Electron-phonon interaction effects on the dielectric response of Si

L. F. Lastras-Martínez
Instituto de Investigación en Comunicación Óptica, Universidad Autónoma de San Luis Potosí Álvaro Obregón 64, 78000, San Luis Potosí SLP, México

M. Cardona
Max-Planck-Institut für Festkörperforschung Heisenbergstrasse 1, 70569 Stuttgart, Germany

(Recibido: 26 de julio de 2005; Aceptado: 8 de diciembre de 2005)

The availability of isotopically pure semiconductors in the last fifteen years has triggered their scientific and technological interest. The effects of the electron-phonon interaction on the band structure can be experimentally investigated by measuring the temperature or the isotopic composition dependence of energy gaps. In this article, we discuss the effects of isotopic composition on the dielectric function of silicon by using spectroscopic ellipsometry in the energy range from 3.1 to 3.7 eV. The silicon crystals investigated are the isotopically pure $^{28}$Si and $^{30}$Si, and the natural Si ($^{nat}$Si, $M_{nat}$=28.14 amu). At low temperatures, the energies of the interband transitions become mass-dependent through the dependence of the electron-phonon interaction and the lattice parameter on the average isotopic mass. We determine the mass dependence of critical point energies and other optical parameters as accurately as possible by analyzing the ellipsometric data in reciprocal (Fourier-inverse) rather than direct (frequency) space.

Keywords: Optical constants; Semiconductors; Theory; Models; Numerical simulation

1. Introduction

The effects of isotopic composition on lattice constants and electronic properties of semiconductors have an increasing scientific and technological interest due to the availability of isotopically pure materials. Experimental measurements of the isotopic mass dependence of lattice constant for diamond [1,2] Ge [3, 4], Si [5], GaAs [4] and ZnSe [4], as well as theoretical calculations for diamond [6], Si [6, 7], Ge [4, 6], GaAs [4, 8] and ZnSe [4, 8] have been reported in the literature. Likewise, there are several reports concerning the effects of the isotopic composition on the interband energies of the lowest direct or indirect gaps of GaAs [4], Ge [9, 10], ZnSe [11], CdS [12], and CuCl [13]. For higher energies, spectroscopic ellipsometry (SE) has been used to measure the isotopic dependences of the gaps and broadening parameters corresponding to the $E_1$ and $E_1+\Delta_1$ transitions of Ge [9, 10] and Si [14, 15].

Physical properties of semiconductors such as thermal conductivity, thermal expansion and gap energies of the interband transitions, are determined by the propagation of vibrations (phonons) through the lattice and their interaction with electronic states. These properties can be modified, for instance, by changing the mass of the atoms that constitute the lattice, i.e., by modifying the isotopic composition of the semiconductor.

In the case of the gap energies, the dependence on the isotopic composition has been attributed to three effects. The simplest one is related to the thermal expansion [16], which is explained by assuming anharmonic potentials for the corresponding oscillators. In this case, the time-average position of the oscillators is no longer equal to $a_0$, but leads to a lattice parameter which is shifted to $a_0+\Delta a_0$. The formalism used to describe thermal expansion is the so-called quasi-harmonic approximation, which assumes that the interatomic potential is still harmonic, with the energy levels of the vibrations given by $E_n=(n+1/2)\hbar\omega_0$, but with the lattice constant dependent on temperature. The effect on the critical point energy is given by [17]:

$$\Delta \omega_g(T) = a B_0 \left( \frac{\Delta V(T)}{V_0} \right)_p,$$

$$= \left( \frac{\partial \omega_g}{\partial P} \right)_T \left( V_0 \frac{\partial P}{\partial V} \right)_T \left( \frac{\Delta V(T)}{V_0} \right)_p$$

(1)

Where $a$ is the volume deformation potential, $B_0$ the bulk modulus, $P$ the pressure and $\Delta V(T)$ the volume expansion. In Eq. (1), the term with the strongest temperature dependence is the thermal expansion $(\Delta V(T)/V_0)_p$.

The other two effects that modify the gap energies, representing explicitly the electron-phonon interaction, are (a) the Fan terms, corresponding to a complex self-energy, and (b) the Debye-Waller terms. These terms are of second order in the phonon amplitude $u$ and the temperature effect appears in the thermodynamical expectation value $\langle u^2 \rangle$, which is given by [17]:

$$\langle u^2 \rangle = \frac{\hbar}{4M_j \Theta_j} \left[ 1 + 2n_B \right]$$

(2)

where $n_B=1/(e^{E_g/\Theta}-1)$ is an average Bose-Einstein statistical factor, and $M_j$ the mass of the $j$th isotope with the corresponding phonon frequency $\Theta_j$. 

©Sociedad Mexicana de Ciencia y Tecnología de Superficies y Materiales
Table 1. Renormalization parameters $a_B$ and unrenormalized gap $E_B$ for the E1 transitions of Si obtained in the present work and from the literature (based on the measured and calculated temperature dependence of the E1 transition of $^{28}\text{Si}$). The value of the renormalization parameter includes both components: the electron-phonon interaction and the volume effect.

<table>
<thead>
<tr>
<th>Source</th>
<th>$a_B$ (meV)</th>
<th>$E_B$ (eV)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Present work</td>
<td>110±5</td>
<td>3.562±0.005</td>
</tr>
<tr>
<td>Other measurementsa</td>
<td>39±13</td>
<td>3.562±0.014</td>
</tr>
<tr>
<td>(up to 350 K)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Theoryb</td>
<td>70.6</td>
<td>3.5266</td>
</tr>
</tbody>
</table>

aRef. [18]
bRef. [27]

The shifts with temperature and the isotopic mass dependence of a given interband energy gap $E_g(T,M_j)$ are proportional to $\langle u^2 \rangle$ and can be described as [17, 18]

$$E_g(T,M_j) = E_B - a_B (M_{nat}/M_j)^{1/2} [1 + 2n_B],$$

where $E_B$ and $a_B$ are the unrenormalized (bare) gap and the renormalization parameter, respectively, at $T=0$. Here, $M_{nat}$ is the average mass of the natural isotopic composition of the semiconductor under consideration, $\Theta=\Theta (M_{nat}/M_j)^{1/2}$ and $\Theta$ is the Debye temperature of the semiconductor under consideration, in our case natural Si. The gap renormalization parameter $a_B$ includes two contributions: that of the electron-phonon interaction and the effect of thermal expansion. Numerical values of $E_B$, $\Theta$, and $a_B$ can be obtained by fitting Eq. (3) to the temperature dependence of the gaps.

In the low temperature limit ($T \ll \Theta$), Eq. (3) reduces to

$$E_g(T,M_j) = E_B - a_B (M_{nat}/M_j)^{1/2}. \quad (4)$$

Note that $a_B$ is the energy difference between the unrenormalized gap ($M_j \to \infty$) and the renormalized value [1, 17]. A similar expression can be written for the Lorentzian broadening parameter $\Gamma$ [9, 10].

In the high-temperature regime ($T \gg \Theta$), Eq. (3) can be written as

$$E_g(T,M_j) = E_B - 2T \frac{a_B}{\Theta}. \quad (5)$$

In this limit, $E_g(T,M)$ is independent of $M_j$ [1, 17]. The extrapolation of Eq. (5) to $T=0$ can be used to determine the unrenormalized gap energy $E_B$, i.e., the value that corresponds to atoms in fixed lattice positions without vibrations, from the measured temperature dependence of $E_g(T)$ in the high temperature (i.e., linear in $T$) region.

Using Eq. (4) the difference in energy $\Delta E_g$ between a given energy gap in isotopically pure and natural material can be written as

$$\Delta E_g = E_g(M_j) - E_g(M_{nat}) = a_B \left[ 1 - \left( \frac{M_{nat}}{M_j} \right)^{1/2} \right], \quad (6)$$

with a similar expression for $\Gamma$. In principle, a reliable value of $a_B$ for a given gap can be obtained by fitting Eq. (3) from low temperatures, where Eq. (4) is valid, to temperatures high enough for the slope of Eq. (5) to be well defined.

In the present work we review the isotopic effects on the dielectric function $\varepsilon$ of Si around the E1 direct interband transitions. We discuss low-temperature (25 K) ellipsometric measurements for natural silicon $^{28}\text{Si}$, $M_{nat}=28.14$ amu) and for isotopically pure $^{28}\text{Si}$ and $^{30}\text{Si}$ samples from 3.1 to 3.7 eV. To determine the critical point parameters as accurately as possible we analyze the spectra in reciprocal (Fourier-inverse) [19-21] rather than direct (frequency) space. The reciprocal space approach has several advantages over the most commonly used direct space, because the information concerning the baseline, the physically relevant information, and the noise contributions appear separately in the low-, medium-, and high-index Fourier coefficients, respectively.

2. Reciprocal Space Analysis

We consider an optical spectrum consisting of $N$ data points $f_j$, $j=1, \ldots, N$, equally spaced in energy from $E_1$ to $E_f$, and with a periodicity such that $f_j = f_{j+N}$. For simplicity, we take $N$ to be an even number. Then the $f_j$’s can be represented by the Fourier series [19-21]

$$f(\theta) = A_0 + \text{Re} \left( \sum_{n=1}^{N/2} C_n e^{-i\theta} e^{in\theta} \right), \quad (7)$$

where $\theta$ and $E$ are related by

$$\theta = \frac{E - E_1}{E_s} \quad (8)$$

and the factor $E_s$ is given by

$$E_s = \frac{N(E_f - E_1)}{2\pi(N-1)}. \quad (9)$$

We assume that the lineshape of the optical structure in $\varepsilon$ due to the transitions associated with a critical point at energy $E_g$ is described in "direct" space by the Lorentzian [19]
L(E) = B(E) + Ce^{i\beta} (E - E_g + i\Gamma)^{-\mu}, \quad (10)

where B(E) is a slowly varying background function, C and \( \beta \) are amplitude and phase factors, and \( \mu \) is the order of the singularity. B(E) will only affect the lowest Fourier coefficients and can therefore be neglected in our analysis.

For the lineshape (Eq. (10)) and its \( \kappa^{th} \) derivative the complex Fourier coefficients are given by [19]

\[
C_n = \frac{C_n^{\kappa+1}e^{-n\delta}}{E_s^{\mu+\kappa}\Gamma(\mu)},
\]

\[
\varepsilon_n = \frac{(\mu - \kappa)\pi}{2} - \beta + n\theta_g,
\]

where \( \delta = \Gamma/E_s \), \( \theta_g = (E_g - E_i)/E_s \), and \( \Gamma(\mu) \) is the gamma function.

According to the shift theorem [22], if the Fourier inversion origin is moved from \( \theta = 0 \) to \( \theta = \theta_0 \) the new Fourier coefficients are related to the unshifted ones by [19]

\[
C_n' = C_n,
\]

\[
\varepsilon_n' = \varepsilon_n - n\theta_0. \quad (12)
\]

From Eqs. (11) and (12) it is clear that if \( \theta_0 = \theta_g \), \( \varepsilon_n' \) will exhibit no dependence on \( n \), in which case \( E_g \) is given by \( E_g = \theta_0 E_s + E_i \).

The relative changes in \( E_g \) and \( \Gamma \) that result from a perturbation such as a change in isotopic mass can be expressed in a very simple way. From Eq. (11) we have

\[
\Delta \ln(C_n) = -\frac{\Delta \Gamma}{E_s} = -\frac{\Gamma - \Gamma_{\text{ref}}}{E_s} n
\]

\[
\Delta \varepsilon_n = \frac{\Delta E_g}{E_s} = \frac{E_g - E_{g,\text{ref}}}{E_s} n. \quad (13)
\]

At this point, it is illustrative to compare the Lorentzian and Gaussian lineshapes in reciprocal space. In direct space a Gaussian is given by:

\[
G(E) = \frac{-iCe^{i\beta}}{\Gamma} e^{-[(E - E_g)/\Gamma]^2}, \quad (14)
\]

and in reciprocal space by [21]

\[
C_n = \frac{1}{\sqrt{\pi}} \frac{Cn^{\mu+\kappa-1}e^{-\mu^2/2}}{E_s^{\mu+\kappa+\Gamma(\mu)}},
\]

\[
\varepsilon_n = \frac{(\mu - \kappa)\pi}{2} - \beta + n\theta_g. \quad (15)
\]
Figure 3. Natural logarithm of the amplitudes $C_n$ of the Fourier transform coefficients of the data of Fig. 1, together with best fit straight lines to the average dependences over coefficients 1 through 10. The number given for each sample are the values of $\Gamma$ (in meV) and their statistical uncertainties as calculated from the slopes of these lines. The white-noise region appears for $n \geq 11$.

Figure 4. Same as Fig. 3, but for the averages of the phases $\xi_n$ of $\varepsilon_1$ and $\varepsilon_2$ for inversion origins at 3.45 and 3.47 eV, as indicated.

From Eqs. (11) and (15) we note that $\ln(C_n)$ is proportional to $n$ and $n^2$ for Lorentzian and Gaussian lineshapes, respectively. Thus, by applying the inverse space formalism, the lineshape of a measured spectra can be determined directly from $\ln(C_n)$.

Reciprocal space has several advantages over the real space for the analysis of optical spectra. Among them we mention:

a) Base line effects, information and noise are separated in the low-, medium- and high-index Fourier coefficients, respectively.

b) The order $\kappa$ of a differentiation, appears in the reciprocal space as the factor $(n/E_s)^\kappa$ and the phase shift $-\kappa \pi/2$ in $C_n$ and $\xi_n$, respectively.

c) The energy of the gap can be estimated by shifting in energy the measured spectrum to obtain a phase $\xi_n$ independent of $n$ (Eq. (12)).

d) Differences in gaps and broadening parameters between samples of different isotopic composition can be determined accurately using reciprocal space. This procedure can be applied also to the determination of relative gap shifts in systems such as quantum wells, mismatched heterostructures, composition in ternary systems, etc.

e) The calculation of relative changes of gaps and broadening parameters is reduced to obtaining the slope of $\Delta \xi_n$ and $\Delta \ln(C_n)$, respectively. Equation (13) will be used to determine the relative dependence of $E_g$ and $\Gamma$ on isotopic composition.

f) Line shapes such as Lorentzian and Gaussian can be directly recognized.

g) The approach works fine only for one or two Lorentzians. The present formalism was specified to only one Lorentzian, however, if the spectrum under analysis consists of two Lorentzian transitions the formalism can be also applied [21].

3. Experimental results

In Ref. [14] three $\langle 111 \rangle$ oriented samples were investigated: natural Si ($M_{nat} = 28.14$) [23] and two nearly isotopically pure crystals, $^{28}$Si and $^{30}$Si. The details of the growth process can be found in Ref. [14]. The experiments were performed with a rotating analyzer ellipsometer [24], attached to a cryostat [14], with the samples at $T = 25 \pm 5$K. At this temperature Eq. (4) is valid since $\Theta \approx 650$K for Si. The surfaces were not prepared in any special manner prior to the measurement. Accordingly, the $\langle \varepsilon \rangle$ spectra were corrected for the presence of SiO$_2$ layers [14] by applying the three-phase (bulk-oxide-ambient) model [25].

The $\varepsilon$ spectra of the $^{28}$Si, $^{29}$Si and $^{30}$Si samples at 25 K from 3.1 to 3.7 eV are shown in Fig. 1 as triangles, diamonds, and circles, respectively. This energy range contains two critical points, $E_0$ and $E_1$ [18], as evidenced by the dominant peak and the shoulder in the $\varepsilon_1$ spectrum of Fig. 1. Actually, the $E_1$ transition consists of an $E_1$ and an $E_1 + \Delta_1$ component, but since their separation, $\Delta_1 = 0.03$ eV, is far too small to be resolved, both transitions will be
treated as a single critical point. Considering also that the $E_1$ transitions are stronger than the $E_{0'}$, the transitions can be treated as a single critical point and the reciprocal space formalism developed for only one critical point can be applied. The information obtained by this procedure should be mainly related to the $E_1$ transitions.

The most common method to obtain precise values for parameters such as $E_g$ and $\Gamma$ from optical spectra, is based on the fit of Eq. (10) to the second derivatives of the measured spectra with respect to energy. However, the noise/signal ratio increases in the derivative spectra and, in general, smoothing procedures are necessary to filter the noise with the consequence that the resolution and the information contained in the spectra deteriorate. This can be seen in Fig. 2, where we took the second energy derivatives of spectra of Fig. 1. The spectra were smoothed by using the Savitzky-Golay filter method \[26\]. The spectra look rather similar for the three samples and it is not possible, for instance, to see any difference among these gaps by fitting the second derivative of Eq. (10). Thus, in order to obtain reliable information we will analyze the spectra in reciprocal space.

Figure 3 shows the dependence of $\ln(C_n)$ on $n$ for each of the three samples investigated, as determined from both $\varepsilon_1$ (squares) and $\varepsilon_2$ (circles), along with best-fit straight lines (dashed) of the coefficients 1 through 10. It is seen that the data up to $n = 10$ are rather well represented by straight lines, showing that, to the accuracy of the data, the critical points are indeed Lorentzian \[21\]. The $\Gamma$ values determined from the slopes of the best-fit lines are $66.7 \pm 3.1$, $64.7 \pm 2.8$, and $62.6 \pm 3.4$ meV for the $^{28}$Si, $^{nat}$Si and $^{30}$Si samples, respectively.

Results for $\xi_n$, taking inversion origins of 3.45 and 3.47 eV, are shown in Fig. 4. The presence of two critical points manifests itself in the beating oscillations, which have a period $\Delta n$ of about 8, indicating a critical point separation of about 85 meV. Again, the white-noise region begins at about $n = 11$ and extends to higher values. The average slopes of the phases are positive and negative for the inversion origins of 3.45 and 3.47 eV, respectively. This places, according to the shift theorem, $E_g$ for the $E_1$ transitions somewhere between these energies. The gap energy of 3.454 ± 0.002 eV at 30 K reported by Lautenschlager et al \[18\] is thus in good agreement with the value determined by using the inverse space method.

The uncertainty in $E_g$ obtained from $\xi_n$ is substantially larger than the expected isotope shift. However, this uncertainty results mainly from the interference oscillations between the contributions of the $E_{0'}$ and $E_1$ constituents of the $E_1$ structure. Assuming that these have the same isotopic mass dependence, we can eliminate the associated inaccuracies by subtracting the $\ln(C_n)$ data (see Fig. 3) of the natural silicon sample from those of the other two specimens.

We show in Figs. 5 and 6 the dependence of $\Delta \ln(C_n)$ and $\Delta \xi_n$ on $n$. According to Eq. (13) the slope of $\Delta \ln(C_n)$ and $\Delta \xi_n$ is proportional to the difference in $\Gamma$ and $E_g$, respectively. From the slope of Fig. 5 determined by the
points 1 through 8 (dashed straight lines) we find that the broadening parameter of the $^{28}\text{Si}$ sample is smaller than that of the natural sample by $0.9 \pm 0.3$ meV. The broadening of the $^{30}\text{Si}$ and $^{nat}\text{Si}$ samples is the same to within experimental uncertainty.

Repeating the $\zeta_0$ analysis for points 1 through 8 (Fig. 6), we found that the critical point energy of the $^{30}\text{Si}$ sample is higher than that of the $^{nat}\text{Si}$ sample by $3.7 \pm 0.7$ meV. The $^{28}\text{Si}$ and $^{nat}\text{Si}$ samples exhibit essentially the same value of $E_g$; the higher value of $0.25 \pm 0.3$ meV obtained for the $^{28}\text{Si}$ sample is smaller than the statistical uncertainty. A theoretical value for $\Delta E_1 = 1.2$ meV has been obtained from pseudopotential calculations of the temperature dependence of the gap [27]. This value is close to our experimental result. Experiment and theory support the fact that the differences between the results for the $^{28}\text{Si}$ and natural samples should be negligible. We note that it would be impossible to obtain this accuracy by direct-space analysis.

From the values of $\Delta E_1$ and Eq. (6), we can calculate values of $a_0$ and $E_g$. These results and those calculated from literature data for $E_g(T)$, are given in table 1. The values of $\Delta E_1$ include both components of the gap shift, the electron-phonon interaction and the thermal expansion effect. The thermal expansion component can be estimated by writing Eq. (1) in the form $\Delta E_1^{\text{vol}} = (dE_1/d\ln V)(d\ln V/dM)\Delta M$ and the values of $dE_1/d\ln V = 5.1$ eV [28], $d\ln V/dM = 6.6 \times 10^{-5}$/amu [7], and $\Delta M = 1.8$ amu. This gives a value of $\Delta E_1^{\text{vol}} = 0.61$ meV.

4. Conclusions

We have discussed the effects of isotopic mass on the $E_1$ critical point parameters of Si by analyzing ellipsometric data for the dielectric function taken from $3.1$ to $3.7$ eV at $25$ K. To obtain the maximum possible accuracy we determined the critical point parameters by reciprocal space analysis rather than using the more common direct (frequency) space formalism. We show that the gap shift with isotopic mass between the $^{30}\text{Si}$ and $^{nat}\text{Si}$ crystals is $3.7 \pm 0.7$ meV.

Acknowledgments

We would like to thank D.E. Aspnes for having introduced us to the reciprocal space method of analysis. Thanks are also due to D. Rönnow, T. Ruf, M. Konuma and S.D. Yoo for their collaboration in earlier versions of this work.

References