Tunneling in thin MOS structures

J. Maserjian

Citation: Journal of Vacuum Science & Technology **11**, 996 (1974); doi: 10.1116/1.1318719
View online: http://dx.doi.org/10.1116/1.1318719
View Table of Contents: http://scitation.aip.org/content/avs/journal/jvst/11/6?ver=pdfcov
Published by the AVS: Science & Technology of Materials, Interfaces, and Processing

Tunneling in thin MOS structures*

J. Maserjian

Jet Propulsion Laboratory, California Institute of Technology, Pasadena, California 91103

(Received 5 August 1974)

Recent results on tunneling in thin MOS structures are described. Thermally grown SiO_2 films in the thickness range of 22–40 Å have been shown to be effectively uniform on an atomic scale and exhibit an extremely abrupt oxide–silicon interface. Resonant reflections are observed at this interface for Fowler–Nordheim tunneling and are shown to agree with the exact theory for a trapezoidal barrier. Tunneling at lower fields is consistent with elastic tunneling into the silicon direct conduction band and, at still lower fields, inelastic tunneling into the indirect conduction band. Approximate dispersion relations ($E-k$, $E-\kappa$) are obtained over portions of the silicon-dioxide energy gap and conduction band.

INTRODUCTION

Recent results[1,2] involving tunneling through thin MOS structures indicate a chemical and electronic structure that is surprisingly well defined. Figure 1 shows the energy-band model which we have used for describing $Au/Cr/SiO_2/Si$ structures with oxide films in the thickness range of 20–40 Å. It is common to view such simple, one-dimensional, band models as gross idealizations when applied to extremely thin amorphous oxide layers. One can argue that the band structure is not defined in a few molecular layers, especially in amorphous material; that the chemical interface is smeared out; that the oxide is seriously nonuniform on this scale of thickness; and that pinholes can dominate the electrical characteristics. However, we find that contrary to such arguments, the MOS structures behave very much as predicted by the ideal one-dimensional model shown in Fig. 1.

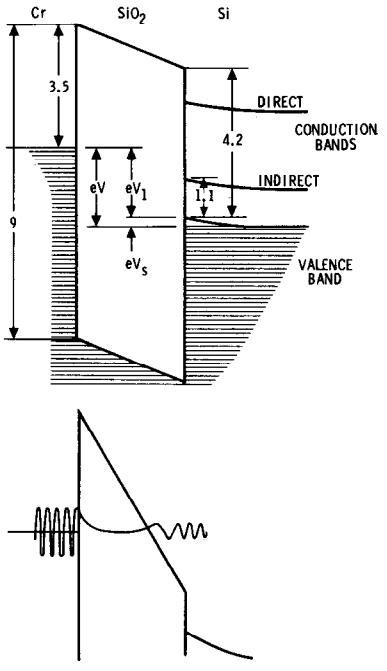

FIG. 1. Band model for thin MOS structure. Lower diagram represents Fowler–Nordheim tunneling.

The tunneling characteristics can be interpreted in terms of four voltage regions with the help of Fig. 1: (I) At low fields, when the voltage across the oxide V_1 is less than 1.1 V corresponding to the silicon band gap, inelastic tunneling occurs into surface states.[3] (II) When V_1 exceeds 1.1 V, inelastic tunneling continues into the indirect conduction band of silicon (a large mismatch in transverse momenta must be accounted for) and the current increases rapidly with voltage.[4] (III) When V_1 exceeds about 3.4 V, elastic tunneling occurs into the direct conduction band of silicon[5] and the current converges on the theoretical dependence predicted for ideal metal–insulator–metal structures.[1] (IV) At high fields, when V_1 exceeds 4.2 V, corresponding to the barrier at the oxide–silicon interface, one observes Fowler–Nordheim tunneling into the oxide conduction band.[6]

In analyzing our results we were confronted with an unknown current–voltage dependence for inelastic tunneling corresponding to regions I and II, and an unknown dispersion relation $\kappa(E)$ for describing the attenuation of the electron waves in the oxide band gap. Lewicki and others[7–10] have shown that $\kappa(E)$ can be measured for MIM structures having trapezoidal energy barriers by the dependence of tunneling current on either voltage or insulator thickness. Results by both methods have been shown to be in good agreement with one another and gave dispersion relations for AlN and GaSe that could be approximately described by a Franz form.[11] In the case of MOS structures, the process of inelastic tunneling into the indirect conduction band introduces an unknown factor into the voltage dependence over much of the range of data. Fortunately, this factor is independent of thickness and we are still able to rely on Lewicki's thickness-dependent method for obtaining $\kappa(E)$ over this range. This approach

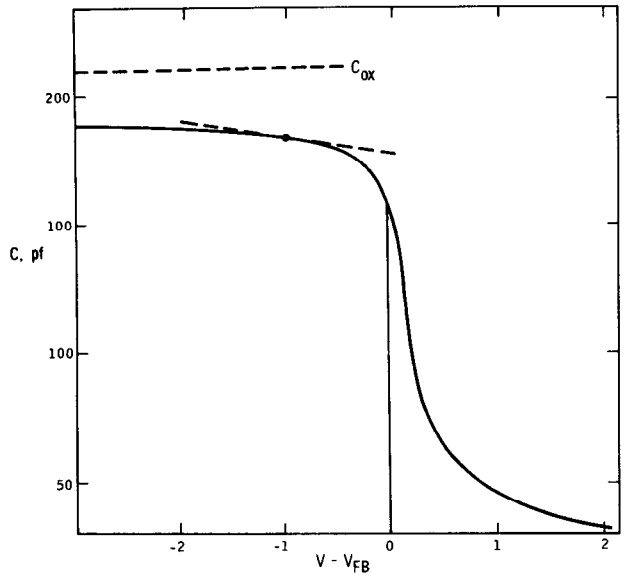

FIG. 2. C–V curve for thin MOS structure.

requires very precise measurements of oxide thickness which is described in the following section.

OXIDE THICKNESS

An oxide with a range of thicknesses was grown on p^+ silicon wafers by placing each wafer in a temperature gradient (700–800°C) in dry O_2 for 15 min. CrAu was evaporated on both sides and patterned on one side into an MOS array of 160×160 μm squares. Ellipsometer measurements provided a guide to the average oxide thickness before metallization but could not accurately resolve the small thickness gradient along the wafer. Also, the chromium metallization appeared to reduce the oxide film by about 4 Å. C–V measurements provided the accuracy required in the thickness, but only by taking into account the accumulation layer in the silicon.

Figure 2 shows that, for a 40 Å oxide film, the capacitance fails to saturate at the oxide capacitance C_{ox} even at high fields corresponding to degenerate accumulation.[12] However, the contribution of the surface capacitance C_s for a degenerate accumulation layer can be expressed in the following convenient form:[2]

$$C_s - (a^2/2)^{\frac{1}{6}} C^{\frac{1}{3}} (dC/dV)^{-1/6}, \quad (1)$$

where C is the measured capacitance in the voltage range of degenerate accumulation ($|V - V_{FB}| \gg kT/e$), V is the total applied voltage, and a^2 is a constant depending on the density of states effective mass, m_c or m_v, for the appropriate band edge [see Eq. (8), Ref. 2]. The values of m_c and m_v, which apply to a narrow degenerate accumulation layer, were determined experimentally using Eq. (1) and the relation

$$1/C = 1/C_{ox} - 1/C_s. \quad (2)$$

By plotting measured values of $1/C$ vs $C^{-\frac{1}{3}}(dC/dV)^{1/6}$, one obtains an intercept $1/C_{ox}$ and a slope $(2/a^2)^{\frac{1}{6}}$. Figure 3 illustrates an example of one such plot, which was repeated for several wafers of different doping and orientation. The results gave $m_c/m = 0.056 \pm 0.004$ and $m_v/m = 0.196 \pm 0.004$, independent of orientation or doping concentration. Therefore, the constant a^2 is known and one can determine C_{ox} and the flat-band voltage V_{FB} from measurements of C and dC/dV at only one point in the degenerate accumulation region, such as is indicated in Fig. 2. Actually, a few points were measured and the average values of C_{ox} and V_{FB} used. Having C_{ox} and V_{FB}, the analysis also provides V_s (or V_1).

The random error in C_{ox} by this method is estimated to be about ±0.5%. In determining the oxide thickness (i.e., $w = \epsilon_{ox} A/C_{ox}$), the uncertainty in the measurement of area A, even with good control of patterning, introduces an additional random error of about ±0.5% so that the total random error of thickness is about ±1%. The systematic error due to uncertainties in the absolute magnitudes of area, dielectric constant, and m_v is greater but does not seriously affect the analysis which is based primarily on relative oxide thickness.

Clearly, the capacitance measurement gives the average oxide thickness of the MOS capacitor. A smooth variation of this measured thickness (e.g., 23–34 Å) was observed across wafers grown in a thermal gradient. Of particular interest is the fact that the tunneling current measured at constant voltage for each adjacent MOS square across the wafer varied in a perfectly smooth and continuous manner. The log of the current is nearly proportional to the oxide thickness (see next section) and, when plotted versus position along the wafer, exhibits a variation consistent with that obtained from the capacitance measurements.[13] The average thickness change between adjacent squares is of the order of 0.2 Å and the effective tunneling thickness follows these small changes. Since the tunneling current increases about one order of magnitude for every 2.5 Å decrease in thickness, it clearly reflects these small changes in average thickness. However, the cur-

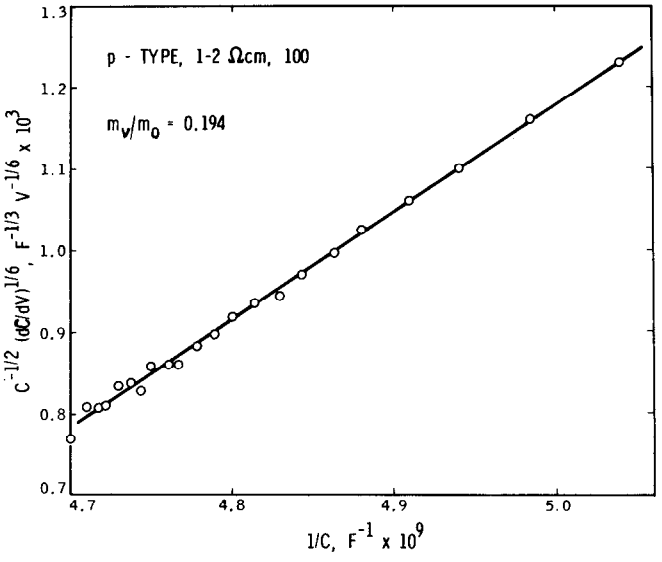

FIG. 3. C–V dependence in degenerate accumulation.

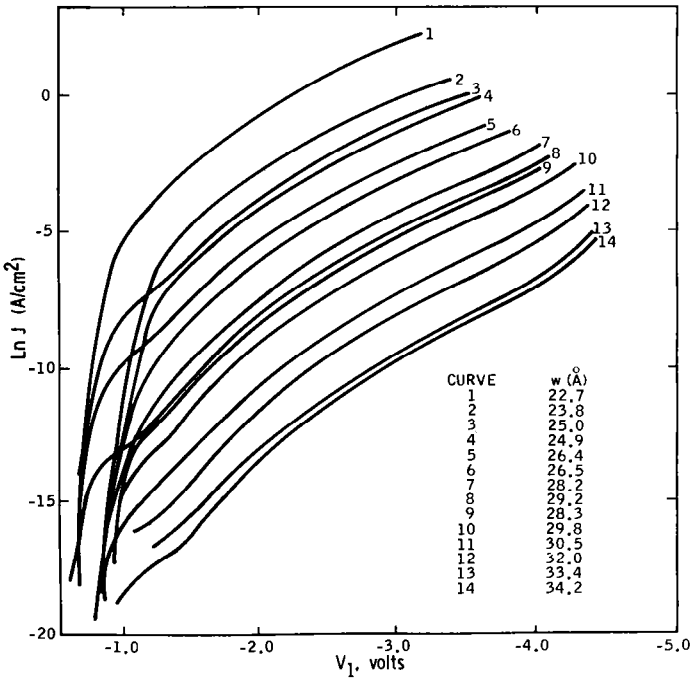

FIG. 4. Tunneling characteristics for different oxide thicknesses. Results were obtained from two wafers with different surface treatment.

rent should also be sensitive to small random deviations from a uniform oxide film in a single MOS device; that is, the average current density will be weighted heavily at the thinnest regions. Since the current follows the average thickness in such a consistent manner, we are inclined to conclude that the oxide film must be very uniform or vary at most only slightly in a regular pattern. The resolution limit of lateral distances over which thickness variations can be detected is the lateral dimension of the tunneling wave packet, which extends over many atomic sites.[14]

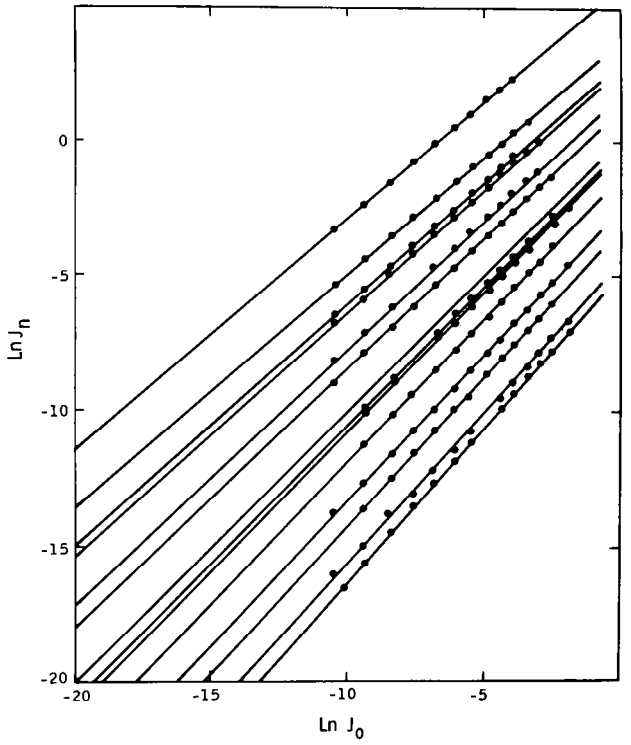

FIG. 5. Interdependence of tunneling current for different oxide thickness.

The results raise interesting questions about the nature of the arrangement of chemical bonds at the interfaces during growth. However, for the purpose of interpreting tunneling, we are confident that the oxide films can be treated as perfectly uniform in thickness with no evidence of pinholes.

TUNNELING CHARACTERISTICS (REGIONS I, II, AND III)

Figure 4 illustrates the dc tunneling characteristics obtained at liquid-nitrogen temperatures from two wafers with overlapping ranges of oxide thickness. A small systematic error in thickness measurements between wafers is present (e.g., small change in area); otherwise, the relative values of thickness indicated are accurate to ± 0.3 Å. The two wafers can easily be distinguished at low voltages, with one exhibiting a significantly larger excess current below 1.1 V due to tunneling into surface states (region I). Both wafers were prepared identically except that the one showing less surface-state tunneling had received a mild preoxidation and strip prior to the final oxide growth. Otherwise, the two sets of curves converge on the same dependence at higher voltages (regions II and III). The characteristics of regions II and III were reproducible from wafer to wafer, including results obtained from wafers prepared and measured at both Chalmers University, Sweden, and JPL. Room-temperature measurements also gave similar characteristics but at slightly high currents, as predicted by tunneling theory. However, breakdown occurred at lower fields at room temperature.

The interdependence of the tunneling currents among devices of different oxide thickness is shown in Fig. 5, where the log current values for different-thickness MOS devices are plotted against the log current of one reference MOS device, all compared at the same values of V_1. For the case of elastic tunneling through a trapezoidal barrier, the points should fall on a straight line with slopes proportional to thickness.[7] However, inelastic tunneling into the indirect band (region II) extends over most of the range of data and has the effect of reducing the slopes by an unknown amount. Therefore, the plots serve only as a qualitative check on the consistency of the data with a trapezoidal barrier.

The effect of the indirect conduction band is shown in Fig. 6. The solid curves compare two MOS devices of nearly the same oxide thickness obtained from wafers of different surface orientations. The small but reproducible difference over the range of inelastic tunneling (region II) is consistent with the different transverse momenta in the indirect conduction band for the two orientations. At higher fields the curves converge with

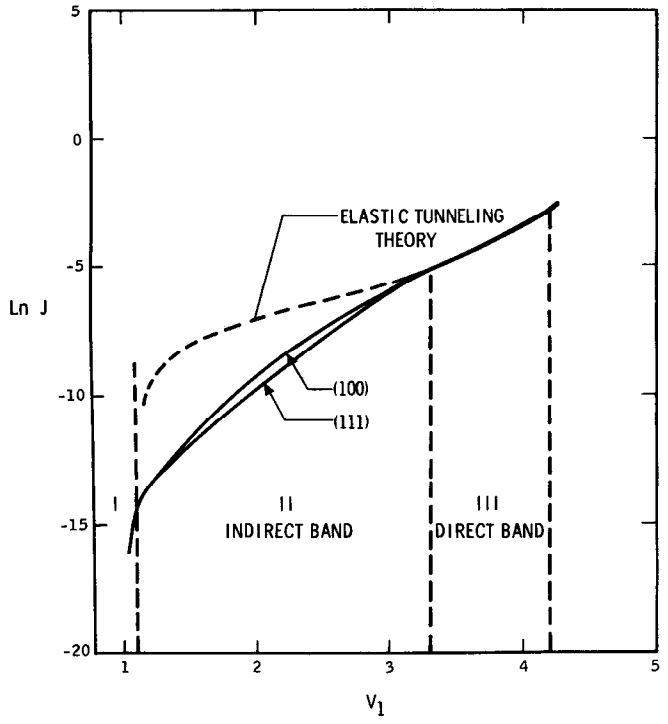

FIG. 6. Effect of inelastic tunneling into the indirect conduction band of silicon.

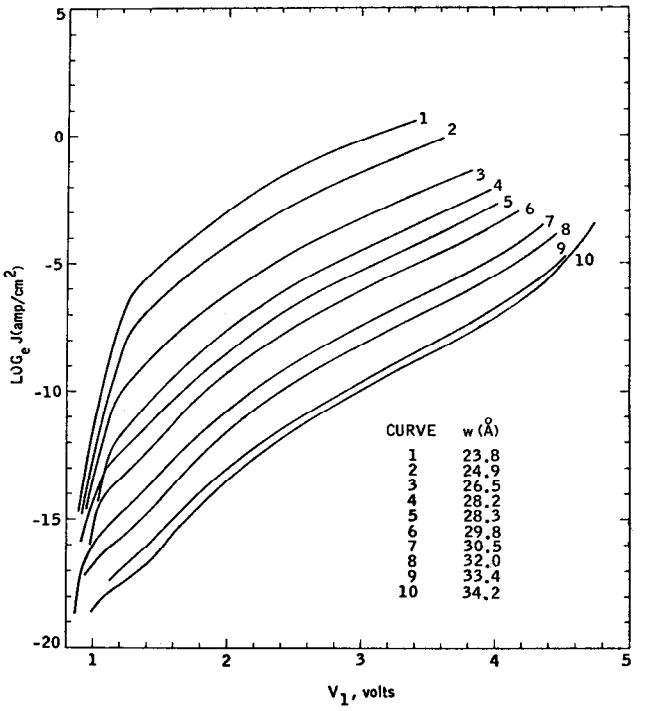

FIG. 7. Tunneling characteristics used for analysis.

the elastic tunneling theory when the dispersion relation $\kappa(E)$ obtained from this work is used (see next section). The effect of the indirect band is to reduce the tunneling current approximately exponentially with the voltage difference below the direct band edge. There is little doubt that this reduction is real because the elastic tunneling current cannot account for it even if we allow for some uncertainty in $\kappa(E)$. Furthermore, the effects of inelastic tunneling via defect states tend to be ruled out by the examples of the contributions of different surface-state densities in Fig. 4. Such effects were seen to be swamped out in region II.

DISPERSION RELATION $\kappa(E)$

For the purpose of determining $\kappa(E)$ it is a sufficiently good approximation to use the following simple expression for tunneling in regions II and III ($V_1 \leq 4.2$ V)[8]:

$$J \simeq B(V_1) w^{-2} \exp[-2\bar{\kappa}(V_1) w], \tag{3}$$

where $B(V_1)$ includes an unknown voltage-dependent factor in region II and $\bar{\kappa}$ is the average value of κ defined by

$$\bar{\kappa} = (1/w) \int_0^w \kappa dx. \tag{4}$$

$\bar{\kappa}$ becomes independent of thickness for a trapezoidal barrier[8] and specifically for the barrier in Fig. 1,

$$\bar{\kappa} = \frac{1}{eV_1 - 0.7} \int_{4.2-eV_1}^{3.5} \kappa(\mathcal{E}) d\mathcal{E}, \tag{5}$$

where $\mathcal{E}$ is the energy difference below the oxide conduction band. Therefore, we have the following linear dependence on w:

$$\ln(Jw^2) = \ln B(V_1) - 2\bar{\kappa}(V_1) w. \tag{6}$$

Using the tunneling characteristics shown in Fig. 7, obtained from a range of thicknesses over a single wafer, the values of $\ln(Jw^2)$ can be plotted against w for fixed values of V_1. Figure 8 shows a few curves plotted according to Eq. (6). The scatter is due primarily to the

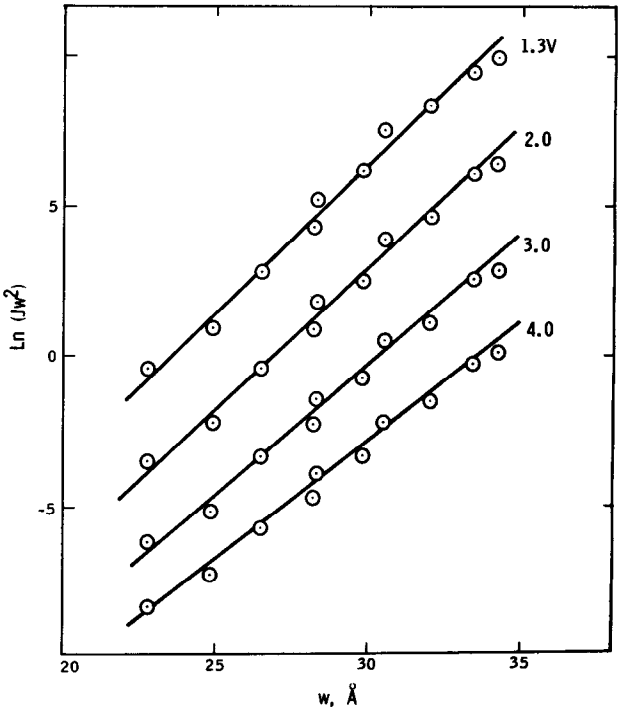

FIG. 8. Dependence of tunnel current on oxide thickness.

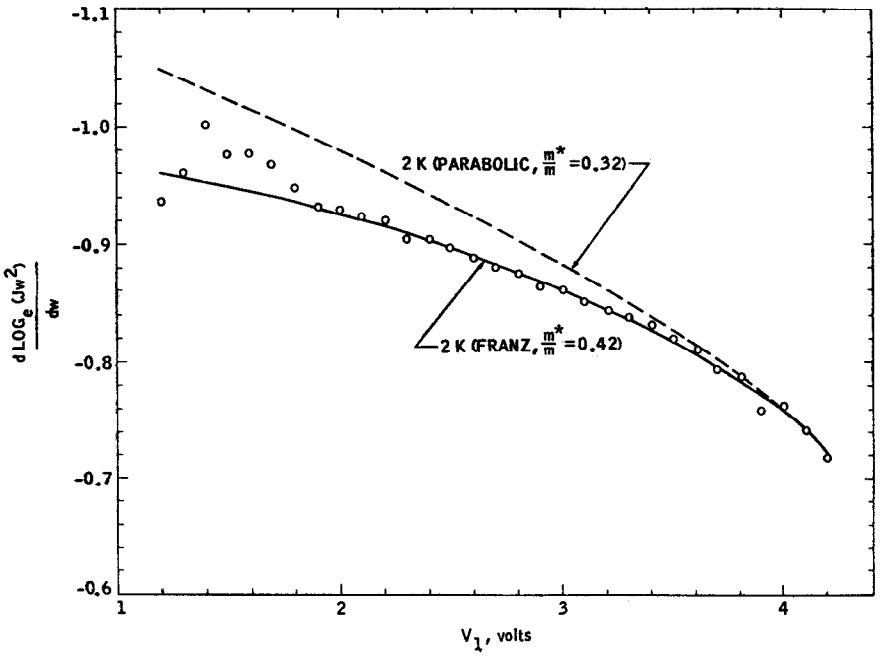

FIG. 9. Comparison of tunneling data with dispersion relations.

random error in the measured thickness (± 0.3 Å). The slopes were actually obtained by a linear regression fit (correlation coefficient $\simeq 0.99$).

Figure 9 shows the slopes so obtained plotted against V_1. Since, from Eq. (6), $d\ln(Jw^2)/dw = 2\bar{\kappa}$, one can make direct comparisons with the values of $2\bar{\kappa}$ calculated from Eq. (5), using both a parabolic and Franz relation; that is,

$$\kappa(\mathcal{E}) = (2m^*/\hbar^2)^{\frac{1}{2}}\mathcal{E}^{\frac{1}{2}} \quad \text{(parabolic)}, \tag{7}$$

$$\kappa(\mathcal{E}) = (2m^*/\hbar^2)^{\frac{1}{2}}\mathcal{E}^{\frac{1}{2}}(1-\mathcal{E}/9)^{\frac{1}{2}} \quad \text{(Franz)}. \tag{8}$$

Since the value of $2\bar{\kappa}$ at 4.2 V corresponds to the constant for Fowler–Nordheim tunneling (see next section), there is a common point of comparison with Lenzlinger and Snow's results[6] for thick oxides.

The Franz dependence clearly gives a better fit to the data in Fig. 9, and also explains the decreasing values of m^* obtained by Lenzlinger and Snow for increasing barrier heights at the metal interface, when a parabolic relation is assumed. The agreement, at the onset of Fowler–Nordheim tunneling, with previous results on thick oxides supports our belief that basically the same MOS structure is appropriate for these thin oxides. The following section on Fowler–Nordheim tunneling gives more convincing evidence.

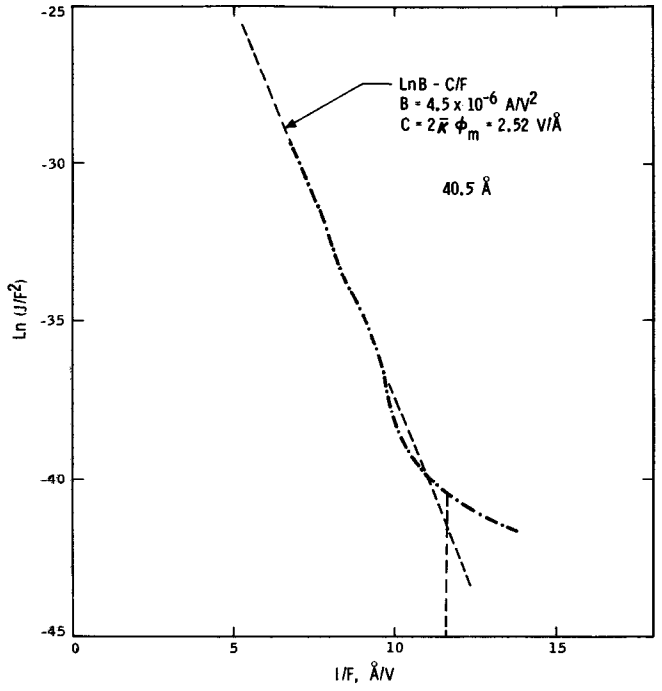

FIG. 10. Fowler–Nordheim plot showing weak oscillations.

FOWLER–NORDHEIM TUNNELING (REGION IV)

The well-known relation for Fowler–Nordheim tunneling[6] based on the WKB approximation can be expressed by

$$J = B_0 F^2 \exp(-2\bar{\kappa}\phi_m/eF), \tag{9}$$

where ϕ_m is the metal barrier height, equal to 3.5 eV for our case, F is the actual oxide field $(V_1 - 0.7)/w$, B_0 is a constant equal to 6.5×10^{-7} A/V^2 (value obtained by numerical integration), and $\bar{\kappa}$ [see, Eq. (5)] becomes, for $V_1 \geq 4.2$ V,

$$\bar{\kappa} = \frac{1}{3.5}\int_0^{3.5} \kappa(\mathcal{E})\,d\mathcal{E}. \tag{10}$$

Measurements of Fowler–Nordheim current were obtained from thin MOS structures using 10 μsec pulses and liquid-nitrogen temperatures. The results for a 40.5 Å device are plotted in the conventional way in Fig. 10, which should be a straight line according to the WKB theory of Eq. (9). The data shows weak oscillations about a straight line, the line having a slope $2\bar{\kappa}\phi_m/e$ equal to the value obtained from the previous lower field results and a constant B about seven times larger than the value of B_0 from the WKB theory.

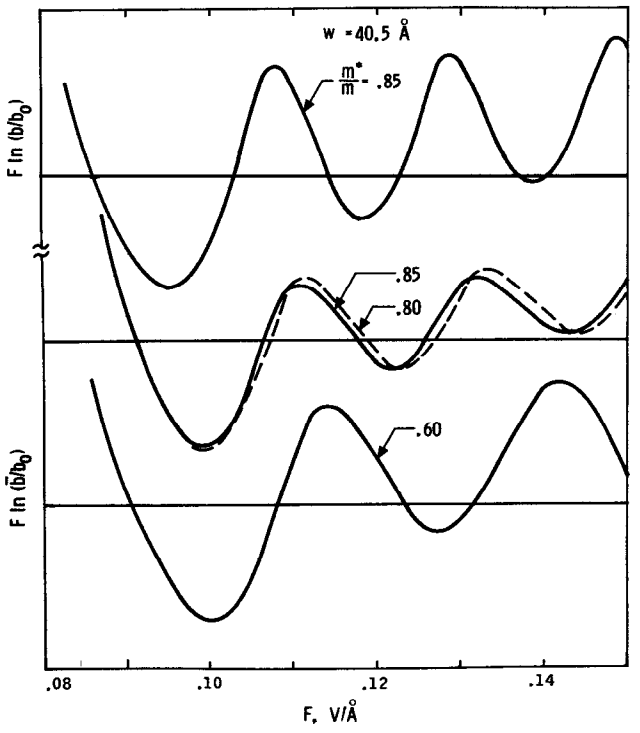

FIG. 11. Theoretical plots of oscillatory term in Fowler–Nordheim tunneling. Upper curve is for monoenergetic electron; lower curves are for actual energy distribution.

Gundlach and others[15,16] have predicted such oscillations using an exact solution of the tunneling transmission coefficient through a trapezoidal barrier. The oscillations arise from interference due to electron-wave reflections at the oxide–silicon interface (see lower diagram in Fig. 1). The exact transmission coefficient for one electron [Eq. (18), Ref. 15] reduces in the range of interest to the simpler form

$$D \sim b \exp\left(-2\int \kappa dx\right), \tag{11}$$

which is just the usual WKB approximation multipled by a modulating factor b given by

$$b=[\mathrm{Ai}^2(-\alpha x_1)+(\alpha/k)^2\,\mathrm{Ai}'^2(-\alpha x_1)]^{-1}, \tag{12}$$

where

$$\alpha=(2\bar{m}^* eF/\hbar^2)^{\frac{1}{3}},$$
$$x_1=w-(\phi_m+\mathcal{E}')/eF,$$

k is the electron wave vector in the silicon direct conduction band, $\bar{m}^*$ is the average effective mass in the oxide conduction band, $\mathcal{E}'$ is the energy difference below the metal Fermi level, and Ai, Ai′ are the Airy function and its derivative.

Since the factor b was derived for a monoenergetic electron, we must include it in the integral over energy for determining the tunnel current. This procedure is equivalent to evaluating the weighted average $\bar{b}$

$$\bar{b}=\frac{2\kappa}{eF}\int_0^{\mathcal{E}_{max}} b\exp(-2\bar{\kappa}\mathcal{E}'/eF)d\mathcal{E}'. \tag{13}$$

The Fowler–Nordheim current then becomes

$$J=\bar{b}B_0F^2\exp(-2\kappa\phi_m/eF). \tag{14}$$

For the purpose of comparison with the experimental results, it is useful to rewrite Eq. (14) into the following form:

$$F\ln(J/b_0B_0F^2)=F\ln(\bar{b}/b_0)-2\bar{\kappa}\phi_m/e, \tag{15}$$

where b_0 is introduced on both sides in order to normalize the right-hand expression by setting b_0 equal to $b(\mathcal{E}'=0)$ at the onset of Fowler–Nordheim tunneling. The right-hand term is then the usual Fowler–Nordheim constant $-2\bar{\kappa}\phi_m/e$ with a small term $F\ln(\bar{b}/b_0)$ added which oscillates around it, and the left-hand term becomes equivalent to the remainder of the WKB expression with b_0B_0 replacing B_0. It is interesting to note that $b_0\simeq 7$ and that $b_0B_0\simeq B$ in Fig. 10; however, this agreement may be somewhat coincidental since we were not rigouous about the constant (the order of unity) in Eq. (11).

In Fig. 11, the oscillatory term for monoenergetic electrons $F\ln(b/b_0)$ is plotted in the upper curve and compared with the term $F\ln(\bar{b}/b_0)$ plotted below, corresponding to the actual energy distribution of the electron current. The effect of the tunneling energy distribution on the phase and damping of the oscillations compared with the monoenergetic case is clearly important and cannot be neglected. Since all parameters are at least approximately known except for $\bar{m}^*$, the effect of varying $\bar{m}^*$ is shown in the lower three curves.

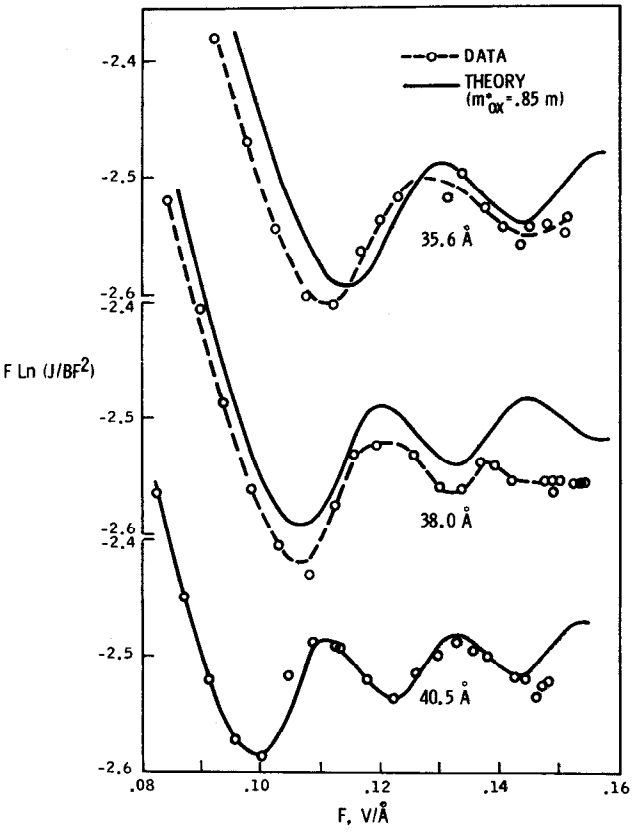

FIG. 12. Comparison of observed vs theoretically predicted oscillations.

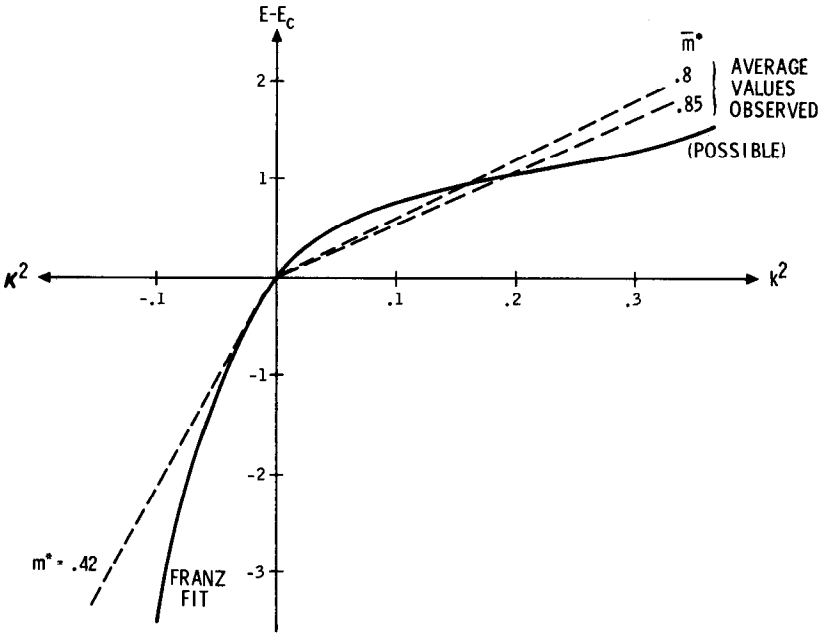

FIG. 13. Approximate dispersion relations $k(E)$ and $\kappa(E)$ for SiO_2.

In fitting the experimental data, $\bar{m}^*$ is constrained fairly tightly as it can be seen to effect phase, period, and amplitude (compare $\bar{m}^*=0.6$ and 0.85). However, it becomes difficult to resolve the difference between 0.8 and 0.85 as shown.

In Fig. 12, the experimental data and theory are compared, using $\bar{m}^*=0.85$. The agreement with the data is remarkably good, especially since we made no attempt to adjust any parameters other than $\bar{m}^*$. Accumulative errors in w, ϕ_m, ϕ_{si}, V_s, k, etc., can account for at least part of the slight differences observed. The value of B used for the experimental plot is the experimental constant obtained in Fig. 10 which, as mentioned above, is equal to b_0B_0 as required by Eq. (15). The plots are compared directly on an absolute scale, using the value obtained for $2\bar{\kappa}\phi_m/e$ of 2.52.

The theory of Eq. (12) assumes a strictly parabolic E–κ dependence in the conduction band which is not likely to be true over the range of energies in the oxide conduction band involved (0–2.5 eV), and therefore $\bar{m}^*$ should represent some average value of m^*. In Fig. 12, a hypothetical E–k dependence is shown connected to the approximate Franz relation obtained for the band gap and plotted in a manner that is consistent with the values $\bar{m}^*/m=0.8$ and 0.85 that agree with the results.

We have neglected the role of image forces which would round the corners of the very narrow trapezoidal barriers.[17] Although it is not obvious just how important the effect of such rounding would be in changing the characteristics of the oscillatory behavior, we believe the effect would be quite significant. If this is the case and in view of the agreement with the trapezoidal band model, we would have to conclude that image forces do not apply to very narrow tunneling barriers. A possible explanation is that the wave packet formed in the oxide conduction band extends over a region which is large (~10 Å) compared with the major variation in image potential. We are presently investigating in detail the effect of image forces on the oscillatory behavior.

CONCLUSIONS

The tunneling characteristics of thin MOS structures are entirely consistent with the simple one-dimensional band model. That is, the oxide film can be considered uniform on an atomic scale and can be characterized by a trapezoidal barrier with an extremely abrupt silicon interface. The energy-dispersion relation for electrons in the oxide-band gap can be better described by a Franz relation rather than the commonly used parabolic relation. The effective mass which describes electrons traversing a range of energy in the oxide conduction band is about 0.85 times the free electron mass. Inelastic tunneling into the indirect band has been observed and is substantially lower than that predicted by elastic tunneling theory. The intrinsic tunneling process into the silicon band dominates surface-state or trap-related tunneling processes in these thin oxide films at moderate to high fields.

ACKNOWLEDGMENTS

The author wishes to acknowledge the contributions and active collaboration of G. Petersson, C. Svensson, and N. Zamani in the course of this work.

*This paper presents the results of one phase of research performed at the Jet Propulsion Laboratory, sponsored by the National Aeronautics and Space Administration under Contract NAS 7-100.

[1]J. Maserjian and G. Petersson, Appl. Phys. Lett. 25, 50 (1 July 1974).
[2]J. Maserjian, G. Petersson, and C. Svensson, Solid-State Electron. 17, 335 (1974).
[3]W. E. Dahlke, Appl. Phys. Lett. 10, 261 (1967).
[4]J. Shewchun, A. Waxman, and G. Warfield, Solid-State Electron. 10, 1165 (1967).

[5]M. L. Cohen and T. K. Bergstresser, Phys. Rev. **141**, 789 (1966).
[6]M. Lenzlinger and E. H. Snow, J. Appl. Phys. **40**, 278 (1969).
[7]G. Lewicki, Ph.D. thesis (California Institute of Technology, 1966) (unpublished).
[8]R. Stratton, G. Lewicki, and C. A. Mead, J. Phys. Chem. Solids **27**, 1599 (1966).
[9]G. Lewicki and C. A. Mead, J. Phys. Chem. Solids **29**, 1255 (1968).
[10]K. H. Gundlach and J. Kadlec, Solid-State Electron. **16**, 787 (1973).
[11]W. Franz, *Handbuch der Physik*, edited by S. Flugge (Springer, Berlin, 1956), Vol. XVIII, p. 155.
[12]K. H. Zaininger and G. Warfield, IEEE Trans. Electron Devices **ED-12**, 179 (1965).
[13]These findings are being investigated further and will be submitted later for publication.
[14]T. H. Di Stefano, private communication.
[15]K. H. Gundlach, Solid-State Electron. **9**, 949 (1966).
[16]M. E. Alferieff and C. B. Duke, J. Chem. Phys. **46**, 938 (1967).
[17]C. B. Duke, *Tunneling in Solids, Solid-State Physics Suppl. 10* (Academic Press, New York, 1969), p. 57.