1-Gb/s/pin Multi-Gigabit DRAM Design with Low Impedance Hierarchical I/O Architecture

Hiroki Fujisawa, T. Takahashi, H. Yoko*, I. Fujii*, Y. Takai, and M. Nakamura

Development Div., Elpida Memory, Inc., Sagamihara, Kanagawa 229-1197, Japan
* Hitachi ULSI Systems Corp., Kokubunji, Tokyo 185-0014 Japan

Abstract

A low impedance hierarchical I/O architecture designed to realize both high-speed and low-voltage DRAMs is presented. In this architecture, use of the divided I/O lines over the memory cells reduces the load of I/O lines by 50% and enables a 2.2 ns reduction of the read/write cycle time. By combining distributed data transfer scheme, we achieved a 4 ns reduction of the access time to 8 ns and 1-Gb/s/pin operation with a 1.8-V power supply in a multi-Gb DRAM.

Introduction

In high-density DRAMs, a hierarchical I/O architecture [1-3] is widely used because of its small chip-size feature. In these conventional architectures, the I/O lines run over the sense amplifier (SA) and sub-word driver (SWD) circuits, and sub-amplifiers are placed in the intersection area so that they have no affect on chip size. However, they are not sufficiently efficient for simultaneous high-speed and low-voltage operations (Fig. 1), because the intersection area decreases with the scaling down of the process technology and this limits the flexibility of the sub-amplifier circuits. In addition, the parasitic capacitance and resistance of the I/O lines over the SA/SWD circuits are increased at smaller feature sizes. The simulated read/write cycle time and access time in the conventional architecture is 7.2 ns and 12 ns, respectively, thus, the data transfer is restricted by the array cycle time and is 556 Mb/s/pin. This paper presents a novel hierarchical I/O architecture that enables high-speed operation with a low-voltage power supply.

Low Impedance Hierarchical I/O Architecture

Figure 2 shows the chip configuration of a 1-Gb 4-bank x16-DQs DDR SDRAM using the low impedance hierarchical I/O architecture. In this architecture, the sub-amplifier circuits are located in the middle of a quarter-bank 64-Mb array, and 16-bit read/write operations are carried out via the 32 main-I/O (MIO) pairs in the quarter-bank. The local-I/O (LIO) lines and MIO lines are routed over the memory cells between the main-word lines (MWL) in place of some meshed power lines. In the conventional architecture, the scaling down of the process technology causes a decrease of the SA/SWD height and tightening of the wiring pitch over the SA/SWD circuits (L/S $\leqq$ 0.4/0.4 μm). On the other hand, this architecture can relax the I/O lines pitch by routing them over the memory cells (L/S = 1.2/1.2μm) and thereby reduce the total load of I/O lines by 50%. Furthermore, coupling noise between adjacent LIO and MIO lines is eliminated by using MWLs for shield lines without incurring an area penalty.

The schematic of the sub-amplifier circuit in this proposed architecture is shown in Fig. 3. In the conventional architecture, the positive feedback NMOS amplifier [3] is used because of the limited intersection area. The delay time of this amplifier is too large because an excessive timing margin is needed to compensate for PVT fluctuations. In this architecture, the non-feedback π-type read amplifier and write buffer circuits are used as a sub-amplifier to enhance the data-path speed in both read and write operations. Figure 4 shows the dependence of the simulated read/write cycle time and the estimated area penalty of the divided MIO/LIO length in this architecture. We designed the MIO length and LIO length to be 8 Mb and 2 Mb, respectively, and the read/write cycle time was 4 ns, which is 3.2 ns faster than with conventional architecture. Using this structure, the area penalty is 2.4% more than that of conventional architecture. Even in the case of a divided 4-Mb LIO length, the area penalty is 1.1% and that can achieve a 5-ns cycle time and 800-Mb/s/pin operation with a 1.8-V power supply.

Distributed Data Transfer Scheme

The block diagram of the distributed data transfer scheme that is used to achieve high-speed data transfer of global-I/O (GIO) lines in a 4-bit prefetch operation is shown in Fig. 5 and Fig. 6. The main-amplifier circuits[4] that reduce the excessive operating margin caused by PVT fluctuations detect the 4-bit prefetch data of the MIO lines simultaneously and their outputs are transferred to the GIO lines. The GIO lines are quite long, up to 12,000 μm, and the amount of simultaneous data transfer in a 4-bit prefetch operation is four times that of conventional SDRAMs, therefore, wide pitch and fully shielded GIO lines are needed. In this proposed scheme, half the GIO lines (F-GIO) are wide pitch and the others (S-GIO) are narrow pitch; they are routed alternately. First, the 1^{st} and 2^{nd} data of a 4-bit prefetch are transferred immediately to the F-GIO lines by the $\phi 1$ signal and prefetch address. Next, the 3^{rd} and 4^{th} data are transferred to the S-GIO lines by the $\phi 2$ signal with a suitable delay. Hence, the S-GIO lines can be used as shield lines during the data transfer of the F-GIO lines and the peak current of GIO output buffer can be reduced to half. In this scheme, the data transfer time of F-GIO lines was 0.7 ns which is 1.2 ns faster than that of the conventional data transfer scheme. The access time of the 3^{rd} and 4^{th} data has a 2.0-ns extra margin in 1-Gb/s/pin operation, thus, there is no access penalty for the 3^{rd} and 4^{th} data in a distributed data transfer operation. In a write operation, the 1^{st} and 2^{nd} input data use the S-GIO lines and the 3^{rd} and 4^{th} data use the F-GIO lines. That is, the F-GIO and S-GIO lines are used as a common data bus for read/write operations in this proposed scheme.

Simulated Results

The simulated waveforms of the write-to-read operation are shown in Fig. 7 and Fig. 8. To achieve an 1-Gb/s/pin operation with column latency of four clocks, a read/write cycle time of less than 4 ns and an access time of less than 8 ns are needed. This proposed architecture achieves a 4-ns cycle time and 8-ns access time simultaneously with a 1.7-V power supply. The data transfer of write-to read cycles seems to operate with a sufficient timing margin although each command is applied 4-ns cycles.

Conclusion

We have proposed a low impedance hierarchical I/O architecture for high-speed and low-voltage operations. This is achieved by using the divided hierarchical I/O lines over the memory cells and the flexibly located sub-amplifier circuits with a 2.4% area penalty. Furthermore, distributed data transfer scheme that reduces the delay time of a 4-bit prefetch operation is used in this architecture. The circuit simulation of a 1-Gb DRAM, assuming 0.13-μm technology, demonstrated that this device can operate at 1-Gb/s/pin with a column latency of four clocks.

Acknowledgments

We are grateful to K. Miyazawa, K. Kajigaya, M. Horiguchi, K. Kimura, T. Sakata, and T. Adachi for their encouragement.

References

[1] M. Nakamura, *et al*., “A 29-ns 64-Mb DRAM with Hierarchical Array Architecture,” *IEEE JSSC*, vol. 31, pp. 1302-1307, Sep. 1996.

[2] C. Kim *et al*., “A 2.5-V, 72-Mb, 2.0-GB/s Packet-Based DRAM with a 1.0-Gbps/pin Interface,” *IEEE JSSC*, vol. 34, pp.645-646, 1999.

[3] T. Sakata, *et al*., “A DDR/SDR-Compatible SDRAM Design with a Three-Size Flexible Redundancy,” in 2000 *VLSI Symp.*, pp. 116-119.

[4] H. Fujisawa *et al*., “A Dual-Phase-Controlled Dynamic Latched Amplifier for High-Speed and Low-Power DRAMs,” *IEEE JSSC*, vol. 36, pp.1120-1126, July 2001.

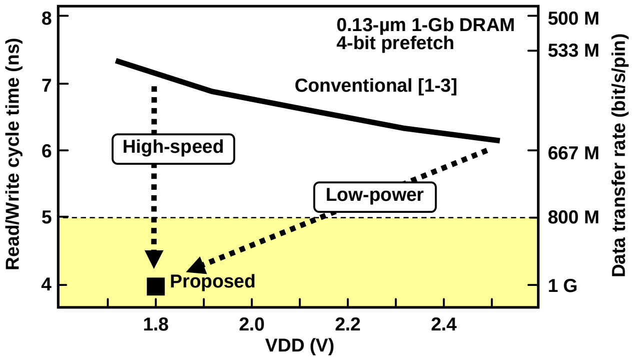


Fig. 1 Operating frequency of hierarchical I/O architecture.

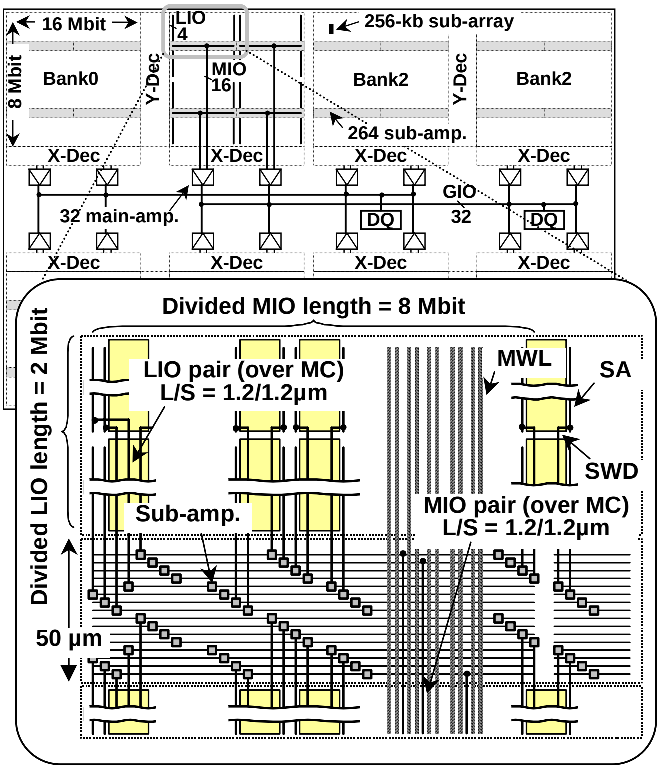
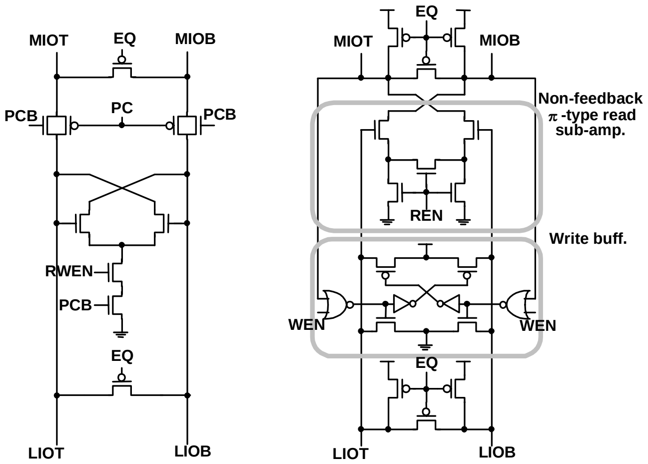


Fig. 2 Low impedance hierarchical I/O architecture of 1-Gb DRAM.



(a) Positive feedback NMOS amplifier [3]

(b) Proposed

Fig. 3 Schematic of sub-amplifier circuit.

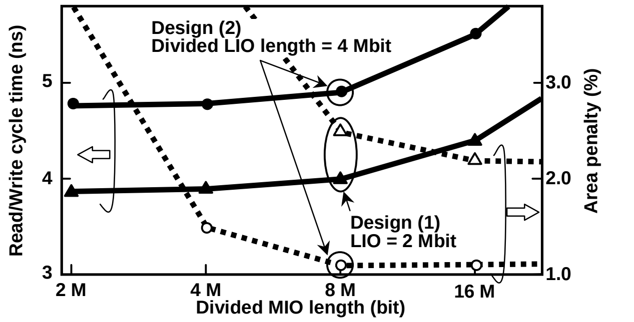


Fig. 4 Simulated operating frequency for proposed architecture.

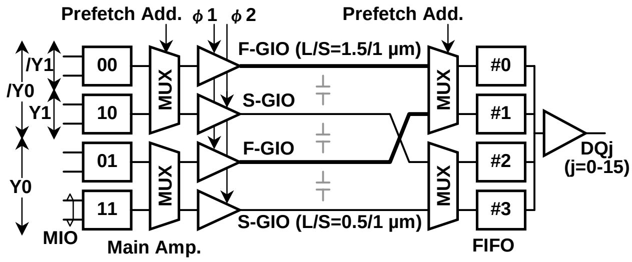


Fig. 5 Block diagram of distributed data transfer scheme.

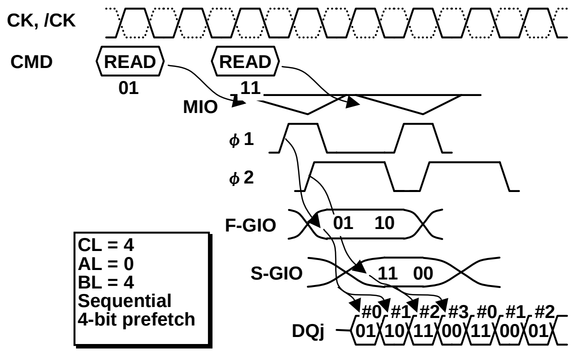


Fig. 6 Timing diagram of distributed data transfer scheme.

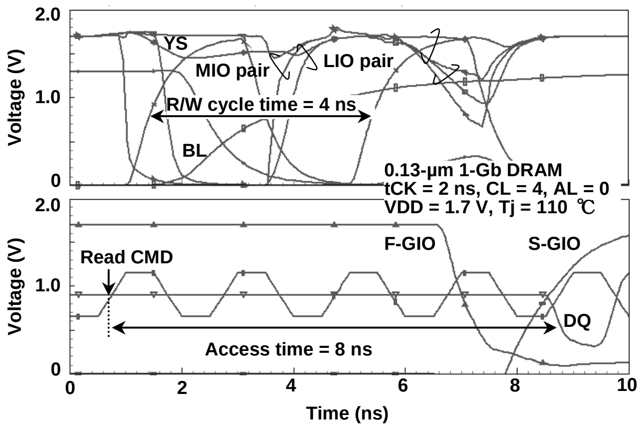


Fig. 7 Simulated waveforms in Write to Read cycles.

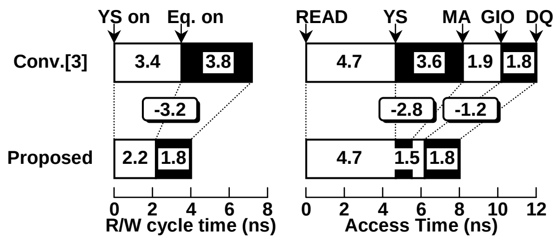


Fig. 8 Improved performance in the proposed architecture.