High-Density Dynamic Random Access Memory Circuit Design

Yincheng Yan *
School of Navigation, Wuhan University of Technology, Wuhan, China
* Corresponding Author Email: yycruheycs@outlook.com

Abstract. Traditional random-access memory (RAM) includes two main types: Static RAM (SRAM) and Dynamic RAM (DRAM). Leveraging the advantage of high integration density, DRAM has found widespread application in System-on-Chip (SoC) systems. With the advancement of integrated circuit technology, embedded DRAM (eDRAM) has been extensively researched due to its advantages of low latency and high bandwidth, these advantages enable it to have great application potential in fields sensitive to storage performance, such as high-performance computing, high-end smartphones, and artificial intelligence chips. This paper reviews the latest research on high-density eDRAM: in terms of cell structure, 4T/3T1C and other multi-transistor designs and dynamic regulation circuits achieve optimized retention time, area and power consumption; in terms of materials innovation, emerging devices such as AOS/IWO/AsymFET reduce leakage current, achieving quasi-non-volatility and second-level retention time; in terms of circuit design for low power consumption optimization, circuits such as negative voltage bootstrap drive improve energy efficiency, and system-level modeling ensures reliability.

Keywords: High-Density DRAM, Multi-Transistor Gain Cells, Oxide Semiconductors, Low Power Consumption Optimization.

1. Introduction

In modern integrated circuit systems, RAM, as a core component for temporary data storage and high-speed reading/writing, plays a crucial role in connecting processors and external storage devices, and its performance directly affects the operational efficiency of the entire system. The traditional random-access memories mainly consist of two types: static random-access memory (SRAM) and dynamic random-access memory (DRAM). Due to the different storage methods, these two types of memories have significant differences in performance and cost: SRAM stores data using a flip-flop structure, thus having the advantages of fast read and write speeds and no need for refreshing. However, the complex unit structure results in low integration density, large chip area occupation, and high manufacturing costs. Therefore, it is more suitable for scenarios that require high-speed but small-capacity storage, such as a processor cache. In contrast, the transistor-capacitor structure used in DRAM is simpler. It represents the data state by the storage of electric charge in the capacitor. It requires regular refreshing to maintain data stability, but compared to SRAM, it has a larger storage capacity per unit area, thus making it less costly to achieve large-capacity storage. Therefore, DRAM has become the best storage solution for SoC. Embedded random-access memory (eDRAM), as an extension of DRAM, integrates memory cells and logic circuits on the same chip, significantly shortening the data transmission path and effectively reducing read/write latency. As a result, it has become a research hotspot in the industry. Furthermore, due to its higher bandwidth compared to traditional DRAM, the storage array structure and interface design of this technology can meet the data transmission rate requirements of high-performance SoC [1,2].

At present, the research on high-density eDRAM mainly focuses on three aspects: improvement of the cell structure, application research of new materials, and adaptation processing for special scenarios (such as low-temperature environments). During this process, many innovative methods emerged, such as multi-transistor gain cell eDRAM (GC-eDRAM), oxide semiconductor (AOS/IWO) devices, and circuit design for low power consumption optimization. As a result, improvements were made in aspects such as storage density, energy efficiency and retention time [3,4]. The paper is organized as follows: Section 2 overviews the structure of DRAM bitcell and the process of its read-write operations. Section 3 describes the latest achievements of eDRAM and Section 4 draws the conclusions.

2. Preliminary

The figure 1 illustrates the 1T1C (1 Transistor + 1 Capacitor) bit cell structure of DRAM (Dynamic Random Access Memory), which serves as the core architectural unit for DRAM to enable data storage. Here is an analysis of each component and its operating principles:

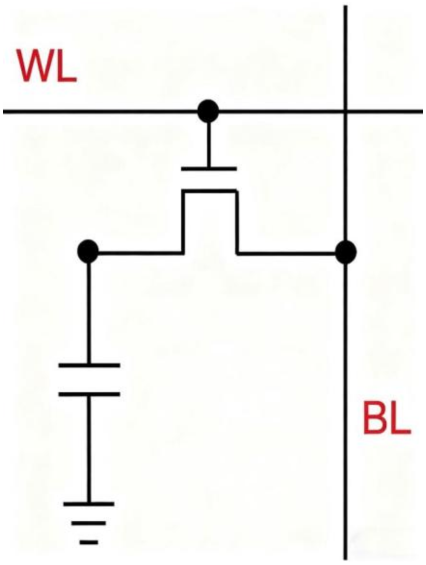

Fig 1. Bit cell structure of DRAM

2.1. Structural Composition and Function of Each Part

WL (Word Line): Used for “row selection”, determining which row of storage cells gets activated. When a high voltage level is applied to the word line, the corresponding MOS transistor is turned on.

BL (Bit Line): Transmits data signals (electrical signals for reading or writing) and acts as a “bridge” between the storage cell and external data circuits.

MOS Transistor (T): Functions as a “switch” to control the on-off state between the storage capacitor (C) and the bit line (BL).

Storage Capacitor (C): The core storage element, representing binary data based on “whether the capacitor stores charge”: if the capacitor holds sufficient charge → it represents “1”; if the charge in the capacitor leaks away (or is very little) → it represents “0”.

2.2. Principles of Read, Write, and Refresh Operations

2.2.1 Write Operation

To write data into the cell:

First, activate the word line (WL) to turn on the MOS transistor, connecting the storage capacitor to the bit line. To write “1”, charge the storage capacitor via the bit line (BL) (the bit line provides a high voltage level, and charge flows into the capacitor); To write “0”, discharge the storage capacitor via the bit line (BL) (the bit line is connected to a low voltage level, and charge flows out of the capacitor).

2.2.2 Read Operation

To read data from the cell:

First, activate the word line (WL) to turn on the MOS transistor, connecting the storage capacitor to the bit line. Second, the charge of the storage capacitor undergoes charge sharing with the parasitic capacitance of the bit line, causing a slight change in the bit line voltage. Third, an external Sense

Amplifier (not shown in the diagram, which is part of the DRAM peripheral circuitry) detects the voltage change of the bit line to determine whether “1” (the voltage change matches the characteristic of “capacitor with charge”) or “0” (the voltage change matches the characteristic of “capacitor without charge”) is stored.

The read operation is “destructive” — charge sharing alters the original charge amount of the storage capacitor. Therefore, after a read operation, “rewriting” (regeneration) is necessary to restore the charge state of the capacitor.

2.2.3 Refresh Operation

Because the storage capacitor exhibits a leakage phenomenon, charge gradually leaks away, leading to data loss. Thus, DRAM requires periodic refresh: Perform “read + rewrite” operations on the storage cells at regular intervals to replenish the charge of the storage capacitor and maintain data validity. This is also one of the core differences between DRAM and SRAM (SRAM, which does not require refreshing).

The 1T1C structure is crucial for DRAM to achieve “large-capacity and low-cost storage” — compared to SRAM (which typically requires 6 transistors), it significantly improves integration density with the minimalist “1 transistor + 1 capacitor” architecture. Meanwhile, through the cooperation of “word line gating + bit line transmission + sense amplification + periodic refresh”, it enables data reading, writing, and long-term retention, and has become the mainstream technical foundation for general-purpose memory (such as computer memory).

3. Latest Research Achievements

3.1. Cell Structure Innovation

Research teams have addressed issues such as short retention time, frequent refreshing, high power consumption, and inefficient area utilization of traditional architectures by optimizing the cell structure, circuit design, and process adaptation of eDRAM, and also met the low-power and high-bandwidth requirements of special scenarios such as quantum computing.

To address issues such as short retention time and frequent refreshing in traditional 2T/3T eDRAM, researchers have achieved balanced retention time and efficient area utilization by adding transistors and designing dynamic adjustment circuits. A team from University College Dublin, Ireland, proposed a 4T nMOS-only gain cell eDRAM. Based on the 3T structure, it incorporates a diode-connected M_4 transistor and a Vadj voltage adjustment line. M_4 is used to compensate for charge leakage at the storage node (SN): when the logic state is ‘1’, Vadj drives M_4 to inject leakage current, offsetting charge loss; when the logic state is ‘0’, it slows down voltage decay. Finally, at Vadj = 650 mV, the retention times for both states reach 25μs, achieving full balance. The cell area is only 0.12μm^2 (consistent with the 3T structure), and the total dynamic power consumption of a 4kb array (22nm FD-SOI process) is 1.1mW, with the power consumption of the balance circuit accounting for <1% [5].

A research team from Bar-Ilan University, Israel, designed a 3T1C GC-eDRAM. This design introduces a replica column to generate self-timed signals, optimizing the timing of read operations. Meanwhile, it employs parallel refresh technology to enable the simultaneous reading and writing of multiple words in the same row, thereby effectively reducing the redundant operations of the row decoder. Under the 65nm process, the area of a 64kb array is 40% smaller than that of 6T SRAM, the retention time reaches 16μs (with a yield of 99.99%), the refresh power consumption is 60% lower than that of traditional schemes, and the memory availability is increased from 47% to 93% [6]

In addition, a team from Fudan University proposed a 3T1C cryogenic eDRAM. It utilizes ultra-threshold (ultra-Vth) transistors to minimize leakage and parasitic capacitance, thereby enhancing SNR charge storage. No refresh is required in a 4K environment (the quantum state read cycle is shorter than the retention time), and an integrated MBIST circuit enables full-bandwidth testing. Under the 28nm process, its energy consumption is 4.9fJ/bit for read operations and 15.9fJ/bit for write operations, the stable retention time reaches 4.8ms, and the data transmission bandwidth is up to 61.4Gb/s. This performance can well meet the low-power and high-bandwidth requirements of quantum error correction systems [7].

A team from Shanghai Tech University proposed a 4T CSDB-GC structure. It compensates for SN charge loss through an internal data regeneration loop (N2/P2) and achieves interference-free dual-port reading by optimizing word line voltages (WLn = 0.75V, WLp = 0.2V). At 4.2K, the retention time is 16.67s (2.6 times longer than that of state-of-the-art (SOTA) cryogenic eDRAM), and the dynamic power consumption is 49.23 μW/Kb, which is 11.4 times lower than that of traditional schemes [8].

3.2. Material Breakthroughs

Research teams have broken through the performance bottlenecks of traditional transistors by introducing transistors made of different materials, addressing the core issues of large leakage current and short retention time. Meanwhile, they have optimized the area, power consumption and latency of eDRAM, meeting the application requirements of large-scale integration and specific scenarios.

To further reduce leakage current and extend retention time, researchers have introduced non-silicon-based devices to overcome the performance bottlenecks of traditional Silicon transistors. The AOS-GC eDRAM, developed by a team from Fuzhou University, utilizes amorphous oxide semiconductor (AOS) field-effect transistors (FETs). The leakage current of these transistors is reduced by 7 orders of magnitude compared with Si FETs, achieving quasi-non-volatile performance. A GCSim simulation platform was developed to quantitatively analyze area, latency, and power consumption. The results show that the area of the vertical channel all-around (CAA) type AOS array is 92.2% smaller than that of SRAM, and its static power consumption is 97.6% lower. In intermittent wake-up DNN scenarios (where the wake-up interval is greater than 1 second), it saves 96.7% more energy than SRAM [3].

A team from the Georgia Institute of Technology used tungsten-doped indium oxide (IWO) BEOL transistors to design a 3T reconfigurable GC-eDRAM. The low leakage of IWO transistors (<1 fA/μm) enables a retention time of over 1s (10^4 times that of Si 3T eDRAM), and M3D integration compresses the cell area to $6F^2$, which is only 40% of that of 10T SRAM [2].

The AsymFET 2T0C eDRAM proposed by a team from Peking University suppresses gate-induced drain leakage (GIDL) and tunnel current through the design of asymmetric source-drain doping and gate spacer layers, achieving a leakage current as low as 10^{-17} A/μm. Under the 55nm process, the retention time is 3.9s (1000 times longer than that of MOSFET 2T0C), the write speed is <5ns with negligible temperature dependence. Wafer-level tests indicate that the full-wafer retention time ranges from 1s to 7s, meeting the requirements of large-scale integration [4].

3.3. Low Power Consumption and Energy Efficiency Optimization

At the circuit level, a team from EPFL proposed a negative-voltage bootstrap driver circuit. Under the 28nm FD-SOI process, no external negative power supply is needed; instead, a -350mV WWL voltage is generated through capacitor charge inversion. This increases the GC-eDRAM retention time by 11.3 times, with an area overhead of only 2.5% (mainly from a 600fF MOSCAP). Monte Carlo simulations show that the WWL voltage fluctuation is <10mV under PVT variations, indicating excellent retention time stability [9].

The bit-serial interleaved storage eDRAM (eDRAM-OESP) proposed by a team from IIIT Bangalore, India, stores operands in a bit-interleaved manner, enabling parallel reading and writing of multiple operands within the same activation window and effectively reducing precharge/activation cycles. Particle swarm optimization (PSO) is used to optimize the transistor sizes of the 28T full adder and 10T multiplexer. The time consumed for 16-bit addition is 31% lower than that of the current state-of-the-art technology. A 1D convolution (256-order filter) takes 2.5ms with an energy consumption of 120nJ, which is 22–24 times faster than an Arduino controller [10].

In the field of system-level optimization, a team from UTM, Malaysia, developed a hybrid threshold voltage (VT) GC-eDRAM. Based on the 130nm process, two schemes (3T with transmission gate write port and 4T with internal feedback) were proposed:

(a) The 3T scheme achieves a retention time of 9.2μs through the self-suppression effect of charge injection in the TG, with a power consumption of 1.51W for a 4Kb array.

(b) The 4T scheme suppresses SN leakage through feedback transistors, achieving a retention time of 25.27μs and a power consumption of 0.597W.

Both indicators are 20–50 times lower than those of traditional 6T SRAM [11,12].

A team from Fudan University constructed a retention/read channel model for 2T0C eDRAM. Through perturbation analysis, it was found that the SN charge follows a log-normal distribution during the retention phase, and the read current follows a Gaussian distribution. Verified by Monte Carlo simulations, this model can accurately predict the error rate, laying a theoretical foundation for memory reliability design [13].

4. Conclusion

High-density eDRAM has formed a trinity technical system of “structural innovation, material breakthroughs, and circuit design optimization”: multi-transistor configurations and dynamic adjustment circuits optimize retention time, area and power consumption; devices such as AOS/IWO/AsymFET drastically reduce leakage current, static power consumption and refresh needs; circuits such as negative voltage bootstrap drive improve energy efficiency, and system-level modeling ensures reliability. High-density eDRAM has become a core embedded memory solution.

References

[1] Yang M, Wang Y, Kulkarni J P. A 118 GOPS/mm² 3D eDRAM TensorCore Architecture for Large-scale Matrix Multiplication[C]//2023 IEEE 30th International Conference on High Performance Computing, Data, and Analytics (HiPC). IEEE, 2023.

[2] Kwak J, Choe G, Lee J, et al. Monolithic 3D transposable 3T embedded DRAM with back-end-of-line oxide channel transistor[C]//2024 IEEE International Symposium on Circuits and Systems (ISCAS). IEEE, 2024.

[3] Chen L, Yang Y, Li W, et al. System-Level Evaluation of AOS Gain Cell eDRAMs for Low-Power Normally-Off Computing[C]//2024 IEEE 17th International Conference on Solid-State & Integrated Circuit Technology (ICSICT). IEEE, 2024.

[4] Wang K, Hao P, Zhang F, et al. Logic-Compatible Asymmetrical FET for Gain Cell eDRAM With Long Retention and Fast Access Speed[J]. Journal of the Electron Devices Society, 2025, 13: 237-244.

[5] Lai K, Yao E, Li Z, et al. A 4kb 4T eDRAM with Balanced Retention Time Adjustment[J]. IEEE Transactions on Circuits and Systems I: Regular Papers, 2025, 72(3): 1345-1356.

[6] Harel O, Casarrubias E N, Eggimann M, et al. 64-kB 65-nm GC-eDRAM with half-select support and parallel refresh technique[J]. IEEE Solid-State Circuits Letters, 2022, 5(6): 170-173.

[7] Zheng J, Nie X, Zhi Z, et al. A 28-nm 10.4-fJ/b Cryogenic embedded DRAM with 3T1C Gain Cell and MBIST at 4-Kelvin[C]//2024 21st International SoC Design Conference (ISOCC). IEEE, 2024.

[8] Shu Y, Zhang H, Sun H, et al. CSDB-eDRAM: A 16Kb energy-efficient 4T CSDB Gain Cell eDRAM with over 16.6s retention time and 49.23 uW/Kb at 4.2 K for cryogenic computing[C]//2023 IEEE International Symposium on Circuits and Systems (ISCAS). IEEE, 2023.

[9] Yigit A, Casarrubias E N, Giterman R, et al. A 128-kbit GC-eDRAM with negative boosted bootstrap driver for 11.3× lower-refresh frequency at a 2.5% area overhead in 28-nm FD-SOI[J]. IEEE Solid-State Circuits Letters, 2023, 6(1): 13-16.

[10] Kabra M, H C P, Deshpande K, et al. eDRAM-OESP: A novel performance efficient in-embedded-DRAM-compute design for on-edge signal processing application[C]//2023 24th International Symposium on Quality Electronic Design (ISQED). IEEE, 2023.

[11] Soo S R, Hamzah A, Alias N E, et al. Design of low power gain-cell eDRAM for 4Kb memory array in 130nm CMOS[C]//2021 International Conference on Electrical Engineering and Informatics (ICEEI). IEEE, 2021.

[12] Hamzah A, Alias N E, Kamisian I, et al. A 2 Kbit Memory Array of Mixed-VT GC-eDRAM Implemented in 130nm Standard CMOS Technology[C]//2021 IEEE Regional Symposium on Microelectronics (RSM). IEEE, 2021.

[13] Xu Y T, Zheng X B, Wang S H. Channel Modeling for 2T0C Gain-Cell eDRAM[C]//2024 9th International Conference on Electronic Technology and Information Science (ICETIS). IEEE, 2024.