UC Irvine
ICS Technical Reports

Title
Trouble management in DRAM fabrication

Permalink
https://escholarship.org/uc/item/8sh437vx

Author
Kim, Jeffrey Y.

Publication Date
1998-11-23

Peer reviewed

eScholarship.org Powered by the California Digital Library
University of California

SL BAR
Z
699
C3
no. 98-40

Trouble Management in DRAM Fabrication

Jeffrey Y. Kim
Information and Computer Science
University of California, Irvine
Irvine, CA 92697-3425
ykim@ics.uci.edu

Technical Report 98-40
23 November 1998

Notice: This Material may be protected by Copyright Law (Title 17 U.S.C.)

Abstract

This paper probes the fabrication process critical to DRAM (Dynamic Random Access Memory) chip productivity (the product yield), describes the technical structure and technical issues of the fabrication process, and explores the practice of DRAM trouble management dealing with the issues in situ. Analysis of the processes of DRAM trouble management indicates that the push for micro-miniaturization with extremely small physical tolerances creates a tension between two technical requirements: smaller chip size and higher chip robustness. The tension disturbs the balance among heterogeneous objectives of engineering groups. This suggests that the tension leads to a socio-technical resolution of engineering problems in DRAM fabrication. The study also questions existing assumptions about the control of engineering problem solving practice.

1. Introduction

The countries of the Pacific Rim, including the U.S., are world leaders in Dynamic Random Access Memory (DRAM) production, holding more than 90% of global share. Although not a huge market in sales (USD19.8 billion in 1997 (E.E.Times,1998)), DRAM is crucial because it is an essential input to the "information economy" based on computers and communications technologies. The DRAM sector leads the fabrication technology for the entire semiconductor industry. It is arguably the "stepping stone" to development of more sophisticated semiconductor and device production, suggesting that countries making DRAM will eventually be able to leverage the knowledge they have gained in that market to compete in higher value-added areas such as processors and application-specific integrated circuits (ASICs). Also, successful DRAM production, often represented as product yield (Leachman, 1994), leads to a potentially wide margin of profit, providing higher competitiveness in the market. Despite its competitive potential, only a few countries have been successful in their DRAM development effort. Among several countries that started the semiconductor industry in 1960s, only two countries, the U.S. and Japan, remain in the competition. The competitive scene is still changing. Newcomers, Korea and Taiwan, successfully entered the market in the 1990s.

It is difficult to initiate and sustain successful DRAM development efforts in this highly competitive environment. DRAM production requires mobilization of considerable resources including a substantial capital investment (the cost of a single DRAM production line can easily exceed USD 1 billion,) knowledge of chip design and process technology, and highly skilled labor. It also needs an infrastructure to mobilize these resources for a sustained period of time before making a profit. Rapid technical change (a new generation of DRAM product every three years) demands continuous updating of equipment and other factory facilities (often, complete replacement and redesign), leaving little time for recovery of the initial investment.

There have been several explanations of the DRAM phenomenon. Successful market entry by Japan and new entrants has been viewed as a result of integrated effort between industry and government (Jang, 1995; Okimoto, Sugano and Weinstein, 1984). This success has been interpreted as an outcome of a nation-specific innovation system creating and sustaining technology development effort (Dosi, 1982; Mowery and Oxley, 1995). Elements forming the national system of innovation (NSI) include government policy, industry structure and government-industry cooperation. Government policy initiates,

directs and coordinates industrial innovation (Okimoto, Sugano and Weinstein, 1984; Yum, 1988). In addition to government policy, industry structure was also found to shape a particular pattern of DRAM innovation effort in Korea (Yoon, 1989) and in Japan (Methe, 1992). Jang (1995) has identified government-industry cooperation as a critical determinant of DRAM success in Korea. These studies focused on conditions facilitating the innovation effort. They do not explain how the conditions are transformed into actual technology development. As exemplified by the weak presence of European countries, a simple collection of all the necessary elements does not guarantee successful industry development. There are other unexamined factors for successful DRAM production.

One such factor is the actual production experience. Engineers learn to control technical problems (known as defects) occurring during the design and production only through the actual production activity. Their newly acquired knowledge helps to lower the occurrence of defects and thus increase the product yield[1] from an initial 30 percent to a competitive 90-plus percent. Arrow (1962) showed that actual production experience creates new knowledge. The knowledge gained through the "learning-by-doing" (LBD) effect leads to productivity increase without additional capital or equipment investment. The effect of LBD was observed to generate endogenous growth of the semiconductor industry by Irwin and Klenow (1994). Irwin and Klenow emphasized the input/output relationship of production. They didn't examine the actual LBD process. This paper focuses on the LBD process as an essential feature of DRAM technology development.

The pursuit of micro-miniaturization is one of the primary LBD activities in DRAM production. Micro-miniaturization pushes the physical limits of technical artifacts and production processes at the quantum level. Working at this micro-level presents new challenges for design and production and creates problems regardless of legitimate designs and production technologies, even when executed within an acceptable margin of error. Moreover, managing the problem doesn't always guarantee elimination of the cause. This level of extreme precision presents problems which may recur at any stage of the design and production. Trouble management -- the ongoing effort to deal with problems arising from extreme precision demanded in their operations -- is a unique characteristic of micro-miniaturization in DRAM production and poses a unique challenge in complex design and production at the frontier of physical limits. In particular, this trouble management raises

[1] Often, the degree of success in DRAM production is represented in "*product yield*". The term *yield* refers to a factor evaluating the performance of the DRAM production activities. The yield is defined as a ratio of working output to the total input (Leachman, 1994).

questions about the traditional, hierarchically organized control of design and production.

This paper presents the issue of DRAM trouble management within a frame illustrated in Figure 1.1. The figure summarizes how the trouble management emerges as the focus through multiple layers of the analysis. The question at the NSI level leads the study to a notion of High Precision Organizations (HPOs) and their technology development effort. Analysis at the HPO level then furthers its focus on the case of DRAM production. Trouble management is positioned as an essential part of successful DRAM production. Within this framework, this paper will probe the fabrication process critical to DRAM productivity (the product yield), describe the technical structure and technical issues of the fabrication process, and explore the practice of DRAM trouble management dealing with the issues in situ.

2. Overview of DRAM Fabrication

2.1 Technical Requirements And The Resulting Tensions

It costs well over 3 billion USD to start a new DRAM production facility. Rapid technical changes in the DRAM industry make it imperative to reach competitive product yield quickly enough to recover the initial investment and make a profit within the first few years of new investment. This targets the goal of DRAM production at producing as many DRAM chips as possible within a given amount of time. The goal is economically meaningful and critical since the initial investment for the facility is already fixed. Decreasing the DRAM unit cost satisfies the DRAM production goal.

One common approach for high product yield is to make the DRAM chip as small as possible. This increases the number of chips produced per silicon wafer. However, the microscopic scale of DRAM chips poses a major challenge to increasing the product yield. DRAM chips are vulnerable to dust particles or small contaminants that may cause a circuit failure. Moreover, the scale of DRAM chips submit the elements of a DRAM circuit to the laws of quantum mechanics. Quantum effects occur in such a way that changes to chip size don't produce the same desired effects as they do at larger scale. The chip size is often represented as minimum feature size, a minimum size opening in photomasking. This minimum feature size requirement affects various dimensions of circuit regions: contact size, line width, distance between line and contact, distance between two interconnection lines, and channel size. Over the past 30 years, DRAM manufacturers have successively

decreased the minimum feature size. However, smaller minimum feature size makes the DRAM circuit more sensitive to contaminants and other process variances. Smaller transistor cells may not be able to hold sufficient electric charge to meet DRAM product specifications. Interconnection wires become thinner, making them more vulnerable to process and operation variations. The demand for smaller minimum feature size leads to a tension between two technical requirements: chip size and chip robustness. This tension permits little margin of error during the DRAM design and production process. Reducing the minimum feature size of the DRAM circuit is a continuous process to manage the inevitable tension created by the conflicting requirements of chip size and chip robustness.

2.2 The Fabrication Process

The DRAM fabrication process requires the circuit design, photomasks, process equipment, and various materials. The materials needed include EMC (resin - basic material for silicon wafers), silicon wafers, lead frame, photo resist, and specialty gases.

The fabrication process transfers a geometric pattern of the circuit layout onto each die on a silicon wafer. The fabrication process will produce a DRAM integrated circuit (or, die) that consist of millions of identical three-dimensional transistor circuit cells and metal wires interconnecting these cells. Currently, CMOS (Complementary Metal Oxide Semiconductor) transistor circuit cells are widely used because of their inherent low power characteristics, comparable speed and better reliability. CMOS transistor cells are made of p-channel and n-channel transistors and multiple insulating layers of silicon dioxide. Metal wires interconnect transistor cells connected through contact holes. The fabrication process consists of three steps repeated in a cyclic fashion to build a three-dimensional DRAM circuit structure. The first fabrication step is to lay a thin film material over the wafer. Next step is to transfer a reduced image of circuit layout pattern onto the thin film applied on the wafer surface. The last step is to dope a selected region of the wafer with ions. The doping step changes conductivity or resistivity of doped region. Fabrication workflow configures a sequence of fabrication steps done to build a three-dimensional circuit structure.

This section describes three elements of DRAM fabrication process: layering, lithography, and doping.

2.2.1 Layering

The first step in circuit image transfer is to prepare a thin film material on the silicon wafer surface. This film serves as a base material to duplicate the geometric pattern of circuit layout during the lithography process. There are three types of layers that can be added to the silicon wafer surface: insulator, semiconductor, and conductor. The insulator layer does not conduct the electricity while the conductor layer carries electric current across the microcircuit. The semiconductor layer forms regions for transistor cells.

Four primary processes are used to produce different layers: thermal oxidation, chemical vapor deposition (CVD), evaporation and sputtering. Thermal oxidation, CVD, and sputtering generate the insulator layer. CVD is used to form the semiconductor. Evaporation or sputtering is used to generate the conductor layer. The following describes the four layering processes.

- Oxidation

Thermal oxidation is a widely used oxidation method, creating a thin film layer on the silicon wafer surface. The process involves growing oxide on a silicon wafer surface in a furnace filled with water vapor or steam atmosphere. The silicon exposed to water accelerates an oxidation process. The oxidation process will form a thin layer of silicon dioxide. The silicon dioxide does not conduct electricity, forming an insulating layer.

Temperature around 1,000 degree centigrade is needed to maintain active oxidation process. As high temperature can result in crystal damage of the silicon wafer, it is necessary to balance oxidation temperature and operation time. In addition to balancing temperature and operation time, it is necessary to maintain contamination-free environment. A contaminant introduced to the wafer surface during the process will produce an uneven oxidation growth. An uneven layer will result in circuit reliability problems. Therefore, cleanness and uniform thickness of the silicon dioxide are key factors used to evaluate the performance of the oxidation process.

- Chemical Vapor Deposition (CVD)

The Chemical Vapor Deposition process relies on a chemical reaction at the atomic level. The CVD device mixes chemical atoms in reaction chamber to release the desired material. This material forms a deposit on the wafer surface. Generally, the CVD method forms

either thin film insulator or semiconductor layers. Precise control of chemical composition and structure is a crucial factor for uniform thickness during the CVD process. A low pressure environment provides more precise control of the chemical compound concentration. In addition, lower concentration of the mixture compounds gives better control over the gas phase reaction (Sze, 1983). Therefore, low pressure and low temperature are preferred for better control of the process. As the density of chips on a wafer has increased, low pressure CVD (LPCVD) has become a common process in the semiconductor industry (Van Zant, 1984).

- Evaporation

Evaporation process heats and evaporates source metal into the gas state. The metal gas is absorbed onto the wafer surface, forming a thin conductor layer. The duration and temperature of evaporation determines thickness. The evaporation method carries out its steps in a very low pressure chamber. It is because aluminum gas combines with the oxygen in the air to form an insulator, breaking the purpose of the layering process. As in other layering methods, uniform layer thickness is a critical factor for the success of the process.

- Sputtering

Sputtering is a physical process that deposits dielectric or semiconductor layers on the wafer surface. This method is widely used during metal wiring (metallization) process. Sputtering accelerates and collides ionized atoms on a plate of target material in a vacuum chamber. Collisions dislodge some atoms of the target material from the surface of the plate. These dislodged atoms settle on the wafer surface as a thin film layer. Sputtering method is a physical process that can be controlled better than chemical processes. The sputtering process offers better control on film composition and layer thickness. As with other layering methods, uniform thickness and cleanness are key determinants for the performance of the sputtering process.

2.2.2 Lithography

A set of photomasks are used to build a three-dimensional circuit structure. Each photomask represents a part of circuit layout. Lithography prints an identical copy of geometric photomask pattern of the microcircuit onto the photoresist coated on the silicon

wafer surface. The lithographic involves a sequence of the operations (Fig.2.1). First, the silicon wafer is cleaned after it is processed from previous fabrication stages (e.g. diffusion, doping). The wafer is fed into an aligner for photo exposure. The photomask is used at this step. After the exposure is done under an intense light source, the photoresist coated on the wafer is developed. The silicon wafer is rinsed and dried to eliminate dust particles and other chemical residue from the wafer surface. The wafer is baked so that the selected regions (e.g. exposed regions in negative exposure) of photoresist are firmly attached to the surface of the wafer. After being cooled, the wafer is sent to the next stage (e.g. etching and diffusion). After etching and doping processes are done., the entire lithographic sequence may be repeated for copying different photomask pattern. The following explains each lithographic step in detail.

- Preparation of Lithographic Process

Application of the photoresist material is the first step in the lithographic process. Two components, photomask and photoresist, are prepared. The photomask carries a geometric pattern of the circuit layout. There are multiple photomasks which represent the fabrication workflow used to build the three-dimensional circuit structure. The opaque area of the photomask blocks light which then passes through clear regions of the photomask onto the photoresist. In addition to photomasks, photoresist is applied for each exposure step. Spinning is a widely used method for the photoresist application. The spinner dispenses an exact quantity of prefiltered photoresist onto the center of the wafer surface. It then spins the silicon wafer at high speed. This will spread the photoresist material uniformly across the surface of the wafer. Any excess amount of photoresist is spun off the edge of the wafer. After the coating of photoresist, the wafer is baked to remove all traces of solvent.

- Align, Exposure and Wafer Development

Once the silicon wafer is coated with the photoresist, it is exposed under high-intensity light to duplicate the pattern of circuit layout mask. The first exposure step is to align the light source, the silicon wafer coated with photoresist and the photomask between the first two. Then, the photomask and photoresist are exposed under the high-intensity visible or ultra-violet (UV) light (Fig. 2.1).

[Fig 2.1 Schematic Diagram of Photo Exposure Process]

The alignment of the silicon wafer and the photomask demands a high degree of accuracy. The projection printing method is widely used since it meets the requirement of high accuracy and resolution while avoiding the defects induced from the exposure process. The projection printing device moves the silicon wafer mechanically one step at a time (in a step-and-repeat fashion, thus the name steppers) so that a projected image of photomask is only exposed on a single die. This stepping exposure is repeated until the whole silicon wafer is exposed. Since the projection printing method focuses the light on a small area of the photomask, this approach reduces a problem caused by exposing the entire photomask and risking the possibility of leaving some areas of the photomask underexposed or unfocused.

[Figure 2.2 Schematics of Lithographic Projection Printer]
(Sze, 1983, p. 275)

The performance measures of the photoexposure process are characterized by the light source and the photoresist. Wavelength and intensity are two primary variables to consider during exposure to the light source. The wavelength determines the feature size of transistor cells and interconnection wires (resolution). The intensity of the light absorbed on the photoresist determines the time of exposure. Sensitivity and resolution characterize the photoresist. Sensitivity of the photoresist is represented by the amount of light energy necessary to change the photoresist. The more sensitive photoresist requires shorter exposure time as the photoresist takes less light energy to create a change. Resolution refers to the smallest feature size that can be printed on the photoresist. Photoresist resolution determines the aspect ratio defined as a ratio of the image opening to the photoresist thickness. The aspect ratio directly affects the density of chip layout pattern copied on the silicon wafer. Therefore, photoresist resolution is one factor used to determine circuit image resolution. During the exposure process, resolution of the photoresist may also be affected by other variables, including an initial resist thickness, prebake conditions, and developer chemistry (Van Zant, 1984).

Photoresist types determine the exposure (negative or positive) made on the silicon wafer. Positive-type photoresist becomes soluble when exposed under light. The longer the positive-type photoresist is exposed under light, the more soluble it becomes. At some threshold energy of the light, the positive-type photoresist becomes completely soluble. Unexposed negative-type photoresist tends to swell by solvent during the development process. This can cause a distortion or a loss of the pattern in the microcircuit. As a result,

the negative-type photoresist may produce more pinholes on its surface. Pinholes will generate defects in the original pattern copied on the underlying layer. However, even though positive-type photoresist provides better resolution, it is more difficult to use than the negative-type photoresist. Therefore, tight control of the lithography process is required to use the positive-type photoresist (Ghandhi, 1994). Even though very little of the exposed photoresist remains soluble by organic solvents, baking of the photoresist and the silicon wafer is done to densify the photoresist. This makes the photoresist more resistant to subsequent chemical etching processes.

The wafer development process dissolves away exposed regions of the photoresist. Only selected parts of the microcircuit design are now printed as a pattern of the photoresist coated on the silicon wafer. A short postbake follows to toughen the remaining photoresist.

During the photoresist exposure, it is essential to maintain a requisite thickness of photoresist. An excessive amount of photoresist results in under-exposure of selected regions. As a result, subsequent steps of development and etching may distort or lose the circuit pattern as photoresist may block some regions of the wafer surface during the processes. Well exposed and developed photoresist defines selected regions more clearly to produce a better result. A clean environment is another critical element. A tiny dust particle is large enough to cover multiple transistor cells or interconnecting wires which cause circuit failure. The dust particle can also distort the image transferred onto the surface of the silicon wafer through the lithographic process.

- Etching and Cleaning

Etching selectively removes unmasked regions of the top layer of the silicon wafer to cut openings in this layer. The photoresist layer protects masked regions and their underlying layer from the etching process. Through etched-out openings, doping (ion implantation or diffusion) is performed to build semiconductor regions. Etching can be done chemically, physically, or as a combination of both.

-- Wet Etching (Chemical Etching)

Chemical etching is an etching process that uses liquid chemical to remove selected regions of the top layer from the silicon wafer. The chemical etching device immerses silicon

wafers in a container of acid etchant. The reaction between acid etchant and unmasked thin film will etch out the exposed regions of the top layer of the silicon wafer. The performance of chemical etching depends on the immersion time, the temperature of the etchant, and the composition of acid etchant. Movement of etchant to the surface and the removal of the etch product are key performance parameters of the chemical etching. Etchant movement characterizes accuracy in pattern definition. It also affects the control of opening dimension due to undercutting. This may become a problem when etching makes openings too wide or distorts the desired shape. Even though chemical etching is the most economical etching process, there are contaminant issues in addition to etching control issues. Rinsing the chemical etchant may leave contaminant particles in the etchant container and on the surface of silicon wafers. These particles disturb subsequent doping and diffusion processes, producing defects.

-- Dry Etching

Dry etching includes plasma etching, ion beam milling, and reactive ion etching (RIE). Plasma etching and reactive ion etching methods use plasmas of noble or molecular gas instead of etchant liquid. Ion beam milling uses an ion blast to etch out unmasked regions of the top layer. A general advantage of dry etching is that it produces uniform, reproducible wide-area etching (Sze, 1990).

In the plasma etching process, etching is carried out by low pressure plasmas of reactive gases. High energy plasma fields excite the gas molecules in the etching chamber which react with the layer in opened regions of the silicon wafer. There are several advantages to use plasma etching. Since plasma etching uses gas, it is easier to control the etching process. Plasma etching also produces fewer particles than liquid etchants used in chemical etching. Also, it is less sensitive to changes in the temperature of the silicon wafer. These factors make plasma etching more repeatable than chemical etching. However, an uncontrolled plasma etcher can make a hole through the silicon wafer. Therefore, it is critical to control etching selectivity during the etching process. The high temperature associated plasma etching also makes photoresist harder to remove. Despite these potential problems, plasma etching is more widely used for its higher yield productivity.

Ion beam milling is a type of plasma etching that does not involve any chemical reaction between the etchant and the top layer material. Ion beam milling accelerates charged noble gas atoms (e.g. argon) to exposed regions of the silicon wafer. Charged gas atoms collide

and break away a part of the top layer material. It is a mechanical process analogous to sand-blasting. Ion beam milling method offers high resolution even though it has less selectivity.

Reactive Ion Etching (RIE) process combines plasma and ion beam milling etching methods. The RIE process uses molecular gas charges instead of the noble gas plasma used in plasma etching. Unmasked regions of the top layer are bombarded by molecular gas ions. The molecular gas plasma also creates a chemical reaction with the unmasked regions. The combination of chemical and physical reactions etch unmasked regions of the top layer. The RIE process is a very controllable etching process due to its anisotropic reaction characteristics. It is also largely insensitive to any photoresist residues left in the openings. This enhances its reproducibility (Kohyama, 1990; Sze, 1990)

The primary concern for any etching process is creating a uniform etching of the unmasked regions of the top layer. This requires a precise control of the degree and direction of etching during the process. The control depends on parameters including the thickness and density of the top layer being etched, plasma gas pressure and field energy.

- Photoresist Stripping

The etching process transfers a copy of the pattern for the microcircuit onto the top layer of the silicon wafer. After the etching process, it is necessary to eliminate the photoresist from the surface of the wafer. A remaining part of the layer forms a part of the DRAM circuit structure. There are multiple ways to strip photoresist from the surface of the silicon wafer. Chemical stripping methods use various solvents to remove photoresist, depending on the category of silicon surface layer (metalized, non-metalized). As in the case of the etching process, remaining particles from the stripping process may result in the failure of the chip. The degree of stripping has to be controlled to avoid over- or under-stripping. The dry stripping method uses excited oxygen atoms to oxidize the components of the photoresist material. The oxidization process converts the photoresist material into a gas which can be vacuumed out. Even though dry stripping does not have any wet chemical to handle, it has a potential source of contamination. The metallic ions contained in photoresist may not be completely stripped in the stripping process. These metallic ions may land on the silicon wafer and cause wafer contamination. They may also cause a failure of the microcircuit produced (Van Zant, 1984)

2.2.3 Doping: Diffusion, Ion Implantation

Doping process creates electronic characteristics of transistor cells by showering ions onto exposed regions of the silicon wafer. Impurity atoms in a single-crystal substrate turns selected regions into positive- or negative-type semiconductor materials. These semiconductor materials determine the primary characteristics of transistors. Diffusion and Ion implantation are two widely-used doping methods.

- Diffusion

The objective of the diffusion is to create junctions by diffusing the dopant atoms to a specific depth into the silicon wafer. This process consists of two major steps: deposition and drive-in (also called diffusion or reoxidation). Both steps of the diffusion process take place in a furnace holding silicon wafer. During the deposition step, a continuous supply of dopant atoms introduce the dopant into the silicon wafer. After the level of dopant concentration at the silicon wafer surface reaches to a specific point, the drive-in step redistributes the dopant atoms to a specific depth in the silicon wafer. The drive-in step continues diffusion at a higher temperature than the deposition step.

Temperature and time are two performance parameters in the diffusion process. These parameters control the amount of dopant atoms entering the silicon wafer. The amount of dopant atoms, in turn, determines the depth of diffusion characterizing the transistor circuit.

The highly concentrated dopants at the wafer surface makes the resulting circuit vulnerable to surface contamination. Since the diffusion doping process uses chemicals, the residue from the diffusion process may remain in the diffusion machine. This residue can become a source of defects for the next batch of the silicon wafers. The diffusion tube and its inner and outer environment can also be a source of contamination, including dust and other impurity particles. The high temperature may drive contaminants into the silicon wafer during the diffusion process and may also cause some crystal damage on the silicon wafer. Because the diffusion approach does not provide a higher level of resolution for circuit construction, it becomes a limiting factor for high density chip fabrication (Sze, 1990; Van Zant, 1984).

- Ion Implantation

Ion implantation is more widely used to introduce chemical impurities into selected regions of the silicon wafer. It creates p-type or n-type regions on the silicon wafer. Devices called ion implanters ionize dopant atoms and accelerate them to a high speed. The ion implanters focus accelerated dopant atoms into a beam that is scanned onto a silicon wafer. The dopant atoms penetrate into selected regions of the silicon wafer to form a semiconductor region.

Two parameters characterize ion implantation performance: ion energy and ion current. Ion energy determines the penetration depth of chemical impurities. Ion current controls the implantation time that sets the dose (amount of ion dopants implanted). These two parameters provide independent control over the doping process.

Ion implantation has advantages in doping and workflow control. Because it is usually carried out at room temperatures, it can use a variety of masking materials (such as silica, silicon nitride, aluminum, and photoresist). Using photoresist as a mask eliminates etching steps and simplifies the fabrication work flow. The low temperature operation also eliminates the problems caused by high temperature oxidation. The ion implantation method also offers better control of doses than thermal diffusion. This enables ion implantation to provide a higher level of accuracy for placing impurities in selected areas of the silicon wafer. Another advantage of ion implantation is that doping is possible without damaging the surface layers. CMOS (Complementary Metal Oxide Semiconductor) chip production uses the ion implantation for most doping tasks.

Ion implantation also has some disadvantages. It does not provide complete control over the penetration depth. This causes a non-uniform distribution of ion dopants in the silicon wafer. Acceleration control of ions is used to remedy this problem. Ion implantation may also result in crystal damage to the semiconductor. This can be particularly troublesome even with the annealing process to remedy the damages done during the process (Ghandhi, 1983; Kohyama, 1990; Sze, 1990).

3. DRAM Fabrication Technical Issues

3.1 List of Technical Issues

There are various sets of technical issues raised during the DRAM fabrication process. The issues are grouped according to their major technical traits: design, interconnection,

process, and manufacturing (SIA, 1994). Within each group, issues are categorized into two subgroups reflecting two technical requirements: chip size and chip robustness.

- Design & Design Tools

-- issues of chip size:
- minimum feature size
- chip sizes
- clock speed and frequencies
- lithography width, space, and alignment capability
- device structure issue with the leakage current
- power dissipation problem
- signal integrity problem (coupling between millions of nets) in the face of high clock rates
- the wiring levels

-- issues of chip robustness:
- smaller transistor operating margins (the precision of setting and controlling threshold voltage)
- etch biases and variations
- diffusion and oxidation parameters and control

- Interconnection

-- issues of chip size:
- complexities of interconnection routing
- voids from electromigration
- voids from stress voiding
- metal and via critical dimensional control
- power dissipation from charging and discharging of the parasitic capacitance of the interconnects
- signal integrity: accurate extraction and calculation of interconnect parastics (timing, proper timing at the system level)

-- issues of chip robustness:
- hot electron

- failures at interfaces of interconnect metalizations
- antireflection layers and resists
- interconnect manufacturing control problem caused by higher speeds
- interconnect reliability: improving interconnect reliability by process or layout optimization requires models relating failure rate to current density, material composition and geometry

• Process:

-- issues of chip size:
- critical dimension, minimum feature size, film thickness, contamination, isolation technologies, process control issues
- high etch-rate etching of small feature sizes with accurate dimensional control and minimal aspect-ratio-dependent etching
- number of mask layers
- layers of metal interconnect
- margin variations (wafer-to-wafer, across-wafer, within the stepper field, from day to day)
- variation in process tool output
- lithography and etching issues related to minimum feature size
- defect reduction in process equipment
- exposure tool - wavelength reduction, off-axis illumination, mask techniques (phase shift mask and optical proximity correction)
- particle detection and control

-- issues of chip robustness:
- process margin
- tool margin
- basic process sensitivities, equipment control, equipment irregularities
- sensor implementation
- process steps
- short-flow process characterization (defect inspection and allocation)
- overlay requirements - intergerometers, alignment systems
- process latitude, etch resistance, implant blocking, mechanical stability
- film characterizations, transferred to pre- and in-process controls (rather than relying on post-process measurements) p.94

- surface planarity, critical process control

- Manufacturing:

-- issues of chip size:
- contamination-free manufacturing
- manufacturing tolerances
- uniformity control

-- issues of chip robustness:
- manufacturing tolerance
- parametric control
- costs in dollars and risk
- reliability (adhesion, migration, ...)
- defect density reduction, reliability learning
- capital equipment utilization

3.2 Technical Issues: Interconnections and Transistor Cells

Since DRAM products consist of millions of identical memory cells and interconnection wires, this paper will examine the defects based on where they occur: interconnection layer and memory cells. There are many types of defects caused by numerous sources. They include operation errors, operation disturbances, contamination, design errors, material failures and equipment failures. One of the most difficult types of trouble with DRAM manufacturing is caused by physical characteristics of DRAM products. They are simply too small to provide any but the most minute margin of error. The following will provide two illustration examples.

3.2.1 Case 1: The Interconnection Layer

- Interconnection Layer Construction

The interconnection layer carries electrical current and the contacts channeling the current between interconnection wires and semiconductor/conductor regions of the DRAM circuit or between adjacent interconnection wires.

The interconnection layer is made of thin metal wires. Aluminum has been widely used for the construction of interconnection layer because it has low resistivity as a medium and has good adhesivity to silicon oxide, making thin film deposition and fine patterning easier.

Aluminum interconnection wires are formed through thin film layering and etching processes. Evaporating aluminum will cover the entire wafer surface with a thin aluminum film. The etching process leaves only selected regions of the thin aluminum film to form wires interconnecting selected regions of the semiconductor and conductor layers. Contacts are filled with aluminum through which electric current flows among semiconductor regions and metal interconnection wires. To build contacts, the holes (contact cuts) are first cut through insulator layer. Aluminum will flow in to fill contact cuts to complete contacts with semiconductor or conductor regions (Van Zant, 1984).

There are two immediate concern about the robustness in using aluminum as interconnection layer (SIA, 1994). The push for smaller chip size leads to thinner metal wires and smaller contacts built in decreasing proximity to one direction with respect to one another. This makes construction of the interconnection layer more complex regarding material density, wire structure, and the overall fabrication process. Also, chip size reduction increases the material density of aluminum wires, causing tress migration and electromigration.

- The Broken Interconnection Wire Problem

Deformation of interconnection wires is one common example of conflicting requirements in DRAM trouble management. Wire deformation is normally found during DRAM fabrication steps. There may be multiple causes resulting in wire deformation. The wire may have been deformed because of an error in controlling fabrication process parameters, such as operation temperature. Wire may have not been formed completely because of contaminants present on the wafer.

An example was provided by a field interview (Interview with T-Two, 1997). This example illustrates a more complex case. The defect caused by the reflective material used in the underlying layer. During the lithographic process, underlying material reflected light to the interconnection metal layer formed above. The photon energy distorted the formation of interconnection wire at the failing spot. The shrinking width of interconnection wires made this kind of effect more damaging.

3.2.2 Case 2: Transistor Cells

- Construction of Transistor Cells

Transistors are the basic units of a DRAM chip. A transistor is a solid state device modulating the flow of binary data stored as electric charge.

A typical CMOS DRAM chip cell has n-type source and drain regions and p-type region, forming a complementary CMOS circuit. A metal gate modulates electric conduction between the source and drain regions by its input voltage. These regions and the gate are separated from one another by an oxide insulator layer. Two regions of transistors (n-type, p-type) are built upon silicon substrate which functions as a base for the multiple layers: conductor, semiconductor, and insulator. There are more than 70 million transistors on a 64 MB DRAM chip. More than 200 fabrication steps (a sequence of layering, lithography, and doping steps) are required to build a 64 MB DRAM transistor cell structure with interconnecting wires linking the cells.

In a DRAM circuit, one bit of data is represented by one binary state of a signal. Binary data are stored by electric charge built up in the capacitors. There are two major types of memory product, based on how they store and access electric charges (i.e., to write and read binary data): DRAM and Static Random Access Memory (SRAM). The SRAM circuit is designed to hold an electric charge without recharging for a long period of time. SRAM chips also read and write data in their circuit faster than DRAM. However, these features limit the number of transistor cells possible on a SRAM chip and constrain the amount of binary data SRAM chip can hold. The DRAM circuit is designed to reduce the size of the memory circuit so that it can hold more binary data. This increases chip density significantly compared to SRAM. As a result, DRAM products are more widely used.

- Characteristics of Refresh Problem Of DRAM Transistor Cells

A major drawback of DRAM circuit is that it can hold electric charge only for a very short time (e.g. 100 nano second). Without reinputting binary information, the charge drains away very rapidly due to an even smaller leakage of electric current around the memory cell structure. DRAM circuit needs an external power to sustain its electric charge stored in DRAM memory cells unlike SRAM. The operation of re-inputting electric binary data is

called "*refresh*" (Kohyama, 1990). DRAM chip refreshing is done hundreds of times per second. The refresh rate is defined as a rate of reinputting the electric data per unit time. The refresh rate is an inverse of the duration of holding the electric charge until a refresh operation reinputs the data. The refresh frequency is determined by how long DRAM cells can hold a charge.

There are three major types of electric current leakage: capacitor dielectric, storage contact junction, and transistor cell (Interview with T-One, 1997). The capacitor dielectric current leakage problem is a drain of charge stored in the capacitor through a deformed part of an insulation layer between a capacitor and its surrounding material. Because the insulation layer is very thin, it is virtually impossible to pinpoint the location of deformation in the insulation layer. The method DRAM engineers use to identify the leakage point is similar to a detective work.

> *... We use an electro-microscope to identify and verify a location of the current leakage. In many cases, there may be deformed points in the insulation layer. However, their small size makes the physical examination not useful. I mean we cannot verify our finding for sure. We use other methods to induce our conclusion. We etch a hole through the insulation layer. Then, we insert a probe to measure the changes of the electric charge stored in a particular capacitor being examined. We measure the time of the charge dissipation to see if the capacitor can hold the charge as designed. If we find the capacitor violating the holding time, then we know that particular capacitor suffers from the capacitor dielectric current leakage. ...*
> (Interview with T-One, 1997)

Uniformly formed layers prevent this type of current leakage problem. However, variations in the fabrication process make it very difficult to obtain uniform thickness of layers because smaller chip size reduces tolerance for variations in the process. Variations include disturbances in the control of equipment and process parameters and produce slight degree of geometric and electric deformations of layers, including non-uniform layer thickness, irregularly shaped region, and misaligned cell structures. Another major factor causing current leakage is the quantum effect. Storage contact junction leakage is a drain of electric charge stored in a storage node through contact junction. This is caused by the electron tunneling effect when the conduction reaches a certain level in the storage node (Gasiorowicz, 1974). Transistor cell current leakage is similar to the storage contact junction leakage in the sense that electrons are driven by tunneling effect.

4. Conclusion

The push for micro-miniaturization with extremely small physical tolerances creates a tension between two technical requirements: smaller chip size and higher chip robustness. Smaller chip size makes the DRAM chip more vulnerable to production and operation variances. Seeking higher chip robustness often relaxes the chip size requirement to provide a wider margin of error in design and production. The broken interconnection layer problem demonstrated that micro-scale circuit design amplifies the damaging effect of high-intensity light. Increasing the physical tolerances may reduce the risk of the interconnection problem. However, as a result, the chip size is increased. Cell structure design also illustrated a similar trade-off between two requirements. Larger cell structure will reduce the negative effect of current leakage while sacrificing chip density. The difficulty with this tension is that while it can be managed, it cannot be completely eliminated as a consequence of the demand for micro-miniaturization.

Extremely small physical tolerances in DRAM micro-miniaturization make the recurrence of problems routine. The small physical tolerances also make the outcome of technical action uncertain. This uncertainty greatly complicates the task of technology decision. The task of finding a satisfying balance between the two requirements becomes particularly difficult. It is central to DRAM trouble management to manage the uncertainty caused by the micro-miniaturization while balancing the two conflicting requirements of chip size and robustness. Therefore, it becomes necessary to examine the actual process of the trouble management as practiced in the DRAM production. Within the context of omni-present tension and recurring problems, the following problems remain empirical questions in DRAM trouble management.

- How do the engineers handle the omni-present tension? Is there any pattern they routinely practice?
- Do DRAM engineers follow a traditional approach -- hierarchically organized and compartmentalized -- to resolve the problems?

The questions will help us to characterize the process of the DRAM trouble management. The study leads us to recognize the knowledge used for the trouble management and to identify what activities are related to creation and accumulation of the knowledge. Subsequently, the findings will offer a description for a set of elements that facilitate/retard the LBD process. The set of elements includes a type and degree of communication among participants of the LBD process and characteristics of information created and shared by them. The study will also enhance our understanding on the role of learning-by-

doing(LBD) capacity to carry out LBD activities by locating the mechanism of changing relationships among learning, strategy and DRAM development knowledge [Figure 1.2].

Furthermore, as framed in Figure 1.1, characterizing the trouble management practice will help us to examine as in the following. A collection of activities, including a nature of resolution practice as political negotiation, will be investigated to understand a process of how inherent technical tension is resolved during production. Findings on the trouble management practice will enrich a description of the conditions and mechanism of successful DRAM production. Various elements, including organizational culture, rules, routines, and technological capacity will be examined for their roles in the tension resolution process. External elements, such as a role of equipment manufacturers, will also be accessed. The lessons learned from the DRAM production cases will lead us to explore emerging technical issues of micro-miniaturization in the HPOs. The technical issues will be studied empirically to characterize the technology development of HPOs outside the DRAM industry. This effort will show us inherent tension caused by the micro-miniaturization. With this understanding, the study will reach to the level of NSI analysis and provide us with a good instance of the mechanism by which various conditions offered by the NSI are materialized toward tension resolution of HPO production. In addition, the explanation on the role of LBD capacity and relevant NSI capacity will enable us to ask empirical questions including what each country's NSI did to make their LBD successful and why there are variances in their outcome. Broadly, the study will provide a ground to examine how the NSI provides a capacity for the firms to initiate and sustain their innovation effort such as LBD.

5. References

Arrow, K., "The Economic Implications of Learning By Doing," *Review of Economic Studies,* 29:155-173, 1962

Dosi, G., "Technological Paradigms and Technological Trajectories: A Suggested Interpretation of the Determinants and Directions of Technical Change," *Research Policy*, 11:147-162, 1982

E.E. Times, July 7, 1998

Gasiorowicz, S., *Quantum Physics*, John Wiley & Sons, Inc. 1974

Ghandhi, S., *VLSI Fabrication Principles: Silicon and Gallium Arsenide*, John Wiley & Sons, Inc. 1994

Irwin, D. and P. Klenow, "Learning By Doing Spillovers in the Semiconductor Industry," *Journal of Political Economy,* 102(6):1200-1227, 1994

Jang, K.C., *A Study on the Background of the Success of DRAM Development in the Industries amd the Roles of the Government Joint R&D Projects*, Master Thesis, Choongang University, 1995

Kohyama, S. (ed.), *Very High Speed MOS Devices,* Clarendon Press, 1990

Leachman, R. (ed.), *The Competitive Semiconductor Manufacturing Survey,* Engineering Systems Research Center, University of California, Berkeley, 1994

Mowery, D. and J. Oxley, "Inward Technology Transfer and Competitiveness: the Role of National Innovation Systems," *Cambridge Journal of Economics*, 19:67-93, 1995

Okimoto, D., T. Sugano and F. Weinstein, *Competitive Edge: The Semiconductor Industry in the U.S. and Japan,* Stanford University Press, 1984

SIA, *The National Technology Roadmap for Semiconductors,* Semiconductor Industry Association, 1994

Sze, S.M. (ed.), *High-Speed Semiconductor Devices,* John Wiley & Sons, 1990

Sze, S.M. (ed.), *VLSI Technology* , McGraw-Hill Book Company, 1983

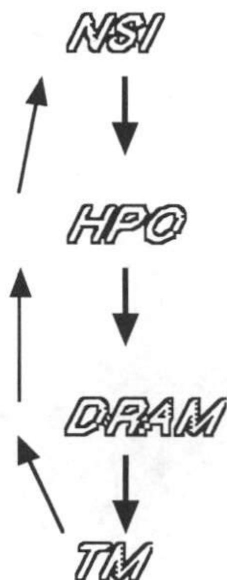
T-One, *Interview Transcription,* 1997

T-Two, *Interview Transcription,* 1997

Van Zant, P., *Microchip Fabrication,* Semiconductor Services, 1984

Yum, J., *A Bureaucratic Organization in a Network Setting: MITI and Japanese Industrial*

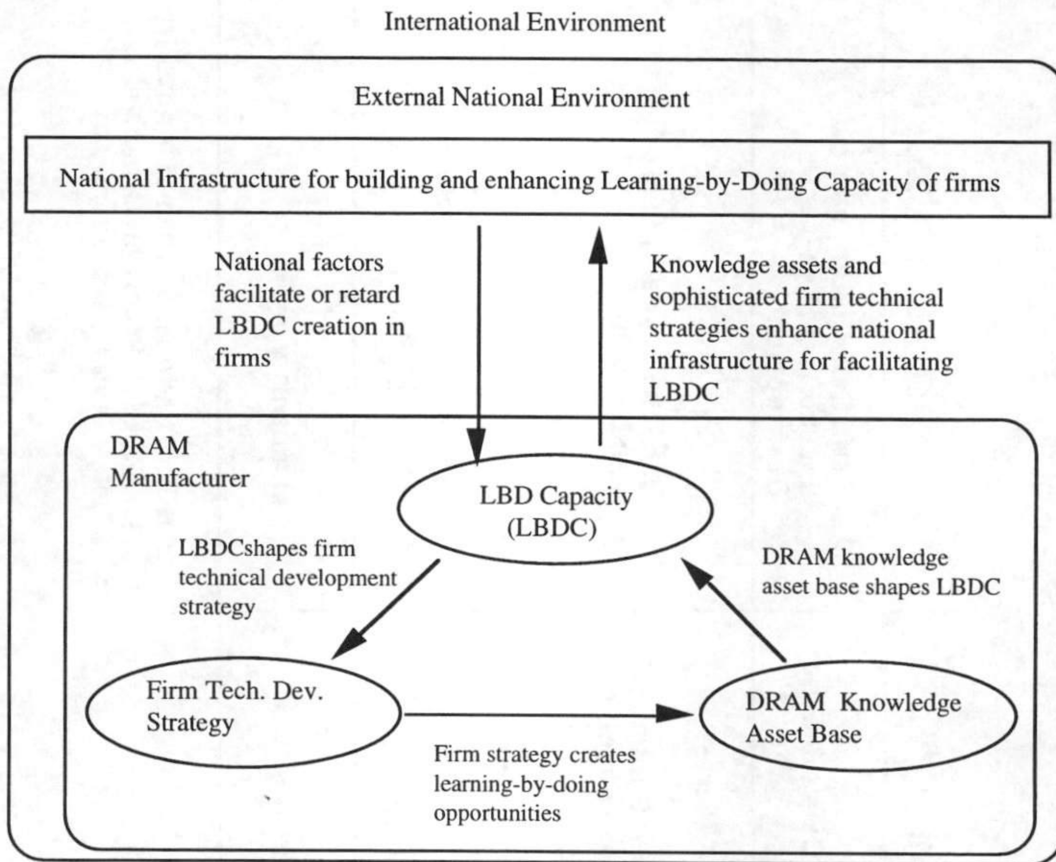
Policy for High Technology, Ph.D. Dissertation, Stanford University, 1989



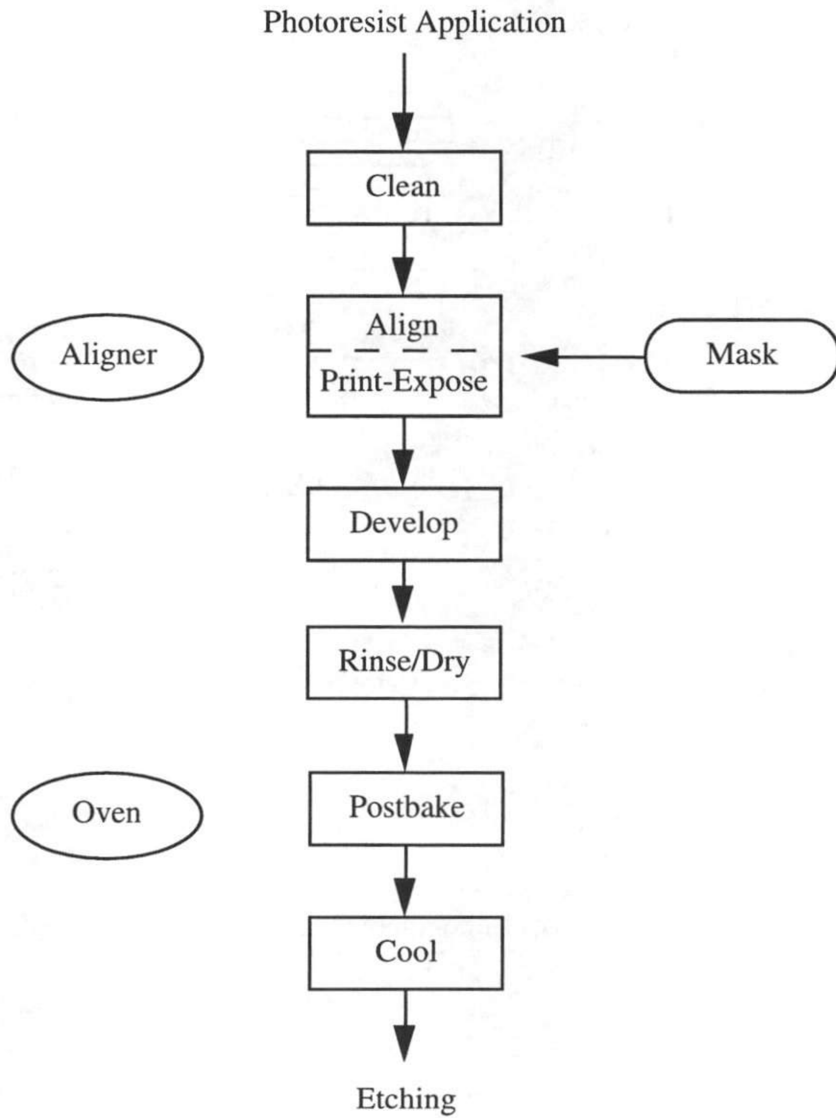
Motivation	Focus of Analysis
Distribution of DRAM Development	Conditions (often, context-specific) for High-Tech. Development
Special Char. of HPO - Micro-Miniaturization Problems in Innovation: Continual Occurence of Problems caused by Micro-scale size	Tension caused by Micro-Miniaturization
Case in Point: - Competitive market - Leading the S.C. industry	Tension between Performance and Robustness
Central to DRAM Production	Mechanism of TM

NSI: National System of Innovation
HPO: High Precision Organization
TM: Trouble Management

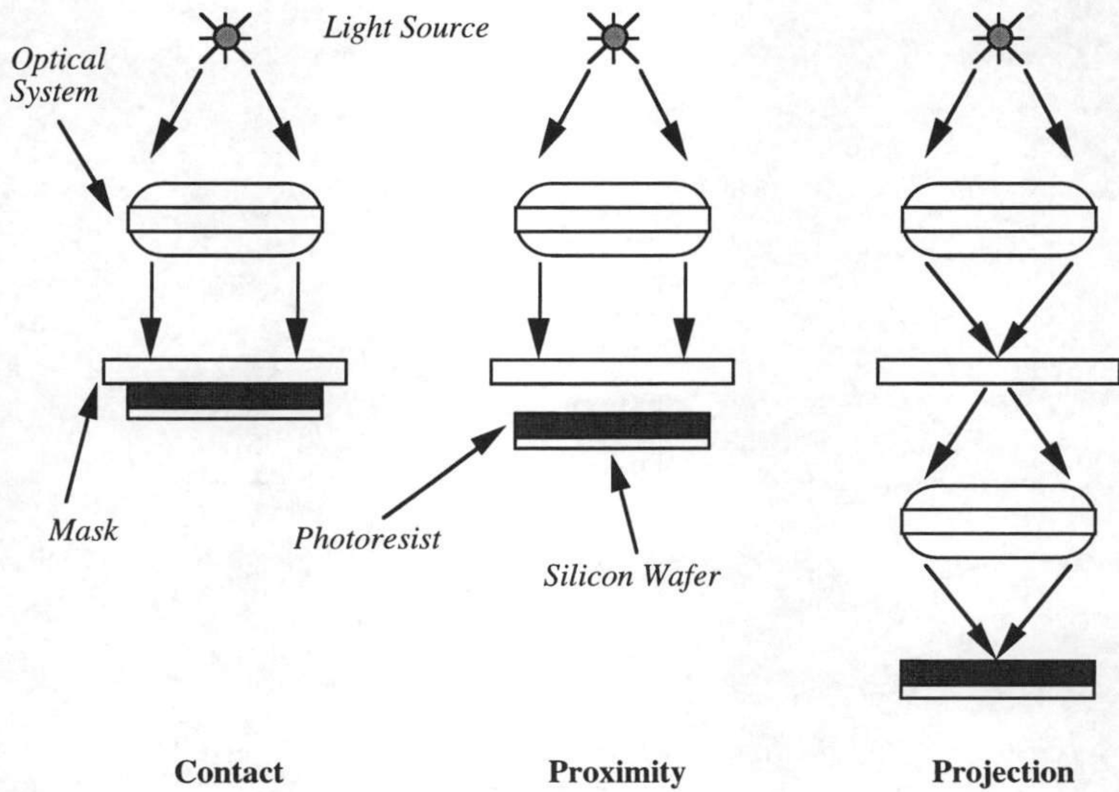
[Figure 1.1 Problem Domain]



[Figure 1.2 Learning-by-doing and National System of Innovation]



[Fig 2.1 Schematic Diagram of Photo Exposure Process]



[Figure 2.2 Schematics of Lithographic Aligners, (Sze, 1983, p. 275)]