High-Resolution Study of the One-Electron Spectrum of Si†

LUIS R. SARAVIA AND DAVID BRUST
Department of Physics, Northwestern University, Evanston, Illinois 60201
(Received 27 December 1967; revised manuscript received 18 March 1968)

In the present paper we reexamine the one-particle interband transition contribution to the optical absorption spectrum in Si. In particular we introduce a $\mathbf{k}\cdot\mathbf{p}$ extrapolation of the pseudopotential method. This allows us to truncate greatly the secular equations appearing in that method. Combining this with a method of zone integration proposed by Gilat and Raubenheimer for phonon spectra, we get $\epsilon_2(\omega)$ curves which show an improvement in computational resolution of $\sim 10^2$. The high speed of our present techniques allows us to examine carefully several models which have been proposed to explain the 3.4-eV fundamental edge in Si. In particular we find dramatic changes in the line shape which depend on the relative position of $\Gamma_{25'} \rightarrow \Gamma_{15}$ and $L_{3'} \rightarrow L_1$. For all models examined, an extremely complex nest of critical points is predicted near the fundamental edge. This indicates that efforts to interpret the fundamental edge from a model of one or two special points in the zone probably cannot be successful.

I. INTRODUCTION

IN the five years since we began our examination of the ultraviolet spectra of semiconductors[1] a great deal of experimental progress has been made. At the time of the previous work the reflectivity spectrum had been taken for a number of semiconductors which showed very rich structure. This could be understood at the time (at least nearly within experimental error). The analysis indicated that in some instances the theoretical, and corresponding experimental, structure could be interpreted in terms of Van Hove singularities originating from critical points in the zone. In other cases, specific volume effects had to be invoked in the interpretation.

The calculations furthermore indicated that some of the structure could be closely identified with symmetry directions in the Brillouin zone (BZ). For example, our study of the energy contours in Ge demonstrated that the origin of the well resolved doublet near 2.2 eV was from states that lay very near the Λ axis ([111] direction), and corresponded to $\Lambda_3 \rightarrow \Lambda_1$ transitions (single group notation). It was pointed out[2] that in general for an M_1 type of critical point the "open" nature of the contours prevented identification with a particular location in the BZ, except for rough arguments. However, because of the cylindrical geometry of the isogap contours, the specification of the Λ axis for the 2.2-eV peak in Ge was quite precise.[2] This was strikingly confirmed by Cardona's work[3] and also by Gerhardt.[4] Further confirmation came through calculation of the deformation potentials[2] and comparison with the effect of hydrostatic pressure on the optical edges.[2,4,5]

The recent experiments utilizing the electro-optic effect[3,6–8] and the piezo-optic effect[4,9,10] have seriously probed the details of the known optical peaks and edges, and discovered previously undetected structure. Because of the inherent uniaxial character of these experiments, one can now, under very favorable circumstances experimentally determine not only the location in k space of the transitions involved, but also the symmetry of the states. Again, of course, one always has to check that the interpretation based on a few points in the zone is really meaningful in terms of the observation being made. For example, the original theory of the Franz-Keldysh effect assumed that an M_0 critical point (c.p.) existed.[11–14] More recent theories have extended this to all types of c.p.[15,16] Still one must be certain that the theory applies to the experiment in question. In particular, one must be sure the energy contours show a clear c.p. over an energy range larger than any broadening which may exist in the data (either inherent in the sample or the experiment).

For the most part, the new experimental studies give results for Si and Ge that are in good accord with our earlier calculations. In the case of Si, however, the very prominent direct absorption edge near 3.4 eV has thus far defied a complete interpretation. Our own earlier work[1] showed a cluster of c.p.'s all with energy fairly near to the $\Gamma_{25'} \rightarrow \Gamma_{15}$ gap, although some of these lie far from the zone center. The experiments of Gerhardt[4] and also of Gobeli and Kane[10] indicate that in particular the Δ axis probably makes a significant contribution to the density of states near this peak. This interpretation of the edge requires that $\Gamma_{25'} \rightarrow \Gamma_{15}$ be quite close to

† Research supported in part by the National Science Foundation and the Advanced Research Projects Agency of the Department of Defense through the Northwestern University Materials Research Center.

[1] D. Brust, Phys. Rev. **134**, A1337 (1964).
[2] D. Brust and L. Liu, Phys. Rev. **154**, 647 (1967).
[3] M. Cardona, K. L. Shaklee, and F. H. Pollak, Phys. Rev. **154**, 696 (1967).
[4] U. Gerhardt, Phys. Rev. Letters **15**, 401 (1965).
[5] R. Zallen and W. Paul, Phys. Rev. **155**, 703 (1967).
[6] B. O. Seraphin and N. Bottka, Phys. Rev. **145**, 628 (1963).
[7] A. K. Ghosh, Phys. Letters **23**, 36 (1966).
[8] A. K. Ghosh (to be published).
[9] W. E. Engeler, H. Fritzsche, M. Garfinkel, and J. J. Tieman, Phys. Rev. Letters **15**, 142 (1965).
[10] G. W. Gobeli and E. O. Kane, Phys. Rev. Letters **15**, 142 (1965).
[11] W. Franz, Z. Naturforsch. **13**, 484 (1958).
[12] L. V. Keldysh, Zh. Eksperim. i Teor. Fiz. **34**, 1138 (1958) [English transl.: Soviet Phys.—JETP **7**, 788 (1958)].
[13] K. Tharmalingham, Phys. Rev. **130**, 2204 (1963).
[14] J. Callaway, Phys. Rev. **130**, 549 (1963).
[15] D. E. Aspnes, Phys. Rev. **147**, 554 (1966).
[16] D. E. Aspnes, Phys. Rev. **153**, 972 (1967).

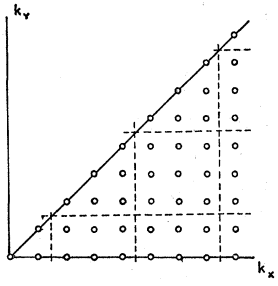

FIG. 1. Two-dimensional example of the arrangement of points in reciprocal space for the $\mathbf{k}\cdot\mathbf{p}$ interpolation method. At the central point of each dotted square the eigenvalues and eigenfunctions are found from the pseudopotential secular equation. The eigenvalues at the other points inside the square are determined by the $\mathbf{k}\cdot\mathbf{p}$ extrapolation.

3.4 eV. Herman *et al.*,[17] however, in their recent calculations find a value of $\Gamma_{25'} \rightarrow \Gamma_{15} < 3.0$ eV. Their conclusion is that the $L_{3'} \rightarrow L_1$ transition is primarily responsible for the observed structure. Unfortunately, no matter which interpretation one adopts, the existing calculations are inadequate to make a sufficiently detailed analysis of the single-particle energy contours, c.p.'s, and concommitant structure, to settle this question. All of these computations[1,18,19] on Si are generally limited to $\sim 10^3$–10^4 independent points in the reduced zone. This is insufficient for a careful examination of the energy spectrum over a small range ~ 0.2 eV. We would like very much to completely resolve the one-particle spectrum so as to be able to separate out any residual effects of electron-hole interaction leading to quasiexciton effects.[20]

In the present paper, therefore, we describe a method whereby one can, with a minimal amount of computer time, get an extremely accurate spectrum. We estimate that to do a comparable job by Monte Carlo[21] methods one would have to have eigenvalues evaluated at considerably more than 10^6 independent points. As a result of this improvement, we can test all of the competing models quite easily. The predicted spectrum of each model is shown in the succeeding sections. The present study is intended to examine fine details which were beyond the earlier analyses by increasing the theoretical resolving power by a factor of $\sim 10^2$. In particular, it is found that some of the models predict features which are not experimentally observed.

II. $\mathbf{k}\cdot\mathbf{p}$ EXTRAPOLATION OF THE PSEUDOPOTENTIAL

As will become apparent in Sec. III, it is necessary for our present studies to have exact eigenvalues at ~ 2000 points in the modular wedge.[22] For this purpose we have found that an application of $\mathbf{k}\cdot\mathbf{p}$ theory greatly expedites the work. First, one computes eigenvalues and eigenfunctions at a small number of points on a coarse cubic mesh in the modular wedge by solving the secular equations generated by the semiempirical pseudopotential approach[1] (~ 100 points in our case). Each of these points is taken as the center of a small cube and surrounded by a refined mesh of points lying on a simple cubic array. The eigenvalues are then determined on this fine-grained mesh by the $\mathbf{k}\cdot\mathbf{p}$ method described in the next paragraphs. Figure 1 illustrates the discussion so far on a two-dimensional plot.

Suppose the central point of the cube has reduced wave vector $\mathbf{k}_0$, and we wish to determine the eigenvalue spectrum at a neighboring point of the little cube situated at

$$\mathbf{k}=\mathbf{k}_0+\delta\mathbf{k}. \tag{1}$$

The pseudopotential wave functions at $\mathbf{k}$ may be written in Bloch form as

$$\varphi_{n,\mathbf{k}}=e^{i\mathbf{k}\cdot\mathbf{r}}U_{n,\mathbf{k}}(\mathbf{r}). \tag{2}$$

The equation satisfied by $U_{n,\mathbf{k}}$ is

$$(H_0(\mathbf{k}_0)+H'(\delta\mathbf{k}))U_{n,\mathbf{k}}=E_{n,\mathbf{k}}U_{n,\mathbf{k}}, \tag{3}$$

where

$$H_0(\mathbf{k}_0)=H_{\text{eff}}+(\hbar^2k_0^2/2m)+(\hbar/m)\mathbf{k}_0\cdot\mathbf{p}. \tag{4}$$

As usual, we have

$$H_{\text{eff}}=\text{kinetic energy}+\text{pseudopotential}. \tag{5}$$

Thus a reasonably complete set of $U_{n,\mathbf{k}_0}$'s can be used to generate the eigenvalue spectrum at the neighboring points. The $U_{n,\mathbf{k}_0}$ themselves are determined by solving $\sim 30\times 30$ secular equations, which means that we have approximately the lowest 30 eigenstates expanded using the first 30 or so plane waves.

We then treat the terms in $\delta\mathbf{k}$ in Eqs. (3)–(5) as a perturbation utilizing the Löwdin technique. One then has to diagonalize the matrix equation

$$|P_{nm}'(\delta\mathbf{k})-E_{n,k}'\delta_{nm}|=0, \tag{6}$$

where

$$E_{n,\mathbf{k}}'=E_{n,\mathbf{k}}-E_{n,\mathbf{k}_0}-\hbar^2\delta k^2/2m-(\hbar^2/m)\mathbf{k}_0\cdot\delta\mathbf{k}. \tag{7}$$

In Eq. (6), δ_{nm} is the Kronecker δ and $n,m\leq N_1$, where N_1 is the number of functions $U_{n,\mathbf{k}_0}$ which we treat exactly in the Löwdin method. The quantity P_{nm}' is defined differently depending on whether it appears on or off the diagonal; specifically,

$$P_{nm}'=P_{nm} \quad \text{if} \quad n\neq m \tag{8}$$

and

$$P_{nm}'=P_{nm}+\sum_{l=N_1+1}^{N}\frac{P_{nl}^*P_{ln}}{E_{n,\mathbf{k}_0}+\hbar^2\delta k^2/2m+\hbar^2\mathbf{k}_0\cdot\delta\mathbf{k}/m-E_{l,\mathbf{k}_0}} \quad \text{if} \quad n=m, \tag{9}$$

with

$$P_{nm}=(\hbar^2/mi)\langle U_{n,\mathbf{k}_0}|\,\delta\mathbf{k}\cdot\nabla\,|U_{m,\mathbf{k}_0}\rangle. \tag{10}$$

N corresponds to the number of eigenfunctions deter-

[17] F. Herman, R. L. Kortum, C. D. Kuglin, and R. A. Short, *Quantum Theory of Atoms, Molecules, and the Solid State* (Academic Press Inc., New York, 1966), p. 381.
[18] E. O. Kane, Phys. Rev. **146**, 558 (1966).
[19] G. Dresselhaus and M. S. Dresselhaus (to be published).
[20] J. C. Phillips, *Solid State Physics* (Academic Press Inc., New York, 1966), p. 56.
[21] D. Brust, Phys. Rev. **139**, A489 (1965).
[22] For a description of the wedge module, see Ref. 1.

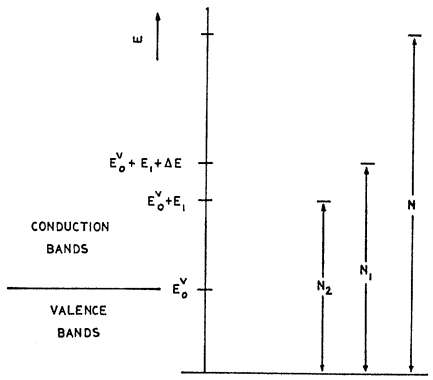

FIG. 2. Approximate relationship of the quantities entering the band calculation. N is the number of eigenvalues calculated by the pseudopotential approach. N_2 is the number of eigenvalues used in the computation of $\epsilon_2(\omega)$. N_1 is the size of the matrix treated exactly by the $\mathbf{k}\cdot\mathbf{p}$ extrapolation method.

mined at the point $\mathbf{k}_0$. Using the plane-wave expansion for the Bloch function, we have explicitly for the matrix element in Eq. (10)

$$\langle U_{n,\mathbf{k}_0}|\,\delta\mathbf{k}\cdot\nabla\,|U_{m,\mathbf{k}_0}\rangle = i\delta\mathbf{k}\cdot\sum_{p=1}^{N}\mathbf{K}_p V_n{}^{p*} V_m{}^{p}, \qquad (11)$$

with

$$U_{n,\mathbf{k}_0} = \sum_{p=1}^{N} V_p{}^{n}\exp i\mathbf{K}_p\cdot\mathbf{r}. \qquad (12)$$

This sum does not depend on $\delta\mathbf{k}$; hence only one evaluation of the matrix element must be performed for each cube. We note that for the diamond structure the coefficients $V_p{}^n$ are real; therefore, the secular equation (6) has only real matrix elements.

In setting up Eq. (6), we retain all terms $O(\delta\mathbf{k})$ on and off the diagonal, whereas terms $O(\delta\mathbf{k}^2)$ are retained only on the diagonal.

In order to determine the cutoff parameter N_1 which determines the number of Bloch states treated exactly in Eq. (6) the following procedure was used: First, it was determined over what range the dielectric constant was to be evaluated. Suppose we wish to find $\epsilon_2(\omega)$ in the range $0\leq\hbar\omega\leq E_1$. Then we must keep all the states $U_{n,\mathbf{k}_0}$ with $E_{n,\mathbf{k}_0}<E_1+E_0{}^V$ (taking $E_{1,\mathbf{k}_0}=E_0{}^V$). However, this is insufficient because we may neglect states with $E_{n,\mathbf{k}_0}$ only slightly larger than $E_1+E_0{}^V$, which would introduce large errors into our evaluation of the energy eigenvalues. This is solved by increasing the upper cutoff to $E_1+E_0{}^V+\Delta E$, where ΔE is taken as a predetermined constant. All of the remaining states with indices between N_1 and N are treated by perturbation theory. Figure 2 illustrates this discussion. Kane[23] has also used a method somewhat similar to the present one.

The convergence error in the present calculation has two sources. First of all, in solving a 30×30 secular equation with the pseudopotential we expect a maximum error ~0.05 eV (in general it will be less than this). We are mainly concerned with the discontinuity errors discussed in Ref. 1 because these will interfere with the smoothness of our computed bands. Fortunately, these are significant at only a few points in the zone, and any small spurious effects can easily be separated out. In practice these prove negligible.

The perturbation treatment of the $\mathbf{k}\cdot\mathbf{p}$ method introduces an additional error in the final eigenvalues. This depends on the size N_1 of the secular equation and on $\delta\mathbf{k}$. In general, we expect the error to increase as $(\delta\mathbf{k})^2$. Tests have been performed to determine the magnitude of these errors. In the present work the maximum value of $\delta\mathbf{k}=0.04167$ and N_1 varies between 8 and 12.

Some of the worst cases tested are shown in Fig. 3. For $\delta\mathbf{k}=0.125$, we see that all errors from the $\mathbf{k}\cdot\mathbf{p}$ secular equation are considerably less than 0.1 eV. Thus we expect the $\mathbf{k}\cdot\mathbf{p}$ errors in our evaluation to be <0.01 eV.

In the computation of ϵ_2, the matrix elements

$$|M_{m,n}(\mathbf{k})|^2 = |\langle U_{\mathbf{k}}{}^m|\,\nabla\,|U_{\mathbf{k}}{}^n\rangle|^2 \qquad (13)$$

are needed. They are calculated at each point k, using the eigenvectors obtained from the pseudopotential secular equation during the evaluation of the momentum matrix elements. Using the $\mathbf{k}\cdot\mathbf{p}$ method, we gain a large saving in computation time. In a typical calculation the final mesh is defined by an interval $1/24\times2\pi/a$ (1600 points in the modular part of the reciprocal cell) and the large cubes by an interval $\frac{1}{8}\times2\pi/a$ (100 points). The respective computing times for the two parts are 15 and 12 min, on a CDC 3400 computer using a FORTRAN program.

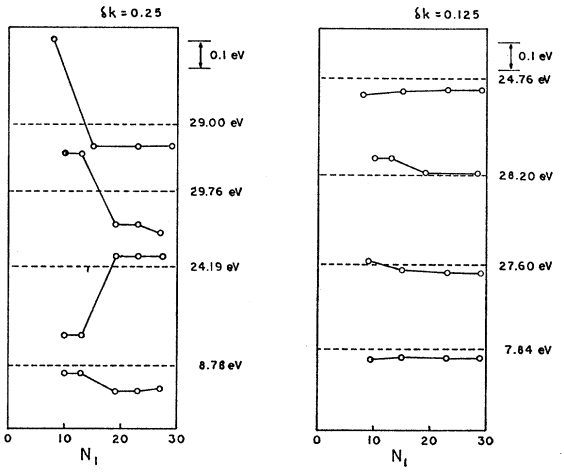

FIG. 3. Convergence properties of the $\mathbf{k}\cdot\mathbf{p}$ extrapolation method when the number N_1 of eigenvalues treated exactly is increased. Results for two values of $\delta\mathbf{k}$ are included. The dashed lines indicate the values obtained at the same point of reciprocal space by the pseudopotential approach.

[23] E. O. Kane, *Semiconductors and Semimetals* (Academic Press Inc., New York, 1966), p. 75.

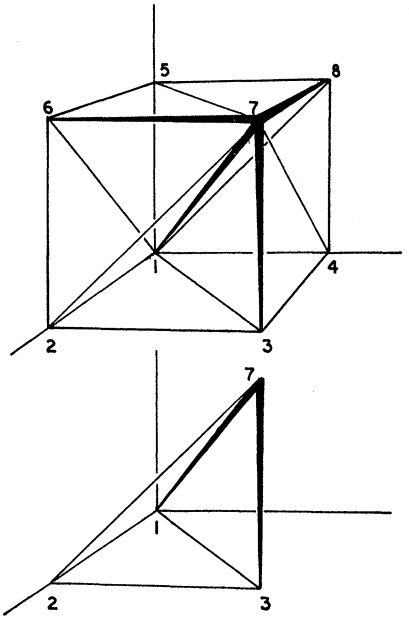

FIG. 4. Division of a cube in reciprocal space into tetrahedrons. Line 1–7 points along the ΓL direction.

III. CALCULATION OF ε_2

The band structure is used to compute the imaginary part of the dielectric constant ϵ_2 according to the formula[1]

$$\epsilon_2(\omega)=\frac{4\pi^2e^2\hbar}{3m^2\omega^2}\sum_{ij}\int_{\mathrm{BZ}}\frac{2}{(2\pi)^3}\times\delta(\omega^{i,j}(\mathbf{k})-\omega)\,|M_{ji}(\mathbf{k})|^2d^3k. \quad (14)$$

The summation is over empty conduction band states i and filled valence band states j.

The integration over the BZ is performed by a linear interpolation on $\omega^{ij}(\mathbf{k})$ as proposed by Gilat and Raubenheimer[24] for phonon spectrum, although our method used for the interpolation is different.

The methods for evaluation of the energy bands give the values of $\omega^{ij}(\mathbf{k})$ in points of a cubic mesh in $\mathbf{k}$ space. In order to perform the interpolation, each cube of the mesh is divided in six equal tetrahedrons, as shown in Fig. 4. All of them have in common the edge 1–7 and the other edges are edges or diagonals of the faces of the cube. The tetrahedron 1-2-3-7 is an example.

In each tetrahedron, $|M_{ij}|^2$ is supposed constant and $\omega^{ij}(\mathbf{k})$ is replaced by a linear function of $\mathbf{k}$. This function has four constants which are completely determined by fitting the given values of $\omega^{ij}(\mathbf{k})$ at the four vertices. The constant-energy surfaces are replaced by planes. The integration is performed by calculating the volume between two isogap planes defined by ω and $\omega+\Delta\omega$ and dividing by $\Delta\omega$. In this way $\epsilon_2(\omega)$ is computed in intervals of $\Delta\omega$.

[24] G. Gilat and L. J. Raubenheimer, Phys. Rev. 144, 390 (1966).

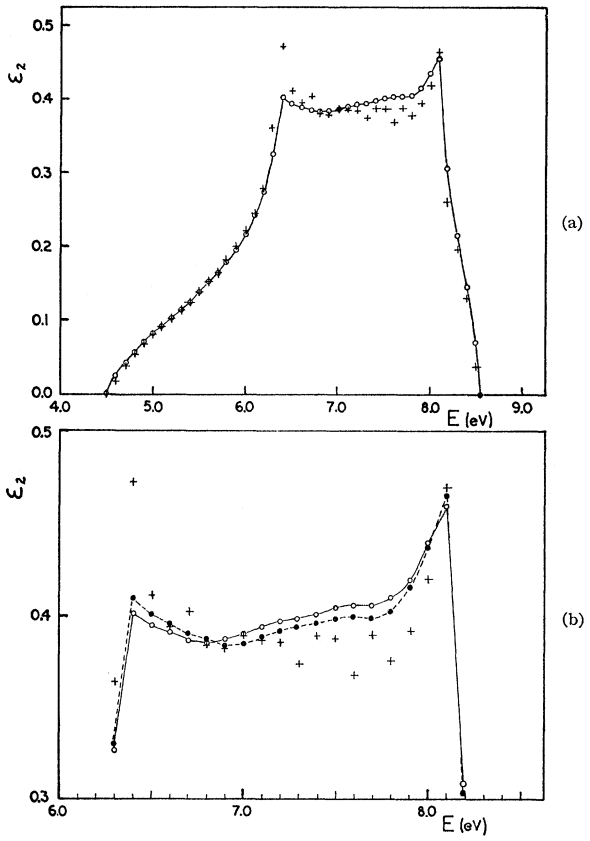

FIG. 5. (a) Joint density of states for the tight binding s band. (b) Results for an fcc lattice corresponding to three meshes with different number of points along the ΓX direction are compared: (i) 10 points (crosses), (ii) 24 points (dots), (iii) 50 points (circles). (a) Shows the complete curve for two cases; (b) shows in detail the region of largest discrepancies for the three cases (note the scale of ϵ_2).

The method has two advantages. First, the planes which form the boundary of the BZ (with the exception of the hexagonal face $KLUW$) are also boundaries for the tetrahedrons of those cubes that cross the faces of the BZ. The elimination of the contribution from the volume of the cubes which is outside the BZ is simply performed by discarding the tetrahedrons which are not part of the BZ. Therefore, the introduction of approximate weighting factors is not necessary. Because of the symmetry of the zone, for each cube crossing the hexagonal face there is another one with the same values of $\omega^{ij}(\mathbf{k})$ and in a complementary position. The contribution from both cubes is computed exactly by taking a $\frac{1}{2}$ weighting factor. As a second advantage, the linear interpolation in adjacent tetrahedrons fits continuously across the common face. The discontinuities that appear in the faces of the cubes, when the same linear interpolation is used in the whole cube, are eliminated.

$\Delta\omega$ can be selected in the program, and the computing time is approximately proportional to the number of points in the $\epsilon_2(\omega)$ histogram. The values of $\epsilon_2(\omega)$ can be

TABLE I. Pseudopotential form factors and energy gaps of some of the principal transitions for the four models. The potential coefficients are in Rydbergs and the gaps are in eV.

	V_{111}	V_{220}	V_{311}	$\Gamma_{25'}$-Γ_{15}	$L_{3'}$-L_1	X_4-X_1
Model I	−0.2100	0.0400	0.0800	3.44	3.13	3.95
Model II	−0.2100	0.0500	0.0800	3.39	3.37	4.05
Model III	−0.2100	0.0600	0.0800	3.34	3.60	4.16
Model IV	−0.1920	0.0434	0.0896	3.39	3.33	3.82

obtained in any interval (ω_1,ω_2), allowing the study of interesting parts of the dispersion curve in detail. Finally, the contribution to $\epsilon_2(\omega)$ from partial regions in the BZ can be calculated, allowing the study of the contribution from individual points, lines, or volumes.

In order to get a good evaluation of $\epsilon_2(\omega)$, the cubic mesh must contain a large number of cubes inside the BZ. The convergence of the method was tested in a simple model where the energies of the interband transition are given by a tight-binding s-band expression for an fcc lattice[25];

$$E_k = 7.0 - 1.0[\cos\pi k_x \cos\pi k_y + \cos\pi k_y \cos\pi k_z + \cos\pi k_z \cos\pi k_x] + 0.2[\cos 2\pi k_x + \cos 2\pi k_y + \cos 2\pi k_z]. \quad (15)$$

$\epsilon_2(\omega)$ was calculated from three meshes corresponding to 10, 24, and 50 points along the ΓX direction. The complete results for $\epsilon_2(\omega)$ are shown in Fig. 5. In particular, we note the rapid improvement in convergence with mesh fineness. The values of $\epsilon_2(\omega)$ near the critical points are the most delicate ones. Nevertheless, differences between the calculation with 24 and 50 points are less than 1% for values which are not near the c.p.'s and less than 2% near them. A typical calculation using 400 values of ω and 1600 points unrelated by symmetry inside the BZ (24 points along the ΓX direction) takes 4 min per interband pair on a CDC 3400 computer using a FORTRAN program. We note that

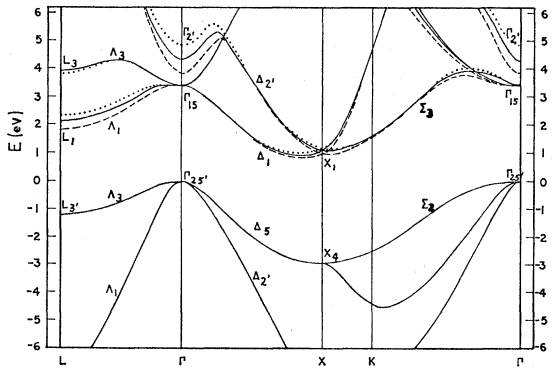

FIG. 6. Energy bands along some principal symmetry lines for models I (dashed lines), II (solid lines), and III (dotted lines). They were computed in a mesh having 25 points along the Δ axis (approximately 1600 points in the asymmetric part of the BZ).

[25] N. F. Mott and H. Jones, *The Theory of the Properties of Metals and Alloys* (Dover Publications, Inc., New York, 1936).

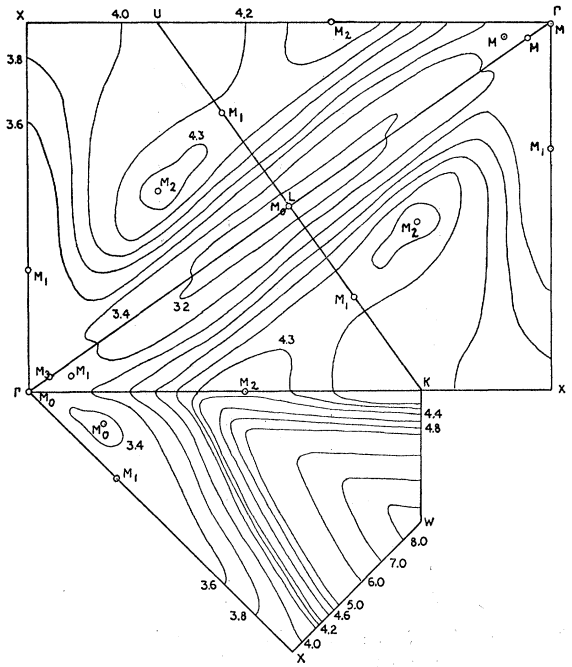

FIG. 7. $E_{4,5}(\mathbf{k})$ contours for model I in the ΓKWX, ΓKL, and ΓXUL planes (in eV).

Higginbotham *et al.* use a method based directly on the Gilat-Raubenheimer approach in their work.[26]

IV. RESULTS

Several authors have discussed the interpretation of the 3.4-eV peak in Si, especially the relative positions of the values for the transition energies $\Gamma_{25'}$-Γ_{15} and $L_{3'}$-L_1. The energy contours have also been studied by Kane,[18] who determined the contribution from different parts of the BZ to $\epsilon_2(\omega)$, obtaining interesting results for the 4.5-eV peak. The methods of calculation discussed in this work allow a detailed study of those peaks in terms of the c.p.'s of the interband transitions.

In order to discuss the influence of the relative positions of the transition energies $\Gamma_{25'}$-Γ_{15} and $L_{3'}$-L_1 in $\epsilon_2(\omega)$, four different models for Si were calculated. The pseudopotential form factors used in each case and the corresponding gaps of some of the principal transitions are listed in Table I.

The energy bands along the Λ, Δ, and Σ symmetry lines, the $E_{4,5}$ energy contours in the ΓKWX, ΓLK, and ΓLUX planes, and the contribution of $E_{4,5}$ to $\epsilon_2(\omega)$ are shown in Figs. 6–10 for the first three models. In Fig. 11, we plot $\epsilon_2(\omega)$ for the fourth model.

Model II will be studied first. The structure of ϵ_2 near the 3.4-eV peak comes mainly from two regions near the Γ and L points. In order to determine their relative importance, the contribution from three partial regions in the BZ were computed separately. The limits of these

[26] C. W. Higginbotham, F. H. Pollak, and M. Cardona, Solid State Commun. 5, 503 (1967).

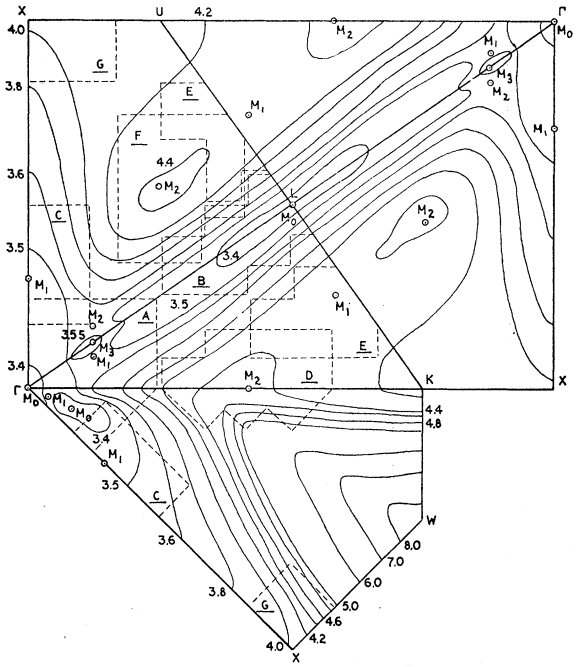

FIG. 8. $E_{4,5}(\mathbf{k})$ contours for model II in the ΓKWX, ΓKL, and ΓXUL planes (in eV). Special volumes which have been studied in detail are marked out by dotted lines in the plane.

regions in the symmetry planes are marked in dotted lines in Fig. 8. Region A surrounds the point Γ and zone B surrounds a large part of the Λ axis including the L point. Volume C encloses a c.p. on the ΓX axis. The results of the computation are shown in Fig. 12.

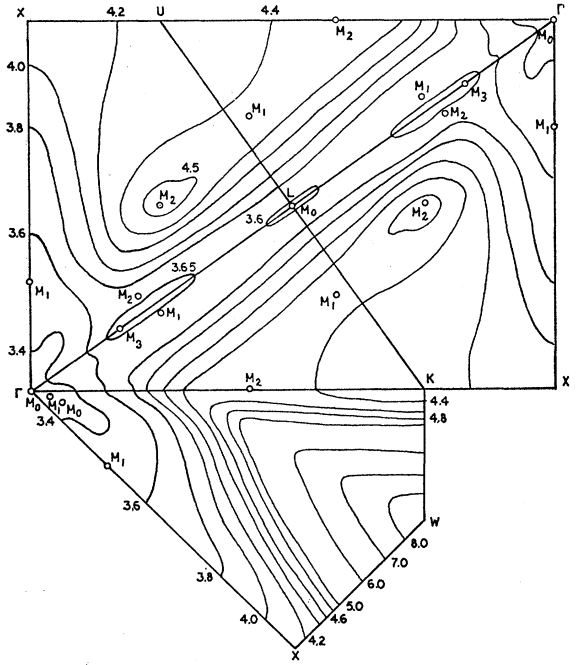

FIG. 9. $E_{4,5}(\mathbf{k})$ contours for model III in the ΓKWX, ΓKL, and ΓXUL planes (in eV).

The strength of the contribution of each c.p. to $\epsilon_2(\omega)$ depends on the particular form of the constant-energy surfaces near the point, but a general remark can be made: The strength depends on the number of points in the star of the k vector. For example, since there are only six points in the star of a point on the Λ axis, the weighting factor is six. For a point in the ΓLK plane the factor is 24.

The main feature in B is a step in $\epsilon_2(\omega)$ ranging from 3.37 to 3.41 eV. In the energy contours there is a segment along the Λ line, from the L point down to $(\frac{1}{3}, \frac{1}{3}, \frac{1}{3})$, where the function is practically constant along the line. We can interpret this as a minimum along the line,

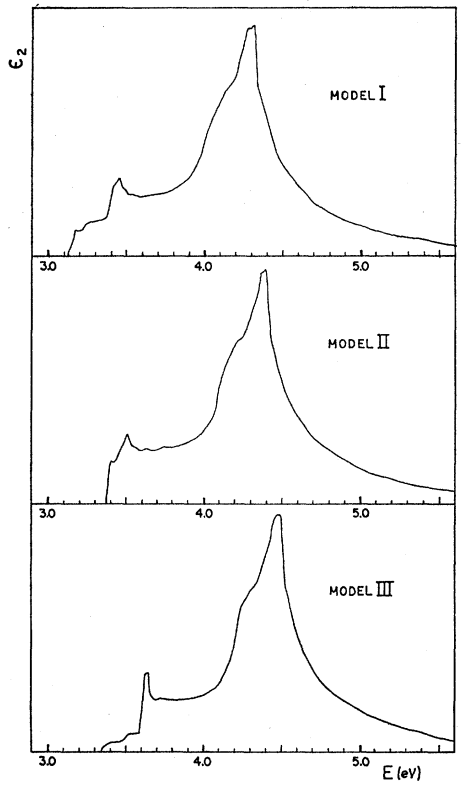

FIG. 10. $\epsilon_2^{4,5}(\omega)$ for models I, II, and III. It was calculated at intervals of 0.01 eV.

producing a step in the density of states. In the limit that the Λ axis does show an entire line segment along which the function is a minimum, we have a two-dimensional c.p. of the type discussed by Kane. Since the values along the line really change from 3.37 eV at the L point to 3.41 eV at $(\frac{1}{3}, \frac{1}{3}, \frac{1}{3})$, the step is slightly canted. The main features for A are a step at 3.40 eV and a peak at 3.52 eV. In order to explain the step we observe that on the ΓXWK plane there are M_0 c.p.'s at (0, 0, 0) and $(\frac{1}{8}, 1/24, 0)$ with energies 3.39 and 3.38 eV. On a line connecting them there is an M_1 c.p. with an energy of 3.54 eV. In order to explain the peak we see that on the ΓKL and ΓXUL planes there exist an M_1 and an M_2 c.p. at $(\frac{1}{8}, \frac{1}{8}, \frac{1}{12})$ and $(\frac{1}{6}, \frac{1}{8}, \frac{1}{8})$, both with

energies of 3.51 eV, respectively. Because the energies are almost the same, the combination produces a sharp peak. This somewhat resembles a two-dimensional logarithmic singularity. The contribution from zone C is small and the critical point at (7/24, 0, 0) does not produce any detectable change in the slope.

For model I the steps associated with the Γ and L points appear separately in Fig. 10. The L step is at 3.13 eV. The M_0 point at (5/24, 1/24, 0) in the ΓXWK plane has an energy of 3.38 eV. It is quite extensive and determines the shape of the second step. The structure associated with the Λ line has approached the Γ point. The M_3 critical point on the Λ line is at (1/24,1/24,1/24) and has an energy of 3.49 eV. The saddle point in the ΓKL plane still exists with an energy of 3.47 eV at $(\frac{1}{12}, \frac{1}{12}, 1/24)$. The other one has disappeared. The contribution of this point is a little displaced in respect to the peak and is not so important as in the other cases. On the contrary, the zone related to the c.p.'s on the Δ axis is quite flat. It goes from (0, 0, 0) to (11/24, 0, 0), with energies ranging from 3.42 up to 3.45 eV and its contribution to $\epsilon_2(\omega)$ determines the small peak.

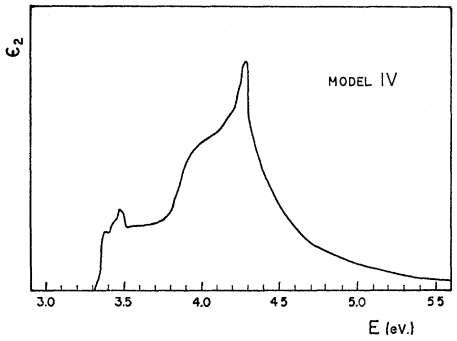

FIG. 11. $\epsilon_2^{4,5}(\omega)$ for model IV. It was calculated at intervals of 0.01 eV.

For model III the Γ point appears separately at 3.34 eV in Fig. 10. The M_0 point in the ΓXKW plane at $(\frac{1}{12}, 1/24, 0)$ still exists but has approached the Γ point, and the total contribution from this zone is considerably reduced. The small change in slope at 3.53 eV is produced by the c.p. at (7/24, 0, 0). The main feature which we wish to examine is the asymmetric peak near 3.6 eV in Fig. 9 for model III. The threshold comes from the M_0 c.p. at L with an energy of 3.60 eV. The very sharp peak is due to the M_1 and M_2 c.p.'s near the Λ axis. These dominate a much larger portion of the BZ than in the previous models.

It must be remarked that in these models we have just tried to change the position of the L, Γ, and X points with respect to one another. As a result, the peaks in the various models are somewhat shifted with respect to one another. Our main concern here is with line shape rather than the absolute position of the peaks.

The other interesting peak is at 4.5 eV. Let us examine model II. Figure 8 shows that the contribution to this peak comes from a zone that begins with an M_2 c.p. at

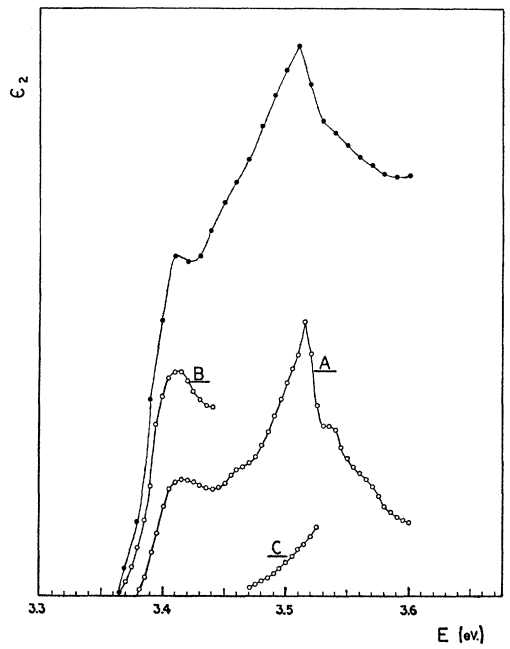

FIG. 12. The contribution to $\epsilon_2^{4,5}(\omega)$ for model II from (a) the regions A, B, and C indicated in Fig. 8 (circles) and from (b) the whole wedge (dots).

(5/12, 5/12, 0) on the Σ axis, contains an M_1 c.p. at $(7/12, 7/12, \frac{1}{4})$ near the KL line, and ends at an M_2 c.p. at $(13/24, \frac{1}{4}, \frac{1}{4})$ in the ΓXUL plane. Note that the

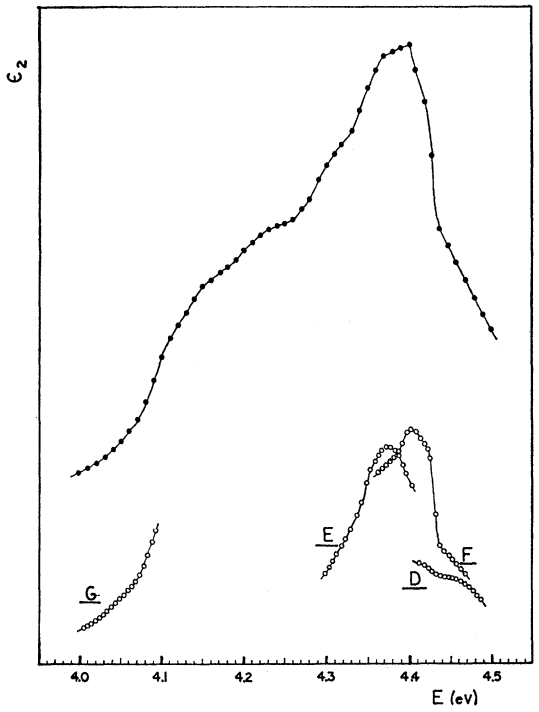

FIG. 13. Shows the contribution to $\epsilon_2^{4,5}(\omega)$ for model II from (a) the regions D, E, F, and G indicated in Fig. 8 (circles) and from (b) the whole wedge (dots).

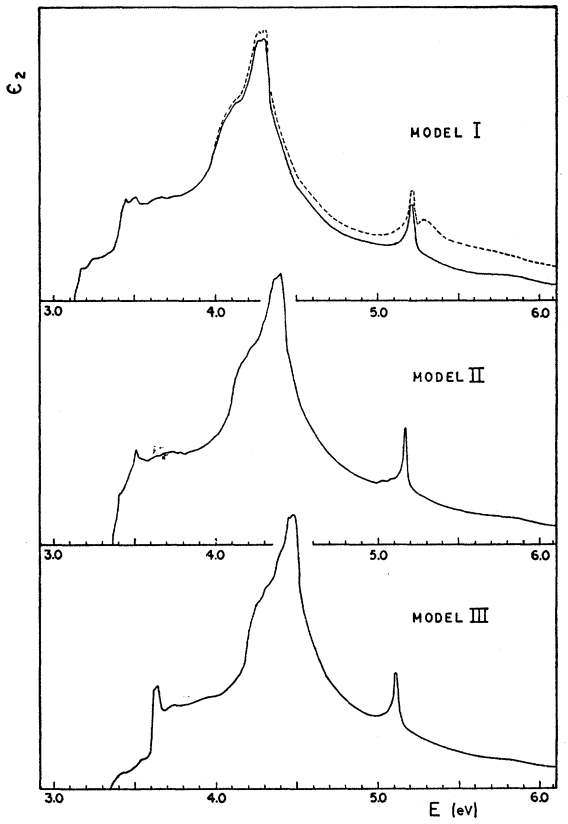

FIG. 14. Shows (a) the contribution from 12 interband gaps to $\epsilon_2(\omega)$ for model I (dashed line), (b) the sum of $\epsilon_2^{4,5}(\omega)$, $\epsilon_2^{3,5}(\epsilon)$, and $\epsilon_2^{4,6}(\omega)$ for models I, II, and III (solid lines).

symmetry of the BZ connects both planes through the lines KL and UL.

In order to study the contributions from the c.p.'s, a partial calculation of $\epsilon_2(\omega)$ was done on four separated zones denoted as D, E, F, and G in Fig. 8. The results are shown in Fig. 13.

The c.p. on the Σ axis has an energy of 4.47 eV. A change in the slope is detected in the corresponding curve, but is insignificant.

We have also separated out the contribution of the region G around the X point. This is shown in Fig. 13 as curve G, where it is seen that $X_4 \rightarrow X_1$ makes a very weak contribution. This fact was also noted by Kane. From symmetry considerations we know that there is no critical point at X. In fact, the critical point on the Δ axis near the zone center is the one that is sometimes mistakenly placed at X (see Fig. 6). Clearly, the two somewhat overlapping regions E and F make the largest contribution to the big peak.

For the other three models the interpretation is completely similar. The only difference is a general small shift in the energies.

The contribution from other interband transitions was computed in model I and shown in Fig. 14. Twelve interband transitions, from conduction bands $E_2E_3E_4$ to valence bands $E_5E_6E_7E_8$ were considered. The contribution from $\epsilon_2^{4,5}$, $\epsilon_2^{4,6}$, and $\epsilon_2^{3,5}$ were found the most important ones and their sum is also plotted. For models II and III the contribution from these bands was also calculated and their sum shown in the same figure.

The main additional features in model I are: a small peak at 3.5 eV, a peak at 5.21 eV, and a small broad peak at 5.28 eV. The peak at 3.5 eV is produced by the $E_{3,5}$ transition. The peak at 5.21 eV is produced by the transition $E_{4,6}$ in a zone near the L point ($L_{3'} \rightarrow L_3$). Two saddle points, in the ΓXUL and ΓKL, planes that are very near to the L point are responsible for the peak. The L point has an energy of 5.24 eV which is a bit displaced with respect to the peak. For models II and III, the energy bands become very flat and the different c.p.'s cannot be separated. The small peak at 5.28 eV comes from the interband transition $E_{3,6}$.

Finally, we mention that for model IV we attempted to determine the effect of lowering the $X_4 \rightarrow X_1$ gap keeping those at Γ and L fixed at the values for model III. As expected, the result is to greatly flatten out the energy contours along the Δ axis compared to those for model III. The peak at 3.4 eV is somewhat enhanced. We have not analyzed the results in detail.

V. SUMMARY AND CONCLUSIONS

We note by examination of Table II or Figs. 7–9 that the interband energy contours of Si are extremely complex, showing a large cluster of c.p.'s in the energy range 3.3–3.6 eV. This is true for all three of the models examined carefully, and presumably for model IV as well. The c.p.'s appear quite far removed from the points of high symmetry in the BZ. This is a clear warning that one must be extremely cautious in making band assignments based on simple interpretations or band calculations based on only a few points in the zone.[17,27]

In all of the present models we have placed $\Gamma_{25'} \rightarrow \Gamma_{15}$ near 3.4 eV. On the other hand, Herman *et al.* claim[17] that this gap must be less than 3 eV. They argue that their first-principle calculations place it very near 3 eV. They allow for the possible shift of this transition by introducing pseudopotential coefficients in a manner analogous to our own. Their conclusion is that $\Gamma_{25'} \rightarrow \Gamma_{15}$ cannot be put above 3 eV without destroying the agreement at other points of the zone. To draw this conclusion, however, it must be assumed that their starting calculation placed $\Gamma_{25'} \rightarrow \Gamma_{15}$ within 0.3 eV of its correct value. This appears beyond the range of present first-principles work. Furthermore, if one places $\Gamma_{25'} \rightarrow \Gamma_{15}$ below 3.0 eV, one expects to find additional structure not only in $\epsilon_2(\omega)$,[19] but much more sharply in the piezo-optic and electro-optic studies. Structure has not been observed at the energies corresponding to Herman's assignment.

Further confirmation for the placement of the

[27] F. Herman, R. L. Kortum, C. D. Kuglin, and R. A. Short, *The Physics of Semiconductors* (Physical Society of Japan, Kyoto, 1966), p. 7.

TABLE II. Critical points in $E_{4,5}$ near 3.4 eV for three models.

Energy	Location	Type
Model I		
3.38	(5/24, 1/24, 0)	M_0[a]
3.44	(0, 0, 0)	M_0[a]
3.42	(7/24, 0, 0)	M_1[a]
3.13	$(\frac{1}{2}, \frac{1}{2}, \frac{1}{2})$	M_0[b]
3.49	(1/24, 1/24, 1/24)	M_3
3.47	$(\frac{1}{12}, \frac{1}{12}, 1/24)$	M_1
Model II		
3.38	$(\frac{1}{8}, 1/24, 0)$	M_0[a]
3.39	(0, 0, 0)	M_0[a]
3.41	$(\frac{1}{12}, 1/24, 0)$	M_1[a]
3.49	(7/24, 0, 0)	M_1
3.37	$(\frac{1}{2}, \frac{1}{2}, \frac{1}{2})$	M_0[b]
3.54	$(\frac{1}{8}, \frac{1}{8}, \frac{1}{8})$	M_3
3.51	$(\frac{1}{8}, \frac{1}{8}, \frac{1}{12})$	M_1[c]
3.51	$(\frac{1}{12}, \frac{1}{8}, \frac{1}{8})$	M_2[c]
Model III		
3.34	(0, 0, 0)	M_0
3.35	$(\frac{1}{12}, 1/24, 0)$	M_0
3.54	(7/24, 0, 0)	M_1
3.60	$(\frac{1}{2}, \frac{1}{2}, \frac{1}{2})$	M_0[b]
3.61	$(\frac{1}{4}, \frac{1}{4}, 5/24)$	M_1[c]
3.64	$(\frac{1}{4}, 5/24, 5/24)$	M_2[c]

[a] The three c.p. produce a step. They can be considered to be a nearly two-dimensional c.p.
[b] A segment along the Λ line with almost constant energy is associated to this c.p. The disposition is a nearly two-dimensional c.p. producing a step in $\epsilon_2(\omega)$.
[c] The two critical points produce a peak.

$\Gamma_{25'} \rightarrow \Gamma_{15}$ transition comes from analysis of the quantum efficiency (QE) data in Si.[28,29] We note that one sees an abrupt rise in the QE (from unity) at 3.4 eV. This may be interpreted as the onset of pair production by electrons. Below 3.4 eV, we may assume that the electron and the hole both share the photon's energy equally following Shockley[30] and hence neither has an energy sufficient to undergo Auger pair production. Above 3.4 eV, the $\Gamma_{25'} \rightarrow \Gamma_{15}$ threshold has been reached and the electron takes nearly all of the energy $\hbar\omega$ supplied by the absorbed photon. If Herman's value of 2.8–3.0 eV were accepted for $\Gamma_{25'} \rightarrow \Gamma_{15}$, one would expect the QE to rise abruptly at this energy, i.e., when the electron is created with an energy $E_{\text{final}} \geq 2E_{\text{gap}}$. Direct calculations confirm that indeed the QE should rise abruptly when $\hbar\omega = \Gamma_{25'} \rightarrow \Gamma_{15}$.[31] We note that studies of the photoelectric yield curves by Cohen and Phillips also place $\Gamma_{25'} \rightarrow \Gamma_{15}$ near 3.4 eV.[32]

Recent studies of Cardona *et al.*[33] on the electro-optic effect in Ge-Si alloys give a value for $\Gamma_{25'} \rightarrow \Gamma_{2'} = 4.1 \pm 0.1$ eV for pure Si. This is very close to where our alloy calculations[34] predicted it would be. That discussion interpreted the break in the optical reflectivity peak observed by Tauc and Abraham[35] as due to the crossing of the Γ_{15} and $\Gamma_{2'}$ levels at 79% Si. If we take a value of $\Gamma_{25'} \rightarrow \Gamma_{2'} = 0.8$ eV in pure Ge, assume that $\Gamma_{25'} \rightarrow \Gamma_{15}$ remains roughly constant over the entire alloy range, and then use our previous interpretation that the break at 79% Si occurs when $\Gamma_{2'}$ and Γ_{15} cross, we have for $\Gamma_{25'} \rightarrow \Gamma_{15} = 3.4$ eV in Si.[36]

Examination of Fig. 14 shows the dramatic change in the appearance of the spectrum that is expected as one allows $L_{3'} \rightarrow L_1$ to vary with respect to $\Gamma_{25'} \rightarrow \Gamma_{15}$. Allowing $L_{3'} \rightarrow L_1$ to be slightly larger than the Γ transition results in the extreme sharpness associated with the edge structures of model III.

Some final comment should be made concerning our choice of models. The emphasis throughout has been on how relative a variation of $\Gamma_{25'} \rightarrow \Gamma_{15}$ and $L_{3'} \rightarrow L_1$ ought to affect the line shape of $\epsilon_2(\omega)$ as far as single-particle interband properties are concerned. In particular, many theoretical studies now place these both fairly close to one another ($\sim\frac{1}{2}$ eV). This, of course, as our present study shows, is just the case when the greatest complexity is likely to occur. We have therefore tested models in which these two important gaps are quasidegenerate. The principal factor in the line shape is the relative position of the levels rather than the precise value of the potential which is used in placing them. Comparison of Fig. 10 (model II) with Fig. 11 amply demonstrates this. Table I indicates that the potential parameters are quite different for the models but that the two gaps of interest are relatively similarly placed, resulting in nearly identical $\epsilon_2(\omega)$ curves near 3.4 eV.

Model I seems almost certainly ruled out on comparison with experiment.[37] According to this model, two strong edges should appear, one near the $L_{3'} \rightarrow L_1$ and the other near $\Gamma_{25'} \rightarrow \Gamma_{15}$. Thus it appears very unlikely that $L_{3'} \rightarrow L_1 < \Gamma_{25'} \rightarrow \Gamma_{15}$ by more than, say, 0.1 eV (an upper limit on lifetime broadening).

In conclusion we note that of all our band models the one closest to experiment is model III. It is interesting that Table II shows c.p.'s of type M_0, M_1, and M_0 at energies of 3.32, 3.54, and 3.60 eV, respectively, for model III. Ghosh finds three peaks at 3.25, 3.42, and 3.55 eV in his data which may correspond to these three critical points. Further work which is in progress should allow us to chose finally between the various possible models.

ACKNOWLEDGMENTS

The authors would like to thank Dr. Evan Kane for useful conversations and L. Liu for several discussions.

[28] A. J. Tuzzolino, Phys. Rev. **134**, A205 (1964).
[29] J. Vavilov, J. Phys. Chem. Solids **8**, 223 (1959).
[30] W. Shockley, Czech. J. Phys. **11**, 81 (1961).
[31] D. Brust (unpublished).
[32] M. L. Cohen and J. C. Phillips, Phys. Rev. **139**, A912 (1965).
[33] M. Cardona, K. L. Shaklee, and F. H. Pollak, Phys. Letters **23**, 37 (1966).
[34] F. Bassani and D. Brust, Phys. Rev. **131**, 1524 (1963).
[35] J. Tauc and A. Abraham, J. Phys. Chem. Solids **20**, 190 (1961).
[36] The small variation of $\Gamma_{25'} \rightarrow \Gamma_{15}$ on going from Ge to Si will not change our result significantly.
[37] H. R. Philipp and E. A. Taft, Phys. Rev. **120**, 37 (1960).