Post-Packaging Auto Repair Techniques for Fast Row Cycle Embedded DRAM

Osamu Wada, Toshimasa Namekawa, Hiroshi Ito, Atsushi Nakayama, and Shuso Fujii

Semiconductor Company, Toshiba Corporation, Kawasaki, Japan

Abstract

A post-packaging auto repair technique is implemented in a 36Mb embedded DRAM macro of 6ns cycle time. It consists of internal compare circuit, redundancy analyzer, and anti-fuses. The internal auto programming of anti-fuse fixes post-packaging failures.

1. Introduction

Embedded DRAM (eDRAM) macro is advantageous in terms of higher bandwidth and lower power consumption compared with commodity DRAM and is suitable for portable systems, network systems and high performance cache memories. Recently, it has been widely adopted in system-on-a-chip (SoC) devices.

In particular, eDRAM macros with fast row cycle operation [1-4] are attracting increasing attention. High-end SoC chips require high-speed performance comparable to SRAM as well as much larger memory capacity than SRAM. The combination of large capacity and high speed operation results in the more frequent occurrence of fail bits in macros, and thus tends to result in a lower test yield, especially in post-package final test.

At-speed test technique using Built-In-Self-Test (BIST) [5, 6] is an efficient method of testing a chip or a memory at the nominal operating frequency. In wafer level tests such as pre-fuse sorting, BIST can be effectively used to screen out fail bits. They may be replaced with redundant bits by laser fuse blow afterward. However another fail bit might appear after packaging due to thermal or electrical stresses during burn-in. Some repair technique after packaging is greatly desired in order to achieve high product yield and assure quality.

Several post-packaging repair techniques using electrical anti-fuse circuitry have been reported: DLL skew calibration [7], DRAM bit-repair scheme with a static latch [8], for instance. In this paper we propose a novel approach to auto repair techniques, including an internal compare circuit, a fail address register and a redundancy analyzer as well as anti-fuses. The techniques are implemented in a high-speed 36Mb eDRAM macro of 6ns cycle time.

Firstly we explain the background in detail: the reason that faster DRAM operation tends to cause fail bits and the necessity of a post-package repair technique. Second the overview of an eDRAM macro with fast (6ns) row cycle operation is described. It is fabricated with a 90nm eDRAM process technology. Thirdly the implementation of the auto repair techniques using anti-fuses in the eDRAM macro is described, and finally an effective testing flow utilizing the auto repair techniques are proposed.

2. Background

In DRAM testing, data retention time is always the biggest issue. Weak cells with short retention time or with large leakage current must be screened out completely. The issue becomes more serious with shorter row cycle time.

Figure 1 shows simulated waveforms of write operation, which is usually the slowest one among DRAM operations and limits the row cycle time tRC. In a conventional scheme of long cycle time, there is enough time to fully restore the memory cell node to the BL high level after sensing. Weak cells can be easily detected in low frequency wafer-level test at high temperature.

At a shorter tRC, the memory cell node might no longer be fully restored at worse conditions, even if write operation is started before sensing. The write-before-sense scheme [2] is applied to our 36Mb eDRAM macro to achieve 6ns tRC. The signal loss at high level occurs on every cell and affects the retention time.

Figure 2 shows the measured result of the retention time characteristics of a previous-generation 130nm eDRAM macro. When tRC became short, the whole distribution curve shifted to the left (shorter retention time), indicating the signal loss, and also one or a couple of new fail bits might appear. The reason for the occurrence of new fail bit is not well understood, but likely due to parasitic resistance in the cells. Fixing such fail bits is the key to achieving the high final test yield of fast eDRAM macro.

We propose an auto repair technique, which incorporates not just BIST and anti-fuse but also an internal comparator, a fail address register and a redundancy analyzer. The technique does not require any external fail bit memory or analyzer, and is primarily intended for post-package tests with conventional testers.

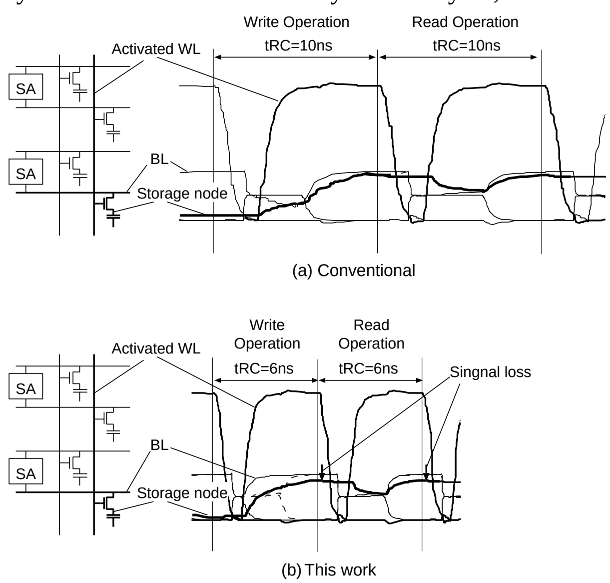

Figure 1: Simulated waveform of memory access operation.

(a) Conventional, (b) This work

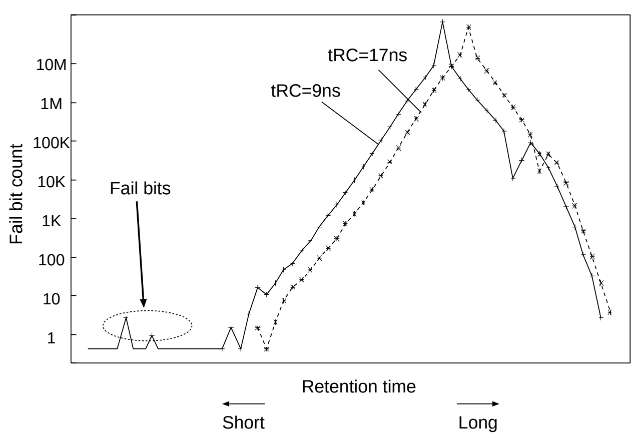

Figure 2: Impact of short tRC on retention time distribution.

3. Overview of eDRAM macro

Figure 3 (a) shows the block diagram of eDRAM macro including the auto repair circuits.

The eDRAM macro consists of 36Mb memory array block, compare circuit, test control circuit and fuse box, and four internal voltage regulators. The 36Mb memory array block consists of four 9Mb sub-arrays including main row decoders and WL drivers, column decoders and secondary sense amplifiers. Each 9Mb sub-array is divided into 512kb cell array mats (AMAT) with local row decoders. The 288bit local data lines (LDI/O) running across the 9Mb sub-macros are multiplexed, and connect to 144bit main data lines (MDI/O). When the macro is tested by BIST or external tester, the 144bit data width is further reduced to 18bit.

The array structure of AMAT is shown in Figure 3(b). AMAT is configured with 512 wordlines (WL) and 8 spare WLs, by 1152 bitline (BL) pairs and 16 spare BL pairs. Both sides of AMAT are connected to 576 sense amplifiers (S/A) and 16 spare S/As. Each S/A is shared between two adjacent AMATs.

The row and column redundancy architecture of the eDRAM macro is shown in Figure 4. Row redundancy unit (RRU) and column redundancy unit (CRU) are the unit regions of row and column replacements respectively. Each RRU contains eight spare rows as shown in Figure 4(a). Any one of 8 spare rows may be used for replacement by programming optical fuses, and

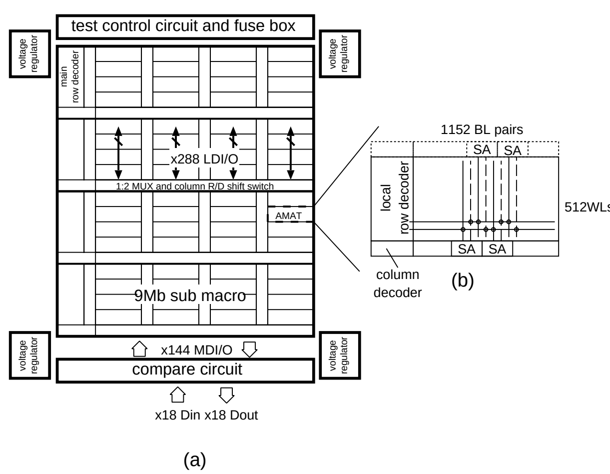

Figure 3: (a) Overview of embedded DRAM macro,

(b) unit memory cell array (AMAT)

the 8th spare row, if unused, may be activated for one more replacement by programming anti-fuses. Each CRU contains two spare columns and so-called shift redundancy scheme is applied as shown in Figure 4(b). One spare column, if unused, is available for anti-fuse replacement.

4. Implementation of auto repair techniques

The block diagram of an auto repair system is shown in Figure 5. It consists of an internal compare circuit, a test control circuit and a fuse box. The fuse box contains optical fuse elements for conventional redundancy and also anti-fuse elements for additional post-packaging repairs. The anti-fuse element is realized by using deep trench (DT) cell, which is identical to DRAM memory cell, and thus no additional process is needed.

The compare circuit is composed of data pattern registers (DPR), fail dataline registers (FDR), fail address register (FAR), and arrays of XOR gates. Before starting at-speed or short tRC test, DPR must be filled with a given data pattern, such as row stripe, column stripe or checkerboard pattern. The true/complement signal is supplied from tester to generate proper expected values from the data pattern depending on address. Read data are compared with expected values, and the results of comparisons are stored in the FDR. Meanwhile the fail address (but only the FIRST one) is stored in FAR.

The timing diagram of the compare circuits is shown in Figure 6. Suppose that the data patterns of "0101…" and "1010…" are written in address "a1" and "a2" respectively, and the expected values of "1010…" were stored in DPR in advance. When the address "a1" is accessed, the read data of "0101…" come out in the next cycle. Here the true/complement signal must be brought high to correctly generate the expected values by inverting the values stored in DPR. If the read data match the expected values, the comparison result remains low.

On the other hand, when the address "a2" is accessed, suppose at this time that a wrong data of "1110…" is read owing to some fail bit. The read data mismatch the expected values that are the same as those stored in DPR this case. The comparison result turns to high and the address "a2" is latched in FAR.

The values in FDR and FAR are serially transferred to the redundancy analyzer after finishing the at-speed test. From the fail address the redundancy analyzer calculates which anti-fuses should be programmed. It also checks the status if the spare rows and columns are still available. If they are not available, the chip is not fixable and the "analyzer result" output goes to high.

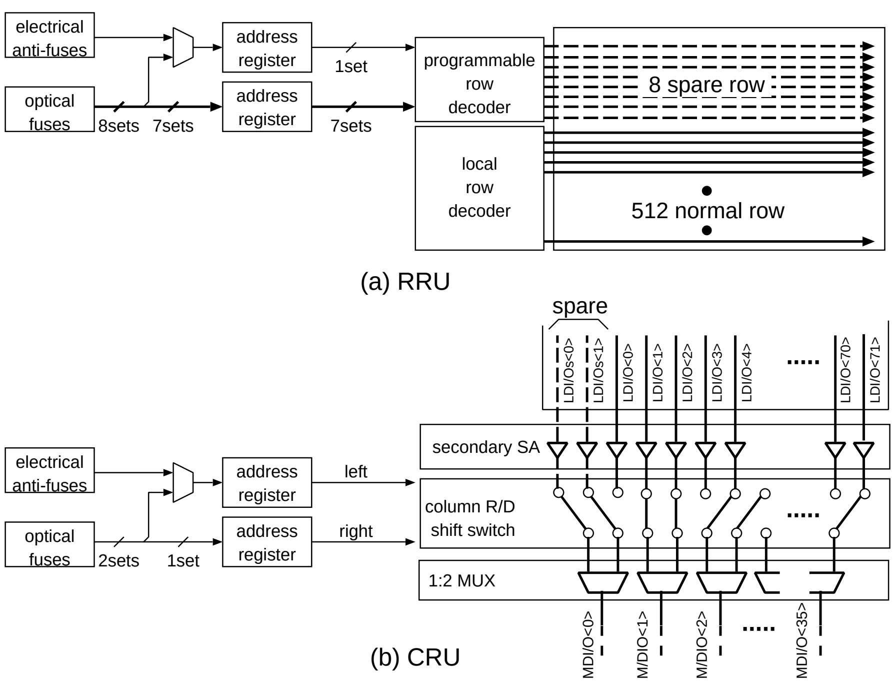

Figure4: Redundancy structure
(a) Row redundancy unit (RRU), (b) Column Redundancy unit (CRU)

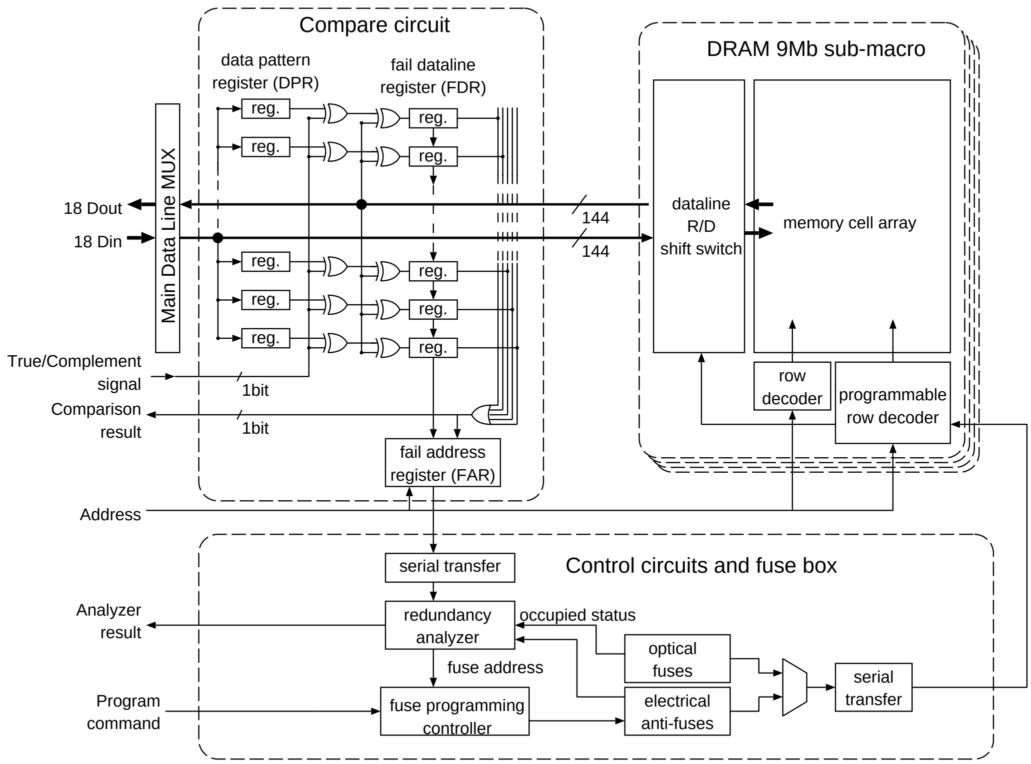

Figure 5: Block diagram of auto repair system

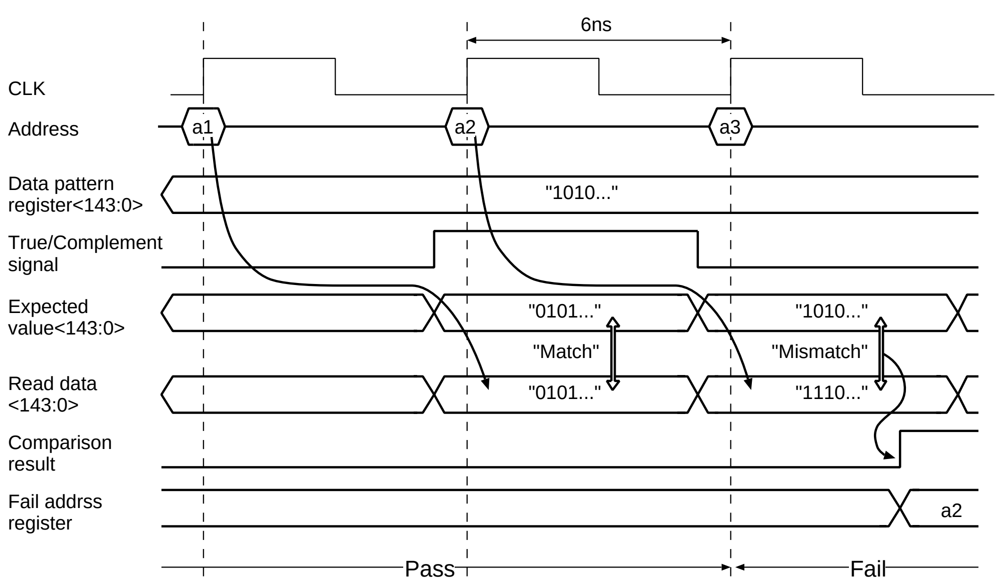

Figure 6: Timing diagram of compare operation.

If the "analyzer result" remains low (fixable), the "program command" from the tester initiates the anti-fuse programming.

A very simplified algorithm is applied to the redundancy analyzer. The FAR stores only the first fail address. Row spare, if available in that redundancy region, is primarily used to fix a failure. If not, column spare is used. With this simplification, the area overhead of the auto repair circuit is minimal.

5. Test flow using auto repair technique

Figure 7 shows two examples of test flows for eDRAM macro. Figure 7(a) shows a conventional test flow and figure 7(b) shows a proposed test flow utilizing the auto repair technique. Although the only difference between them is the position of at-speed test, the auto repair technique improves the yield of speed sorting; besides it pushes up the speed specification.

In the case of conventional test flow, since all of the fail bits including slow cells must be detected in the wafer-level test, BIST and DLL techniques are required in order to avoid the speed limitation of wafer probing. The detected fail bits should be replaced with spare cells by blowing optical fuses, and then, package and burn-in follow. Although some unexpected fail bits might occur during burn-in, they can't be fixed within this test flow because the following final at-speed test has only a function to detect the slow cells. The chips that have such a slow cell go to waste. To avoid these losses, it is necessary to keep a large margin between the speed specification and cut-point during the wafer test. Therefore, the conventional test flow degrades the yield of speed sorting and the speed specification itself.

In the case of the proposed test flow, post-package at-speed test utilizing the auto repair technique is adopted instead of at-speed test at the wafer level. The slow bits whose speed might be degraded within the burn-in can be replaced even at the post-package. When a failure is detected after running a final at-speed test program, the external tester just needs to let the internal circuit of the device under test (DUT) calculate the repair address of the first fail bit, and to initiate a command breaking anti-fuse. This DUT completes the fixing of the first failure or does nothing when no failure is reported, and

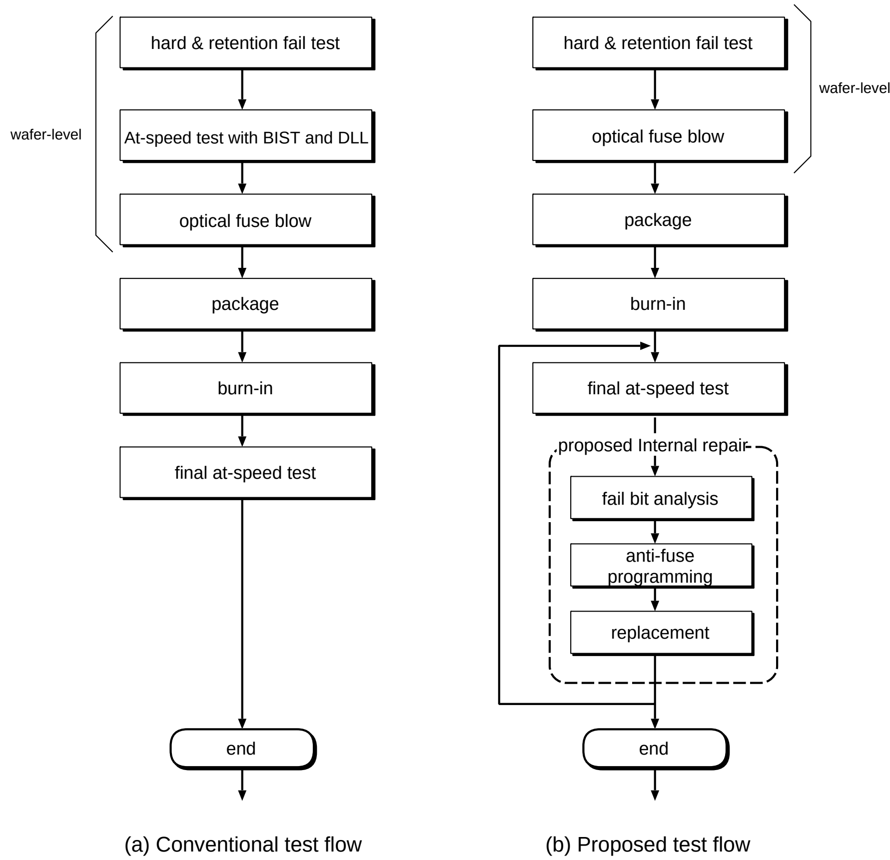

Figure 7: Examples of test flow for embedded DRAM macro.

then is re-powered to reset the redundancy registers. The same at-speed test is iterated to check if another failure remains. This sequence may be repeated until a good product is obtained, or may be terminated after just one or two iterations to save the test time.

The proposed test flow has some advantages. First, since no fail address analyzer is required on the at-speed tester, the cost of equipment can be suppressed. Secondly, multiple DUT can be allowed because the at-speed tester just runs the test program and initiates a command to break anti-fuses regardless of whether these DUT have slow cells or not. This function shortens the test time because of multiple DUT. Besides, it stabilizes the at-speed test because it does not require catching any hi-speed signals. Finally, much less speed-margin is required because the slow cells can be replaced. That means the cut-point on the wafer test can be relaxed. Furthermore the at-speed test on the wafer may not be necessary assuming that few slow cells appear, or the speed specification can be pushed up.

The proposed auto repair circuit has the limitation of fixing just one failure at a time. This limitation may not be serious because the number of fail bits tends to be small, usually just one or a few at most, and the iterative procedure described above should still be efficient.

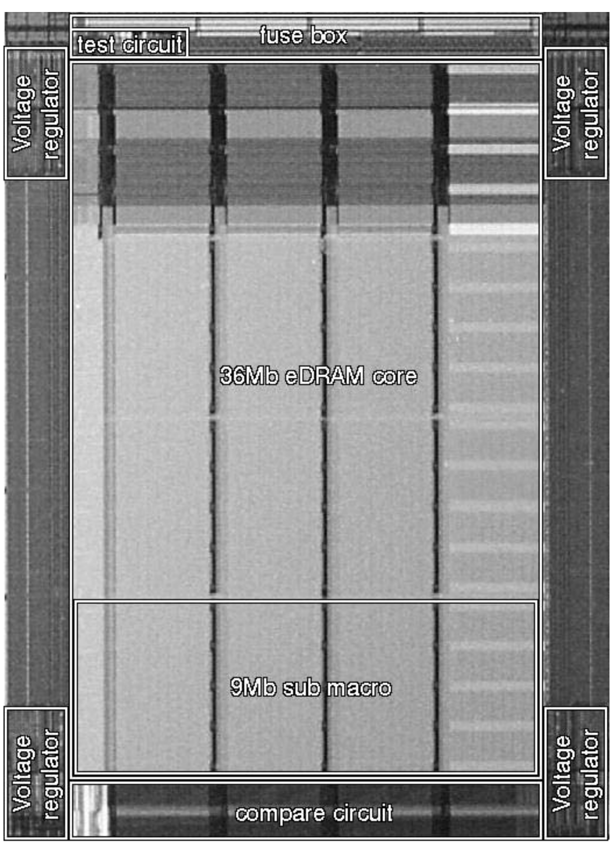

Figure 8: A microphotograph of embedded DRAM macro.

6. Measured results

Figure 8 shows a microphotograph of the eDRAM macro test chip with the proposed auto repair circuit. The features of the eDRAM macro are shown in Table 1. This macro is fabricated in a 90nm eDRAM technology. The area of test circuit and compare circuit is 0.237mm^2. The area overhead of test circuit and compare circuit is almost 1% for a 36Mb eDRAM macro. The macro operates with 2.5V power supply. The internal voltage of 3.3V (VPP) mainly supplies wordline voltage, while that of 1.2V supplies the voltage for peripheral circuitry and sense amplifiers. Figure 9 shows a shmoo plot of random cycle time versus internal voltage of VPP. It shows 6ns of random cycle operation can be achieved under the typical voltage condition and 85 degrees. According to this shmoo plot, we can decide VPP level of a wafer test condition. To detect weak cells, VPP should be 2.7V or less because the clock frequency of wafer test is up to 50MHz.

Figure 10 shows the measured result of the retention time characteristics of the test chip where some weak cells have already screened out by a wafer test as mentioned above. Figure 10 shows two curves. The first curve shows the retention time characteristics for a slow

Table 1: eDRAM macro features.

Technology	90nm embedded DRAM process, 5-layer
Memory Cell size	0.75um x 0.26um = 0.195um^2
Macro size	4.12mm x 5.58mm = 22.99mm^2
Organization	x144 @36Mb
Supply voltage	2.5V (3.3V, 1.2V internal)

```
     tRC(nS) 4.0       8.0      12.0      16.0
              +----+----+----+----+----+----+----+
VPP: 3.500V   *****...............................
VPP: 3.400V   *****...............................
VPP: 3.300V   *****...............................
VPP: 3.200V   *****R..............................
VPP: 3.100V   *****RR.............................
VPP: 3.000V   ******RRRR..........................
VPP: 2.900V   *******RRRRRR.......................
VPP: 2.800V   **********RRRRRRRRRR................
VPP: 2.700V   **************RRRRRRRRRRRRRRRRR.....
VPP: 2.600V   **********************RRRRRRRRRRRRRR
VPP: 2.500V   ********************************RRRR
        R: Repairable bits by blowing optical fuses
```

Figure 9: A shmoo plot of tRC versus internal voltage of VPP.

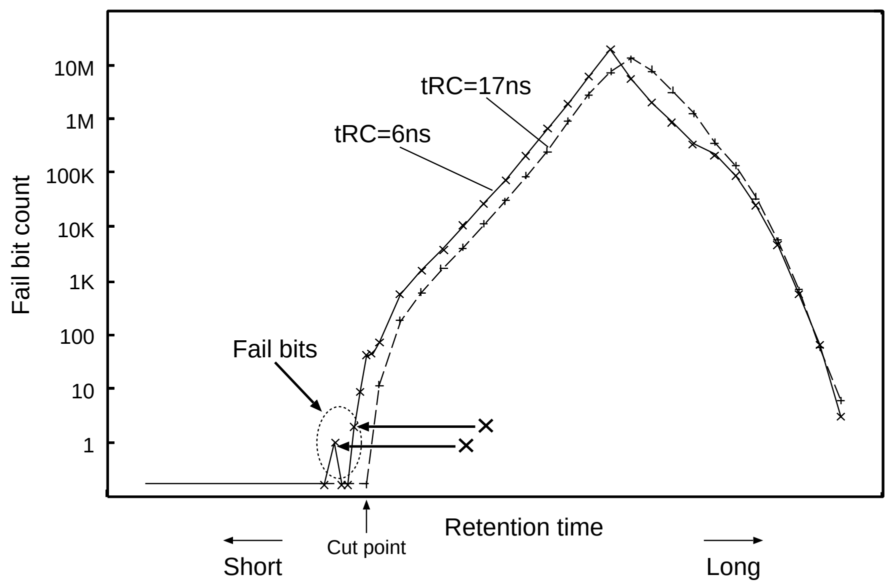

Figure 10: The measured result of the retention time characteristics of the test chip.

random cycle, 17ns of tRC; in the meantime some weak cells have been screened out at the cut-point of retention time. As tRC becomes shorter, the whole distribution curve shifts to the left side. Moreover, other new fail bits also appear in this test chip. The retention times of these fail bits have not been bat in the case of the slower random cycle test. This measured result suggests that the proposed test flow is efficient in ensuring the fast random cycle operation of eDRAM.

7. Conclusions

We have proposed a post-packaging auto repair system for eDRAM macro. It consists of an internal compare circuit, a redundancy analyzer and anti-fuses. In the post-packaging at-speed test, if failures exist, the internal compare circuit can store only the first fail address. The redundancy analyzer calculates which anti-fuses should be programmed and the "program command" from the tester initiates the anti-fuse programming to fix a failure. We have also proposed a test flow using on auto repair technique. It assumes a conventional wafer testing, but puts much weight on the post-package test by utilizing the proposed auto repair technique.

The proposed test flow is advantageous in terms of the capability to fix post-package failure and wide test coverage in screening out fail bits.

The test chip of a 36Mb eDRAM macro has been fabricated using a 90nm eDRAM technology. The area overhead of test circuit and compare circuit is almost 1% for a 36Mb eDRAM macro. The proposed auto repair technique can fix failures that might appear by final at-speed test and contributes to yield improvement.

Acknowledgments

We would like to thank N.Yanagiya and K.Hashimoto for their support of this work.

References

[1] Harold Pilo, et al., "A 5.6ns Random Cycle 144-Mb DRAM With 1.4Gb/s/pin and DDR3-SRAM Interface," *IEEE Journal of Solid-State Circuits*, Vol.38, No.11, pp.1974-1980, Nov.2003.

[2] Yasihiro Agata, et al., "An 8-ns Random Cycle Embedded RAM Macro With Dual-Port Interleaved DRAM Architecture(D2RAM)," *IEEE Journal of Solid-State Circuits*, Vol.35, No.11, pp.1668-1672, Nov.2000.

[3] Yasuhiko Taito, et al., "An Embedded DRAM With a 143-MHz SRAM Interface Using a Sense-Synchronized Read/Write," *IEEE Journal of Solid-State Circuits*, Vol.38, No.11, pp.1967-1973, Nov.2003.

[4] Tohru Kimura, et al., "64Mb 6.8ns Random ROW Access DRAM Macro for ASICs," *ISSCC Digest of Technical Papers*, pp.416-417, Feb.1999.
[5] Hideki Sakakibara, et al., "A 750MHz 144Mb Cache DRAM LSI with Speed Scalable Design and Programmable at-Speed Function-Array BIST," *ISSCC Digest of Technical Papers*, pp.458-459, Feb.2003.
[6] Peter Jakobsen, et al., "Embedded DRAM Built In Self Test and Methodology for Test Insertion," *Proc. of International Test Conference*, pp.975-984, 2001.
[7] Se Jun Kim, et al., "A Low-Jitter Wide-Range Skew-Calibrated Dual-Loop DLL Using Antifuse Circuitry for High-Speed DRAM," *IEEE Journal of Solid-State Circuits*, Vol.37, No.6, pp.726-734, Jun.2002.
[8] Jae-Kyung Wee, et al., "A Post-Package Bit-Repair Scheme Using Static Latches With Bipolar-Voltage Programmable Antifuse Circuit for High-Density DRAMs," *IEEE Journal of Solid-State Circuits*, Vol.37, No.2, pp.251-254, Feb.2002.