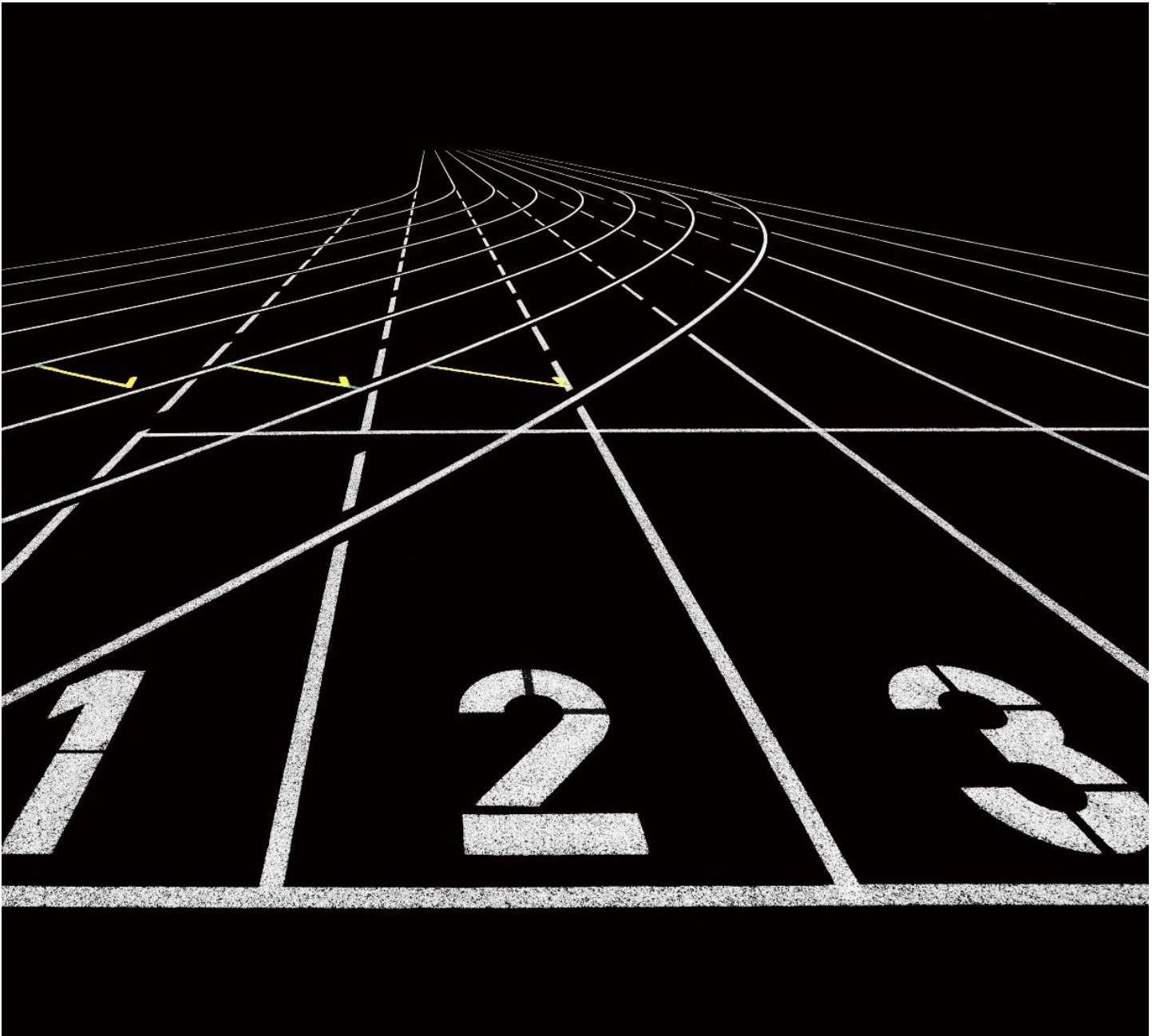


A STUDY ON RAM (RANDOM-ACCESS- MEMORY)

E SAILESWARA REDDY

cs21b1078@iiitdm.ac.in

CS21B1078



RAM

is a form of computer memory that can be read and changed in any order, typically used to store working data and machine code. A random-access memory device allows data items to be read or written in almost the same amount of time irrespective of the physical location of data inside the memory, in contrast with other direct-access data storage media (such as hard disks, CD-RWs, DVD-RWs and the older magnetic tapes and drum memory). RAM contains multiplexing and demultiplexing circuitry, to connect the data lines to the addressed storage for reading or writing the entry.

HISTORY

Early computers used relays, mechanical counters or delay lines for main memory functions. Ultrasonic delay lines were serial devices which could only reproduce data in the order it was written. Drum memory could be expanded at relatively low cost but efficient retrieval of memory items required knowledge of the physical layout of the drum to optimize speed. Latches built out of vacuum tube triodes, and later, out of discrete transistors, were used for smaller and faster memories such as registers. Such registers were relatively large and too costly to use for large amounts of data; generally only a few dozen or few hundred bits of such memory could be provided.

The first practical form of random-access memory was the Williams tube starting in 1947. It stored data as electrically charged spots on the face of a cathode-ray tube. Since the electron beam of the CRT could read and write the spots on the tube in any order, memory was random access. The capacity of the Williams tube was a few hundred to around a thousand bits, but it was much smaller, faster, and more power-efficient than using individual vacuum tube latches. Developed at the University of Manchester in England, the Williams tube provided the medium on which the first electronically stored program was implemented in the Manchester Baby computer, which first successfully ran a program on 21 June 1948. In fact, rather than the Williams tube memory being designed for the Baby, the Baby was a testbed to demonstrate the reliability of the memory

Magnetic-core memory was invented in 1947 and developed up until the mid-1970s. It became a widespread form of random-access memory, relying on an array of magnetized rings. By changing the sense of each ring's magnetization, data could be stored with one bit stored per ring. Since every ring had a combination of address wires to select and read or write it, access to any memory location in any sequence was possible. Magnetic core memory was the standard form of computer memory system until displaced by solid-state MOS (metal–oxide–silicon) semiconductor memory in integrated circuits (ICs) during the early 1970s.

Prior to the development of integrated read-only memory (ROM) circuits, permanent (or read-only) random-access memory was often constructed using diode matrices driven by address decoders, or specially wound core rope memory planes.

Semiconductor memory began in the 1960s with bipolar memory, which used bipolar transistors. While it improved performance, it could not compete with the lower price of magnetic core memory.

MOS RAM

The invention of the MOSFET (metal-oxide-semiconductor field-effect transistor), also known as the MOS transistor, by Mohamed M. Atalla and Dawon Kahng at Bell Labs in 1959, led to the development of metal-oxide-semiconductor (MOS) memory by John Schmidt at Fairchild Semiconductor in 1964. In addition to higher performance, MOS semiconductor memory was cheaper and consumed less power than magnetic core memory. The development of silicon-gate MOS integrated circuit (MOS IC) technology by Federico Faggin at Fairchild in 1968 enabled the production of MOS memory chips. MOS memory overtook magnetic core memory as the dominant memory technology in the early 1970s.

An integrated bipolar static random-access memory (SRAM) was invented by Robert H. Norman at Fairchild Semiconductor in 1963. It was followed by the development of MOS SRAM by John Schmidt at Fairchild in 1964. SRAM became an alternative to magnetic-core memory, but required six MOS transistors for each bit of data. Commercial use of SRAM began in 1965, when IBM introduced the SP95 memory chip for the System/360 Model 95.

Dynamic random-access memory (DRAM) allowed replacement of a 4 or 6-transistor latch circuit by a single transistor for each memory bit, greatly increasing memory density at the cost of volatility. Data was stored in the tiny capacitance of each transistor, and had to be periodically refreshed every few milliseconds before the charge could leak away. Toshiba's Toscal BC-1411 electronic calculator, which was introduced in 1965, used a form of capacitive bipolar DRAM, storing 180-bit data on discrete memory cells, consisting of germanium bipolar transistors and capacitors. While it offered improved performance over magnetic-core memory, bipolar DRAM could not compete with the lower price of the then dominant magnetic-core memory.

MOS technology is the basis for modern DRAM. In 1966, Dr. Robert H. Dennard at the IBM Thomas J. Watson Research Center was working on MOS memory. While examining the characteristics of MOS technology, he found it was capable of building capacitors, and that storing a charge or no charge on the MOS capacitor could represent the 1 and 0 of a bit, while the MOS transistor could control writing the charge to the capacitor. This led to his development of a single-transistor DRAM memory cell. In 1967, Dennard filed a patent under IBM for a single-transistor DRAM memory cell, based on MOS technology. The first commercial DRAM IC chip was the Intel 1103, which was manufactured on an 8 μm MOS process with a capacity of 1 kbit, and was released in 1970.

Synchronous dynamic random-access memory (SDRAM) was developed by Samsung Electronics. The first commercial SDRAM chip was the Samsung KM48SL2000, which had a capacity of 16 Mbit. It was introduced by Samsung in 1992, and mass-produced in 1993. The first commercial DDR SDRAM (double data rate SDRAM) memory chip was Samsung's 64 Mbit DDR SDRAM chip, released in June 1998. GDDR (graphics DDR) is a form of DDR SGRAM (synchronous graphics RAM), which was first released by Samsung as a 16 Mbit memory chip in 1998.

Uses of RAM

In addition to serving as temporary storage and working space for the operating system and applications, RAM is used in numerous other ways.

Virtual memory

Most modern operating systems employ a method of extending RAM capacity, known as "virtual memory". A portion of the computer's hard drive is set aside for a paging file or a scratch partition, and the combination of physical RAM and the paging file form the system's total memory. (For example, if a computer has 2 GB (10243 B) of RAM and a 1 GB page file, the operating system has 3 GB total memory available to it.) When the system runs low on physical memory, it can "swap" portions of RAM to the paging file to make room for new data, as well as to read previously swapped information back into RAM. Excessive use of this mechanism results in thrashing and generally hampers overall system performance, mainly because hard drives are far slower than RAM.

RAM disk

Software can "partition" a portion of a computer's RAM, allowing it to act as a much faster hard drive that is called a RAM disk. A RAM disk loses the stored data when the computer is shut down, unless memory is arranged to have a standby battery source, or changes to the RAM disk are written out to a nonvolatile disk. The RAM disk is reloaded from the physical disk upon RAM disk initialization.

Shadow RAM

Sometimes, the contents of a relatively slow ROM chip are copied to read/write memory to allow for shorter access times. The ROM chip is then disabled while the initialized memory locations are switched in on the same block of addresses (often write-protected). This process, sometimes called shadowing, is fairly common in both computers and embedded systems.

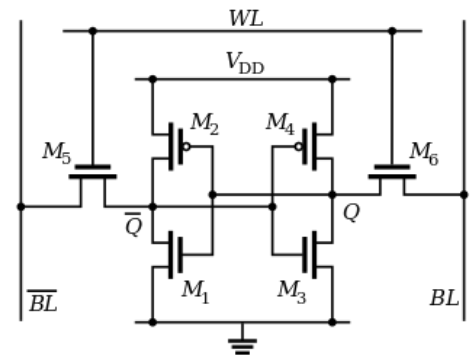
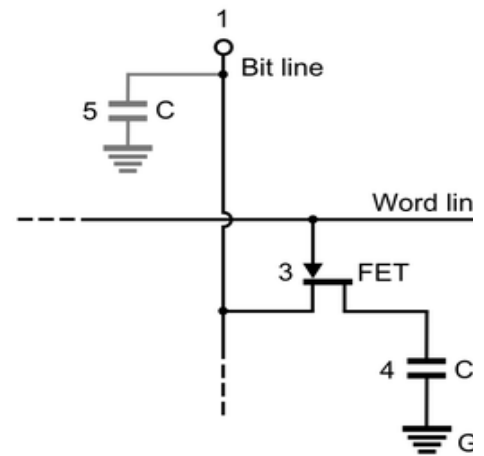
As a common example, the BIOS in typical personal computers often has an option called "use shadow BIOS" or similar. When enabled, functions that rely on data from the BIOS's ROM instead use DRAM locations (most can also toggle shadowing of video card ROM or other ROM sections). Depending on the system, this may not result in increased performance, and may cause incompatibilities. For example, some hardware may be inaccessible to the operating system if shadow RAM is used. On some systems the benefit may be hypothetical because the BIOS is not used after booting in favor of direct hardware access. Free memory is reduced by the size of the shadowed ROMs.

Memory cell

The memory cell is the fundamental building block of computer memory. The memory cell is an electronic circuit that stores one bit of binary information and it must be set to store a logic 1 (high voltage level) and reset to store a logic 0 (low voltage level). Its value is maintained/stored until it is changed by the set/reset process. The value in the memory cell can be accessed by reading it.

In SRAM, the memory cell is a type of flip-flop circuit, usually implemented using FETs. This means that SRAM requires very low power when not being accessed, but it is expensive and has low storage density.

A second type, DRAM, is based around a capacitor. Charging and discharging this capacitor can store a "1" or a "0" in the cell. However, the charge in this capacitor slowly leaks away, and must be refreshed periodically. Because of this refresh process, DRAM uses more power, but it can achieve greater storage densities and lower unit costs compared to SRAM.



TYPES

The two widely used forms of modern RAM are static RAM (SRAM) and dynamic RAM (DRAM). In SRAM, a bit of data is stored using the state of a six-transistor memory cell, typically using six MOSFETs. This form of RAM is more expensive to produce, but is generally faster and requires less dynamic power than DRAM. In modern computers, SRAM is often used as cache memory for the CPU. DRAM stores a bit of data using a transistor and capacitor pair (typically a MOSFET and MOS capacitor, respectively), which together comprise a DRAM cell. The capacitor holds a high or low charge (1 or 0, respectively), and the transistor acts as a switch that lets the control circuitry on the chip read the capacitor's state of charge or change it. As this form of memory is less expensive to produce than static RAM, it is the predominant form of computer memory used in modern computers.

Both static and dynamic RAM are considered volatile, as their state is lost or reset when power is removed from the system. By contrast, read-only memory (ROM) stores data by permanently enabling or disabling selected transistors, such that the memory cannot be altered. Writeable variants of ROM (such as EEPROM and NOR flash) share properties of both ROM and RAM, enabling data to persist without power and to be updated without requiring special equipment. ECC memory (which can be either SRAM or DRAM) includes special circuitry to detect and/or correct random faults (memory errors) in the stored data, using parity bits or error correction codes.

In general, the term RAM refers solely to solid-state memory devices (either DRAM or SRAM), and more specifically the main memory in most computers. In optical storage, the term DVD-RAM is somewhat of a misnomer since, unlike CD-RW or DVD-RW it does not need to be erased before reuse. Nevertheless, a DVD-RAM behaves much like a hard disc drive if somewhat slower.

Recent developments

Several new types of non-volatile RAM, which preserve data while powered down, are under development. The technologies used include carbon nanotubes and approaches utilizing Tunnel magnetoresistance. Amongst the 1st generation MRAM, a 128 kbit (128 × 210 bytes) chip was manufactured with 0.18 μm technology in the summer of 2003.[citation needed] In June 2004, Infineon Technologies unveiled a 16 MB (16 × 220 bytes) prototype again based on 0.18 μm technology. There are two 2nd generation techniques currently in development: thermal-assisted switching (TAS) which is being developed by Crocus Technology, and spin-transfer torque (STT) on which Crocus, Hynix, IBM, and several other companies are working. Nantero built a functioning carbon nanotube memory prototype 10 GB (10 × 230 bytes) array in 2004. Whether some of these technologies can eventually take significant market share from either DRAM, SRAM, or flash-memory technology, however, remains to be seen.

Since 2006, "solid-state drives" (based on flash memory) with capacities exceeding 256 gigabytes and performance far exceeding traditional disks have become available. This development has started to blur the definition between traditional random-access memory and "disks", dramatically reducing the difference in performance.

Some kinds of random-access memory, such as "EcoRAM", are specifically designed for server farms, where low power consumption is more important than speed.

Memory wall

The "memory wall" is the growing disparity of speed between CPU and memory outside the CPU chip. An important reason for this disparity is the limited communication bandwidth beyond chip boundaries, which is also referred to as bandwidth wall. From 1986 to 2000, CPU speed improved at an annual rate of 55% while memory speed only improved at 10%. Given these trends, it was expected that memory latency would become an overwhelming bottleneck in computer performance.

CPU speed improvements slowed significantly partly due to major physical barriers and partly because current CPU designs have already hit the memory wall in some sense. Intel summarized these causes in a 2005 document.

First of all, as chip geometries shrink and clock frequencies rise, the transistor leakage current increases, leading to excess power consumption and heat... Secondly, the advantages of higher clock speeds are in part negated by memory latency, since memory access times have not been able to keep pace with increasing clock frequencies. Third, for certain applications, traditional serial architectures are becoming less efficient as processors get faster (due to the so-called Von Neumann bottleneck), further undercutting any gains that frequency increases might otherwise buy. In addition, partly due to limitations in the means of producing inductance within solid state devices, resistance-capacitance (RC) delays in signal transmission are growing as feature sizes shrink, imposing an additional bottleneck that frequency increases don't address.

The RC delays in signal transmission were also noted in "Clock Rate versus IPC: The End of the Road for Conventional Microarchitectures" which projected a maximum of 12.5% average annual CPU performance improvement between 2000 and 2014.

A different concept is the processor-memory performance gap, which can be addressed by 3D integrated circuits that reduce the distance between the logic and memory aspects that are further apart in a 2D chip. Memory subsystem design requires a focus on the gap, which is widening over time. The main method of bridging the gap is the use of caches; small amounts of high-speed memory that houses recent operations and instructions nearby the processor, speeding up the execution of those operations or instructions in cases where they are called upon frequently. Multiple levels of caching have been developed to deal with the widening gap, and the performance of high-speed modern computers relies on evolving caching techniques. There can be up to a 53% difference between the growth in speed of processor and the lagging speed of main memory access.

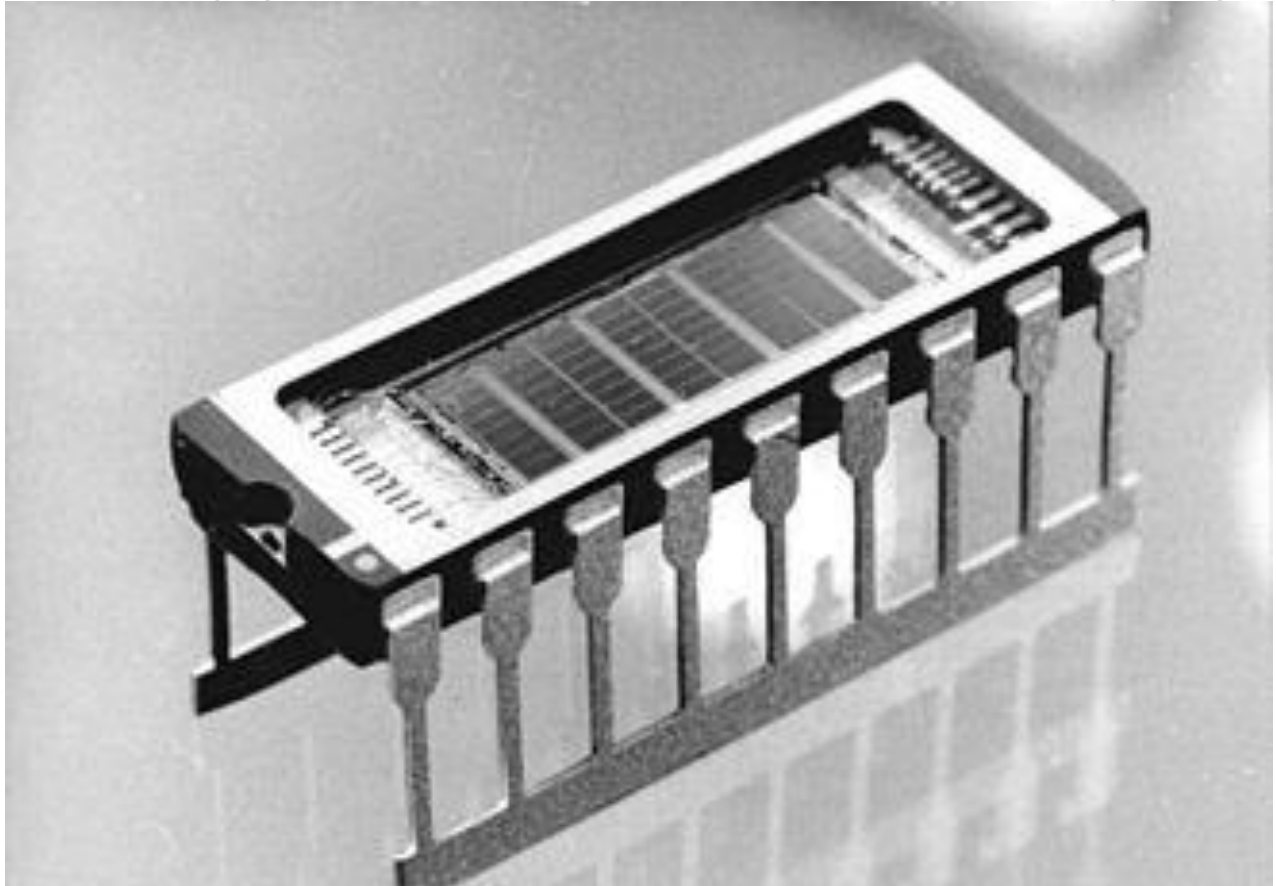
Solid-state hard drives have continued to increase in speed, from ~400 Mbit/s via SATA3 in 2012 up to ~3 GB/s via NVMe/PCIe in 2018, closing the gap between RAM and hard disk speeds, although RAM continues to be an order of magnitude faster, with single-lane DDR4 3200 capable of 25 GB/s, and modern GDDR even faster. Fast, cheap, non-volatile solid state drives have replaced some functions formerly performed by RAM, such as holding certain data for immediate availability in server farms - 1 terabyte of SSD storage can be had for \$200, while 1 TB of RAM would cost thousands of dollars.



RAM NOW(UP)



RAM THEN(DOWN)



SOURCES USED

<https://dictionary.cambridge.org/dictionary/english/ram>

<https://arstechnica.com/information-technology/2013/04/memory-that-never-forgets-non-volatile-dimms-hit-the-market/>

https://www.ibm.com/ibm/history/reference/faq_0000000011.html

<https://web.archive.org/web/20120504133240/http://www.computer50.org/>

<https://www.computerhistory.org/siliconengine/metal-oxide-semiconductor-mos-transistor-demonstrated/>

<https://www.computerhistory.org/siliconengine/silicon-gate-technology-developed-for-ics/>

https://en.wikipedia.org/wiki/Random-access_memory

<http://hardwarehell.com/articles/shadowram.html>

<https://hpi.de/plattner/home.html>

<https://www.minitool.com/news/ssd-prices-fall.html>

<https://www.slashgear.com/samsung-50nm-2gb-ddr3-chips-are-industrys-smallest-2917676/>

<https://books.google.com/books?id=Fa0kAQAAIAAJ>

<https://patents.google.com/patent/US4808857>

X-----X-----X