

11.1 Universal-Vdd 0.65-2.0V 32kB Cache using Voltage-Adapted Timing-Generation Scheme and a Lithographical-Symmetric Cell

Kenichi Osada, Jin-Uk Shin*, Masood Khan*, Yu-de Liou*, Karl Wang*, Kenichi Shoji**, Kenichi Kuroda**, Shuji Ikeda**, and Koichiro Ishibashi
Central Research Laboratory, Hitachi, Ltd., Kokubunji, Tokyo, Japan

*Hitachi Semiconductor America Inc., San Jose, CA, USA

**Semiconductor & Integrated Circuits Group, Hitachi, Ltd., Tokyo, Japan

Low-power-dissipation microprocessors, which vary the frequency and power-supply voltage depending on system loads, have recently been developed for portable electric devices. They require cache technology to ensure a wide operating-voltage range to enhance performance. A universal-Vdd 32kB four-way-set-associative embedded cache test chip uses 0.18 μ m enhanced CMOS technology and continuously operates from 0.65V to 2.0V. Operating frequency and power are from 120MHz at 1.7mW and 0.65V to 1.04GHz at 530mW and 2.0V. Cache performance is attained using voltage-adapted timing generation with plural dummy cells and a lithographical-symmetric memory cell (LS-cell).

As power supply voltage is reduced, a lower threshold voltage becomes preferable for high-speed operation. However, the threshold voltage of memory cells must be kept at a minimum of 0.5V to reduce the leakage current and to maintain static-noise margins. On the other hand, the threshold voltage of peripheral circuits is reduced to 0.4V to attain high-speed operation at low voltage operation. This threshold voltage difference could cause activation failure of a sense amplifier in wide-supply-voltage operation. To avoid such failure, voltage-adapted timing generation uses plural dummy cells. The timing pulse is used for activating the sense amplifier and for resetting word lines.

Figure 11.1.1a shows a block diagram of a half side of the cache quadrant. The quadrant is composed of 256 word lines by 256 bit columns. When the clock signal changes from 'L' to 'H', the signal (dec_en), generated in a D-flip flop, activates predecoders and a word line. Then signal voltages appear on bitline pairs. The signal (dec_en) also activates the dummy word line, which runs parallel to the bitline, and the twelve dummy cells (DCs) on the dummy column drive the dummy bitline, whose capacitance is identical to that of the regular bitline. In these circuits, no extra area is needed for the dummy word line even if plural dummy cells are used. The detailed circuits of the dummy-column cell and edge-column cell are shown in Figure 11.1.1b. The dummy column is used as the electrical dummy column while the edge column is used as the optical dummy. As a result, the layouts of the diffusion layer and poly silicon layer in the SRAM array, the dummy column, and the edge column are kept regular. The dummy-cell current is identical to the memory-cell current in a regular SRAM array. A timing pulse, (voltage-adapted pulse) is used as the sense-amplifier enable signal, (sa_en), the precharge reset signal, (pc_en), and the word-line-reset signal. The word-line-reset is only in the predecoder, so the area for reset is minimized.

Figure 11.1.2 shows timing diagrams of the cache. At 2.0V, bitline-drive time is 24% of the total access time. However, bitline-drive time increases to 48% at 0.65V because of the memory cell with high-Vt MOS. Here, high-Vt dummy cells drive a part of the timing-generation path, so the sense amplifier is activated suitably even at low voltage. Although a dummy cell structure is previously developed, a single dummy cell suffers from cell-current fluctuation. Figure 11.1.3 shows calculated fluctuation of the plural dummy cells at 1.5V. Dummy-bitline-drive-time

fluctuation normalized by total access time is shown as a function of the number of dummy cells. Fluctuations of memory-cell current are calculated using 4% standard deviation, obtained from another test chip. The fluctuation is reduced from 17.5% to 5% when the number of dummy cells increases from 1 to 12. This fluctuation reduction reduces the timing margin for activating the sense amplifier and the delay is reduced by 12.5% compared to the delay of conventional circuits with a single dummy cell [1]. Figure 11.1.2 and Figure 11.1.3 show that the influences of threshold voltage difference and variation of dummy-cell current on circuit performance are reduced.

For low-voltage operation, it is important that an SRAM cell has adequate noise margin. However this margin is difficult because the layouts of generally used SRAM cells are not immune to mask misalignment, so they have suffered from imbalance [2][3]. The lithographical-symmetric memory cell (LS-cell) addresses this problem. Figure 11.1.4 shows SEM micrographs of the lithographical-symmetric memory cells (LS-cell). The upper micrograph shows a diffusion layer and a poly-silicon layer. The lower micrograph shows first and second metals. LS-cell has low aspect. Each layer of the LS-cell is a point of symmetry and the patterns of the poly-silicon layer and diffusion layer are straight with no bends because the well contacts are placed every 64 rows. The LS-cell is not influenced by photolithography misalignment and has good electrical balance. There are other advantages of the LS-cell; its bit lines are shorter because of the low-aspect ratio, so parasitic capacitance of bit lines is reduced; crosstalk between bit lines is drastically reduced because bit lines are shielded by vdd lines and gnd lines; and the vdd lines and gnd lines run on the orthogonal word lines and each memory-cell read current flows on each vdd line and each gnd line. Noise on the vdd lines and gnd lines is reduced. These advantages lead to 13% reduction in total access delay, compared to that of a conventional SRAM cell at 1.5V.

The cache uses 0.18 μ m quadruple-metal enhanced CMOS technology. The pitch from the first to fourth metal is 0.52 μ m. A micrograph of a test chip is shown in Figure 11.1.5. The chip array is composed of four 8kB banks and each bank is composed of one double word. The cache is 3x1.25mm². Well taps are placed every 64 rows in each memory array. This array structure forms a four-way-set-associative 32kB cache. The on-chip PLL is used to obtain the schmo plot in Figure 11.1.6. The test chip operates from 120MHz at 0.65V to 1.04GHz at 2.0V. Power dissipation at 120MHz and 0.65V is 1.7mW and at 1.04GHz and 2.0V is 530mW. Figure 11.1.7 shows specifications of the cache.

Acknowledgements:

The authors thank I. Naka, T. Kitahara, J. Zim-mer, D. Xu, K. Barberg and J. Sun of HSA, A. Hasegawa, M. Aoki and K. Noguchi of SICG, K. Sasaki of CRL, C. Hwang of HAL and D. Henoff and H. Nurser of ST microelectronics for support and discussions.

References:

- [1] Bharadwaj S., et al., "A Replica Technique for Wordline and Sense Control in Low-Power SRAM's," IEEE J. Solid-State Circuits, vol. 33, pp1208-1219, Aug., 1998.
- [2] T. Uetake, et al., "A 1.0ns Access 770MHz 36Kb SRAM Macro," VLSI Circuits Digest of Technical Papers, pp109-110, June, 1999.
- [3] K. J. Kim, et al., "A Novel 6.4 μ m² Full-CMOS SRAM Cell with Aspect Ratio of 0.63 in a High-Performance 0.25 μ m-Generation CMOS Technology," VLSI Technology Digest of Technical Papers, pp68-69, June, 1998.

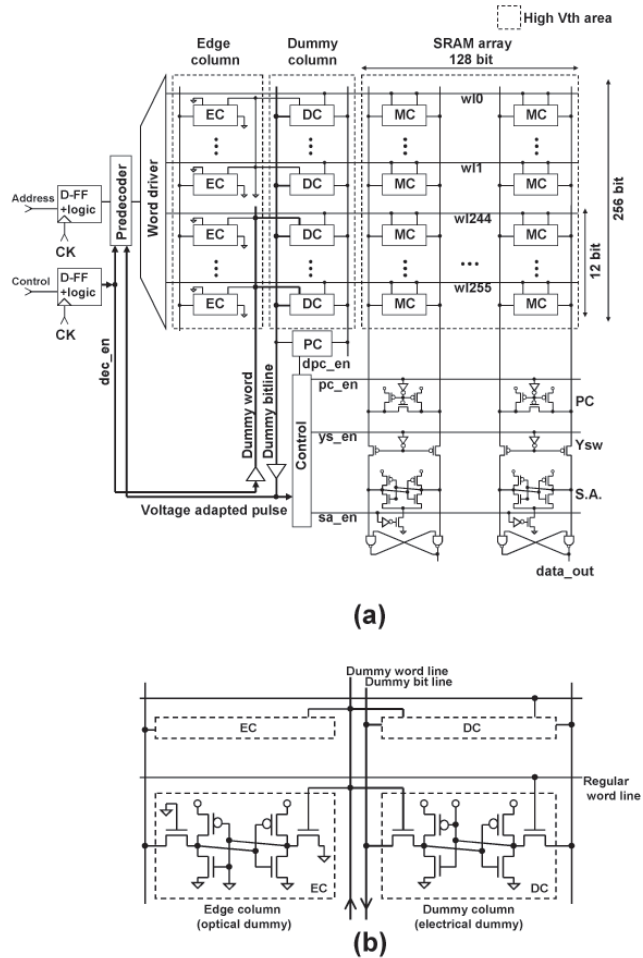


Figure 11.1.1: (a) Block diagram of half side of a bank. (b) Dummy and edge column circuits.

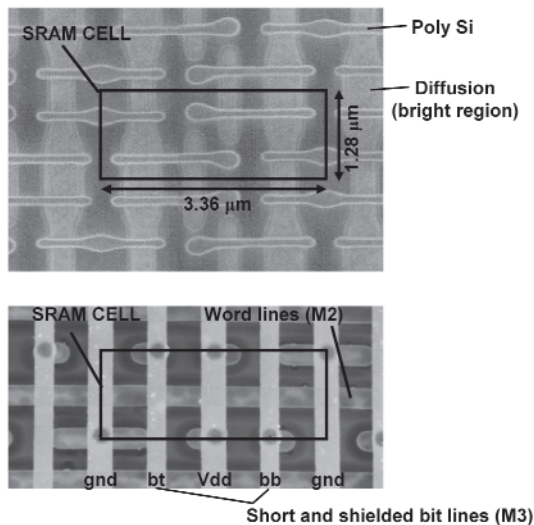


Figure 11.1.4: SEM micrograph of LS-cell.

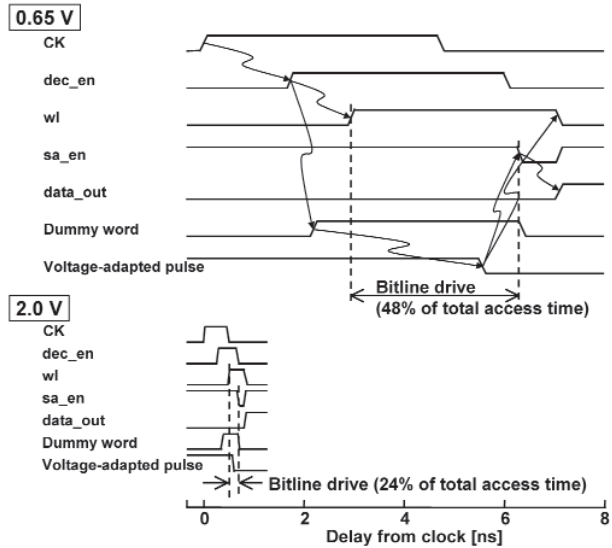


Figure 11.1.2: Timing diagrams.

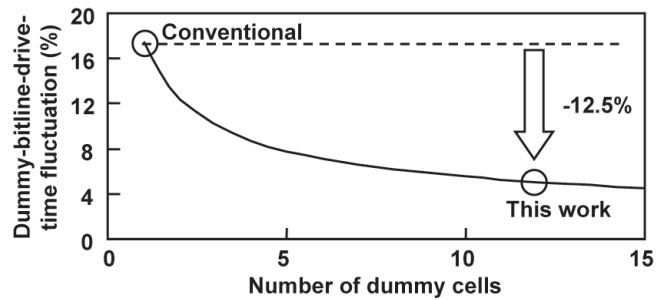


Figure 11.1.3: The effect of plural dummy cells.

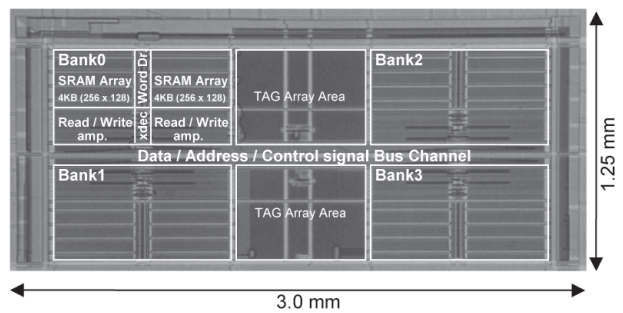


Figure 11.1.5: Chip micrograph.

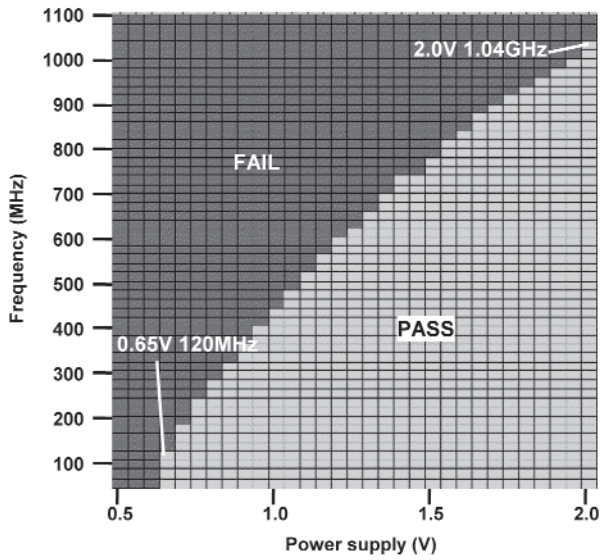


Figure 11.1.6: Shmoo plot.

Process	4-metal 0.18μm enhanced CMOS
Threshold voltage	0.5V: memory cell 0.4V: peripherals
Metal pitch	0.52μm
Cell size (6T)	1.28 x 3.36μm
Cache size	3.0 x 1.25mm
Supply voltage	0.65V-2.0V
Frequency	120MHz-1.04GHz
Power dissipation	1.7mW-530mW
Organization	Four-way-set associative 1024 index x 64 bit x 4 set

Figure 11.1.7: Chip specifications.