similar size. The on-resistance, which is proportional to the reciprocal of the aspect ratio $L/6NZ$, is only 0.05 Ω. By increasing the aspect ratio, lower on-resistance can be obtained.

8.6 NONVOLATILE MEMORY DEVICES

When the gate electrode of a conventional MOSFET is modified so that semipermanent charge storage inside the gate is possible, the new structure becomes a nonvolatile memory device. Since the first nonvolatile memory device proposed by Kahng and Sze[71] in 1967, various device structures have been made, and nonvolatile memory devices have been extensively used in integrated circuits[72] such as the electrically alterable read-only memory (EAROM), the erasable-programmable read-only memory (EPROM), and the nonvolatile random-access memory (NVRAM).

The two groups of nonvolatile memory devices are the floating-gate devices and the MIOS (metal–insulator–oxide–semiconductor) devices (Fig. 60). In both devices, charges are injected from the silicon across the

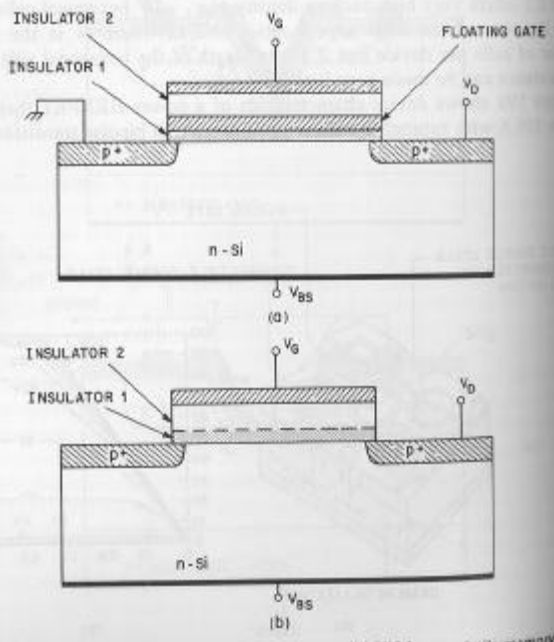

Fig. 60 (a) Floating-gate nonvolatile memory. (b) MIOS nonvolatile memory.

first insulator and stored in the floating gate or at the insulator–oxide interface of the MIOS device. The stored charge gives rise to a threshold voltage shift, and the device is at a higher-threshold voltage state. For a well-designed memory device, the charge retention time can be over 100 years. To erase the stored charge and return the device to a "lower-threshold voltage state," a gate voltage or other means (such as ultraviolet light) can be used.

8.6.1 Floating-Gate Device

The energy-band diagram of the first floating-gate device has an n-type silicon substrate, a layered gate structure of a thin oxide I(1), a floating metal gate M(1), a thick insulator I(2), and an external metal gate M(2) (Fig. 61). Upon application of a positive voltage V_G to the external gate, an electric field is established in each of the two insulators (Fig. 61a). We have, from Gauss's law, that

$$\epsilon_1 \mathscr{E}_1 = \epsilon_2 \mathscr{E}_2 + Q \tag{105}$$

and

$$V_G = V_1 + V_2 = d_1 \mathscr{E}_1 + d_2 \mathscr{E}_2 \tag{106}$$

where ϵ_1 and ϵ_2 are the dielectric permittivities of insulators 1 and 2; $\mathscr{E}_1$ and $\mathscr{E}_2$ are the corresponding fields, V_1 and V_2 are the voltages developed across insulators 1 and 2, and Q is the stored charge on the floating gate. From Eqs. 105 and 106 we obtain

$$\mathscr{E}_1 = \frac{V_G}{d_1 + d_2(\epsilon_1/\epsilon_2)} + \frac{Q}{\epsilon_1 + \epsilon_2(d_1/d_2)}. \tag{107}$$

During the application of V_G, the charge on the floating gate changes with time, provided that the currents in the two insulators are not equal, that is,

$$Q(t) = \int_0^t [J_1(\mathscr{E}_1) - J_2(\mathscr{E}_2)]\, dt \qquad \text{C/cm}^2 \tag{108}$$

where $J_1(\mathscr{E}_1)$ and $J_2(\mathscr{E}_2)$ are the current densities in insulator 1 and 2.

The current transport in insulators is generally a strong function of the electric field. When the transport is Fowler–Nordheim tunneling, the current density has the form

$$J = C_1 \mathscr{E}^2 \exp(-\mathscr{E}_0/\mathscr{E}) \tag{109}$$

where $\mathscr{E}$ is the field, and C_1 and $\mathscr{E}_0$ are constants in terms of effective mass and barrier height. This type of current transport occurs in SiO_2 and Al_2O_3 as discussed in Chapter 7. When the transport is of the Frankel–Poole type, which occurs in Si_3N_4, the current density follows the form

$$J = C_2 \mathscr{E} \exp\left[-q\left(\phi_B - \sqrt{q\mathscr{E}/\pi\epsilon_i}\right)\Big/ kT\right] \tag{110}$$

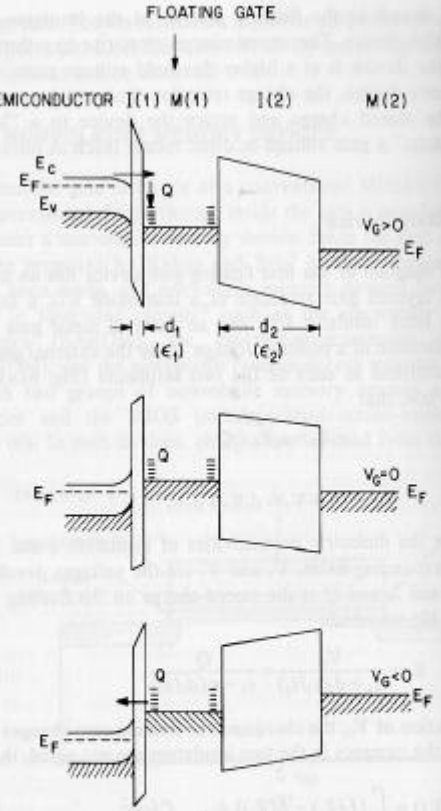

Fig. 61 Energy-band diagram of a floating-gate memory device. (a) Charging (writing mode). (b) Charge storage on the floating gate. (c) Discharging (erasing mode). (After Kahng and Sze, Ref. 71.)

where C_2 is a constant in terms of the trapping density in the insulator, ϕ_B the barrier height, and ϵ_i the dynamic permittivity.

After time t, the applied V_G is removed (Fig. 61*b*) and the stored charge Q causes a shift of the threshold voltage by the amount

$$\Delta V_T = -\frac{d_2}{\epsilon_2} Q. \tag{111}$$

To erase the stored charge, one can apply a negative gate voltage (Fig. 61*c*).

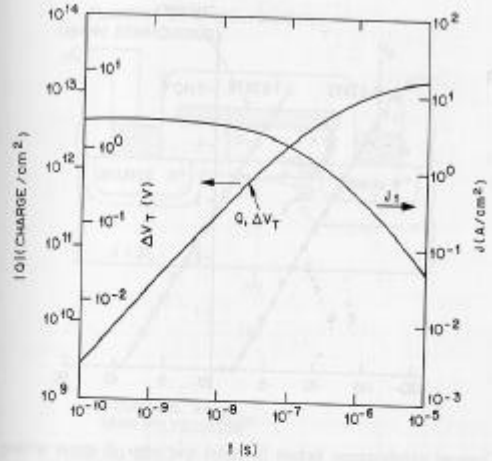

Fig. 62 Calculated charging current and stored charge as a function of charging time. (After Kahng and Sze, Ref. 71.)

Figure 62 gives the results of a theoretical computation, using Eqs. 106 through 109 with the following parameters: $d_1 = 50$ Å, $\epsilon_1 = 3.85\epsilon_0$ (for SiO_2), $d_2 = 1000$ Å, $\epsilon_2 = 30\epsilon_0$ (for ZrO_2), $V_G = 50$ V, and the current J_2 is assumed to be zero. Note that initially the stored charge increases linearly with time and then saturates. For a short time the current is almost constant and then decreases rapidly. The results above can be explained as follows: when a voltage pulse is applied at $t = 0$, the initial charge Q is zero and the initial electric field across I(1) has its maximum value $\mathcal{E}_1 = V_G/[d_1 + (\epsilon_1/\epsilon_2)d_2]$, from Eq. 107. When Q (which is negative for electrons) is sufficiently small so that $\mathcal{E}_1$ remains essentially the same, the current in turn remains the same, and Q increases linearly with time. When Q is large enough to reduce the value of $\mathcal{E}_1$ substantially, the current decreases rapidly with time and $|Q|$ begins to saturate.

Figure 62 also shows the threshold voltage shift based on Eq. 111. For this particular device, to increase the threshold voltage by 1 V, less than 0.1 μS of charging time (also called writing time) is required. Experimentally, the threshold voltage shift can be measured from the drain conductance. The change in V_T results in a change in the channel conductance g_D of the MOSFET. For small drain voltages, the channel conductance of a p-channel MOSFET is given by

$$g_D = -\frac{Z}{L}\mu_p C_i (V_G - V_T), \qquad V_G > V_T. \tag{112}$$