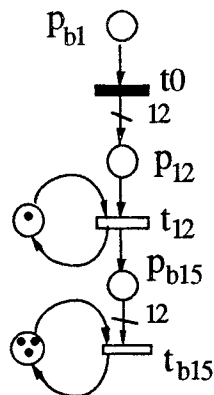
Fig. 5.14 Equivalent graph of two-chip testing process

In the chip-test process, the delay time is dominated by a critical loop, shown in Fig. 5.14. Followed the tool developed for multiple-weighted marked marked graph in the last chapter, the delay time of this process is:

Table 5.23 Result of 2-chip test processes

τ_b(ideal)	probability	π_s
25	98%	24

5.5.3 Models of C-4 Process

It is modeled as a marked graph, shown in Fig. 5.9. First, look at the process of placing chips onto the circuited substrate in Fig.5.10, which is a subprocess in the C-4 process. The path $t_{e5}p_{e7}t_{e7}$ can be reduced to a transition subnet as that of Fig. 5.15(b), and $\tau_{57}=\tau_{e5}+\tau_{e7}=0.1+0.4=0.5$. Furthermore, the path $t_{e3}p_{e5}t_{e5}p_{e7}t_{e7}p_{e13}t_{e9}$ is reduced to a transition subnet too, Fig. 5.15(c), and $\tau_{3-9}=\tau_{e3}+\tau_{57}\times12+\tau_{e9}=7.2$ units, referred to Section 4.2.3.
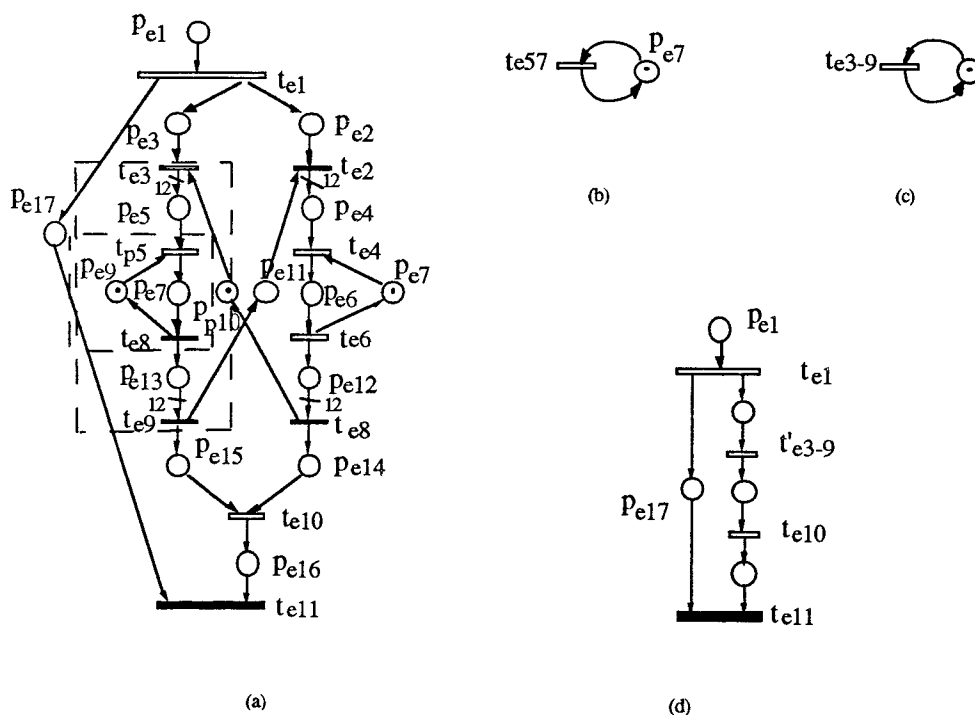

(a)

(b)

(c)

(d)

Fig. 5.15 Reduction process for Place-chip-on-board model

Similarly, the path $t_{e2}p_{e4}t_{e4}p_{e6}t_{e6}p_{e12}t_{e8}$ can be reduced. They are equally likely and parallel-concurrent, such that the cycle including these two paths can be reduced to a single transition with a self loop holding a token. Now the graph can be drawn in Fig. 5.15(d), a sequence process with the cycle time $\pi_e=\tau_{e1}+2\tau_{3-9}+\tau_{e11}=16.1$ units, and $\pi_s=2\tau_{3-9}=14.4$ units.

Then the delay time τ_c of C-4 subnet is $\Sigma\tau_i=\tau_{c1}+\tau_{c2}+\tau_{c3}+\tau_{c4}+\tau_{c5}+\tau_{c6}+\tau_{c7}=48.9$ units with $\pi_s=\tau_{c3}=\pi_e=16.1$ units.

5.5.4 Models of testing and packaging processes

In this net, there are two pairs of mutually conflicting transitions and a pair of multi-weighted arcs. Reduction can be done in four steps shown in Table 5.24, based on the reduction theory in Chapter 4.

Table 5.24 Stepwise reduction in final testing and packaging model

Step	Resources	Macro	τ_s	π_s
1	p_{d11}--t_{d14}	p_2-t_2	--	6.5
2	t_{d8}-p_2-t_2	t'_2	83	78
3	p_{d4}--t'_2	p_1-t_1	94.5	78
4	p_{d1}--t_1	p_d-t_d	99.6 with 96%	78

5.6 Overall Reduction

With the above results, the model of MCM processes can be fixed, shown in Fig. 5.16 and Table 5.25. As known, the delay time of t_b, t_c, and t_d are much smaller than that of ta,

even smaller than π_s in p_a subnet (100.5 units). Thus, the cycle time recorded in these subnets can be ignored. Hence, the cycle time π_{mcm} is

$$\pi_{mcm} = \max(\frac{\tau_a+\tau_c+\tau_d}{a}, \frac{\tau_b+\tau_c+\tau_d}{b}, \pi_{sa})$$
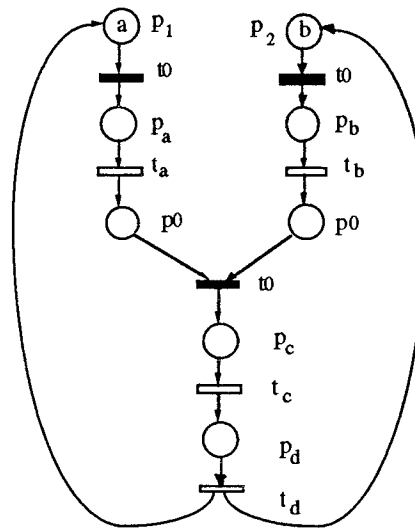
$$= \max(\frac{554.2}{a}, \frac{173.5}{b}, 100.5)$$

Fig. 5.16 Overview of MCM processes (3)

Table 5.25 Overview of MCM processes (3)

p_1: wafers available	p_0: idle place
p_2: chips available	t_0: immediate transition
p_a: substrate fabricate and test	t_a: operation on substrate
p_b: chips test and prepare	t_b: test chips
p_c: mount chips onto circuited substrate	t_c: C-4 process operation
p_d: final test and package	t_d: final operation

Now the final reduction is in Fig. 5.17 with τ_{mcm} listed in Table 5.26. As seen, it is a model of MISO (multiple-input-single-output) case. During these analyses, we start with an abstract point, then the detailed systems, and finally, back to the abstract point with the information of time. Furthermore, during the reduction, we also set up the time information for the different levels in the MCM manufacturing systems.

Table 5.26 Time table of π_{mcm}

a	b	π_{mcm}(ideal)	prob. of ideal
1	≥1	554.2	93%
2	≥1	227.1	93%
3	≥1	184.7	93%
≥4	1	173.5	94%
4	≥2	138.7	93%
5	≥2	110.8	93%
≥6	≥2	100.5	93%
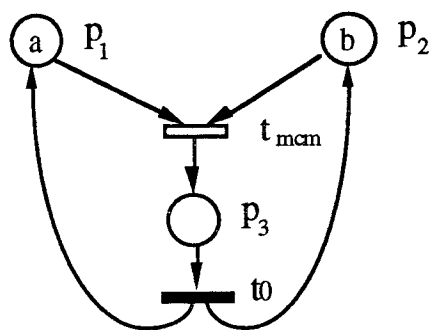
Fig. 5.17 An MISO model of MCM processes

This chapter expounds the whole procedure and results of MCM manufacturing processes. First, with top-down approaches, the model of MCM manufacturing systems is set up. Then, a clear view is given during the analysis and reduction the model with bottom-up approaches. This work using the reduction theory developed in Chapter 4 approaches to a MISO case. Being a complicated model, the MCM manufacturing systems are introduced and analyzed clearly using Petri nets.

CHAPTER 6

CONCLUSIONS

This chapter summarizes the contribution of the research and discusses the results obtained. Thereafter, some suggestions to enhance the work done are given and directed to further research.

6.1 Summary

The reduction theory for deterministic timed Petri nets is applied to multichip module (MCM) fabrication systems. The concepts, properties and methodologies of Petri nets have been expounded. Also reduction theory has been developed during the analyses.

The reduction theory for deterministic timed Petri nets developed in this thesis is based on marked graphs. A method for analysis of a class of timed multiple-weighted marked graphs is developed. The results are successfully applied to the analysis of an FMS cell and MCM processes.

Two basic Petri net techniques are applied, hybrid approaches and reduction approaches. The hybrid method is applied to set up the models of MCM manufacturing processes, from abstract to concrete. Then, the deterministic time information is incorporated into the model. It is a necessary step to build control systems.

71

Reduction technologies are applied to the analysis of complex systems. For the timed nets, calculation of their cycle time becomes very important and necessary in this process. The present methodology of delay time calculation is based on timed marked graphs. But the problem is that a complicated net may not be a marked graph. The most efficient way is to approximate the subnets with marked graphs.

This thesis shows that Petri nets are a powerful and handy tool for manufacturing systems. As a graphical tool, they make procedures easy to read and the relations among the operations clear. As a mathematical tool, they supply an opportunity to modify the models without any influence on the intrinsic properties. Time information is very important in the design of control systems, although modifying models with timing is much more difficult than that of ordinary models, especially in the reduction approaches.

6.2 Suggestions

Application of Petri nets to MCM fabrication systems is very important in their understanding and control.

The following suggestions are made for the future research:

First, MCMs are repairable modules. Hence, disposing of unqualified cells in manufacturing processes may get two results: repairable and unrepairable. Unrepairable cells are discarded, but the repairable should be reworked. In the further study, this phenomena should be modeled.

Second, stochastic time information can be augmented to Petri net models. During computer processing, the state-explosion problem may be faced, such that the model reduction is important and necessary. This is a hot topic in current studies.

Third, one can enhance control systems by programming on the basis of a Petri net model. This is an exiting avenue to validate the current work.

Finally, reduction methods need to be extended to more complicated Petri nets.

REFERENCES

[1] R. C. Aubusson and I. Catt, "Wafer-Scale Integration - a Fault-tolerant Procedure," IEEE Solid State Circuits, 13, pp. 339-344, 1978.

[2] John W. Balde, "Trends in Advanced Packaging Technology," Computer, 25, pp. 67-68, December 1992.

[3] C. J. Bartlett, J. M. Segelken, N. A. Teneketges, "Multichip Packaging Design for VLSI-Based Systems," IEEE Transactions on Components, Hybrids and Manufactures Technologies, 10, pp. 647-653, 1987.

[4] Daniel Y. Chao, M. Zhou and David T. Wang, "Multiple-Weighted Marked Graphs," Proceeding of 12th IFAC World Congress, Sydney, Australia, July 1993.

[5] Bernard C. Cole, "Wafer Scale Integration finally Goes Commercial (Wafer Stack)," Electronics, 62, pp. 72-73, December 1989.

[6] Clyde F. Coombs,Jr., Printed Circuits Handbook, 3rd Edition, New York, McGraw-Hill Book Company, 1988.

[7] D. Crockett, A. Desrochers, F. DiCesare and T. Ward, "Implementation of a Petri Net Controller for a Machine Workstation," IEEE international Conference on Robotics and Automation, pp. 1861-1867, March 1987.

[8] David J. Elliott, Integrated Circuit Fabrication Technology, New York, McGraw-Hill Book Company, 1982.

[9] R. L. Geiger, P. E. Allen and N. R. Strader, VLSI - Design Techniques for Analog and Digital Circuits, New York, McGraw-Hill Book Company, 1990.

[10] R. W. Harcourt, O. Zorba and A. G. Whitaker, "Competing Technologies of MCMs," Electronic Engineering, 64, pp. 57-58, February 1992.

[11] A. Ichikawa and K. Hiraishi, <u>Analysis and Control of Discrete Event Systems Represented by Petri Nets</u>, Springer-Verlag, pp. 115-134, 1991.

[12] R. R. Johnson, "Multichip Modules: Next-Generation Packages," IEEE Spectrum, 27, pp. 34-36, March 1990.

[13] R. W. Johnson, "Thin Film Multichip Hybrids; an Overview," Proceeding NEPCON West, pp. 655-682, March 1989.

[14] David B. Knorr, "Mechanical Aspects of Multichip Module Reliability," JOM, 44, pp. 29-35, July 1992.

[15] A. H. Kumar and R. R. Tummala, "The Past, Present and Future of Multilayer Ceramic Multichip Modules in Electronic Packaging," JOM, 44, pp. 10-14, July 1992.

[16] R. M. Lea, Editor, <u>Wafer Scale Integration</u>, second Edition, Amsterdam, North Holland, 1988.

[17] L. Liang, J. D. Wilson, N. Brathwaite, L. E. Mosley and D. Love, "High Performance VLSI through Package-Level Interconnects," proceeding to 39th Electronic Components Conference, pp. 518-523, 1989.

[18] Jerry Lyman, "Multichip Modules Aim at Next-Generation VLSI," Electronic Design, 37, pp. 33-34, March 1989.

[19] J. Ma and M. Zhou, "Performance Evaluation of Discrete Event Systems via Stepwise Reduction and App.roximation of Stochastic Petri Nets," IEEE International Conference on Decision and Control, Tuscon AZ, December 1992.

[20] J. Ma, M. Zhou and D. Y. Chao, "Applying Petri Net Approximation to Performance Analysis of Manufacture System with Nonexponential Transition Functions."

[21] N. MacDonald, G. Neish, A. Sinclair F. Baba, T. Tatematsu, K. Hirawa and K. Miyasaka, "200Mb Wafer Memory," Digest: 1989, IEEE International Solid-State Circuits Conference, pp. 240-241, 1989.

[22] W. Maly, <u>Atlas of IC Technologies: an Introduction to VLSI Processes</u>, Manlo Park, CA, Benjamin/Cummings Book Company, 1987.

[23] Jonah Mcleod, "Multichip Modules Vie for PC Business," Electronics, 65, pp. 10, November 9, 1992.

[24] J. R. Moyne and L. C. McAtee, Jr., "A Generic Cell Controller for Automated VLSI Manufacture Facility," IEEE Transactions on Semiconductor Manufacture, pp.77-81, May 1992.

[25] T. Murata, "Petri Nets: Properties, Analysis and Application," Proceeding IEEE, 77, pp. 541-580, April 1989.

[26] Joe Prang, "MCMs: the Reality of Today," Electronic Engineering, 64, pp. 33, January 1992.

[27] M. K. Premkumar, W. H. Hunt, R. R. Sawtell, "Aluminium Composite Materials for Multichip modules," JOM, 44, pp. 24-28, July 1992.

[28] P. J. Ramadge and W. M. Wonham, "The Control of Discrete Event Systems," Proceedings IEEE, 77, pp. 81-98, January 1989.

[29] Gail Robinson, "Multichip Modules - an Alternative to PCBs," Design News, 46, pp. 23-24, September 3 1990.

[30] W. S. Ruska, <u>Microelectronic Processing</u>, New York, McGraw-Hill Book Company, 1987.

[31] M. Silva and S. Velilla, "Programmable Logic Controllers and Petri Nets: a Comparative Study," IFAC Conference on Software for Computer Control, pp. 83-88, Pergamon Press, 1983.

[32] R. K. Spielberger, C. D. Huang, A. H. Mones, D. L. Tett and F. L. Hampton, "Silicon-on-Silicon Packaging," IEEE Transactions on Computer, 7, pp. 193-196, 1984.

[33] Tim Studt, "Chip Advances Mean Faster and Smaller IC Devices," R&D magazine, May 1991.

[34] I. Suzuki and T. Murata, "A Method for Stepwise Refinements and Abstractions of Petri Nets," Journal of Computer and System Science, 27, pp. 51-76, 1983.

[35] James N. Sweet, "Integrated Test Chips Improve IC Assembly," IEEE Circuits and Devices, September 1990.

[36] T. G. Tessier, I. Turlik, G. M. Adema, D. Sivan, E. K. Yung and M. J. Berry, "Process Considerations in Fabricating Thin Film Multichip Modules," International Electronic Packaging Symp., pp. 248-270, 1989.

[37] S. K. Tewkbury and L. A. Hornak, "Wafer Level System Integration: a Review," IEEE Circuits and Devices, pp. 22-30, September 1989.

[38] Y. Tsukada, S. Tsuchida and Y. Mashimoto, "Surface Laminar Circuit Packaging" IEEE Transactions on Solid State, 1992.

[39] Rao R. Tummala, "Multichip Packaging - a Tutorial," Proceedings of the IEEE, 80, pp. 1924-1941 December 1992.

[40] Samuel Weber, "For VLSI, Multichip Modules may Become the Packages of Choice," Electronics, 62, pp. 106-108, April 1989.

[41] M. Zhou, F DiCesare and A. A. Desrochers, "A Hybrid Methodology for Synthesis of Petri Net Models for Manufacturing Systems," IEEE Transactions on Robotics and Automation, 8, pp. 350-361, June 1992.

[42] M. Zhou and F. DiCesare, "Adaptive Design of Petri Net Controllers for Error Recovery in Automated Manufacture Systems," IEEE Transactions on Systems Manufactures Cybernestics, 19, pp. 963-973, 1989.

[43] M. Zhou, K. McDermott, P. A. Patel and T. Tang, "Construction of Petri Net Based Mathematical Models of An FMS Cell," Proceedings of 1991 IEEE International

Conference on Systems, Manufactures and Cybernetics, Charlottesville, Virginia, pp. 367-372, October 1991.

[44] --, notes from NJIT course, EE698-103, "Discrete Event Dynamic System," 1992.