특별강연

대 용량 메모리 기술 및 동향

High Density Memory Technology and Trend

윤흥일, 김창현, 황창규

Hongil Yoon, Changhyun Kim, and Chang-Kyu Hwang

Abstract

Over the years of decades, the memory technology has progressed a long, marble way. As we have evidenced from the Intel's 1Kb DRAM in 1970 to the Gigabit era of 2000's, the road further ahead towards the Terabit era will be unfolded. The technology once perceived inconceivable is in realization today, and similarly roadblocks as we know of today may become trivial issues for tomorrow. For the inquiring mind, the question is how the "puzzle" of tomorrow's memory technology is pieced-in today. The process will take place both in evolutionary and revolutionary ways. Among these, note-worthy are the changes in DRAM architecture and the cell process technology. In this paper, some technical approaches will be discussed to bring these aspects into a general overview and a perspective with possibilities for the new memory technology will be presented.

Key Words(중요용어) : memory, high density, high speed, low voltage, DRAM

1. Introduction

The memory technology today is in the forefront of the evolution and revolution of electronics industry. The scaling of electronic device has evidenced the exponential growth of chip density which is described by the Moore's law with the state-of-the-art processing technology on the verge of photolithography limit. The moving thrust is fueled by the growing market demand for faster and cost-effective hardwares as well as computing-heavy softwares. The market is diverging fast while the products are converging to provide an unified solution for multi-purpose tasks. The time-to-market has placed a stringent requirement for the product development cycle.

The increasing complexity of designing, fabricating, and testing integrated circuit (IC) calls for suitable solutions in all areas related to the flow of IC designing and manufacturing hierarchy: device and circuit design, modeling, fabrication, and testing. Approaches utilizing computer aided design (CAD) have found a growing role in the design and verification stages due to the increased complexity of devices. Process architecture and integration have to insure manufacturability with good process margin and sound defect control. Device testability and test coverage have become important factors.

The generic objective for memory has always been a quest for high density, high speed, and low power. The half-bitline pitch, the peak data bandwidth, and the operating voltage have been the key barometers for these measures. DRAM as the most prevalent memory type have evidenced a variety of evolutionary and revolutionary architectures such as Synchronous, DDR and RAMBUS DRAMs in attaining these goals with varied emphases. Recently, 1Gb DDR DRAMs[1-4] having half-

bitline pitches of 0.11~0.17mm, peak bandwidths of 200~333Mbps, and internal operating voltages of 1.8~2.0V are introduced, and 128Mb RAMBUS DRAM with peak bandwidths of 800Mbps is currently in volume production. Varied applications have sought other alternative memory types such as SRAM, NVM, and FRAM. Besides these standalones, embedded memory macros are addressing varied and application-specific needs with many opportunities for mixed technology.

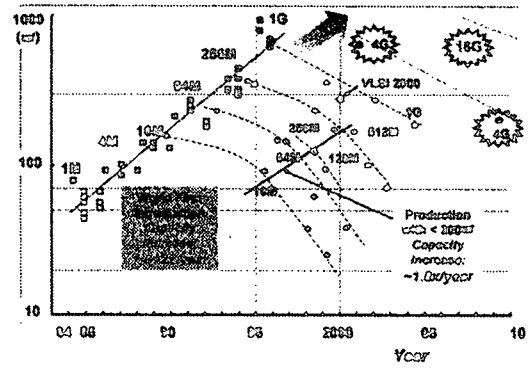
2. High Density Perspective

The focus has been to produce smaller memory cell possibly by extending the current and established process technology. The KrF lithography technique can be extended down to the 0.11mm regime[4]. But beyond, advanced photolithography techniques using ArF excimer, e-beam or X-ray sources will be needed for sub-0.1mm line and space patterning.

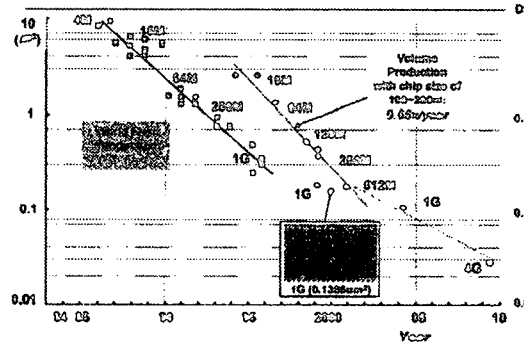
Device isolation has been performed using the shallow trench isolation (STI) technique, and further silicon on insulator (SOI) and selective epitaxial-layer growth (SEG) technologies will be available to significantly reduce the junction capacitance. Low resistance gate material such as Ti-Silicide and Cobalt-Silicide will be developed to reduce propagation delay and replace currently prevalent W-Silicide.

The number of metallization steps will be increased to 4~5 layers to accommodate the high packing density, especially for embedded memory and logic applications. Copper interconnect material will be used to improve reliability and speed. Also, low resistance bitline using W or Al will be effective in DRAM in reducing the sensing delay time.

Figure 1 shows the technology development trend for the introduction of DRAM density generations and the production in Samsung with the chip size of around 100~200mm^2. The time difference between them is about 5~6 years and is getting longer as the memory density increases due to the limit of process and device technology development. The current development trend may be continued up to 16Gbit DRAM and the production of 4Gbit DRAM with the chip size of <200mm^2 will take place in the next 10 years.



(a)



(b)

Fig. 1: Technology development trend of (a) DRAM chip size and (b) DRAM cell size.

As break-throughs in achieving much higher device density, several conceptual techniques can be proposed. First, the vertical memory cell process technique using trench process[5] can achieve very small effective cell area by merging an access transistor and a capacitor. The effective cell area can be reduced to 4F^2 compared to the 6F^2 in conventional open bit line architecture where F denotes the minimum feature size.

The use of a multi-bit cells by using multi-level logic circuit techniques[6] can increase the data storage without physically increasing the number of memory cells. In the future, the single electron memory cell technology and bio-cell development using cloning and genetics technology could be used to realize the extremely small memory cell and multi-bit cell.

Stacking technology has been a practical method to obtain larger density with multiple, stacked packages with similar volume constrain. Another novel way to

obtain high density is to utilize chip-to-chip interconnections with 10~20mm thick silicon ribbon cables by micro-machining technology. Using this technology, a folded or rolled three dimensional memory chip can be realized consisting of internally connected multiple chips.

New material approaches such as magnetic resistive layers are also being actively researched. The layers are written by passing data-dependent current through orthogonal conductors and read by discriminating between the different resistances from different magnetic polarization states of the TMR layers[7]. Using these materials, nonvolatile memory can be claimed to have access speed and cell size comparable to those for DRAM and there will be no need of high voltages required to write such as in flash memories.

3. Low Power Perspective

Traditionally low voltage trend of memory has been set to meet the requirements for the reliability and power consumption of previous generations. Especially, this need is critical for mobile and portable applications to prolong the battery lifetime. The technology for secondary batteries such as Ni-Cd, Ni-MH, or Li-Ion has been developed dramatically to cope with the extension of lifetime. The operating voltage of 1V~1.5V may be accommodated from a single Ni-Cd (~1.2V) battery. Further reduction of power consumption will require solar-cell-operated system (~0.5V).

SOI technology is one of the promising candidates for high performance devices such as CPU and memory controller in low voltage (<1.0V) applications due to numerous technology benefits such as low junction capacitance, dynamic threshold logic capability, reduced short channel effect, and reduced latch-up and SER.

In conventional DRAM operation, robust sense amplifier with large sensing voltage margin and high speed and efficient DC generators are required to achieve proper read and write operations while maintaining fundamental DRAM refresh time characteristics. These needs are getting stronger as the power supply voltage is reduced. Large sensing voltage margin can be achieved using size-effective but large capacitive cell capacitor using high dielectric materials such as BST (er=150~200) and Ta_2O_5 (er =20~25) in fin or crown type structures. A possible way to effectively decrease the threshold barrier of sense amplifier is to use the gated boosting sensing scheme to amplify the sensing signals as well as to develop signals quickly at initial sensing period even at VDD of 1V[8].

The other challenge still remains in DC generators to maintain the ON/OFF ratio of about 10^{-7} in DRAM characteristics (fast access time vs. long refresh time) with boosted voltage level with high pumping efficiency at low voltage. Other attractive peripheral circuit approaches to achieve low power consumption are the use of power-efficient logic circuits and adiabatic circuit techniques.

4. High Speed Perspective

Figure 2 shows the trend of the memory and CPU operating frequencies showing the diverging speed gap due to the process technology difference between logic and memory. However, the gap can be decreased by adopting high speed interface circuits and I/O signaling scheme to increase the external data transfer rate. Even though silicon has relatively low fundamental frequency limitation compared to III-V materials, the I/O speed of around 1.5~2GHz/pin can be attained for the point-to-point communication. Due to the rapid development of high speed interface circuit techniques (double- or quadruple-data rate and pulse-amplitude-modulation techniques), the data transfer rate of 3~4Gbps/pin will be achieved in the next 5 years, resulting in realization of nominal PC systems with a few high density memories.

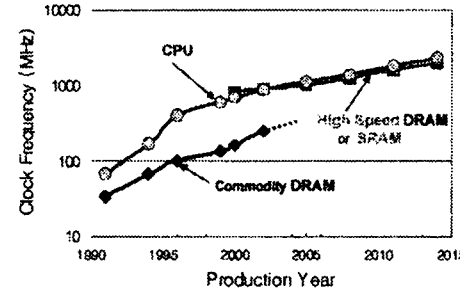


Fig. 2: Trend of memory and CPU operating frequencies.

The DRAM architecture changes for low random latency have been evolved to improve the system efficiency and bandwidth from fast page mode to synchronous clock control mode. Nowadays, several application-specific DRAMs are proposed to increase the system performance by adding row caches (ESDRAM, VCRAM) and to significantly reduce the random access time (latency of 7~15ns) by dividing memory arrays in smaller size such as in FCRAM. This high bandwidth and low latency trend will be continued in the embedded memory by implementing large internal data bus (up to 32~128Byte) with drastically reduced I/O loading (2~3pF) and non-multiplexed addressing schemes.

To make high speed devices to be more adaptable to production stage, skew minimization, process-voltage-temperature (PVT) insensitive design, low dielectric and resistive materials development, small parasitic packages, and good signal integrity on a system channel should be considered.

As the speed of device is increased, the skew in a chip coming from PVT variations is a critical factor to minimize. Especially, intra-die variation as well as inter-die variation have to be carefully accounted for at the design stage to improve the device productivity. Design approaches should employ extensive CAD analysis method, implementation of on-chip self-calibration circuits, uniform clock tree routing techniques, and timing circuit to produce an internal clock synchronized with the external clock.

One feasible process approach to reduce intra-chip variation is the use of an optical proximity correction technique by generating correction rules during mask generation to produce uniform critical dimension patterns such as for transistor gate lengths in a chip.

5. Conclusion

For the future into the new millenium, technology from areas previously conceived to be remotely related can give rise to the major change in the memory technology. The memory technology may be envisioned to broaden its scope to encompass the fields of quantum electronics, opto-electronics, bio-engineering, and materials engineering. Micro-machining technology and multi-chip stacking techniques can be integrated to the existing memory technology. The future roadmap would be multi-branched to reflect these multi-disciplinary opportunities but will require hurdling over many undefined challenges.

References

[1] H. Yoon, et al., ISSCC, pp.412, 1999.

[2] Y. Takai, et al., ISSCC, pp.418, 1999.

[3] T. Kirihata, et al., ISSCC, pp.422, 1999.

[4] K. N. Kim, et al., Symp on VLSI Tech., pp.10, 2000.

[5] W. F. Richardson, et al., IEDM, pp.714, 1985.

[6] T. Murotani, et al., ISSCC, pp.74, 1997.

[7] R. Scheuerlein, et al., ISSCC, pp.128, 2000.

[8] K. C. Lee, et al., Symp. on VLSI Circuits, pp.104, 1996.