Interacting quasi-band model for electronic states in alloy semiconductors: Relation to average t-matrix approximation and band anticrossing model

Yuzo Shinozuka
Department of Materials Science and Chemistry, Faculty of Systems Engineering, Wakayama University, Wakayama 640-8510, Japan
E-mail: yuzo@sys.wakayama-u.ac.jp
Received May 30, 2014; accepted June 10, 2014; published online June 27, 2014

A new variational theory is proposed for electronic states in alloy semiconductors with arbitrary diagonal and off-diagonal randomness and any concentration. In A_{x}B_{1-x} substitutional alloys, the theory allows the mutual interaction between two quasi-A and -B electronic bands whose effective band widths are proportional to c_{A} and c_{B}, respectively, i.e., an interacting quasi-band model. The model provides satisfactory results near-band-edge states, especially for band bowing. For diagonal randomness, the theory corresponds to the average t-matrix approximation, and in the dilute limit, a formula similar to the band anticrossing model, which has been frequently applied to GaN-related alloys, is obtained.

© 2014 The Japan Society of Applied Physics

Alloy semiconductors are useful materials, and desirable physical properties can be obtained by changing the constituent concentrations. Of these materials, GaN-related alloys have attracted considerable attention in connection with efficient LEDs and laser diodes (LDs). Theoretical research on electronic states in random system has a long history going back to the 1960s. The coherent potential approximation (CPA) successfully explains the overall electronic properties of A_{1-x}B, alloy systems that have diagonal \(^{6} \) and off-diagonal \(^{7} \) randomness. If the parameters of A and B are not markedly different, then the electronic properties of the alloy change continuously in response to a change in c; this is known as amalgamation-type behavior. When the differences are large, the alloy properties are based on components originating from each constituent, even though the system is statistically homogeneous; this is persistence-type behavior. However, the application of CPA is limited to particular situations, and progress in its extension is rather slow.\(^{8,9} \) First-principles calculations using density functional theory are another effective method for alloys. By considering quasi-periodic supercells, each containing several hundred atoms, and recovering the Bloch theorem, the electronic properties can be explained to some extent, but progress is limited by the size of available computer memory. Thus, such calculations can be expected to establish qualitative and semi-quantitative theories for electronic states in alloy semiconductors. In 1999, Shan et al.\(^{10,11} \) proposed the band anticrossing model (BAM) to address electronic states in GaInNAs. This phenomenological theory takes into account mutual interactions between the band states and a localized state. The BAM successfully explains band bowing and the concentration dependence of the electronic energies of various III–V alloys at dilute concentration limits.\(^{12,13} \) However, the underlying physics of the model is not yet clear, and attempts to further extend the model to finite concentrations have proven to be difficult.

In the present paper, we propose another approach to random alloy systems: a variational theory that can be applied to various alloys having arbitrary diagonal and off-diagonal randomness at any strength and concentration. In a particular limit, this new theory corresponds to the average t-matrix approximation (ATA).\(^{4,6,8} \) and in another limit, a form similar to the BAM is produced. Here, we discuss the general properties, validity, and shortcomings of this new model.

Let us consider a substitutional binary random alloy A_{x}B_{1-x} (c_{B} = 1 - c_{A}) of which the electronic Hamiltonian is given by

\[ H = \sum_{n} |n\rangle \varepsilon_{n} \langle n| + \sum_{n} \sum_{m} |n\rangle t_{nm} \langle m|. \] (1)

Here, \(|n\rangle\) denotes a Wannier state localized at the site \(n\) located at \(R_{n}\) in a regular lattice. Each site is randomly occupied by either an A or B atom with a probability of \(c_{A}\) or \(c_{B}\), respectively. The site energy, \(\varepsilon_{n}\), takes a value of either \(\varepsilon_{A}\) or \(\varepsilon_{B}\) depending on the type of atom occupying the site, and the difference \(\Delta = \varepsilon_{A} - \varepsilon_{B}\) denotes the strength of the diagonal randomness. The transfer energy \(t_{nm}\) between sites \(n\) and \(m\) depends on the relative positions of \(R_{n}\) and \(R_{m}\) as well as their atomic occupations:

\[
t_{nm} = \begin{cases} 
\varepsilon_{A} & n = A \text{ site, } m = A \text{ site} \\
\varepsilon_{AB} & n = A \text{ site, } m = B \text{ site} \\
\varepsilon_{BA} & n = B \text{ site, } m = A \text{ site} \\
\varepsilon_{B} & n = B \text{ site, } m = B \text{ site}
\end{cases}
\] (2)

In pure limits (\(c_{A} = 1\) and \(c_{B} = 0\), i.e., pure crystal A or \(c_{A} = 0\) and \(c_{B} = 1\), i.e., pure crystal B), the eigenstates are Bloch-type states that are characterized by the wave vector \(k\),

\[ |\psi_{k}\rangle = \frac{1}{\sqrt{N}} \sum_{n} \exp(ikR_{n}) |n\rangle, \] (3)

where \(N\) is the total number of sites. The corresponding eigenenergies \(E_{A}\) and \(E_{B}\) in crystals A and B, respectively, are given by

\[ E_{A}(k) = \varepsilon_{A} + \varepsilon_{A}^{k}, \] (4a)

\[ E_{B}(k) = \varepsilon_{B} + \varepsilon_{B}^{k}, \] (4b)

In a similar manner, we can write \(t_{AB}^{k} = (1/N) \sum_{n} \sum_{m} t_{nm} \exp[-ik(R_{n} - R_{m})]\) and likewise for \(t_{BA}^{k}\). However, note that these two quantities do not correspond to any real situations.

Before presenting the new method, it is instructive to start with a simple cubic lattice in which the transfer energy is only nonzero (\(-V\)) between nearest neighbors, irrespective of
The virtual crystal [Eq. (5)] is given by

$$E(k) = \tilde{E} = -2V(\cos k_1 + \cos k_2).$$

At the bottom of the band, the energy $E(k = 0) = \tilde{E} - 6V$ linearly changes with $c_A$ or $c_B$ (= $1 - c_A$), and the corresponding wave function is $|\psi_\text{la}(k)| = (1/\sqrt{N}) \sum_n |n|$, where an electron has equal amplitudes extending over all sites. The effects of the randomness, $H' = \sum_n |n\rangle (\delta_e - \delta_i) |n\rangle$, i.e., the deviation from the average potential, can be taken into account using a second-order perturbation. The correction to the virtual crystal [Eq. (5)] is given by

$$\Delta E = c_A c_B (\varepsilon_e - \varepsilon_B)^3 \frac{1}{N} \sum_k \frac{1}{E(k) - E'(k)}.$$  

For the bottom of the three-dimensional band, $\Delta E$ is always negative finite. By using the lattice Green’s function in the simple cubic system, we find that the resulting band bowing is $\Delta E \sim -0.247 c_A c_B (\varepsilon_e - \varepsilon_B)^3/V$. Since the wavefunctions in the virtual crystal are uniformly extended, the band bowing suggests a nonuniform amplitude distribution in the alloy, i.e., if $\varepsilon_e > \varepsilon_B$, then the lowest state favors the lower-potential B sites. Consider the trial wavefunction $|\tilde{\psi}_{k=0}\rangle = (1/\sqrt{N}) \sum_n |n\rangle$, in which the variational parameter $\beta_n$ can take one of two values, $\beta_A$ or $\beta_B$, depending on the atomic occupation at $n$. The expectation value of the Hamiltonian $\langle \tilde{\psi}_{k=0}|H|\tilde{\psi}_{k=0}\rangle$ should be minimized under the condition $\langle \tilde{\psi}_{k=0}|\tilde{\psi}_{k=0}\rangle = c_A |\beta_A|^2 + c_B |\beta_B|^2 = 1$. After a statistical average over random occupations of A and B atoms and introducing the Lagrange multiplier, $E$, we find that the quantity

$$\langle \tilde{\psi}_{k=0}|H - E|\tilde{\psi}_{k=0}\rangle = (c_A - E) c_A \beta_A^2 + (\varepsilon_B - E) c_B \beta_B^2 - 6V(c_A \beta_A^2 + c_B \beta_B^2),$$.  

should be minimized with respect to $\beta_A$ and $\beta_B$. Thus, we have

$$\begin{pmatrix} \varepsilon_A - 6c_A V - E & -6c_B V \\ -6c_A V & \varepsilon_B - 6c_B V - E \end{pmatrix} \begin{pmatrix} \beta_A \\ \beta_B \end{pmatrix} = 0,$$

which has two solutions for nonzero $\beta_A$ and $\beta_B$ values only when

$$E_{\pm} = \frac{1}{2} \left[ (\varepsilon_A + \varepsilon_B - 6V \pm \sqrt{(\varepsilon_A - \varepsilon_B)^2 + 12(c_A - c_B)(\varepsilon_A - \varepsilon_B)V + 36V^2}]^{1/2} \right].$$

Figure 1(a) shows the concentration dependence of the lower branch $E_-$. The deviation from the virtual crystal (straight dotted line) shows asymmetrical bowing with respect to the interchange $c_A \leftrightarrow c_B$, which is contrary to the second-order perturbation given in Eq. (6). Figure 1(b) shows the probabilities $|\beta_A|^2$ and $|\beta_B|^2$, and as expected, $|\beta_A|^2 < 1$ and $|\beta_B|^2 > 1$ when $\Delta = \varepsilon_A - \varepsilon_B > 0$. Figure 1(c) shows the quantity $P_B = c_B |\beta_B|^2$, which is the total participation on all B sites ($c_B N$ in total). The large deviation from $P_B = c_B$ (the virtual crystal) indicates that an electron has amplitudes at lower-potential B sites that are larger than those at A sites even when no spatial localization occurs and the wavefunction extends over the alloy. When $c_B$ is very small, a localized state at an isolated B impurity splits off from the bottom of the host A band ($\varepsilon_e - 6V$) when $\Delta/6V > 0.5$ (the correct condition is shown to be $\Delta/6V > 0.34$ based on the Koster–Slater method).
Fig. 2. Examples of calculated energy bands obtained from Eq. (13) for typical cases: (a)-(c) diagonal randomness, (d)-(g) off-diagonal randomness, and (h)-(j) diagonal + off-diagonal randomness. Here, $T^i = \max(t^i_j) = -\min(t^i_j)$ for $i, j = A, B$. The dark red colored regions have a second solution $(dE_x/dt_0 < 0)$. Note that the parameters in (i) and (j) satisfy the condition $\alpha^{AB} = \beta^{AB} = \sqrt{\alpha^{AA} \alpha^{BB}}$, which was studied in Ref. 7.

The present theory has the following properties:

1) It provides a useful formula that is symmetric in atoms A and B and can be applied to arbitrary types of diagonal and off-diagonal randomness at any strength and concentration.

2) It is a good approximation for near-band-edge states, i.e., near the bottom of the conduction band and near the top of the valence band. For weak disorder, Eq. (13) gives the correct behavior corresponding to a virtual crystal. Depending on the parameters, positive or negative band bowing appears (see the discussion below). However, the theory is less reliable in the middle of a band because it may give a spurious gap, even in weak disorder cases.

3) As can be seen in Eq. (13), $E_{\pm}(k)$ depends on $k$ through the values of $t^{AA}_{ik}, t^{BB}_{ik}, t^{AB}_{ik}$, and $t^{BA}_{ik}$. Hence, if all the maxima or minima of the $t^i_k$'s ($i, j = A, B$) are given at the same $k$-point in the Brillouin zone, then the resulting allowed electronic bands have a distribution over the common colored area in Fig. 2, irrespective of the form of $t^i_k$.

4) Except for cases (d) and (j) in Fig. 2, the obtained $E_{\pm}(k)$ is located within the band area of pure crystal A and pure crystal B. Thus, the Saxon–Hutner theorem \cite{Saxon} which has been proven for diagonal randomness, \cite{Hutner} holds.

5) The trial wave function, Eq. (10), has two solutions $+$/$-$ for each $k$, irrespective of the concentration; hence, the total number of states appears to be doubled $2N$. This inconsistency will be resolved in the discussion below.

Let us discuss these properties in detail, starting with band bowing. Various types of band bowing, i.e., deviations from a virtual crystal, can be obtained depending on the character and strength of the randomness. In a purely diagonal random case ($t^{AA}_{ik} = t^{BB}_{ik} = t^{AB}_{ik}$), it can be easily shown that $E_{\pm}(k)$ is a monotonically increasing or decreasing function of the concentration. Anomalous large bowing that violates the Saxon–Hutner theorem, such as $\min E_-(k) < \min[E_+(k), E_B(k)]$ or $\max E_+(k) > \max[E_-(k), E_B(k)]$, is caused by off-diagonal randomness. In a purely off-diagonal case ($\varepsilon_A = \varepsilon_B$), a necessary condition for the former is $|\min t^{AB}_{ik}| > |\min t^{AA}_{ik}|, |\min t^{BB}_{ik}|$ (a similar inequality exists for the latter). The anomalous large band-gap bowing found in several nitride alloys \cite{Shinozuka} can be explained if these conditions are fulfilled in at least one of the conduction and valence bands of the alloys.

Now, let us consider the relationship between the new model and the ATA. We concentrate on a case in which all $t^i_k$'s ($i, j = A, B$) have the same dispersion, i.e., proportional to $t_k$: $t^i_k = T^i t_k$. Figure 3(a) shows the typical behavior of $E_{\pm}(k)$ obtained from Eq. (13). We can attempt to calculate the density of states per site using the normal procedure as $\rho_\pm(E) = N^{-1} \int \delta(E - E_{\pm}(k)) \rho_0(t_0(E_{\pm})) dE_{\pm}/dt_k)^{-1}$, where $\rho_0(E) = N^{-1} \int \delta(E - t_k), and$ $t_k(E_{\pm})$ is the inverse function of Eq. (13). In this way, we always have $\int \rho_-(E) dE = \int \rho_+(E) dE = 1$ irrespective of the concentration, as mentioned above. On the other hand, it would be plausible to consider that the states $|\psi_k\rangle$ within an interval $dt_k$ are modified to form $|\psi_{k+}\rangle$ and $|\psi_{k-}\rangle$ within an interval $dE_+$ and $dE_-$, respectively. Since $dE_+/dt_k + dE_-/dt_k = (c_A t^{AA} + c_B t^{BB})$ from Eq. (13), the degrees of freedom should be respectively reduced by the factors $(c_A t^{AA} + c_B t^{BB})^{-1} dE_+/dt_k$ and $(c_A t^{AA} + c_B t^{BB})^{-1} dE_+/-dt_k$. Thus, the renormalized density of states will be given by

$$\tilde{\rho}_\pm(E) = \frac{1}{c_A t^{AA} + c_B t^{BB}} \rho_0(t(E_{\pm})).$$

(14)

In this case, $\int \tilde{\rho}_-(E) + \tilde{\rho}_+(E) dE = 1$ holds, and we recover the correct number of states. However, the sum rule $\int \tilde{\rho}_-(E) dE = c_0$ and $\int \tilde{\rho}_+(E) dE = c_A$ in case of $\Delta = \varepsilon_A - \varepsilon_B > 0$ only holds in the split band limit $\Delta \gg |t^i_k|$. A
self-consistent theory such as CPA is necessary to obtain the sum rule in persistence-type alloys. In pure crystals, for which $\varepsilon_A = \varepsilon_B$ and $t_{kA}^A = t_{k}^A = t_{k}^{AB} = \epsilon A = \epsilon B = 1$, Eq. (13) has $N$ normal solutions ($\beta_A = \beta_B = 1$, $E_A(k) = \varepsilon + T_k$) and $N$ flat solutions ($\beta_A = \sqrt{c_A/c_B}$, $\beta_B = -\sqrt{c_B/c_A}$, $E_A(k) = \varepsilon$). Since $d E_A/dk = 0$ in the latter branch, these improper solutions provide null contributions to the density of states. Thus, the reduction factor $dE_A/dk$ (or $dE/k$) is important especially for the top (bottom) of $E_A(k)$ ($E_B(k)$). Figures 3(b) and 3(c) show the calculated energy density of states using a semi-elliptic form $\rho_0(t_k) = (2/\pi)\sqrt{1 - t_k^2}$ for $-1 \leq t_k \leq 1$. The figures show that the renormalized density of states (solid line) is exactly the same as that in ATA (refer to Fig. 3 in Ref. 4).

Finally, we consider the dilute and isolated limits. In the dilute limit in which one of the constituent concentrations, say $c_A$, is small, B atoms are isolated in the host crystal A. We can set $c_A \sim 1$ and $c_B \ll 1$. Then, assuming that $t_{kA}^{AB} = t_{k}^A = \text{const.} = V^{AB}$ and neglecting the concentration dependence of the diagonal terms in Eq. (12), we obtain an equation similar to Eq. (1) in Ref. 10, which was proposed in regard to the BAM. Note that one of the off-diagonal terms in Eq. (12) is proportional to $E_A$ and the other is not, but in the BAM both are proportional to $\sqrt{E_A}$. The BAM has been used to discuss the behavior of electronic states in dilute GaN-related alloys, and large band bowing is considered to be generated by anticrossing between band states and a localized state. In contrast, the present theory implies that it would be more plausible to explain large band bowing as a result of mixing host band states and the impurity band states. This picture is consistent with a recent first-principles calculation of GaAs$_{1-x}$N$_x$ alloys.

In summary, a new variational theory has been proposed to explain the electronic states in alloy semiconductors having arbitrary randomness at any strength for any lattice structure and concentration. The theory captures the overall concentration dependence, while positive or negative band bowing depends on the parameters. The theory provides good approximations for near-band-edge states, i.e., it is applicable to those electronic states near the top of the valence band and the bottom of the conduction band. In the present paper, we have assumed $A_BA_B$ binary alloys with simple unit cell structures, but extensions to compound semiconductors such as ternary In$_x$Ga$_{1-x}$N would not be difficult if we consider $A = \text{In}$ and $B = \text{Ga}$. The present theory is now being studied through applications to various alloys of compound semiconductors by using their real energy dispersions.

Acknowledgments The present study was partially supported by a JSPS KAKENHI. The author would like to thank Professor H. Yaguchi of Saitama University for valuable discussions.