Consideration of Poly-Si Loaded Cell Capacity Limits for Low-Power and High-Speed SRAM's

H. Kato, K. Sato, M. Matsui, H. Shibata, K. Hashimoto, T. Ootani, and K. Ochii

***Abstract*—The maximum bit capacity of poly-Si loaded SRAM's is estimated, based on cell stability limits. When SRAM density increases, the voltage level of a storage node in the high state decreases more quickly because of MOS drain leakage current that flows in the poly-Si load; this can prevent regular cell operation. The poly-Si load resistance and the drain leakage current distribution are measured by using special 0.8-μm 1-Mb SRAM test chips. The maximum bit capacity is then calculated for low-power and high-speed SRAM's. The limit is 4 Mb for low-power SRAM's and 4 Gb for high-speed SRAM's.**

I. Introduction

High-density SRAM's are being fabricated with fine lines on monolithic silicon with capacities reaching 16 Mb. One of the most important enabling technologies for high-density SRAM's is the use of poly-Si loads on NMOS transistors; this reduces die area significantly [1], [2]. However, the poly-Si loads are not active elements and Lu *et al.* [3] argued that this may limit the scaling of SRAM's. Their paper discussed design criteria based on the conduction mechanism of the poly-Si loads. Equivalently, it is appropriate to consider how cell stability is determined by the drain leakage current at the cell driver transistor. The purpose of our paper is to understand the role of the drain leakage current and to estimate the bit capacity limit of the poly-Si loaded SRAM.

The storage node of a poly-Si loaded cell in the high state holds its voltage level by current through the poly-Si load. With time, the storage node voltage is pulled down due to two kinds of leakage current flowing out from the node: leakage current of the backward p-n junction and subthreshold leakage. Fig. 1(a) and (b) shows the equivalent circuit and the cell structure of the typical poly-Si loaded SRAM cell, respectively. Because the leakage current flows through the poly-Si load, the voltage drop Δ is expressed as

$$\Delta = Ri_{DL}. \quad (1)$$

Here R is the resistance of the poly-Si load and i_{DL} is the drain leakage current.

The voltage drop prevents regular cell operation if the leakage current reaches a level comparable to the poly-Si resistor current.

Manuscript received September 4, 1991; revised December 17, 1991.

The authors are with the Semiconductor Device Engineering Laboratory, Toshiba Corporation, 1, Komukai, Toshiba-cho, Saiwai-ku, Kawasaki 210, Japan.

IEEE Log Number 9106419.

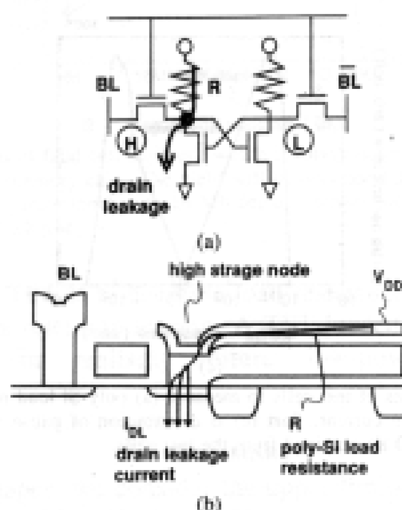

Fig. 1. Mechanism of voltage drop. (a) Equivalent circuit and (b) cross section of typical poly-Si loaded cell. Leakage current appears at high storage node.

The two kinds of leakage current previously mentioned tend to increase when the fabrication process is scaled. The higher impurity dosage and the lower temperature of heat treatment in a deep-submicrometer process cause the p-n junction leakage per unit area to increase. Moreover, lowering the threshold voltage to meet scaled supply voltage operation causes the MOSFET's subthreshold leakage to increase unit channel width. The resistance of the poly-Si load must be increased for future higher density SRAM's if they are to maintain the 1-μA retention current that is now typical. Thus, higher poly-Si load resistance and higher leakage currents are associated with higher density SRAM's and both tend to prevent regular cell operation.

The variance of poly-Si resistance R and drain leakage i_{DL} is also important. If the resistance R or the leakage current i_{DL} occasionally takes a large value, the voltage drop Δ becomes significant. This possibility is not negligible even if the variance is ultimately reduced, because a high-density SRAM will contain billion of transistors.

In this paper, we estimate the upper limit of memory bit capacity for low-power SRAM's whose retention current is no more than 1 μA. We also estimate it for high-speed SRAM's whose retention current can be more than 1 μA. Our estimate is based on measured distribution data of poly-Si load resistance and leakage current of cell nodes on a 0.8-μm 1-Mb CMOS SRAM specially designed for this purpose.

0018-9200/92$03.00 © 1992 IEEE

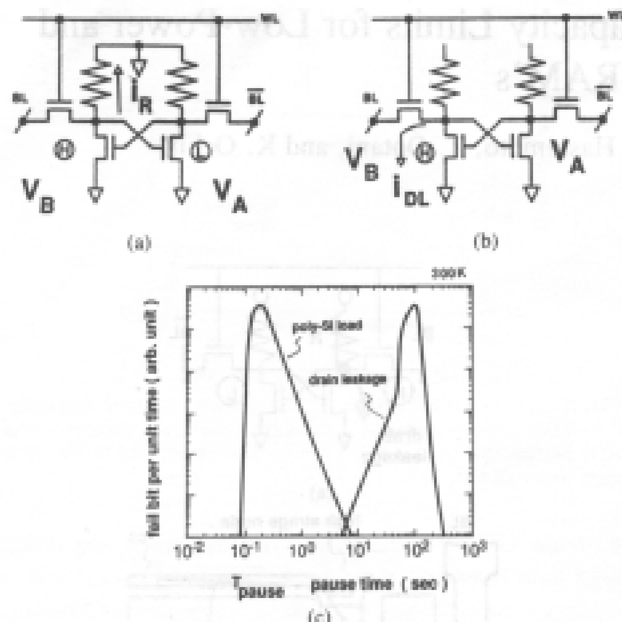

Fig. 2. Structures of test cells to measure (a) poly-Si load resistance and (b) drain leakage current. Part (c) is distribution of pause time at room temperature (300 K) obtained from the test cells.

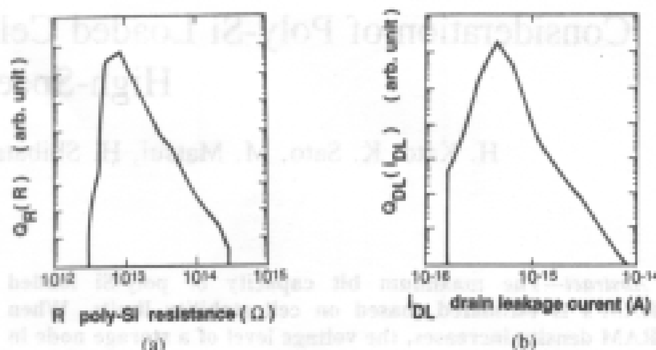

Fig. 3. Distribution of (a) poly-Si load resistance Q_R and (b) drain leakage current Q_{DL} obtained from the pause time distributions in Fig. 2(c).

II. Distribution of Poly-Si Resistance and Leakage Current

We fabricated two kinds of SRAM test chips, with cell structures as illustrated in Fig. 2(a) and (b). Both cells have poly-Si load terminals; those of Fig. 2(a) were connected to V_{SS} (ground line) and those of Fig. 2(b) were connected to nowhere, that is, opened. They are originally charged to V_{CC}. After data are written, those cells cannot hold the data statically because once written charge escapes from their storage nodes, it cannot be compensated. They can hold data for only a certain period, defined as T_{pause}. The discharging path of the cell of Fig. 2(a) is through a poly-Si load toward V_{SS} and that of the cell of Fig. 2(b) is toward the substrate as junction leakage and also toward V_{SS} as NMOS driver subthreshold leakage. The leakage current of the cell of Fig. 2(a) includes all of the components of the cell of Fig. 2(b); however, all of these components are negligibly small compared to the leakage current from poly-Si load.

The pause time distributions at room temperature (300 K), P_R and P_{DL}, of the cells of Fig. 2(a) and (b) are shown in Fig. 2(c), which are obtained from 1-Mb SRAM's fabricated by a 0.8-μm process. The activation energy obtained from the shift of the distribution peak is 0.51 eV for the poly-Si load and 0.67 eV for drain leakage.

With the hypothesis that the poly-Si load can be represented by a linear resistance R, the pause time for the cell in Fig. 2(a) is expressed as

$$T_{pause} = RC \cdot \ln (V_0/V_{cr}). \tag{2}$$

And with the hypothesis that the drain leakage current has constant value i_{DL}, the pause time for the cell in Fig. 2(b) is given by

$$T_{pause} = C(V_0 - V_{cr})/i_{DL}. \tag{3}$$

Here V_0 is the storage node high voltage just after the write operation has finished, V_{cr} is the critical voltage below which data upset occurs, and C is the storage node capacitance. Transforming the distribution P_R or P_{DL} into R or i_{DL} with the use of (2) or (3), we obtain the distribution of poly-Si load Q_R and leakage current Q_{DL} as follows:

$$Q_R(R) = C \ln (V_0/V_{cr}) \cdot P_R(CR \ln (V_0/V_{cr})) \tag{4}$$

$$Q_{DL}(i_{DL}) = C(V_0 - V_{cr})/i_{DL}^2 \cdot P_{DL}(C(V_0 - V_{cr})/i_{DL}). \tag{5}$$

The value V_{cr} reflects the cell stability and this value is determined by SPICE simulation and measured data from the test element group (TEG).

Fig. 3(a) and (b) shows the distributions of poly-Si load resistance and leakage current obtained from the data of Fig. 2(c). Note that these are distributed over a range as wide as two orders of magnitude.

III. Degradation of Yield

The drain leakage current of the cell driver transistor causes a decrease in the cell storage node voltage when it starts in the high state. The amount of the voltage drop, Δ, is given by $\Delta = Ri_{DL}$. The distribution of Δ is derived from the distributions of Q_R and Q_{DL} as follows:

$$Q_\Delta(\Delta) = N \int_0^\infty dR\, Q_R(R) Q_{DL}(\Delta/R). \tag{6}$$

Here N is the normalization constant. Fig. 4 shows the distribution of Δ. The solid line corresponds to a 1-Mb density SRAM. The most frequent value of Δ is 3.8 mV, which is so small that it does not affect cell stability. From this result, we are convinced that a 0.8-μm 1-Mb SRAM should have a good immunity to the drain leakage current. Will this immunity be maintained for a 4-Mb SRAM or on even higher density SRAM? To analyze this problem, the distributions Q_R and Q_{DL} are needed.

It is assumed that the distributions Q_R and Q_{DL} will change as follows, when the density is increased by 4 times for each generation:

a) For a low-retention-power SRAM:

$$Q_R(R) \rightarrow \frac{1}{4} Q_R\left(\frac{R}{4}\right), \quad Q_{DL} \rightarrow Q_{DL}(i_{DL}). \tag{7}$$

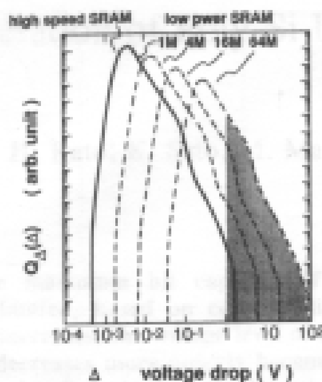

Fig. 4. Distribution of voltage drop Δ that is calculated from the distributions Q_R and Q_{DL} in Fig. 3(a) and (b).

b) For a high-speed SRAM:

$$Q_R(R) \rightarrow Q_R(R), \quad Q_{DL} \rightarrow Q_{DL}(i_{DL}). \tag{8}$$

For a low-retention-power SRAM with a 4 times higher density, the median of the poly-Si load resistance must be four times larger to maintain a retention current of 1 μA. On the other hand, for a high-speed SRAM with a 4 times higher density, the distribution of Q_R is not changed because the retention current of 1 μA is not restricted to 1 μA. For both low-power and high-speed SRAM's, the drain leakage distribution Q_{DL} is assumed constant because the increase of leakage density is compensated by the reduction of leakage area.

The dashed curves in Fig. 4 are Δ distribution for low-power SRAM, which are obtained by the assumption of (7) and the formula (6). The distribution for high-speed SRAM is the same over several generations because of the assumption (8). So the solid line in Fig. 4 also represents the Δ distribution for high-speed SRAM's.

The fatal cell number f is formulated as

$$f = B \cdot \int_{\Delta_B}^{\infty} d\Delta Q_\Delta(\Delta). \tag{9}$$

Here B is the memory bit capacity. The value f corresponds to the number of cells in a chip that have a voltage drop greater than Δ_B, where normal cell operation is not possible.

If the probability of a fatal cell is dominated by the Poisson distribution, the yield Y of SRAM chip is represented by $Y = e^{-f}$. Fig. 5 shows the dependence of f and Y on the memory bit capacity B. Here the critical value of Δ_B is set to 1 V for all values of memory bit capacity B. The integral in (9) corresponds to the shaded area in Fig. 5. In a strict sense, this value must be changed if the supply voltage (V_{CC}) is scaled down with a new generation, but this is not necessary since the result of this paper is insensitive to V_{CC}. Note that the fatal cell number f increases exponentially and the yield Y decreases catastrophically as the memory bit capacity B reaches 16 Mb for a low-retention-power SRAM and 16 Gb for a high-speed SRAM.

To reduce the voltage drop, a thin-film transistor (TFT) can be employed for the load of a SRAM [3]–[5]. The role of the TFT is to decrease load resistance and realize a large data retention margin. A TFT loaded cell will be necessary in realizing future low-power-retention SRAM's.

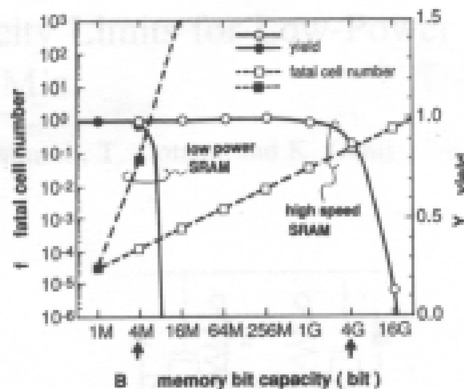

Fig. 5. Trends of fatal cell number f which involves pause failure and the yield Y of the memory chip. The yield will be decreased drastically when memory bit capacity reaches to 16 Mb for low-power SRAM and 16 Gb for high-speed SRAM.

IV. Summary

In this paper, we consider the upper limits of memory capacity for SRAM's intended for low-power and for high-speed use. As density is increased, the voltage of a cell storage node in the high state tends to decrease more because the value of poly-Si load resistance increases and/or the drain leakage current increases. To determine the density limits, we propose a method to measure the distribution of poly-Si load resistance and the drain leakage current. The maximum bit capacity in an SRAM with a 1-μA retention current is 4 Mb, and the maximum bit capacity feasible for high-speed SRAM is 4 Gb, where the minimum retention current is about 4 mA.

The drain leakage current due to crystal defects is another component responsible for the voltage drop of high storage node. If this component exists, the upper limit of memory capacity becomes lower than estimated above. Practically, the defect leakage current should be also considered in addition to that discussed in this paper.

Acknowledgment

The authors would like to thank S. Watanabe for his encouragement and critical reading of the manuscript.

References

[1] T. R. O'Connel, J. M. Hartman, E. D. Errett, G. S. Leach, and W. C. Dunn, "A 4 K static clocked and non-clocked RAM design," in *ISSCC Dig. Tech. Papers*, Feb. 1977, pp. 14–15.

[2] V. G. McKenny, "A 5 V-only 4 K static RAM," in *ISSCC Dig. Tech. Papers*, Feb. 1977, pp. 100–101.

[3] N. C. Lu, L. Gerzberg, and J. D. Meindl, "Scaling limitations of monolithic polycrystalline-silicon resistors in VLSI static RAM's and logic," *IEEE J. Solid-State Circuits*, vol. SC-17, no. 2, pp. 312–320, Apr. 1982.

[4] T. Yamanaka *et al.*, "A 25 μm^2, new poly-Si PMOS load (PPL) SRAM cell having excellent soft error immunity," in *IEDM Tech. Dig.*, 1988, pp. 48–51.

[5] T. Ootani *et al.*, "A 4-Mb CMOS SRAM with a PMOS thin-film-transistor load cell," *IEEE J. Solid-State Circuits*, vol. 25, no. 5, pp. 1082–1090, Oct. 1990.