A New Road to Terabit Era
(Evolution of Transistors and a New Paradigm for Memory Devices)

Yang-Kyu Choi
Department of Electrical Engineering, KAIST, Daejeon, Korea, 305-701
Tel: (82) 42-350-3477, Fax: (82) 42-350-8565, E-mail: ykchoi@ee.kaist.ac.kr

Over the course of the past four decades, we have witnessed dramatic changes in our lifestyle resulting from an unprecedented information technology (IT) revolution. Continuing development of semiconductor technology has been a key element of the IT revolution. A main driving force in semiconductor technology has been the silicon technology, which allows the integration of more devices within a given area by miniaturizing transistors. There are several concerns regarding when and where miniaturization will end. The fundamental limit may be attributed to von Neumann and Landauer theorem, which introduced the concept of a minimum switching energy. This limit is based on the minimum signal energy transfer during binary switching transition. This is represented by $E_{min}=kT\cdot ln2$, where k is Boltzmann's constant and T is the absolute temperature. The minimal energy, which is independent of the material, structure, and circuit configuration, can be estimated by the Heisenberg's uncertainty principle $(\Delta x\Delta p \geq h/2\pi)$. According to this prediction, the ultimately scaled device size operating at room temperature is roughly 1.5nm. The 3nm size transistor we demonstrated has reached the fundamental limit for any arbitrary structured and material. Thus it is assured that the 3 nm transistor, the world's smallest field-effect transistor, is the end-point that human can reach. In addition to being employed in logic transistors, silicon memory has been at the center of an ongoing battle to create the smallest, most density-rich, and most innovative memory products. Silicon based memory devices have advanced at a noticeable pace. DRAM, SRAM, and Flash have played core roles in the domains of density, speed, and non-volatility, respectively. However, the traditional scaling approach is now pushing these memories toward physical and technological limits. We have developed the world smallest flash memory cell, which employs an eight-nanometer design rule. This state-of-the-art Flash memory approaches a gate length of 8nm, which would be the fundamental limit of currently developed charge storage materials.

As semiconductor technology arrives at its physical and technological end, revenue from downscaling tends to be decreased. Therefore, an entirely new concept is required in order for silicon memory technology to remain competitive. According to this stringent requirement, we exploited a new paradigm, which is a concept of a multi-functional device that can create value and continue the use of the infrastructure of the semiconductor industry.

An ideal memory device should satisfy three requirements: high speed, high density, and non-volatility. Unfortunately, a memory that can satisfy all these requirements has yet to be developed. Consequently, memory devices have been advanced by pursuing one of those requirements. Therefore, it is expected that if a single transistor can perform a variety of functions, new markets can be created, sustaining the growth momentum in the silicon memory industry. For example, DRAM and Flash memory functions can be operated in a single chip. This fusion memory has been developed to meet the demands of multi-functions and high-performance digital applications. A fusion memory chip combines various types of memory such as DRAM, Flash memory, and SRAM, etc. The first type of fusion memory is a physical combination of chips, which are combined by a multi-chip package (MCP) or system-in-package (SIP) technology. However, in MCP/SIP, the time delay caused by wiring can degrade the data transmission speed. This implies that effective communication between the CPU and memory will be difficult in a multimedia chip. In a notable breakthrough, two or more kinds of memory chips have been combined on one chip, which is implemented by system-on-chip (SOC) technology. However, the hybrid integration of those functional memories is obstructed by process complexity and cost. Furthermore, the fundamental limitation in the aforementioned fusion memories is that a portion of the different memory blocks is fixed, and hence users cannot maximize the strength of each memory density and speed. But if a single memory transistor can perform different memory functions, users can optimize specifications of the memory to fit their demands. Based on intensive experience, a prototype of a unified-RAM (URAM) is demonstrated.

In URAM technology, a single memory transistor can process non-volatile memory and high speed capacitorless 1 transistor-DRAM (1T-DRAM). These operations are identified by the distinctive operational voltage domains. By combining a floating body as a hole storage zone for DRAM and an oxide/nitride/oxide stacked gate dielectric as an electron trapped zone for non-volatile memory, URAM operation is realized in a single transistor. In summary, this paradigm shift from a multi-bit cell to a multi-function cell can make Moore's law be alive and continue the evolution of silicon technology.