Toward a Theory of Isoelectronic Impurities in Semiconductors*

ROGER A. FAULKNER
Palmer Physical Laboratory, Princeton University, Princeton, New Jersey 08540
and
Bell Telephone Laboratories, Murray Hill, New Jersey† 07974

(Received 19 March 1968; revised manuscript received 9 September 1968)

Isoelectronic impurities in semiconductors are investigated using a Koster-Slater one-band–one-site approximation. Good agreement with experiment is found for the optical absorption induced by nitrogen in gallium phosphide. Binding energies of excitons to single nitrogen impurities and to double nitrogen impurities are calculated using the complete structure of the GaP conduction bands. Very limited success is achieved in these calculations, which ignore correlation effects and the lattice relaxation and electronic polarization of the host crystal. The indications are that the response of the host to the impurity is of crucial importance in the short-range-interaction problem in semiconductors.

I. INTRODUCTION

THE term "isoelectronic impurity" is used to denote an impurity center in a crystal arising when an atom from the same column of the periodic table as one of the constituents of the host crystal substitutes for that constituent. Isoelectronic substitutions in certain wide band gap semiconductors have a profound effect on the optical properties of the materials for photon energies in the vicinity of the band gap. In particular, the impurity systems GaP:N, GaP:Bi, and ZnTe:O exhibit sharp absorption lines and CdS:Te exhibits an absorption band lying within the band gap of the pure crystal.[1–4] These lines can be attributed to exciton bound states at the impurities.

These systems have been classified as isoelectronic acceptors and isoelectronic donors on the basis of whether the impurity is, respectively, attractive for electrons or attractive for holes.[2] For isoelectronic acceptors, the mechanism for binding an exciton is conceived to be as follows: An electron in the conduction band is attracted to the uncharged impurity by a short-range potential and becomes bound to it, the result being a charged system that can subsequently bind a hole. Isoelectronic donors would operate in the inverse sequence, first binding a hole and then an electron.

In one system, GaP:N, in addition to the principal bound state (A line), other bound states have been observed in both absorption and fluorescence which lie lower in energy than the A line and form a sequence of levels converging to the A line.[1] These levels have been identified as bound states of excitons to two nitrogen impurities at various interatomic spacings. They are labeled NN_1, NN_2, NN_3, etc., in increasing energy (diminishing binding energy). There is no apparent regularity in these lines except that they converge to the A line.

In GaP:N, the A line is actually a doublet, called A and B, arising from the degenerate valence band. The A-B separation is 0.8 meV. Lying above the A-B doublet by 1.8 meV is a pair of doublets (seen weakly in fluorescence) lying quite close together and which are clearly replicas of the A-B doublet.[5] Figure 1 shows the fluorescence from these states. GaP has a three-valley conduction band and these higher states can be identified as the antisymmetrical pair split off from the symmetrical A-B doublet by intervalley mixing.

GaP is an indirect band gap semiconductor and as such, the optical absorption above the band gap is dominated at low temperatures by the presence of nitrogen.[6] Purely electronic processes (assisted by the nitrogen) which produce optical absorption are (1) creation of an unbound electron and an unbound hole,

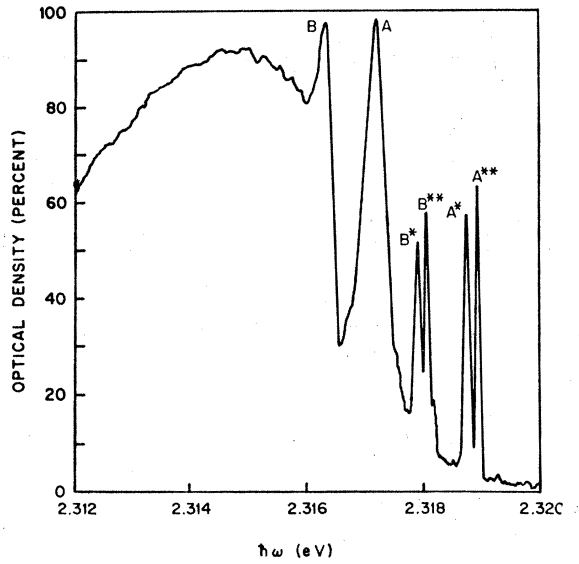

FIG. 1. Fluoresence spectrum of the A, B lines in GaP:N showing the excited states split off from A,B by intervalley mixing. (Spectrum taken by P. J. Dean.)

* Work supported in part by the National Science Foundation.
† Present address.
[1] D. G. Thomas and J. J. Hopfield, Phys. Rev. **150**, 680 (1966).
[2] J. J. Hopfield, D. G. Thomas, and R. T. Lynch, Phys. Rev. Letters **17**, 312 (1966).
[3] J. D. Cuthbert and D. G. Thomas, Phys. Rev. **154**, 763 (1967).
[4] A. C. Aten, J. H. Haanstra, and H. de Vrier, Philips Res. Repts. **20**, 395 (1965).
[5] P. J. Dean (private communication).
[6] J. J. Hopfield, P. J. Dean, and D. G. Thomas, Phys. Rev. **158**, 748 (1967).

(2) creation of a bound electron and an unbound hole,
(3) creation of an unbound exciton.

These processes are discussed in Sec. III, where the consequences of making a Slater-Koster one-band–one-site approximation will be investigated. This simple model has the advantages of being soluble analytically and of being able to correlate a great deal of information with essentially only one adjustable parameter. In Secs. IV and V, the results of more extensive computer calculations are presented.

II. THEORETICAL FORMALISM

A. Notation

Before proceeding further, an aside on notation would be in order.

Almost always, units with $\hbar=1$ will be used.

Bloch functions (BF) will be denoted either as $\psi_n(\mathbf{k},\mathbf{x})$ or abstractly as a ket vector with a round bracket: $|n\mathbf{k})$. Here, n refers to the band and $\mathbf{k}$ refers to the reduced wave vector lying in the first Brillouin zone (BZ). The normalization for the BF's will be taken to be a Dirac δ function:

$$(n\mathbf{k}|n'\mathbf{k}')=\delta_{nn'}\delta(\mathbf{k}-\mathbf{k}'). \tag{2.1}$$

In terms of the periodic part $U_n(\mathbf{k},\mathbf{x})$, $\psi_n(\mathbf{k},\mathbf{x})$ may be written

$$\psi_n(\mathbf{k},\mathbf{x})=[\Omega^{1/2}/(2\pi)^{3/2}]e^{i\mathbf{k}\cdot\mathbf{x}}U_n(\mathbf{k},\mathbf{x}), \tag{2.2}$$

where Ω is the volume of a primitive cell of the crystal. The normalization of $U_n(\mathbf{k},\mathbf{x})$ is

$$\int d^3x\, U_n^*(\mathbf{k},\mathbf{x})U_{n'}(\mathbf{k},\mathbf{x})=\langle n\mathbf{k}|n'\mathbf{k}\rangle=\delta_{nn'}. \tag{2.3}$$

Pointed brackets, as above, refer to integrations over the primitive cell of the periodic parts of the BF's. Round brackets refer to integrations of the complete BF's over all space.

Wannier functions (WF) are defined in terms of the BF's:

$$w_n(\mathbf{x}-\mathbf{R})=\frac{\Omega^{1/2}}{(2\pi)^{3/2}}\int d^3k\;\theta(\mathbf{k})e^{-i\mathbf{k}\cdot\mathbf{R}}\psi_n(\mathbf{k},\mathbf{x}) \tag{2.4}$$

or

$$|n\mathbf{R})=\frac{\Omega^{1/2}}{(2\pi)^{3/2}}\int d^3k\;\theta(\mathbf{k})e^{-i\mathbf{k}\cdot\mathbf{R}}|n\mathbf{k}), \tag{2.5}$$

where $\mathbf{R}$ is a direct lattice vector.

The function $\theta(\mathbf{k})$ is defined to be one if $\mathbf{k}$ lies in the first BZ and is zero otherwise. This is simply a device for restricting $\mathbf{k}$-space integrals to the BZ.

The WF normalization is

$$(n\mathbf{R}|n'\mathbf{R}')=\delta_{nn'}\delta_{\mathbf{R}\mathbf{R}'}. \tag{2.6}$$

BF's can be written in terms of the WF's defined above:

$$|n\mathbf{k})=\frac{\Omega^{1/2}}{(2\pi)^{3/2}}\sum_{\mathbf{R}}e^{i\mathbf{k}\cdot\mathbf{R}}|n\mathbf{R}). \tag{2.7}$$

Reciprocal-lattice vectors are denoted as $\mathbf{G}$, and $\mathbf{G}\cdot\mathbf{R}=2\pi\times$(an integer).

B. Formalism

Let the interaction between electrons in the crystal and the isoelectronic impurity be describable by a potential $V(\mathbf{x})$, which consists of both the bare interaction and the interaction due to the self-consistent rearrangement of the charge density in the crystal. Let us also adopt the independent-particle approach for the moment.

The Hamiltonian for a single electron then becomes

$$H=H_0+V, \tag{2.8}$$

where

$$H_0|n\mathbf{k})=\epsilon_n(\mathbf{k})|n\mathbf{k}) \tag{2.9}$$

and $\epsilon_n(\mathbf{k})$ are the band energies.

We may write Schrödinger's equation

$$(E_\lambda-H_0)\psi_\lambda=V\psi_\lambda \tag{2.10}$$

and use the BF representation

$$[E_\lambda-\epsilon_n(\mathbf{k})](n\mathbf{k}|\psi_\lambda)=\sum_{n'}\int d^3k'\theta(\mathbf{k}')\times(n\mathbf{k}|V|n'\mathbf{k}')(n'\mathbf{k}'|\psi_\lambda). \tag{2.11}$$

Writing the amplitude $(n\mathbf{k}|\psi_\lambda)$ as $\psi_\lambda(n\mathbf{k})$, the momentum-space wave function, we have the integral equation

$$\psi_\lambda(n\mathbf{k})=\delta_{nn_\lambda}\delta(\mathbf{k}-\mathbf{k}_\lambda)+[E_\lambda-i0-\epsilon_n(\mathbf{k})]^{-1}\sum_{n'}\int d^3k'\theta(\mathbf{k}')\times(n\mathbf{k}|V|n'\mathbf{k}')\psi_\lambda(n'\mathbf{k}'). \tag{2.12}$$

λ denotes the quantum numbers n_λ and $\mathbf{k}_\lambda$ of the unperturbed wave function.

The unperturbed wave function is $\varphi_\lambda=\delta_{nn_\lambda}\delta(\mathbf{k}-\mathbf{k}_\lambda)$ and satisfies the equation

$$[E_\lambda-\epsilon_n(\mathbf{k})]\varphi_\lambda(\mathbf{k})=0.$$

The term $-i0$ is attached to E_λ in Eq. (2.12) specifies what to do with the singularity when $E_\lambda=\epsilon_n(\mathbf{k})$. In particular,

$$\frac{1}{[E_\lambda-i0-\epsilon_n(\mathbf{k})]}=P\frac{1}{[E_\lambda-\epsilon_n(\mathbf{k})]}+i\pi\delta(E_\lambda-\epsilon_n(\mathbf{k})),$$

where P denotes the principal part.

If the eigenvalue E_λ lies in the band gap, the first term is omitted and ψ_λ becomes square integrable. If the

eigenvalue lies in the continuum of the unperturbed crystal, then $E_\lambda=\epsilon_{n_\lambda}(\mathbf{k}_\lambda)$ and the normalization of the states ψ_λ is

$$\sum_n \int d^3k\ \theta(\mathbf{k})\psi_\lambda^*(n\mathbf{k})\psi_{\lambda'}(n\mathbf{k})=\delta(n_\lambda,n_{\lambda'})\delta(\mathbf{k}_\lambda-\mathbf{k}_{\lambda'}). \quad (2.13)$$

One must be careful to distinguish the argument of the wave function $(n\mathbf{k})$ from the labeling of the wave function $(n_\lambda\mathbf{k}_\lambda)$.

The ground state of the crystal consists of electrons occupying these states up to the band gap between the conduction bands and the valence bands. If the admixture of conduction band states to the valence band states is not great, we can take the ground state to be the same as the unperturbed ground state, i.e., electrons filling Bloch states up to the band gap.

An excited state of the crystal consisting of one electron in the conduction bands and one hole in the valence bands can be written

$$|\lambda\rangle=\sum_c\sum_v\int d^3k_1\theta(\mathbf{k}_1)\int d^3k_2\theta(\mathbf{k}_2)\psi_\lambda(c\mathbf{k}_1,v\mathbf{k}_2) \times C_c^\dagger(\mathbf{k}_1)C_v(\mathbf{k}_2)|G\rangle, \quad (2.14)$$

where $C^\dagger$ and C are Bloch-state creation and destruction operators and $|G\rangle$ is the state vector of the ground state. C and $C^\dagger$ satisfy the anticommutation relations

$$\{C_n(\mathbf{k}),C_{n'}^\dagger(\mathbf{k}')\}=\delta_{nn'}\delta(\mathbf{k}-\mathbf{k}'). \quad (2.15)$$

The Schrödinger equation for ψ_λ consists of the unperturbed part, the interaction of the electron and of the hole with the impurity, and the Coulomb attraction of the electron and the hole:

$$\begin{aligned}[E_\lambda-\epsilon_c(\mathbf{k}_1)+\epsilon_v(\mathbf{k}_2)]\psi(c\mathbf{k}_1,v\mathbf{k}_2)&\\ =\sum_{c'}\int d^3q\ \theta(\mathbf{q})(c\mathbf{k}_1|V|c'\mathbf{q})\psi(c'\mathbf{q},v\mathbf{k}_2)&\\ -\sum_{v'}\int d^3q\ \theta(\mathbf{q})(v'\mathbf{q}|V|v\mathbf{k}_2)\psi(c\mathbf{k}_1,v'\mathbf{q})&\\ -\frac{4\pi e^2}{\kappa}\frac{1}{(2\pi)^3}\int d^3q\frac{1}{q^2}\psi(c\mathbf{k}_1+\mathbf{q},v\mathbf{k}_2+\mathbf{q}).& \quad (2.16)\end{aligned}$$

If the state is discrete, it can be normalized:

$$\langle\lambda|\lambda\rangle=1=\sum_c\sum_v\int d^3k_1\theta(\mathbf{k}_1)\int d^3k_2\theta(\mathbf{k}_2) \times|\psi_\lambda(c\mathbf{k}_1,v\mathbf{k}_2)|^2. \quad (2.17)$$

In Born approximation (linear response), the transition rate for optical absorption is given by

$$w_{fi}=2\pi|T_{fi}|^2\delta(E_f-E_i), \quad (2.18)$$

where

$$T_{fi}=\frac{e}{mc}\sum_v\sum_c\int d^3k_1\theta(\mathbf{k}_1)\int d^3k_2 \times\theta(\mathbf{k}_2)\psi_\lambda^*(c\mathbf{k}_1,v\mathbf{k}_2)(c\mathbf{k}_1|\mathbf{A}\cdot\mathbf{P}|v\mathbf{k}_2). \quad (2.19)$$

$\mathbf{A}(\mathbf{x})$ is the vector potential for the incoming radiation, $\mathbf{A}(\mathbf{x})=\mathbf{A}_0e^{i(\mathbf{k}\cdot\mathbf{x}-\omega t)}$. $\mathbf{P}$ is the momentum operator $-i\hbar\nabla$.

As usual, since the wave vector of the incoming light is so small compared to the extension of the BZ, we may take it to the zero:

$$T_{fi}=\frac{e}{mc}\sum_v\sum_c\int d^3k\ \theta(\mathbf{k})\psi_\lambda^*(c\mathbf{k},v\mathbf{k})\langle c\mathbf{k}|\mathbf{A}_0\cdot\mathbf{P}|v\mathbf{k}\rangle. \quad (2.20)$$

The incoming radiation flux is

$$\langle S\rangle=\omega^2|\mathbf{A}_0|^2n/8\pi c, \quad (2.21)$$

where n is the index of refraction at frequency ω.

The cross section for optical absorption by a single impurity center is then

$$\sigma(\omega)=\sum_{\substack{\text{final}\\ \text{states}\\ (\lambda)}}4(2\pi)^2\frac{e^2}{m^2c\omega n} \times\left|\sum_v\sum_c\int d^3k\ \theta(\mathbf{k})\psi_\lambda^*(c\mathbf{k},v\mathbf{k})\langle c\mathbf{k}|p|v\mathbf{k}\rangle\right|^2 \times\delta(E_\lambda-\hbar\omega). \quad (2.22)$$

The absorption coefficient α is given by

$$\alpha(\omega)=\sigma(\omega)\times(\text{No. of impurities per unit volume}).$$

The formula for $\sigma(\omega)$ can be simplified under special conditions, as will be discussed in the next chapter.

III. ONE-BAND–ONE-SITE APPROXIMATION

A. Model

One can expand the matrix element appearing under the integral sign in Schrödinger's equation (2.16) in terms of WF matrix elements[7]:

$$(n\mathbf{k}|V|n'\mathbf{k}')=\frac{\Omega}{(2\pi)^3}\sum_{\mathbf{R}\mathbf{R}'}e^{-i\mathbf{k}\cdot\mathbf{R}}(n\mathbf{R}|V|n'\mathbf{R}')e^{i\mathbf{k}'\cdot\mathbf{R}'}. \quad (3.1)$$

If $V(\mathbf{x})$ is of shorter range than the interatomic spacing in the crystal, and if the WF's are well localized, the dominant term in the expansion above comes from $\mathbf{R}=\mathbf{R}'=0$ [if $V(\mathbf{x})$ is centered at the origin of coordinates].

A simple model may be obtained if we take only this principal term into account and if, further, we consider only one conduction band (c) and only one valence

[7] G. F. Koster and J. C. Slater, Phys. Rev. 96, 1208 (1954).

band (v). Then, in this model,

$$\begin{aligned}&(c\mathbf{k}|\,V\,|c\mathbf{q})\\&\quad=[\Omega/(2\pi)^3](c\mathbf{R}=\mathbf{0}|\,V\,|c\mathbf{R}=\mathbf{0})\equiv J_c\Omega/(2\pi)^3,\\&(v\mathbf{k}|\,V\,|v\mathbf{q})\\&\quad=[\Omega/(2\pi)^3](v\mathbf{R}=\mathbf{0}|\,V\,|v\mathbf{R}=\mathbf{0})\equiv J_v\Omega/(2\pi)^3,\end{aligned}\tag{3.2}$$

and, if the impurity is located at a lattice site $\mathbf{R}$ instead of the origin,

$$(c\mathbf{k}|\,V(\mathbf{x}-\mathbf{R})\,|c\mathbf{q})=e^{-i(\mathbf{k}-\mathbf{q})\cdot\mathbf{R}}J_c\Omega/(2\pi)^3.\tag{3.3}$$

This second form will be useful when we consider the case of two impurities separated by a lattice vector $\mathbf{R}$.

Measuring the excited-state energy from the band gap, the electron-band energy $\epsilon_c(\mathbf{k})$ up from the conduction-band minimum, and the hole energy down from the valence-band maximum, the Schrödinger equation for an electron and a hole is

$$\begin{aligned}&[E_\lambda-\epsilon_c(\mathbf{k}_1)-\epsilon_v(\mathbf{k}_2)]\Psi_\lambda(\mathbf{k}_1,\mathbf{k}_2)\\&=\frac{J_c\Omega}{(2\pi)^3}\int d^3q\;\theta(\mathbf{q})\Psi_\lambda(\mathbf{q},\mathbf{k}_2)-\frac{J_v\Omega}{(2\pi)^3}\int d^3q\;\theta(\mathbf{q})\Psi_\lambda(\mathbf{k}_1,\mathbf{q})\\&\quad-\frac{4\pi e^2}{\kappa}\frac{1}{(2\pi)^3}\int d^3q\frac{1}{q^2}\Psi_\lambda(\mathbf{k}_1+\mathbf{q},\mathbf{k}_2+\mathbf{q}).\end{aligned}\tag{3.4}$$

Let us restrict ourselves, for the remainder of this section, to the case of GaP:N. In this case, the nitrogen impurity is attractive for electrons and repulsive for holes. The parameters J_c and J_v appearing in Eq. (3.4) are negative.

We can argue that we can ignore the interaction of the hole with the impurity as follows: We expect a resonance behavior for conduction-band electrons due to the existence of a shallow bound state at the impurity, making a large scattering phase shift even though the attraction is rather weak. For holes, the potential is repulsive and no resonance will exist. The analogy here is obviously to neutron-proton scattering and the existence of the weakly bound deuteron. In the case of the bound state, which we think of as the electron bound by the short-range impurity potential and the hole bound by the long-range Coulomb attraction of the electron, the short-range repulsive core the hole experiences due to the impurity can be ignored for the most part. We then have Schrödinger's equation for the electron-hole-impurity system:

$$\begin{aligned}&[E_\lambda-\epsilon_c(\mathbf{k}_1)-\epsilon_v(\mathbf{k}_2)]\Psi_\lambda(\mathbf{k}_1,\mathbf{k}_2)=\frac{J\Omega}{(2\pi)^3}\int d^3q\;\theta(\mathbf{q})\Psi_\lambda(\mathbf{q},\mathbf{k}_2)\\&\quad-\frac{4\pi e^2}{\kappa}\frac{1}{(2\pi)^3}\int d^3q\frac{1}{q^2}\Psi_\lambda(\mathbf{k}_1+\mathbf{q},\mathbf{k}_2+\mathbf{q}),\end{aligned}\tag{3.5}$$

where J is a negative energy parameter yet to be determined.

B. Bound-State Problem in the One-Band–One-Site Approximation

We shall approximate $\Psi_\lambda(\mathbf{k}_1,\mathbf{k}_2)$ for the bound state by a product wave function $\Psi_e(\mathbf{k}_1)\Psi_h(\mathbf{k}_2)$ and first look only at the electron part of Schrödinger's equation:

$$[E-\epsilon_c(\mathbf{k})]\Psi(\mathbf{k})=\frac{J\Omega}{(2\pi)^3}\int d^3q\;\theta(\mathbf{q})\Psi(\mathbf{q}).\tag{3.6}$$

The interaction between the electron and the impurity is a truncated momentum-space interaction reminiscent of the BCS model for superconductivity. If the integral over the BZ were extended over all space, the interaction would be a δ-function potential. In three dimensions, of course, a δ-function potential has no bound state.

The Fourier transformation of the model cutoff momentum-space interaction leads to a potential in coordinate space that is almost a square well with a range equal to the interatomic spacing. The deviation from a square well is principally a small, decaying, oscillatory tail.

Equation (3.6) can be solved immediately by noting that the right side of the equation is a constant and that for $E<0$, the multiplier on the left is never zero:

$$\Psi(\mathbf{k})=N/[E-\epsilon_c(\mathbf{k})],\tag{3.7}$$

$$N=J\frac{\Omega}{(2\pi)^3}\int d^3q\;\theta(\mathbf{q})\Psi(\mathbf{q}).\tag{3.8}$$

Substituting (3.7) into (3.8) gives us the eigenvalue equation

$$1+J\frac{\Omega}{(2\pi)^3}\int d^3q\;\theta(\mathbf{q})\frac{1}{[\epsilon_c(\mathbf{q})-E]}=0.\tag{3.9}$$

We are measuring both E and $\epsilon_c(\mathbf{k})$ from the lower edge of the conduction band. It becomes quite clear that the effective-mass approximation is not good in this problem, because if we take $\epsilon_c(\mathbf{q})=q^2/2m^*$ and extend the integration over all momentum space, the integral diverges. However, we can use the relation

$$\frac{1}{(\epsilon-E)}=\frac{1}{\epsilon}+\frac{1}{\epsilon}\frac{E}{(\epsilon-E)}\tag{3.10}$$

to obtain the equation

$$1+J\langle 1/\epsilon\rangle=-J\frac{\Omega}{(2\pi)^3}\int d^3q\;\theta(\mathbf{q})\frac{E}{\epsilon(\mathbf{q})(\epsilon(\mathbf{q})-E)},\tag{3.11}$$

where

$$\langle 1/\epsilon\rangle=\frac{\Omega}{(2\pi)^3}\int d^3q\;\theta(\mathbf{q})\frac{1}{\epsilon(\mathbf{q})}\tag{3.12}$$

is the average over the BZ of $1/\epsilon(\mathbf{k})$.

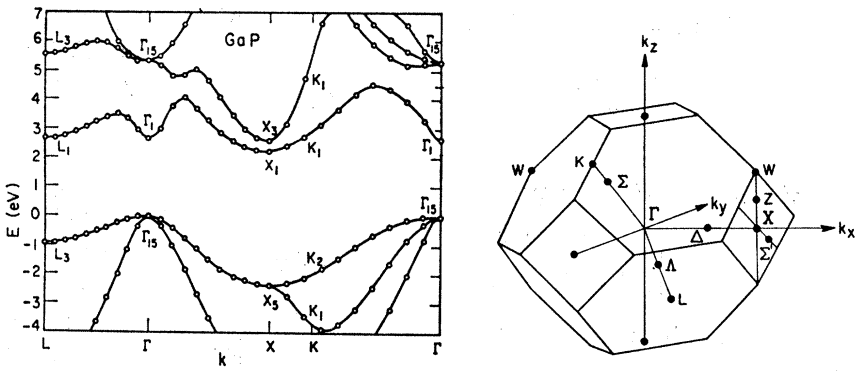

FIG. 2. Band structure and Brillouin zone of GaP. (After Cohen and Bergstresser, Ref. 8.)

The second integral can be extended over all space and effective masses can be used if $|E|$ is small.

We now need to know the band structure of GaP. Figure 2 shows the GaP band structure as given by the empirical pseudopotential calculations of Cohen and Bergstresser[8] along with the Brillouin zone for a face-centered cubic crystal with the symmetry notation of BSW.[9]

The (indirect) band gap is between points Γ_{15} and X_1 and equals 2.3 eV. There are three inequivalent points X in the Brillouin zone, giving GaP a three-valley conduction band. The effective masses in the valleys are not isotropic, the longitudinal mass being 1.49 m and the two transverse masses being 0.25 m (directions relative to the line from X to Γ). Of special note is that the first conduction band bends over in the center of the zone and comes within 0.5 eV of the minimum at X.

Denoting the effective masses at X as (m_1,m_1,m_2) and extending the integral of Eq. (3.11) over all space, we obtain

$$1+J\langle 1/\epsilon\rangle = J(3\Omega/2\pi)m_1(2m_2|E|)^{1/2}. \quad (3.13)$$

For the bound state to exist, we must have $1+J\langle 1/\epsilon\rangle < 0$. Therefore, there is a critical value of $|J|$ below which no bound state exists. The special combination of J and $\langle 1/\epsilon\rangle$ appearing in Eq. (3.13) occurs so often in what follows that we shall give it a name:

$$Q \equiv \frac{1+J\langle 1/\epsilon\rangle}{J} > 0. \quad (3.14)$$

The normalization constant appearing in Eq. (3.7) can be determined by requiring

$$\int d^3k\, \theta(\mathbf{k})|\Psi(\mathbf{k})|^2 = 1 \quad (3.15)$$

or

$$\int d^3k\, \theta(\mathbf{k}) \frac{|N|^2}{|\epsilon(\mathbf{k})+|E||^2} = 1. \quad (3.16)$$

For small $|E|$, we can use the effective-mass approximation again, and $|N|$ is then

$$|N| = \frac{(2m_2|E|)^{1/4}}{2\pi(3m_1m_2)^{1/2}} \quad (3.17)$$

or, letting $m^* = (m_1m_1m_2)^{1/3}$ be the geometric mean of the masses,

$$|N| = \frac{(2m^*|E|)^{1/4}}{\sqrt{3}(2\pi m^*)}. \quad (3.18)$$

Equation (3.7) for the electron wave function then becomes

$$\Psi_e(\mathbf{k}) = \frac{(2m_e^*|E_1|)^{1/4}}{\sqrt{3}(2\pi m_e^*)} \frac{1}{[\epsilon_c(\mathbf{k})+|E_1|]}, \quad (3.19)$$

with $|E_1|$ given by

$$\frac{3\Omega}{(2\pi)} m_e^*(2m_e^*|E_1|)^{1/2} = Q \equiv \frac{1+J\langle 1/\epsilon\rangle}{J}. \quad (3.20)$$

In the spirit of the previous discussion, we now take the hole to be in a hydrogenic $1S$ state

$$\Psi_h(\mathbf{k}) = \frac{2\sqrt{2}|E_2|(2m_h^*|E_2|)^{1/4}}{(2\pi m_h^*)[\epsilon_v(\mathbf{k})+|E_2|]^2}, \quad (3.21)$$

where, for purposes of maintaining a simple model, the hole mass is taken to be isotropic around a valence-band maximum at the center of the BZ (Γ):

$$\epsilon_v(\mathbf{k}) = k^2/2m_h^*. \quad (3.22)$$

The hole wave function is normalized so that

$$\int d^3k|\Psi_h(\mathbf{k})|^2 = 1. \quad (3.23)$$

We can now use the two wave functions we have obtained to form a product trial wave function for the bound-exciton problem, using E_1 and E_2 as variational parameters.

Solving numerically gives the value of Q, the one important parameter of the simple model, using the

[8] Marvin L. Cohen and T. K. Bergstresser, Phys. Rev. 141, 789 (1966).

[9] L. P. Bouckaert, R. Smoluchowski, and E. Wigner, Phys. Rev. 50, 58 (1936).

experimentally known value of E (0.02 eV),

$$Q=\frac{1+J\langle 1/\epsilon_c\rangle}{J}=0.022\ \text{eV}^{-1} \tag{3.24}$$

along with the wave-function parameters

$$\begin{aligned}|E_1|&=0.0104\ \text{eV},\\ |E_2|&=0.0096\ \text{eV}.\end{aligned} \tag{3.25}$$

The masses have been taken to be

$$\begin{aligned}m_e^*&=0.35m,\\ m_h^*&=0.20m.\end{aligned}$$

C. Optical Absorption at the Bound-State Energy

Thus far, nothing new has been learned; we have simply substituted one parameter for another. However, we can go on to calculate the integrated optical absorption strength due to the bound state using the wave-function parameters calculated above.

The expression given in the last section for the optical-absorption cross section can be simplified considerably, using the properties of the wave function. First, since we are considering only one valence band and one conduction band, we may write the general expression as

$$\sigma(\omega)=8(2\pi)^2\frac{e^2}{nc\omega}\left|\int d^3k\ \theta(\mathbf{k})\psi_e^*(\mathbf{k})\psi_h(\mathbf{k})\langle c\mathbf{k}|p/m|v\mathbf{k}\rangle\right|^2 \times\delta(\hbar\omega-(E_g-|E_b|)), \tag{3.26}$$

where the sum over the final states has given a factor of 2 due to spins.

Because the hole wave function is concentrated about $\mathbf{k}=\mathbf{0}$, we can make the further approximation of setting $\mathbf{k}=\mathbf{0}$ in the BF matrix element and in the electron wave function:

$$\sigma(\omega)=8(2\pi)^2\frac{e^2}{nc\omega}\left|\left\langle c\mathbf{k}=\mathbf{0}\left|\frac{p}{m}\right|v\mathbf{k}=\mathbf{0}\right\rangle\right|^2|\psi_e(\mathbf{0})|^2 \times\left|\int d^3k\ \theta(\mathbf{k})\psi_h(\mathbf{k})\right|^2\delta(\hbar\omega-\hbar\omega_0). \tag{3.27}$$

Using Eq. (3.21) for the hole wave function, the integral over $\mathbf{k}$ space can be performed, and we obtain

$$\sigma(\omega)=\frac{16}{3}(4\pi)^2\left(\frac{e^2}{\hbar c}\right)\frac{1}{n\hbar\omega}|X|^2\left[\frac{E_g+\Delta}{\Delta}\right]^2 \times(m_h^*/m_e^*)^{3/2}|E_1|^{1/2}|E_2|^{3/2}\delta(\hbar\omega-\hbar\omega_0), \tag{3.28}$$

where

n (index of refraction) $=3.45$,

$\Delta=\epsilon_c(\Gamma)-\epsilon_c(X)=0.5$ eV,

E_g (indirect energy gap) $=2.3$ eV,

$\hbar\omega_0=(E_g-|E|\text{(bound state)})$,

$|X|^2=|\langle c\mathbf{k}=\mathbf{0}|x|v\mathbf{k}=\mathbf{0}\rangle|^2$
$=|\langle c\mathbf{k}=\mathbf{0}|p/m|v\mathbf{k}\equiv\mathbf{0}\rangle|^2/(E_g+\Delta)^2$
$=$ optical absorption matrix element.

Because no broadening processes have been considered, the cross section as a function of frequency is represented by a δ function. The expression for $\sigma(\omega)$ can be integrated and this integrated absorption can be compared to the area under the experimental absorption curve.

$$\int\sigma(\omega)d(\hbar\omega)=\frac{16}{3}(4\pi)^2\left(\frac{e^2}{\hbar c}\right)\frac{1}{n\hbar\omega_0}|X|^2\left[\frac{E_g+\Delta}{\Delta}\right]^2 \times(m_h^*/m_e^*)^{3/2}|E_1|^{1/2}|E_2|^{3/2} =1.01\times|X|^2\ (\text{meV}). \tag{3.29}$$

Experimental areas give

$$\int\sigma(\omega)d(\hbar\omega)=(4.5\pm1.0)\ \text{Å}^2\ \text{meV}. \tag{3.30}$$

If $|X|=(2.1\pm0.3)$ Å, not at all an unreasonable value for the optical matrix element, the area under the experimental absorption curve is well accounted for.

D. Optical Absorption Above the Band Edge

Yet another experimental observation can be understood using the simple model: the optical absorption above the band edge induced by the nitrogen impurities.

Considering only purely electronic transitions, the optical absorption occurs principally in three processes: (a) the creation of a bound electron and an unbound hole, (b) the creation of an unbound electron and an unbound hole, and (c) the creation of an unbound exciton. Processes (a) and (c) have one-particle densities of final states which lead to square-root absorption thresholds. Process (b) has a two-particle density of final states and, consequently, an E^2 threshold.

Process (a) is the simplest to calculate because the wave function for the electron+impurity bound state has already been determined. The hole can be taken to be free if the Coulomb interaction is ignored. This is undoubtedly a bad approximation at the absorption threshold, but it improves for higher energy. Making this approximation gives us the electron-hole wave function for this process:

$$\psi(\mathbf{k}_1,\mathbf{k}_2)=\psi_e(\mathbf{k}_1)\delta(\mathbf{k}_2-\mathbf{k}_0), \tag{3.31}$$

where $\mathbf{k}_0$ is the final hole momentum, and $\psi_e(\mathbf{k}_1)$ is given

by Eq. (3.19) with $|E_{el}|=8$ meV, the binding energy of the electron by itself to the nitrogen impurity.

Using the effective-mass approximation for the hole and the same approximations made before in the case of the absorption at the exciton bound-state energy, we obtain the square-root absorption curve for frequencies greater than, but near, the absorption edge:

$$\sigma(\omega)=\frac{16}{3}(4\pi)\left(\frac{e^2}{\hbar c}\right)|X|^2\left[\frac{E_g+\Delta}{\Delta}\right]^2\frac{|E_{el}|^{1/2}}{n\hbar\omega}\times(m_h^*/m_e^*)^{3/2}(\hbar\omega-\hbar\omega_0)^{1/2},$$
$$\hbar\omega_0=E_g-|E_{el}|. \tag{3.32}$$

Process (b), the production of unbound electrons and unbound holes, is one step more complicated. We will now need the continuum-electron wave function.

Schrödinger's equation for the electron in interaction with the impurity, Eq. (3.6), can be written for $E>0$:

$$\psi_{\mathbf{k}_0}(\mathbf{k})=\delta(\mathbf{k}-\mathbf{k}_0)+\frac{1}{[E+i0-\epsilon_c(\mathbf{k})]}\frac{J\Omega}{(2\pi)^3}\int d^3q\,\theta(\mathbf{q})\psi_{k_0}(\mathbf{q}). \tag{3.33}$$

Here, $\mathbf{k}$ is the argument of the wave function and $\mathbf{k}_0$ is the label of the state represented by ψ; $\delta(\mathbf{k}-\mathbf{k}_0)$ is the unperturbed wave function; and $E=\epsilon_c(\mathbf{k}_0)$ is the energy of the state.

Writing ψ in the form above preserves the normalization,

$$\int d^3k\,\theta(\mathbf{k})\psi_{\mathbf{k}_1}{}^*(\mathbf{k})\psi_{\mathbf{k}_2}(\mathbf{k})=\delta(\mathbf{k}_1-\mathbf{k}_2). \tag{3.34}$$

Integrating Eq. (3.33) over the BZ yields

$$\int d^3k\,\theta(\mathbf{k})\psi_{\mathbf{k}_0}(\mathbf{k})=1+\frac{J\Omega}{(2\pi)^3}\int d^3k\times\theta(\mathbf{k})\frac{1}{[E+i0-\epsilon_c(\mathbf{k})]}\int d^3q\,\theta(\mathbf{q})\psi_{\mathbf{k}_0}(\mathbf{q}). \tag{3.35}$$

Defining

$$f(E)=\frac{\Omega}{(2\pi)^3}\int d^3k\,\theta(\mathbf{k})\frac{1}{[\epsilon_c(\mathbf{k})-E-i0]}, \tag{3.36}$$

we can solve for $\int d^3q\,\theta(\mathbf{q})\psi_{k_0}(\mathbf{q})$:

$$\int d^3q\,\theta(\mathbf{q})\psi_{k_0}(\mathbf{q})=\frac{1}{[1+Jf(E)]}. \tag{3.37}$$

The electron wave function then becomes

$$\psi_{\mathbf{k}_0}(\mathbf{k})=\delta(\mathbf{k}-\mathbf{k}_0)+\frac{J\Omega}{(2\pi)^3}\frac{1}{[1+Jf(E)]}\frac{1}{[E+i0-\epsilon_c(\mathbf{k})]}. \tag{3.38}$$

The function $f(E)$ has appeared before in the bound-state problem where the eigenvalue equation for $E<0$ was

$$[1+Jf(E)]=0. \tag{3.39}$$

For energies E above the indirect minima but below the direct relative minimum of the conduction band

$$\psi_{\mathbf{k}_0}(\mathbf{0})=\frac{J\Omega}{(2\pi)^3}\frac{1}{[1+Jf(E)]}\frac{1}{[E-\Delta]},\quad E=\epsilon_c(\mathbf{k}_0). \tag{3.40}$$

Had we used the Born approximation for this calculation, setting $\psi_{\mathbf{k}_0}(\mathbf{k})=\delta(\mathbf{k}-\mathbf{k}_0)$ in the right side of Eq. (3.33), we would have obtained

$$\psi_{\mathbf{k}_0}(\mathbf{k})=\delta(\mathbf{k}-\mathbf{k}_0)+\frac{J\Omega}{(2\pi)^3}\frac{1}{[E+i0-\epsilon_c(\mathbf{k})]}, \tag{3.41}$$

$$\psi_{\mathbf{k}_0}(\mathbf{0})=\frac{J\Omega}{(2\pi)^3}\frac{1}{[E-\Delta]}\ \text{(Born approx.)}. \tag{3.42}$$

But because of the existence of a shallow bound state, $|1+Jf(E)|\ll 1$ for E near zero (recall that $[1+Jf(0)]/J=0.022$ eV^{-1}). Since the optical absorption is approximately proportional to $|\psi(\mathbf{0})|^2$, we would have lost several orders of magnitude had we used the Born approximation for the continuum-electron wave function.

If the Coulomb interaction between the electron and the hole is ignored, we can take the hole to be a free particle and take its wave function to be a momentum-space δ function. The electron+hole wave function is then

$$\psi(\mathbf{k}_1,\mathbf{k}_2)=\psi_{\mathbf{k}_e}(\mathbf{k}_1)\delta(\mathbf{k}_2-\mathbf{k}_h), \tag{3.43}$$

where $\mathbf{k}_h$ is the final hole momentum.

Using Eq. (2.22) for the optical absorption cross section,

$$\left|\int d^3k\,\theta(\mathbf{k})\psi^*(c\mathbf{k},v\mathbf{k})\langle c\mathbf{k}|p/m|v\mathbf{k}\rangle\right|^2 \simeq|\langle c\mathbf{0}|p/m|v\mathbf{0}\rangle|^2|\psi_{\mathbf{k}_e}(0)|^2\left|\int d^3k\,\delta(\mathbf{k}-\mathbf{k}_h)\right|^2 \simeq(E_g+\Delta)^2|X|^2\left|\frac{J\Omega}{(2\pi)^3}\frac{1}{\Delta}\right|^2\frac{1}{|1+Jf(E)|^2}, \tag{3.44}$$

where E has been assumed small compared with Δ.

Summing over the final electron and hole states gives

$$\sigma(\omega)=\frac{32}{2}\left(\frac{e^2}{\hbar c}\right)\left[\frac{E_g+\Delta}{\Delta}\right]^2|X|^2\left(\frac{m_h^*}{m_e^*}\right)^{3/2}\frac{1}{n\hbar\omega}\times\int_0^{(\hbar\omega-E_g)}dE\frac{E^{1/2}(\hbar\omega-E_g-E)^{1/2}}{[\varphi(E)+E]}, \tag{3.45}$$

where

$$\varphi(E)=\frac{[1+J\,\mathrm{Re}(f(E))]^2}{36\pi^4(2m_e^*)^3[J\Omega/(2\pi)^3]^2} \tag{3.46}$$

and the term E accompanying $\varphi(E)$ comes from $\mathrm{Im}[f(E)]$ which is proportional to the density of states ($E^{1/2}$).

We now need to know $\mathrm{Re}[f(E)]$ for $E>0$. The trick used before for $E<0$ does not work in this case, so that we must resort to something else.

If the density of states for the entire conduction band were known, $f(E)$ could be calculated. In the absence of this knowledge, and in order to keep the calculations simple, a model density of states can be used as shown in Fig. 3. $\mathrm{Re}[f(E)]$ can be calculated from this density of states and is also shown in Fig. 3.

The essential features of $\mathrm{Re}[f(E)]$ illustrated here which would hold true for any density of states beginning as $E^{1/2}$ and having a finite total integral are

(1) $\mathrm{Re}[f(E)]$ has a square-root singularity as $E\to 0$ from below:

$$f(E)\to\langle 1/\epsilon\rangle-\text{const}\times|E|^{1/2}. \tag{3.47}$$

(2) $\mathrm{Re}[f(E)]$ is analytic about $E=0$ for $E\to 0$ from above:

$$\mathrm{Re}[f(E)]\to\langle 1/\epsilon\rangle+\text{const}\times E+\text{const}\times E^2+\cdots. \tag{3.48}$$

$1+J\,\mathrm{Re}[f(E)]$ is sketched in Fig. 3 also. There is a point for $E>0$ at which it equals zero. At this point, $\varphi(E)=0$ in Eq. (3.45), greatly augmenting the optical absorption. We can, in fact, obtain an upper limit on $\sigma(\omega)$ for this process by setting $\varphi(E)=0$ for all E in the integral of Eq. (3.45):

$$\int_0^{\hbar\omega-E_g}dE\frac{E^{1/2}(\hbar\omega-E_g-E)^{1/2}}{[\varphi(E)+E]}$$
$$\leq\int_0^{\hbar\omega-E_g}dE\frac{(\hbar\omega-E_g-E)^{1/2}}{E^{1/2}}=\tfrac{1}{2}\pi(\hbar\omega-E_g). \tag{3.49}$$

The resonance effect acts to change the $(\hbar\omega-E_g)^2$ form for the optical absorption arising from the two-particle density of final states to the linear form $(\hbar\omega-E_g)$.

Using the linear form $\mathrm{Re}[f(E)]=\langle 1/\epsilon\rangle(1-\rho E)$, the integral can be performed and the optical absorption cross section expressed in closed form. The results are not sensitive to the value of ρ. Figure 4 shows the cross section calculated for $\rho\langle 1/\epsilon\rangle=0.82/(\mathrm{eV})^2$. The curve goes as $(\hbar\omega-\hbar\omega_0)^2$ at threshold and approaches $(\hbar\omega-\hbar\omega_0)^{1/2}$ asymptotically.

The final process, the production of unbound excitons, is the most complicated of all but it can be calculated in much the same way as the previous process.

We must now consider Schrödinger's equation for both the electron and the hole:

$$[E-\epsilon_c(\mathbf{k}_1)-\epsilon_v(\mathbf{k}_2)-V_{\mathrm{Coul}}]\psi(\mathbf{k}_1,\mathbf{k}_2)$$
$$=\frac{J\Omega}{(2\pi)^3}\int d^3q\;\theta(\mathbf{q})\psi(\mathbf{q},\mathbf{k}_2). \tag{3.50}$$

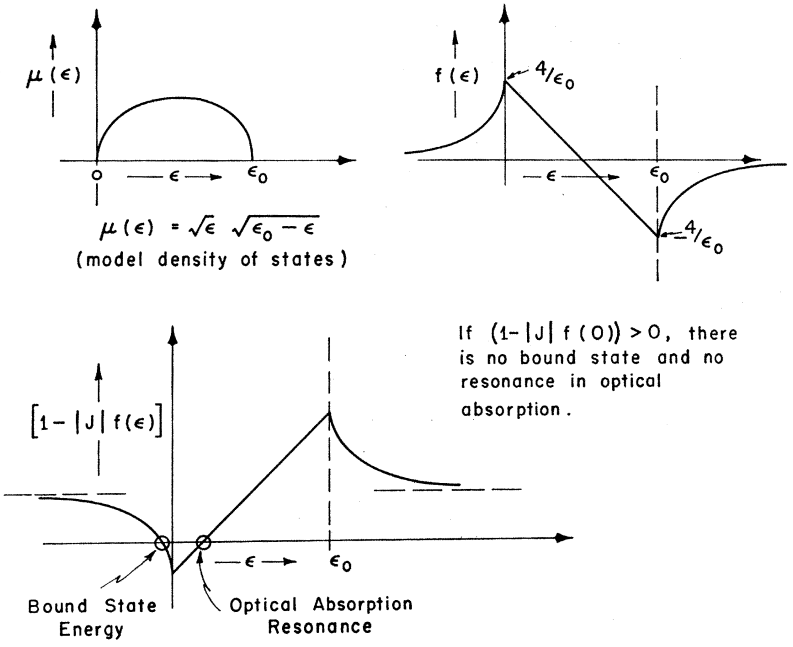

FIG. 3. Model density of states for the GaP conduction band and the function $f(E)=\langle 1/(\epsilon(\mathbf{k})-E)\rangle_{\mathrm{BZ}}$. Only the real part of $f(E)$ is shown here. The imaginary part is proportional to $\mu(E)$, the density of states.

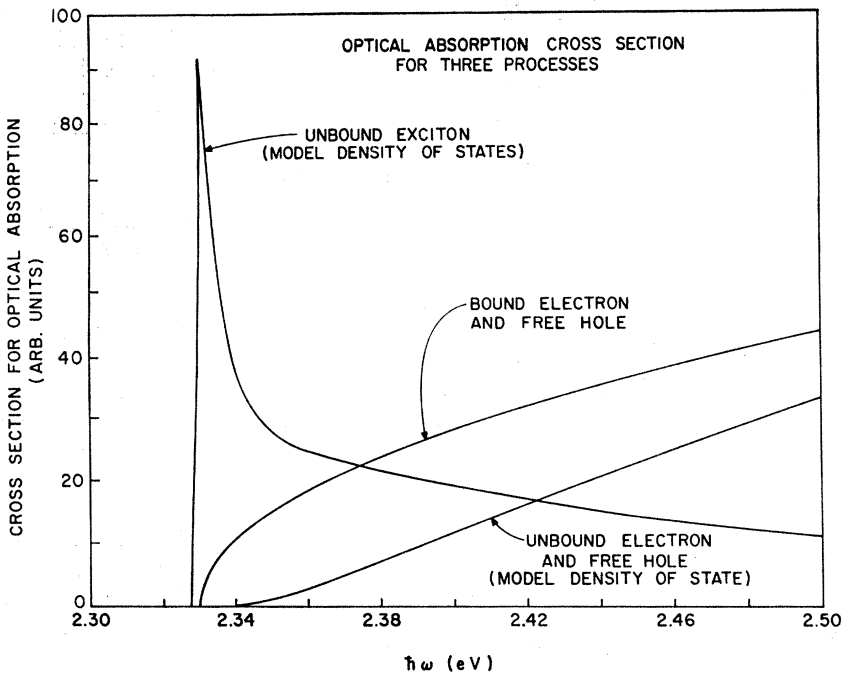

FIG. 4. Theoretical optical absorption cross sections for three processes in GaP for electrons and holes in interaction with nitrogen impurities.

We can write this equation formally:

$$\psi(\mathbf{k}_1,\mathbf{k}_2)=\psi_0(\mathbf{k}_1,\mathbf{k}_2)+\frac{1}{[E+i0-\epsilon_c(\mathbf{k}_1)-\epsilon_v(\mathbf{k}_2)]}\times\frac{J\Omega}{(2\pi)^3}\int d^3q\,\theta(\mathbf{q})\psi(\mathbf{q},\mathbf{k}_2)+(\text{Power series in } V_{\text{Coul}}),\quad(3.51)$$

where $\psi_0(\mathbf{k}_1,\mathbf{k}_2)$ is the wave function of the unperturbed exciton.

If we ignore all Coulomb interactions except that necessary for $\psi_0(\mathbf{k}_1,\mathbf{k}_2)$, that is, if we drop the power series in V_{Coul} from Eq. (3.51), we obtain a soluble equation. This approximation, along with the same approximations made for the other two processes, is undoubtedly bad near the absorption edge, but it improves for higher energies.

Making this approximation, integrating the resulting equation on $\mathbf{k}_1$, and solving algebraically gives

$$\int d^3q\,\theta(\mathbf{q})\psi(\mathbf{q},\mathbf{k}_2)=\frac{1}{[1+Jf(E-\epsilon_v(\mathbf{k}_2))]}\int d^3q\,\theta(\mathbf{q})\psi_0(\mathbf{q},\mathbf{k}_2).\quad(3.52)$$

This gives us the exciton wave function

$$\psi(\mathbf{k}_1,\mathbf{k}_2)=\psi_0(\mathbf{k}_1,\mathbf{k}_2)+\frac{1}{[E+i0-\epsilon_c(\mathbf{k}_1)-\epsilon_v(\mathbf{k}_2)]}\times\frac{1}{[1+Jf(E-\epsilon_v(\mathbf{k}_2))]}\frac{J\Omega}{(2\pi)^3}\int d^3q\theta(\mathbf{q})\psi_0(\mathbf{q},\mathbf{k}_2),\quad(3.53)$$

again exhibiting the resonant denominator

$$[1+Jf(E-\epsilon_v(\mathbf{k}))].\quad(3.54)$$

The unperturbed exciton wave function for energies near the exciton band edge is, assuming spherical masses,

$$\psi_0(\mathbf{k}_1,\mathbf{k}_2)=\delta(\mathbf{k}_1-\tfrac{1}{2}\mathbf{G}-\mathbf{k}_2-\mathbf{K})\times\varphi_0\left(\frac{\mu}{m_e^*}(\mathbf{k}_1-\tfrac{1}{2}\mathbf{G})+\frac{\mu}{m_h^*}\mathbf{k}_2\right).\quad(3.55)$$

Here, μ is the electron-hole reduced mass, $m_e^*m_h^*/(m_e^*+m_h^*)$; $\frac{1}{2}\mathbf{G}$ is the position of one of the conduction-band minima; $\mathbf{K}$ is the momentum of the exciton. $E=-|E_{\text{ex}}|+K^2/2M$, $M=m_e^*+m_h^*$. $\varphi_0(\mathbf{k})$ is the Coulomb wave function of the electron-hole system in the c.m. frame,

$$(2\sqrt{2}/\pi)a^{3/2}(1+k^2a^2)^{-2},$$

$a^2=1/2\mu|E_{\text{ex}}|$. $|E_{\text{ex}}|$ is the exciton binding energy, 0.010 eV. Because of the δ function in $\psi_0(\mathbf{k}_1,\mathbf{k}_2)$, $\psi_0(\mathbf{k},\mathbf{k})=0$. Therefore,

$$\psi(\mathbf{k},\mathbf{k})=\frac{1}{[E-\epsilon_0(\mathbf{k})-\epsilon_v(\mathbf{k})]}\frac{1}{[1+Jf(E-\epsilon_v(\mathbf{k}))]}\times\frac{J\Omega}{(2\pi)^3}\varphi_0\left(\mathbf{k}+\frac{\mu}{m_e^*}\mathbf{K}\right),\quad(3.56)$$

where the δ function has been used to perform the integral on ψ_0.

For small values of $\mathbf{K}$, $E=-|E_{\text{ex}}|+K^2/2M$ is small compared to $\Delta=0.5$ eV and $\mathbf{k}$ is restricted to small

values by φ_0. We may then approximate:

$$\psi(\mathbf{k},\mathbf{k})=-\frac{1}{\Delta}\frac{1}{[1+Jf(E-k^2/2m_h^*)]}\times\frac{J\Omega}{(2\pi)^3}\varphi_0\left(\mathbf{k}+\frac{\mu}{m_e^*}\mathbf{K}\right). \quad (3.57)$$

Integrating $\psi(\mathbf{k},\mathbf{k})$ and performing the sums over the final states yields

$$\sigma(\omega)=768\left(\frac{e^2}{\hbar c}\right)\left[\frac{E_g+\Delta}{\Delta}\right]^2|X|^2\frac{1}{n\hbar\omega}\frac{\Omega^2}{(2\pi)^3}\frac{1}{Q^2}\times(m_h^*/m_e^*)^{3/2}|E_{ex}|^{3/2}(2M)^3\times(\hbar\omega-\hbar\omega_0)^{1/2}|w(\hbar\omega-\hbar\omega_0)|^2, \quad (3.58)$$

where

$$\hbar\omega_0=E_g-|E_{ex}|,$$

$$w(x)=\int_0^\infty dz\,(z)^{1/2}\frac{JQ}{[1+Jf(x-|E_{ex}|-|E_{ex}|z)]}\times\left\{\left[1-\frac{m_h^*}{m_e^*}\frac{x}{|E_{ex}|}+\frac{M}{m_e^*}z\right]^2+4\frac{m_h^*}{m_e^*}\frac{x}{|E_{ex}|}\right\}^{-1}. \quad (3.59)$$

The evaluation of $w(x)$ must be done separately for three regions:

(1) $x<|E_{ex}|-|E_{el}|$, $|E_{el}|$ = binding energy of an electron to the nitrogen.
(2) $|E_{ex}|-|E_{el}|<x<|E_{ex}|$.
(3) $x>|E_{ex}|$.

These boundary values represent the thresholds for the two competing processes: (a) bound electrons+free holes and (b) unbound electrons+free holes, both of which fall within the range of optical absorption due to unbound excitons.

In evaluating $w(x)$, the argument of $f(x-|E_{ex}|-|E_{ex}|z)$ can be positive for $x>|E_{ex}|$. $f(\omega)$ is taken to be

$$f(\omega)=\left\langle\frac{1}{\epsilon_c}\right\rangle-\frac{3\Omega}{4\pi}(2m_e^*)^{3/2}(-\omega)^{1/2}, \qquad \omega<0 \quad (3.60)$$

$$f(\omega)=\left\langle\frac{1}{\epsilon_c}\right\rangle(1-\rho\omega)+i\frac{3\Omega}{4\pi}(2m_e^*)^{3/2}(\omega)^{1/2}, \quad \omega>0 \quad (3.61)$$

where ρ is the same parameter introduced earlier in the consideration of process (b).

The analytical results are lengthy and are not presented here.

Figure 4 shows a plot of the optical absorption cross section for this process using $\rho\langle 1/\epsilon\rangle=0.82/(\text{eV})^2$, the same value as for competing process (b).

The sharp peak in this curve occurs at the threshold for the production of bound electrons and free holes and is a resonance effect due to the appearance of this cross-channel square-root threshold. If the Coulomb interaction were treated better, this peak would become much less pronounced.

Figure 5 shows the sum of the three processes for optical absorption considered here superimposed on the experimental curve, which has had the background intrinsic absorption subtracted out.

The first thing that one is struck by on comparing the two curves is the lack of any structure on the theoretical curve such as is present on the experimental curve. The reason, of course, is that phonons have been totally ignored in this treatment.

Several resonance effects are present in this absorp-

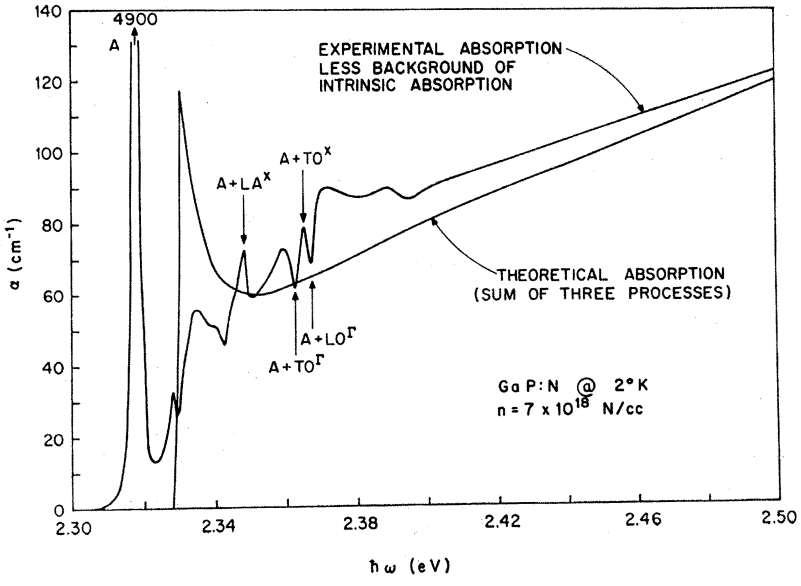

FIG. 5. Comparison of the theoretical optical absorption coefficient for GaP:N with experiment. (Experimental spectrum by Dean and Thomas, Ref. 6.)

tion curve. Hopfield *et al.* have discussed these effects.[6] Of particular note are two positions marked $A+\mathrm{TO}^{\Gamma}$ and $A+\mathrm{LO}^{\Gamma}$, where ordinary phonon replicas of the principal absorption line (A line) would be expected. Instead, due to cross-channel interference, valleys occur instead of peaks.

The second thing that one is struck by is the sharp peak in the theoretical curve, which is much less pronounced in the experimental curve. As was discussed previously, a better treatment of the Coulomb interaction should wash the theoretical peak out, bringing it more into line with experiment.

With these exceptions noted, the theoretical curve is seen nonetheless to account very well for almost all of the absorption induced by the nitrogen impurity well above the band edge.

The value of the optical matrix element used for the theoretical curve is $|X|=1.66$ Å, slightly less than that determined from the area under the principal absorption peak (2.1 ± 0.3 Å) but still quite reasonable.

It is satisfying to see that the tail of the theoretical curve is practically parallel to the experimental curve since the approximations concerning the Coulomb interactions are much better in this region than near the thresholds.

It is also worthwhile to point out again at this point that if the Born approximation had been used for the electron wave functions, the magnitude of the theoretical optical absorption curve would be four orders of magnitude smaller than is shown here.

E. Electron Scattering Cross Section

The scattering cross section for Bloch electrons in the conduction-band scattering from a nitrogen impurity can be calculated from the electron continuum wave function (3.38).

The total cross section for scattering out of a pure Bloch state is

$$\sigma_{\mathrm{SCT}}(E)=\frac{3}{\pi}(m_e^*J\Omega)^2\frac{1}{|1+Jf(E)|^2} \tag{3.62}$$

or

$$\sigma_{\mathrm{SCT}}(E)=\frac{4\pi}{3}\Big/\left[2m_e^*E+\left(\frac{2\pi(1+J\,\mathrm{Re}[f(E)])}{3m_e^*J\Omega}\right)^2\right]. \tag{3.63}$$

For electrons near the minimum of the conduction band, $E\to 0$,

$$\begin{aligned}\sigma_{\mathrm{SCT}}(E)&=(3/\pi)(m_e^*\Omega)^2(1/Q^2)\\&=6300\ (\text{Å})^2.\end{aligned} \tag{3.64}$$

We can do the same thing for the hole except that now $1+J\langle 1/\epsilon\rangle$ is no longer near zero as it was for the electron but is of order 1. This reduces the scattering cross section by four orders of magnitude. This can be considered an *a posteriori* justification for ignoring the hole-impurity interaction in the beginning.

F. Double-Impurity Bound States

We return to the bound-state problem in the one-band–one-site approximation, now considering the case of an electron bound to a pair of impurity atoms separated by a lattice vector $\mathbf{R}$. We let one atom be at the origin of coordinates and the other be centered about the lattice site at $\mathbf{R}$. Then, according to Eq. (3.3), the potential matrix element should be

$$(c\mathbf{k}|V_{\mathrm{II}}|c\mathbf{q})\simeq[J\Omega/(2\pi)^3][1+e^{-i(\mathbf{k}-\mathbf{q})\cdot\mathbf{R}}], \tag{3.65}$$

and Schrödinger's equation for the electron becomes

$$[E-\epsilon_c(\mathbf{k})]\psi(\mathbf{k})=\frac{J\Omega}{(2\pi)^3}\int d^3q\,\theta(\mathbf{q})[1+e^{-i(\mathbf{k}-\mathbf{q})\cdot\mathbf{R}}]\psi(\mathbf{q}). \tag{3.66}$$

This equation has an immediate solution for $E<0$:

$$\psi(\mathbf{k})=\frac{A+Be^{-\mathbf{k}\cdot\mathbf{R}}}{(\epsilon_c(\mathbf{k})-E)}, \tag{3.67}$$

where

$$A=-\frac{J\Omega}{(2\pi)^3}\int d^3q\,\theta(\mathbf{q})\psi(\mathbf{q}), \tag{3.68}$$

$$B=-\frac{J\Omega}{(2\pi)^3}\int d^3q\,\theta(\mathbf{q})e^{i\mathbf{q}\cdot\mathbf{R}}\psi(\mathbf{q}). \tag{3.69}$$

Let us define a new function related to $f(E)$:

$$f(E,\mathbf{R})\equiv\frac{\Omega}{(2\pi)^3}\int d^3q\,\theta(\mathbf{q})\frac{e^{i\mathbf{q}\cdot\mathbf{R}}}{(\epsilon_c(\mathbf{q})-E)}, \qquad f(E)=f(E,0). \tag{3.70}$$

Substituting the wave function Eq. (3.67) into Eqs. (3.68) and (3.69) yields two equations for the two unknowns A and B:

$$\begin{aligned}[1+Jf(E)]A+Jf(E,\mathbf{R})B&=0,\\Jf^*(E,\mathbf{R})A+[1+Jf(E)]B&=0.\end{aligned} \tag{3.71}$$

Requiring the determinant of the coefficients to vanish gives the eigenvalue equation

$$[1+Jf(E)]^2=|Jf(E,\mathbf{R})|^2. \tag{3.72}$$

Because of the crystal symmetry, $f(E,\mathbf{R})$ is real and we can write the two eigenvalue equations

$$\begin{aligned}1+Jf(E)&=Jf(E,\mathbf{R}), &&\text{(case I)}\\1+Jf(E)&=-Jf(E,\mathbf{R}). &&\text{(case II)}\end{aligned} \tag{3.73}$$

One of these equations gives an energy level below the single-impurity level and the other gives a level above the single-impurity level (if it can be satisfied at all for negative energies). These are the familiar bonding and antibonding states of molecular calculations.

In case I, by the eigenvalue equation,

$$\psi_{\mathrm{I}}(\mathbf{k})=\frac{1-e^{-i\mathbf{k}\cdot\mathbf{R}}}{[\epsilon_c(\mathbf{k})-E]} \tag{3.74}$$

and in case II,

$$\psi_{\mathrm{II}}(\mathbf{k})=\frac{1+e^{-i\mathbf{k}\cdot\mathbf{R}}}{[\epsilon_c(\mathbf{k})-E]}. \tag{3.75}$$

We have seen before that the optical absorption cross section is approximately proportional to $|\psi(0)|^2$. At $\mathbf{k}=0$, $\psi_{\mathrm{I}}(0)=0$, and $\psi_{\mathrm{II}}(0)\simeq 2/\Delta$. Therefore, the states of case I are first-order forbidden optical transitions and the states of case II are allowed.

If $f(E,\mathbf{R})>0$, the case-II state lies lower in energy than case I. If $f(E,\mathbf{R})<0$, the case-I state lies lower.

We already have an expression for $(1+Jf(E))$ for small negative energies from the work done on the single-impurity bound state. To do a good job of evaluating $f(E,\mathbf{R})$, knowledge of the band structure throughout the Brillouin zone is again necessary. However, a good insight into the nature of the wave-function interference effects in this problem can be gained by simply using the effective-mass approximation and extending the integral over all $\mathbf{k}$ space. The oscillatory complex exponential causes the integral to converge. Let

$$\begin{aligned}\mathbf{G}_1&=(2\pi/a)(2,0,0),\\ \mathbf{G}_2&=(2\pi/a)(0,2,0),\\ \mathbf{G}_3&=(2\pi/a)(0,0,2).\end{aligned} \tag{3.76}$$

The $\mathbf{G}$'s are reciprocal-lattice vectors pointing toward the three valleys of the GaP conduction band.

If we were to use isotropic masses for each valley ($m_1=m_2=m^*$), we would obtain

$$f(E,\mathbf{R})=m^*\frac{\Omega}{(2\pi)}\frac{\exp\{-(2m^*|E|)^{1/2}R\}}{R}\Gamma(\mathbf{R}), \tag{3.77}$$

$$\Gamma(\mathbf{R})=\exp(i\tfrac{1}{2}\mathbf{G}_1\cdot\mathbf{R})+\exp(i\tfrac{1}{2}\mathbf{G}_2\cdot\mathbf{R})+\exp(i\tfrac{1}{2}\mathbf{G}_3\cdot\mathbf{R}). \tag{3.78}$$

Each of the exponentials in $\Gamma(\mathbf{R})$ can be either $+1$ or -1, so that $\Gamma(\mathbf{R})$ can be ± 1 or ± 3. In fact, it turns out that $\Gamma(\mathbf{R})$ is either $+3$ or -1 for different values of $\mathbf{R}$.

$\Gamma(\mathbf{R})=3$ yields a lower-lying state than $\Gamma(\mathbf{R})=-1$, other things being equal, and speaking only of the state lying below the single-impurity level, $\Gamma(\mathbf{R})=-1$ yields an optically forbidden state, and $\Gamma(\mathbf{R})=3$ yields an allowed state.

If we use the fact that the masses in each valley are anisotropic, $m^*=(m_1,m_1,m_2)$, we get further striking effects:

$$f(E,\mathbf{R})=f_1+f_2+f_3, \tag{3.79}$$

where

$$f_1=m_1\frac{\Omega}{(2\pi)}\frac{1}{R}\exp(i\tfrac{1}{2}\mathbf{G}_1\cdot\mathbf{R})\frac{\exp\{-(2m_2|E|)^{1/2}R[\cos^2\theta_1+(m_1/m_2)\sin^2\theta_1]^{1/2}\}}{[\cos^2\theta_1+(m_1/m_2)\sin^2\theta_1]^{1/2}}. \tag{3.80}$$

Here, θ_1 is the angle between $\mathbf{R}$ and $\mathbf{G}_1$; f_2 is the same function with $\mathbf{G}_2$ substituted for $\mathbf{G}_1$; and f_3 is the same function with $\mathbf{G}_3$ substituted for $\mathbf{G}_1$.

Using the pseudopotential calculation of Bergstresser and Cohen,[8] the effective masses of GaP are $m_1=0.26m$ and $m_2=1.49m$. With such a great difference between the masses, $f(E,\mathbf{R})$ becomes quite sensitive to the orientation of $\mathbf{R}$ in addition to its ordinary dependence on the magnitude of $\mathbf{R}$.

This scheme for the double nitrogen-impurity bound states, due partly to the crudeness of the evaluation of $f(E,\mathbf{R})$, does not yield energy levels in agreement with experiment. Computer calculations have been performed that use a realistic potential and that take the band structure of GaP more fully into account. These will be presented in the next section.

IV. COMPUTER CALCULATIONS OF THE BOUND-STATE PROBLEM

The one-band–one-site approximation cannot be considered an adequate description of isoelectronic impurities despite its success in certain areas. Most notable among its failures is its inability to predict the energies of any bound states or, given the energy of the single-nitrogen bound state, to predict the energies of the double-nitrogen lines correctly. Furthermore, the excited states of the single-nitrogen level do not appear, even incorrectly.

Therefore a more ambitious program was undertaken, to calculate the bound states from more or less first principles.

To do this requires a number of items of information. First, the band structure and BF's of the host crystal must be known throughout the BZ. This can be generated from the empirical pseudopotential form factors given for GaP and many other semiconductors by Cohen and Bergstresser.[8] Second, a pseudopotential representing the difference between the impurity and the atom it replaces is needed. The atomic structure calculations of Herman and Skillman[10] provide the core-electron wave functions necessary for this task.

[10] Frank Herman and Sherwood Skillman, *Atomic Structure Calculations* (Prentice-Hall, Inc., Englewood Cliffs, N. J., 1963).

It also involves the problems of choice of phase for the numerically calculated BF's and the band assignments to be made for the different energy levels.

A. Band Structure and Bloch Functions of GaP

The periodic parts of the Bloch functions can be written as

$$U_n(\mathbf{k},\mathbf{x})=\sum_G a_n(\mathbf{k},\mathbf{G})e^{i\mathbf{G}\cdot\mathbf{x}}/(\Omega)^{1/2} \tag{4.1}$$

and therefore the complete BF's are

$$\psi_n(\mathbf{k},\mathbf{x})=[1/(2\pi)^{3/2}]\sum_G a_n(\mathbf{k},\mathbf{G})e^{i(\mathbf{G}+\mathbf{k})\cdot\mathbf{x}}. \tag{4.2}$$

Here, $\mathbf{k}$ is a reduced wave vector in the first BZ and the $\mathbf{G}$'s are reciprocal-lattice vectors.

The crystal pseudopotential Hamiltonian is

$$H=-\hbar^2\nabla^2/2m+V(\mathbf{x}), \tag{4.3}$$

$$V(\mathbf{x})=\sum_{\mathbf{G}}[S^S(\mathbf{G})V_{\mathbf{G}}{}^S+iS^A(\mathbf{G})V_{\mathbf{G}}{}^A]e^{-i\mathbf{G}\cdot(\mathbf{x}-\boldsymbol{\tau})}, \tag{4.4}$$

where the notation of Cohen and Bergstresser has been used. The pseudopotential $V(\mathbf{x})$ differs from that of Bergstresser and Cohen in that the origin of coordinates is taken at a phosphorous site rather than half-way between a phosphorous site and a gallium site. This is done with an eye to the future because the nitrogen impurity will be located at a phosphorous site.

$$\boldsymbol{\tau}=\tfrac{1}{8}(1,1,1)a, \tag{4.5}$$

where a is the length of a unit cube.

$$\begin{aligned} S^S(\mathbf{G})&=\cos\mathbf{G}\cdot\boldsymbol{\tau},\\ S^A(\mathbf{G})&=\sin\mathbf{G}\cdot\boldsymbol{\tau}. \end{aligned} \tag{4.6}$$

$V_{\mathbf{G}}{}^S$ and $V_{\mathbf{G}}{}^A$, the pseudopotential form factors, are given for GaP by Cohen and Bergstresser.[8] Here, $\mathbf{G}$ is measured in units of $(2\pi/a)$.

To find the energy levels and the coefficients $a_n(\mathbf{k},\mathbf{G})$ in Eq. (4.1), the basis states $e^{i(\mathbf{G}+\mathbf{k})\cdot\mathbf{x}}/\Omega^{1/2}$, with $(\mathbf{G}+\mathbf{k})^2\leq 7$, were used to form the Hamiltonian matrix with perturbation corrections applied using states with $7<(\mathbf{G}+\mathbf{k})^2\leq 21$ according to Löwdin's method.[11] The matrix is then diagonalized to find the eigenvalues and eigenvectors. The computer program that does all of this was originally written by Brust.[12]

The complex matrices involved are approximately 20×20 and the entire calculation takes approximately 10 sec of IBM 7094 computer time for each point $\mathbf{k}$.

B. Impurity Pseudopotential

The assumption contained in the band-structure calculation, and continued here, is that when atoms are placed in a regular array forming a crystal only the outer electrons are modified; the electrons making up the filled shells underneath are not appreciably affected. One can then go on to talk only about the valence electrons if, somewhere in the formalism, account is taken of the Pauli exclusion principle. Formally, this simply requires that all wave functions be orthogonal to the unchanged core wave functions.

If we write a Schrödinger equation for a true wave function,

$$(H_0+V)|\psi\rangle=E|\psi\rangle, \tag{4.7}$$

and require $|\psi\rangle$ to be orthogonal to a number of core states $|C\rangle$,

$$\begin{aligned} \langle C|\psi\rangle&=0, \quad \text{all } |C\rangle\\ \langle C|C'\rangle&=\delta_{CC'}, \end{aligned} \tag{4.8}$$

then we can write

$$|\psi\rangle=[1-\sum_C |C\rangle\langle C|]|\varphi\rangle \tag{4.9}$$

and no matter what $|\varphi\rangle$ we choose, $|\psi\rangle$ will automatically be orthogonal to all the $|C\rangle$'s.

The energy-dependent nonlocal pseudopotential is defined to be

$$V_p=V-\sum_C (E_C-E)|C\rangle\langle C|, \tag{4.10}$$

so that

$$(H_0+V_p)|\varphi\rangle=E|\varphi\rangle \tag{4.11}$$

with no restrictions on $|\varphi\rangle$ imposed by the Pauli exclusion principle.

This procedure is implied in the band-structure calculations, which do not yield the true BF's but the "pseudo" BF's which must yet be orthogonalized to the core wave functions.

One can ignore this fact and continue working with the "psuedo" BF's if an appropriate pseudopotential representing the difference between the nitrogen impurity and the host phosphorous can be computed.

Consider an isolated ion core: The potential seen by an external electron consists of the electrostatic attraction of the nucleus and the repulsion of all the core electrons:

$$V(\mathbf{x})=-\frac{Ze^2}{|x|}+\sum_C e^2\int d^3y\frac{|\varphi_C(\mathbf{y})|^2}{|\mathbf{x}-\mathbf{y}|}. \tag{4.12}$$

The pseudopotential acting on the wave function would be

$$\begin{aligned} V_p\varphi(\mathbf{x})&=V(\mathbf{x})\varphi(\mathbf{x})\\ &\quad-\sum_C (E_C-E)\varphi_C(\mathbf{x})\int d^3y\,\varphi_C{}^*(\mathbf{y})\varphi(\mathbf{y}). \end{aligned} \tag{4.13}$$

If $\varphi(\mathbf{x})$ is nonzero and smooth in the neighborhood of $\mathbf{x}=\mathbf{0}$, much more slowly varying than any of the φ_C's, we can write the pseudopotential as a local

[11] P. Löwdin, J. Chem. Phys. **19**, 1396 (1951).
[12] David Brust, Phys. Rev. **134**, A1337 (1964).

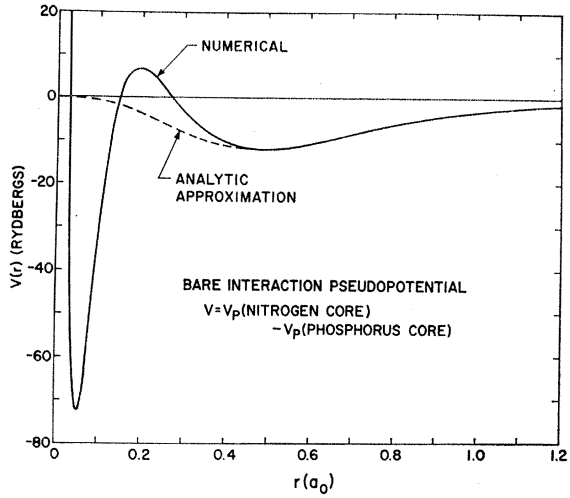

FIG. 6. Pseudopotential for nitrogen substituting for phosphorus from the core wave functions of Herman and Skillman, Ref. 10.

potential by observing

$$\int d^3y\ \varphi_C^*(\mathbf{y})\varphi(\mathbf{y}) \approx \varphi(\mathbf{0}) \int d^3y\ \varphi_C^*(\mathbf{y}) \quad (4.14)$$

and

$$\varphi_C(\mathbf{x})\varphi(\mathbf{x}) \approx \varphi_C(\mathbf{x})\varphi(\mathbf{0}). \quad (4.15)$$

Then

$$V_p\varphi(\mathbf{x}) \approx \Big[V(\mathbf{x}) - \sum_C (E_C - E) \times \varphi_C(\mathbf{x}) \int d^3y\ \varphi_C^*(\mathbf{y}) \Big] \varphi(\mathbf{x}). \quad (4.16)$$

This local pseudopotential can be calculated for phosphorous and for nitrogen using the core wave functions calculated by Herman and Skillman.[10] Because the core states are rather deep, $|E_C| \gg |E|$, and V_p is not sensitive to E, which can be taken to be an average of the atomic energy levels for the outer S and P states.

The bare potential for the nitrogen impurity in GaP is the difference between the nitrogen and the phosphorous core pseudopotentials.

Figure 6 shows this potential calculated numerically from Eq. (4.16) together with the analytic approximation actually used in the computer calculations of the bound-state problem. The approximation ignores the innermost structure of the pseudopotential because that region does not contain sufficient volume to affect the wave functions appreciably. The greatest strength of the potential comes from the broader well farther from the origin.

Because the first conduction band of GaP is predominantly S-like in character, the approximation of Eqs. (4.14) and (4.15) should not be too bad.

C. Bound-State Problem

The technique for finding bound states is simple in theory but more complicated in practice.

We write Schrödinger's equation

$$(E - H_0)|\psi\rangle = V|\psi\rangle, \quad (4.17)$$

where H_0 is the crystal Hamiltonian and V is the pseudopotential presented in the previous section.

For an energy which lies outside the spectrum of H_0, we can write

$$(1 - GV)|\psi\rangle = 0, \quad G = (E - H_0)^{-1}. \quad (4.18)$$

Because we have a short-range potential, the appropriate set of basis functions to use for Eq. (4.18) is the set of WF's. Expressing $(1-GV)$ as a matrix in this basis, the condition that a bound state exist is that the determinant of $(1-GV)$ vanishes for some $E = E_b$, the bound-state energy. The matrix considered is finite because $(n\mathbf{R}|V|n'\mathbf{R}')$, the matrix element of V with WF's, becomes negligible for sufficiently large $\mathbf{R}$ or $\mathbf{R}'$.

The problem one faces now is how to generate WF's from BF's. The relevant formula is straightforward enough:

$$w_n(\mathbf{x} - \mathbf{R}) = \frac{\Omega^{1/2}}{(2\pi)^{3/2}} \int d^3k\ \theta(\mathbf{k}) e^{-i\mathbf{k}\cdot\mathbf{R}} \psi_n(\mathbf{k},\mathbf{x}). \quad (4.19)$$

However, numerical calculations of $\psi_n(\mathbf{k},\mathbf{x})$ do not specify the phase relative to BF's with different values of $\mathbf{k}$. An even greater problem is a useful definition of an energy band. Does one require, e.g., band No. 4 to lie lower in energy than band No. 5 everywhere in the BZ or does one allow them to be inverted in certain regions?

Callaway and Hughes have investigated these problems.[13] They treated the case of the neutral vacancy in silicon. Since silicon contains a center of inversion, they were able to work with real matrices in the generation of their BF's, and so the choice of phase reduced to a choice of sign. Their method of defining energy bands is based on considerations of continuity of symmetry through the BZ rather than strict continuity of energy. Their method yields WF's that are well localized and that possess the symmetries of the one-dimensional representations of the point group.

GaP has no center of inversion, and the numerical BF's have complex components. Also, the energy bands do not cross so neatly in GaP as in silicon. Whether for these reasons or because of a lack of fortitude on the part of the investigator, simpler phase and band assignments were chosen for this problem. The BF's were taken to be real and positive at $\mathbf{x} = \mathbf{0}$ and band assignments were made straightforwardly according to increasing energy. The WF's obtained in this way were not as well localized as those of Callaway and Hughes, so more lattice sites were needed for the matrix $1-GV$. A total of 19 was used as opposed to Callaway's 10. To keep the numerical labor within reason, only the

[13] Joseph Callaway and A. James Hughes, Phys. Rev. **156**, 860 (1967).

first two conduction bands were considered. This resulted in 38×38 matrices for the single-nitrogen problem and as large as 76×76 for the double-nitrogen problem.

The matrix $(1-GV)$ in the WF basis is

$$(n\mathbf{R}|(1-GV)|n'\mathbf{R}')=\delta_{nn'}\delta_{\mathbf{R}\mathbf{R}'}-\sum_{n''\mathbf{R}''}(n\mathbf{R}|G|n''\mathbf{R}'')(n''\mathbf{R}''|V|n'\mathbf{R}'). \quad (4.20)$$

To find the G matrix, we expand the WF's in BF's:

$$(n\mathbf{R}|G|n'\mathbf{R}')=\frac{\Omega}{(2\pi)^3}\int d^3k\,\theta(\mathbf{k})\int d^3k'\theta(\mathbf{k}')\times e^{i\mathbf{k}\cdot\mathbf{R}_e}\delta^{i\mathbf{k}'\cdot\mathbf{R}'}(n\mathbf{k}|(E-H_0)^{-1}|n'\mathbf{k}'). \quad (4.21)$$

But

$$(n\mathbf{k}|(E-H_0)^{-1}|n'\mathbf{k}')=[E-\epsilon_n(\mathbf{k})]^{-1}\delta_{nn'}\delta(\mathbf{k}-\mathbf{k}'), \quad (4.22)$$

therefore

$$(n\mathbf{R}|G|n'\mathbf{R}')=\delta_{nn'}\frac{\Omega}{(2\pi)^3}\int d^3k\,\theta(\mathbf{k})\frac{e^{i\mathbf{k}\cdot(\mathbf{R}-\mathbf{R}')}}{[E-\epsilon_n(\mathbf{k})]}. \quad (4.23)$$

We recognize this, of course, as being an old friend: the function $f(E,\mathbf{R}-\mathbf{R}')$ of Sec. III.

The calculation of the Green-function (GF) matrix G requires only knowledge of the energy-band structure and is relatively straightforward to calculate numerically once the $\epsilon_n(\mathbf{k})$ are known for a dense enough set of points in the BZ.

To find the potential matrix elements $(n\mathbf{R}|V|n'\mathbf{R}')$ is a bit more difficult, because the WF's are not known directly, but through an integral formula involving the numerically known BF's:

$$(n\mathbf{R}|V|n'\mathbf{R}')=\frac{\Omega}{(2\pi)^3}\int d^3k\,\theta(\mathbf{k})\int d^3k'\times\theta(\mathbf{k}')e^{i(\mathbf{k}\cdot\mathbf{R}-\mathbf{k}'\cdot\mathbf{R}')}(n\mathbf{k}|V|n'\mathbf{k}'). \quad (4.24)$$

But the BF's themselves are known only through the coefficients of the basis set of plane waves: $e^{i(\mathbf{G}+\mathbf{k})\cdot\mathbf{x}}$,

$$(n\mathbf{k}|V|n'\mathbf{k}')=\frac{1}{(2\pi)^3}\sum_{\mathbf{G}}\sum_{\mathbf{G}'}a_n^*(\mathbf{k},\mathbf{G})a_{n'}(\mathbf{k}',\mathbf{G}')\times\int d^3x\,e^{-i(\mathbf{G}+\mathbf{k}-\mathbf{G}'-\mathbf{k}')\cdot\mathbf{x}}V(\mathbf{x}). \quad (4.25)$$

This formula illustrates why an analytic approximation for $V(\mathbf{x})$ was used. If a numerical evaluation of the Fourier transform of V had been done, along with all the other computer operations necessary to calculate $(n\mathbf{R}|V|n'\mathbf{R}')$, the cost would have been prohibitive.

For the potential

$$V(\mathbf{r})=-J(r/r_0)^4e^{-4r/r_0}, \quad (4.26)$$

we have the Fourier transform

$$V(\mathbf{k})=\int d^3r\,e^{-i\mathbf{k}\cdot\mathbf{r}}V(\mathbf{r})=-J\frac{15\pi}{8}\left(\frac{r_0}{4}\right)^5\frac{\{6-20(kr_0/4)^2+6(kr_0/4)^4\}}{[1+(kr_0/4)^2]^6} \quad (4.27)$$

and the BF matrix element

$$(n\mathbf{k}|V|n'\mathbf{k}')=\frac{1}{(2\pi)^3}\sum_{\mathbf{G}}\sum_{\mathbf{G}'}a_n^*(\mathbf{k},\mathbf{G})\times V(\mathbf{G}+\mathbf{k}-\mathbf{G}'-\mathbf{k}')a_{n'}(\mathbf{k}',\mathbf{G}'). \quad (4.28)$$

The numerical evaluation of the BF's yields approximately 20 of the $a_n(\mathbf{k},\mathbf{G})$'s for each function.

Returning to the WF matrix elements, use of the point group of GaP, T_d, simplifies Eq. (4.24) somewhat. Denoting by $\{\alpha\}$ one of the 24 rotations of T_d,

$$\{\alpha\}\psi_n(\mathbf{k},\mathbf{x})\equiv\psi_n(\mathbf{k},\alpha^{-1}\cdot\mathbf{x}). \quad (4.29)$$

With the phase convention adopted here ($\psi_n(\mathbf{k},\mathbf{0})$ = real and positive), the BF's are symmetrical for all rotations (eigenvalue = 1) and, for time inversion,

$$\{\alpha\}|n\mathbf{k})=|n\alpha\mathbf{k}),\qquad \psi_n(-\mathbf{k},\mathbf{x})=\psi_n^*(\mathbf{k},\mathbf{x}). \quad (4.30)$$

The origin of coordinates has been taken at the impurity site, so that $V(\mathbf{x})$ is invariant under all rotations:

$$\{\alpha\}V\{\alpha^{-1}\}=V. \quad (4.31)$$

Therefore,

$$(n\mathbf{k}|V|n'\mathbf{k}')=(n\alpha\mathbf{k}|V|n'\alpha\mathbf{k}') \quad (4.32)$$

for all $\{\alpha\}$ in T_d, and

$$(n\mathbf{R}|V|n'\mathbf{R}')=\frac{\Omega}{(2\pi)^3}\int_{[\mathrm{BZ}]}d^3k\int_{[1/48\text{ of BZ}]}d^3k'\times 2\,\mathrm{Re}\{\Sigma(\mathbf{k},\mathbf{R};\mathbf{k}',\mathbf{R}')(n\mathbf{k}|V|n'\mathbf{k}')\}, \quad (4.33)$$

where

$$\Sigma(\mathbf{k},\mathbf{R};\mathbf{k}',\mathbf{R}')=\sum_{\alpha}e^{i(\alpha\mathbf{k}\cdot\mathbf{R}-\alpha\mathbf{k}'\cdot\mathbf{R}')}. \quad (4.34)$$

The complex function $\Sigma(\mathbf{k},\mathbf{R};\mathbf{k}',\mathbf{R}')$ can be expressed in closed form.

This use of symmetry reduces the labor of computation by a factor of 48 but there still remains a double integral to perform, once over a basic 1/48 of the BZ and once over the entire zone.

We therefore need $(n\mathbf{k}|V|n'\mathbf{k}')$ for $\mathbf{k}'$ in a set of points within a basic 1/48 of the zone and for $\mathbf{k}$ on 48 times that number of points. The number of points actually used in the basic 1/48 of the zone was 16. This means the evaluation of $(n\mathbf{k}|V|n'\mathbf{k}')$ a total of $3\times48\times16\times16=36\,864$ times. The factor of 3 arises because two bands are involved. Although 16 points do

not thoroughly cover the basic volume of the zone, they should be just adequate for the present work.

The evaluation of 36 864 matrix elements required approximately one-half hour of high-speed computer time. Having these numbers in storage, the evaluation of four matrix elements $(n\mathbf{R}|V|n'\mathbf{R}')$ for n and $n'=1,2$ (first and second conduction bands) required from 10–30 sec of computer time, depending on the values of $\mathbf{R}$ and $\mathbf{R}'$.

Symmetry can be used to reduce the number of WF matrix elements which must be computed. From Eq. (4.35) it is seen that the matrix elements are real:

$$(n\mathbf{R}|V|n'\mathbf{R}')=(n'\mathbf{R}'|V|n\mathbf{R}). \tag{4.35}$$

Also, using Eq. (4.30) for the transformation properties of the BF,

$$\begin{aligned}\{\alpha\}|n\mathbf{R})&=\frac{\Omega}{(2\pi)^3}\int d^3k\,\theta(\mathbf{k})e^{-i\mathbf{k}\cdot\mathbf{R}}|n\alpha\mathbf{k})\\&=\frac{\Omega}{(2\pi)^3}\int d^3k\,\theta(\mathbf{k})e^{-i\mathbf{k}\cdot\alpha\mathbf{R}}|n\mathbf{k})\\&=|n\alpha\mathbf{R}).\end{aligned} \tag{4.36}$$

Therefore,

$$\begin{aligned}(n\mathbf{R}|V|n'\mathbf{R}')&=(n\mathbf{R}|\{\alpha\}^{-1}V\{\alpha\}|n'\mathbf{R}')\\&=(n\alpha\mathbf{R}|V|n'\alpha\mathbf{R}').\end{aligned} \tag{4.37}$$

Using values of $\mathbf{R}$ out to the second neighbor [$\mathbf{R}\sim(0,0,0),(1,1,0),(2,0,0)$, where $\sim$ indicates the inclusion of all symmetry-equivalent sites], the total number of pairs $(\mathbf{R},\mathbf{R}')$ which must be calculated reduces from $19\times19=361$ to 17. Table I lists these matrix elements.

TABLE I. Wannier-function potential matrix elements for the first two conduction bands of GaP: $(n\mathbf{R}|V|n'\mathbf{R}')$ (eV).

R	R′	$(n,n')=$ (1,1)	(1,2)	(2,1)	(2,2)
(000)	(000)	−2.26051	−1.11251	−1.11251	−0.54756
(110)	(000)	−0.18857	−0.09280	−0.06120	−0.03013
(200)	(000)	+0.06466	+0.03182	−0.04950	−0.02437
(110)	(110)	−0.02314	−0.01065	−0.01065	−0.00865
(110)	($\bar{1}\bar{1}$0)	−0.01592	−0.00304	−0.00304	+0.00321
(110)	(1$\bar{1}$0)	−0.01535	−0.00498	−0.00498	−0.00136
(110)	(101)	−0.01594	−0.00811	−0.00811	−0.00693
(110)	(10$\bar{1}$)	−0.01741	−0.00508	−0.00508	−0.00191
(110)	($\bar{1}$01)	−0.01308	−0.00330	−0.00231	+0.00149
(110)	($\bar{1}$0$\bar{1}$)	−0.01308	−0.00231	−0.00330	+0.00149
(200)	(200)	−0.00200	+0.00141	+0.00141	−0.00138
(200)	($\bar{2}$00)	−0.00179	+0.00145	+0.00145	−0.00080
(200)	(020)	−0.00183	+0.00141	+0.00141	−0.00108
(110)	(200)	+0.00591	−0.00436	+0.00210	−0.00207
(110)	($\bar{2}$00)	+0.00514	−0.00398	+0.00153	−0.00065
(110)	(002)	+0.00487	−0.00475	+0.00164	−0.00174
(110)	(00$\bar{2}$)	+0.00538	−0.00334	+0.00161	−0.00087

D. Single-Impurity Bound State

Using the matrix elements of Table I, the $(2\times19)\times(2\times19)$ matrix $(n\mathbf{R}|V|n'\mathbf{R}')$ can be set up. Given a value of E, the GF matrix $(n\mathbf{R}|G(E)|n'\mathbf{R}')$ can be calculated fairly quickly.

The energy eigenvalue of the bound state can be found by plotting $\det[1-G(E)V]$ as a function of E. The point where $\det[1-G(E)V]=0$ is the bound-state energy E_b.

The entire process of calculating $G(E)$, setting up $(1-GV)$, and taking the determinant required less than 5 sec of computer time, most of which was spent calculating the determinant.

It was found that the potential used gave a bound state well into the forbidden gap, approximately 1 eV. A reduction factor λ was used to reduce the strength of the potential to fit the experimentally observed single-nitrogen level in the hope that the same parameter would give reasonable values for the double-nitrogen levels. The energy of the single-nitrogen state was found to be extremely sensitive to the value of this parameter, giving -0.0133 eV at $\lambda=0.504$ and -0.0068 eV at $\lambda=0.50$. A similar situation existed in the one-band–one-site approximation in its sensitivity to the potential strength J. The calculations of Callaway and Hughes[13] exhibit a similar sensitivity. Also in the computer calculations, excited states of the single-nitrogen problem never appeared. This is a property of the short range of the potential chosen for the calculation. A longer-range potential would reduce the intervalley matrix elements responsible for splitting off the excited states, and would bring them below the band edge. This, coupled with the sensitivity of the energy level to the potential strength, suggests that the self-consistency effects, which would be of longer range than the present bare potential, are of prime importance in this problem.

A multiplying factor of 0.501 puts the single-nitrogen level at approximately -8 meV.

E. Double-Impurity Bound States

For the bound states arising from a pair of nitrogen atoms, one located at the origin and the other located at the lattice site $\mathbf{R}_0$, we can write the potential

$$V_{II}(\mathbf{x})=V(\mathbf{x})+V(\mathbf{x}-\mathbf{R}_0) \tag{4.38}$$

and the WF matrix elements of this potential

$$\begin{aligned}(n\mathbf{R}|V_{II}|n'\mathbf{R}')&=(n\mathbf{R}|V|n'\mathbf{R}')\\&\quad+(n\mathbf{R}-\mathbf{R}_0|V|n'\mathbf{R}'-\mathbf{R}_0).\end{aligned} \tag{4.39}$$

The size of the GV_{II} matrix is controlled by the conditions that $\mathbf{R}$ and $\mathbf{R}'$ must belong to either or both of the sets of vectors

$$\begin{aligned}S_1&=\{(000),\sim(110),\sim(200)\},\\S_2&=\{\mathbf{R}|\mathbf{R}-\mathbf{R}_0\in S_1\}.\end{aligned}$$

There are 19 vectors in each set, and the sets overlap

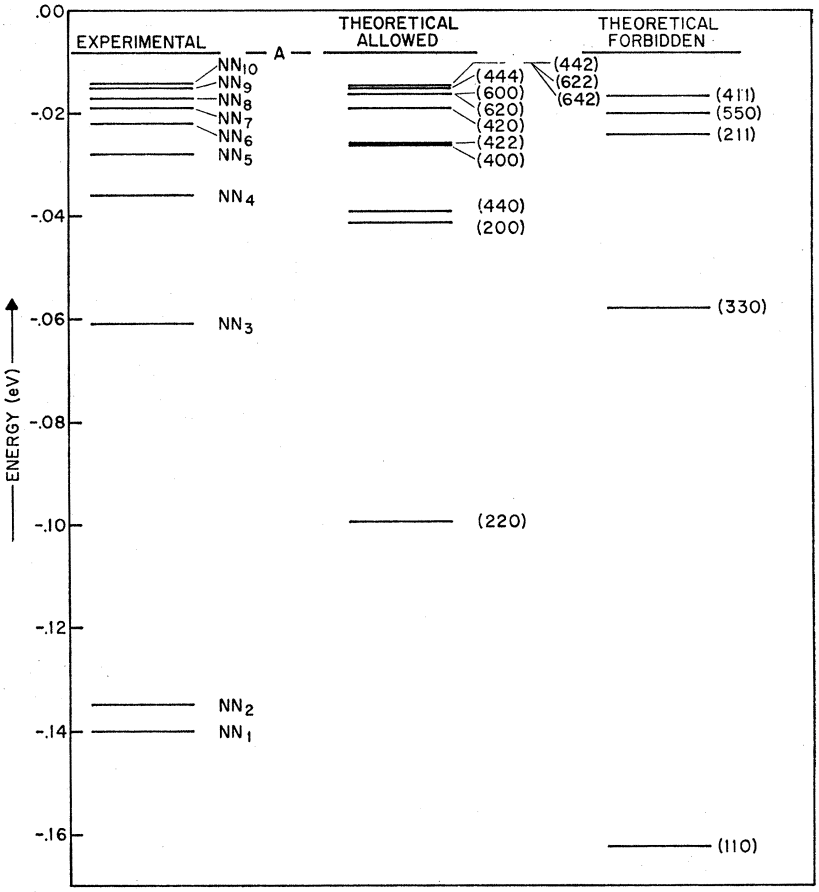

FIG. 7. Energy levels of the double nitrogen states as calculated from the pseudopotential of Fig. 6 reduced by a factor of 0.501. Only the theoretical electron binding energies are presented here and the experimental levels have been raised by 13 meV. (Experimental levels after Thomas and Hopfield, Ref. 1.)

for small $\mathbf{R}_0$. Thus, the GV_{II} matrix will be a size ranging from $(2\times19)\times(2\times19)$ to $(2\times38)\times(2\times38)$, depending on the extent of the overlap of S_1 and S_2.

In any case, $(n\mathbf{R}|GV_{\text{II}}|n'\mathbf{R}')$ can be calculated almost as easily as $(n\mathbf{R}|GV|n'\mathbf{R}')$. In fact, several values or $\mathbf{R}_0$ can be considered at the same time if they are close enough together because many of the matrix elements of G are the same for different but close values of $\mathbf{R}_0$.

Again, the evaluation of the $\det(1-\lambda GV_{\text{II}})$ averaged about 5 sec.

Using a multiplying factor of 0.501, which puts the single-nitrogen level at -0.0083 eV, the double-nitrogen levels come out as shown in Fig. 7. They are arranged into two groups, allowed and forbidden, according to whether $\psi(\mathbf{k}=\mathbf{0})\neq0$ or $\psi(\mathbf{k}=\mathbf{0})=0$, respectively. The leftmost column contains the experimental levels for comparison. The numbers in parentheses represent the lattice vector separating the two nitrogen impurities.

Although these energy levels obviously do not agree with experiment, the range of energy is correct and the average spacing of the levels is about right.

Of special note is the fact that the theoretical levels appear in almost a random ordering according to the magnitude of $\mathbf{R}$, the separation of the two nitrogens.

This great shuffling of the energy levels is, of course, due to the wave function interference effects discussed before in the case of the one-band–one-site approximation. The calculated energy levels are obviously very sensitive to the orientation of $\mathbf{R}$.

V. PHENOMENOLOGY: EFFECTIVE DEPTH AND RANGE

One certain conclusion can be drawn from the results of the calculations presented in the preceding sections: The bare, very-short-range potential of one atom substituting isoelectronically for another in a semiconductor is not adequate to describe all aspects of the problem.

It is clear that the effective potential that one should use must include lattice relaxation and electronic polarization of the host crystal. These effects are of longer range than the bare potential and their inclusion would decrease the sensitivity of the binding energy to the potential strength and reduce the intervalley matrix elements responsible for splitting off the excited states and for the orientational dependence of the double-impurity bound states. On the other hand, there must remain a sufficiently strong short-range part of the potential to give strong optical absorption as observed in GaP:N.

This section will be concerned with an attempt to

find a characteristic potential depth and range which will reproduce the experimental binding energies observed in GaP:N without worrying about the ultimate origin of such a potential.

A. Procedure

One could choose, say, a Gaussian potential with adjustable range and run through the calculation described in Sec. IV for several values of the range. Such a procedure would waste a great deal of computer time.

In lieu of recalculating the WF matrix elements for each value of the range, it was decided to make a truncation of the problem and simply assign these matrix elements reasonable values and proceed from there with the rest of the calculation of Sec. IV.

For the determinant of $(1-GV)$, the matrix elements of V were taken to be zero for interband coupling, thereby reducing the problem to the consideration of only the lowest conduction band. Within this band, the matrix elements of V were taken to be

$$(\mathbf{R}|V|\mathbf{R}')=\delta_{\mathbf{R},\mathbf{R}'}V(R),$$
$$V(R)=-Je^{-(R^2/A^2)},$$

where J is the adjustable depth and A is the adjustable range.

The matrix elements of G were calculated exactly, using the band structure of Cohen and Bergstresser[8] as described in Sec. IV.

This procedure is obviously bad for several reasons: (1) It ignores interband coupling; (2) it ignores off-diagonal elements of $(\mathbf{R}|V|\mathbf{R}')$; and (3) it possesses inversion symmetry, while GaP does not have inversion symmetry.

However, it has advantages: (1) It can be done in a reasonable amount of time; (2) it becomes exact in the limit of very-long-range potentials.

B. Results

With two parameters to adjust, two experimental binding energies can be fitted exactly or several experimental binding energies can be fitted approximately. The former procedure was adopted here. The A line and the valley-orbit excited states (A^* line) were fitted using a depth and range

$$J=0.64 \text{ eV},$$
$$A=1.15a,$$

where a (cubic lattice constant of GaP) $=5.44$ Å.

With these parameters, the binding energy of an electron in the ground state (Γ_1) is 8 meV and the binding energy in the doubly-degenerate first excited state (Γ_{12}) is 6.5 meV. The splitting between Γ_1 and Γ_{12} is a fairly sensitive function of the range, being essentially zero at $A=1.5a$ and being greater than 10 meV at $A=1.0a$. The binding energy of the A line is much less sensitive to the potential strength. With the range as chosen, $J=0.6$ eV gives a binding energy of 2.5 meV and $J=0.7$ eV gives a binding energy of 19 meV.

Having chosen a depth and a range, the double-impurity bound states can be calculated. These are shown in Fig. 8 along with the experimental levels for comparison.

In this sequence, as opposed to the sequence of Fig. 7, the levels appear in more nearly the "normal" sequence, i.e., monotonically diminishing binding energy with increasing separation. However, there is still enough intervalley coupling to cause the splitting shown between the (330) state and the (411) state, which have the same magnitude of separation but different orientations. Also, as contrasted to Fig. 7, the levels do not converge to the A line nearly so fast as the experimental levels.

It appears that the double-impurity levels are quite dependent on the shape of the potential and cannot be calculated simply knowing a depth and a range.

VI. CONCLUSIONS

Short-range interactions in solids have been a traditional source of frustration. In some cases, namely, "central cell corrections" to donor and acceptor binding energies, their effects are often small compared to the more tractable aspects of the problem. For isoelectronic

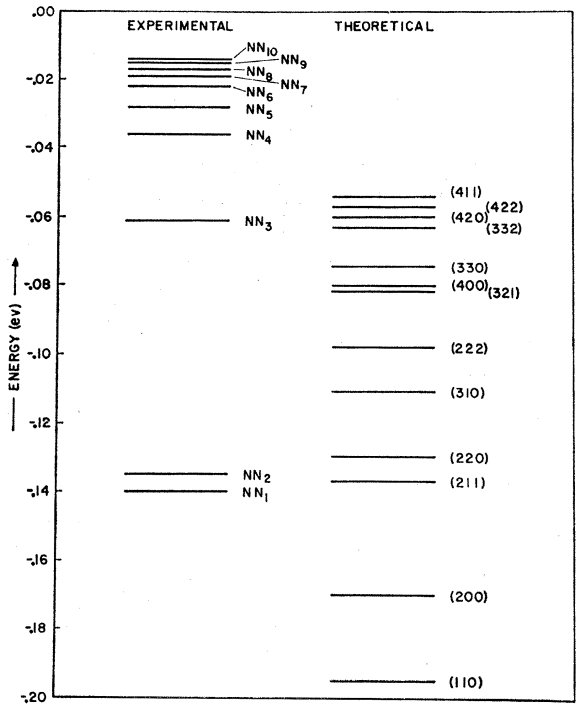

FIG. 8. Energy levels of the double nitrogen states calculated semiempirically after first fitting the single nitrogen state (A line) and its excited states from intervalley mixing (A^* line). Only the theoretical electron binding energies are presented here and the experimental levels have been raised by 13 meV. (Experimental levels after Thomas and Hopfield, Ref. 1.)

impurities, however, the short-range interaction is the dominant, if not the entire, problem.

Much has been explained here about isoelectronic traps using the Slater-Koster one-band–one-site approximation. In regard to GaP:N, this approximation correlates the position of the principal bound state (A line) with its oscillator strength in optical absorption and with the absorption strength for frequencies above the band edge using essentially only one adjustable parameter.

In both the one-band–one-site approximation and the more elaborate computer calculations using a very-short-range potential, the binding energy was found to be extraordinarily sensitive to the potential strength and, contrary to experiment, the excited states due to intervalley mixing were found to be unbound. Taking a phenomenological potential of intermediate range both reduced the binding energy sensitivity substantially and brought the excited states below the band edge. This indicates that a large measure of lattice relaxation and electronic polarization of the host crystal must be included in the short-range potential problem in future calculations. The extra potential from such effects might be exponentially decaying or it might go as $1/r^n$ for large distances. However, the central core cannot be neglected for any such long-range potential for $n>1$. Only the Coulomb problem of all the singular potential problems of this form becomes insensitive to additional structure in a small central region when the size of that region is allowed to become very small.

Callaway and Hughes[13] have shown that practical calculations involving WF's and short-range potentials are possible. The present work, which uses many of their methods, indicates that the next major step must be a first-principles calculation of the response of the host crystal to the short-range perturbation. In the case of GaP:N, such a calculation could be tested by its ability to predict the double-nitrogen levels. Electron-hole correlation effects, also ignored here, could prove quite important to a thoroughly successful theory.

A better understanding of isoelectronic traps would lead to a better understanding of the related problem of central cell corrections to donor and acceptor states. It would also open the way for calculations of more complicated systems such as nearest-neighbor donor-acceptor pairs. With vision, one can imagine a whole chemistry of complexes within the vacuum represented by the perfect crystal.

Note added in proof. Those who compare Fig. 7 of this article with Fig. 6 in the previous paper by Faulkner and Hopfield[14] will note that they are quite different. In the previous calculation, the shift of origin which was made in order to compute BF's with the origin of coordinates at the impurity site was inadvertently taken in the wrong direction. Thus, that calculation was actually a calculation of an impurity on a gallium site instead of on a phosphorus site in GaP.

ACKNOWLEDGMENT

The author wishes to thank Professor John J. Hopfield for originally suggesting the topic of this work and for valuable advice and counsel during its preparation.

[14] R. A. Faulkner and J. J. Hopfield, in *Localized Excitations in Solids*, edited by R. F. Wallis (Plenum Press, Inc., New York, 1968), p. 218.