Technology Strategy for the Semiconductor Memory Market

by

Tomohiko Nakamura

B.S. Electrical and Electronics System Engineering, Osaka Prefecture University (2001)
M.S. Electrical Engineering and Information Science, Osaka Prefecture University (2003)

Submitted to the System Design and Management Program
in Partial Fulfillment of the Requirements for the Degree of

Master of Science in Engineering and Management

at the

Massachusetts Institute of Technology

June 2012

Signature of Author____
Tomohiko Nakamura
System Design and Management Program
May 2012

Certified by____________________
James M. Utterback
Thesis Supervisor
David J. McGrath jr (1959) Professor of Management and Innovation
and Professor of Engineering Systems

Certified by____________
Henry Birdseye Weil
Thesis Reader
Senior Lecturer, MIT Sloan School of Management

Accepted by______________________________
Patrick Hale
Director
System Design & Management Program

Technology Strategy for the Semiconductor Memory Market

by

Tomohiko Nakamura

Submitted to the System Design and Management Program on May 11th, 2012
in Partial Fulfillment of the Requirements for the Degree of
Master of Science in Engineering and Management

ABSTRACT

Solid state memories are used in a variety of applications as data and code storages. A non-volatile memory is a memory that retains information when its power supply is off. Flash memory is a type of nonvolatile memory that can be erased and programmed by data called blocks. Flash memory is used to store system control codes in various applications. A recent representative example is Solid State Drive (SSD). A SSD is a type of computer storage that is composed of a number of flash memories. SSDs store data electronically unlike hard disk drives, which store data magnetically. The advantages of a solid-state drive are: no mechanical parts, less power and weight than hard disk drives.

This thesis investigates existing technologies, markets, emerging technologies, and applications in the solid state memory market. It analyzes the current solid state memory industry structure using market data and frameworks. The emerging technologies and applications are researched in order to deliver technological innovations to the semiconductor memory market. It also studies and suggests how strategies of firms might influence technologies, value chains, and future evolutions using system dynamics models. The system dynamics models are based on a conventional commodity market model because of similarity between the semiconductor memory market and the commodity market. The simulation results provide insights into future market evolutions caused by new technologies. This research will lead to recommendations for companies currently struggling to survive in the industry, and for companies considering entry.

Thesis Advisor: James M. Utterback

Title: David J. McGrath jr (1959) Professor of Management and Innovation and Professor of Engineering Systems

Thesis Advisor: Henry Birdseye Weil

Title: Senior Lecturer, MIT Sloan School of Management

ACKNOWLEDGEMENT

I would like to thank to all who supported my thesis and SDM life in Cambridge.

Professor James Utterback, my thesis supervisor, provided thoughtful suggestions and advices on my research activity and on my thesis. His expertise in management and technology help achieve the goal of this thesis. Without his warm support and encouragement, this thesis would not be possible.

I would also like to express my gratitude to Professor Henry Weil, my thesis reader. He provided helpful suggestions on my system dynamics models and data collections based on his experiences.

Lastly, I would also like to thank my family who helped me to get through my time at SDM. Without their understanding and support, I could not have completed my work.

Contents

List of Figures

List of Tables

1. Introduction

Companies make strategic decisions on structures of value chains, architectures, interfaces, and standards in order to deliver value to their customers. Each decision is not completely isolated from other decisions. Shifts in value chains are happening within the semiconductor industry value chain. The role of vendors and suppliers are changing to increase the value to end customers.

According to Michael Porter's classic value chain model, the goal of activities in a value chain is to add more value for the customer than the cost of carrying out the activity. The activities to add more value depends ultimately on obtaining cost competitiveness, differentiation, or the flexibility to adapt to customer needs.

The semiconductor value chain has many activities and players; for example material suppliers, equipment suppliers, manufacturing service suppliers, integrated device manufacturers, fabless chip makers, original equipment manufacturers, and end users. The structure of the value chain changes every decade, depending on the complexity of technologies, number of players, end user needs, and associated degree of vertical integration.

Solid state memories are used in a variety of applications as data or code storages. A non-volatile memory is a memory that retains information when its power supply is off. This definition generally includes all types of Read Only Memory (ROM), such as Programmable Read Only Memory (PROM), Erasable Programmable Read Only Memory (EPROM), Electrically Erasable Programmable Read Only Memory (EEPROM), and flash memory.

Flash memory is a type of nonvolatile memory that can be erased and programmed in units of data called blocks. It is an evolution from EEPROM, which is erased and programmed by byte. Flash memory obtained its name because the memory cells are erased in a single operation. Some companies offer a flash memory that holds more than two bits in each memory cell, thus doubling the capacity of memory without increasing the chip size. Flash memory is used to store system control codes such as the Basic Input-Output System (BIOS) in personal computers, digital cellular phones, digital cameras, LAN switches, PC Cards for notebook computers, digital set top boxes, embedded controllers, and other devices. There are two types of flash memory: NAND flash and NOR flash.

A Solid State Drive (SSD) is a type of computer storage that is composed of a number of flash memories. SSDs store data electronically, unlike hard disk drives, which store data magnetically. The notable advantage of a solid-state drive is that it does not have mechanical parts, which allows writing/erasing data at a much higher speed and has a more predictable lifespan for the storage drive. In addition, it requires less power and weight than hard disk drives.

This thesis investigates the existing technologies in solid state memories, and suggests evolution scenarios in the near future. It analyzes the current solid state memory industry structure using value chain analysis. It also studies and suggests how firms might influence technologies, value chains, and future evolution using system dynamics models. This research will lead to recommendations for companies currently struggling to survive in the industry, and for companies considering entry.

2. Market Analysis – Applications of solid state memory

The main goal of this thesis is to describe the possible evolution in memory industry in future. This thesis makes recommendations that will assist investment in technologies, standardization, product development, and strategic decisions for technology and alliances. To begin with, the market analysis is necessary to know current situation of the memory industry.

2.1 Technological trend

Generally speaking, there are two types of semiconductor memory: volatile memory and non-volatile memory. A volatile memory, in other words RAM, loses data when its power is off. On the other hand, a Non-Volatile Memory, in other words ROM, can retain data even though its power is off. There are several types of RAM and ROM. Table.2.1 shows a classification of the types of memory devices.

Table.2.1 Classification of memory devices

Volatile Memory (RAM)	SRAM	Asynchronous SRAM	
		Synchronous SRAM	
	DRAM	Asynchronous DRAM	
		Synchronous DRAM	SDRAM
			DDR/DDR2/DDR3 DRAM
			XDR-DRAM
Non-Volatile Memory (ROM)	Mask ROM		
	Programmable ROM	OTP ROM	
		EPROM	
		EEPROM	
		Flash ROM	NOR type
			NAND type

SRAM

SRAM (Static Random Access Memory) is a type of random access memory which has a memory cell composed of flip-flop circuits with several transistors. SRAM can retain storage data without a refresh operation while its power is on. SRAM has faster access and lower power consumption than DRAM. However, SRAM needs bigger area because it has a number of transistors in a cell. The biggest disadvantage of SRAM is its price when it is compared to DRAM.

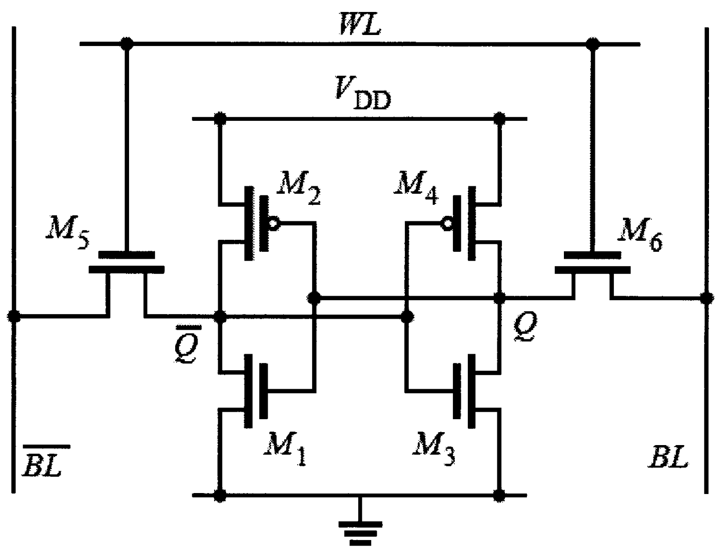


Figure.2.1 SRAM with 6-transistor cell

DRAM

DRAM (Dynamic Random Access Memory) has a memory cell which is composed of a capacitor and a transistor. The data can be changed either charging or discharging the capacitor. DRAM needs a refresh operation to keep the programing data because it stores the data into the capacitor. Capacitors leak currents in steady state conditions, thus the information is eventually evaporated if the capacitor is not refreshed periodically. This is why DRAM is named dynamic, and the main difference between DRAM and SRAM.

DRAM is mainly used in personal computers. It is in laptop and workstation computers as well as in video game machines. An advantage of DRAM over SRAM is its bit per cost. DRAM is structurally simple and takes smaller space. This feature allows DRAM to reach higher densities than SRAM.

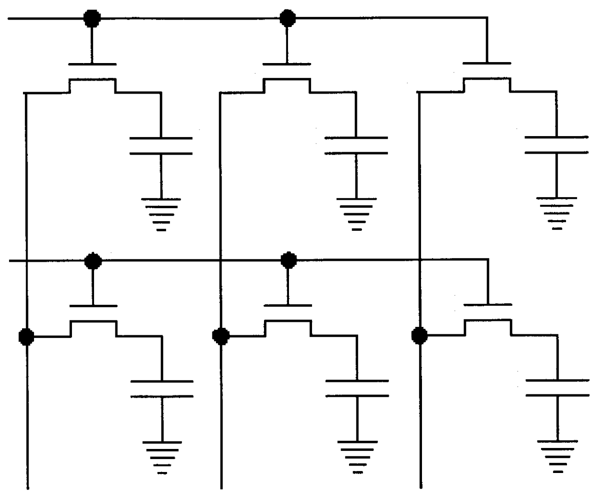


Figure.2.2 A Conventional DRAM cell

Mask ROM

Mask ROM is a type of memory whose data is written in a factory. Contents of Mask ROM are pre-programmed by mask ROM manufacturers, not by users. Because the data is programmed in manufacturing process, users cannot erase or reprogram the data. The word "Mask" means a mask used in photolithography of a fabrication process.

Mask ROM is suitable for mass productions because it has lower cost. Mask ROM has more compact memory cells than any other types of semiconductor memories. If development and validation are already completed, there is no need to change the data again. In such a case, Mask ROM is the best choice to reduce manufacturing costs. In contrast, a mask cost is high although it is only for once, and there is a long Turn-Around Time from design to production. If design errors are found in a validation phase, mask changes cost very high. In addition, if an error in the data or code is found during the actual use of mask ROMs, it must be replaced and reproduced with new costly masks.

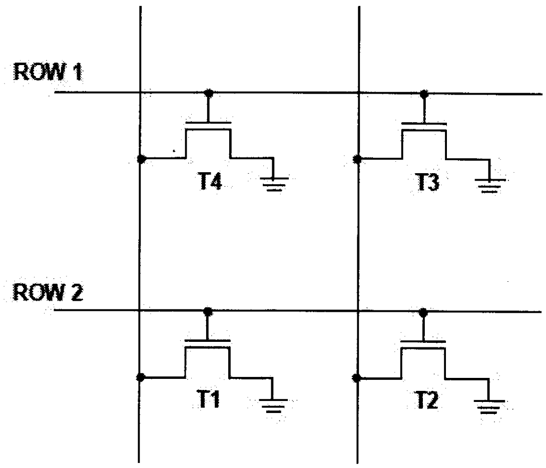


Figure.2.3 Cell structure of Mask ROM

Programmable ROM

Programmable ROM (Read Only Memory) is a type of ROM which is reprogrammable by users. This type of memory is suitable for developing new products or products which will not be produced at a large number of times. OTP ROM, EPROM, EEPROM, and Flash memory are a kind of programmable ROMs.

OTP ROM

OTP ROM (One Time Programmable ROM) is a ROM which is programmable only once. Actually this type of memory could have data erase functions, but it is not erasable because erase function is disabled internally.

EPROM

EPROM (Erasable Programmable Read Only Memory) is a type of memory that can be programmed by users and erased by ultra violet light. EPROMs are visually recognizable by its window on top of the package. Users can program data using an electronic device that supplies higher voltages. Once it is programmed, EPROM can be erased by exposing it to strong ultraviolet light. The window allows EPROM to expose its silicon die to UV light.

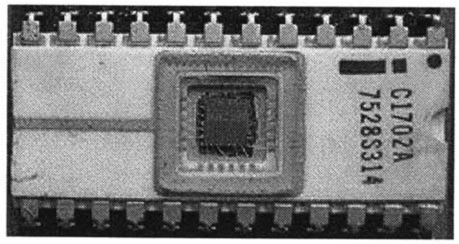


Fig.2.4 EPROM package

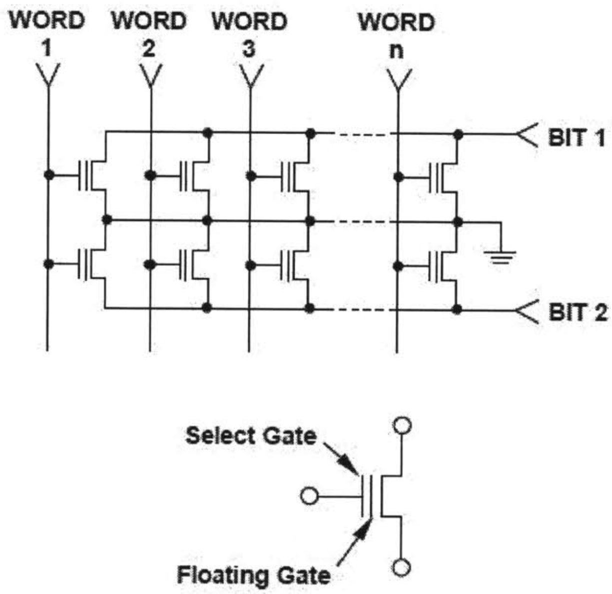


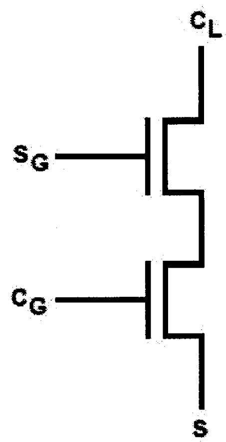
Figure.2.5 Cell structure of EPROM

EEPROM

EEPROM (Electronically Erasable and Programmable Read Only Memory) is erasable and programmable electronically. It is a type of non-volatile memory used to store small amounts of data that must be saved at its power-off condition. Typically system parameters or device configuration data are stored in EEPROM. EEPROMs have arrays of floating-gate transistors as memory core cells.

EEPROM is suitable for applications which require repetitive program/erase cycles. Although EPROMs must be physically removed from the device for erasing and programming, EEPROMs are programmable and erasable on a circuit board.

EEPROMs are mainly used with a serial bus interface because EEPROMs used to be limited to single byte operations. However recent EEPROMS allow multi-page operations in serial interfaces, such as SPI and I^2C.



	S_G	C_L	C_G	S
Erase	V_{PP}	0	V_{PP}	0
Program	V_{PP}	V_{PP}	0	0
Read	V_{CC}	1	V_{CC}	0
Unselected	0	X	0	0

Figure.2.6 Cell structure of EEPROM

Flash memory

Flash memory is a non-volatile memory that can be electrically erased and reprogrammed. It is evolved from EEPROM, but the difference is that the data must be erased in units of blocks. Flash memory is a better choice than EEPROMs in terms of cost if large amounts of data are to be stored. There are two representative types of flash memory: NOR type and NAND type.

NOR flash memory

NOR type allows a byte (8-bits) or word (16-bits) operation to be read or programmed independently or simultaneously. The NOR allows faster random access than NAND, thus direct code execution from flash memory is possible. It can be used as a replacement for EPROMs and mask ROMs. Representative applications are BIOS data storage for personal computers, PDAs, mobile phones, video games, industrial robotics, and medical electronics.

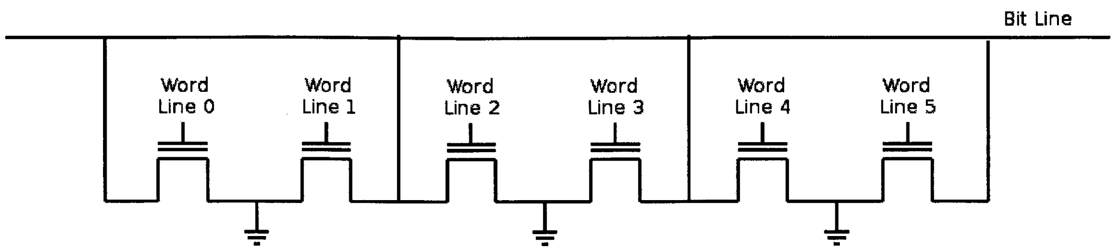


Figure.2.7 Cell structure of NOR flash

NAND flash memory

The high density NAND type must also be programmed and read by blocks, or pages. The NAND type is primarily used in memory cards, USB flash drives, and solid-state drives. NAND are used to store large amounts of data, and is suitable for data storage media. However, NAND flash memory cells are easily broken because of the proximity of memory cells. Software controls, such as wear leveling and garbage collections, are used to extend a life cycle of the product.

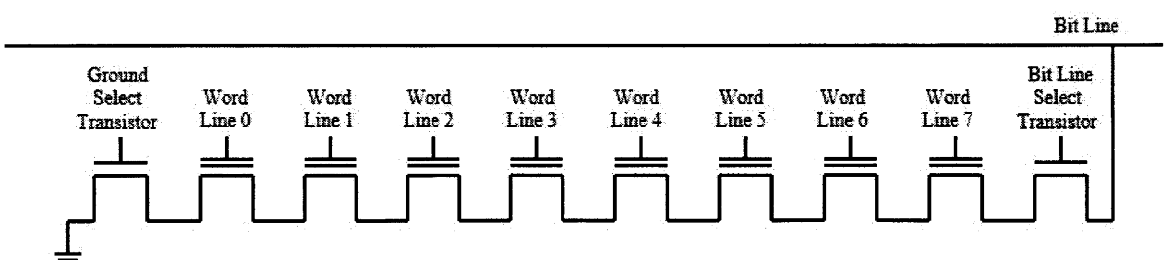


Figure.2.8 Cell structure of NAND flash

2.2 Business trend

The Semiconductor market

The semiconductor industry has experienced many changes for the last several decades. One of the biggest changes is the rise of the consumer electronics market as a demand driver for semiconductors. The growing demand for information technologies, automotive products and consumer electronics products, and mobile phones are driving the demand of the semiconductors market. The major segments of semiconductor applications have been identified as wireless communications, internet, personal computers and consumer electronics. Microprocessors, memory chips, discrete devices, application-specific integrated circuits (ASIC), and Field Programmable Gate Arrays (FPGA) are the major semiconductor proroducts.

Convergence of applications and technologies is the major trend in the semiconductor industry. Semiconductor technologies are researched for continuously scaling down to smaller sizes using smaller scales. Thus more functionality is to be implemented on a chip. A number of chips on a Printed Circuit Board today will be merged onto a single chip in future.

The semiconductor industry is historically cyclic: growing steadily with downward cycle and upward cycle. Figure 2.9 shows the annual change in semiconductor revenue (y axis) by calendar quarters for 1996–1999, 2000–2002, 2008–2010. The semiconductor industry currently experiencing the down cycle started in early 2011. In all cycles, the starting point is the first calendar quarter because 4th quarter involves Thanksgiving Day and Christmas. Sales of semiconductor go down in conjunction with the peak of consumer products. According to a prediction from IDC, the current semiconductor cycle is expected to bottom in early 2Q12. Capital spending for factories will also pick up at that time.

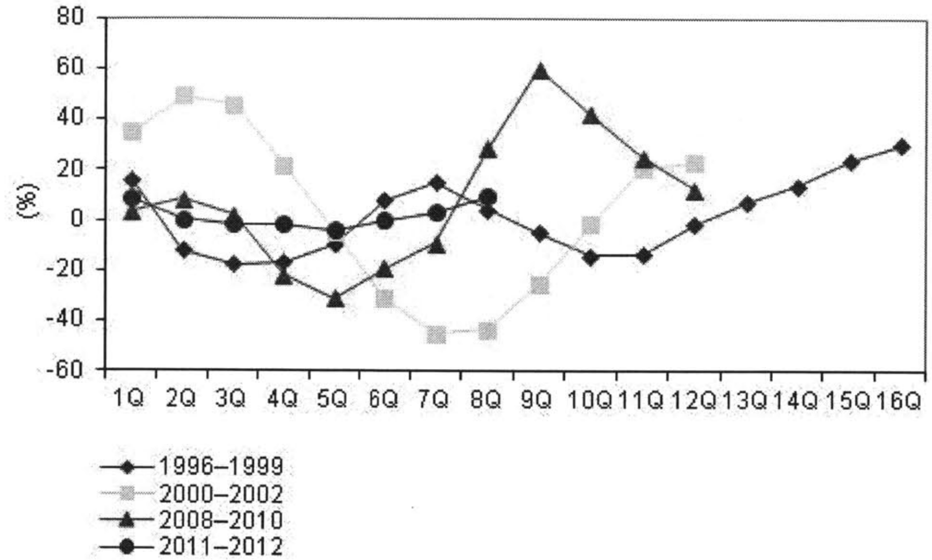


Figure 2.9 Worldwide Semiconductor Revenue Growths by Quarters, 1996–2012
(Source: IDC, 2012)

Memory market

Dynamic random access memory, DRAM had been the main memory component of most personal computers and consumer electronics products. DRAM was "a workhorse" of IC industry.

However, DRAM vendors are currently facing chronic oversupply issues due to overproduction. Such supply issues forces several vendors to reduce or terminate their DRAM business. In addition, weak demand and pricing environment in coming years will force DRAM vendors into more consolidation. Most vendors know current pricing cannot meet the cost for advanced technology. All DRAM vendors will reduce their capital expenditures in 2012.

Instead of the weak demand for DRAM, flash memory market has experienced rapid growth for the last decade. In particular, the flash memory market growth has either outpaced or equaled to the total semiconductor market, and expected to continue the growth for the next several years. The flash memory market has become a significant part of the semiconductor market. There are some arguments that it will soon replace random access memories which are suitable for fast but temporary storage. Because of its market size and the future growth, flash memory is going to have more impacts on the global semiconductor industry.

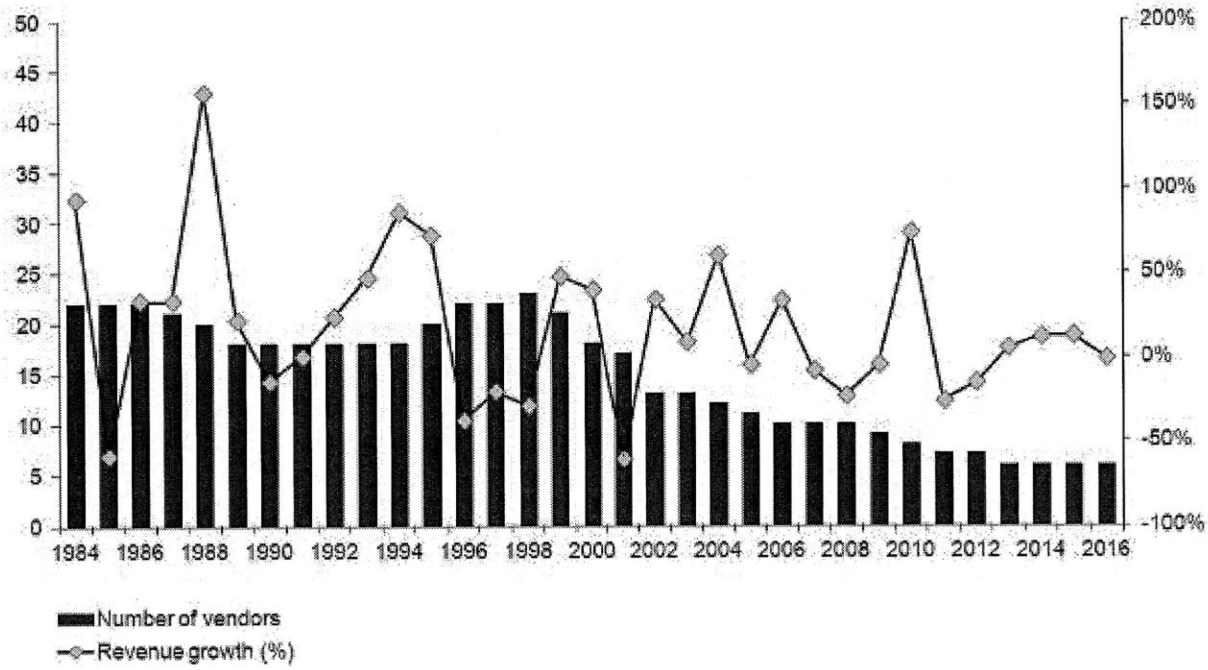


Figure.2.10 History and Outlook on DRAM Revenue Growth, 1984–2016
(Source: IDC, 2012)

Non Volatile Memory market

Flash memory manufacturers are producing more chips to meet the market demand. As a result, price of memory goes down day by day because of overcapacity. Some manufacturers have changed production from DRAM to flash memory, while some have both productions to absorb the variance of market demands.

Non-volatile memory was first developed in the early 1980s, but it was looking for new applications which require nearly permanent date storage. It took almost a decade to fully emerge as a market in the early 1990s, and the industry was confident about the future growth potential. In the early 1990s, demand for the first non-volatile memory, EPROM, is led by laptop PCs with the support of a long-term growing trend in the semiconductor industry. It also stimulated the emergence of the consumer electronics products and portable devices, which later become the primary driver of end-use demand.

In the middle of 1990s, EEPROM began competing against EPROM. In terms of cost competitiveness and its functionality to enable erase operations, the market grew to the next stage. Those two technologies, EPROM and EEPROM are primarily led by Intel, the leader in flash memory production at that time.

They kept lowering flash memory price-per-bit ratio. Regarding the future demand for EEPROM, Intel began shifting their production to EEPROM and away from DRAM business.

NOR flash gained bigger attention in the end of 1990s. Because NOR type flash memory has faster random read operation while it has longer erase and write operation times. It is a suitable replacement for traditional mask ROMs. ROMs are at that time used to store program codes that usually do not need to be updated. NOR flash memory found effective uses in computer's BIOS, base-band processing in mobile phones, and parameter storage in set-top boxes. NOR-based flash was the basis of early flash-based removable media. CompactFlash was originally based on NOR flash, but later NAND flash replaced it because of its lower cost.

In the early 2000s, sales of NAND flash grew rapidly. This is mainly because price of NAND flash is lower than the price of NOR flash. It requires less chip area per cell due to smaller memory cell size, allowing greater storage density and lower cost per bit. It also possesses shorter erase and write times, and up to ten times higher endurance than NOR flash. However, the I/O interface of NAND flash does not allow a random-access unlike SRAM, Mask ROM, and NOR flash. Moreover, the data in NAND flash is easily broken, which means cannot be used as a programming code storage media. This made NAND flash unsuitable for simple replacement of NOR flash. Thus NAND flash tried to seek other applications, such as hard disks and optical media. NAND flash is first used in a removable media in 1995. Other types of storage media, such as USB memory, Multi-Media-Card, Secure Digital, and Memory Stick followed later.

The most recent use of NAND flash memory is as a replacement of hard disks. A Solid-State Drive outweighs traditional Hard Disk Drives in terms of speed, noise, power consumption, and reliability. This is because flash memory does not have mechanical devices and latencies which hard disk drives have. SSDs are gaining attention as mobile secondary storage devices. They are also used as substitutes for hard disk drives in high-performance desktop computers and some cloud servers.

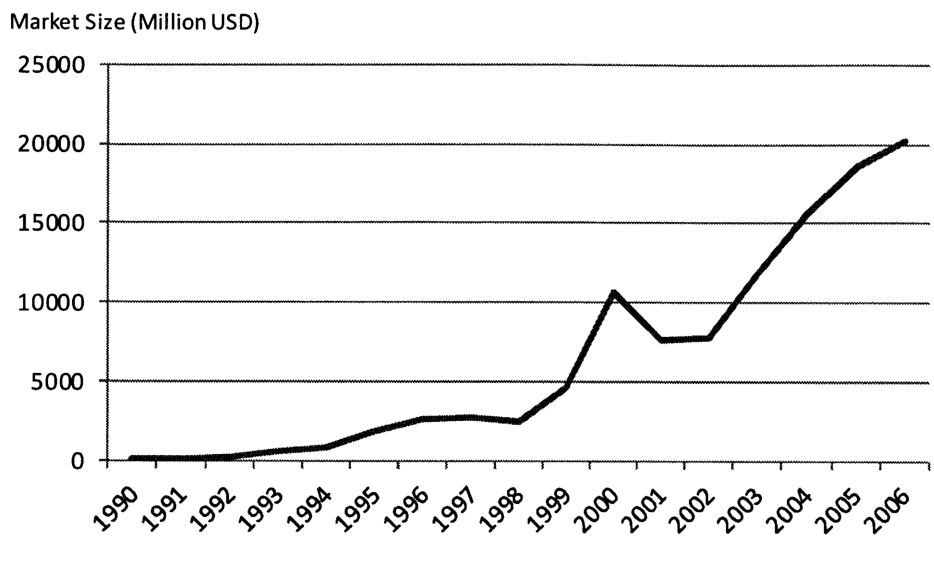


Figure 2.11 Historical market size of flash memory

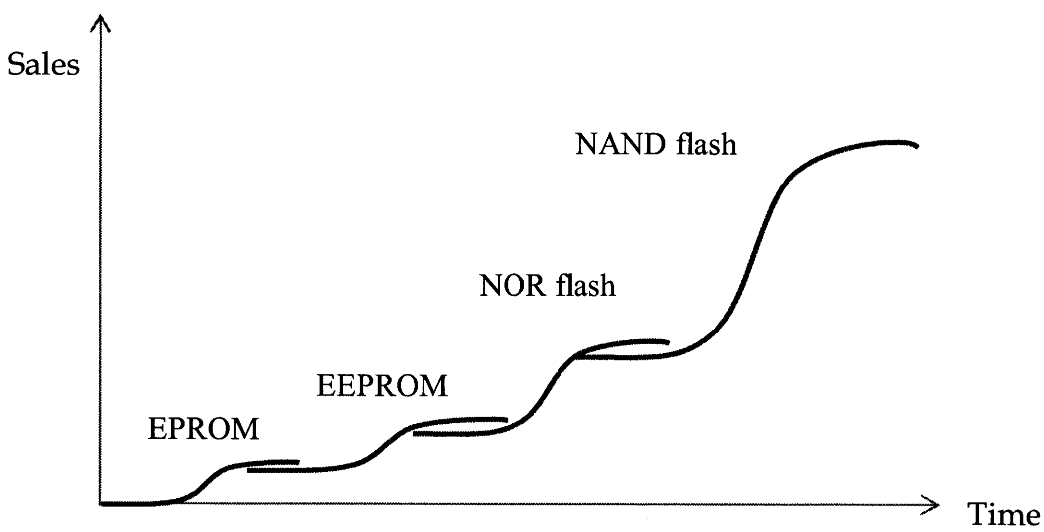


Figure.2.12 S-curve representation of the flash memory market

2.3 Design Value Chain

This section analyzes the current value chain structure in the semiconductor industry and memory market, and describes changes taking place in them. This section also mentions what suppliers are taking critical roles in the value chain. There are issues arising from changes in the semiconductor value chain: Overall changes in the semiconductor value chain, and changes within specific segments of the value chain.

According to Michael Porter's value chain model, the goal of each activity in a value chain is to add more value for the customer than the cost of running the business. The ability to add value depends on a cost advantage, differentiation, or providing state of the art technologies. The structure of the semiconductor value chain is shown in Figure.2.13. There are many players in the value chain because semiconductor business is a large industry. Although Integrated Device Manufacturers (IDMs) are playing the most critical roles in the industry, other players, such as CAD tool vendors, IP vendors, materials vendors, equipment suppliers, foundries, fabless vendors, packaging services, set makers, distributors, and original equipment manufacturers (OEMs) are not small businesses.

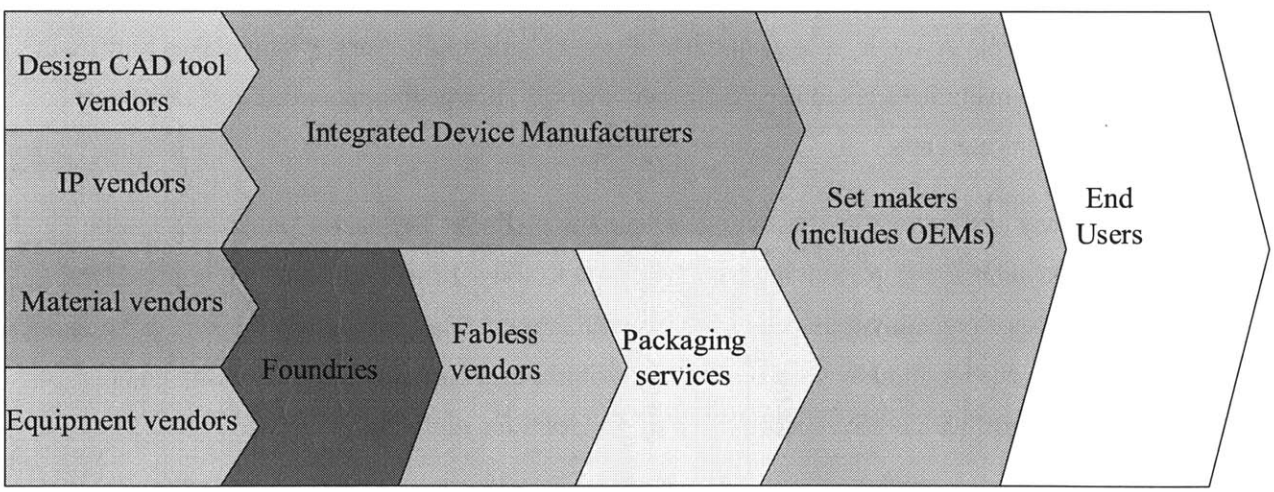


Figure.2.13 Value Chain of the Semiconductor Industry

The value chain of the semiconductor industry has been experiencing rapid changes and will likely cause a major restructuring in the near future. The main value in semiconductor devices is shifting from manufacturing into design, especially logic products such as microprocessors, ASICs, and FPGAs. The main reason is that maintaining factories inside of a company is becoming more challenging due to rising prices of manufacturing equipment. However, in memory industry it is very hard to separate semiconductor design and manufacturing facilities since memory products are a type of analog products: process parameters greatly affect design methodology.

As the industry matures, understanding the values and profits of customers becomes a critical issue for survival. The value of products is shifting from hardware to software, particularly in services, contents or user interface designs. Industry players should consider impacts of developments based on influence from up and down in the value chain, and need to deal with various groups of stakeholders. They will have to adopt new business models, for example outsourcing some operations to maintain their cost competitiveness.

Therefore the major factors changing the industry can be predicted by considering requirements from demand sides and changes from supply sides in the value chain.

Demand-Side Changes

As growth in key demand drivers such as PCs and cellular phones slows, the industry is looking for new. The demand now is driven by a large number of low-volume applications, which makes it difficult for integrated device manufacturers to keep their own factories. It is getting more difficult to support design and manufacturing costs.

In addition, many applications are becoming commodity products. The values and powers on the demand side of the value chain are moving from hardware to software services and content providers. As a result, semiconductor manufacturers need business with a new group of customers beyond existing ones. To survive, vendors need to think beyond technology features they currently have. Usability and the ability to make platform with other vendors will play key roles for survivals.

Supply-Side Changes

Given the enormous capital expenditures and operating costs involved in semiconductor manufacturing, profitability of semiconductor manufacturers are in question. Today, the presence and importance of fabless vendors are increasing, such as nVIDIA, Qualcomm and broadcom. Financial reality and industry maturity will likely continue to drive the consolidations of IDMs into a few virtual IDMs and a few strong alliances. Most chip vendors will adopt the fabless model while a few big companies will remain as IDMs. This will inevitably lead to an increase in importance of foundries in the value chain. Semiconductor vendors are fighting among themselves to get priority status in foundry service companies, to ensure foundries make their chips as many as they need when market demand is high and capacity is tight. Semiconductor vendors will also have to make decisions on which industry participants to form alliances.

As industry costs continue to increase, the most critical issue to resolve will be which players are profitable enough to bear the burden of investing into capacity and R&D. If a major semiconductor manufacturer is driving to achieve large-scale adoption of new technology, the investment in process R&D will never produce a return. This burden has fallen into the left side of the value chain. Equipment makers and materials manufacturers have to take on major risks to keep up the state of the art technologies. For them, if a lead time for a new product development is too long, then they are exposed to higher risks. Thus, product development decisions on partnering with specific customers can have a grave consequence. Increasing costs are thus likely to further drive industry consolidation or powerful alliance of many guaranteed clients.

Faced with changes in both the supply and demand side of the semiconductor industry value chain, device vendors need a better understanding of upstream and downstream changes. They must make strategic business decisions on business models involving outsourcing, IP strategy, partnerships, target end users and markets. To survive, semiconductor vendors must adapt to changing dynamics of the value chain.

2.4 Incumbent Firms

Changes in Firm Behavior and Industry Structure

Semiconductor producers have utilized various strategies to meet increasing demand for flash memory and to obtain market share. At the beginning of flash memory growth, producers of flash memory had to decide whether to switch production from other nonvolatile devices to flash memory. When flash memory growth exploded in the beginning of 2000s, existing firms increased production and firms producing DRAM began non-volatile memory production. More recently, firms have partnered to gain an advantage in the competition. The following describes firm behaviors and considers impacts on the semiconductor industry.

Figure.2.14 shows market share of incumbent non-volatile memory players. For firms producing flash memory in the early 1990s, they were making their products over uncertainty of technology. That uncertainty makes decisions to enter a new market and produce new products. In addition, many companies had more immediate priorities on DRAM than to focus on new technology with little demand. Decisions to produce a new product are very simple: all in, partially in, and all out.

Intel, at that time the NO.1 Non-volatile memory maker, was one of the only manufacturers that decided to commit to flash memory business. In 1991, the company made a strategic decision to shift focus from EPROM to NOR flash memory. Other firms, such as AMD, decided to produce flash memory partially.

Some were motivated by Intel's announced go out of EPROM production, but they also do not want to stop development on NOR flash memory because of possibilities on higher densities and lower bit per cost. Eventually, some companies were unable to compete in the NOR flash memory market because they stopped the development. Intel's leap into the flash memory market proved critical in a technology that would soon dominate the nonvolatile memory market. By 1992, Intel had captured 75 percent market share of the flash memory market.

Toshiba, while it is the first adaptor of flash memory, had not been produced much market share because they focused on developing NAND flash memory. This is because they believed that NAND flash memory has better possibility than NOR flash memory in terms of scaling and cost per bit, and Samsung followed that strategy. In the early 2000s, with the advent of USB flash memory, memory cards, and digital music players, NAND flash memory took over the majority of market share. By 2005, Samsung and Toshiba took almost 50 percent of market share while Intel's market share has gone down to below 10 percent with disintegrations of its flash memory business.

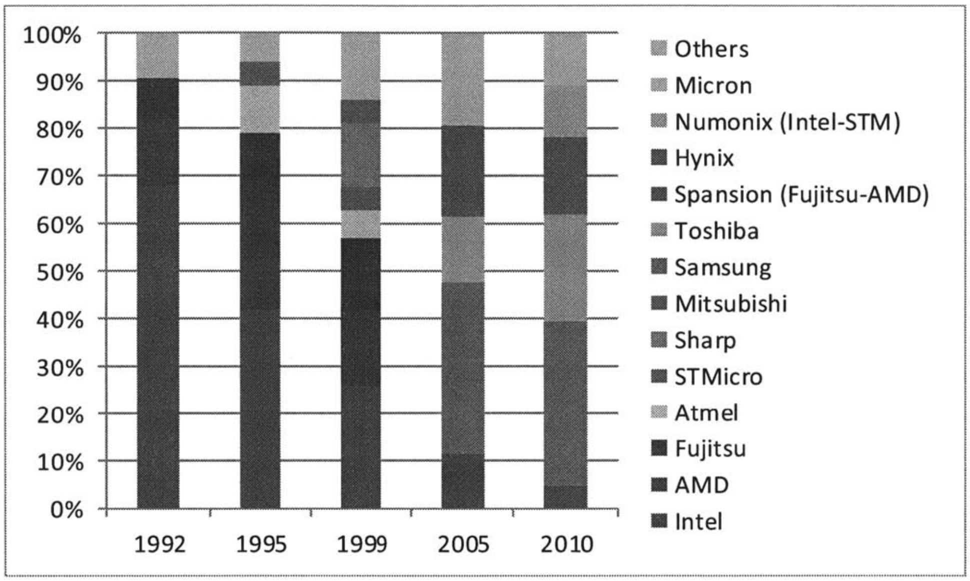


Figure.2.14 Market Share of Incumbent players in the Non-Volatile Memory Market

(source: iSupply)

2.5 Applications and Demand from End Users

According to Marco Iansiti, technological potential and technological yield can be measured. He introduced a concept of them, and established measurement methodology for quantifying the relative impacts on system-level architectures component.

Technological potential means an upper bound of feasibility in terms of performance, based on physical limitation of the chosen technologies. Technological yield defines product performance to technological potential of a product. In other words, it is a measure of effectiveness of product architectures.

Iansiti also defined product performance and project performance differently. Product performance describes the performance of a resulting physical product. Project performance describes the performance of a development team responsible for producing product designs, in terms of the technological yield that the product design team was able to produce in a given amount of time or with a certain amount of engineering effort.

Fine provides a description and examples of industry shifts from vertical integration to horizontal stratification and back, and proposes a "Double Helix" cycle as well as forces that drive the cycle. Fine also discusses the importance of supply chain design as a part of the overall product design process and as a core competence of the firm.

Historically, the growth of semiconductor demand is primarily fueled by new types of applications: personal computers, digital camera, mobile phones, and music players. Semiconductor manufacturers have provided products with higher performance to meet needs from end-users. While ASICs and microprocessors specialized in certain types of applications, discrete devices and memories are used in various types of applications, in other words those are commodities. This rise in the consumer electronics market has accelerated flash memory market growth and helped to make flash memory a prominent segment within the semiconductor industry.

Flash memory suits the consumer electronics market, because it contributes to two main trends in consumer electronics market: mobility and miniaturization. Continuous improvements in smaller die size, cost down, low power consumption, and higher performance have been leading the electronics and wireless markets. For example, cell phones, a major application of flash memory, require code and data storage to save and read data frequently. The performance improvements and other useful functions are now invading traditional hard dirk drive market. Following are main reasons for replacement:

1) Flash memory is smaller in terms of physical size
2) Less consumption power without mechanical parts
3) More reliable than magnetic storage technology

Because of the market size and future possibilities, flash memory has expanded its applications and has grown quickly.

3. Value chain of new memory technology

This section investigates the value chain dynamics of non-volatile memory to recognize potentials of innovations. This section also explores new technologies in memory products, and develops a value chain to predict the future revolution.

3.1 Critical steps for success

3.1.1 Value and innovation

Innovation is a term for new technological knowledge and market knowledge to provide a new product or service which customers will want. If a product is new, it means its cost is lower, its performances are improved, or it had never existed in the market before. As discussed so far, lower cost will bring more profits and higher competitiveness in the commodity market. New technology is a key to offer new products or services to customers. The understanding of linkage between technology and market is the first step to offer an innovation.

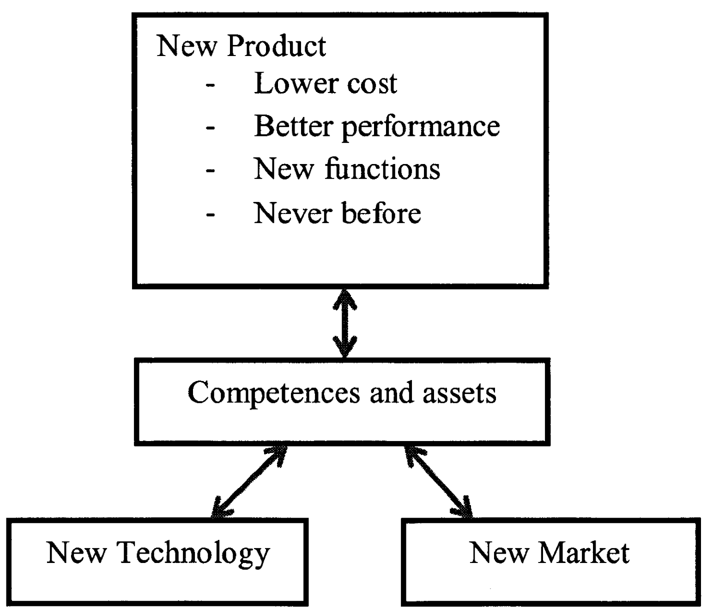


Figure.3.1 Market Innovation and Technology Innovation

3.1.2 Sustainable-disruptive innovation

Innovation can be defined as impacts on a company's capability because it implies a change in ability to offer new products. Innovation is disruptive if it is very different from existing knowledge and makes existing knowledge obsolete. On the other hand, innovation is sustainable if it is based upon existing knowledge. Most innovations existing in the semiconductor market are incremental. Figure.3.2 shows a classification of emerging memory technologies.

		Innovation	
		Sustainable	Disruptive
Type of firm	Incumbent	SONOS 3D-NVRAM	FeRAM MRAM ReRAM PRAM
	New entrant	ZRAM TRAM TTRAM	Nano-RAM Others

Figure.3.2 Sustainable vs Disruptive New Memory Technologies by Type of Firms

Although bit-per-cost is the utmost importance for those new technologies, new memory technologies also have to meet performance requirements because computer systems have increased its performance significantly in processing capability, such as multi-core architectures, high-speed I/O interfaces, and high-speed wireless communications.

3.1.3 Emerging memory technologies

There are many emerging memory technologies which are trying to replace DRAM and Flash in market. The ideal characteristic of new technology is to meet the performance of SRAM and the density of NAND Flash. The emerging mainstream memory technology should be multipurpose, higher performance and suitable for markets.

SONOS

SONOS stands for "Silicon/Oxide/Nitride/Oxide/Silicon", is a type of non-volatile memory which has silicon nitride (Si_3N_4) gate instead of poly-silicon gate. Data is stored into the nitride gate, and it provides lower programming voltages and higher endurance than conventional flash memories.

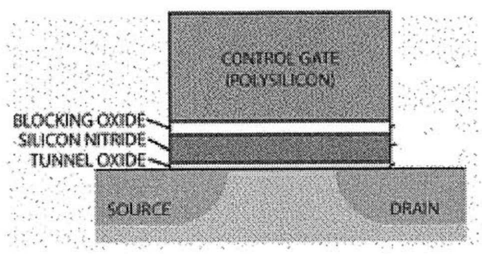


Fig.3.3 SONOS Cell Structure

(source: http://chipdesignmag.com/display.php?articleId=1752)

3D-NVRAM

3-dimensional non-volatile memory is one of the most promising candidates for ultra-high density storage devices. There are several candidates for the future three dimensional non-volatile memories, such as cross point cell, stacked two dimensional array structures, and bit-cost scalable flash. Although those technologies realize a multi-stacked memory array which enables continuous bit cost reduction, the cost to build one layer of memory array is still relatively high.

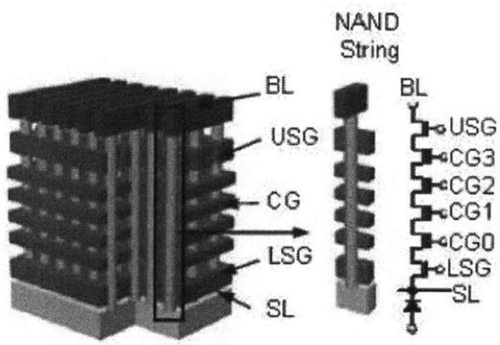


Figure.3.4 Cell Structure of 3-D Non Volatile Memory

FeRAM

Ferroelectric RAM, called FeRAM or FRAM, is a random-access memory which uses a ferroelectric gate cell instead of a poly-silicon cell. FRAM is one of a growing number of alternative non-volatile random-access memory technologies that offer the same functionality with flash memory. FRAM has some advantages over flash memory:

1. lower power usage
2. faster programming time
3. Greater endurance, more than 10^{16} times of program-erase cycles

Although it has some disadvantages, such as lower storage densities, capacity limitations, and higher cost, FRAMs are used as an embedded memory in microprocessors and ASICs as a secondary-cache.

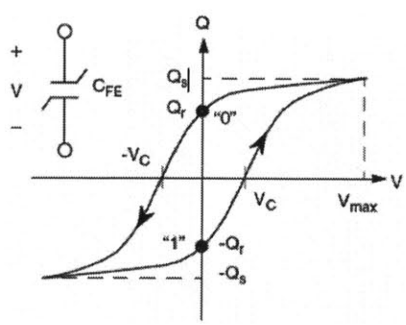


Figure.3.5 An Operation Concept of FRAM

MRAM

Magneto-resistive random-access memory (MRAM) is a non-volatile memory technology that has magnetic storage cells. It has several advantages in performances, and it is considered as an universal memory.

The two ferromagnetic plates hold the storage information. Each of those plates can retain a magnetic field, and an insulating layer separates them. A spin valve configuration, which composes the memory cell, is the featuring structure of MRAM. One of the magnet plates has a polarity, and the other's field can be changed to match the field to store data.

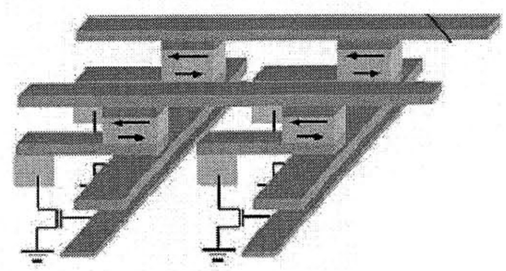


Figure.3.6 Cell Structure of MRAM

ReRAM

Resistive random-access memory (RRAM or ReRAM) is a new non-volatile memory type first developed by Sharp, and some of other types have patented. Different forms and dielectric materials of ReRAM have been developed so far.

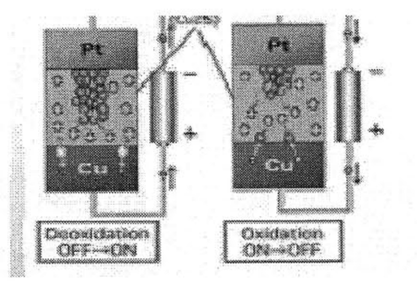


Figure.3.7 Cell Structure of ReRAM

PCRAM

Phase Change memory is a type of non-volatile memory which stores data into chalcogenide glass. The storage mechanism is similar to optical storage devices. An electric current generate heat to change the state of the glass to crystalline or amorphous. PRAM is one of universal memories close to stable mass production.

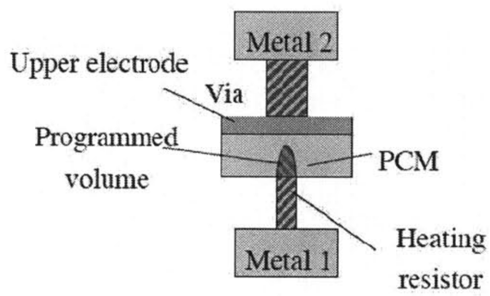


Figure.3.8 Cell Structure of PCRAM

ZRAM

Zero-capacitor RAM is a type of random access memory which stores data on the floating body effect of Silicon-On-Insulator (SOI) process technology. It provides memory access speeds similar to the standard SRAM cell used in cache memory, and ease of use on SOI process for microprocessors and ASICs.

Z-RAM relies on the floating body effect, a sub-product of the SOI process technology which places transistors in isolated layers. The floating body effect causes a variable capacitance in saturation region of a transistor, and is generally a problem for analog circuit designs.

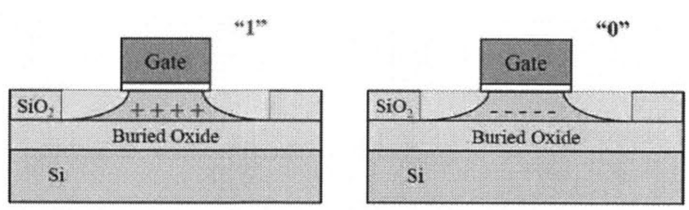


Figure.3.9 Cell Structure of ZRAM

TRAM, TTRAM

Thyristor RAM (T-RAM) is a type of memory which has an unusual structure of memory cells. In this technology, capacitively-coupled thyristor uses a negative differential resistance. It makes its memory cells capable of high densities. T-RAM is highly scalable than conventional.

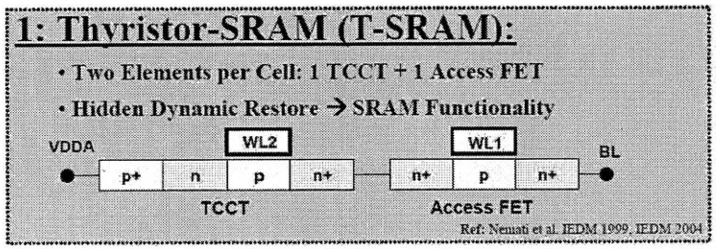


Figure.3.10 An Operation Concept of T-RAM

3.2 Emerging New Market

The semiconductor memory market has been expanding its own size by lowering its cost. Although there are many fields which require more storage data, energy consumption of semiconductor chips is another problem emerging in some applications.

3.2.1 Cloud service

Cloud computing is the delivery of computing as a service by sharing resources, software, and information. It provides software as a service, without requiring users to know the location and other details of the computing infrastructure.

In cloud service, users can access cloud based applications from a laptop or a smart phone to the data which are stored on servers at a remote location. This type of data center structure allows businesses to obtain applications easier and less maintenance. Thus it makes the internet environment more quickly adjust resources to meet uncertain end user demand.

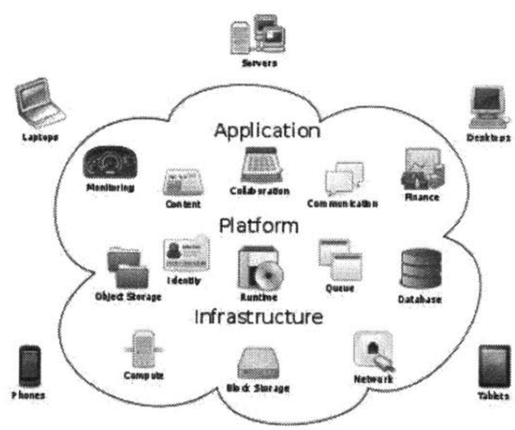


Figure.3.11 A Concept of Cloud Service

3.2.2 Energy in Data centers

A data center has computer systems and associated components, such as telecommunications and storage systems. It includes backup power supplies, data communications connections, environmental controls and security devices.

Many people are trying to modernize data centers in order to improve its performance and energy efficiency due to the increasing new internet technologies, equipment and capabilities, such as cloud computing. Data centers gradually transform themselves by standardization, consolidation, virtualization, automation and securitization. One way is to reduce the number of data centers and to replace equipment. This also includes reduction of the number of hardware, software tools and operation processes within a data center. Businesses replace old equipment with newer ones that provide better capacity and performance. Another way is to consolidate multiple servers. Virtualization helps to lower capital expenditures and operational expenses, and energy consumption.

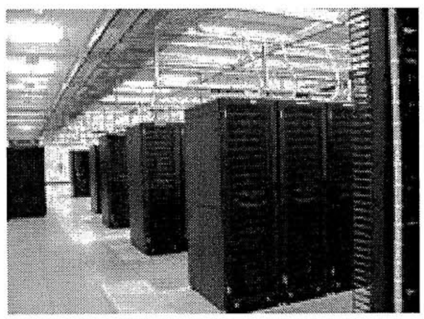


Figure.3.12 An example of a data center

In cloud computing, saving energy and increasing performance are critical issues because applications are provided as accessible data. The cost of running a data center is increasing and slowly dominating the total energy consumption. Cooling equipment, server components, processors, memories and hard disk drives are contributing to the total energy consumption. Servers and data centers account for 23% of IT energy demand. Data centers are one of energy demand where a lot of focus has been on to reduce the power, by introducing cooling systems, and improving energy efficiency of servers. Although hardware becomes smaller and less expensive, energy costs in servers or data centers are growing steadily.

Data center systems are usually designed to have significant margins to its maximum capacity. Typically, a stand-alone server will only operate at around 20% of its capacity limitation. In addition, emerging technologies place heavier burdens on processing and storage hardware. Accessing to huge data sets for each operation has huge workloads.

Optimization of servers is the main concern for improving energy efficiency. The efficiency of servers can be measured through diagnostics tests, and an energy demand model can be derived from the measurement result. The power consumption is directly related to the cooling cost. Thus reducing energy per work will bring significant cost reductions. Power management software can identify servers which are consuming more energy in a data center. It can also provide servers whose processes can be consolidated to improve efficiency, and predict cyclical power demand. The power management also provides which servers are in lower power state. This software approach helps manage energy consumption.

Consumption energy in data centers is contributing to costs of companies significantly. Figure.3.13 shows spending on servers from 1996 to 2010. Although the money spent on new equipment and infrastructures stays between $50 and $70 billion for the last 15 years, the spending on reducing power and cooling equipment goes up from $10 billion to $40 billion. Based on this data, reducing the cost for power and cooling is the utmost importance in data centers and state of the art internet technologies.

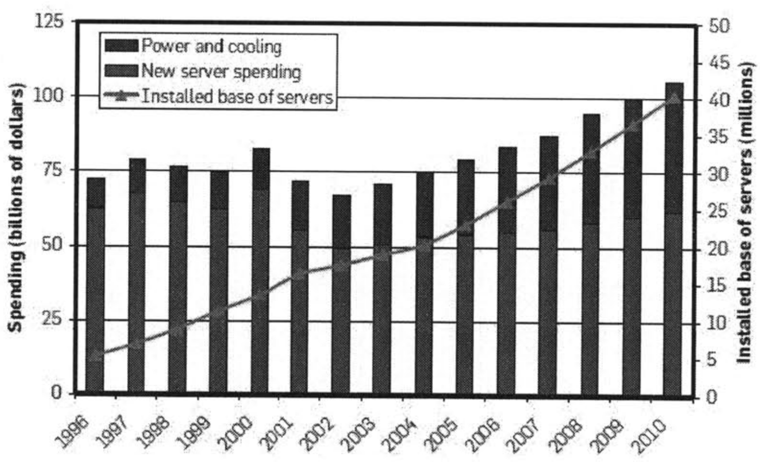


Figure.3.13 The History of Spending on Power and Cooling, and New Servers

3.3 Memory Value Chain for Cloud Computing

In this chapter, we have discussed emerging technologies and market. This section focuses on how to provide values of next generation memory technologies to the cloud service, especially focusing on how to deliver a low power solution to the cloud computing industry.

3.3.1 Non-Volatile memory for servers

Historically Hard Disk Drives have been the major storage device for servers in data centers. HDDs have competitive bit-per-cost and reliable magnet storage.

However, NAND flash based Solid State Disks have been in market for several years, and being implemented to servers to replace HDDs. Flash memory can reduce the energy consumption because it does not have mechanical components. The use of flash memories for both main memory and permanent storage reduces energy consumption and data transfers between layers of hierarchy. Although processors dominate system power consumption, data storage devices are as important as processors because those are working together to process vast amount of data. Servers can benefit from integrating flash memory instead of using HDDs in

1) Reducing stand-by power
2) Improving throughput by reading and programing

A drawback for the use of flash memory as a storage is its programming endurance. Flash memory has limitation for its program/erase cycles, which means it is not programmable forever. It allows about 10,000 times for program/erase operations, but after that the operation is not guaranteed. This is due to the endurance of the oxide layer in flash memory cell. The oxide layer will not function as an insulator after 10,000 times of program/erase operations because an electron leaks into poly-silicon gate. Although software techniques, such as wear leveling and fault tolerance, can extend the life of memory cells, there is a certain structural and physical limitation for flash memory.

Power consumption of memory and disk takes about 40% of overall power consumption in data centers. In addition, the current trend says that this will continue to go up by improving performance and throughput. New technologies are expected to reduce energy consumption, and costs at data centers.

Emerging technologies, such as PCRAM and MRAM, have higher possibilities for scalabilities and lower energy than flash memory. Although most of those devices also have program/erase limitations, it is better than flash memory. In addition, bit-per-cost is the most important point for emerging memory technologies. Some technologies, for example PCRAM is in market, but its price is still higher and its density is lower than flash memory because the technology is still premature. Many companies have been putting a lot of efforts on improving the performance of flash memory, so new technologies cannot catch it up easily. However, since the limitation of flash memory is obvious, development of new memory technologies is unavoidable.

3.3.2 Supply chain structure

For the study of value chain for cloud computing memory, supply chain analysis is necessary to know how to provide the value from new technology development to cloud service providers. Figure.3.14 shows the memory supply chain for cloud computing. Since the growth of DRAM market has been stagnant for years, non-volatile memory is an essential player for next data storage era.

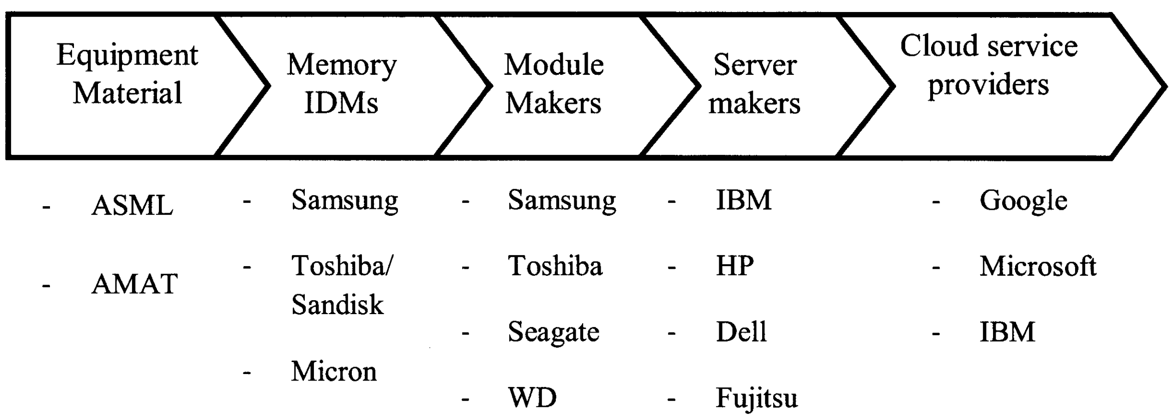


Figure.3.14 The Memory Supply Chain for Cloud Computing

Semiconductor equipment market has been dominated by Applied Materials, but recently ASML has been growing its market share because of its immersion lithography technology. Currently NAND flash memories are primarily provided by Samsung and Toshiba/Sandisk, but the next generation technology providers are expected to replace them. Memory module makers, currently its equivalent to SSD makers, integrate a number of flash memory chips and provide storage systems to server makers. Although flash memory makers, Samsung and Toshiba also makes SSDs, traditional HDD makers, such as Seagate and Western Digital, also entered SSD market as well as supplying their own hard disk drives. Server makers, such as IBM, HP, and Dell, implement SSDs into their own servers to improve performance and power consumption of their servers. Finally, cloud service providers buy those servers with SSDs, and their data center performance will be improved and equipment costs for cooling will be reduced.

3.3.3 Performance comparison

Figure.3.15 shows a performance portfolio for incumbent technologies and emerging technologies in terms of scalability and power consumption. In this figure flash memory is placed on the center of diagram for the comparison with other technologies. Hard Disk Drives have better scalability than other incumbent and emerging technologies, but its high power consumption is a drawback for using HDDs. Although 3D flash memory is one of the prospective technologies because of the ease of migration from flash memory, its power consumption is still comparable to flash memory.

Therefore, if we focus on reducing power consumption at data centers, the best candidates for disk storage are RRAM and PCM. In addition, MRAM and FRAM can also be considered as a cash memory for microprocessors. All of them use new materials for its memory cells, unlike flash memory. Thus for the next generation memory development, new material is a key for reducing power consumption.

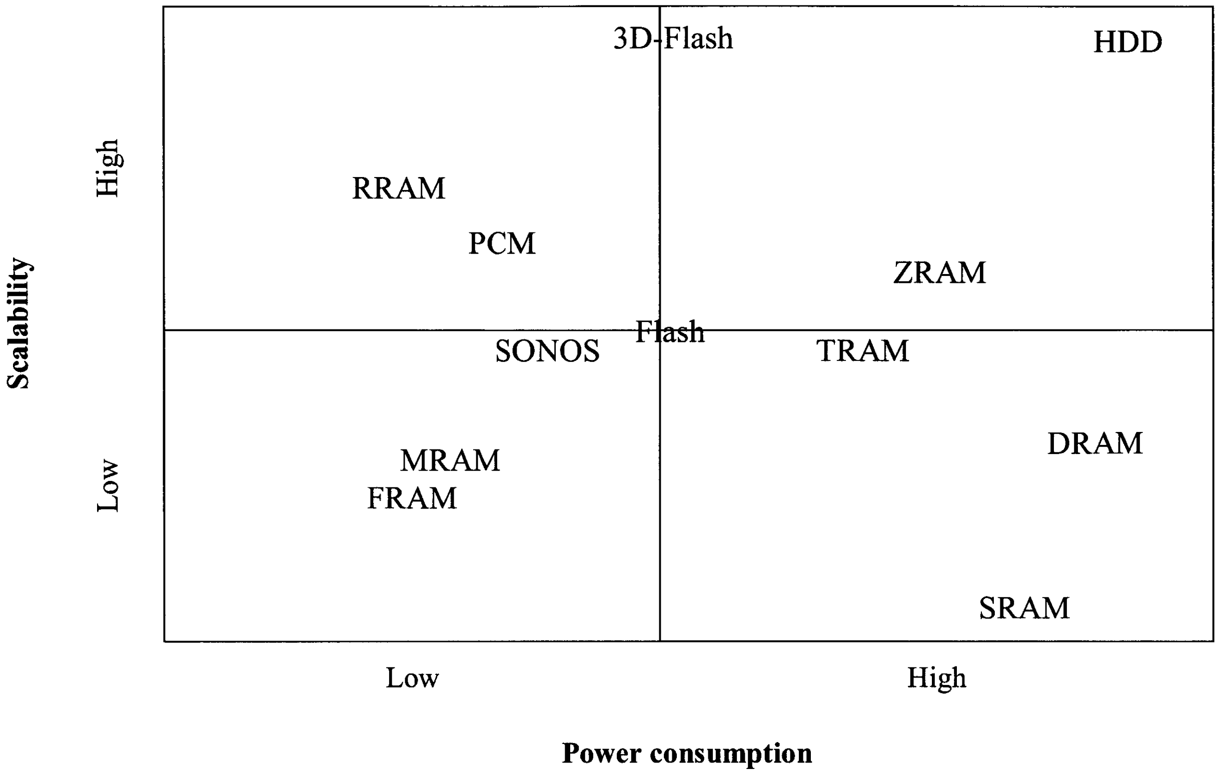


Figure.3.15 Technology Portfolio for Emerging Memory Technology in view of Scalability and Power Consumption

4. Creating changes in a competitive market

This thesis investigates non-volatile memory industry and hypothesizes the future growth of the industry. The memory market is considered as a commodity market. There are three important factors that affect economic activities: economic cycles, price competition and technology evolution. Following system dynamics models are developed in order to understand the causes and effects in the industry based on historical data in the memory market. The purpose of these models is to grasp dynamics of the industry with the simplest structure.

4.1 Approaches for a deflating market

4.1.1 Modeling approach

In general, semiconductor memory products are commodity products. The market is routinely repeating a downturn and an upturn with uncertainties and price competitions. Players in the market make wrong business decisions because of undesirable financial results, market bubbles, excessive investments, and fast commoditization. To understand the dynamics of the market accurately, a model should provide realistic future outcomes as a result of management decisions: investments, technology choices and customer demands.

Figure.4.1 gives an overview of a causal loop for the memory market. The model is designed to provide a hypothetical firm operating in competition with other companies in the industry. Strategic forecasting models are not intended to be used in a particular firm or a specific product. The fundamental concept of the figure is how demands and prices move in accordance with investment decisions on production capacities and technology. Investment decisions of market players affect future demand and price. Apparently the most important element in a commodity market is the price. The market price greatly affects all the company's revenues and profits, thereafter investments decisions. Investment decisions are based on profits of the last year and a demand forecast for the next year. This decision is very important as a manager of a company because if managers overestimate the demand and invests into production excessively, the company eventually suffers from overcapacity. However, if managers did not invest into production and capacity, they cannot supply new products with competitive prices and enough quantities when the demand is high. The market price is one of the most important exogenous factors, and the investment decision is a critical indigenous factor.

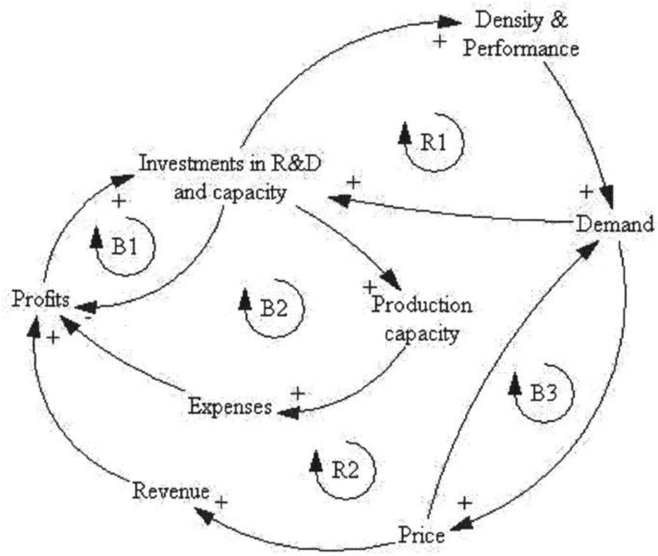


Figure.4.1 Causal Loops for the Memory Market

4.1.2 Demand and capacity effects

In commodity markets, price is subject to supply and demand. If the supply does not meet the market demand, the price will go up because many customers want products even though the price is higher than an usual level. If the supply is excessive to the market demand, the price will go down because many companies want to sell their own product although a margin of the product is relatively low. Companies want to reduce their inventory as much as possible to reduce holding costs.

Demand and supply can be considered as a function of the price in commoditiy markets. However, the demand is affected not only the price but also many other factors: market condition, political situation, and branding. The capacity depends on how much a company invested into capical expenditure. It is not a complicated function. A company can buy expensive equipment as much as they want if there is enough money.

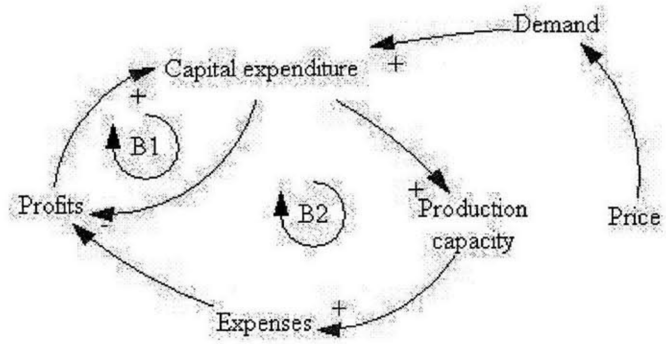


Fig.4.2 Causal Loops for Demand and Capacity Effects

4.2 A Model structure for Evolution Scenario

4.2.1 Basic simulation model

Figure.4.3 shows the simulation model structure for the memory market. The overall architecture is based on a model in commodity markets because memory products are a type of commodity products. It is also a cyclic market with steady growth: demand repeats an upturn and a downturn while the size of the market is basiTacally growing. An investment decision on capex affects the size of capacity. Managers have delays from profit to capex because they need to understand a trend of the market and make decisions periodically. After investing into capex, there is implementation time from capex to capacity because they need to buy equipment, build a new factory, and hire new employees to expand the capacity. Since technology is not useful forever, there is a retirement rate for the capacity. Old products are replaced by new products, so they will lose market values eventually. A company always needs to invest into capex because a certain rate of capacity is retiring at a certain rate.

If demand growth accelerates, managers will expect a higher growth rate in the future. They invest into R&D and capacity to fulfill orders from customers. When the new capacity is ready to provide enough supplies to the market, there is no guarantee that customers still want same products with the same price and quantity. In the model, the demand is an exogenous factor. Thus it is independent from capacity because in a commodity market, price is decoupled from supply side if a competition is very intense.

Table.4.1 Simulation Parameters for the Memory Market Evolution Model

Initial Capacity	10,000 (pieces)
Technology Life Cycle	18 (month)
Initial price	5 (dollars)
Implementation time	6 (months)
Holding cost	0.2 (dollar/pieces)
Investment rate	0.3 (dmls)

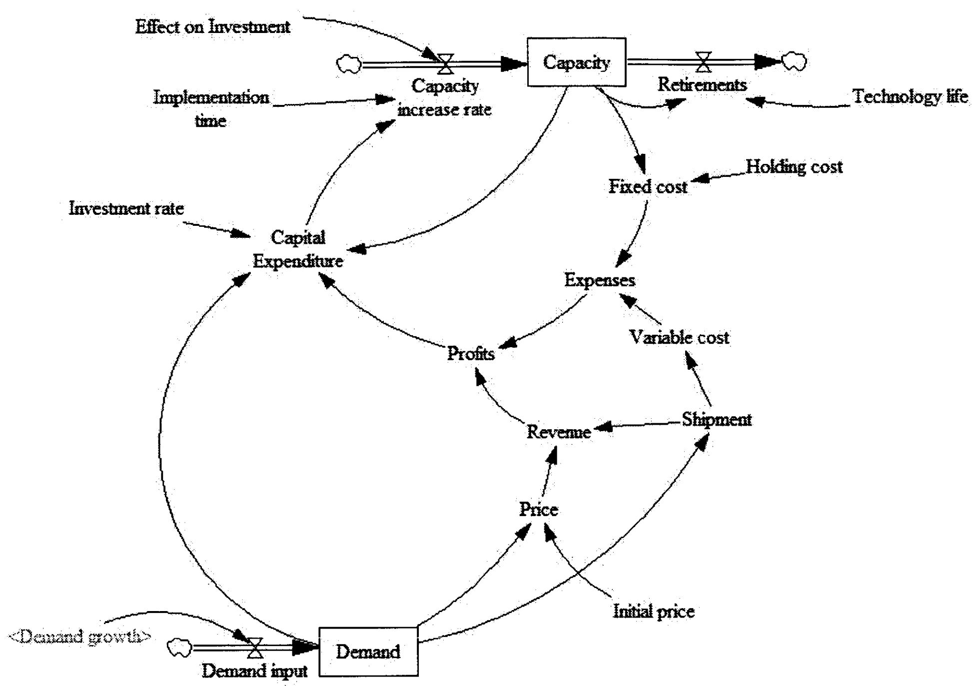


Figure.4.3 The Simulation Model for Base Case

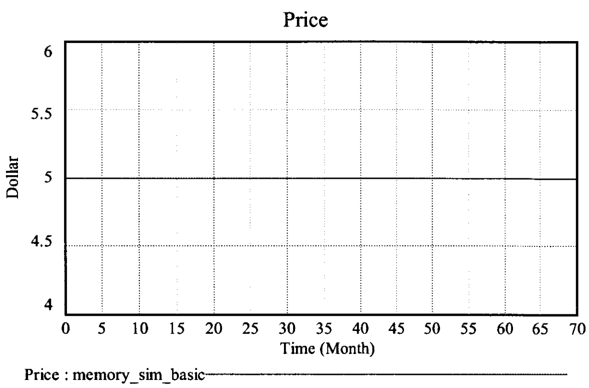
4.2.2 Simulation results with basic model

Figure.4.4 shows simulation results using the model in Figure.4.3. In this simulation, price is fixed to \$5, and demand is also fixed to 30,000 as shown in Figure.4.4 (a) and (b).

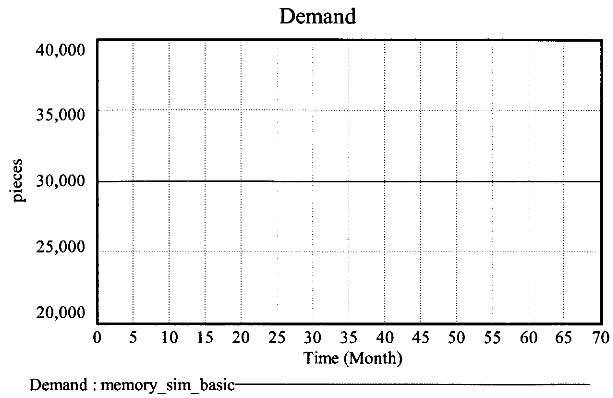
Figure.4.4 (c) shows shipment and capacity. In the beginning shipment goes up with capacity because the capacity is lower than the demand. When the capacity reaches the demand, the shipment is tied to the demand but the capacity keeps growing because there is still investment in capex.

Figure.4.4 (d) shows expenses. The variable cost and fixed cost go up while the capacity is lower than the demand. On the other hand, if the capacity exceeded the demand, the variable cost is not going up anymore because customers do not buy products. The fixed cost keeps growing because of excessive capacity.

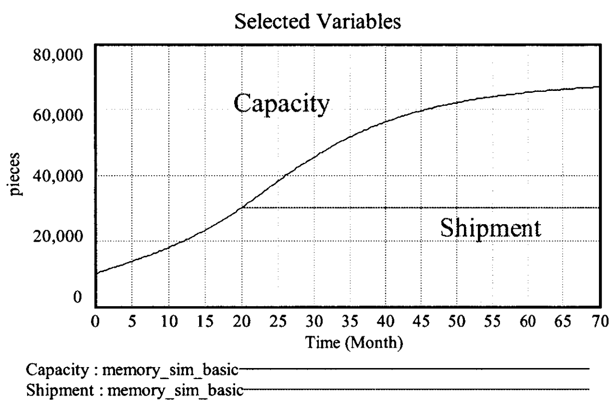
Figure.4.4 (e) shows profits and capital expenditures. In this simulation, 30% of profits are allocated to capital expenditure. Profits and capex are growing while the demand is higher than the capacity. The profits are slightly decreasing because of the increase in fixed cost.



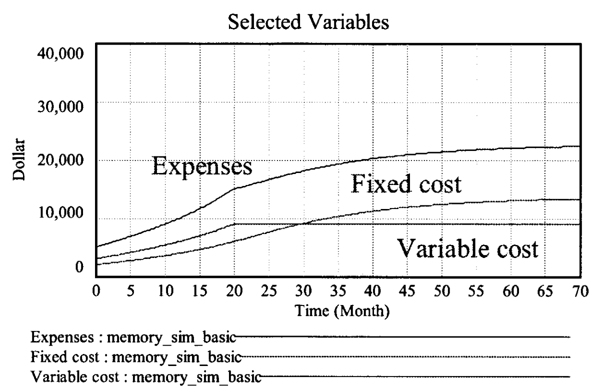
(a) Price



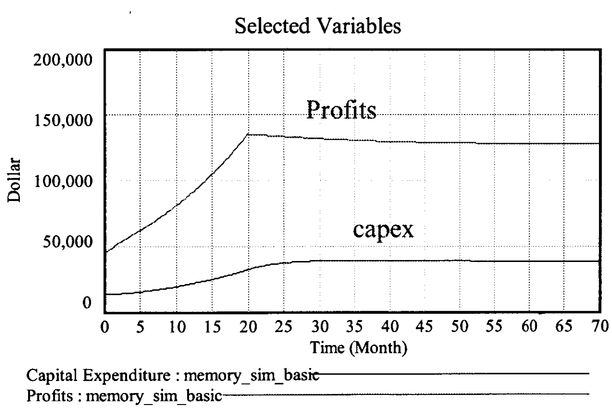
(b) Demand



(c) Capacity/Shipment



(d) Expense



(e) Profits

Figure.4.4 Simulation Results with constant price/demand

4.3 Investment decisions

4.3.1 Investment decision model

An investment decision is one of the most important decisions in a fast changing commodity market. A capacity expansion plan affects market price, and eventually company's profits. There are always uncertainties for future demands and technology directions. Although no market is going up forever, managers want to see future growth in their own company. Thus they spend too much profits on investment. If managers invested into assets and new technologies excessively, the company will suffer from overinvestment and overcapacity. However, if the managers did not invest into new technologies, the company will not be able to supply enough chips for customers in future. The investment into capacity is necessary, but it should be an appropriate amount to sustain a company.

Figure.4.5 shows the simulation model structure for the memory market. The overall architecture is still same with Figure.4.3, but an investment decision function has added to the model. The decision criterion is simple as follows:

If Capacity = shipment, then capex = 30% (invest 30%)

If Capacity > shipment, then capex = 0% (don't want to invest)

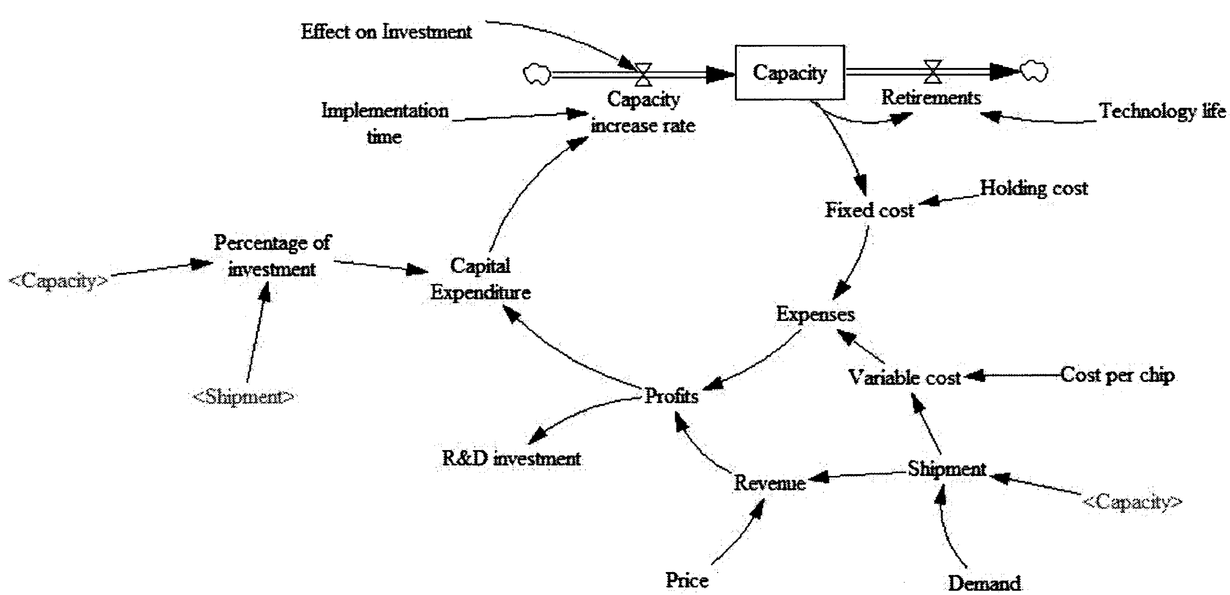


Figure.4.5 The Simulation Model with Investment Decisions

The reason to use capacity and shipment for the investment decision is that it is impossible to know actual demands in market. Generally speaking, demand is willingness of people to pay for a product. It is not quantifiable, thus the one possibility to know the demand of a product is to see the number of shipment. If there is enough number of shipments, there is more demand for the product.

Although managers do not want to invest into capacity if the capacity is higher than the shipment, the company stops capex immediately. However, there is a decision delay from profits to capex. The capex will gradually go down if the capacity is higher than the shipment.

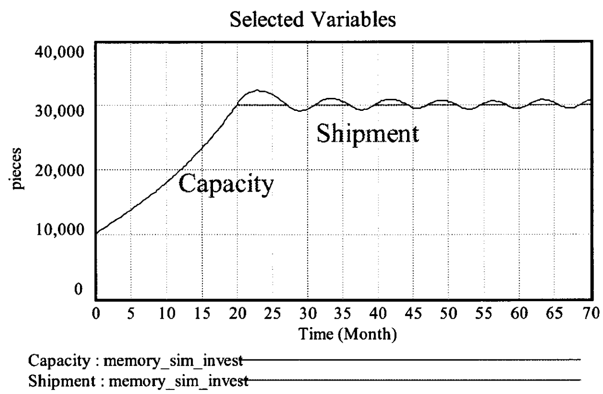
4.3.2 Simulation results with investment decisions

The model with investment decisions is tested to see how investment decisions affect future evolutions in the market. In this simulation, price is fixed to \$5 and demand is fixed to 30,000 to make a comparison with the basic model.

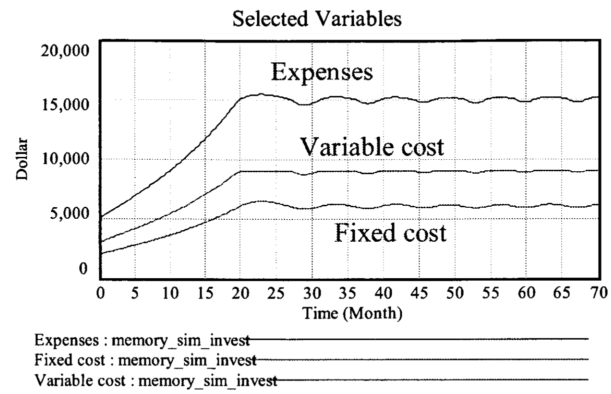
Figure.4.6 (a) shows shipment and capacity. When the capacity reaches the demand, the capacity overshoots because of decision delay, but it goes down to lower than the shipment because of retirement rate of the capacity. Eventually the capacity settles down to the same value with the actual demand, 30,000.

Figure.4.6 (b) shows expenses. The fixed cost stops at a value of (Capacity x holding cost) because the capacity settles to 30,000.

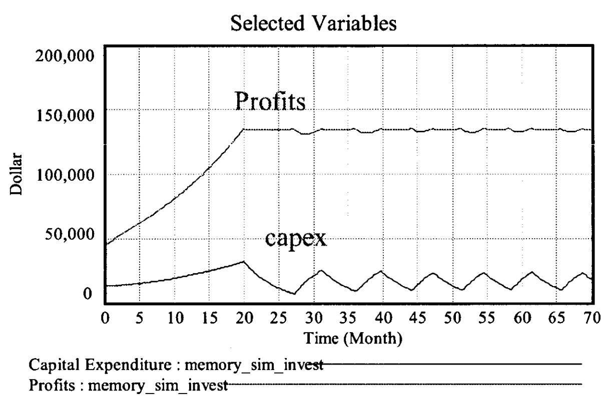
Figure.4.6 (c) shows profits and capital expenditure. If the capacity is lower than the shipment, the capex goes up gradually to increase the capacity. However, if the capacity exceeds the shipment, the capex goes down slowly to save the investment. Capex is always necessary to make up for capacity losses by retirement rate.



(b) Capacity/Shipment



(b) Expense



(c) Profits

Figure.4.6 Simulation Results with Variable Investment Decisions

4.4 Price Declining Model

The average price of semiconductor memory is usually dropping day by day. Although the initial price of the highest density has the highest price, semiconductor memory will be commoditized quickly because other competitors catch up the state of the art technology immediately.

Figure.4.7 shows flash memory pricing trajectories from 2009 to 2015. This data primarily shows NAND flash pricing for SSD. In 2009, the overall flash memory price was $5/Gigabyte. At the end of 2011, the price has been down to $2.2/GB. Based on history and experience in the memory market, the price is expected to go down to less than $1 in 2014 or 2015.

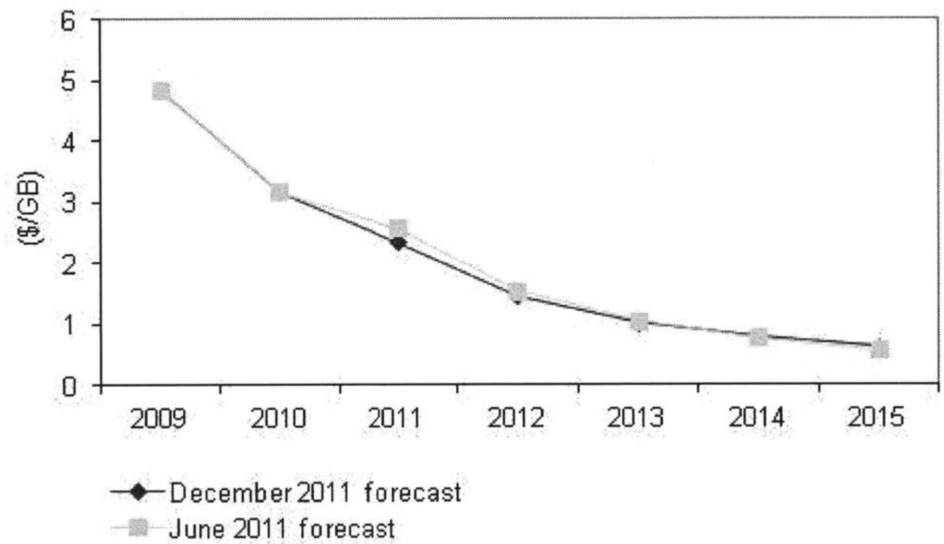


Figure.4.7 History and Forecast of Flash Memory Pricing

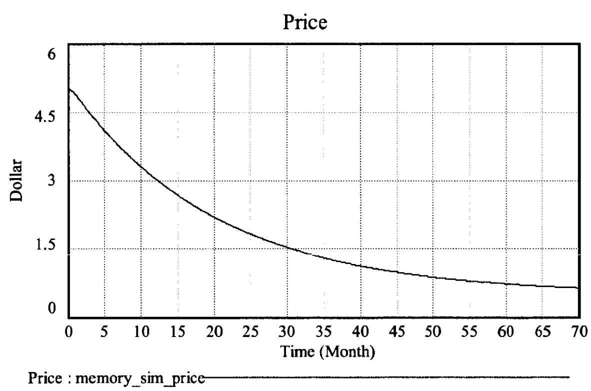
(Source: IDC, 2011)

Figure.4.8 shows simulation results using the model in Figure.4.7. The price trajectory in Fgure.4.8 (a) follows the price in reference data from IDC. It starts from $5 in 2009 and goes down to $0.50 in 6 years.

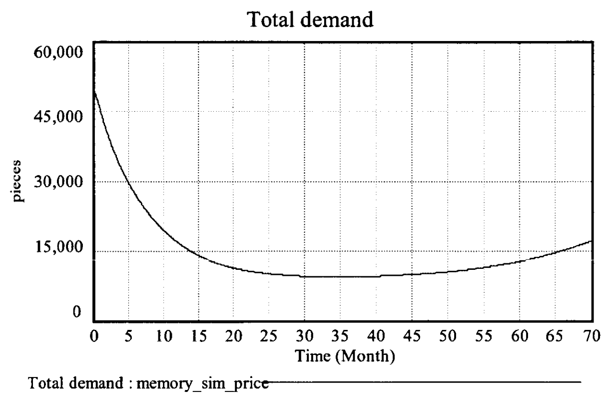
The demand quickly goes down as shown in Figure.4.8 (b). This is because in this model, the demand is calculated by price/capacity. In the beginning capacity goes up to meet the initial demand, but the effect of decreasing price and increasing demand causes steep declines in demand.

Figure.4.8 (c) shows capacity and shipment. The capacity catches up the shipment after 12 months, but the capacity overshoots because the demand is going down quickly. Capacity and shipment eventually becomes the same level in 45 months.

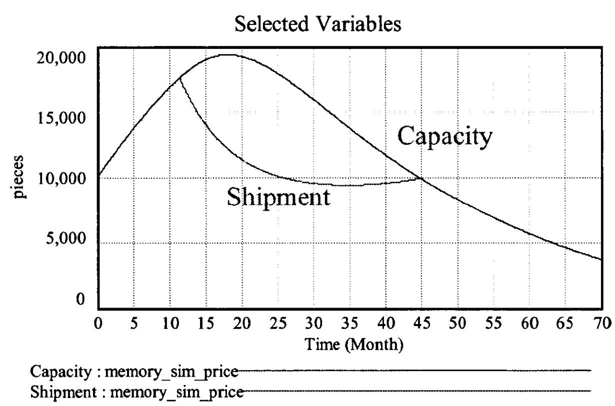
Figure.4.8 (d) shows profits. Initial profit is very high, but after 12 months profits goes down steeply because of excessive capacity, declining demand, and decreasing price.



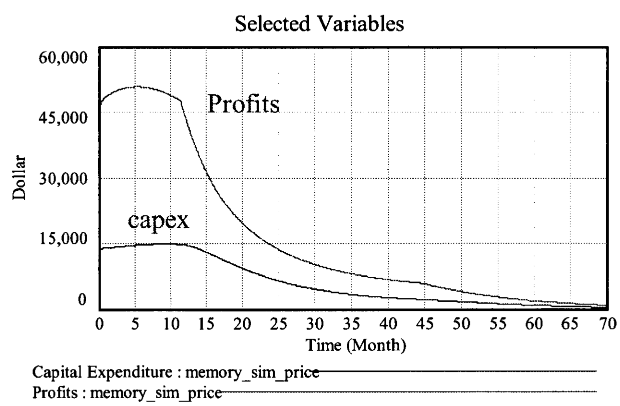
(a) Price



(b) Demand



(c) Capacity/Shipment



(d) Profit/capex

Figure.4.8 Simulation Results with Declining Price

4.5 Technology effect on cost

In the memory market, every company is struggling to reduce the size of their own chips. The size of a silicon chip directly affects yields from a silicon wafer. Although there are customer requirements for performances, functions, and power consumptions, customers basically just worry about price. Thus the cost-per-bit is the most important factor in the memory market. Technology evolution is important to reduce costs because smaller technology nodes lead smaller die size, which means more chips from a silicon wafer. Because the price goes down day by day, engineers and researchers need to develop new technologies which can reduce total cost of memory. Figure.4.9 shows the causal loop for technology effects on cost.

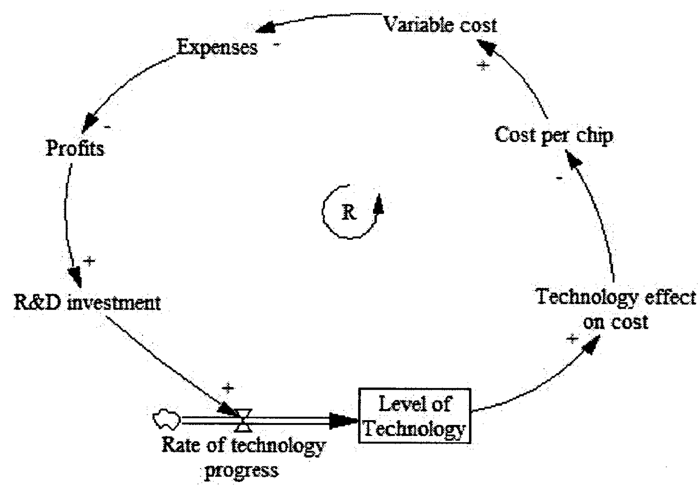


Figure.4.9 The Causal Loop for Technology Effects on Cost

Figure.4.10 shows a simulation model for investigating technology effects on cost reduction. This model is based on the model in Figure.4.3. All the companies are investing on the state of the art technologies to make smaller nodes. Investment into developing smaller nodes is inevitable for memory manufacturers. Otherwise the company is behind of their competitors in terms of cost competitiveness.

If profits are high, it is easy to invest into R&D and capacity. It is simple because they have enough cash on hand. After a certain investment, this model assumes that new low cost technology will be developed. The development of low cost technology reduces the cost per chips, thus variable cost will be lower. It eventually leads to lower expenses for a company.

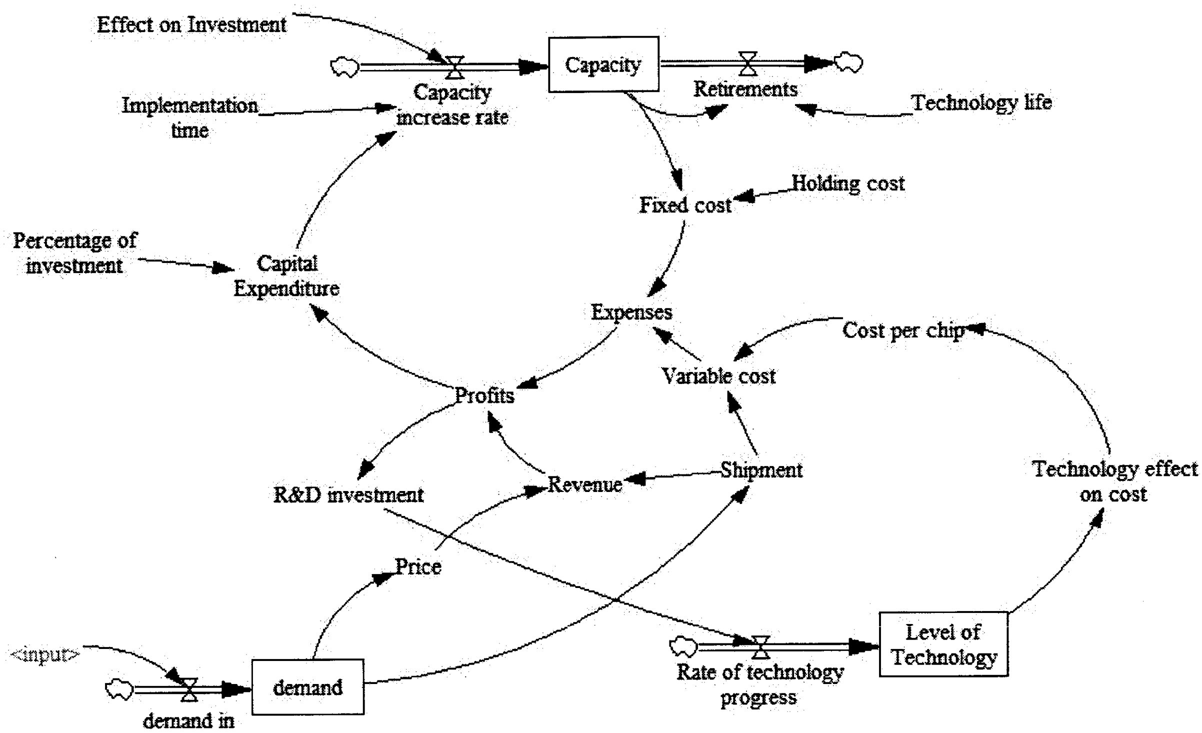


Figure.4.10 The Simulation Model including Technology Effects on Cost

The level of technology is a maturity rate between 0 and 1. It indicates how much technology is matured as a result of investment. Figure.4.11 shows the relationship between R&D investment and technology maturity. The model assumes technology has a trip point somewhere in an early stage of development and a saturation point somewhere in a later stage of development.

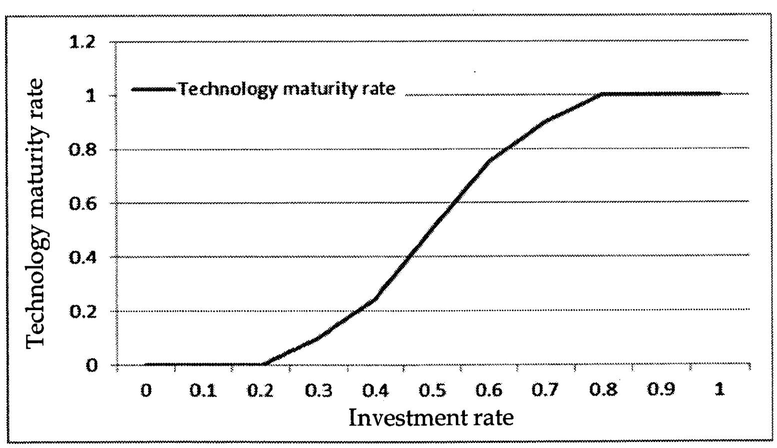
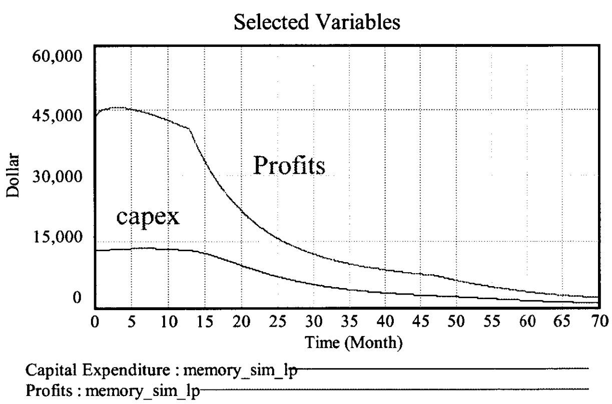


Figure.4.11 The Relationship between Investment and Technology Maturity

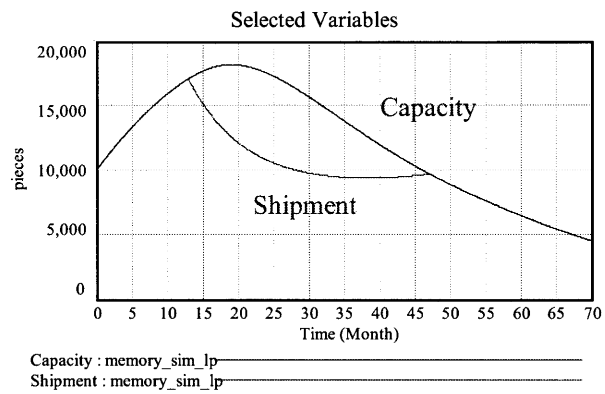
Figure.4.12 shows simulation results based on the model in Figure.4.10. The price and demand profile is same with Figure.4.8 (a) and (b), which follows the price in reference data from IDC.

Figure.4.12 (a) and (b) show profits and capacity. The curves are very similar to the simulation results in Figure.4.8, but are slightly higher because of lower cost per chip.

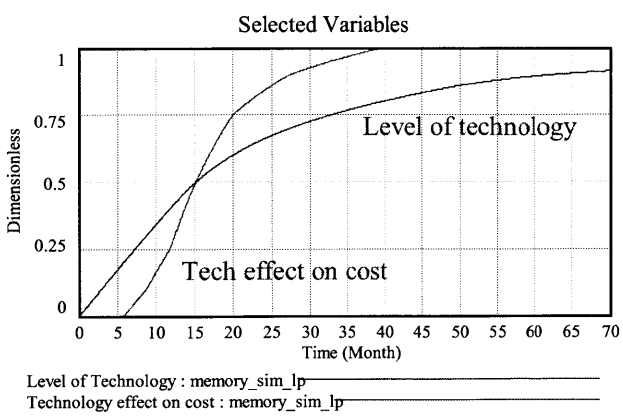
Figure.4.12 (c) and (d) show technology rate and cost per chip. In this simulation, the costs are set to go down from \$0.6 to \$0.1 per chip because of the investment into technology.



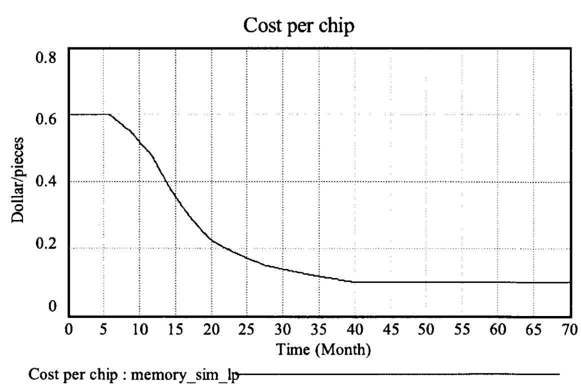
(a) Cost and profits



(b) Capacity and Shipment



(c) Technology rate



(d) cost per chip

Figure.4.12 Simulation results of Technology Effects on Cost

4.6 Technology effect on low power for server market

Chapter 3 concluded that low power products for servers are a solution for the next growing market. By reducing energy consumptions in data centers, cloud service providers can reduce their operating costs. Therefore, memory manufacturers should invest in development of low power products to expand the overall market size.

Figure.4.13 shows a causal loop for technology effects on low power technology. The model assumes that R&D managers made a decision to invest into new low power technology. If low power technology matures enough to be in the market after a certain amount of investment, the demand will go up because a new market has developed. The bigger demand will cause higher revenues, and it will also cause a higher profits.

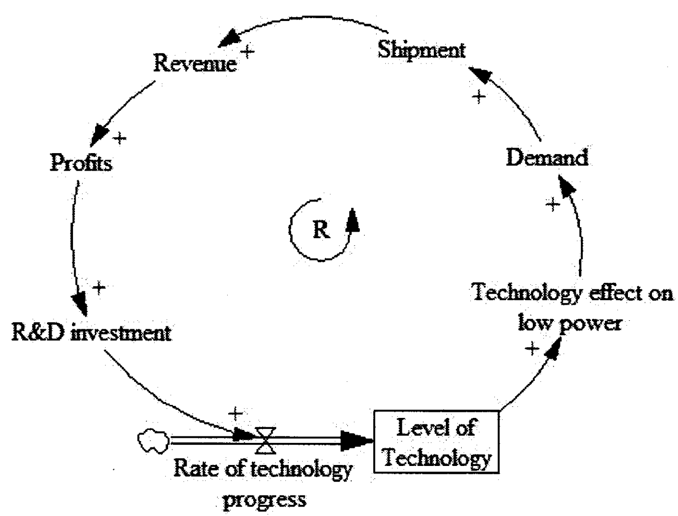


Figure.4.13 The Causal Loop for Technology Effects on Low Power

Figure.4.14 shows a simulation model for investing in low power technology. The new low power technology affects to total demand, which is the sum of new demand and existing demand. The R&D investment is determined by the profits. In this simulation it is 30% of profits.

Figure.4.15 estimates demands of flash memory by applications. According to IDC, flash memory shipments for servers will be about 100 million pieces in 2015, while shipments for enterprises and commercial will remain less than 20 million in total.

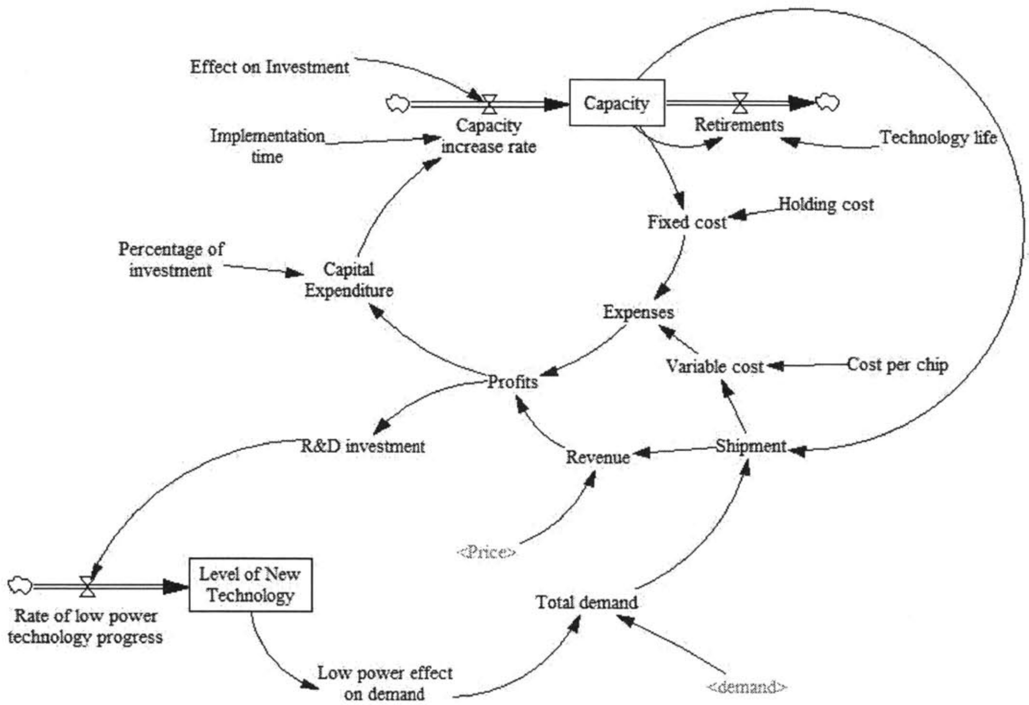


Figure.4.14 The Simulation Model including Low Power Effects

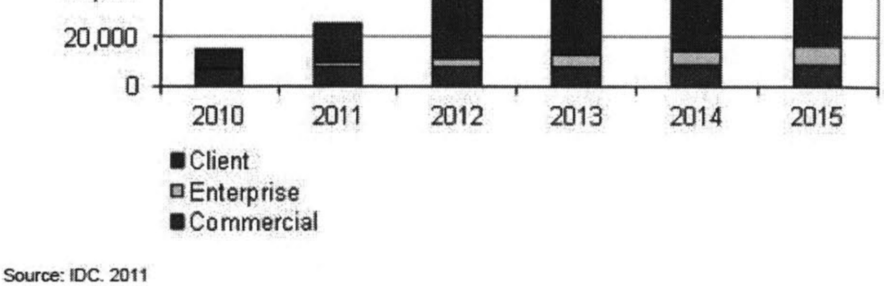


Figure.4.15 Market Estimation on Flash Memory Shipment by Application

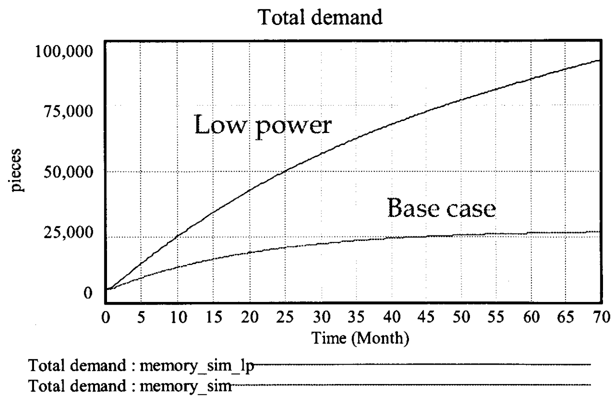
Figure.4.16 is simulation results based on the model in Figure.4.14. Figure.4.16 shows 2 simulation results: One is with low power solution, and the other is without investment in low power technology.

Figure.4.16 (a) shows the demand profile of low power effects. These curves follow the market growth estimation in Figure.4.15. If there is no low power solution for servers, there are markets for enterprise and commercial only, which in total are about 20,000 demands. However, taking into server market in demand, the total demand will be around 100,000 in 2015.

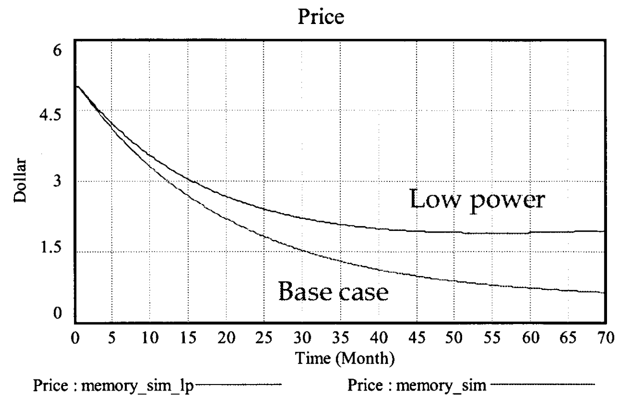
Figure.4.16 (b) shows price profiles of the low power effects. Because the demand went up, the price remains higher than no low power solution.

Figure.4.16 (c) shows profits. Due to the increase of price, the profit is also increased. Figure.4.16 (d) shows capacity. Capacity is increased because of more investments from profits. However, it does not meet all the demand. Investments should be higher than 30% if the company needs to make more revenues and profits,

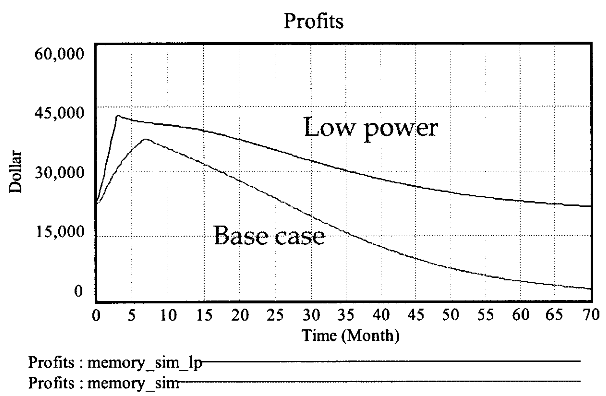
Figure.4.16 (e) shows technology effect on demand. In this simulation, the new technology development affects to the demand almost linearly. The effect on demand almost linearly goes up as the company invests into low power solution.



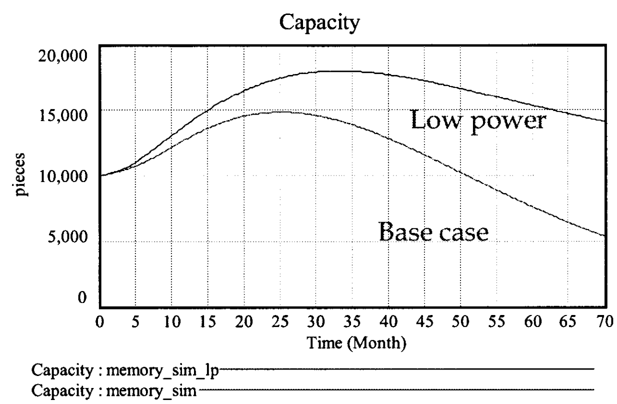
(a) Demand



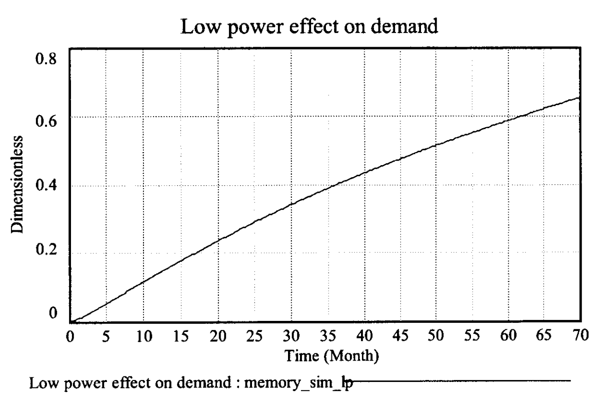
(b) Price



(c) Profits



(d) Capacity



(e) Technology effect on demand

Figure.4.16 Simulation Results of Low Power Effects

4.7 Evolution cases

In this section, several demand growth patterns are tested to see how the system dynamics model responds to various types of market demand. Although the market estimation is shown in the last section, demand is not predictable in nature. Many types of demand evolution scenarios should be considered to respond to unexpected market situations.

4.7.1. Linear growth

Figure.4.17 shows a simulation model for different cases of demand. In this simulation model, all models discussed so far are included: investment strategy, cost reduction, and low power technology. The demand is now an exogenous variable to test several demand growth patterns. The company is investing in low cost technology and low power technology simultaneously. The price is equal to demand divided by Capacity.

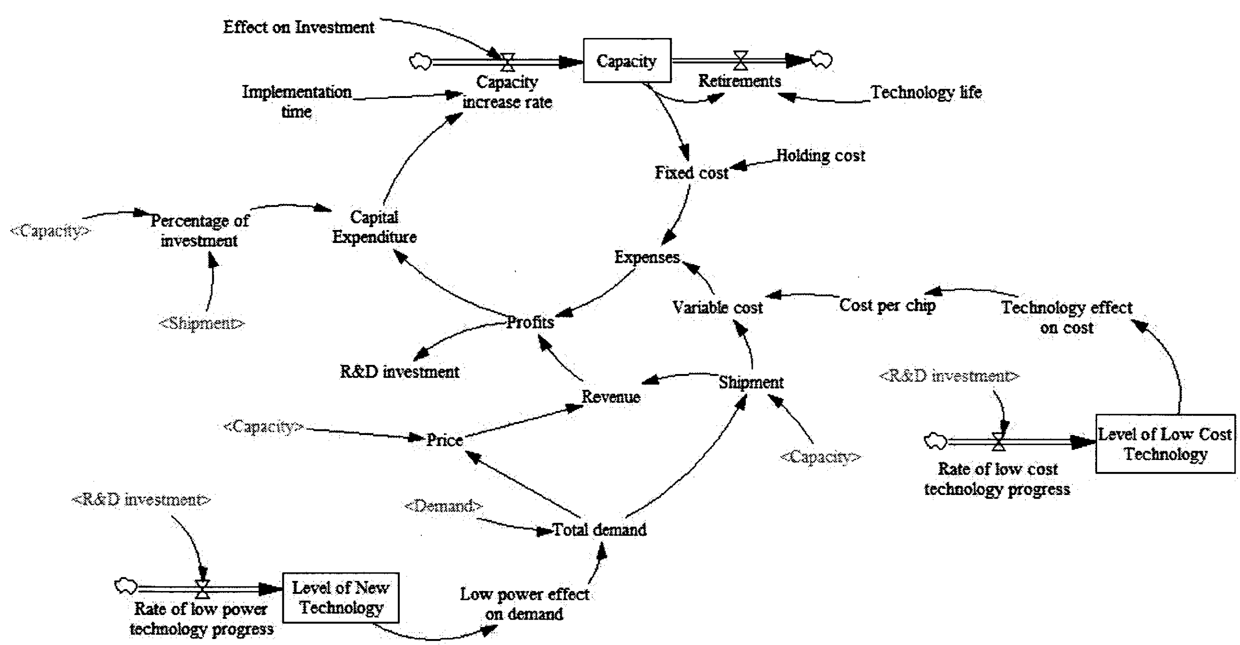


Figure.4.17 The Simulation Model for Demand Growth Patterns

Figure.4.18 is simulation results based on the model in Figure.4.17. Figure.4.18 shows 3 simulation results for three types of demand growth profiles.

Figure.4.18 (a) shows demand growth patterns. There are three patterns: (1) is a linear growth pattern starting from 0, (2) is a slow growth pattern with initial offset in 30,000 demand, and (3) is linear declining pattern from 60,000.

Figure.4.18 (b) shows price trajectories. In (1), the price goes up because demand goes up and capacity goes down from the initial level, but after an overshoot it will stay in the range between $3 and $4 because of the increasing demand and capacity. In (2) the price stays around $3 because of constant growth in demand and capacity. (3) shows a quick price decline in the beginning, and after that price goes down constantly due to the decrease in demand.

Figure.4.18 (c) shows profits of three simulation results. (1) does not make a profit in the beginning because it does not have any demand. The profit grows until the shipment reaches the capacity. In (2), the profits increase steadily as demand grows. In (3), the profits go down linearly with some fluctuations. Although some disturbances are observed, the profits of three simulations are very similar to demand profiles. It means that the profits are primarily affected by demands, rather than price fluctuations

Figure.4.18 (d) shows capacity and shipment. In (1), the capacity goes down from the initial level because there is no demand. When the demand catches up the capacity, it will maintain the same level for about two years, but after that it will go up because the increase in demand. (2) exhibits stead growth in capacity and shipment. In (3), the capacity goes up in the beginning because the demand is much higher. However, after 30 months, it starts to decrease because of lower demand and profits.

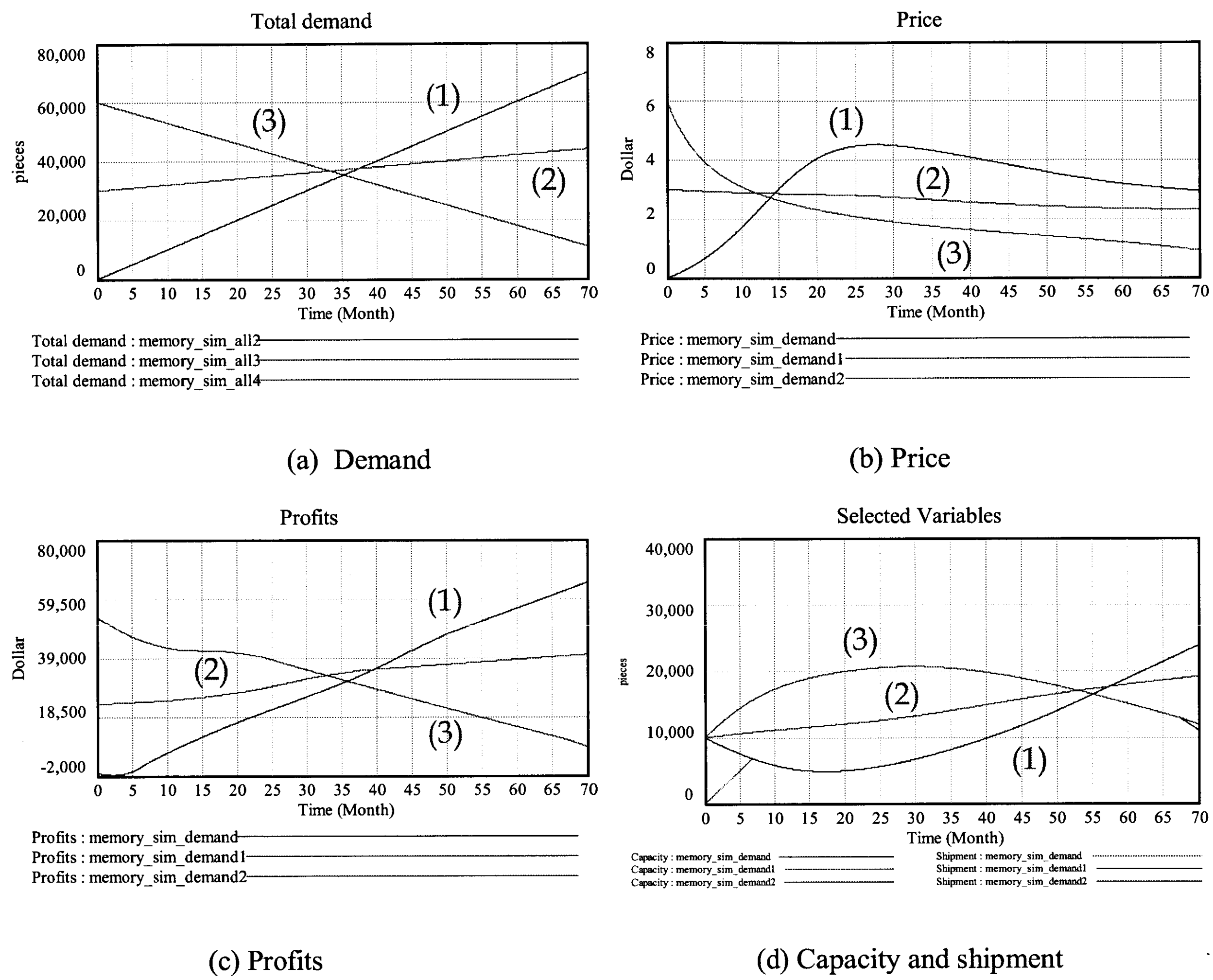


Figure.4.18 Simulation Results of Linear Demand Growth Patterns

4.7.2 Cyclic growth pattern

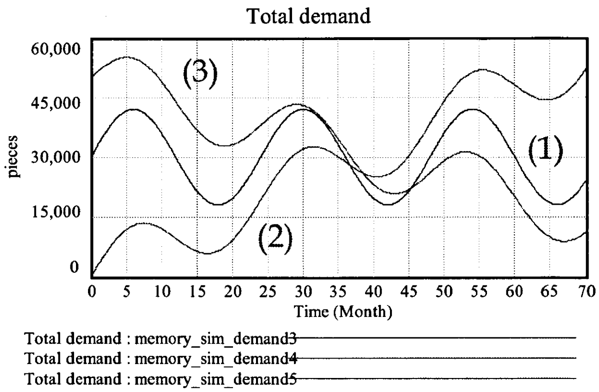
Figure.4.19 is simulation results based on the model in Figure.4.17. Figure.4.19 shows 3 simulation results for three types of demand growth with cycles.

Figure.4.19 (a) shows cyclic demand growth patterns. There are three patterns: (1) is a cyclic pattern with initial offset in demand (30,000), (2) is a cyclic growth pattern from 0, and (3) is cyclic declining pattern from 50,000. The period of cycles is 24 months.

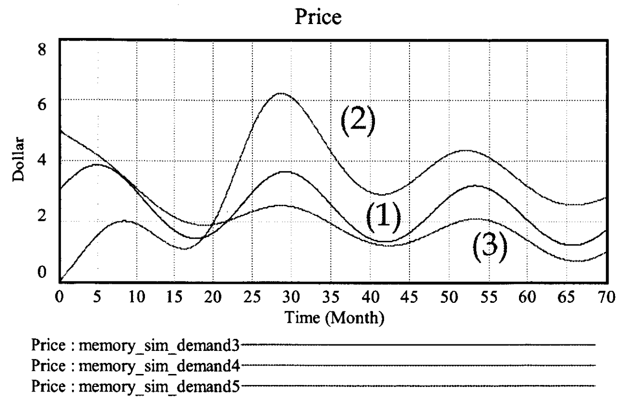
Figure.4.19 (b) shows price trajectories. In (1) the price stays from $1to $4 because of cyclic fluctuation in demand. In (2), the price goes up because demand increases while capacity decreases because of the low initial demand, but after the peak it will stay around $3. (3) shows a quick price decline in the beginning, but price fluctuates around $2 after the initial decrease.

Figure.4.19 (c) shows profits of three simulation results. (1) shows slight increase with cycles in profits. (2) starts from negative profits due to the low initial demand. In (3), profits go down with fluctuation. All profits waveforms show similar shapes with demand.

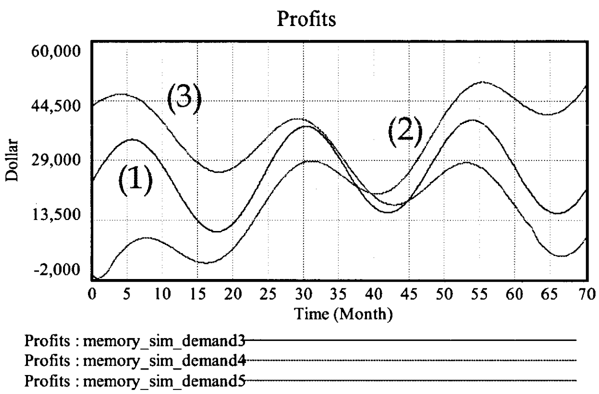
Figure.4.19 (d) shows capacity and shipment. (1) shows stead increase with low fluctuation in capacity and shipment. In (2), the capacity goes down in the beginning because the initial demand is low. But after the second peak it will go up because of the increase in demand. In (3), the capacity goes up in the beginning due to the high initial demand, and it will maintain higher level for long time, but it goes down eventually because of the decreasing demand.



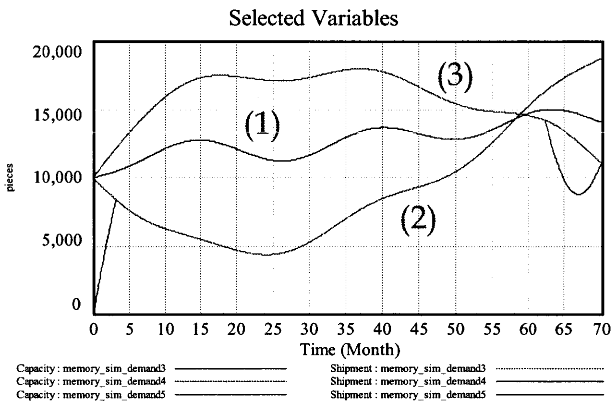
(a) Demand



(b) Price



(c) Profits



(d) Capacity and shipment

Figure.4.19 Simulation Results of Cyclic Demand Growth Patterns

5. Conclusions and Recommendations

The semiconductor market is becoming a more complex and intense market in which to compete. According to Worldwide Semiconductor Application Forecaster, worldwide semiconductor revenue will grow 3.2% year over year in 2012, and will achieve a compound annual growth rate of 4.8% for the 2011 – 2016 period.

The semiconductor memory market has been a very competitive market because of the volatility of price. In the memory market, as it is considered as a commodity market, price is the most important factor for survival. In some periods the price falls below the cost, which means market players cannot make a gross profit. If the economy is in a downturn, such as in 2008 or 2009, even the market leader cannot make a profit. It is indeed a very competitive market and difficult to operate a memory company.

There are some general beliefs in the semiconductor memory market. First, the continuous investment into new technology is a major driver of growth. Since many companies are working hard to reduce the cost of memory chips, companies that cannot follow the state of the art technology immediately lose their cost and performance competitiveness in the market. Companies that fall behind in developing new technology and aggressively putting money into production will quickly find themselves exposed to severe price competition because of lower manufacturing cost of the latest technologies. Any company that is ahead of developing new technology is also subjected to severe price competition due to quick commoditization effects. In spite of that, all companies aggressively move into new technologies to lower costs.

Another belief is that for management of the memory companies, capacity expansion is the most important factor for survival in the market. Apparently the investment decision in capacity is not an easy task because it is a cyclic market. Even though everyone knows that semiconductor demand is cyclic, no one knows when an upturn or downturn will come. The companies that overinvested in capacity will be exposed to a financial crisis, while the companies that underinvested will lose market share in the next upturn. Global inflations and recessions cause managers across industries to make similar investment decisions. The excess of capacity will decrease the price, but companies cannot meet the market demand if they do not have enough capacity.

In this thesis, system dynamics models are developed to understand the dynamics of the memory market and to predict future growth. The model predicted capacity and profit growth in accordance with demand and price movements. Two types of technology effects are introduced in the 2009 - 2015 timeframe: low cost effect and low power effect. The model assumed that new technologies will ramp up as a company invests in them. However, that is not always the case. New technologies require a lot of human resources and equipment which need capital expenditures. However, a few technologies are successful in the market while the majority is unsuccessful in business.

Several conclusions can be drawn from the simulation results of the system dynamics model. First, managers in memory companies should make investment decisions quickly based on demand/price fluctuation and capacity level. If the price declines it can severely affect the revenues of semiconductor companies that were counting on a continuation of the capacity expansion. They will not meet the customer demand for a product because of lower profits. The effects of economic cycles are no longer limited to the semiconductor industry. Managers prefer to be aggressive in adding semiconductor manufacturing capacity because they expect growth to continue. Most companies are also looking to reduce costs in factories. One way is to have strategic partnerships vertically. Foundries allow fabless chip makers to hedge against the capacity investment. Because companies are more specialized to their own strengths, vertical alliances in the value chain are more important to provide a product to end customers. While a few companies, such as Intel and Samsung, can remain as Integrated Device Manufacturers, other IDMs are disintegrating into fabless design companies or semiconductor foundries.

Pricing will remain a key indicator for the memory market. In a competitive commodity market, the unit price for a product will fluctuate according to the quantity demanded by consumers and the quantity supplied by producers. In the memory market, price is usually going down every day because a product is quickly commoditized. Higher capacity flash memory-based devices, such as smartphones, tablet PCs, and SSDs, have been primary applications driving increasing demand. Since applications for flash memory are expanding, the future increase of demand is apparent. However, overall increase of capacity will cause near-term aggressive price declines because the price is inversely proportional to the capacity. Historically, the DRAM industry has been experiencing similar price erosion because of a supply/demand imbalance. Eventually it leads to company consolidations. Given that the memory is the main component of cost in many new applications, the assumption for the memory pricing has resulted in a more aggressive decline.

This thesis points out the importance of low power features for the next generation of memory development and how to tackle the price-competitive market. Although the price will still remain an important factor, power consumption is becoming a problem in servers at data centers. Low power devices have been playing an important role in mobile devices; such as mobile phones, digital still cameras, and MP3 players. In those devices, power consumption determines how long the system runs because the power is supplied from a battery. The total supply of power is limited and efficiency of the use of power determines how long users can use the devices. Low power solutions are increasing their presence in semiconductor industry, and data centers require innovations in energy control. If the advantages of emerging technologies are used effectively, servers in data centers can reduce energy consumptions and cooling costs drastically. There are some software solutions: wear leveling, garbage collection, and system level optimization. However, innovation in hardware is necessary to completely change the game of the business. CPU, main memory, power supply, and peripheral buses take the majority of power consumption. For main memory, the use of new materials, such as MRAM, FRAM, PCM and ReRAM, is a better way to reduce leakage current rather than evolving from current structures. It is also possible to use multiple types of memory for main memory and permanent storage. Future growth in demand can be accounted for by a technology effect which increases demand for new products. New technologies will create new products with new functions. In this thesis a new technology effect is introduced to develop low power products.

While investment in low power technology will be important, traditional efforts to reduce cost will also remain a key factor. However, the cost of staying on the leading edge of technological progress is increasing because of increasing price of manufacturing equipment. The effect of commoditization is also caused by equipment suppliers. Even though a company developed the highest density product with the smallest node, other companies quickly catch up to the technology. Equipment suppliers are playing key roles for this commoditization because they sell the same equipment to many companies. The latest product developed for a market leader will also be sold to other competitors. The price decline due to technology commoditization is unavoidable in the current semiconductor value chain structure.

The volatility in semiconductor demand is related to the effects of price movements and capacity utilization. In this thesis, several demand growth patterns are tested because demand is inherently unpredictable. The unpredictability of demand is due to several reasons: variability of new products, varying life cycles, and dependence on customers' adoption. While companies have become increasingly sophisticated in forecasting demand and investment plans based upon past experiences, there still is a great amount of unpredictability in future. Adaptability to the changing demand needs to be tested to measure the robustness of the model. The price movements are also affected by investments by semiconductor manufacturers. If manufacturers want to increase their market share, they will invest more heavily in manufacturing capacity than in new technologies. Over capacity will cause price decline, and eventually it leads to scarcity of investments. Manufacturers who experience huge losses would make the same type of strategic investment with other companies. Managers of memory manufacturers should closely watch the availability of manufacturing capacity and market demand in the worldwide industry and use market data in making investment decisions. The statistical data should be used to prove the soundness of expansion plans for factories as a basis for investment decisions on future demand growth of semiconductor memory.

Bibliography

- Iansiti, Marco, 1997. "Technology Integration: Making Critical Choices in a Dynamic World", Harvard Business School Press

- Fine, Charles, 1998. "Clock Speed", Basic books

- Christensen, Clayton, 1997, "Innovator's Dilemma", Harvard Business Press

- Afuah, Alan, Bahram, Nik, 1995. "The hypercube of innovation", Research Policy 24, pp.51-76

- Kawashima, Masahisa, 2002, "Telecom Value Chain Dynamics and Carriers' Strategies in Converged Markets". MOT thesis, MIT

- Mollick, Ethan. 2006. "Establishing Moore's Law", IEEE Annals of the History of Computing, July-September, pp.62-75

- Henry, Weil Birdseye, Utterback, James M, 2005, "The dynamics of Innovative Industries".

- Kameda, Mitsuhiro, 2004, "Disruptive Innovation – Value Change and Complementary Change". MOT thesis, MIT

- Weber, Charles, 2003, "Rapid Learning in High Velocity Environments". PhD Thesis, MIT

- Bass, Michael J, 2000, "A Strategic Study of Disruption, Dis-integration, and Modularity in the Microprocessor Industry", MOT thesis, MIT

- Trent, Tracy, 1997, "Changing the rules on market leaders: Strategies for survival in the high-performance workstation Industry". MOT thesis, MIT

- Clayton, Christensen, Suarez, Fernando, Utterback, James, 1998, "Strategies for Survival in Fast-Changing Industries". Management Science, Vol.44, No.12 Part 2 of 2

- Acee, Happy, Utterback, James, 2005, "Disruptive Technologies: An Expanded View". International Journal of Innovation Management, Vol.9, No.1 pp.1-17

- George Escobido, Matthew, 2009, "Toward a more dynamic model of competition". SDM thesis, MIT

- Pistorius, C W. I, Utterback, James, 1995, "The Death Knells of Mature Technologies". Technological Forecasting and Social Change 50, 133-151

- Kai, Hidetaka, 1992, "Competitive strategy under standardization in the personal computer industry and its influences on new entrants". MOT thesis, MIT

- West, Hugh, 1998, "Technology strategy in commodity industries". MOT thesis, MIT

- VanBree, Kenneth, 1995, "Silicon cycles: An analysis of the patterns of growth in the semiconductor industry using system dynamics methods". MOT thesis, MIT

- Sosa, Lourdes, 2009, "Application-Specific R&D Capabilities and the Advantage of Incumbents: Evidence from the Anticancer Drug Market". Management science, Vol.55, No.8, pp.1409-1422

- Glenn, Carroll, et al, 1996, "The fate of de novo and de alio producers in the American automobile industry 1885-1981". Strategic management journal, Vol.17, 117-137

- Roberts, David, 2001, "The Efficient Data Center Architectures Using Non-Volatile Memory and Reliability Techniques" Ph.D thesis, The University of Michigan

- Weil, Henry, 2010, "Why Market Makes Mistakes", Kybernates, Vol.39 No.9/10 pp.1429-1451

- Fazan, Pierre, 2010 "Future RAM Emerging Technologies and Their Applications" Global Semiconductor Alliance memory conference

- Yang, Michael, 2011, "IHS iSuppli Data Flash Market Tracker"

- Yang, Michael, 2010, "iSuppli Flash Memory Market Shares"

- IDC, 2012, "Worldwide Semiconductor 2012 Top 10 Predictions"

- IDC, 2011, "Worldwide Solid-State Storage 2011-2015 Forecast update"

- Aochi, Hideaki, 2009, "BiCS Flash as a Future 3D Non-Volatile Memory Technology for Ultra High Density Storage Devices" International Memory Workshop 09, pp.1-2

- Integrated Memory Engineering Corporation, 1997, "Memory 1997"

- Yinug, Faran, 2007, "The Rise of the Flash Memory Market: Its Impact on Firm Behavior and Global Semiconductor Trade Patterns", Journal of International Commerce and Economics, United States International Trade Commission