21.6 A 100nm Double-Stacked 500MHz 72Mb Separate-I/O Synchronous SRAM with Automatic Cell-Bias Scheme and Adaptive Block Redundancy

Kyomin Sohn, Young-Ho Suh, Young-Jae Son, Dae-Sik Yim, Kang-Young Kim, Dae-Gi Bae, Ted Kang, Hoon Lim, Soon-Moon Jung, Hyun-Geun Byun, Young-Hyun Jun, Kinam Kim

Samsung Electronics, Hwasung, Korea

As multi-core processors become mainstream, the demand for high-density cache memories has increased. Conventional 6T-cell-based SRAMs do not provide enough density for this trend, although they do have the desirable feature of high-speed access. To overcome the density limitation, an SRAM using a double-stacked S^3 (stacked single-crystal Si) SRAM cell was introduced for mobile applications [1]. This work demonstrates a high-speed SRAM using double-stacked-cell. From the process point of view, our design uses fully proven technologies for mass production at the sacrifice of cell size, from $25F^2$ to $36F^2$ [2]. In addition, peripheral only CoSix layers and a tungsten shunt wordline scheme are used for high-speed operation. From a circuit-design perspective, three schemes are introduced. They are automatic cell bias (ACB) for managing the current of SRAM cell transistors by controlling cell bias, adaptive block redundancy (ABR) for dealing with various defects from the new cell technology, and wordline pulse-width regulation (WPR) for adjusting wordline pulse-width according to cycle time.

There is previous research on controlling cell bias adaptively for high performance and stability [3]. Based on that, our on-chip cell-biasing scheme, including detector, compensates for PVT variations with aging effect. Since transistors M1 to M4 of Fig. 21.6.1 are SOI transistors grown by epitaxial growth, it is difficult to control their characteristics compared with a bulk-based pull-down NMOS. Automatic cell-bias (ACB) scheme controls voltages to regulate the performance of those transistors. Figure 21.6.2 shows the concept of current measurement of each transistor and the scheme of ACB for controlling the current of load PMOS transistor. The method of measuring the current of the NMOS pass transistor is similar but the current is measured from the source of pre-charge transistors. ACB controls the voltage level of cell power (Vddc) and wordline enable (Vpp) automatically at initial settling time or by JTAG inputs. The measured current of each cell is compared with LSL (lower specification limit) and USL (upper specification limit), which are modeled by R_{USL} and R_{LSL}. In the case of ACB for load PMOS, the number of cells over USL or under LSL is counted and Vddc is adjusted to minimize the number of cells out of specification. It has 8 steps and each step is about 50mV. The graph of Fig. 21.6.2 shows the measured current distribution of before and after ACB. In this case, power consumption decreases by lowering Vddc and cell-on current increases by raising Vpp. With this procedure, we set the cell on current for high-speed operation and while limiting power consumption within the specification, increasing chip yield.

To further improve yield, an efficient redundancy scheme is critical in memory device. Figure 21.6.3 shows defect analysis of double-stacked SRAM and a repair policy. Unlike planar-type SRAM, the ratio of lump-type defects is relatively high, as much as 30%, because the cell has a complex vertical structure. Lump-type defects cannot be effectively bypassed by the conventional redundancy scheme. We design an adaptive block redundancy (ABR) to deal with lump-type defect and to also improve repair efficiency and flexibility. The conventional redundancy scheme is restricted in flexibility because of the shape of redundant cell array, as shown in Fig. 21.6.3. ABR adds a redundant block to a group of blocks that share same I/Os and uses that block as a redundant cell array for replacing both row- and column-type defect cells. Figure 21.6.4 shows the composition of a redundant block of ABR. In the case of row repair, replaced normal cells are allocated to cell columns of the redundant cell block by switching the row address to a column and block address in MRD_RED of Fig. 21.6.4. In other words, a row repair is treated as a column repair. The advantages of ABR are as follows. First, the ratio of redundant cells for row and column is easily adjusted. Second, the shape a group of cells replaced by a single repair is easily modified by addressing mapping in the redundant cell block. We add two row addresses to MRD_RED so that a single row repair replaces 16 rows across 2 blocks. Finally, this scheme is easy to implement because a normal cell block is used for redundant cell array with the partial change of address inputs. The ABR scheme repairs 30% more defects than the conventional redundancy scheme.

ACB makes the cell on current distribution of a chip close to the expected value, but the speed of a SRAM is determined by the slowest cell in the chip. For this fine-tuning, we design a wordline pulse-width regulation (WPR) scheme. The pulse-width of the wordline is one of critical factors that determines the shortest cycle time. When the wordline pulse width is fixed at a certain value, the shortest cycle time is also fixed and chips that have slow cells do not operate at all. WPR helps SRAM to operate at as short cycle time as possible. It also increases the yield of SRAM by providing a wide range for speed binning, according to the speed of each chip. The left side of Fig. 21.6.5 shows the waveform and the schematic of wordline pulse-width regulator. In the cycle time of over 3ns, wordline pulse-width maintains 2ns. When the cycle time is shorter than 3ns, pulse-width becomes narrower by K_NXT signal, which is generated by next clock. In WPR, core precharge time is about 0.8ns and the wordline pulse-width is given by subtracting precharge time from cycle time.

It is difficult to know the required core operation time of each chip in a wafer level because of the limit of wafer test equipment. To know the speed of each chip at the wafer level, a short cycle test (SCT) scheme is proposed. Various built-in self-test methods for at-speed test have already been introduced [4]. The features of SCT are simplicity and interaction with test equipment. The right side of Fig. 21.6.5 shows the timing and the simplified block diagram of SCT. At the rising edge of the external clock, a high-frequency clock, K_{SC}, of 4 or 8 pulses is generated by a short cycle generator (SCG). According to K_{SC}, a pattern generator produces read and write control signals, R/W_{SC}, 3 LSBs of addresses, ADD_{SC}, and the write-data-pattern selector, $DinP_{SC}$. $DinP_{SC}[0:2]$ are signals for selecting from data patterns saved in write-data registesr. Test results of each cycle are given by MDQ[0:8] and MP_{EXT}. MDQ[0:8] are merged data output, which means 4 I/Os are merged to 1 MDQ I/O. It shows the last read-data in a compressed way. MP_{EXT} is used to integrate the data output of 4 or 8 cycles. If all data meet the expected data, MP_{EXT} becomes high. By using SCT, wafer test at a short cycle time can be executed and the test result is used for known good die.

Figure 21.6.6 shows measured output signals and shmoo data of cycle time versus supply voltage. It shows the data output waveforms with 1Gb/s/pin at 500MHz and 2ns cycle time at 1.8V supply voltage. Figure 21.6.7 shows the chip microphotograph of 72Mb SRAM in a 100nm double-stacked S^3 SRAM cell technology and its performance summary table.

References:

[1] Y. Suh, H. Nam, S. Kang et al., "A 256Mb Synchronous-Burst DDR SRAM with Hierarchical Bit-Line Architecture for Mobile Applications," *ISSCC Dig. Tech. Papers*, pp. 476-477, Feb. 2005.

[2] S. Jung, H. Lim, C. Yeo et al., "High Speed and Highly Cost Effective 72M bit Density S3 SRAM Technology with Doubly Stacked Si Layers, Peripheral only CoSix layers and Tungsten Shunt W/L Scheme for Standalone and Embedded Memory," *Dig. Symp. VLSI Technology*, pp. 82-83, Jun. 2007.

[3] H. Yu, N. Kim, Y. Son et al., "A SRAM Core Architecture with Adaptive Cell Bias Scheme," *Dig. Symp. VLSI Circuit*, pp. 128-129, Jun. 2006.

[4] H. Sakakibara, M. Nakayama, M. Kusunoki et al., "A 750MHz 144Mb Cache DRAM LSI with Speed Scalable Design and Programmable At-Speed Function-Array BIST," *ISSCC Dig. Tech. Papers*, pp. 458-459, Feb. 2003.

 978-1-4244-2010-0/08/$25.00 ©2008 IEEE

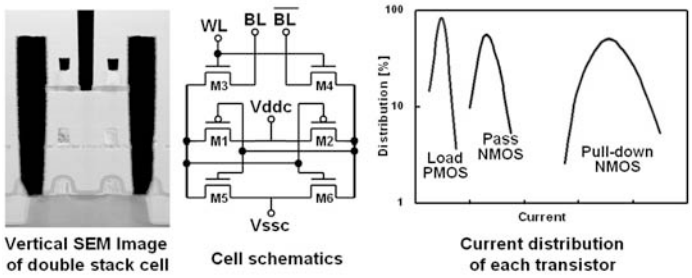


Figure 21.6.1: Double-stacked S3 (stacked single-crystal Si) SRAM cell.

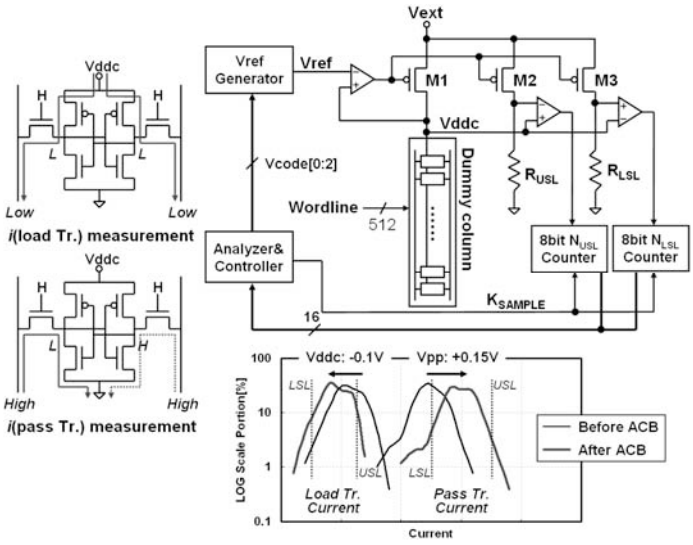


Figure 21.6.2: Automatic cell-bias (ACB) scheme and test results.

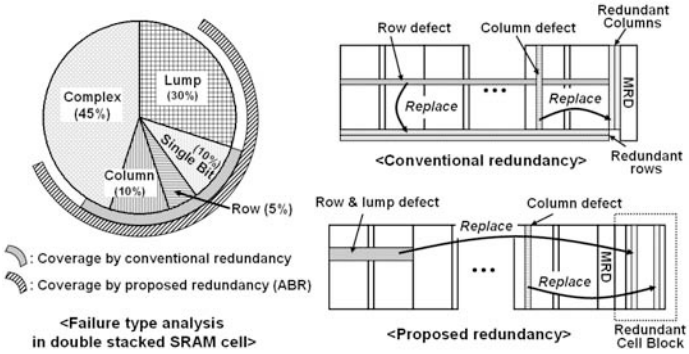


Figure 21.6.3: Defect analysis of double-stacked SRAM and redundancy scheme.

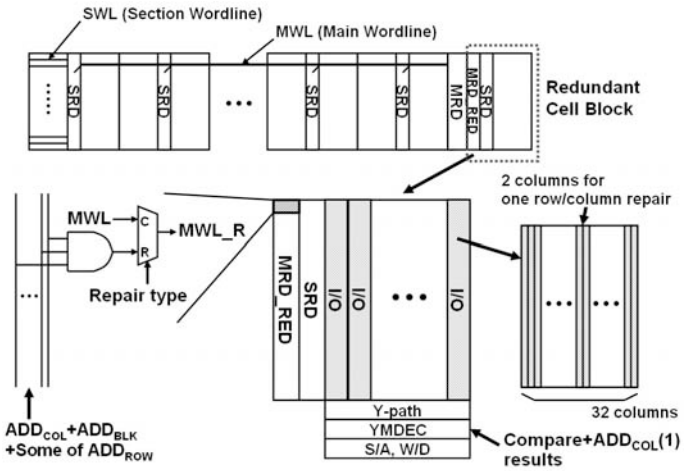


Figure 21.6.4: Adaptive block redundancy (ABR) scheme.

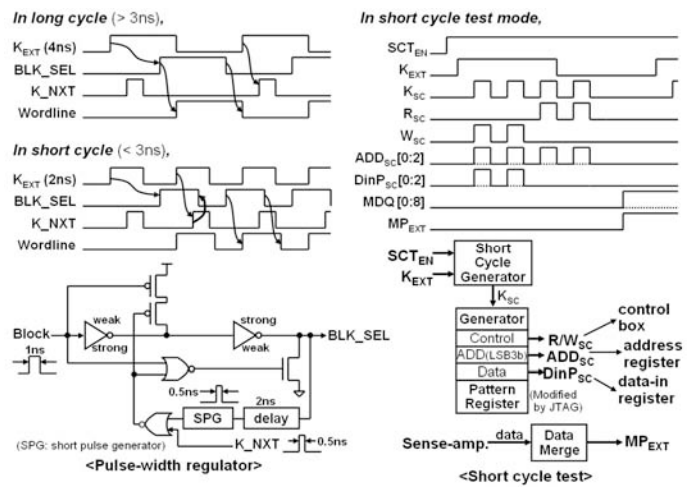


Figure 21.6.5: Wordline pulse-width regulator (WPR) and short cycle test (SCT) scheme.

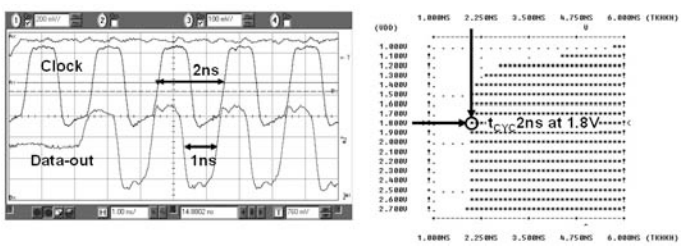
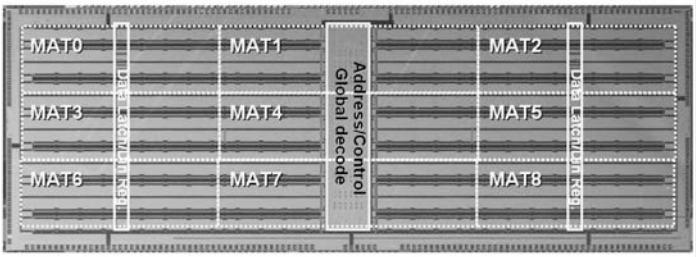


Figure 21.6.6: Measurement results.

Continued on Page 622



Organization	2M x 36, 4M x 18, 8M x 9
Cycle Time	500MHz
Data Rate	72Gb/s (Double Data Rate, Separate I/O)
Latency	5ns @ 1st Data
Power Supply	External 1.8V, Internal 1.2V
Average Power	1.32W Core
Technology	100nm Double Stacked S^3 SRAM Cell
Cell Size	0.36mm^2
Chip Size	14.1mm x 5mm = 70.5mm^2

Figure 21.6.7: Chip micrograph and performance summary.

 978-1-4244-2010-0/08/$25.00 ©2008 IEEE