35 E. O. Johnson, *J. Appl. Phys.*, **28**, 1349 (1957).
36 J. S. Brugler and P. G. A. Jespers, *IEEE Transact. Electron Devices*, **ED-16**, 297 (1969); J. Golder and E. Baldinger, *Helv. Phys. Acta*, **44**, 387 (1971); M. Declercq and P. Jespers, *Revue HF (Belg.)*, **9**, 244 (1974).
37 W. Shockley and W. T. Read, Jr., *Phys. Rev.*, **87**, 835 (1952).
38 M. Zerbst, *Z. Angew. Phys.*, **22**, 30 (1966).
39 C. T. Sah, R. N. Noyce, and W. Shockley, *Proc. IRE*, **45**, 1228 (1957).
40 T. W. Collins and J. N. Churchill, *IEEE Transact. Electron Devices*, **ED-22**, 90 (1975).
41 P. U. Calzolari and S. Graffi, *Solid-State Electron.*, **15**, 1003 (1972).
42 R. F. Pierret, *IEEE Transact. Electron Devices*, **ED-19**, 869 (1972).
43 R. F. Pierret and D. W. Small, *IEEE Transact. Electron Devices*, **ED-22**, 1051 (1975).
44 D. K. Schroder and H. C. Nathanson, *Solid-State Electron.*, **13**, 577 (1970).
45 Y. Kano and A. Shibata, *Jap. J. Appl. Phys.*, **11**, 1161 (1972).
46 D. K. Schroder and J. Guldberg, *Solid-State Electron.*, **14**, 1285 (1971).
47 D. W. Small and R. F. Pierret, *Solid-State Electron.*, **19**, 505 (1976).
48 M. V. Whelan, IEDM Washington, D.C., October 1969 and *Philips Research Reports*, Supplement No. 6 (1970).
49 R. F. Pierret, *Solid-State Electron.*, **17**, 1257 (1974).
50 R. F. Pierret and R. J. Grossman, *Solid-State Electron.*, **20**, 373 (1977).

10

Charges, Barrier Heights, and Flatband Voltage

10.1 INTRODUCTION

In this chapter we discuss (1) the contribution of charges and of work function differences to the flatband voltage, (2) how these contributions are determined, and (3) how flatband voltage is measured. This chapter lays the groundwork for parts of Chapters 11 and 15.

Section 10.2 describes how oxide charge affects MOS capacitor characteristics and how its polarity and magnitude can be determined. Section 10.3 describes how internal photoemission can be used to determine the energy barrier height between silicon and SiO_2 and between the gate and the SiO_2. Work function differences can be determined from these barrier heights. Section 10.4 describes how the work function difference between gate material and silicon affects C-V characteristics and how it is measured. Section 10.5 describes how flatband voltage is measured when interface traps are present and the uncertainties involved in the extraction of oxide charge from flatband voltage caused by charged interface traps and doping nonuniformities.

10.2 OXIDE CHARGE

There are two basic types of charge in the SiO_2 layer: (1) interface trap charge, discussed in Chapters 5–7, and (2) oxide charge. The feature that distinguishes interface trap charge from oxide charge is that interface trap charge varies with gate bias, whereas oxide charge is independent of gate bias. We ignore interface trap charge in this section and concentrate on oxide charge.

There are three types of oxide charge Q_o, that are technologically important. The first type, oxide fixed charge Q_f, is the charge density remaining after interface trap charge is annealed out (see Section 15.4.3). Oxide fixed charge is located at or very near the Si–SiO_2 interface (see Chapters 11 and 15). In electrical measurements Q_f can be regarded as a charge sheet located at the Si–SiO_2 interface.

The second type of oxide charge, oxide trapped charge Q_{ot}, usually is located either at the metal-SiO_2 interface or at the Si–SiO_2 interface. One exception is when the oxide traps are introduced by ion implantation. In this case Q_{ot} can be distributed within the oxide layer. Oxide trapped charge is commonly produced by the injection of hot electrons or holes from an avalanche plasma in a high field region in the silicon, injection of carriers by photoemission or by exposure to ionizing radiation. These effects are discussed in Chapter 11.

The third type, mobile ionic charge Q_m, most commonly is caused by the presence of ionized alkali metal atoms such as sodium or potassium. This type of charge is located either at the metal-SiO_2 interface, where it originally entered the oxide layer, or at the Si–SiO_2 interface, where it has drifted under an applied field. Drift can occur because such ions are mobile in SiO_2 at relatively low temperatures.

Immobile oxide charge can be distinguished from mobile ionic charge by a bias-temperature aging experiment. Gate bias is applied for a given length of time while the sample is held at a moderately elevated temperature. The density of oxide fixed charge may change under this treatment, but the centers responsible for this charge do not move. However, the mobile ionic charge can be cycled back and forth between the metal-SiO_2 interface and the Si–SiO_2 interface by this treatment without discharge, and the resulting ionic current can be detected.

10.2.1 Measurement of Oxide Charge

Oxide charge is an important parameter in devices. For example, it can alter the threshold voltage of a MOSFET, alter the silicon surface potential and thereby change reverse surface leakage current in a p-n junction, alter the avalanche breakdown voltage of a p-n junction, and alter common emitter current gain at low collector current in a bipolar transistor. It can invert the silicon, resulting in unwanted current paths between elements in

an integrated circuit. It also must be understood for the correct interpretation of C-V curves measured on an MOS capacitor.

The simplest and most widely used method for measuring oxide charge density is to infer this density from the voltage shift of a C-V curve. We start by calculating the effect of oxide charge on the C-V curve. Let $n_o(x)$ be the *volume* density of oxide charge. This charge induces an *image* charge in the silicon surface. Therefore, the capacitance at any given gate bias will be different from what it would be if $n_o(x) = 0$. The gate bias required to compensate the image charge produced in the silicon surface by this oxide charge and produce the flatband condition ($\psi_s = 0$) is given by integration of Poisson's equation in the oxide. Poisson's equation is

$$\frac{d^2\psi}{dx^2} = -\frac{qn_o(x)}{\epsilon_{ox}} \tag{10.1}$$

where x is measured from the metal-SiO_2 interface, $n_o(x)$ is the volume density of positive oxide charge, and ψ is the band bending in the oxide. Integrating (10.1) from a point x in the oxide to the silicon surface, $x = x_o$, where $d\psi/dx$ vanishes in the flatband condition,* (10.1) becomes

$$-\frac{d\psi}{dx} = -\frac{q}{\epsilon_{ox}}\int_x^{x_o} dx'\, n_o(x'). \tag{10.2}$$

Integrating (10.1) a second time, from the gate $x = 0$ to the silicon surface, (10.2) becomes

$$\psi(0) = -\frac{q}{\epsilon_{ox}}\int_0^{x_o} dx \int_x^{x_o} dx'\, n_o(x'). \tag{10.3}$$

At the gate $\psi(0)$ is related to the gate bias at flatbands V_{FB} by

$$\psi(0) = V_{FB} - W_{ms} \tag{10.4}$$

where W_{ms} represents the *work function difference* discussed in detail in Section 10.4; W_{ms} is a nonelectrical contribution to the gate bias, resulting from work done against chemical forces in moving an electron from the silicon to the gate electrode. The double integral in (10.3) can be reduced to a single integral by interchanging the order of integration

$$\int_0^{x_o} dx \int_x^{x_o} dx'\, n_o(x') = \int_0^{x_o} dx'\, n_o(x') \int_0^{x'} dx = \int_0^{x_o} dx'\, x'\, n_o(x'). \tag{10.5}$$

*For nonuniform doping, $(d\psi/dx)$ does not vanish at the silicon surface when ψ_s vanishes. In this case the left-hand-side of (10.2) becomes $Q_s(0)/\epsilon_{ox} - d\psi/dx$, where $Q_s(0)$ is the silicon surface charge per unit area at zero band bending (see Section 10.5).

When (10.5) and (10.4) are used in (10.3), (10.3) becomes

$$V_{FB} - W_{ms} = -\frac{q}{\epsilon_{ox}} \int_0^{x_o} dx'\, x'\, n_o(x'). \tag{10.6}$$

From (10.6), $V_{FB} - W_{ms}$ is the portion of the flatband voltage due to oxide charge.

The relationship (10.6) between V_{FB} and the charge distributed in the oxide is clarified by introducing an average distance $\bar{x}$ called the *centroid* of the charge distribution. The charge centroid is defined by

$$\bar{x} = \frac{\int_0^{x_o} x' n_o(x')\, dx'}{\int_0^{x_o} n_o(x')\, dx'}. \tag{10.7}$$

Because $q \int_0^{x_0} n_o(x')\, dx' = Q_o$, (10.7) becomes

$$\bar{x} Q_o = q \int_0^{x_o} x' n_o(x')\, dx'. \tag{10.8}$$

From (10.6) and (10.8), we obtain

$$V_{FB} - W_{ms} = \frac{\bar{x} Q_o}{\epsilon_{ox}}. \tag{10.9}$$

The quantity $\bar{x}Q_o$ is called the *first moment* of the charge distribution. Equation (10.9) shows that $V_{FB} - W_{ms}$ is a measure of the first moment of the charge distribution. Thus the oxide charge distribution cannot be obtained from a measurement of V_{FB} alone. In fact, not even $\bar{x}$ and Q_o can be obtained from such a measurement. To separate $\bar{x}$ and Q_o, either the oxide charge density or the centroid must be known independently. For the common and important case of thermally grown oxides, where oxide charge is located in a sheet at the Si–SiO_2 interface, Q_o is obtained directly from a measurement of V_{FB} because the centroid is known to be the oxide thickness (within about 30 Å or less). For this case, (10.8) or (10.9) becomes

$$V_{FB} - W_{ms} = -\frac{x_o Q_o}{\epsilon_{ox}} = -\frac{Q_o}{C_{ox}} \tag{10.10}$$

where $\epsilon_{ox}/x_o = C_{ox}$, the oxide layer capacitance per unit area. Equation (10.10) still shows that $V_{FB} - W_{ms}$ is proportional to the first moment of the charge distribution. But now we can calculate Q_o from (10.10) because $\bar{x}$ is independently known from C_{ox}.

Apart from a discussion of mobile ionic charge, the remainder of this chapter is restricted to the case of a charge sheet located at the Si–SiO_2 interface where V_{FB} and oxide charge density are related by (10.10). This case applies to oxide charge density characteristic of thermal oxidation. The case of bulk charge distributed in the oxide is discussed in Section 11.4, where photo I-V measurements are used to separate Q_o and $\bar{x}$.

Equation (10.6) or (10.10) shows that the entire C-V curve is shifted along the voltage axis with respect to the ideal C-V curve [$n_o(x) = 0$] by the amount $V_{FB} - W_{ms}$. The capacitance at each value of bias is shifted by $V_{FB} - W_{ms}$, so that the shape of the C-V curve is unaltered, and the C-V curve is parallel to the ideal C-V curve.

Figures 10.1*a*,*b* illustrate the shift along the voltage axis of a high frequency C-V curve when positive Q_o is present in the oxide. The voltage shift is measured with respect to an ideal high frequency C-V curve where $Q_o = 0$ and W_{ms} corresponds to the gate material of the experimental device. The C-V curve for a p-type substrate is shown in Fig. 10.1*a* and for an n-type substrate, in Fig. 10.1*b*. In both cases positive Q_o causes the C-V curve to shift to more negative values of gate bias with respect to the ideal C-V curve. Positive Q_o is the most commonly observed process

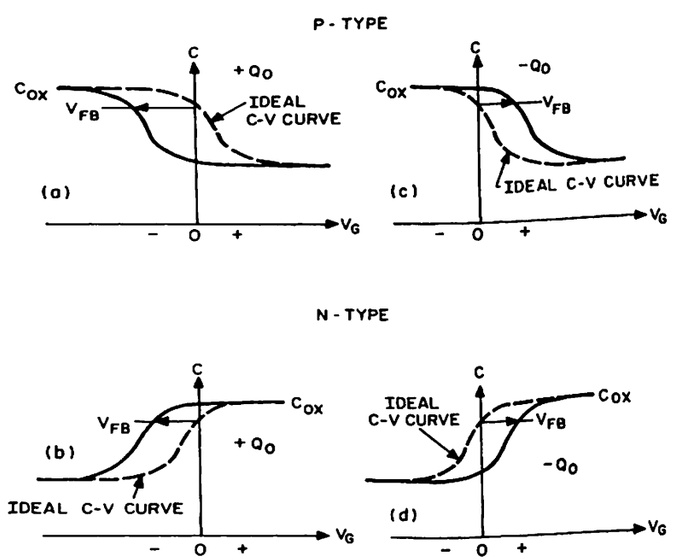


Fig. 10.1 High frequency capacitance as a function of gate bias showing the effect of oxide charge. The C-V curves marked "ideal" have no oxide charge. (*a*) Positive oxide charge, p-type; (*b*) positive oxide charge, n-type; (*c*) negative oxide charge, p-type; (*d*) negative oxide charge, n-type.

related charge. Although negative Q_o is not as common, it can be produced deliberately by the injection of electrons into the SiO_2, for example. To show that the C-V curves can be used to determine the polarity as well as the magnitude of oxide charge, Figs. 10.1c, d show the shift of a high frequency C-V curve along the voltage axis with respect to the ideal C-V curve for negative Q_o. Figure 10.1c shows a p-type substrate and Fig. 10.1d, an n-type substrate. The C-V curve is shifted to more positive bias values for both p-type and n-type substrates for negative oxide charge. Thus, from the direction of the voltage shift of the C-V curve, the polarity of Q_o can be determined, provided that Q_o is the sole cause of the shift. From the magnitude of the voltage shift of the C-V curve, the value of Q_o can be determined from (10.10). A low frequency C-V curve also could be used for this purpose.

The voltage shift of the C-V curve caused by Q_o is explained as follows, using an n-type substrate for illustration. For a given gate bias and $Q_o = 0$, the depletion layer width is such that negative charge on the gate is balanced by positive dopant ion charge in the depletion layer. If positive Q_o is introduced into the Si–SiO_2 interface, this charge balance is upset. Image charge is introduced in the gate and in the silicon. For the gate, this image charge increases the negative gate charge. For the silicon, additional electrons are added at the depletion layer edge, and the depletion layer width is reduced. Thus the capacitance for n-type is greater than for the ideal capacitor with $Q_o = 0$. The capacitance is greater for all values of gate bias, except strong accumulation where the capacitance in both cases equals C_{ox}, and in strong inversion where the capacitance reaches the same saturation value at high frequency or reaches C_{ox} again at low frequency. The result is a shift of the C-V curve to more negative gate bias as seen in Fig. 10.1b.

10.2.2 Measurement of Mobile and Oxide Fixed Charge

Consider how to distinguish between flatband voltage shifts due to mobile ionic charge and those due to oxide fixed charge.

Consider an experiment where the only oxide charge is oxide fixed charge. The initial high frequency C-V curve measured is labeled (i) in Fig. 10.2a. After heating at 180°C for half an hour with a positive gate bias producing a field of a few million volts per centimeter across the oxide, and cooling back to room temperature, the curve labeled (f_+) in Fig. 10.2a is obtained. Repeating the bias-temperature aging with negative gate bias yields curve (f_-) in Fig. 10.2a. Figure 10.2a shows that no shift in the C-V curve was produced. Therefore, from (10.9), the first moment of the oxide fixed charge density did not change under this treatment. Because it is improbable that the density of oxide charge and its centroid both changed in a manner to keep the first moment constant, we conclude that oxide fixed charge centers are immobile.

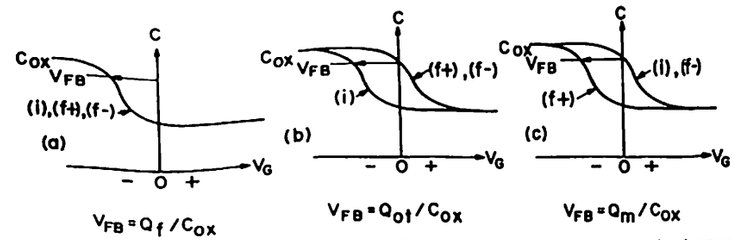


Fig. 10.2 Diagram illustrating how oxide fixed charge, oxide trapped charge, and mobile ionic charge can be distinguished from a bias-temperature aging experiment. The Symbols (i) denotes the initial C-V curve; (f_-), after positive bias-temperature aging, and (f_-), after negative bias-temperature aging. (a) Oxide fixed charge; (b) oxide trapped charge; (c) mobile ionic charge.

Repeating this bias-temperature aging experiment with oxide trapped charge of the type that anneals out at low temperatures, curves (i), (f_+), and (f_-) in Fig. 10.2b show that the first moment of oxide trapped charge is reduced. Gate bias polarity has no effect. Again, it is most likely that the oxide trapped charge centers are immobile.

Finally, Fig. 10.2c shows the results of repeating this experiment on an oxide contaminated by mobile ions. Initially V_{FB} is low, and after positive bias aging it increases. With negative bias aging, V_{FB} returns to its original value. These results are attributed to mobile ion movement that alters the centroid of the mobile ion distribution $\bar{x}$ and so alters V_{FB} according to (10.9). The alternative explanation, that the charge density also varies, is ruled out by other measurements, described in Section 10.2.3.

Assuming the ions are mobile, Fig. 10.2c is explained as follows. Initially, the ions were concentrated at the metal-SiO_2 interface where they entered the oxide. Because most of the ionic charge is imaged in the gate rather than the silicon, the initial C-V curve has a low V_{FB}. During positive gate bias aging, ions drift from the metal-SiO_2 interface to the Si-SiO_2 interface so that almost all the ionic charge is imaged in the silicon producing the shift shown by curve (f_+) in Fig. 10.2b. Aging with negative gate bias aging, ions drift from the metal-SiO_2 interface to the Si–SiO_2 condition; that is, curves (f_-) and (i) are identical.

10.2.3 Measurement of Mobile Ions

The most important mobile ion in SiO_2 is sodium for reasons given in Section 15.3. Potassium ions also are mobile. Because they are metallic, these ions are positive in SiO_2. This section describes the most practical methods of measuring mobile ion density in SiO_2. The technology for minimizing the effects of sodium on device characteristics is described in Section 15.3.4.

(a) *The Sodium Profile*

To better understand the ion density measurement techniques, it is necessary to know how sodium and sodium ions are distributed in SiO_2. In a classic paper, Yon et al.[1] determined the total and mobile ion distribution in SiO_2. Figure 10.3 illustrates the sodium profile they found in an as-grown oxide where no pains were taken to minimize sodium contamination. This curve was obtained by neutron activation analysis described in Section 10.4.1(a). The sodium contamination is greatest near the air-oxide interface where sodium entered the oxide.

The total sodium profile is shown in Fig. 10.3. However, not all the sodium is incorporated in the oxide as mobile ions. Some of the sodium is chemically bound in the silica lattice. This sodium is electrically inactive, as it is uncharged and immobile. Several other workers[2–4] have reported electrically inactive sodium. It remains unclear how the conditions of oxide growth or the method of contamination affects the amount of sodium that remains electrically inactive.* However, our interest is in the mobile ionized sodium that affects device stability.

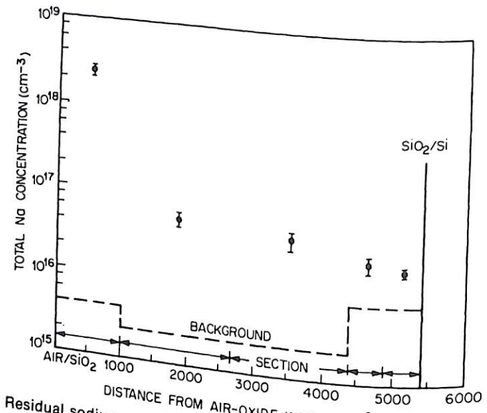


Fig. 10.3 Residual sodium concentration in thermally grown SiO_2 versus distance into the oxide measured from the air-oxide interface. These results are typical of nine oxides grown in wet oxygen and two oxides in "dry" oxygen. There was no measurable sodium in the silicon. The dashed curve is background activity shown as equivalent concentration. Arrows show oxide steps successively etched off to obtain the profile. Initial oxide thickness was 5400 Å and acceptor density was $5 \times 10^{14}\,cm^{-3}$. After Yon et al.[1] Copyright (1966), IEEE.

*It has been established that sodium originating from sodium halides evaporated onto the oxide before metallization all becomes mobile. This is a common way of deliberately introducing sodium contamination.[5]

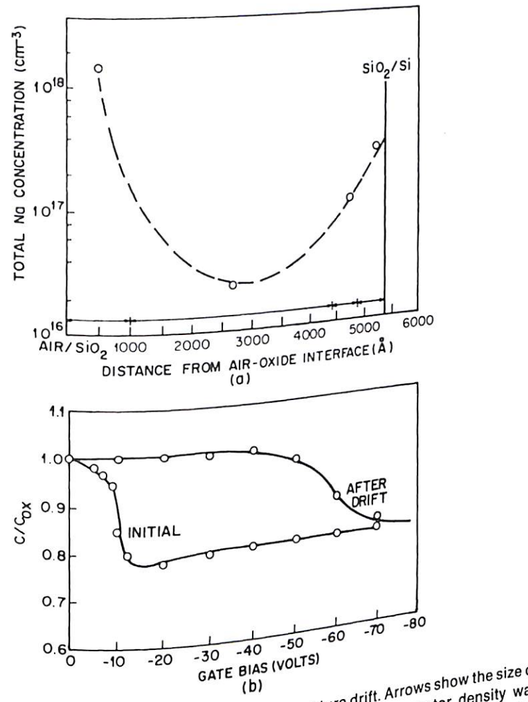


Fig. 10.4 (a) Sodium profile obtained after bias-temperature drift. Arrows show the size of the oxide etch-off steps. Initial oxide thickness was 5400 Å and acceptor density was $5 \times 10^{14}\,cm^{-3}$. After Yon et al.[1] Copyright (1966), IEEE. (*b*) Diagram illustrating *C-V* characteristics measured at 1 MHz normalized to the capacitance at $V_G = 0$ before and after bias-temperature drift at 200°C for 10 min with an applied oxide field of 5×10^5 V/cm. The initial oxide fixed charge density was $4 \times 10^{11}\,cm^{-2}$ and the change after drift was 1.9×10^{12} sodium ions/cm². The sample is the same as in (*a*) before etch-off. After Yon et al.[1] Copyright (1966), IEEE.

The mobile ionized sodium in the profile of Fig. 10.3 was drifted toward the Si–SiO_2 interface by heating the sample to 200°C for 10 min with a positive gate bias applied corresponding to an electric field of 5×10^5 V/cm. Figure 10.4*a* shows the sodium profile after this treatment. Notice the pile up of sodium at the Si–SiO_2 interface as a result of the drift of mobile sodium under the applied field. Assuming that all the charged sodium is in the last etch-off step, that is, within 1000 Å of the Si–SiO_2 interface, reasonable agreement is obtained between the sodium density in this etch-off step, $N_m = 9.7 \times 10^{11}\,cm^{-2}$, and the charge density as calculated

(a) *The Sodium Profile*

To better understand the ion density measurement techniques, it is necessary to know how sodium and sodium ions are distributed in SiO_2. In a classic paper, Yon et al.[1] determined the total and mobile ion distribution in SiO_2. Figure 10.3 illustrates the sodium profile they found in an as-grown oxide where no pains were taken to minimize sodium contamination. This curve was obtained by neutron activation analysis described in Section 10.4.1(a). The sodium contamination is greatest near the air-oxide interface where sodium entered the oxide.

The total sodium profile is shown in Fig. 10.3. However, not all the sodium is incorporated in the oxide as mobile ions. Some of the sodium is chemically bound in the silica lattice. This sodium is electrically inactive, as it is uncharged and immobile. Several other workers[2–4] have reported electrically inactive sodium. It remains unclear how the conditions of oxide growth or the method of contamination affects the amount of sodium that remains electrically inactive.* However, our interest is in the mobile ionized sodium that affects device stability.

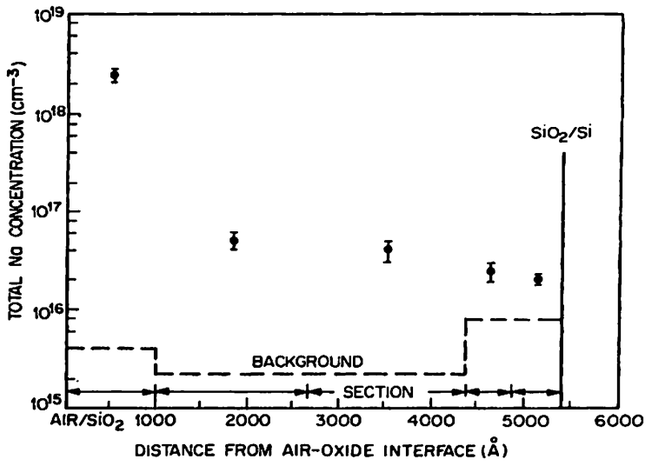
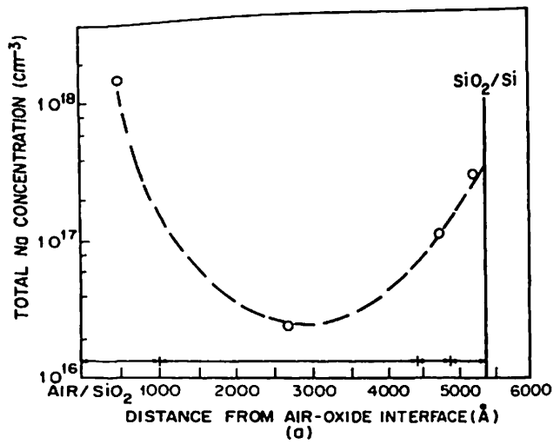


Fig. 10.3 Residual sodium concentration in thermally grown SiO_2 versus distance into the oxide measured from the air-oxide interface. These results are typical of nine oxides grown in wet oxygen and two oxides in "dry" oxygen. There was no measurable sodium in the silicon. The dashed curve is background activity shown as equivalent concentration. Arrows show oxide steps successively etched off to obtain the profile. Initial oxide thickness was 5400 Å and acceptor density was $5 \times 10^{14}\,cm^{-3}$. After Yon et al.[1] Copyright (1966), IEEE.

*It has been established that sodium originating from sodium halides evaporated onto the oxide before metallization all becomes mobile. This is a common way of deliberately introducing sodium contamination.[5]



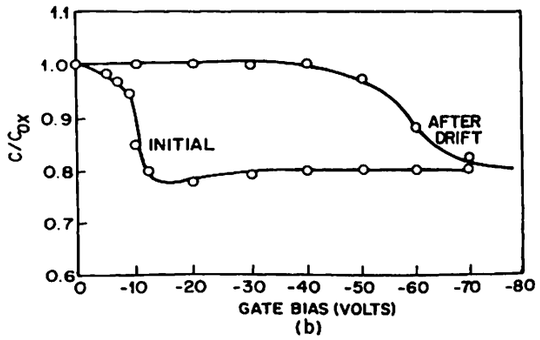


Fig. 10.4 (a) Sodium profile obtained after bias-temperature drift. Arrows show the size of the oxide etch-off steps. Initial oxide thickness was 5400 Å and acceptor density was $5 \times 10^{14}\,cm^{-3}$. After Yon et al.[1] Copyright (1966), IEEE. (*b*) Diagram illustrating *C-V* characteristics measured at 1 MHz normalized to the capacitance at $V_G = 0$ before and after bias-temperature drift at 200°C for 10 min with an applied oxide field of 5×10^5 V/cm. The initial oxide fixed charge density was $4 \times 10^{11}\,cm^{-2}$ and the change after drift was 1.9×10^{12} sodium ions/cm². The sample is the same as in (a) before etch-off. After Yon et al.[1] Copyright (1966), IEEE.

The mobile ionized sodium in the profile of Fig. 10.3 was drifted toward the Si–SiO_2 interface by heating the sample to 200°C for 10 min with a positive gate bias applied corresponding to an electric field of 5×10^5 V/cm. Figure 10.4*a* shows the sodium profile after this treatment. Notice the pile up of sodium at the Si–SiO_2 interface as a result of the drift of mobile sodium under the applied field. Assuming that all the charged sodium is in the last etch-off step, that is, within 1000 Å of the Si–SiO_2 interface, reasonable agreement is obtained between the sodium density in this etch-off step, $N_m = 9.7 \times 10^{11}\,cm^{-2}$, and the charge density as calculated

from the flatband voltage shift of the C-V curves measured before and after high temperature drift shown in Fig. 10.4*b*, $N_m = 1.9 \times 10^{12}\,cm^{-2}$.

The drifted sodium ions are actually much less than 1000 Å from the Si–SiO_2 interface, Williams[6] performed a photoemission experiment (see Section 11.4.3) in which sodium ions were drifted to the Si–SiO_2 interface, inverting the p-type silicon substrate, so that photoemission from the conduction band could be observed. From this experiment, an order of magnitude estimate can be made, placing the sodium $\approx$10–100 Å from the Si–SiO_2 interface. Using the photo I-V method described in Sections 11.4.2 and 11.4.3, DiMaria[7] found the drifted sodium ions to be within about 50 Å of the Si–SiO_2 interface. In this experiment sodium ions in concentrations up to $2.6 \times 10^{12}\,cm^{-2}$ were drifted alternately to the Si–SiO_2 interface and to the Al–SiO_2 interface with oxide fields between 2×10^6 and 4.5×10^6 V/cm at temperatures between 20 and 40°C. Photo I-V measurements then were done at 77°K to avoid ionic motion under gate bias. Ionized sodium was found to pile up within about 50 Å of both interfaces. Therefore, the sodium ions after drift can be considered to be a sheet of charge at the interface to which they were drifted.

(*b*) *C-V Method*

There are four steps in the sodium story: (1) the discovery that mobile sodium ions in the oxide were responsible for the drift in device characteristics universally experienced in the early days of the MOS technology, (2) the identification of the most common sources of sodium and the steps taken to minimize or eliminate sodium contamination during manufacture, (3) the introduction of monitoring at various steps in the fabrication to maintain control of the processing, and (4) the development of protective coatings and encapsulants to keep sodium contamination out of the oxide during device operation. All these steps are more fully described in Section 15.3.

In all four historical steps, the C-V method was used to measure mobile sodium ion density. Snow et al.[8] were the first to use the C-V method for studying sodium contamination in SiO_2. The C-V method still is the most widely used method.

In the C-V method, a high frequency (usually 1 MHz) C-V curve of an MOS capacitor is measured. Then the MOS capacitor is heated to various temperatures up to 300°C and held there for up to 30 min, which is long enough to ensure that all of the available sodium ions at the given temperature drift completely across the oxide. A positive gate bias is applied sufficient to cause an oxide field of a few MV/cm. After the given time period at elevated temperature and high field, the MOS capacitor is cooled back to room temperature and another C-V curve measured. The flatband voltage shift between the C-V curve before and after bias-temperature drift is a measure of the mobile ion concentration drifted at the given temperature.

Because mobile ions pile up in a charge sheet at the Si–SiO_2 interface, (10.10) can be used to calculate charge density from the flatband voltage shift. This calculation will give an accurate value of sodium ion density because the error in flatband voltage, discussed in Section 10.5, is unaffected by bias-temperature drift provided the change in interface trap level density is negligible, and gross nonuniformities in interfacial charge are not produced (see Chapter 6). Therefore, these errors cancel when the flatband voltage of the curve after drift is subtracted from the flatband voltage of the curve before drift to obtain the flatband voltage shift. In many applications bias-temperature treatment causes negligible changes in either interface trap level density or interfacial charge uniformity. In these cases the C-V curves before and after bias-temperature drift are parallel to each other. Then the flatband voltage shift is simply the parallel voltage shift. For those cases where significant changes occur in interface trap level density or interfacial charge uniformity, a more accurate method for obtaining sodium ion density is described in Section (c) below.

Figure 10.5 illustrates the effect of sodium ion drift on the C-V characteristics of an MOS capacitor with a negligible change of interface trap level density and no gross interfacial charge nonuniformity. Figure 10.5*a* shows the shift of the C-V curve along the voltage axis before and after bias-temperature aging, and Fig. 10.5*b* shows the charged sodium distribution in the oxide corresponding to each C-V curve; C-V curve 1 is the original C-V curve measured at 1 MHz, and C-V curve 2 is measured after heating for 5 min at 150°C with -10 V applied to the gate. For both cases, ionized sodium is piled up at the metal-oxide interface so that the image charge is induced primarily in the metal and there is virtually no effect on the C-V curve. Then C-V curve 3 is measured after heating for 5 min at 150°C with $+10$ V on the gate. Now, the ionized sodium has drifted to the Si–SiO_2 interface, where it induces a large image charge in the silicon, causing a large shift of the C-V curve toward more negative bias values. Finally, C-V curve 4 is measured after heating for 5 min at 150°C with the gate shorted to the substrate. The C-V curve has returned to its initial bias range because the sodium ions have drifted back to the metal electrode. The sodium ions end up at the metal-oxide interface under a zero applied field because the activation energy there is greater than at the Si–SiO_2 interface.

A very important observation is that the same number of sodium ions can be cycled back and forth between the silicon and metal electrodes a number of times.* Therefore, sodium ions do not exchange charge with either electrode and do not become discharged even though they pile up to within tunneling distance at each electrode.† As a consequence, the ion distribution for moderate concentrations is frozen in at room temperature

*An exception is in oxides grown in an ambient containing chlorine (see Section 15.3).
†An electrode at which ions do not discharge is called a *blocking electrode.*

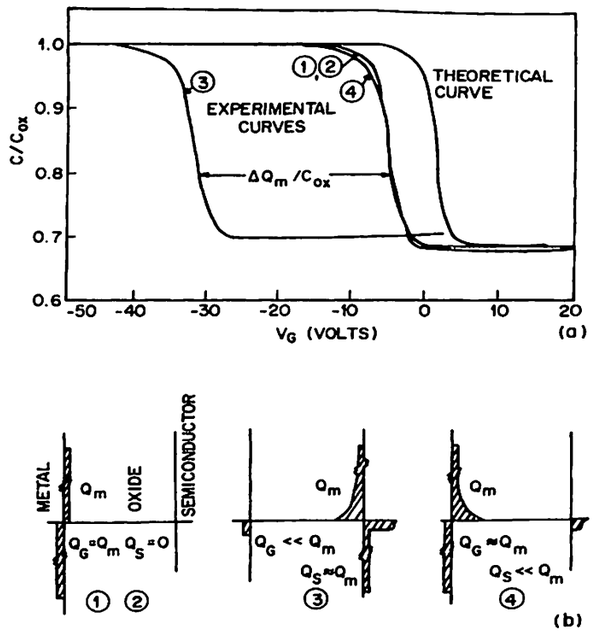


Fig. 10.5 (a) Plot of C/C_{ox} versus gate bias measured before and after sodium drift. Oxide is thermally grown in "dry" oxygen to a thickness of 2000 Å. Acceptor density was $1.5 \times 10^{16}\,cm^{-3}$. After Snow et al.[8] (*b*) Charged sodium distribution in the oxide corresponding to the various *C-V* curves in (*a*); Q_m is the ionized sodium charge density, Q_G is the image charge density induced in the metal by Q_m, and Q_s is the image charge density induced in the silicon by Q_m. After Snow et al.[8]

after drift at higher temperatures. This fact is the basis for the *C-V* method.

This failure to discharge, along with the high mobility of sodium ions in SiO_2, causes large instabilities in silicon devices. In contrast, mobile ions that become neutralized at either electrode can contribute only a transient instability during transit from one interface to the other and once discharged at an electrode, contribute no further instability. The observation that sodium ions do not discharge at the electrodes is limited to temperatures below 390°C. Chou[9] has shown experimentally that sodium ions will begin to discharge at temperatures above 390°C.

The number of sodium ions that can drift in the SiO_2 is limited by emission from the interface where the sodium initially was located (see Section 15.3 and Chapter 16). The sodium ions must overcome energy barriers at this interface before they can drift across the oxide. Therefore,

there is a distribution of activation energies associated with ion drift.[10] The fraction of the total sodium concentration that becomes ionized and drifts is temperature dependent.

After measuring the amount of ionizable sodium in the oxide, the next problem is to relate this measurement to device stability. Sodium ion densities in the low $10^{10}\,cm^{-2}$ range drifted at 200°C are considered tolerable for most integrated circuit applications, and such low sodium ion densities routinely are achieved in manufacture as described in Section 15.3.4.

The sodium ion concentration usually is measured at a temperature above the operating temperature of the device. This is done to detect smaller sodium ion densities than are detectable at operating temperature. It is difficult to predict from high temperature data the amount of sodium that would drift at operating temperature because there is a distribution of activation energies for sodium ion emission from the interfaces. For example, extrapolation using an Arrhenius plot of the drifted amount of sodium against the reciprocal of aging temperature is unjustified. In practice, the processing is adjusted until the amount drifted at aging temperature is sufficiently small that even if that amount drifted at operating temperature, device characteristics would stay within specifications. Because at operating temperature the amount that drifts is significantly lower than at aging temperature, this approach is conservative.

(c) *Triangular Voltage Sweep Method*

In the triangular voltage sweep (TVS) method a triangular voltage ramp is applied to the gate and ionic displacement current measured at an elevated temperature. This method has four advantages over the *C-V* method:

1 The mobile ion density is accurately obtained even if interface trap level density changes significantly by heating or if gross nonuniformity in interfacial charge occurs (e.g., as a result of defect decoration). An example where significant changes in interface trap level density occur is an MOS capacitor having an aluminum gate deposited by electron-gun evaporation. X-Rays generated when electrons strike the molten aluminum target cause ionizing radiation damage in the SiO_2 as described in Section 11.6. Although the electrical effects of this damage can be annealed out, subsequent bias-temperature treatments will cause large increases in interface trap level density.

2 Different mobile ionic species such as sodium and potassium can be separated at a given temperature because the peak in gate current occurs at a different gate bias. The density of each can be determined.[11] The *C-V* method would measure the sum total of all mobile ionic species in the oxide.

3 Ion densities as low as $10^9\,cm^{-2}$ can be detected, making this method more sensitive than the C-V method. The greater sensitivity of the TVS method makes it possible to measure sodium ion drift at lower temperatures than possible with the C-V method.
4 The TVS method is faster than the C-V method because only one curve is needed and the sample can remain heated for many measurements.

The one disadvantage of the TVS method is that the analysis is more complicated than that of the C-V method. For the major application of monitoring the fabrication process, the C-V method, because of its simpler analysis, is used rather than the TVS method. However, the advantages of the TVS method warrant its further discussion.

Yamin[12] was the first to use the TVS method for detecting mobile ions in SiO_2. Kuhn and Silversmith[13] and Chou[9] discussed the application of the TVS method for determining mobile ion charge density in SiO_2.

In the TVS method the MOS capacitor is held at a constant temperature while a linear voltage ramp is applied to the gate and the resulting gate current measured against gate bias as the mobile ions drift from one electrode to the other. The triangular ramp voltage then sweeps the ions back to the other electrode and the measured gate current changes direction. The circuit shown in Fig. 12.5 and described in Section 12.5.2 can be used to measure gate current, and the pedestal design shown in Fig. 12.18 can be used to heat and contact the sample.

The idealized gate current as a function of applied bias with and without mobile ionic charge is shown in Fig. 10.6. At room temperature, gate current is

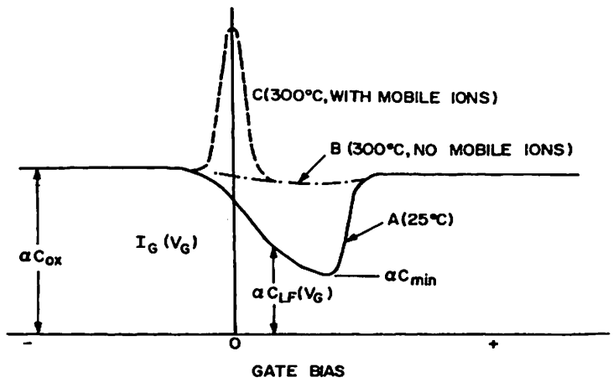


Fig. 10.6 Gate current as a function of gate bias in response to a linear voltage ramp in an MOS capacitor having a p-type substrate. Curves A (25°C) and B (300°C) would be measured when there are no mobile ions in the oxide. Curve C is the displacement current due to mobile ion drift from the oxide bulk to an interface. After Kuhn and Silversmith.[13] Reprinted by permission of the publisher, The Electrochemical Society.

proportional to the low frequency differential capacitance shown in curve A. An increase in temperature results in an increase of C_{min}/C_{ox} as described in Section 3.2.4. At temperatures typically about 300°C, gate current of an uncontaminated device becomes nearly constant and equal to C_{ox}, as shown in curve B. This constant capacitance results from the large value of n_i at 300°C, which makes the silicon surface capacitance large compared to C_{ox}. However, at 300°C with mobile ions present, a gate current peak near $V_G = 0$ is superposed on the C-V characteristic, as shown in curve C.

The peak in current of curve C arises as follows. At large negative gate bias all the mobile ions are at the metal-oxide interface, and the gate current that flows is proportional to C_{ox}. As gate voltage increases, mobile ions begin drifting toward the Si–SiO_2 interface, attracting an increasing number of electrons to the silicon surface. That is, the ionic movement causes extra electrons to flow from the gate to the silicon through the external circuit, thereby increasing gate current. This excess current peaks when the largest number of mobile ions is traversing the oxide layer. As gate voltage is increased further, mobile ions pile up at the Si–SiO_2 interface. Fewer ions flow, and the excess gate current falls. Ultimately, all mobile ions that will drift at the given temperature have piled up at an interface. Then the gate current again becomes proportional to C_{ox}.

The mobile ion density drifted at a given temperature is proportional to the area under the peak in the gate current caused by the ionic motion. To show this, we determine the gate current per unit area I_G defined as

$$I_G \equiv \frac{dQ_G}{dt} \tag{10.11}$$

where Q_G is the gate charge, given by charge neutrality as

$$Q_G = -Q_m - Q_{ot} - Q_f - Q_{it} - Q_s. \tag{10.12}$$

Although the mobile charge Q_m moves, the total amount is time independent. Therefore, taking time derivatives of (10.12), we find

$$I_G = -\frac{dQ_{it}}{dt} - \frac{dQ_s}{dt}. \tag{10.13}$$

If V_G varies slowly enough, the equilibrium expressions relate Q_{it} and Q_s to the band bending $\psi_s(t)$. Under these conditions

$$-\frac{dQ_{it}}{dt} = C_{it}(\psi_s)\frac{d\psi_s}{dt} \tag{10.14a}$$

and

$$-\frac{dQ_s}{dt} = C_s(\psi_s)\frac{d\psi_s}{dt}. \tag{10.14b}$$

To find $(d\psi_s/dt)$, we use Gauss's law in the form

$$C_{ox}(V_G - V_{FB} - \psi_s) = -Q_s(\psi_s) - Q_{it}(\psi_s) \tag{10.15}$$

where V_G, V_{FB}, and ψ_s all are time varying. Taking time derivatives of (10.15), we find

$$\frac{d\psi_s}{dt} = \frac{d(V_G - V_{FB})}{dt} \frac{C_{ox}}{C_{ox} + C_s + C_{it}}. \tag{10.16}$$

Substituting (10.14) in (10.13) and using (10.16), we obtain I_G

$$I_G = C_{LF}(V_G) \frac{d(V_G - V_{FB})}{dt} \tag{10.17}$$

where, as usual, the low frequency MOS capacitance per unit area C_{LF} is defined by

$$C_{LF} \equiv C_{ox} \frac{C_s + C_{it}}{C_{ox} + C_s + C_{it}}. \tag{10.18}$$

As shown in Fig. 10.6, to a fair approximation at elevated temperatures

$$C_{LF} \approx C_{ox}. \tag{10.19}$$

Also, we use a linear voltage ramp, so

$$\frac{dV_G}{dt} = \alpha \tag{10.20}$$

where α is the constant voltage sweep rate. If we substitute (10.19) and (10.20) in (10.17), I_G becomes

$$I_G \approx C_{ox} \left(\alpha - \frac{dV_{FB}}{dt}\right). \tag{10.21}$$

If we take αC_{ox} to the left of (10.21) and integrate from a gate bias $(-V_G)$ to a gate bias V_G, (10.21) becomes

$$\begin{aligned} \int_{-V_G}^{V_G} dV_G[I_G - \alpha C_{ox}] &= -C_{ox} \int_{-V_G}^{V_G} dV_G \left(\frac{dV_{FB}}{dt}\right) \\ &= -\alpha C_{ox} \int_{t(-V_G)}^{t(V_G)} dt \left(\frac{dV_{FB}}{dt}\right) \\ &= -\alpha C_{ox}\{V_{FB}[t(V_G)] - V_{FB}[t(-V_G)]\}. \end{aligned} \tag{10.22}$$

The integral over gate bias on the right of (10.22) is made into an integral over time using (10.20). The integral on the left of (10.22) is the area between the I_G-V_G curve and the straight line $I_G = \alpha C_{ox}$, representing the gate current of the MOS capacitor when no ions move. The right side of (10.22) can be evaluated by using (10.9). We suppose that at the time when the gate bias was $-V_G$, at time $t(-V_G)$, the mobile charge centroid was $\bar{x}(-V_G)$. Similarly, when the gate bias was V_G, at time $t(V_G)$, the centroid

was $\bar{x}(V_G)$. Then (10.9) provides

$$V_{FB}[t(V_G)] - V_{FB}[t(-V_G)] = \frac{qN_m}{\epsilon_{ox}} [\bar{x}(V_G) - \bar{x}(-V_G)] \tag{10.23}$$

where N_m is the mobile ion density per unit area. Therefore, (10.22) becomes

$$\int_{-V_G}^{V_G} dV_G[I_G - \alpha C_{ox}] = \alpha q N_m \left[\frac{\bar{x}(V_G)}{x_o} - \frac{\bar{x}(-V_G)}{x_o}\right]. \tag{10.24}$$

For sufficient biases and times, the centroids in (10.24) are bias and time independent. Consequently, (10.24) shows that for a linear voltage ramp at temperatures sufficient to make $C_{LF} \approx C_{ox}$, the area under an I_G-V_G curve in excess of $I_G = \alpha C_{ox}$ is proportional to the mobile ion density per unit area.

To use (10.24) to estimate N_m, one ordinarily assumes $\bar{x}(V_G) = x_o$ (all positive ions drifted to the silicon surface) and $\bar{x}(-V_G) = 0$ (all positive ions drifted to the gate). These assumptions probably are valid to within 100 Å.

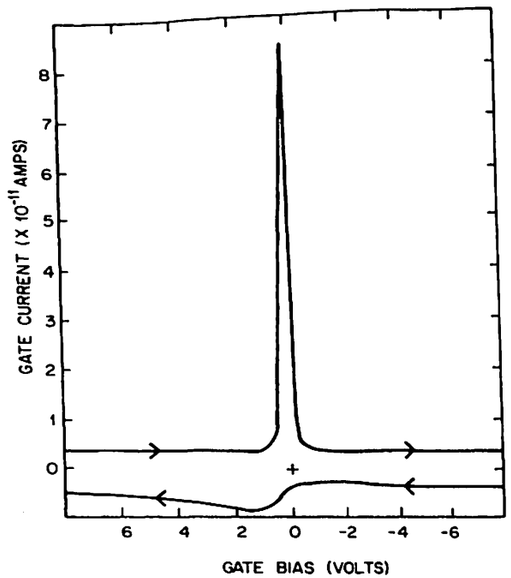


Fig. 10.7 Gate current as a function of gate bias measured at 202°C on a heavity contaminated oxide. The substrate of the MOS capacitor was an *n*-type silicon epitaxial layer 5 μm thick with a resistivity of 1 Ω-cm and a (111) orientation. The oxide was grown in an oxygen plasma to a thickness of 1000 Å. The gate was Cr–Au, with an area of 1.3×10^{-3} cm^2. Voltage sweep rate was 77.5mV/sec. After Kuhn and Silversmith.[13] Reprinted by permission of the publisher, The Electrochemical Society.

Figure 10.7 shows an I_G-V_G curve measured at 202°C on an oxide heavily contaminated with sodium. The sweep from positive to negative gate bias (left to right) where the ions drift from the Si–SiO_2 interface to the metal-oxide interface is shown in the upper curve. The area under this curve corresponds to 2.5×10^{12} ions/cm^2. The sweep from negative to positive gate bias (right to left) where the ions drift from the metal-oxide interface to the Si–SiO_2 interface is shown in the lower curve.

The most striking feature of these curves is the strong asymmetry of the I_G-V_G characteristics in the two sweep directions. Reasons for this asymmetry are described briefly in Section (e) below. The areas under both peaks in Fig. 10.7 are identical so that all ions transported toward the metal-oxide interface can be returned to the Si–SiO_2 interface. The areas under the peaks are independent of voltage sweep rate if the range of bias variation is sufficient and if the time spent at extreme bias is long enough to achieve a time independent pile up at each interface.

Our derivation of (10.24) does not require a detailed understanding of the time dependence of V_{FB}. This is fortunate because ionic transport alone is not sufficient to model V_{FB}. As discussed later, mobile ion motion is dominated by interfacial barrier effects.

(d) *Other Measurement Methods*

The electrical methods described in Sections (b) and (c) above are the most sensitive of the methods used to detect sodium. However, the electrical methods do not measure the neutral sodium profile or the sodium in chemical reagents. Analytic methods used for sodium profiling are the radioactive tracer method,[4,14] neutron activation analysis[1,3,14] described briefly in Section 11.4.1(a), flame photometry,[15,16] and secondary ion mass spectroscopy[17] (SIMS).

In SIMS the mass spectroscopic analysis of secondary ions proceeds simultaneously with the sputter etching of the oxide by primary ions. However, sodium becomes *mobile*[17] in the electric field created by the primary ions, so that only the total amount of sodium and not its profile can be measured. Secondary ion mass spectroscopy is the only analytical method whose limit of sensitivity is comparable to the electrical methods, but as work with this method has been confined to small sample areas, its full sensitivity has yet to be exploited. This method requires complex instrumentation and is destructive to the sample. Therefore, it is not used as widely as the other analytical methods.

Among the methods that require a large sample area, the radioactive tracer method combines sensitivity with simplicity. Neutron activation analysis is possible because sodium incorporated "in process" can be activated and detected. However, this method requires access to a nuclear pile.

The least sensitive but simplest analytical method is flame photometry, which is used in monitoring sodium contamination of chemical reagents

Table 10.1 Commonly Used Techniques for Detection of Sodium in SiO_2 Films.[a]

Method	Limiting Factors	Sensitivity: Na Atoms	Sensitivity: Na Ions	Ref.
Neutron activation analysis Isotope, ^{24}Na; half-life, 15 hr; γ-ray emission at 1.4 MeV and 2.75 MeV	1. Neutron flux 2. Background radiation	10^{11}–10^{12}		1, 3, 14
Radioactive tracer Isotope, ^{22}Na; half-life, 2.6 yr, γ-ray emission at 0.51 MeV	1. $^{22}Na/^{23}Na < 10^{-2}$ (usually) 2. Background radiation	10^{11}–5×10^{11}		4, 14
Flame photometry Sodium emission 5890–5896 Å	1. Minimum volume of liquid sample required 2. Sodium contamination of HF solvent	8×10^{11}		15, 16
Secondary ion mass spectroscopy ^{23}Na; O_2^+ primary beam	1. Ionization efficiency of Na 2. Usually small sample area ($\sim 10^{-5}$ cm^{-2})	$\sim 10^6$		17
C-V method			10^{10} cm^{-2}	8
Triangular voltage sweep			10^9 cm^{-2}	13

[a] After Kriegler.[18]

used in integrated circuit processing. In this method sodium atoms in solution are atomized into a flame* that excites the characteristic sodium emission line. The intensity of the emitted light is directly proportional to the number of sodium atoms excited by the flame and thus to the sodium concentration in the solution. The sodium emission line is resolved by a monochrometer and detected by a photomultiplier tube. The instrument is calibrated with standard solutions containing a known amount of sodium to establish a relationship between the emission intensity of a known spectral line and the concentration of sodium in solution. Similar intensity measurements are made on the sample solution and converted into concentration by using the calibration curve.

These analytical methods are summarized in Table 10.1, which also shows, for comparison, the two electrical methods. The sensitivity of the C-V method in Table 10.1 is based on $x_o = 1000$ Å and a detectable voltage shift of 0.1 V assuming no change in interface trap level density. Although a smaller voltage shift could be measured, the extra effort required would nullify the simplicity of this method; thus the TVS method might as well be used.

(e) *Kinetics of Ion Drift*

Hofstein[4] was the first to observe that the bias-temperature drift of sodium ions from the silicon to the metal was much faster than the drift from the metal to the silicon. The mobility of sodium ions in the silica bulk has been estimated to be about 4×10^{-12} cm^2/V-sec at room temperature.[18,19] In the MOS capacitor, the rate limiting step appears to be emission over an energy barrier from the interface. The energy barrier appears to be higher at the metal-oxide interface than at the Si–SiO_2 interface, accounting for the asymmetry observed in Fig. 10.7.

Ionic transport through the oxide can be studied without the complications of interfacial detrapping.[18,19] Such studies are based on observation of the time derivative of flatband voltage, as observed using the current given by (10.21), for example. As explained earlier, as ions detrap and drift across the oxide, the current rises until all the available ions detrap, and then saturates while all the ions are in transit across the oxide. When the ions begin to collect at the opposite electrode, the current drops toward the baseline value again. The time period between the current's initial rise and its drop is the time for the leading ions to traverse the oxide.

On the basis of crude assumptions of constant drift velocity and negligible influence of the ionic charges themselves on the field, Stagg[19] found that his transit time measurements implied an activated bulk mobility in dry oxides of the form $\mu = \mu_o \exp(-qE_A/kT)$. Here, for sodium, $\mu_o \approx$ 1 cm^2/V sec and $E_A = 0.66$ eV and for potassium, $\mu_o \approx 0.03$ cm^2/V sec and $E_A = 1.1$ eV.*

*An oxyhydrogen flame is commonly used.

Sodium ion emission from the electrodes has been studied using two methods:[10,20,21] the isothermal transient ionic current (ITIC) method and the thermally stimulated ionic current (TSIC) method. In the ITIC method ionic displacement current is measured as a function of time while the MOS capacitor is held at an elevated temperature with a field applied across the oxide. In the TSIC method ionic displacement current is measured as a function of temperature at a given applied oxide field. These two methods are useful mainly for research studies of ion emission from the silicon and the gate electrodes. The understanding of ion emission from the electrodes gained from these methods is at an early stage. These methods are not discussed further.

10.3 INTERNAL PHOTOEMISSION

Internal photoemission is used to determine the energy barrier between silicon and the conduction band of SiO_2† and the energy barrier between various gate metals and the SiO_2 conduction band. The barrier heights between the gate metal and SiO_2 and between silicon and SiO_2 can be used to estimate the work function difference between the gate and the silicon.

This measurement, described in Sections 11.4.2 and 11.4.3, provides a check of the directly measured work function differences described in Section 10.4. The measurement of the SiO_2 bandgap energy is described in Section 11.6.5(a).

In this section the use of internal photoemission to determine electrode energy barriers is described. Of the many approaches to the use of photoemission,[22] we restrict ourselves to those used on the MOS system. Sections 11.2, 11.3, and 11.4 discuss the use of avalanche injection and internal photoemission to study trapping phenomena in SiO_2.

No appreciable current flows between the metal and the silicon in the MOS system when fields below the dielectric breakdown strength of the oxide are applied because (1) SiO_2 is an insulator with a wide bandgap (8.8 eV)[23] (therefore, free hole and electron densities intrinsic to SiO_2 are negligible at all temperatures of interest) and (2) electrons (or holes) in each electrode must overcome large energy barriers (several electron volts) to enter the SiO_2 as free carriers (therefore, injection of holes and electrons into the SiO_2 is difficult, limiting current flow).

*These measurements were made using a voltage step method rather than the TVS method. When a voltage step is used, the oxide field in the absence of ionic charges is constant, rather than linearly increasing as with the ramp method.

†The term "conduction band" is applied loosely to SiO_2. In an energy band an electron is not localized at a particular lattice site but is free to move about the entire crystal. The band assumption may not apply to thermally grown SiO_2.