Surface roughness at the Si(100)-SiO_2 interface

S. M. Goodnick, D. K. Ferry,* and C. W. Wilmsen
Department of Electrical Engineering, Colorado State University, Fort Collins, Colorado 80523

Z. Liliental,† D. Fathy,‡ and O. L. Krivanek
Center for Solid State Science, Arizona State University, Tempe, Arizona 85281
(Received 6 May 1985)

We have studied the statistical properties of random surface roughness at the Si-SiO_2 interface using high-resolution transmission electron microscopy (HRTEM). The spectral properties of the HRTEM roughness on normally prepared and intentionally roughened samples appears to be well characterized as a first-order autoregressive or Markovian process which corresponds to an exponential decay in the autocovariance function rather than the usual Gaussian approximation which has been widely used. Such an exponential decay is characterized by tails in the spectrum which are directly attributable to the discrete or steplike nature of the interface roughness which is restricted to occur on crystalline atomic sites. Using a simplified model, we have estimated the effect of projecting the two-dimensional interface roughness through the cross-section thickness to form the one-dimensional boundary studied here. For an isotropic medium, we find that the statistical character of the roughness is preserved during this transformation, but that the rms fluctuation of the roughness is attenuated so that the actual interface is rougher than indicated by the HRTEM technique. After correcting for such averaging, the parameters estimated from the HRTEM are more in agreement with the same parameters used to fit the surface-roughness-limited Hall mobility in metal-oxide-semiconductor field-effect transistor devices.

I. INTRODUCTION

The electronic properties of two-dimensional electrons in a Si inversion layer may be significantly affected by the oxide-semiconductor interface. Scattering of the electrons by fixed charges, surface states, and interface roughness are all believed to be major contributions to the channel mobility, especially at low temperatures.[1] Other phenomena, such as the observed valley splitting in Si(111) devices[2] and anomalous optical properties,[3] are sometimes associated with peculiarities of the Si-SiO_2 interface. However, little is actually known of the microscopic nature of the Si-SiO_2 interface or how it interacts with the electrons in the inversion layer. In the present paper we investigate the statistical properties of random fluctuations in the interface boundary studied using cross-sectional high-resolution transmission electron microscopy (HRTEM) and compare these results to the present statistical model assumed in the theoretical description of surface-roughness scattering.

In the past decade there have been numerous investigations into the chemistry and structure of the Si-SiO_2 interface.[4] Initial investigations of the interface using Auger-spectroscopy sputter profiling indicated interface widths in excess of 20 Å which were attributed to roughness at the interface.[5–7] However, due to sputter broadening effects, this value is somewhat greater than the actual width.[8] Investigations of the initial oxidation of Si using Auger spectroscopy without sputter profiling imply that the transition from Si to SiO_2 actually occurs over only one or two monolayers with SiO comprising the intermediate layer.[9] X-ray photoemission (XPS) studies also imply an abrupt interface and additionally identify bonding states corresponding to Si_2O_3 and Si_2O.[10] The apparent abruptness of the interface is also born out by Rutherford backscattering experiments.[11]

Studies of the Si-SiO_2 interface using cross-sectional HRTEM are in agreement with the above results.[12–15] From HRTEM the measured rms fluctuation of the interface is on the order of 2 Å, which is approximately one monolayer of interface width. In these studies the parameters corresponding to a Gaussian model for the roughness autocovariance function were estimated from the HRTEM cross sections and compared to similar parameters used to fit the low-temperature Hall mobility of *n*-channel metal-oxide-semiconductor field-effect transmitter (MOSFET) devices. The agreement in these studies was reasonably close, although the HRTEM roughness appeared to be consistently smaller than that required to fit the Hall mobility data. This is discussed in more detail in the present work. Studies using normal-incidence TEM (Ref. 16) have shown similar results for the interface width, and under certain oxidation conditions showed the presence of Si inclusions in the oxide near the interface. A tendency towards a rougher surface with increased oxidation rate was also observed. Investigations of the atomic step density at the Si-SiO_2 interface have been performed by Hahn *et al*, using low-energy-electron diffraction (LEED) spot profiling.[17–19] In their work a correlation was found between the atomic step density as estimated from the broadening of the LEED pattern and the low-temperature hole mobility in *p*-channel Si(111) devices. More recently, similar atomic step densities have been estimated for Si(100) and Si(111) using XPS and

comparing the ratio of Si^{2+} (ideally terminated surface) with Si^{1+} and Si^{3+} whose appearance is believed to correspond to steps in the interface.[20]

Based on the literature, the model one draws of the Si-SiO_2 interface is an almost atomically abrupt interface with the presence of steps and discontinuities which gave rise to suboxide species due to the differing bonding configuration at the points of discontinuity [for instance, a step of one monolayer on the Si(100) surface results in a Si^{1+} oxidation state]. For quantized channel electrons localized at the Si side of the Si-SiO_2 interface, steps and discontinuities in the interface itself will perturb the electronic energy levels through fringing of the surface electric field. In the usual models for surface-roughness scattering, one assumes an abrupt boundary between the Si and SiO_2 which randomly varies according to a quasicontinuous function $\Delta(\mathbf{r})$, where $\mathbf{r}$ represents the two-dimensional position vector in the plane of the interface.[1,21–23] In view of the present knowledge of the Si-SiO_2 interface, this assumption of an abrupt termination of the Si lattice is probably not unreasonable. Assuming that the surface potential may be expanded about the deformed interface as

$$V(x+\Delta(\mathbf{r}))\simeq V(z)+\Delta(\mathbf{r})\frac{\partial V(z)}{\partial z}\ , \tag{1}$$

the matrix element for scattering in the Born approximation is then given as

$$|\langle \mathbf{k}\,|\,V_{\mathrm{SR}}(\mathbf{r})\,|\,\mathbf{k}'\rangle\,|^2=e^2F_s^2\,|\,\Delta(\mathbf{q})\,|^2\ ,\quad \mathbf{q}=\mathbf{k}-\mathbf{k}'\ , \tag{2}$$

where SR denotes surface roughness, F_s is the average surface field, $\Delta(\mathbf{q})$ is the Fourier transform of $\Delta(\mathbf{r})$, and $\mathbf{q}$ is the scattered wave vector. Such a model is implicitly macroscopic in nature as the atomic variation of the surface field is neglected. We will not attempt to justify this assumption, nor that of the Born approximation in the present work. Rather, we will concentrate on the statistical properties of $\Delta(\mathbf{r})$, the relation of this to $\Delta(\mathbf{q})$, and the corresponding effect on the scattering rate.

In the Born approximation only the magnitude squared of $\Delta(\mathbf{q})$ (referred to as the power spectrum) is needed, and thus the phase of $\Delta(\mathbf{q})$ can be neglected. The usual procedure has been to assume that the autocovariance function of $\Delta(\mathbf{r})$ is isotropic and Gaussian such that[21]

$$C(\mathbf{r})=\langle\Delta(\mathbf{r}')\Delta(\mathbf{r}'-\mathbf{r})\rangle\simeq\Delta^2e^{-r^2/L^2}\ , \tag{3}$$

where $C(\mathbf{r})$ is the autocovariance function, Δ is the rms value of $\Delta(\mathbf{r})$, and L, which governs the decay of the autocovariance, is referred to as the correlation length. By convolution, the power spectrum is just the transform of the autocovariance (3), which is given as

$$S(\mathbf{q})=\,|\,\Delta(\mathbf{q})\,|^2=\pi\Delta^2L^2e^{-q^2L^2/4}\ . \tag{4}$$

This assumption of a Gaussian autocovariance has been used for many years with no real justification. In fact, as we will show in the present paper, this model does not appear to provide an accurate representation of the spectrum in terms of a two-parameter model.

In the present work we attempt to quantify the statistical properties [i.e., $C(\mathbf{r})$ and $S(\mathbf{q})$ in (3) and (4)] of the interface roughness of the Si(100) sample by measuring the roughness observed in HRTEM cross sections of the Si-SiO_2 interface. In Sec. II we describe the sample preparation and cross-sectioning technique used to generate HRTEM micrographs. In Sec. III we discuss the procedure followed in digitizing the interface boundary and estimation of the autocovariance sequence and the power spectrum of the roughness. In Sec. IV the results for different devices will be compared and the roughness parameters Δ and L for both models above will be estimated. In the past the most serious criticism of the HRTEM technique in analyzing surface roughness is that of the averaging effect of projecting the actual two-dimensional roughness into a one-dimensional boundary. To understand this effect, we present in Sec. V a simple model for the projection effect which is used to estimate the correlation between the actual two-dimensional roughness and the one-dimensional roughness measured from HRTEM. Finally, in Sec. VI we discuss the mathematical model for the power spectrum, $S(\mathbf{q})$, and the corresponding effect on the electron scattering rate in the inversion layer.

II. SPECIMEN PREPARATION

In the present studies a comparison between the roughness-limited mobility and the physical roughness of Si(100) devices is made. In order to make such a comparison, cross sections were made of the Si-SiO_2 interface on large-area optical devices adjacent to the Hall bars on which the mobility was measured. The devices used for the present study were fabricated at the Naval Research Laboratory. Typical Si devices were fabricated with a gate oxide grown at 1000 °C in dry oxygen to approximately 0.2 μm in thickness. Cross-sectional HRTEM and Hall mobility measurements were made on devices from one such wafer referred to as sample 1. The addition to normally prepared samples, one wafer, sample 2, was oxidized with 5 vol % HCl in the ambient in order to increase the expected roughness. Another wafer, sample 3, was tilted 1.2° from the [100] direction toward the [111], and thus is expected to show peculiarities in the roughness to accommodate this tilt. All devices were oriented along cleavage planes in the [110] direction, and cross sections were made parallel to the principal axis of the Hall devices.

High-resolution images of the Si-SiO_2 interface have previously been reported[12–15] in which the cross-sectional specimen preparation for HRTEM is essentially the same as the method outlined by Bravman *et al.*[24] Here the samples are cut into strips $\sim$1 mm thick in the direction parallel to the Hall channel. Then the samples are turned vertically and glued together to form a surface with several interfaces. This sample is then mounted on glass and mechanically polished from both sides to a thickness of approximately 50 μm. The sample is then mounted on a metal support ring and ion-thinned using an argon sputter gun. This is done until the central region of the sample is thinned to approximately 100–200 Å, which is sufficiently thin for TEM examination.

The cross sections were examined in a JEOL 200 CX electron microscope at a primary voltage of 200 kV and a point-to-point resolution of 2.5 Å. Two-dimensional lat-

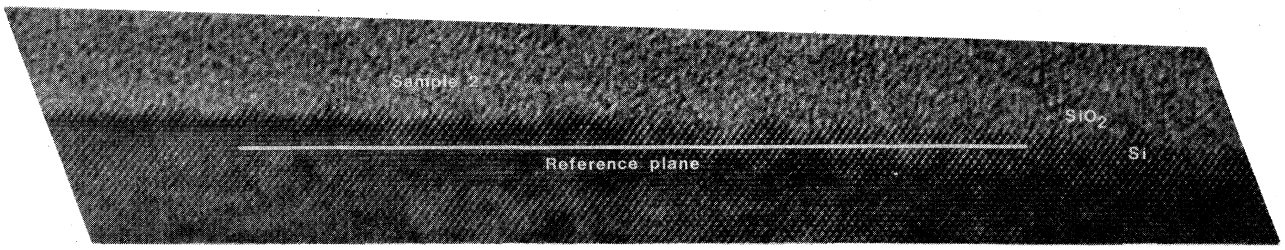

FIG. 1. Si(111) lattice images for sample 2 (intentionally roughened showing interface between the Si and SiO_2.

tice images were recorded with the beam parallel to the Si [011] direction, i.e., with the (100) interface and two sets of (111)-type planes in Si edge on. An example of the lattice images produced for a Si(100)-SiO_2 interface is shown in Fig. 1 for sample 2. The image shown in Fig. 1 is oriented along the [011] direction, and the Si(111) lattice planes in the lower half of the picture are clearly visible. The structureless image in the top half of the picture is the amorphous SiO_2 layer, and the interface between crystalline Si and the amorphous SiO_2 is readily discerned.

III. ANALYSIS

A. Digitization of the interface boundary

In order to estimate the statistical properties of the interface roughness, it is first necessary to digitize the apparent interface boundary. In the present analysis the interface boundary is chosen as the last discernible lattice fringe corresponding to the periodicity of the Si. This procedure is somewhat arbitrary at many points, as an abrupt change from crystalline Si to noncrystalline SiO_2 is not always apparent. Such an analysis also neglects the contribution to the interface width of intermediate bonding state and the existence of a transition layer between Si and SiO_2. However, in the context of the simple interface model used to fit inversion-layer mobility, and the apparent abruptness ascertained in previous studies, such a procedure is deemed acceptable.

In the actual digitization process a reference lattice plane is chosen as shown in Fig. 1. The position of the last discernible lattice image is estimated from the number of intermediate lattice sites between the reference plane and the interface. The primary difficulty in the digitization process arises from the sometimes ambiguous choice of the interface boundary, especially in less resolved pictures. This uncertainty may be modeled more or less as a random error to the roughness and is expected to contribute a "white"-noise component to the spectrum. In comparing pictures which have been redigitized at different times, the relative error in the estimated roughness parameters (Δ and L) is on the order of 10%.

In Fig. 2 we show the result of digitizing the interface boundary of the HRTEM picture of Fig. 3. In the inset of this figure an expanded view is given of the Si(111) lattice fringes at the Si(100) interface. A step (as seen by HRTEM) in the surface (corresponding to a Si^{1+} state) occurs in the horizontal direction at a spacing of 1.92 Å. However, when no step occurs the spacing is 3.84 Å. As it is much more convenient to work with an equally spaced sequence, we simply insert a data point of the same value when the interface is constant so that the sampling increment is uniformly 1.92 Å. In the vertical direction the roughness is forced to assume discrete values consistent with the Si lattice spacing. Here, changes in the position of the interface occur in steps of 2.71 Å, corresponding to the allowed positions of the unresolved basis as viewed by the cross-sectional technique.

In this particular data sequence, which represents perhaps the most extreme example from this wafer, a very large background "trend" is observed to be superimposed on the random fluctuations of the interface. Such trends have been observed in previous HRTEM studies,[12] and appear to be long-wavelength (>200 Å) fluctuations in the interface. The analysis of the statistical properties of the interface roughness in this case is complicated, as such long-wavelength fluctuations represent, at least over the length of the HRTEM picture, a nonstationary (i.e., nonconstant mean) contribution to the roughness. In sequential analysis[25] such trends in a random sequence are often referred to as "deterministic" components in the data, even though, at least in the present case, they may simply be part of a much longer stochastic process. At best, one can only remove such components from the data, either through differentiation or by fitting and subtracting the observed trend. For the present we will attribute such trends in the data to very-long-wavelength fluctuations in the surface which are uncorrelated with the shorter-range fluctuations observed in Fig. 1. As will be discussed elsewhere, the contribution to the electronic scattering rate due to this long-wavelength process is very minimal and

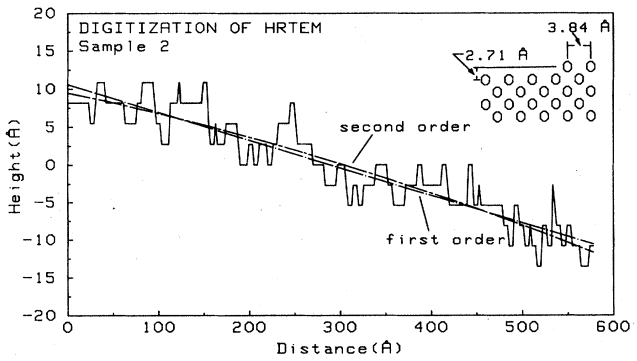

FIG. 2. Digitization of interface boundary including first- and second-order fits to the background. Inset shows the relevant dimensions for steps occurring in the interface.

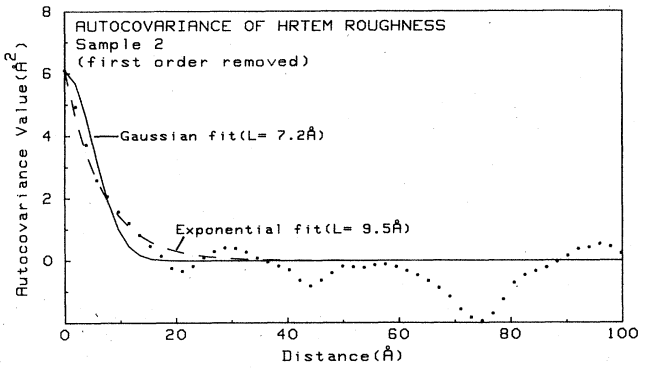

FIG. 3. Autocovariance of the sequence shown in Fig. 2 using (5). The solid line is a fit using a Gaussian model while the dashed line represents an exponential fit.

can be safely ignored. Thus we subtract out the observed trend to form a stationary sequence by least-squares-fitting the background with a low-order polynomial. In Fig. 2 we show the fits corresponding to first and second order. Although the second-order fit appears better, the exact order to subtract is not well established and may introduce unnecessary error in the data. This is discussed further in Sec. V.

B. Autocovariance sequence

From the digitized sequence described above in Sec. II A, the autocovariance function of the HRTEM interface may be estimated. The autocovariance function may be estimated for a stationary process with zero mean as[25]

$$C(m)=\frac{1}{N}\sum_{n=0}^{N-m-1} y(n)y(n+m)\ , \tag{5}$$

where $y(n)$ is the nth value of the roughness sequence, N is the total number of digitized points, and $C(m)$ is the mth autocovariance coefficient. For higher values of m the estimate becomes increasingly poor due to the smaller number of points from which (5) is calculated. Thus, only the first few coefficients are considered to be truly good estimates for the actual autocovariance function. The calculated covariance function of the sequence in Fig. 2 is shown in Fig. 3 (after subtracting a first-order background). The autocovariance function in Fig. 3 is observed to monotonically decay to zero, after which fluctuations are observed related to the statistical uncertainty in the autocovariance estimate [Eq. (5)]. The rms value of the data corresponds to the zeroth coefficient of the autocovariance sequence

$$\Delta_m^2=C(0)=\frac{1}{N}\sum_{n=0}^{N-1} y^2(n)\ . \tag{6}$$

Here Δ_m denotes the estimate of the rms fluctuation of the HRTEM interface.

As discussed earlier, it is usually assumed that the autocovariance function is described by a Gaussian as given in (3). To see how this compares with the estimated covariance, (5), a least-squares fit of a single Gaussian is performed using the correlation length L as a fit parameter and Δ given by (6). The result of this fit is shown by the solid line in Fig. 3. As can be seen, the fit is only fair for short distances and quite poor for larger distances. The disagreement at large distance is associated with the statistical error inherent in (5).[25]

Near the origin, the decay of the autocovariance appears more exponential than Gaussian. Thus, as an alternative two-parameter model we consider the exponential function for the autocovariance,

$$C_e(n)=\Delta_m^2 e^{-(\sqrt{2}x/L_m)}\ ,\quad x=n\,\Delta x \tag{7}$$

where Δx is the sampling interval (1.92 Å) and L_m is the estimated correlation length. The factor of $\sqrt{2}$ in the exponential is included to yield the same prefactor as (4) in the two-dimensional power spectrum. This is discussed in more detail in Sec. VI. The fit using the exponential model above is shown by the dashed line in Fig. 3. This fit is observed to be much closer to the estimated covariance near the origin than the Gaussian model, a fact which is true for all the cross sections studied. Thus, in terms of a two-parameter model for the covariance, exponential decay appears more appropriate than the Gaussian function.

C. Spectral estimation

It is not the autocovariance function that enters into the electronic scattering matrix element, but rather the magnitude squared of the Fourier transform of $\Delta(\mathbf{r})$ referred to as the power spectrum. Thus, for comparison with the actual scattering rate estimated from the Hall mobility, it is desirable to estimate the power spectrum of $\Delta(\mathbf{r})$ itself rather than the autocovariance. The determination of the true power spectrum of a semi-infinite random process based on a finite sampling of data is a classic problem in the field of estimation theory.[25] Thus, many techniques are available for the optimal estimation of the true power spectrum.

In preliminary studies,[13–15] we have calculated the power spectrum from the fast-Fourier-transform (FFT) or "periodogram" of the original sequence. However, as an estimate of the true spectrum, calculating the periodogram is a poor procedure which inherently leads to a large variance between the calculated and true spectrum independent of the length of the sequence due to forcing the data to zero outside the picture length. Due to the large variance, the spectrum obtained by directly transforming the data will fluctuate rather wildly about the true spectrum, giving a very noisy appearance.

As a better estimate of the roughness spectrum, we have fitted the random roughness sequence using an Mth-order autoregressive (AR) model[25,26]

$$y(n)=-\gamma_1 y(n-1)-\gamma_2 y(n-2)+\cdots+\gamma_M y(n-M)+a(n)\ , \tag{8}$$

where γ_n is the nth prediction-error filter coefficient, $a(n)$ is the nth component of a white-noise process (referred to as the residue), and $y(n)$ is the original data. Equation (8) shows the result of starting with a white-

noise sequence $a(n)$ (which has a uniform spectral density), and passing this sequence through a linear, stationary filter to obtain the correlated sequence $y(n)$. Multiplying (8) by $y(n+k)$ and taking the expectation value results in an equivalent difference equation for the autocovariance sequence. With knowledge of the first M values of covariance, the coefficients γ_1 through γ_M may then be solved. The spectrum corresponding to (8) is given by

$$S_E(q)=\sigma_a^2 \Big/ \left|1+\sum_{n=1}^{M-1}\gamma_n e^{-iqn\Delta x}\right|^2, \tag{9}$$

where the coefficients γ_n have the same meaning as in (8) and σ_a^2 is the variance of the white-noise sequence $a(n)$. The spectrum given by (9) has been shown to be formally identical to that arising from maximization of the information entropy, which has interesting physical interpretations as well.[26–28]

In estimating the power spectrum of the surface-roughness function, one begins with the actual data and not the autocovariance coefficients. Therefore, the autocovariance coefficients must be estimated, which, as discussed earlier, becomes increasingly poorer as the coefficient number increases. As a criterion for choosing the optimum order [M in (8)] of the AR process, we have used the final-prediction-error (FPE) method of Akaike,[29] in which the increasing accuracy of higher-order AR processes is weighed against the increasing error of higher covariance estimates as given by (5). Thus, the FPE tends to a minimum value for a certain choice of M in (8) which is taken as the optimal order for representing the spectrum. As a check on the validity of the AR model, one may use the difference equation (8) and the calculated coefficients $\gamma_1-\gamma_M$ to generate the residues $a(n)$ from the original sequence $y(n)$. If the AR model is to be consistent, then the residue sequence given by $a(n)$ should in fact correspond to a white-noise sequence for which various statistical tests exist. For nonstationary sequences such as the one shown in Fig. 3, we find that the AR model fails to generate a white-noise residue spectrum. In every case, however, subtracting a linear fit to the background appears to result in a consistent AR model corresponding to a stationary sequence. This fact seems to indicate that removal of only the first-order background is sufficient to remove the nonstationary effects discussed earlier.

In Fig. 4(a) we show estimates of the power spectrum using the periodogram (direct FFT) and the AR method for the HRTEM micrograph shown in Fig. 1. Note that we plot $|\Delta(q)|$ rather than $|\Delta(q)|^2$ in order to emphasize features in the spectrum. As discussed earlier, the direct FFT of the data sequence of Fig. 2 results in the noisy appearance of the spectrum shown in Fig. 4(a). The variance is in fact proportional to the magnitude of the spectrum, and thus the largest spikes in the spectrum occur close to $q=0$, where the spectrum is largest. In contrast, the fourth-order AR estimate of the same data appears as a smoothed version of the periodogram (the order of the process is determined using the FPE discussed above). Thus, the variance of the AR spectra is considerably reduced compared with that of the FFT spectra due to the low order of the AR model (fourth). In the AR estimate we also observe poorly resolved shoulders or side lobes which may represent a weak periodicity in the data, here on the order of 12–15 Å. At present it is not clear whether such side lobes are artifacts of the sampling techniques or whether they arise from true periodicities in the surface. In fact, as discussed in the next section, these side lobes tend to disappear when different sequences from the same sample are averaged.

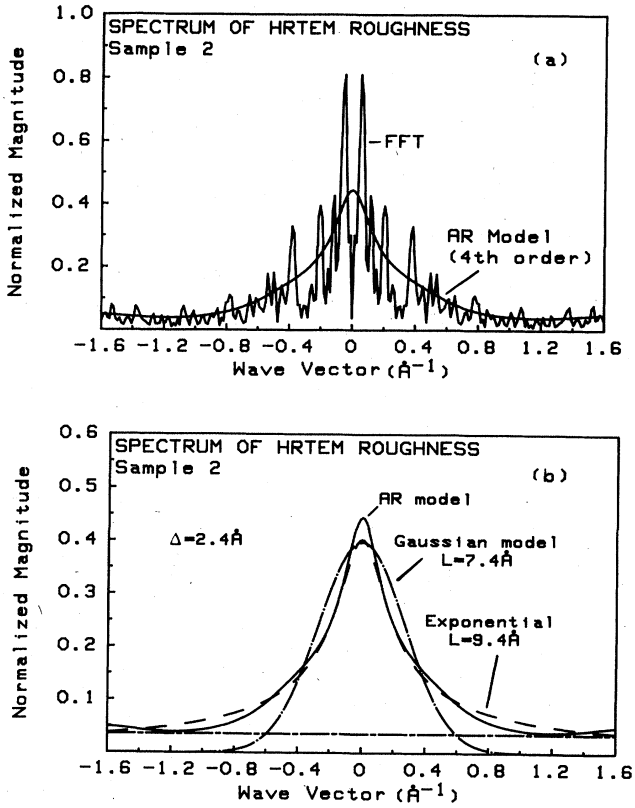

FIG. 4. (a) Direct FFT (fast-Fourier-transform) of sequence shown in Fig. 2 compared to the fourth-order AR (autoregressive) spectrum. (b) Comparison of the fourth-order AR spectrum with the fits arising from the exponential and Gaussian models. The horizontal line is the white-noise spectrum associated with the lattice quantization.

Using the parameters obtained from the fits to the autocovariance sequence, the one-dimensional spectra corresponding to the two-parameter Gaussian and exponential models may be calculated. As seen in Fig. 4(b), the spectral fit due to the exponential model is much better over the entire range of the spectrum than that of the Gaussian model. The Gaussian spectrum is observed to decay slower than the actual spectrum at low wave vector and then falls to zero much too rapidly at high wave vector. This comparison is typical of all the pictures we have studied, as will be seen in Sec. III. The tails in the spectrum, which do not appear in the Gaussian model, appear to be part of a white-noise background in the roughness. Part of this background could arise from digitization error during the sampling process, as discussed earlier. However, even in the most highly resolved pictures, in

which there is little uncertainty concerning the interface boundary, such tails persist. A more likely effect which gives rise to tails in the spectrum of the roughness function is that due to quantization of the interface onto discrete lattice sites. In the present work we have considered the interface "boundary" as the last crystalline lattice site corresponding to the unresolved basis of the Si lattice. Using such a criterion, the actual values of the roughness sequence can only assume values that differ (in the [110] direction at a (100) interface) by increments of $a_0/2$, where $a_0=5.43$ Å is the lattice constant for the Si diamond lattice. The effect of quantization error has been well characterized in the field of communication systems[30] and can typically be modeled as an added white-noise signal with a variance value of $\delta^2/12$ (δ is the quantizing interval) if one assumes the quantization error is uncorrelated with the original signal. The spectrum for a white-noise process is constant for all frequencies (wave vectors) with an amplitude given by the variance. The white-noise spectrum for the present situation is plotted in Fig. 4(b). Here it is observed that the spectrum approaches this level as the wave vector becomes large. This behavior suggests that, at least asymptotically, the roughness spectrum approaches that of a white-noise process with variance $\sigma^2=(a_0/2)^2/12$, due to the discrete nature of the lattice, a fact not accounted for in the assumption of a Gaussian covariance. In the atomic roughness, the interface boundary occurs on actual atomic sites, rather than lattice sites which correspond to the two-atom Si basis. Thus, the effect of quantization is somewhat different in the actual interface than that apparent in the HRTEM roughness and is complicated by the nonrectangular bonding configuration of the tetrahedral Si lattice. However, even in the two-dimensional interface it is expected the quantization effects will influence the spectrum causing deviations from the band-limited Gaussian model given by (4).

IV. RESULTS

A. Normal Si(100)

It is expected that the interface of normally prepared Si surfaces should be relatively smooth, a fact which is borne out by the HRTEM results. In Fig. 5 we show a high-magnification picture of the oxide-semiconductor interface from sample 1. The interface in this picture is observed to be quite smooth with very few steps in the Si surface. In these pictures no evidence for Si clusters in the oxide as measured by Sugano[16] are found. By dividing the number of step discontinuities by the total number of lattice sites available along the [110] direction, a comparison with the step-density measurements of Hahn *et al.*,[18,19] can be made. From the cross sections of sample 1, this value is approximately 14%, which compares to the lower values measured by Hahn *et al.*[19]

To develop a composite picture of the statistical properties of interface roughness, it is necessary to average over the results of pictures from different regions of the wafer. Initially, an effort was made to use cross sections immediately adjacent to the Hall device in which the mobility was measured. However, for the wafers we studied the channel mobility was fairly uniform from device to device, and thus cross sections were used from different regions so that some sort of statistical average of the roughness could be obtained. To perform this average we segment the roughness sequences of different pictures to a length of approximately 500 Å (256 points) and estimate the average parameters Δ and L from these segments. This length was chosen for comparison with the numerical results of the next section. In Table I we tabulate the results from three cross sections of sample 1 with various orders of polynomial background removed. With only the average removed, there is a very large variance in the estimated correlation lengths [both Gaussian (L_g) and exponential (L_e)]. This variance is reduced as successive backgrounds are removed. As discussed earlier, due to large wavelength fluctuations in the roughness, the sequences often appear nonstationary, and the analysis discussed previously is then questionable. However, for the normally oxidized Si samples the original sequence did not seem to suffer particularly from nonstationary effects as judged visually and by the residues of the AR model (see the preceding section). This is evidenced in Table I by the rather monotonic decrease in the correlation length as higher-order backgrounds are removed rather than an abrupt decrease in this parameter from zero to first order.

To properly average the statistical properties of the various segments, we first estimate the autocovariance coefficients of the individual sequences using (7), and then average these coefficients to obtain a composite autocovariance function. From this composite function we then

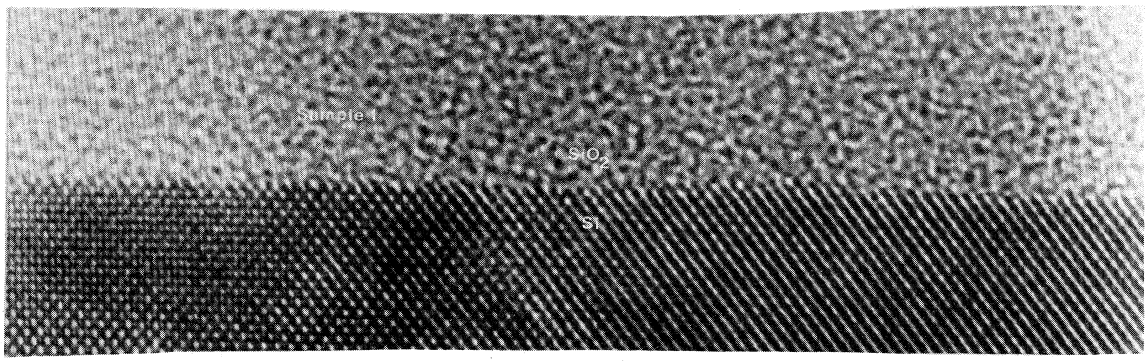

FIG. 5. HRTEM micrograph of Si-SiO_2 interface of sample 1.

TABLE I. Sample 1, normally prepared, Si(100), seven segments of 256 points each, step density of 14%.

Background order removed	Δ_m (Å)	L_g (Å)	L_e (Å)	Average AR order
0	1.99±0.28	24.9±14.2	36.7±23.4	8
1	1.78±0.24	14.9± 9.8	22.0±12.7	8
2	1.73±0.25	14.0±10.1	19.8±13.7	6
3	1.62±0.21	10.3± 6.6	14.0± 8.8	6
4	1.54±0.16	8.3± 4.2	10.9± 6.2	6
5	1.45±0.04	6.6± 2.2	8.5± 3.2	5
6	1.39±0.06	5.7± 1.6	7.1± 2.0	6

estimate Δ and L as discussed earlier and compute the appropriate autoregressive model for the spectrum. In Fig. 6 we show the estimated autoregressive spectrum (with the first order removed for comparison with the following results) along with the fits due to both the Gaussian and exponential models. It is interesting to note that while in Table I the average AR order of the individual segments is greater than 4, the order of the averaged process is only first order. This suggests that features in the individual spectrum, such as the side lobes in Fig. 4, are really artifacts of the individual sequence and are not part of the global structure. As seen in Fig. 6, the two-parameter exponential model is a closer fit over the whole range to the AR model than that of the Gaussian, as noted earlier. This is not surprising as the exponential model turns out to be formally identical to that of a first-order AR process, which will be further discussed in Sec. VI.

B. Intentionally roughened Si(100)

In Fig. 1 we showed the HRTEM micrograph corresponding to sample 2, which was intentionally roughened with the introduction of HCl into the oxidizing ambient. Visually, the interface in Fig. 1 appears rougher than the normal Si case, which is verified from the average step density of 22% measured for the pictures of this sample, although there is considerable variance between pictures, with some pictures appearing nearly as smooth as the normally prepared wafer. However, in contrast to the normal sample, many pictures from sample 2 exhibit nonstationary components, as was discussed in Sec. III. Thus, part of the effect of the HCl vapor could also be in the creation of long-range inhomogeneities in the surface.

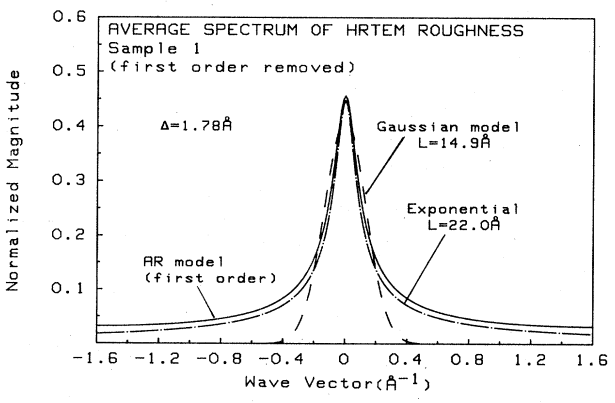

FIG. 6. Composite spectrum represented by a first-order AR model of different cross sections from sample 1, together with the fits using the Gaussian and exponential models.

As discussed in Sec. IV A, we section the data of different pictures from cross sections on the same wafer and average them to obtain the estimate of the roughness parameters. In Table II we compile the average estimate of Δ_m and L_m from eleven sections (some overlapping) taken from two cross sections of sample 2. Here the effect of subtracting the background is much more dramatic. With only the mean removed, the variance in the measured correlation length is quite large. However, upon subtracting a first-order background from each segment, the correlation length is reduced considerably, as is the variance between sections. This behavior is quite suggestive of the presence of an uncorrelated long-wavelength component in the roughness that is being effectively eliminated by the subtraction process. Again, when various sections from sample 2 are averaged, the order of the AR process is found to be unity, which suggests that the spectrum is in fact well represented by a two-parameter model. As with sample 1, as shown for this sample in Fig. 4, the exponential model closely follows the AR spectrum, while the Gaussian fit is much poorer. In comparison with the normally prepared wafer, the average rms height Δ is slightly larger for the intentionally roughened case, while the average correlation length is somewhat less than that of the normal wafer, although for both samples these parameters lie within experimental uncertainty of each other. The results suggest, however, that the interface fluctuations in the roughened sample are, on the average, larger and occur more frequently than in a normally prepared sample.

C. Tilted Si(100)

The HRTEM micrograph of sample 3 is shown in Fig. 7, where the cross section is along the tilt direction. This is evident from the apparent angle of 1.5° between the apparent interface boundary and the Si(100) lattice planes shown in Fig. 8, which is somewhat larger than the nominal tilt angle of 1.2°. From the digitization of the interface shown in Fig. 8, it appears that this tilt occurs over successive terraces in the Si(100) planes rather than a uniform decrease in the interface. Here the tilt angle is taken

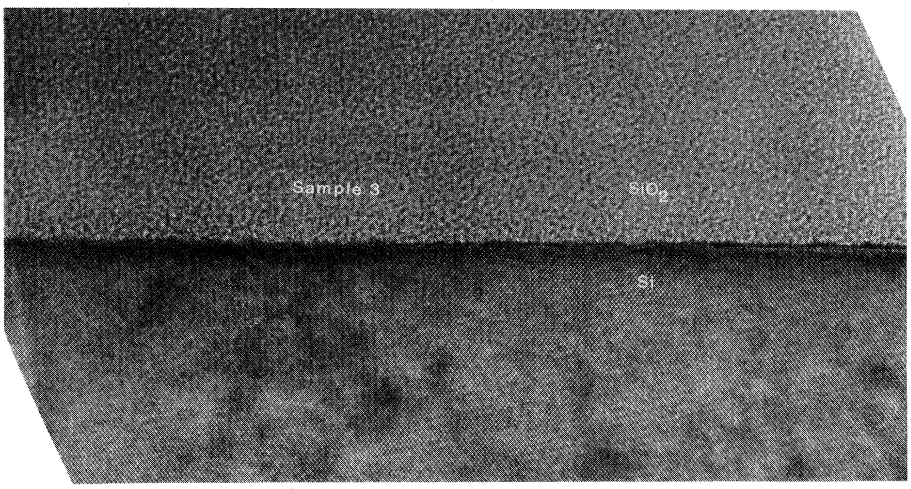

FIG. 7. HRTEM micrograph of sample 3.

as the angle between a first-order fit to the background and (100) surface.

The apparent nonstationarity in the roughness sequence of this sample is related to the intentional tilt in the interface, and thus we have a physical basis for removing at least a linear background from the data. In Table III we again list the estimated roughness parameters for various orders of background removal. We have studied only one cross section from this wafer, so that the number of independent sections is relatively small (4) and thus discussion of the variance of the measured parameters is rather forced. However, we do note a significant reduction of the rms height and correlation length after removing the first-order background that is indicative of the degree of the nonstationarity of this particular data set caused by the tilt.

The average spectrum for the sample 3 shown in Fig. 9 (first-order background removed) is found to be fourth order compared to the previous first-order AR cases. This higher order gives rise to the side lobes shown in Fig. 9, which correspond to a weak periodicity of 11.6 Å. As the number of sections is low, we cannot be certain whether such a periodicity is real or an artifact which could be averaged away with the inclusion of more data. As evidenced from this figure, even for the fourth-order spectrum, the exponential covariance model is a much better fit than the Gaussian. Here the average roughness parameter of the tilted wafer is intermediate between the smooth and intentionally roughened wafers, although the difference again is within the variance of the data.

V. RELATION OF THE ONE-DIMENSIONAL HRTEM ROUGHNESS TO THE SEMI-INFINITE TWO-DIMENSIONAL SURFACE ROUGHNESS

The purpose of the present work is to compare the one-dimensional roughness measured by HRTEM to the two-dimensional roughness comprising the Si-SiO_2 interface. In this respect we need to understand how the roughness parameters Δ_m and L_m estimated in the preceding section from a finite-length one-dimensional sequence compare to a similar pair of parameters characterizing the semi-infinite two-dimensional surface. To accomplish this we have investigated two effects. One is the relative information loss associated with viewing a finite segment of an infinite stochastic process and attempting to estimate the statistical properties of the infinite sequence. This effect we characterize in Sec. V A by synthesizing random data with appropriate spectral characteristics and then attempting to estimate the original spectrum from sectioned

TABLE II. Sample 2, 5 vol % HCl, Si(100), 11 segments of 256 points each, step density of 22%.

Background order removed	Δ_m (Å)	L_g (Å)	L_e (Å)	Average autoregressive order
0	2.82±1.20	24.9±19.4	42.1±34.7	5
1	2.04±0.40	7.1± 2.0	9.9± 4.0	5
2	1.95±0.40	6.0± 1.5	7.7± 2.4	4
3	1.91±0.40	5.6± 1.2	6.8± 1.2	4
4	1.88±0.40	5.4± 1.1	6.5± 1.0	5
5	1.87±0.39	5.3± 1.1	6.4± 1.1	5
6	1.85±0.38	5.2± 1.1	6.2± 1.0	5

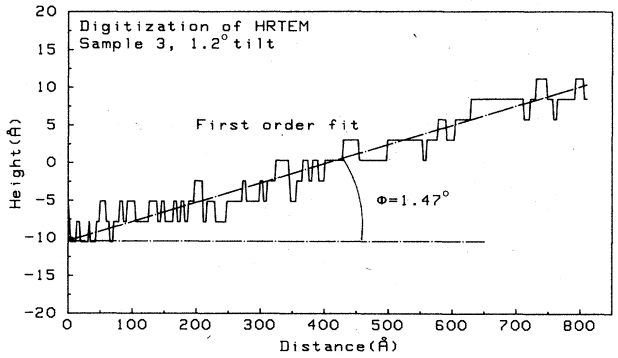

FIG. 8. Digitization of the interface of Fig. 7 with first-order fit showing tilt of the interface from the Si(100) surface.

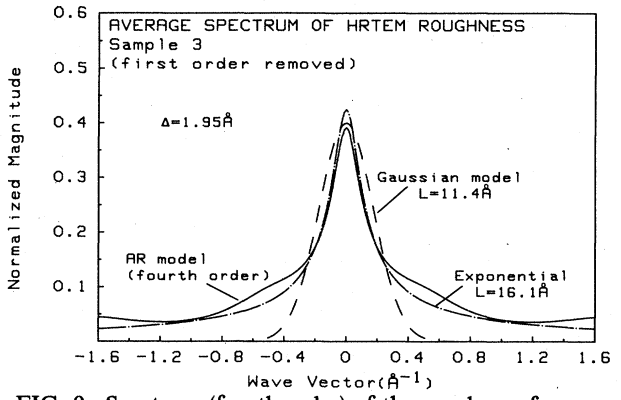

FIG. 9. Spectrum (fourth-order) of the roughness from sample 3 (slope removed) and two-parameter fits using the values shown.

portions of this data. The other effect we investigate is the projection of the two-dimensional roughness via cross-sectional electron microscopy into the one-dimensional sequence digitized and analyzed in the preceding portion of this paper. In Sec. V B we use a simple model for the projection effect to estimate the magnitude of this averaging process by numerically generating a two-dimensional isotropic random surfaces with either Gaussian or exponential statistics and "projecting" through this surface to simulate the HRTEM pictures. From this comparison and the results of Sec. V A we are able to make an estimate of the two-dimensional roughness which enters into the surface-roughness formalism.

A. Finite picture length and nonstationary effects

The typical picture length used in estimating the roughness spectrum varies from 400 to 800 Å along the channel. Random fluctuations in the interface which have wavelength on the order of or greater than the picture length itself will be attenuated as they are not well represented. Due to this fact, the measured values of Δ_m and L_m are compressed somewhat, compared to their true values averaged over the semi-infinite channel length. In Figs. 10—12 we have quantified this effect somewhat by numerically generating long one-dimensional stationary sequences of "colored" random noise with spectral characteristics given by either a Gaussian or exponential model. These data are then "windowed" at various lengths and the parameters Δ_m and L_m are estimated as detailed earlier for comparison with the original parameters used to generate the sequence.

In Fig. 10 we show the effect of picture length on the estimated values of Δ_m and L_m for original values (using the Gaussian model) of $\Delta=3$ Å and $L=15$ Å. Here we average over 100 sequences to find the variance associated with the estimation method. As may be expected, the estimates of Δ_m and L_m become increasingly smaller as the window length decreases, reflecting the reduction in long-wavelength components of the original long sequence. Note also the increase in the statistical variance in Δ_m and L_m as the window length decreases due to the loss of information with which to predict the original spectrum.

In Fig. 11(a) we show the estimate of Δ_m versus the original correlation length (using the exponential model) for various orders of background removed and a fixed window length of 490 Å (which corresponds to the sequence length chosen in Sec. IV). As the original correlation length L increases, the estimate for Δ_m decreases with increasing variance. This trend results from the fact that as L is increased, more of the spectral power is concentrated in the long-wavelength components of the roughness. Thus, due to the finite length of the window,

TABLE III. Sample 3, 1.2° tilt from Si(100) toward Si(111), four segments of 256 points each, step density of 20%.

Background order removed	Δ_m (Å)	L_g (Å)	L_e (Å)	Average autoregressive order
0	4.72±0.88	53.3±7.8	99.0±25.2	5
1	1.97±0.28	11.4±6.1	16.1± 8.3	6
2	1.91±0.24	9.5±3.3	13.4± 5.1	4
3	1.66±0.11	5.4±1.4	7.1± 1.5	7
4	1.60±0.06	4.9±1.5	5.9± 1.7	7
5	1.59±0.06	4.9±1.5	5.9± 1.6	7
6	1.56±0.09	4.5±1.0	5.3± 1.1	7

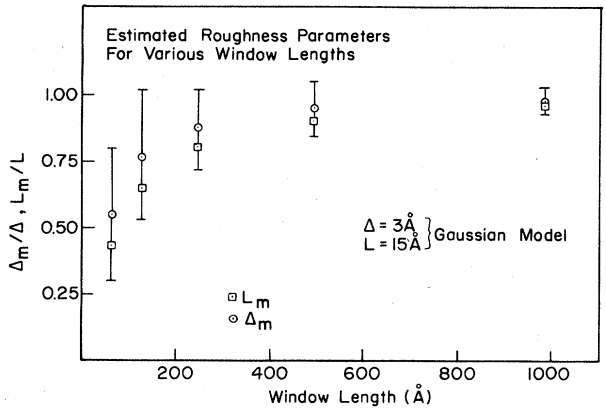

FIG. 10. Effect of window length on the estimated roughness parameters Δ_m and L_m (using a Gaussian model) sectioned from a long Gaussian process with a rms height of 3 Å and a correlation length of 15 Å.

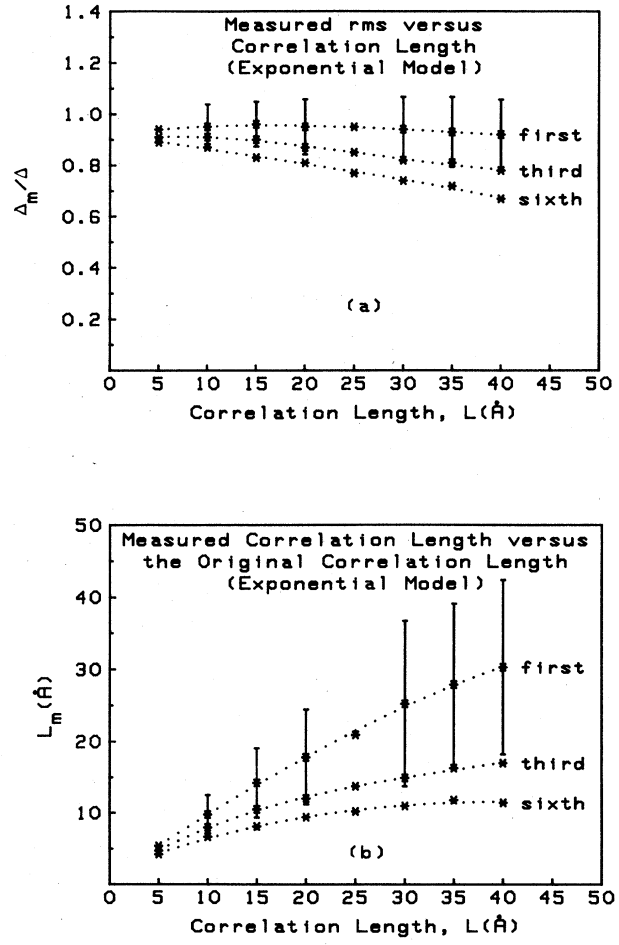

FIG. 11. (a) Estimate of the rms height Δ_m versus correlation length of the original sequence (now using an exponential covariance) for a fixed window length of 490 Å. (b) Estimate of the correlation length L_m versus the correlation length for several different orders of removed background.

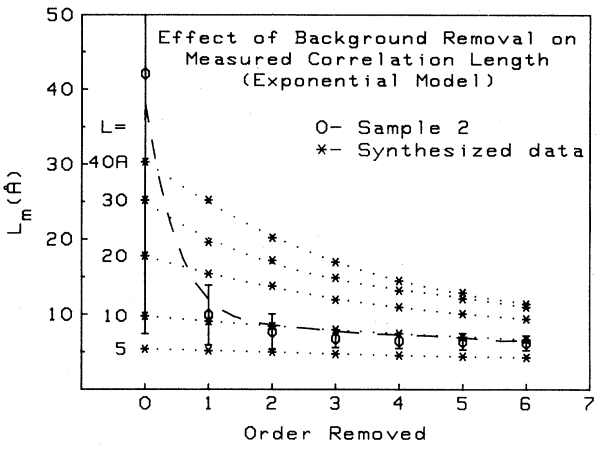

FIG. 12. Estimated correlation length versus order of the polynomial background that is removed for several different values of the original correlation length along with experimental data for sample 2. Dashed line shows the effect of adding a long-correlation-length ($L=400$ Å) Gaussian process to the original sequence in which $L=15$ Å.

fluctuation on the order of or greater than this length will be attenuated, reducing the estimate of Δ_m and increasing the uncertainty. As can also be observed, as increasing higher-order backgrounds are removed, the estimate of the original Δ is reduced. Thus, in correcting for nonstationarity by removing the background, one is also losing information concerning the actual statistics if the original sequence is stationary.

In Fig. 11(b) we plot the estimated correlation length versus original correlation length as in Fig. 11(a). Again, the trend is similar, with the estimated L_m becoming increasingly poorer as the original correlation length is increased and as successively higher-order backgrounds are removed. In Fig. 11(b) it appears that the estimate of L_m saturates for increasing L and a fixed picture length. This implies a limit to which one may estimate the correlation length given a fixed-length picture. Thus, for very-long-correlation-length processes, we cannot predict from a short HRTEM segment what the original correlation length is.

In Fig. 12 we plot the estimated correlation length L_m versus the order of background removed for different values of the original correlation length of the stationary semi-infinite sequence. The effect is a rather monotonic decrease in the estimate which becomes increasingly stronger as the correlation length is increased. Superimposed on this plot are the results from sample 2 (see Sec. IV) shown by the data points. The experimental behavior deviates strongly from the expected behavior of a stationary sequence between zeroth and first order, but then follows the trend quite well for higher orders. We find that the behavior of the estimate for sample 2 is fitted quite well if we add a very-long-correlation-length component ($L=400$ Å) to the original stationary sequence as shown in Fig. 12. This added component then appears as a nonstationary trend when segmented to 500 Å and is virtually eliminated after removing the first-order background.

B. Projection effect

In comparing the roughness parameters obtained from HRTEM pictures to those arising from fitting the mobility, the main problem is relating the one-dimensional parameters in the former case to the two-dimensional values appearing in the scattering matrix element. In particular, one needs to understand what the effect is of "projecting" the interface roughness through the nonzero thickness of the HRTEM cross section to form the actual picture. In the multibeam calculations of Desseaux *et al.*[31] for Ge, the image contrast increases monotonically from zero for sample thicknesses less than 25 Å. Thus, at a critical crystalline material thickness a lattice image becomes "visible" when the contrast is sufficiently large compared to the background noise. As a simple model for the projection effect, we consider a lattice image to be "produced" when a sufficient number of crystalline lattice points (larger than some critical thickness) are aligned along the direction of the electron beam. Usually there is an amorphous layer present following sample preparation which may degrade the resolution, and thus may require a thicker crystalline cross section in order to obtain a clear lattice image. Therefore, we chose (somewhat arbitrarily) a critical thickness of ten atomic layers, which is in excess of the calculated region in which one would expect to see an image. However, as will be shown, the results are rather insensitive to the choice of critical thickness, somewhat justifying this arbitrary choice of thickness.

To characterize the averaging effect, we generate a two-dimensional rough surface with either an isotropic Gaussian or exponential autocovariance by generating the magnitude in the Fourier domain with random phase and inverse-transforming. This random function is taken to represent the interface roughness function $\Delta(\mathbf{r})$ of the Si-SiO_2 interface. A plot of such a random surface obtained using the Gaussian model is shown in Fig. 13(a) for $L=25$ Å. Qualitatively, we see that the mean distance between the "bumps" along this surface is approximately the same as the correlation length L. The surface appears smooth due to the rapid falloff of the Gaussian spectrum with wave vector which eliminates short-range fluctuations. For the exponential model shown in Fig. 13(b), the generated surface has a much noiser appearance due to the tails in the spectrum of this model, which allows for short-range fluctuations as well.

Using the procedure above, a lattice "image" is produced if a sufficient thickness of crystalline material is aligned along the beam. In this way, a one-dimensional interface boundary is produced which corresponds to the actual boundary quantified from HRTEM pictures. This synthesized roughness is analyzed as described in the preceding section to obtain Δ_m and L_m, which are subsequently compared to the original values of the two-dimensional surface. In order to determine the statistical fluctuations of the estimated parameters, we average over 50 different surfaces to calculate the mean and variance.

The results of this calculation for a fixed cross-section thickness of 245 Å is shown in Fig. 14. This thickness represents an extreme case, as usually the cross-section thickness is somewhere between 100 and 200 Å for the

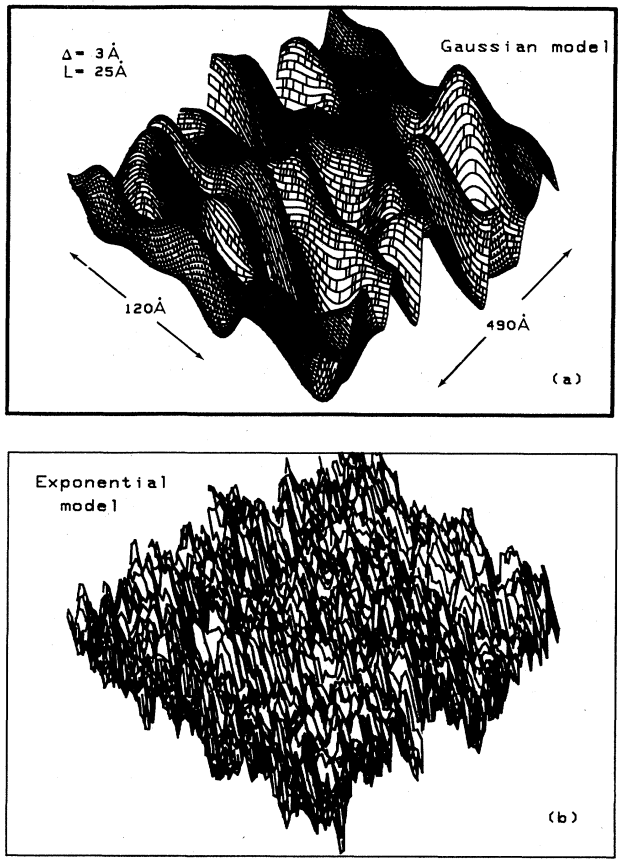

FIG. 13. (a) Two-dimensional random surface generated from a Gaussian model with $\Delta=3$ Å and $L=25$Å. Two-dimensional surface generated from an exponential model with the same parameters as (a).

pictures shown here, and thus the averaging will not be as severe. In Fig. 14(a) we plot the measured Δ_m [normalized to the original two-dimensional (2D) value] versus L for both autocovariance models. For small values of L the averaging effect is severe and a large attenuation of the measured Δ_m occurs. Thus, for short-range fluctuations in the surface, the averaging is more severe. Thus the attenuation of Δ_m is larger for the exponential model, where more of the power is located in the tail of the spectrum. For larger L the measured Δ_m approaches a constant value determined by the finite picture length, which reduces the contribution of long-wavelength fluctuations. As can be seen, the variance of the estimated Δ_m increases with increasing L due to this latter effect. In the above analysis we have neglected the quantization of the roughness onto discrete lattice sites. We have included this effect in the projected two-dimensional surface roughness and the results are qualitatively the same. Quantization effects were not found to be important until the rms fluctuations of the interface are somewhat less than the quantization level itself. The main effect is a reduction of the estimated value of L due to the "whitening" effect of the quantization error. However, quantization of the continuous surface (Fig. 13) changes the statistics of the orig-

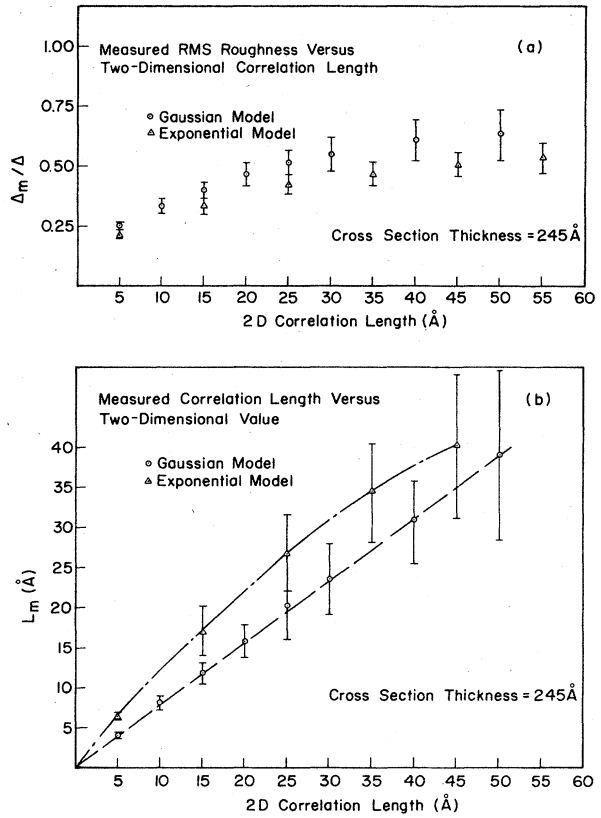

FIG. 14. (a) Estimated Δ_m for the projected one-dimensional sequence versus the correlation length of the original two-dimensional sequence for both the Gaussian and exponential models. (b) Estimated one-dimensional correlation length versus the two-dimensional values as in (a).

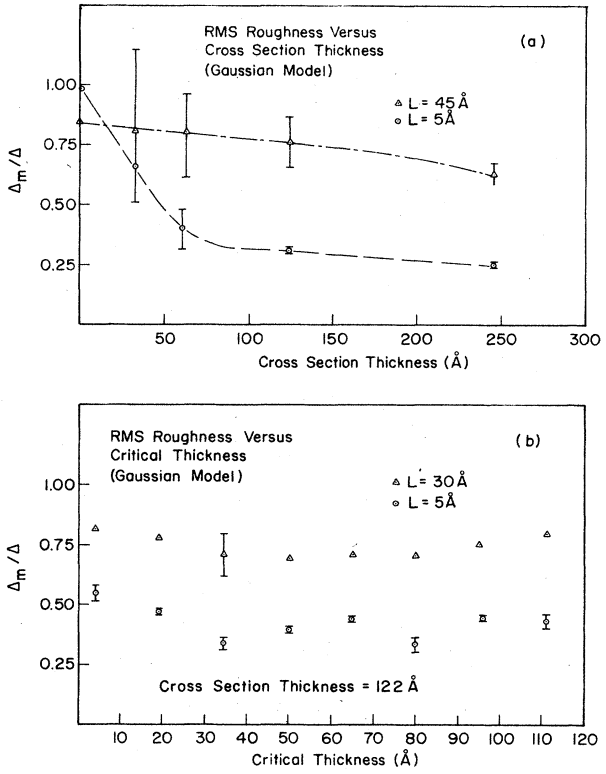

FIG. 15. (a) Effect of cross-section thickness on the estimated rms height for two different correlation lengths of the original surface. (b) Effect of the critical thickness for lattice-image production on the estimated rms height Δ_m.

inal two-dimensional surface, which complicates the comparison. For Δ greater than the lattice spacing, the results shown in Fig. 14 are not strongly effected by quantization.

In Fig. 14(b) we plot the measured L_m versus L of the original surface for both the exponential and Gaussian models. For large correlation lengths, large variance in the estimated values occur and both models underestimate the two-dimensional values of L. This phenomena is related more to the finite-picture-length effect discussed previously than to effects due to projection. For small L the exponential and Gaussian models give opposite effects, with the former overestimating L and the latter underestimating it. The difference in the former case could be due to the functional difference between the one-dimensional and two-dimensional spectra calculated for the exponential model. We find, however, in comparing the projected surface spectrum to the original 2D spectrum, that the shape does not change substantially when projected; that is, a Gaussian autocovariance for the 2D surface still appears to give a Gaussian-like projected interface. Thus, the one-dimensional spectrum still seems to maintain the character of the original 2D surface.

The effect of sample thickness on Δ_m is shown in Fig. 15(a) for short (5 Å) and long (45 Å) correlation lengths. For small correlation lengths, Δ_m decreases substantially as the sample thickness increases. This result is directly attributable to averaging of the roughness via the projection effect, which reduces the rms values of the fluctuations observed in HRTEM. Note that the statistical variation in the estimate is also reduced as the sample thickness increases. Thus, the averaging effect tends to smooth out the fluctuations in the estimate of Δ_m. At higher correlation lengths, the averaging effect is not as pronounced for the range of thicknesses typical for producing lattice images (100—200 Å), so that the estimate for Δ_m in this case does not depend strongly on thickness.

In Fig. 15(b) we plot Δ_m versus the critical thickness for the image production described earlier. This is calculated for both short ($L = 5$ Å) and long ($L = 30$ Å) correlation lengths and, as may be observed, over a broad range of thicknesses there is really little effect due to the exact choice of the critical thickness. Thus, the somewhat arbitrary choice of this parameter does not really change the results presented here.

VI. DISCUSSION OF RESULTS

A. Relation of AR model to two-parameter models

In Sec. IV we showed that for the averaged autocovariance function of the roughness sequence the spectrum is

well characterized by a first-order autoregressive (AR) model, even though the spectrum of individual sequences may be of higher order. To see how this first-order AR model compares to the simple two-parameter models for the covariance, one starts with the difference equation (8), which represents an Mth-order AR process for the realization $y(n)$ of the nth point in the roughness sequence. If we multiply (8) by $y(n+k)$ and take the expectation value, then by definition of the autocovariance we obtain

$$C(k)=-\gamma_1 C(k-1)-\gamma_2 C(k-2) - \cdots -\gamma_M C(k-M) , \quad k>0 \tag{10}$$

where $C(k)$ is the kth autocovariance coefficient, and the γ_i's are the prediction-error filter coefficients given in (8) and (9). The term involving $a(n)$ in (8) (the white-noise sequence) vanishes when the expectation value is taken. Knowing the first M autocovariance coefficients, (10) may be solved to yield

$$C(k)=A_1 G_1^{-|k|}+A_2 G_2^{-|k|}+\cdots , \tag{11}$$

where the G_i's are the roots of the characteristic equation

$$p^M+\gamma_1 p^{M-1}+\cdots+\gamma_M=0 , \tag{12}$$

and the A_i's are obtained from the boundary conditions given by the first M autocovariance coefficients. In general, the roots of (12) may occur in complex-conjugate pairs which combine to give decaying sinusoidal solutions. Such solutions are responsible for the side-lobe structures observed in higher-order spectrums such as Fig. 10. For a first-order AR process, the root is real (and for all the sequences analyzed here, positive as well), and thus the autocovariance becomes

$$C(k)=AG^{-|k|}=Ae^{-\ln(G)|k|} . \tag{13}$$

If we consider (13) as a sampled version of a continuous function with a sampling increment Δx, then the first-order AR autocovariance is formally identical with the exponential model discussed previously, with the correlation length given as

$$L=-\sqrt{2}\,\Delta x/\ln(G) , \tag{14}$$

and the rms value

$$\Delta=A^{1/2} . \tag{15}$$

Thus the exponential model is just a special case of a general autoregressive process, of first order. In fact, the autocovariance of high-order AR processes can be decomposed into a sum of exponentially decaying solutions as shown by (11), some terms containing oscillatory contributions as well. The values of L calculated from the AR model using (14) are, in general, found to be slightly smaller than the fits using the exponential model. This fact arises because the first-order AR model considers all the higher autocorrelation coefficients irrelevant past the first two, while the exponential model was fitted to the whole autocovariance sequence.

First-order AR processes are characteristic of a variety of random processes such as speech-recognition models, ultrasonic waveforms, etc., and thus it is not unreasonable that this is characteristic of the random roughness as well. The first-order AR process is of special significance, however, as this is a special case of a general Markovian process. This is apparent from the first-order difference equation describing the first-order AR process,

$$y(n)=\gamma y(n-1)+a(n) , \tag{16}$$

in which the future value of the random process $y(n)$ depends only on the present value and is not affected by information concerning previous values. Markovian processes appear quite often in statistical physics, for instance, in the description of Brownian motion using the Fokker-Planck equation. However, in the present case the roughness is not a Markovian process in time, but rather one expressed as a function of the distance along the interface.

B. Two-dimensional roughness

In analyzing the spectrum of the one-dimensional roughness sequence, the representation using a low-order autoregressive model is probably the most reasonable approach. However, in the present work we are interested in estimating the two-dimensional roughness, which, when projected through the cross-section thickness, results in the one-dimensional sequence we measure from the HRTEM pictures. In fact, one may define an autoregressive or all-pole model to describe the two-dimensional random surface as[32]

$$y(m,n)=\sum_{p=0}^{N-1}\sum_{q=0}^{M-1}\gamma(p,q)y(m-p,n-q)+a(m,n) , \tag{17}$$

where $y(m,n)$, $a(m,n)$, and $\gamma(p,q)$ are the two-dimensional analogs of the same parameters appearing in (8). However, in contrast to the one-dimensional case, the spectrum arising from (17) does not correspond to the two-dimensional maximum-entropy spectrum.

In the present work, rather than pursuing more sophisticated representations of the two-dimensional roughness as described by (17), we simply assume that the decay of the autocovariance is exponential and isotropic. For this model the matrix element for scattering is proportional to the two-dimensional transform of (7), which is

$$S(\mathbf{q})=\pi\Delta^2 L^2/[1+(q^2L^2/2)]^{3/2} . \tag{18}$$

Through the factor of 2 appearing in the exponent of (7), (18) is equal to 4 at $\mathbf{q}=\mathbf{0}$, but decays more slowly than the Gaussian model, especially at large wave vectors. Using the Ando model[23] for surface-roughness scattering, the momentum relaxation time is given as

$$\frac{1}{\tau(k)}=\frac{e^2F_s^2m^*}{2\pi\hbar^3}\int_0^{2\pi}d\theta(1-\cos\theta)S(q) \times\left[\frac{\Gamma(q)}{\epsilon(q)}\right]^2 , \quad q=2k\sin\theta/2 \tag{19}$$

where m^* is the effective mass parallel to the interface, $\hbar$ is Planck's constant, k is the electron wave vector, θ is the scattering angle, $\Gamma(\mathbf{q})$ represents corrections for image potential and electric field modification at the deformed in-

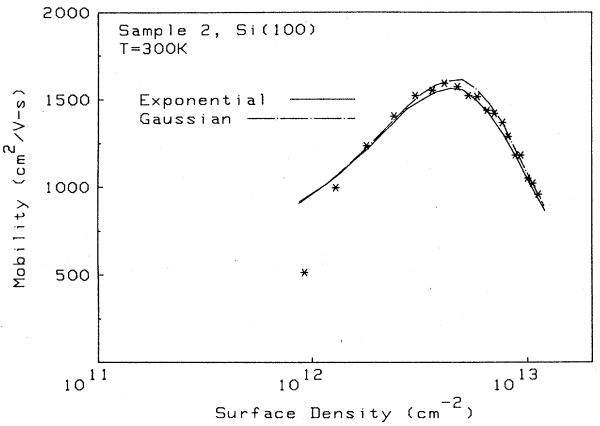

FIG. 16. Hall mobility for sample 2 at 52 K and the corresponding fits using the exponential (solid line) and Gaussian (dashed line) models. For the exponential model, $\Delta=4.8$ Å and $L=13$ Å, while for the Gaussian model, $\Delta=3.5$ Å and $L=15$ Å. The interface-impurity charge density is $N_i=1.1\times10^{12}$ cm^{-2} in both cases.

terface, $\epsilon(\mathbf{q})$ is the electron dielectric function to account for free-electron screening, and other parameters are defined in (1)—(4).

To illustrate the effect of the roughness spectrum $S(\mathbf{q})$ on the channel mobility, we fit the Hall mobility of a device from sample 2 using (19) as well as contributions from impurity and acoustic-phonon scattering with full temperature-dependent screening.[1] This fit is shown in Fig. 16, where the ionized-impurity concentration, the rms height, and the correlation length L are used as fit parameters with $S(\mathbf{q})$ given by both (18) and (4). Qualitatively, the fit to the mobility is the same for both models. The difference is in the magnitude of the roughness parameters used in the fit. For the exponential model (18), a larger value of Δ is required, as well as a shorter value of L, which qualitatively implies that a rougher interface is required in this model to match the scattering rate of the Gaussian model. This behavior may result from the fact that more of the spectral power in the exponential model is located in the high-frequency tails of the spectrum where scattering is not as efficient. A detailed discussion of the effect of the roughness spectrum on the electron mobility and comparison of the measured Hall mobility with the HRTEM results will be reported elsewhere.[33]

The mobility-fit parameters shown in Fig. 16 are somewhat larger than the HRTEM roughness parameters given in Table II. However, the HRTEM parameters are estimated from the one-dimensional roughness, and, as discussed earlier, are compressed from their corresponding two-dimensional values through averaging and picture-length effects. To quantify this averaging effect, we simulate the effect of averaging on the measured roughness using a simple projection model as discussed in Sec. V. The results of the simulation show that there is in fact a direct relation between the one-dimensional roughness measured by HRTEM and the effective two-dimensional roughness. The rms height is reduced, particularly as the correlation length becomes small (less than 10 Å), due to the projection effect through the cross section. On the other hand, the estimated correlation length is reduced and greater error introduced as the original correlation length is increased due to limitations associated with the finite picture length. This latter effect is really the more serious effect, as for long-correlation-length processes one cannot unambiguously estimate the roughness spectrum from a short picture length. This question of picture length is intimately related to the problem of the nonstationarity of the roughness sequence discussed earlier. Long-wavelength components in the roughness are present which span the picture length and thus appear as a nonstationary component. At best, we can assume that these long-wavelength fluctuations are uncorrelated with the shorter processes and that we effectively eliminate these components by subtracting the first-order fit to the background from the overall sequence. We chose first order because this is the minimum order necessary to force the sequences to obey a stationary model as judged by the residue spectrum discussed in Sec. III. As seen in Fig. 12, the behavior of the experimental data of sample 2 mimics that of the sum of two uncorrelated stationary signals, one with a long correlation length. Thus, such a model is plausible, but does not include the possibility that the roughness is really a single long-correlation-length process that is inadequately represented due to the picture length.

To compare the HRTEM roughness to the actual two-dimensional roughness appearing in (19), the one-dimensional roughness parameters are corrected for the effect of background subtraction, using Fig. 12, and then compared to the effective two-dimensional parameters using the results of Sec. V. In Sec. V we presented the results of the projection effect of a thick cross section of 245 Å, which is rather thick. The usual thickness actually varies somewhat monotonically from 100 to 200 Å across the HRTEM picture. This thickness variation is difficult to quantify experimentally, and due to the simplicity of the projection model it is probably not meaningful to include. Therefore, we simply assume an average cross-section thickness of 122 Å upon comparison to the two-dimensional roughness. The estimated two-dimensional roughness parameters should not, therefore, be taken too literally, and should instead be used as a gauge of the projection effect on the roughness.

Here, a comparison of the average one-dimensional roughness parameters (with first order removed) to the effective two-dimensional parameters estimated as outlined above is in order. The results for sample 3 are included only for comparison, as the expected anisotropy due to the tilt angle invalidates the simple isotropic model used. We see that the difference in Δ between samples 1 and 2 is further widened due to the shorter correlation length of sample 2, which increases the averaging effect due to projection. The difference in correlation length is more pronounced as well, although the experimental uncertainty in the estimation of this parameter is much greater. As shown in Fig. 16 and also in previous work,[14,15] the one-dimensional HRTEM roughness parameters are invariably smaller than the corresponding transport results. After correcting for projection effects, however, this trend is reduced and even reversed, with the Hall parameters

and the HRTEM parameters agreeing somewhat better.[33]

A direct comparison of the present results with roughness estimates based on LEED and XPS is somewhat difficult as these results are interpreted in terms of step density along the surface. The estimates of the step density from the HRTEM were found to average 14% for the normal wafer, sample 1, and 22% for the roughened sample 2, thus falling within the lower and upper limits measured by Hahn and Henzler.[18] The trend seems to be that pictures with a shorter correlation length have a high step density, regardless of the rms height. Thus the step density measured by LEED seems to relate more to the correlation length L rather than the rms height of roughness. Overall, the magnitude of the roughness here is still in agreement with previous studies, even when corrected for projection effects which indicate interface widths of 1 or 2 monolayers. The idea of an exponential covariance which has high-"frequency" (wave vector) tails in the spectrum is certainly more compatible with a step-density model, which also contains high-frequency components due to the discontinuous changes in the surface. A Gaussian model is essentially band-limited at the inverse of the correlation length and thus appears rather smooth, as demonstrated by the comparison in Fig. 14. Thus, a Gaussian model for the covariance is not consistent with the periodcity of the crystalline lattice, which requires the interface boundary to assume discrete values.

In comparing the roughness of the present Si(100) devices to that of other MOS systems, we find the present samples to be quite smooth. HRTEM cross sections have now been made of the oxide-GaAs interface[34] and the SiO_2-InP interface,[35,36] both of which show substantially more roughness than the Si-SiO_2 interface. For these oxide—III-V-compound interfaces, rms roughnesses sometimes in excess of 20 Å have been measured, depending on the surface preparation. Evidence for intermediate oxides and dislocations at the interface are also found in these III-V systems which deviate substantially from the abrupt boundary model used in interpreting Si mobility data.

VII. SUMMARY

In conclusion, we have attempted to estimate the statistical properties of the randomly varying roughness at the Si-SiO_2 interface using cross-sectional high-resolution transmission electron microscopy (HRTEM). Comparison between normally oxidized, intentionally roughened, and tilted Si(100) substrates indicate that the one-dimensional roughness is characterized as a first-order autoregressive (AR) or Markovian process, albeit in the presence of a nonstationary component representative of very-long-wavelength fluctuations in the interface. This first-order AR model is characteristic of an exponential decay in the autocovariance model rather than the usual Gaussian model. Such a model is more compatible with the discrete nature of the roughness which occurs at atomic steps along the interface than is the assumption of a Gaussian covariance.

The effect of projection of the two-dimensional roughness through the nonzero thickness of the cross section to form the HRTEM micrograph has been characterized using a simple model. Here, a one-to-one relation appears to exist within statistical uncertainty between the one-dimensional roughness and the original two-dimensional roughness. The main reduction occurs in the rms or variance of the roughness which is decreased by the averaging effect of projection. Although the correlation length does not appear to be substantially affected by the projection process, the finite picture length of the HRTEM micrographs may severely effect the measured value of this latter parameter.

In using the exponential model to calculate the scattering rate of inversion-layer electrons by surface roughness, larger values of the rms height Δ are required to fit experimental mobility data than are required using a Gaussian model for the autocovariance. This results in a larger error in the former case between the mobility parameters and the one-dimensional roughness parameters obtained from the HRTEM pictures. However, after correcting for projection effects in the HRTEM process, the agreement with the exponential model is in fact closer and thus does not appear to provide an unreasonable model for surface-roughness scattering.

ACKNOWLEDGMENTS

The authors would like to thank Martin Peckerar at the U.S. Naval Research Laboratory for device fabrication and Bruce McCombe at the University of New York at Buffalo for many useful discussions. We would also like to express our thanks to C. Loeffler and S. Udpa for useful discussions relating to spectral estimation techniques. This research was supported by U.S. Office of Naval Research Grants No. N00014-78-C-0124 and No. N00014-76-C-0387 and National Science Foundation Grant No. DMR-81-17052.

---

*Present address: Center for Solid State Electronics, Arizona State University, Tempe, AZ 85281.

†Present address: Lawrence Berkeley Laboratory, University of California, Berkeley, CA 94720.

‡Present address: Microelectronic Center of North Carolina, Research Triangle Park, NC 27709.

[1]T. Ando, A. B. Fowler, and F. Stern, Rev. Mod. Phys. **54**, 437 (1982).

[2]T. Ando, Surf. Sci. **98**, 327 (1980).

[3]T. W. Nee, Phys. Rev. B **29**, 3225 (1984).

[4]See, for instance, *The Physics of SiO_2 and its Interfaces,* edited by S. Pantelides (Pergamon, New York, 1977).

[5]J. S. Johannesen, W. E. Spicer, and Y. E. Strausser, J. Appl. Phys. **47**, 3028 (1976).

[6]J. F. Wager and C. W. Wilmsen, J. Appl. Phys. **50**, 874 (1979).

[7]C. R. Helms, N. M. Johnson, S. A. Schwartz, and S. E. Spicer, J. Appl. Phys. **50**, 7007 (1979).

[8]H. Frenzel and P. Balk, in *The Physics of MOS Insulators,* edited by G. Lucovsky, S. Pantelides, and F. L. Galeener (Pergamon, New York, 1980), p. 246.

[9]R. S. Bauer, R. Z. Bachrach, and L. J. Brillson, in *The Physics of MOS Insulators,* Ref. 8, p. 221.

[10]F. J. Grunthaner, P. J. Grunthaner, R. P. Vasquez, B. F. Lewis, J. Moserjion, and A. Madhukar, J. Vac. Sci. Technol. **16**, 1443 (1979).
[11]M. W. Cheung, L. C. Feldman, P. J. Silverman, and I. Steinsgaard, Appl. Phys. Lett. **35**, 859 (1979).
[12]O. L. Krivanek and J. H. Mazur, Appl. Phys. Lett. **37**, 392 (1980).
[13]S. M. Goodnick, R. G. Gann, D. K. Ferry, C. W. Wilmsen, and O. L. Krivanek, Surf. Sci. **113**, 145 (1982).
[14]S. M. Goodnick, R. G. Gann, J. R. Sites, D. K. Ferry, C. W. Wilmsen, D. Fathy, and O. L. Krivanek, J. Vac. Sci. Technol. B **1**, 803 (1983).
[15]Z. Liliental, O. L. Krivanek, S. M. Goodnick, and C. W. Wilmsen, Mater. Res. Soc. Symp. Proc. **35**, 193 (1985).
[16]T. Sugano, Surf. Sci. **98**, 145 (1980).
[17]M. Henzler, Surf. Sci. **73**, 240 (1978).
[18]P. O. Hahn and M. Henzler, J. Appl. Phys. **52**, 4122 (1981).
[19]P. O. Hahn and M. Henzler, J. Vac. Sci Technol. A **2**, 574 (1984).
[20]M. H. Hecht, P. J. Grunthaner, and F. J. Grunthaner, in *Proceedings of the* 17th *International Conference on the Physics of the Semiconductors, San Francisco, 1984,* edited by J. D. Chadi and W. A. Harrison (Springer, New York, in press).
[21]R. E. Prange and T. W. Nee, Phys. Rev. **168**, 779 (1968).
[22]R. F. Greene, in *Molecular Processes on Solid Surfaces,* edited by E. Drauglis, R. D. Getz, and R. J. Jaffee (McGraw-Hill, New York, 1967), pp. 239—263.
[23]T. Ando. J. Phys. Soc. Jpn. **43**, 1616 (1977).
[24]J. C. Bravman and R. Sinclair, J. Electron. Micro. Technol. **1**, 53 (1984).
[25]See, for example, G. M. Jenkins and D. G. Watts, *Spectral Analysis and its Applications* (Holden-Day, San Francisco, 1968).
[26]T. J. Ulrych and T. N. Bishop, Rev. Geophys. Space Phys. **13**, 183 (1975).
[27]A. van den Bos, IEEE Trans. Inform. Theory **IT-17**, 493 (1971).
[28]J. P. Burg, Geophysics **37**, 375 (1972).
[29]H. Akaike, Ann. Inst. Statist. Math. **22**, 203 (1970).
[30]W. R. Bennett, Bell Syst. Tech. J. **27**, 446 (1978).
[31]J. Desseaux, A. Renault, and A. Bourret, Philos. Mag. **35**, 357 (1972).
[32]D. E. Dudgeon and R. M. Marsereau, *Multidimensional Digital Signal Processing* (Prentice-Hall, Englewood Cliffs, 1984), pp. 315—339.
[33]S. M. Goodnick, J. R. Sites, D. K. Ferry, and C. W. Wilmsen (unpublished).
[34]O. L. Krivanek and S. L. Fortner, Ultramicrosc. **14**, 121 (1984).
[35]Z. Liliental, O. L. Krivanek, J. F. Wager, and S. M. Goodnick, Appl. Phys. Lett. **46**, 889 (1985).
[36]O. L. Krivanek, Z. Liliental, J. F. Wager, R. G. Gann, S. M. Goodnick, and C. W. Wilmsen J. Vac. Sci. Technol. B **3**, 1081 (1985).

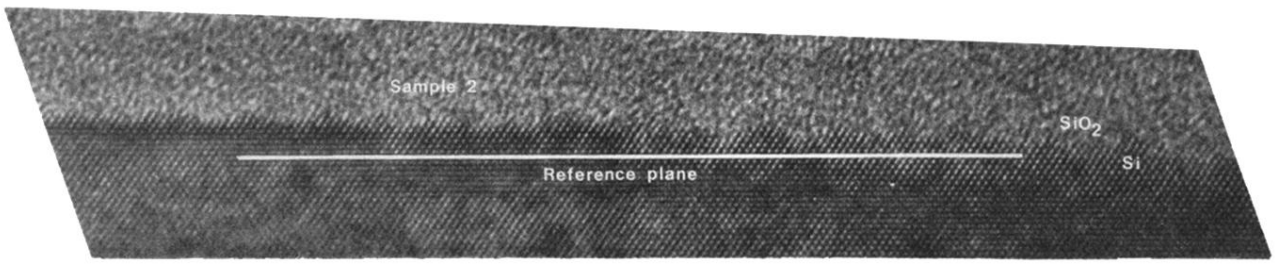

FIG. 1. Si(111) lattice images for sample 2 (intentionally roughened showing interface between the Si and SiO_2.

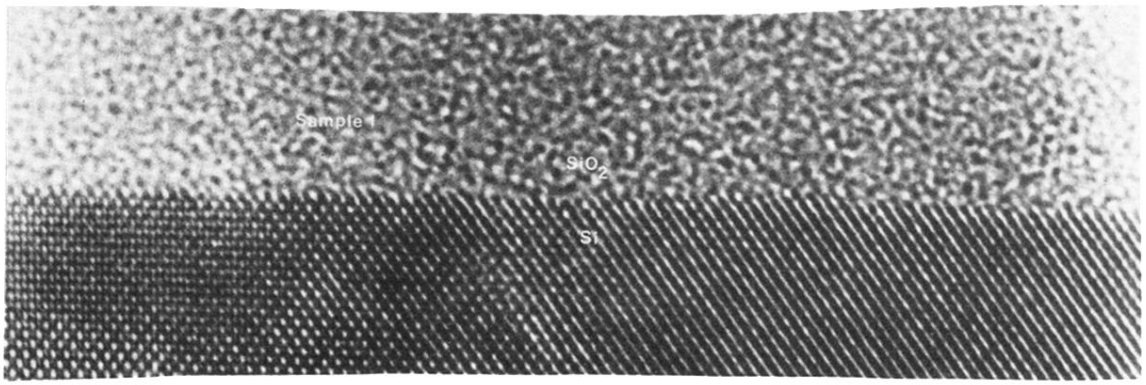

FIG. 5. HRTEM micrograph of Si-SiO_2 interface of sample 1.

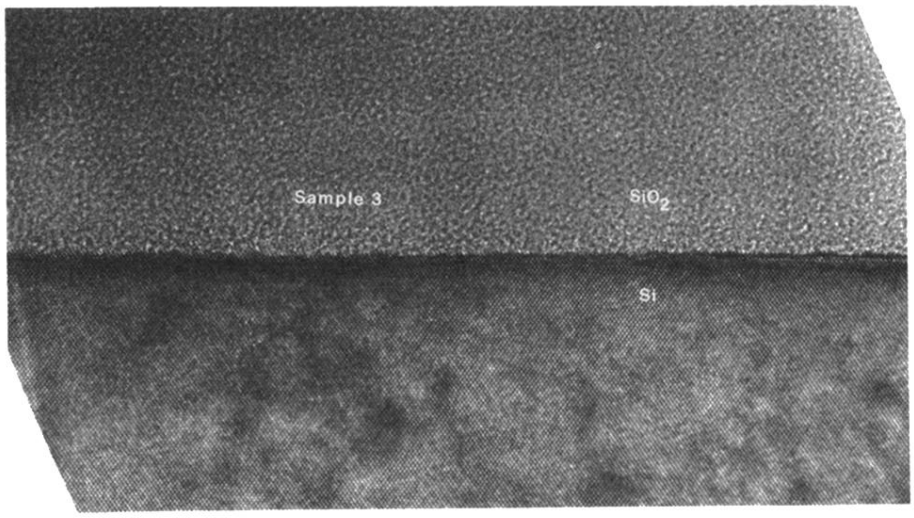

FIG. 7. HRTEM micrograph of sample 3.