1.2 Semiconductor Memories for IT Era

Chang-Gyu Hwang

Samsung Electronics Co., Ltd., Yongin City, Kyunggi-Do, Korea

I. Introduction

Since the beginning of history, civilization has sought ways to enrich lives with more comfort. The IT industry has gained momentum with the invention of computers, global internet, and wireless communication in connecting the world in a more intelligent and informative environment without much limitation in space or/and time. The memory industry has prospered mainly by the standardization and massive production for the EDP market according to Moore's law by continuously focusing on design automation and developing more scaled processing technology.

Recently various technologies are merging for more synergy, enhancing the cultural association. For example, mechatronics has come from a merge between mechanical engineering and electrical engineering, and nanotechnology is combining both physics and mechanics to encompass more varied application fields. An analogy exists even in the semiconductor IT industry with the merge of frontend (design and process) and backend (package and module) technologies. Future memory technology will be driven by the combined fusion technologies to provide a total solution. Also, the memory market will be greatly diversified for numerous customized applications in addition to current mainstream EDP market (such as digital cellular, digital TV, server appliance, handheld computers, digital camcorder, set top box, auto navigation system, automotive and medical electronics, etc.) to meet various IT customer demand with technology capability, fully utilizing and combining the existing technologies. Also, memory bit cost is continuously decreasing, which stimulates the creation of IT users, and semiconductor memories are continuously improving to enhance efficiency through smarter design techniques and more advanced process technology. Thus, the memory business will prosper further as the market has grown in the past decade.

Technical approaches are discussed here to bring these diversified memory aspects into a general overview and recent perspective, showing a variety of possibilities for the memory technology in terms of design, process, and backend technology that will unfold in the next decade.

II. Future Memory Technology and Market Trends

Thus far, the trend of memory technology development has been driven mainly by efficiency improvement in terms of high density, high speed, and low power through smart design techniques and scaled advanced process technology development for the mass production of standard products commonly used for personal computers. However, the future memory market with the advanced process, especially, for IT industry will also have specialized and diversified products that can be supported using each available technology area. Figure 1.2.1 shows the diversified memory demands from various groups of users (for server, workstation, desktop, network, communication, game, consumer, etc.) which are different in memory bandwidth, latency, system performance, power, data holding degree, and so on. Technology development now focuses on providing system solutions to the various customers with highly differentiated value-added products.

The total solution for a high-performance memory system may be provided through both component and system aspects. In the component aspect, generic memory goals such as bit density, speed (latency, bandwidth), and power consumption are the main technology aims to improve system performance [1]. Firstly, large density area is one of the natural moving trends to produce a smaller memory cell. Figure 1.2.2 shows the technology trend of different memories from past to near future. The trend has been characterized by Moore's law since the birth of memory with Intel's 1kb DRAM in 1970. The same integration density trend will hold until 2010 for the server applications to connect wide-spread internet communication and networking environments. Secondly, the low-voltage trend of memory devices is traditionally set to meet the reliability and power consumption requirements as shown in Figure 1.3.3. This need is particularly critical for mobile and portable applications (mobile phone, PDA, notebook PC, wrist-watch, heart pacer, etc.) to prolong the battery life to support either ultra-low standby power or real-time transmission of motion pictures. To reduce active current and data-retention power of DRAM to <μA/Mb for mobile communication systems, multi-partitioning of cell arrays and optimal refresh interval control schemes with monitored device temperature and memory cell leakage current are effective methods [2]. Thirdly, the memory speed has been steadily improved by process advances. Figure 1.2.4 shows the trend of the memory and CPU operating frequency. The speed gap between logic and memory is increasing. However, recent implementation of high-speed interface circuits and new I/O signaling schemes like in RDRAM with 1~2Gb/s/pin data transfer can significantly narrow the gap [3]. Further increased data transfer speed to 2~4Gb/s/p is possible with a multi-level data transfer scheme [4], while SRAM and graphics DRAM with a double-data-rate scheme working at 500~600Mb/s/pin and present commodity SDRAM are moving at a slower pace from 100MHz to 200MHz. This high-bandwidth device is attractive for high-performance small systems with small granularity, especially for graphics.

To bridge the gap between those items for the component solution, evolutionary and revolutionary system solutions including package and module technology are further sought to overcome limits of overall speed, density, size, power, and performance through the pursuit of differentiated value-added development. Hence, more cost-effective differentiated higher-performance systems to meet delicate customer requirements can be realized to provide value-added total memory solutions. The challenges are the provision of top-edge technology combined solutions with fully-utilized system with memory (SWM) technologies like cost-effective single-chip integration of different memories [5-6], memories integrated with high-performance logic [7-8], multi-chip package with low pin parasitics, small high-bandwidth multi-channel module, and so on.

Figure 1.2.5 describes the future technology and demand trend in IT areas such as PC's, servers, graphics, networking, and communication fields [9-11]. The common technical needs for each area are high speed, low power, and high density. DRAM and SRAM will mainly focus on high speed and low power for mainly EDP and networking applications, while NVM (Flash) will mostly target large density mass storage for digital mobile consumer products. From the application point of view, handheld consumer electronics and communication-oriented products will gain the new market from the conventional memory intensive-products such as personal computers. For the near future, a wristband type gadget, so-called "information bank system" (with such special features as 1.0V, 1GHz, MPU and 15GB memory with voice recognition, 10W peak power, 0.1W standby power, 4-hour battery life at full speed, 3x5x0.7cm^3 system size, and 300g total weight including battery) would be realized due to the improvement of memory performance with the help of wireless communication electronics to accommodate many useful functions such as detection of disease or environmental contamination, automatic check-in at hotel, substitutional credit card (e-money), information and time table scheduler, game and e-book, and GWS identification recognition system, etc. Also, nanotechnology-based memory may be designed and fabricated

 0-7803-7335-9 ©2002 IEEE

that mimics the function of an individual neuron. Therefore, it is possible to fabricate a neuro-biochip containing many of these transistors to simulate the function of brain neurons. This can someday replace damaged or malfunctioning brain circuitry and moreover, an artificial brain will be possible.

III. New Design Area

In the IT era, the number of memory systems for communication and mobile applications will increase so that design concepts will change to accommodate a specialized diverging memory market. To achieve this, the CAD tool environment and high-performance (power, speed) circuit techniques will be activated to meet time-to-market for memory-intensive designs such as graphics function intensive high-performance devices and various unified memory products. Power reduction techniques will be one of the highest priorities as the handheld market grows exponentially. Strong demand for high-performance servers and networks stimulates the need of high-speed and high-capacity devices to properly communicate with the terminals of many users at the same time.

For future cost-effective specialized products, design challenges should be considered because of the limitation of further improvement with conventional shrinkage methods in terms of speed enhancement, bit density increase, and power reduction. First, to make a high-speed device to be more adaptable for production, several critical circuit techniques should be considered. The reduction of process variations on a wafer becomes extremely difficult and costly to control at the current level of device scaling. Thus, there is a strong demand to establish a design environment that reflects variation of process parameters including transistors and contact resistance as the device scales. Secondly, the speed can deteriorate as the device shrinks, because of cross-talk noise and interconnect delay. Thirdly, the restricted pin RLC specifications on the parasitic of I/O circuitry including package require design efforts such as synchronized time-delay compensation, on-chip adaptive self-calibration, and signal integrity on a system channel. Fourthly, careful estimation of device degradation of scaled device during a high-speed operation is becoming more important. Hence, it is inevitable to set up the overall CAD environment for extracting the model parameters considering process fluctuation, for building the statistical simulation system, for designing circuits robust against variations in voltage and temperature and noise, and for inclusion of reliability considerations.

Many dedicated unified chips will also be proposed in the future to provide total memory solutions instead of achieving high-performance devices obtained by optimizing chip architecture. They range from small modifications such as SRAM interface with DRAM cells and DRAM with row caches, to large-scaled integration such as small computer memory chips by combining several different memory chips or memories merged into the logic devices. The total memory solution providers are able to deliver differentiated high-performance chips (in terms of functions, speed, and power) in a short time owing to the exploitation of an established environment of memory MACROs.

Therefore, to develop high-performance products, effective CAD tool environments for efficient design methodology and high-performance circuit design techniques with adoption of optimal process technology are required to make providing specialty products more profitable, while meeting the particular customer specifications.

IV. Process Technology Evolution

For the new IT era, the trend of memory process technology is moving in two directions: a traditional high density approach for memory intensive EDP applications as a mainstream and a user-specific special process approach suitable for mobile and portable applications such as ultra low standby power and/or high speed and multifunctional unified memory chips.

Firstly, the development of miniaturized process technology has great momentum, pursuing higher density with a smaller chip. The trend of bit density increase is characterized by 1.5x~1.6x/year from 64kb to 512Mb DRAM since 1984 and the trend might be continued until 2005 with a slightly decreased rate due to increased difficulties and challenges in process and device technology. At that time 1Gb DRAM with ~120mm^2 chip (0.8~1Gb/cm^2) and 1Gb NAND Flash with ~80mm^2 chip (1.2~1.4Gb/cm^2) will be available with 300mm wafers. Advanced technology will be used to realize higher-scale integration with several high-performance functions on a chip and production of 4Gb DRAM with <200mm^2 chips will take place in the next 10 years [12,13,14]. Several techniques for realizing high-bit density are possible with relatively less advanced process technology available at present and in the near future: multi-bit cell (2~3b) [14-17], fabrication of vertical transistor and memory capacitor for a 6F^2 to 4F^2 memory cell (F=minimum feature size) [18], multi-chip stacking connected with conventional wires using wafer thinning to ~50~100µm. Figure 1.2.6 shows a prospective technology roadmap up to 2010 for different memory devices.

Secondly, in addition to traditional process technology enhancement, a user-specific special process will be developed with such emerging technologies to provide higher-performance products (ultra-high speed and/or low power) and specialized converging products (high-performance small computer memory): Cu or other low-resistive interconnects (W or salicide) and low dielectric constant insulators to improve reliability and speed, new storage materials (high-dielectric materials like BST (ε_r=150~200) and Ta_2O_5 (ε_r=20~25), magnetic resistive layers like GMR and TMR) to significantly increase the bit density, SOI or selective epitaxial-layer growth (SEG) process and strained silicon (like SiGe) process technology to reduce parasitic and achieve ultra-high speed. Multilevel metal processed (up to 4~5 layers) will be popular in the near future to accommodate reduction of delay time, improvement of noise characteristics, and high packing density, especially useful for embedded memory and logic applications. One ultimate unified memory approach is the development of all memory integration processes to combine all memories (DRAM, SRAM, and NVM) in a single chip (MIC) even with logic circuits to take advantage of each memory's merits and interfaces, resulting in creation of new high-performance small computing market. The other unification way is the use of special storage materials like magnetic layers [19-22]. Nonvolatile DRAM memory can make access speed and cell size comparable to those for DRAM without need of high voltages required to write in Flash memory, resulting in creation of new memory markets for IT applications.

V. Backend (package & module) Areas

Packaging technology has developed remarkably for the last decade and development rate will accelerate in the next decade for rapidly creating IT industry to provide a system solution. Figure 1.2.7 shows a prospective package technology roadmap up to 2005 for diversified memory devices. To accommodate an increasing number of I/Os in the device, grid array packages are adopted in many applications. Although present wire-bonding technology is stable and mature, it is not adequate for an interconnection method for high-frequency devices with large I/O pin counts. Chip-scale package/chip-size package (CSP) is another example providing excellent electrical performance at package level due to small dimensions (low pin parasitics). As mobile applications such as handheld communications and other small-form-factor digital systems emerge, more compact memory package solutions with relatively small pin counts are required for system miniaturization. The most popular to reduce the mounting area is the multi-chip package (MCP), which is popular for higher density in one package or multiple packages in a small volume using multiple dies stacked vertically or horizontally.

 0-7803-7335-9 ©2002 IEEE

Combination of different memories is now the most popular form of stacking. Also, memory and logic and/or memory and ASIC combinations will grow in popularity in the future due to reduced mounting area and low I/O pin counts. The system in package (SIP) is a recent terminology that incorporates combination of different memories or logic and ASIC. SIP can be a transient solution or can replace where SOC is necessary by incorporating the advanced interconnection technology like flip chip for less space. A more complex die stacking structure composed of various dies soon will be available for mobile applications.

Module technology has played a major role providing a memory solution for the EDP market and especially for server applications requiring a large memory capacity. Recently, for portable handheld systems, a total high-performance small-system solution is becoming more demanding. Thus, a small-form-factor memory module is getting more important for portability. Also, high-bandwidth modules (for high-rate data transfer in a given volume) for EDP and consumer markets are considered for high-speed and small system with the challenges of precise PCB design technology and good impedance matching techniques on channels through mother board, module PCB, and socket to provide good signal integrity during high-speed clocking. A 10GB/s module bandwidth will be available in 2005 with the development of multi-channel modules.

VI. Conclusion

For the next century, semiconductor memories will be improved continuously to enhance the efficiency with smarter design techniques and more advanced process technology and to provide differentiated value-added system solutions. Figure 1.2.8 shows trends for total memory demand (~1.7x per year) and cost reduction rate over the last 2 decades and beyond. The memory bit cost ($/MB) has continuously decreased by ~0.65x per year and the population of IT users is rapidly increasing. Total memory revenue will increase by ~1.1x per year during the next decade. Thus, the memory business will prosper further as IT researchers and engineers have in the past with four momentum words: love (zeal), jealousy (differentiation), truth (satisfaction), and freedom (customization).

Acknowledgements:
The author thanks C. Kim, T. S. Jung, K. N. Kim, D. H. Lee, K. D. Suh, H. Yoon, and Y. J. Choi for providing data and discussions and all colleagues for efforts and contributions for this paper.

References:
[1] C.G. Hwang, Symp. on VLSI Circuits, pp. 88-91, Jun 2000.
[2] K. Sato, et al., ISSCC, pp. 268-269, Feb. 1991.
[3] C. Kim, et al., IEEE J. Solid-State Circuits, vol. 34, pp. 645-652, May 1999.
[4] Jared L. Zerbe, et al., Symp. on VLSI Circuits, pp. 128-131, Jun. 2000.
[5] Sawada, K. et al., JSSC, vol. 23, pp. 12-19, Feb. 1988.
[6] Sato, K. et al., JSSC, vol. 26, pp. 1556 -1562, Nov. 1991.
[7] O. Takahashi, et al., JSSC, pp. 1673-1679, Nov. 2000.
[8] Y. Yokoyama, et al., JSSC, pp. 503-509, March 2001.
[9] P. Gelsinger, ISSCC, pp. 22-25, 2001.
[10] G. Weinberger, ISSCC, pp. 20-24, 2000.
[11] B. Merritt, Platform Conference, 2001.
[12] K. N. Kim, et al., Symp. on VLSI Technology, pp. 7-8, Jun. 2001.
[13] H.S. Jeong, et al., IEDM, pp. 353-356, 2000.
[14] T. Murotani, et al.," ISSCC, pp. 74-75, Feb. 1997.
[15] M. Bauer, et al.," ISSCC, pp. 132-133, Feb. 1995.
[16] T. S. Jung, et al.," JSSC, vol. 31, pp. 1575-1583, Nov. 1996.
[17] T. Cho, et al.," JSSC, vol. 36, pp. 1700-1706, Nov. 2001
[18] W. F. Richardson, et al., IEDM, pp. 714-717, Dec. 1985.
[19] S.S.P. Parkin, et al., J. Applied Physics, vol. 85, no. 8, pp. 5828-5833, April 1999.
[20] R. Scheuerlein, et al., ISSCC, pp. 128-129, Feb. 2000.
[21] B. G. Jeon, et al., ISSCC, pp. 272-273, Feb. 2000.
[22] C. Paz de Araujo, et al., ISSCC, pp. 268-269, Feb. 2000.

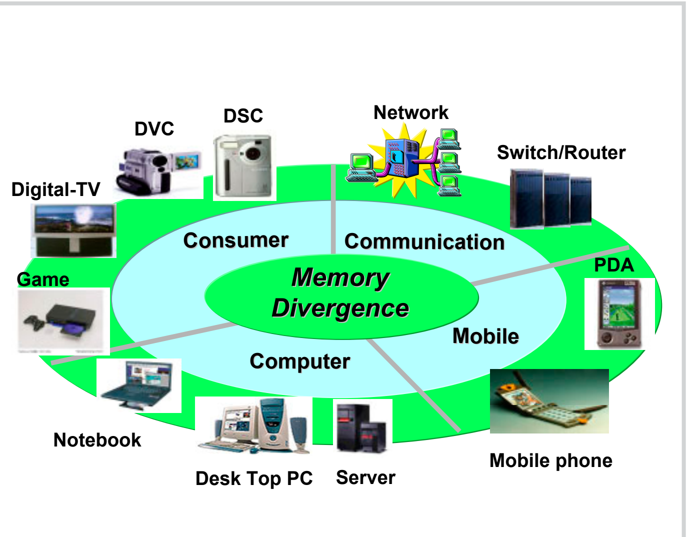

Figure 1.2.1: Diversified memory demands from various groups of users.

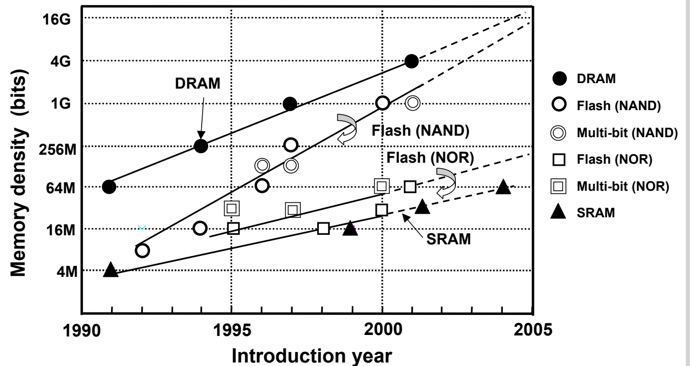

Figure 1.2.2: Technology trend introduction of different memories from the past to the near future.

Year	1999-2000	2001~2004	2005~2010	> 2010
DRAM DR (F)	0.2~0.18µm	0.15 ~0.11µm	0.1~0.07µm	< 0.07µm
Density	256Mb	512Mb~2Gb	4~16Gb	>32Gb
Voltage	5~2.5V	2.5~1.5V	1.5~1.2V	1.2~1.0V
SRAM	0.18µm	0.15 ~0.1µm	0.1~0.07µm	< 0.07µm
Density	16Mb	32~64Mb	144~256Mb	>512Mb
Voltage	~1.8V	1.5~1.0V	1.0~0.8V	<0.8V
NVM (Flash)	0.2~0.18µm	0.15 ~0.09µm	0.08~0.05µm	< 0.05µm
Density	256Mb	512Mb~2Gb	4~16Gb	>32Gb
Voltage	3.3V	~1.8V	~1.8V	1.3~1.0V

Figure 1.2.3: Trend of memory device operating voltage migration.

 ©2002 IEEE

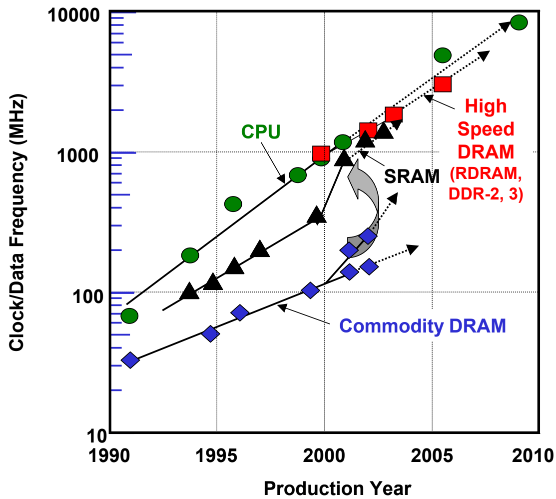

Figure 1.2.4: Trend of the memory and CPU operating frequency.

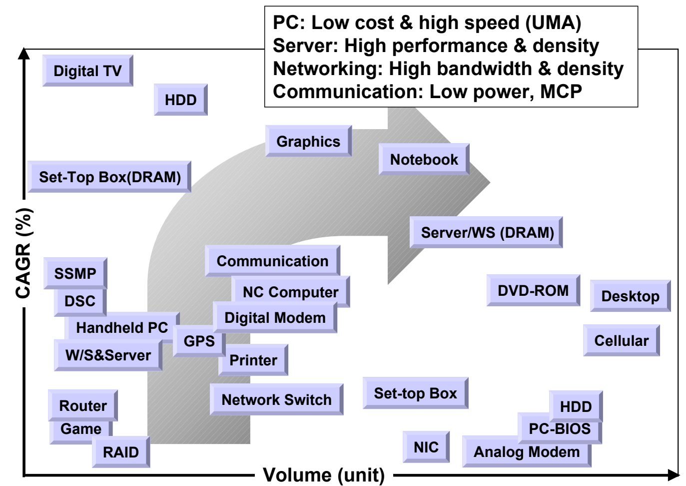

Figure 1.2.5: Future technology and demand trend of each IT area: PC, server, networking, consumer, and communication fields.

Year	1999-2000	2001~2004	2005~2010	> 2010
Litho-graphy	KrF λ=248nm	KrF+ RET: Resolution Enhancement Technique	ArF+Next λ=193nm	NGL, E-Beam, X-Ray
Isolation	STI			STI/SOI/SEG
Metal	W-Plug + Al		Al, Cu	Cu
Package	TSOP,CSP,COB,FCBGA, MCM, Stacking		Bare die stack	Stacking,mach'ng
DRAM DR (F)	0.2~0.18um	0.15 ~0.11um	0.1~0.07um	< 0.07um
Capacitor	Cylinder/HSG + NO	HSG + Ta_2O_5/ Al_2O_3	Ta_2O_5 Cylinder	High κDielectric
Cell Dim.	$8F^2$		$6F^2$	$4F^2$
Transistor	Gox=60~70A WSix, CoSix	Gox= 40~50A WSix, CoSix, W	Gox=30~40A Metal	New dielectric Midgap metal gate
SRAM	0.18um	0.15 ~0.1um	0.1~0.07um	< 0.07um
Cell Tr.	6-Tr, Conventional	6-Tr, Conv/Borderless	6-Tr, Borderless Interconnection	
Transistor	Gox=20~35A, NO	Gox=13~20A, NO	Gox=8~15A	Gox<12A, High κ
NVM (Flash)	0.2~0.18um	0.15 ~0.09um	0.08~0.05um	< 0.05um
Cell Dim.	$6F^2$		$4F^2$	
Cell Tr.	Floating Gate	High Coupling FG, Charge Trapping		

Figure 1.2.6: Prospective technology roadmap for different memories up to 2010 .

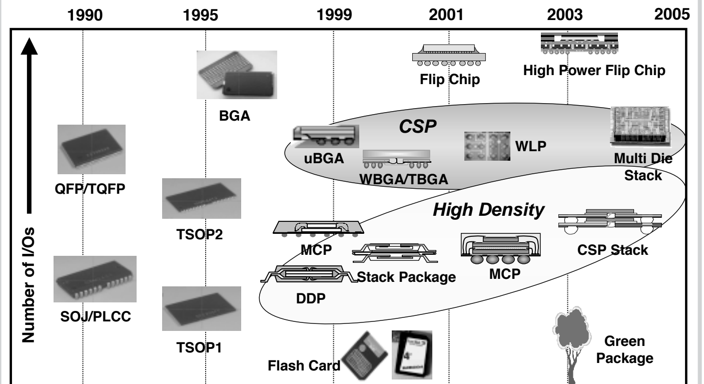

Figure 1.2.7: Prospective package technology roadmap for diversified memory devices.

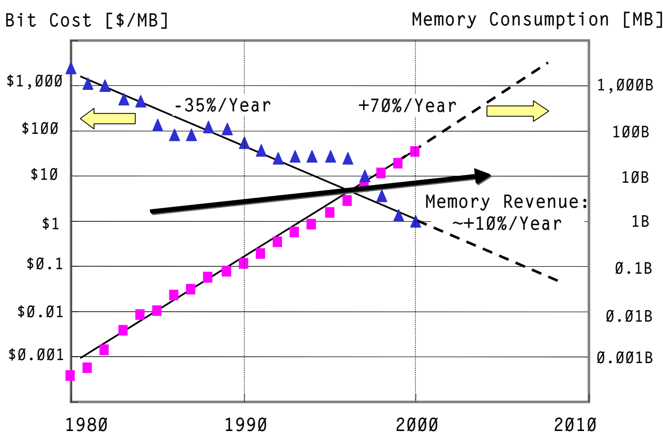

Figure 1.2.8: The trends of the total memory size demanded and memory cost reduction rate for the last two decades and into the future.

0-7803-7335-9
©2002 IEEE

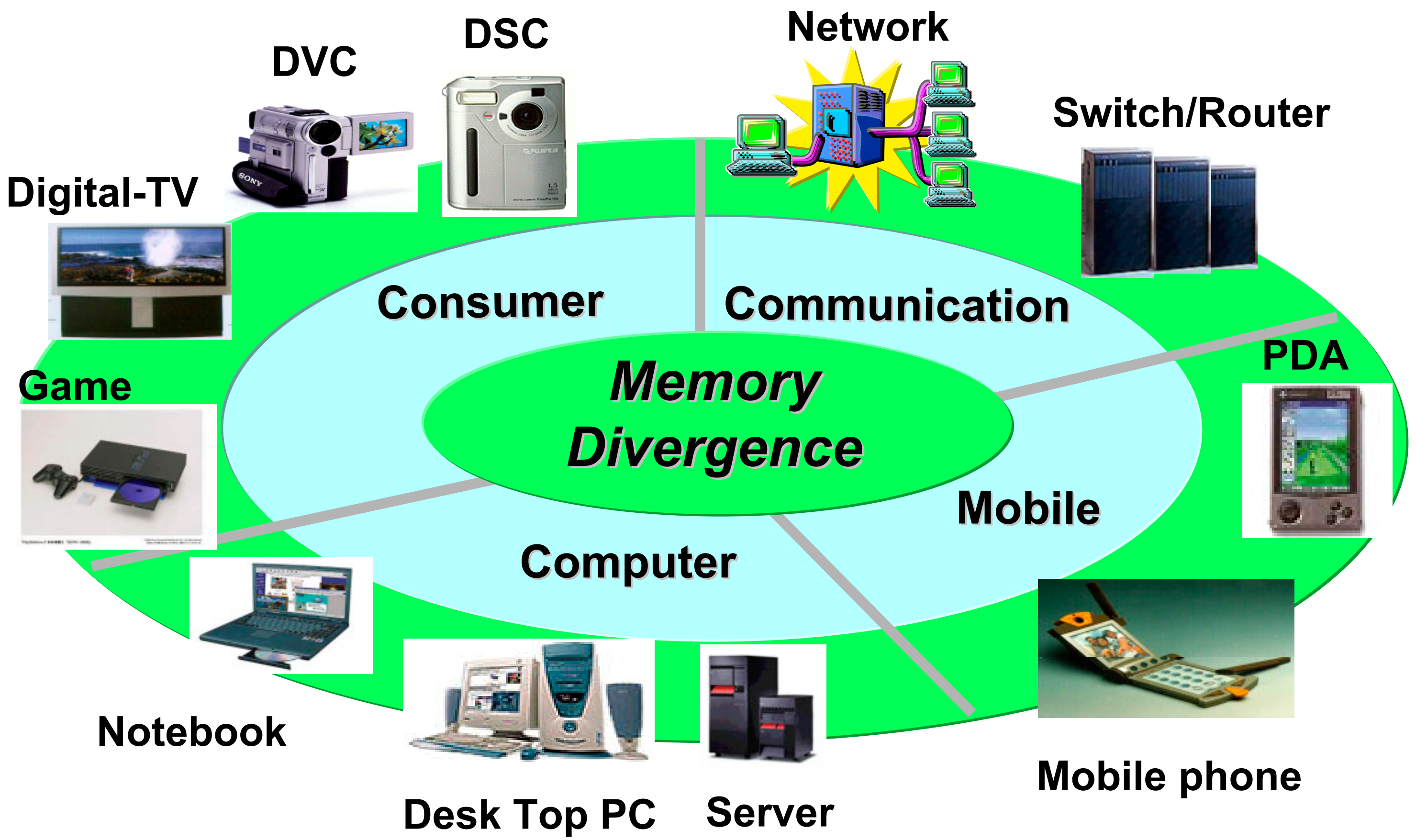

Figure 1.2.1: Diversified memory demands from various groups of users.

©2002 IEEE

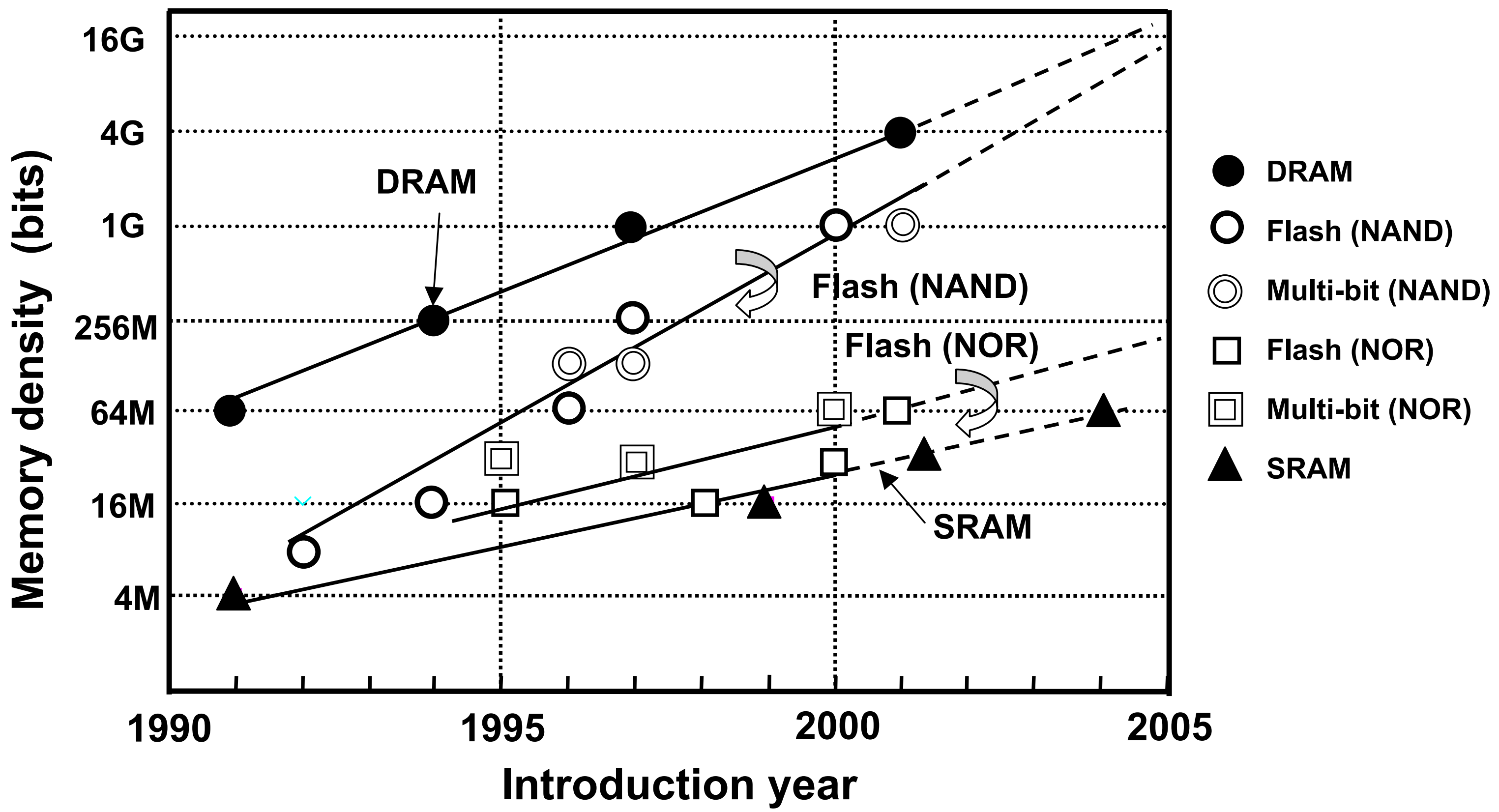

Figure 1.2.2: Technology trend introduction of different memories from the past to the near future.

 ©2002 IEEE

Year	1999-2000	2001~2004	2005~2010	> 2010
DRAM DR (F)	0.2~0.18µm	0.15 ~0.11µm	0.1~0.07µm	< 0.07µm
Density	256Mb	512Mb~2Gb	4~16Gb	>32Gb
Voltage	5~2.5V	2.5~1.5V	1.5~1.2V	1.2~1.0V
SRAM	0.18µm	0.15 ~0.1µm	0.1~0.07µm	< 0.07µm
Density	16Mb	32~64Mb	144~256Mb	>512Mb
Voltage	~1.8V	1.5~1.0V	1.0~0.8V	<0.8V
NVM (Flash)	0.2~0.18µm	0.15 ~0.09µm	0.08~0.05µm	< 0.05µm
Density	256Mb	512Mb~2Gb	4~16Gb	>32Gb
Voltage	3.3V	~1.8V	~1.8V	1.3~1.0V

Figure 1.2.3: Trend of memory device operating voltage migration.

©2002 IEEE

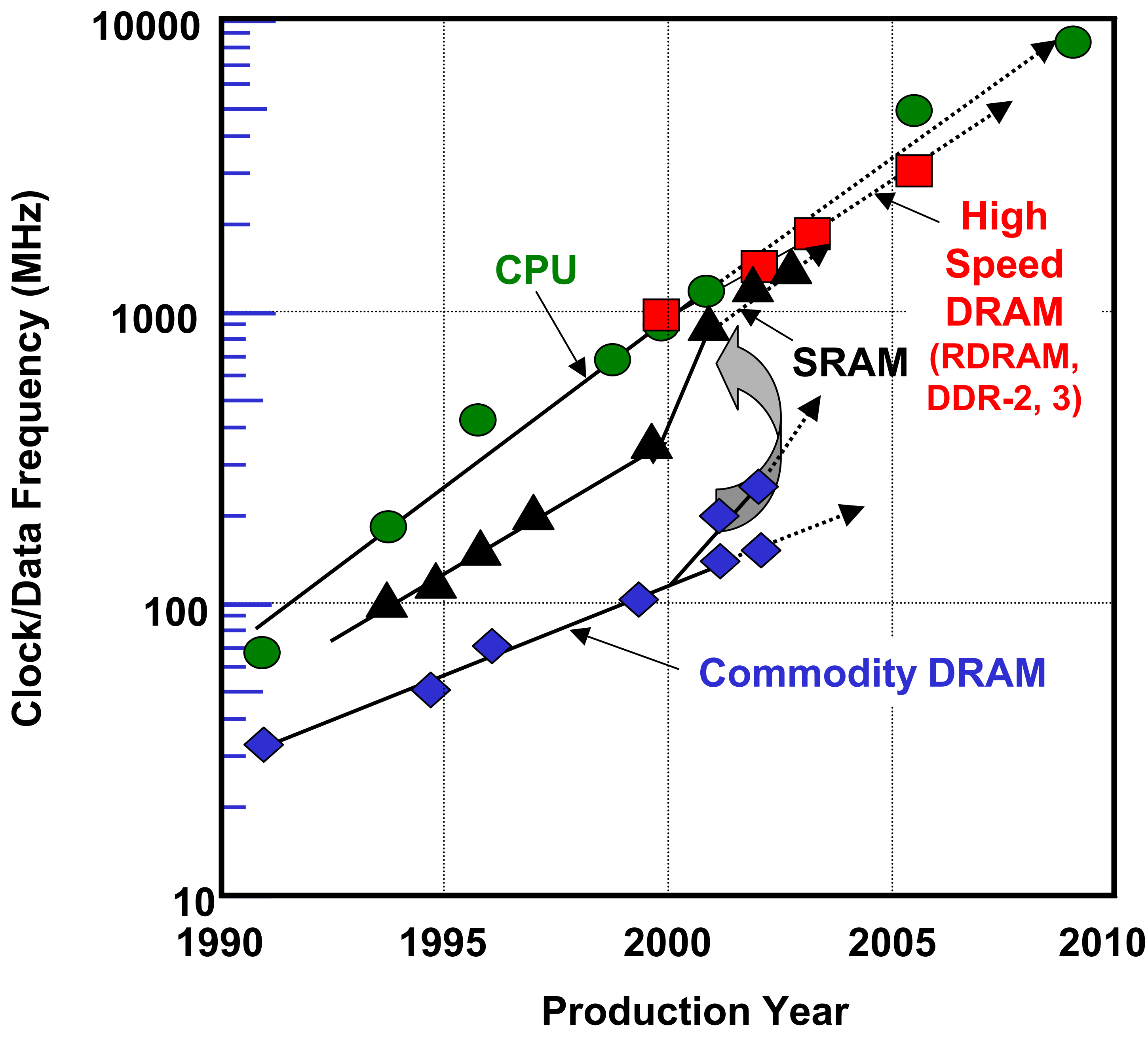

Figure 1.2.4: Trend of the memory and CPU operating frequency.

 ©2002 IEEE

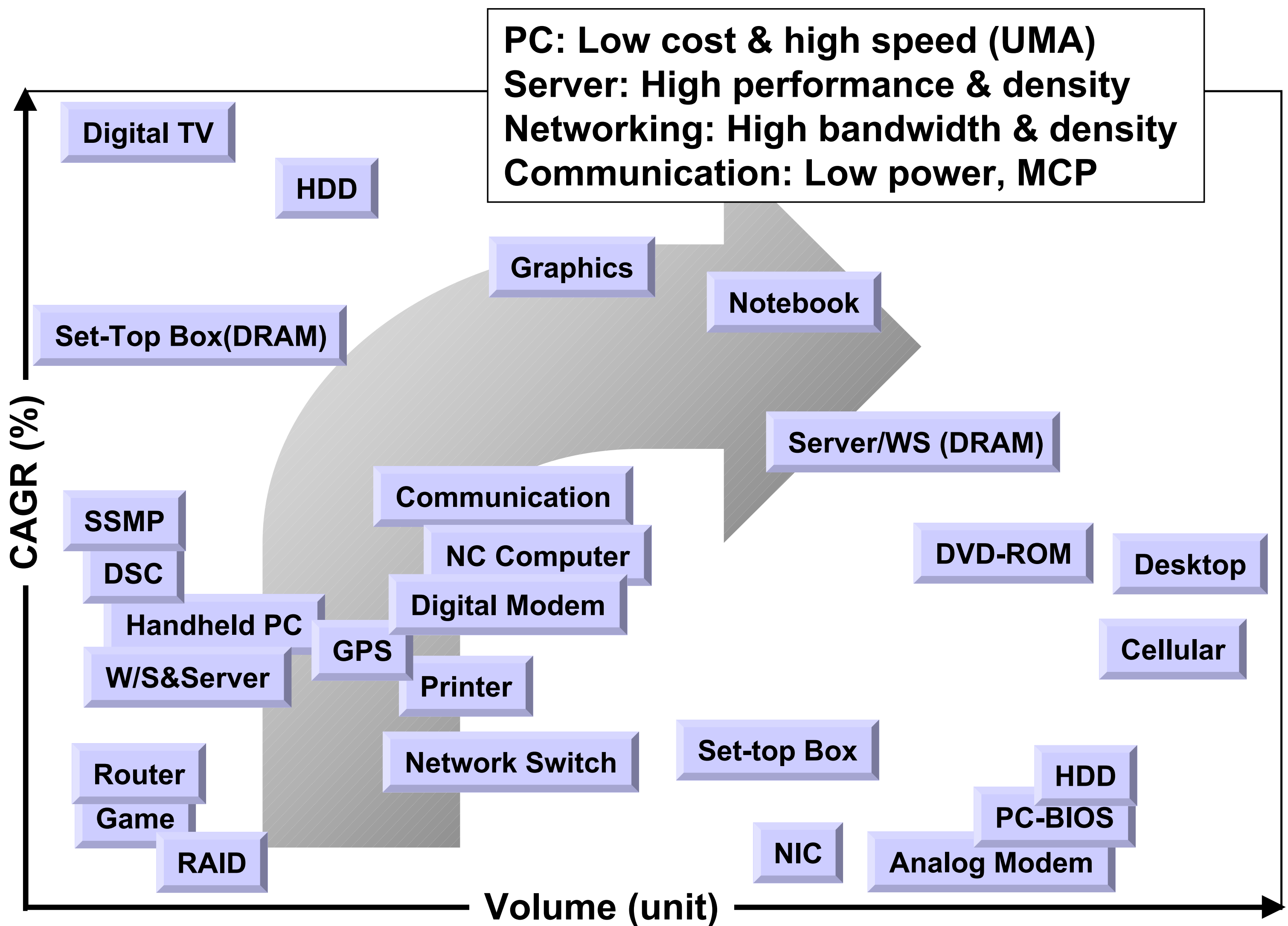

Figure 1.2.5: Future technology and demand trend of each IT area: PC, server, networking, consumer, and communication fields.

 ©2002 IEEE

Year	1999-2000	2001~2004	2005~2010	> 2010
Litho-graphy	KrF λ=248nm	KrF+ RET: Resolution Enhancement Technique	ArF+Next λ=193nm	NGL, E-Beam, X-Ray
Isolation	STI			STI/SOI/SEG
Metal	W-Plug + Al		Al, Cu	Cu
Package	TSOP,CSP,COB,FCBGA, MCM, Stacking		Bare die stack	Stacking,mach'ng
DRAM DR (F)	0.2~0.18um	0.15 ~0.11um	0.1~0.07um	< 0.07um
Capacitor	Cylinder/HSG + NO	HSG + Ta_2O_5/ Al_2O_3	Ta_2O_5 Cylinder	High κDielectric
Cell Dim.	$8F^2$		$6F^2$	$4F^2$
Transistor	Gox=60~70A WSix, CoSix	Gox= 40~50A WSix, CoSix, W	Gox=30~40A Metal	New dielectric Midgap metal gate
SRAM	0.18um	0.15 ~0.1um	0.1~0.07um	< 0.07um
Cell Tr.	6-Tr, Conventional	6-Tr, Conv/Borderless	6-Tr, Borderless Interconnection	
Transistor	Gox=20~35A, NO	Gox=13~20A, NO	Gox=8~15A	Gox<12A, High κ
NVM (Flash)	0.2~0.18um	0.15 ~0.09um	0.08~0.05um	< 0.05um
Cell Dim.	$6F^2$		$4F^2$	
Cell Tr.	Floating Gate		High Coupling FG, Charge Trapping	

Figure 1.2.6: Prospective technology roadmap for different memories up to 2010 .

©2002 IEEE

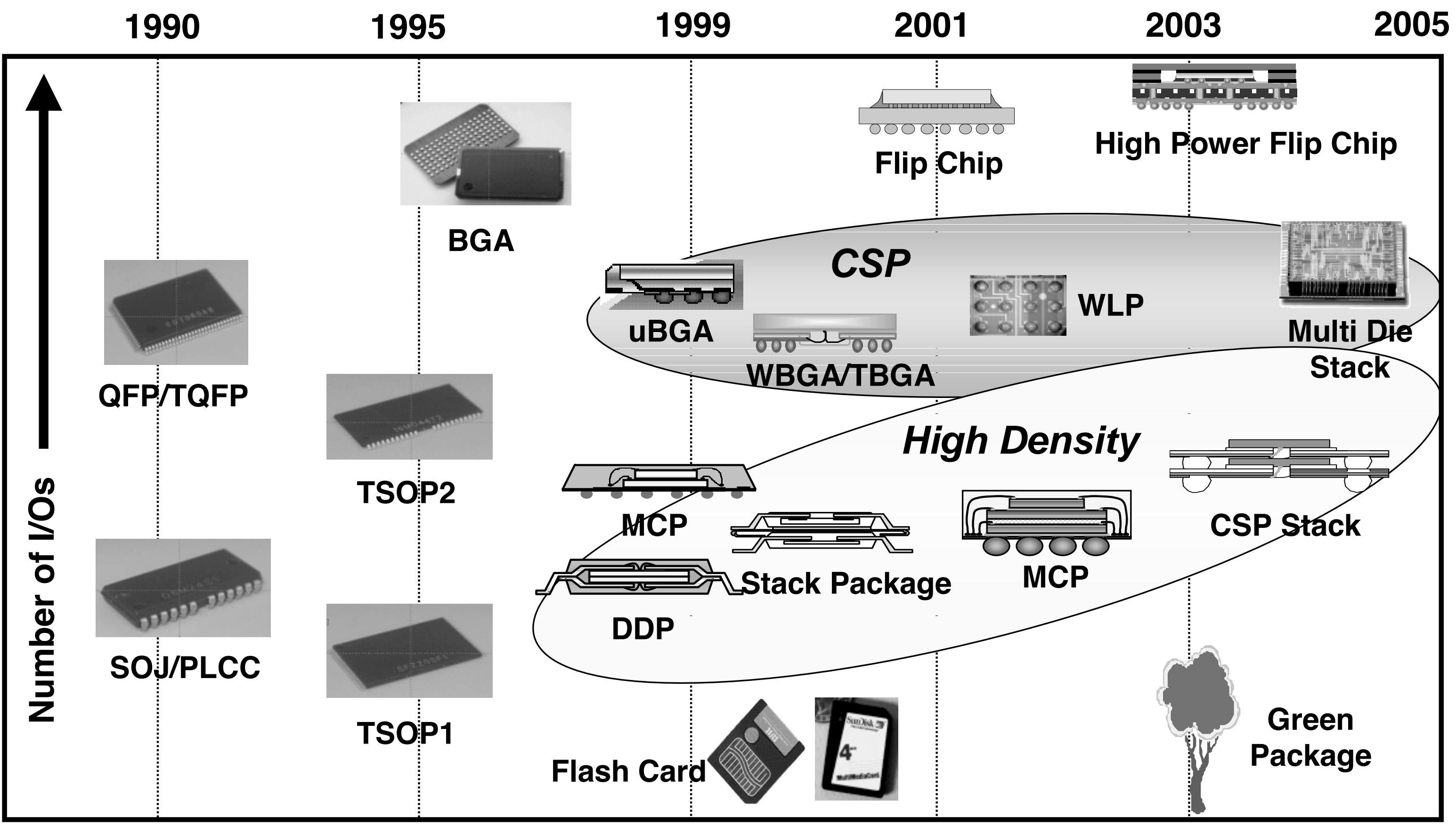

Figure 1.2.7: Prospective package technology roadmap for diversified memory devices.

©2002 IEEE

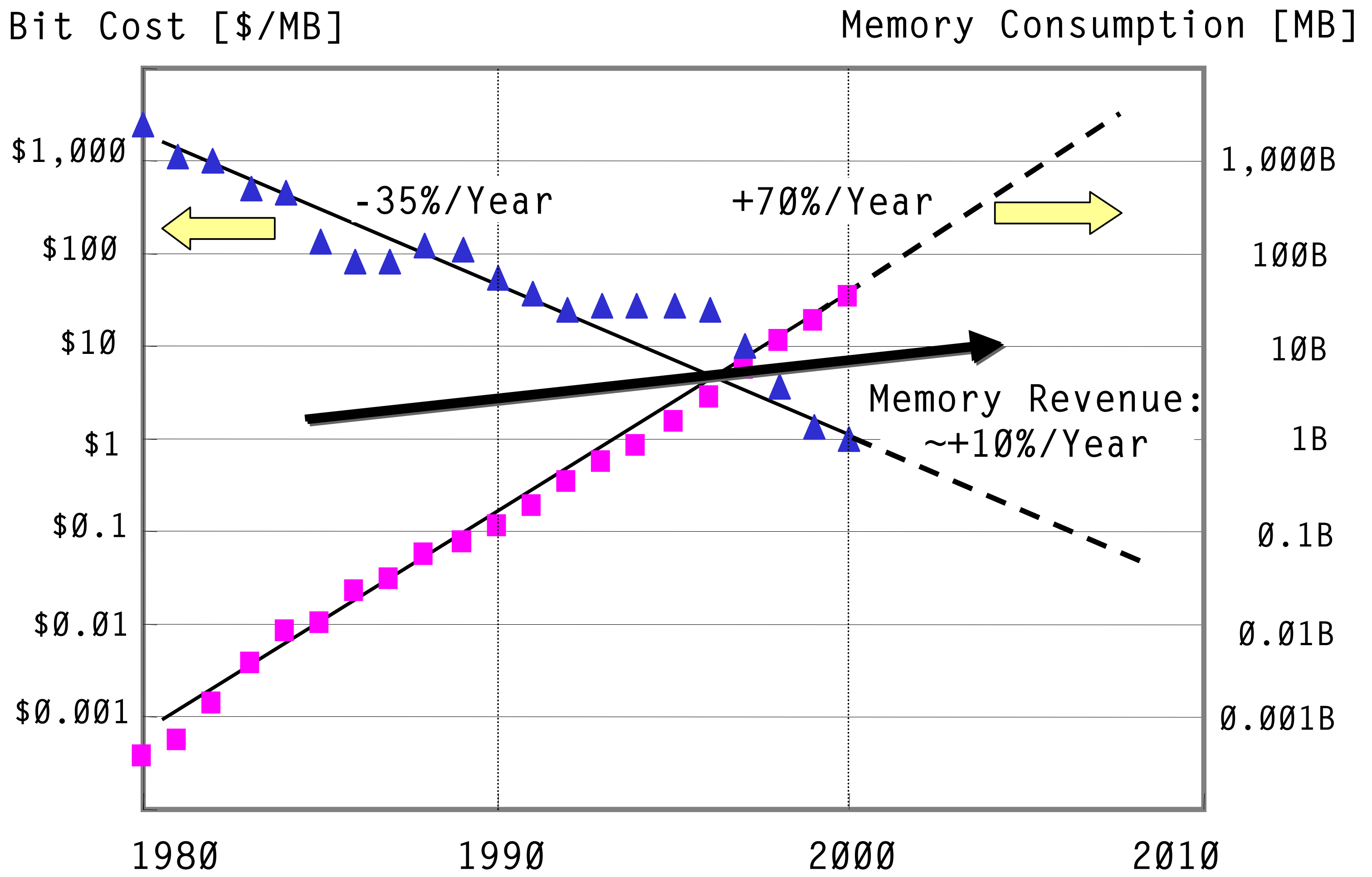

Figure 1.2.8: The trends of the total memory size demanded and memory cost reduction rate for the last two decades and into the future.

 ©2002 IEEE