Theory of Defects in Semiconductors: Recent Developments and Challenges

by Stefan K. Estreicher

Our present understanding of the properties of defects in semiconductors comes from a combination of microscopic experiment and first-principles theory. Over the past few decades, theory has evolved from a qualitative tool to the point where many properties of defects can be predicted quantitatively. The following perspective outlines some recent developments and remaining challenges.

Native defects and impurities profoundly affect the mechanical, electrical, optical, and magnetic properties of materials, especially semiconductors.[1,2] In silicon for example, O and N are introduced during crystal growth. They pin dislocations introduced during device fabrication and improve the mechanical strength of the material. Shallow dopants such as B (or P) provide holes (or electrons) to the valence (or conduction) band. Intrinsic, or native, defects (vacancies and silicon self-interstitials) are by-products of ion implantation and high-temperature anneals and promote the diffusion of dopants. Extrinsic defects, in particular transition metal impurities, may be present in the source material or introduced inadvertently during IC fabrication processes such as chemomechanical polishing or metallization. Most of these defects introduce deep levels in the forbidden energy gap, or simply gap, and are (unwanted) electron-hole recombination centers, sometimes with large cross sections. Their presence often gives rise to characteristic photoluminescence (PL) bands. Hydrogen is almost unavoidable. It passivates the electrical activity of dopants and of many deep-level defects, but interstitial H is electrically active by itself. H-related local vibrational modes (LVMs) are seen by Fourier-transform infrared (FTIR) absorption or Raman spectroscopy. Magnetic impurities such as Mn can render a semiconductor ferromagnetic. The list goes on. Many defects are found in more than one configuration and/or charge state, and their properties may change with the position of the Fermi level.

For the production of devices, all these defects must be controlled. This implies a detailed understanding of the stable and metastable configurations, solubility and precipitation sites, formation and binding energies, diffusion path and barrier, interactions with dopants, native defects, impurities, and electrical (optical and/or magnetic) activity. Some of that information may be obtained from experimental techniques such as FTIR, Raman, PL, or deep-level transient spectroscopy (DLTS). But experiments have limitations. The specific state of the defect under study must be present in sufficiently high concentrations to be detectable. Then, the defect must have an IR-active LVM to be seen by FTIR, levels in the appropriate region of the gap for DLTS analysis, etc.

First-principles theory has very different limitations. It has evolved into an essential tool in the identification and characterization of defects in semiconductors.[2] Much of the development of theory has been made possible by the vast amount of experimental information available about defects in silicon.[3] This summary contains an overview of today's most utilized methodology, a discussion of some recent developments, and comments about present and future challenges.

Defect Properties at $T = 0$ K

Periodic supercells approximate the host crystal. These are large unit cells (64 to a few hundred host atoms) repeated in all directions of space. Since the defect is also repeated periodically, the supercell must be as large as possible in order to minimize unrealistic defect-defect interactions. Indeed, one impurity in a Si_{64} cell corresponds to a concentration of about 1.5 atomic percent! At the same time, the supercell must also be as small as possible because the computational expense grows substantially with size. Compromises vary with the author and the defect under study. Figure 1 shows the substitutional-interstitial $\{Cu_s,Cu_i\}$ pair in the Si_{64} cell.

The electronic states are calculated in a single cell using first-principles density-functional theory within the local-density or generalized-gradient approximations.[4] "First-principles" means that none of the input parameters are fitted to an experimental database. The atomic core regions are replaced by norm-conserving, angular-momentum dependent, *ab initio* pseudopotentials.[5] The choice of a particular single-particle basis set for the valence states dictates the type of pseudopotential to be used. Typical basis sets consist of plane waves[6] or, for real-space methods, carefully selected atomic-like orbitals[7,8] or Gaussian-shaped wavefunctions.[9]

Classical molecular-dynamics simulations[6-10] allow geometry optimizations (conjugate gradients), simulated quenching or annealing, or constant-temperature runs to study defect diffusion or reactions. This 'temperature' is determined by the nuclear kinetic energy: the electrons remain in the ground state and the free energy is not included. The time step used when solving Newton's laws

(continued on page 30)

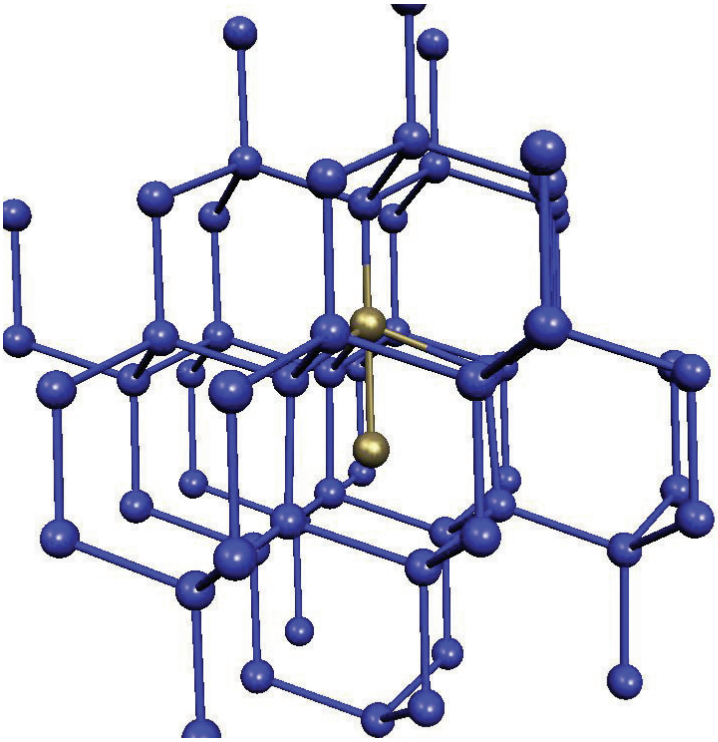

FIG. 1. Substitutional-interstitial copper pair (gold circles) in the Si_{64} supercell.

Estreicher

(continued from page 28)

for nuclear motion is of the order of a femtosecond.

Geometries (in stable and metastable states), binding, formation, migration, or reorientation energies, spin densities, and other properties of a wide range of defects have been calculated with this approach. One can obtain impurity-related LVMs (*e.g.* a Si-H stretch mode) by calculating the total energy for adjacent positions of a given atom (*e.g.* H along the Si-H bond direction). The desired frequency is extracted from the curvature of the energy. This works, but provides only one specific frequency. A better approach is to calculate and then diagonalize the entire dynamical matrix by calculating all the force constants or using linear response theory.[11] This is computer-intensive but the dynamical matrix provides the entire vibrational spectrum.[12,13]

So far, the defect properties that can be calculated quantitatively do not include the precise position of defect-related levels in the gap. Being able to predict accurately these levels would provide a much-needed connection between first-principles theory and electrical data (DLTS). Two approaches are used to predict the existence and approximate position of levels in the gap.

One method involves scaling the ionization energy of an unknown defect to that of a defect with a known donor level, usually in the upper half of the gap. This can also be done for acceptor levels using electron affinities.[14] The accuracy of the prediction depends on the availability of a suitable marker—a defect similar to the one under consideration and for which DLTS data exist.

Another method involves plotting the formation energies of the various charge states of a defect as a function of the electron chemical potential,[15] *i.e.*, as a function of the Fermi level. The intersections of the lines show the location of the gap levels. However, since the supercell is periodic, the energies of charged defects include an unrealistic Madelung energy term, which should be removed. The correction can be as large as 0.3 eV in a Si$_{64}$ supercell.[16,17] Precisely how to evaluate this Madelung correction for delocalized charge distributions is unclear. Another difficulty is that the gap calculated with a local exchange potential is about half the size of the measured one. Simply rescaling the calculated values by stretching the gap ignores the fact that defect levels may be pinned to a specific band. One solution is to use a non-local exchange potential,[18] but this comes at a high computational cost since this calculation scales as N^4, where N is the basis set size.

Finite Temperatures

The energetics of defects calculated as described above are strictly valid only at T = 0 K. Further, the total zero-point energy is ignored (although that of high-frequency LVMs is easy to include). Extending first-principles theory to finite temperatures has only been done in a few instances.

One approach is to do thermodynamic integration.[19,20] The method requires extensive Monte Carlo or molecular-dynamics simulations at various fixed temperatures and the knowledge of the free energy of a reference system.

An alternative approach is to evaluate the dynamical matrix (calculated at T = 0 K) at many q points in the Brillouin zone of the supercell. The result is a phonon density of state, which can be used to calculate of the Helmholtz vibrational free energy in the harmonic approximation. Comparisons of the measured and calculated phonon densities of state, specific heats, and vibrational entropies show that this constant-volume, harmonic approach works very well up to at least 800 K for semiconductors such as C, Si, Ge, or GaN.[21] At low temperatures, the approximations are excellent, as demonstrated by accurate predictions of subtle isotope effects in the specific heat of elemental semiconductors.[22]

This method can be extended to defects in supercells.[21,23] The perturbed phonon density of states is constructed from the dynamical matrix and used to calculate the Helmholtz vibrational free energy. Other contributions, such as the configurational entropy, are added analytically. The thermodynamic effects can be large. For example, the dissociation energy of the substitutional-interstitial copper pair in Si decreases by almost 0.3 eV between 0 and 400 K.[21] Since these energies often appear in exponential Boltzmann factors, such corrections are substantial. Figure 2 shows the calculated binding energy of the {Cu$_s$,Cu$_i$} pair in the Si$_{64}$ cell as a function of temperature. The binding enthalpy 0.84 ± 0.09 eV measured in the 333-417 K range is the dissociation enthalpy (1.02 ± 0.09 [24]) minus the migration energy (0.18 ± 0.02 [25]). It matches the calculated binding free energy at 400 K.

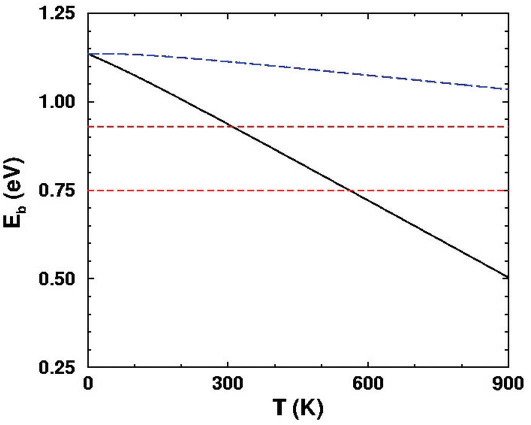

FIG. 2. *Calculated binding free energy of the {Cu$_s$,Cu$_i$} pair in Si compared to the binding enthalpy 0.84 ± 0.09 eV measured in the 333-417K range. The dashed curve shows the contribution of the vibrational free energy and the solid curve also contains the configurational entropy term.*

Conclusion

First-principles theoretical studies of defects and impurities in semiconductors are capable of predicting many measurable quantities: geometrical configurations in the stable and metastable states, electronic structures, formation, binding and activation energies, charge and spin densities, and local and pseudo-local vibrational modes. The most commonly used method employs (local) density-functional theory and molecular-dynamics simulations in periodic supercells. It can be extended to finite temperatures provided that the dynamical matrix of the supercell is known. The methodology is applicable to a wide range of materials with reasonably large gaps, which includes defects in the bulk of insulators or semiconductors, molecules, surfaces, and nanostructures. The calculation of reliable and accurate gap levels is one example of the remaining challenges for theorists. Much more difficult situations involve both short- and long-range interactions, as for magnetic impurities in semiconductors. ■

References

1. H. J. Queisser and E. E. Haller, *Science*, **281**, 945 (1998).
2. S. K. Estreicher, *Materials Today*, **6** (6), 26 (2003).
3. J. Chelikowsky, *MRS Bull.*, **27**, 951 (2002).
4. W. Kohn, *Rev. Mod. Phys.*, **71**, 1253 (1999).
5. B. Fornberg and D. M. Sloan, *Acta Numerica*, **94**, 203 (1994).
6. R. Car and M. Parrinello, *Phys. Rev. Lett.*, **55**, 2471 (1985).
7. O. Sankey and D. J. Niklewski, *Phys. Rev. B*, **40**, 3979 (1989).
8. D. Sánchez-Portal, O. Ordejón, E. Artacho, and J. M. Soler, *Int. J. Quant. Chem.*, **65**, 453 (1997).
9. R. Jones and P. R. Briddon, in *Semiconductors and Semimetals,* Vol. 51A, M. Stavola, Editor, p. 287, Academic, New York (1998).
10. S. K. Estreicher and P. A. Fedders, in *Computational Studies of New Materials*, D. A. Jelski and T. F. George, Editors, World Scientific, Singapore (1999).
11. S. Baroni, S. de Gironcoli, A. Dal Corso, and P. Gioannozzi, *Rev. Mod. Phys.*, **73**, 515 (2001).
12. J. M. Pruneda, S. K. Estreicher, J. Junquera, J. Ferrer, and P. Ordejón, *Phys. Rev. B*, **65**, 075210 (2002).

13. S. K. Estreicher, D. West, J. Goss, S. Knack, and J. Weber, *Phys. Rev. Lett.*, **90**, 035504 (2003).
14. A. Resende, R. Jones, S. Öberg, and P. R. Briddon, *Phys. Rev. Lett.*, **82**, 2111 (1999).
15. S. B. Zhang and J. E. Northrup, *Phys. Rev. Lett.*, **67**, 2339 (1991).
16. P. A. Schultz, *Phys. Rev. B*, **60**, 1551 (1999); *Phys. Rev. Lett.*, **85**, 1942 (2000).
17. J. Lento, J.-L. Mozos, and R. M. Nieminen, *J. Phys: Condens. Matter*, **14**, 2637 (2002).
18. J. Lento and R. M. Nieminen, *J. Phys: Condens. Matter*, **15**, 4387 (2003).
19. E. Smargiassi and R. Car, *Phys. Rev. B*, **53**, 9760 (1996).
20. M. Kaczmarski, R. Rurali, and E. Hernández, *Phys. Rev. B*, **69**, 214105 (2004).
21. S. K. Estreicher, M. Sanati, D. West, and F. Ruymgaart, *Phys. Rev. B*, **70**, 125209 (2004).
22. M. Sanati, S. K. Estreicher, and M. Cardona, *Solid State Commun.*, **131**, 229 (2004).
23. O. K. Al-Mushadani and R. J. Needs, *Phys. Rev. B*, **68**, 235205 (2003).
24. A. A. Istratov *et al.*, *Appl. Phys. Lett.*, **72**, 474 (1998).
25. A. A. Istratov *et al.*, *Phys. Rev. Lett.*, **81**, 1243 (1998).

About the Author

Stefan K. Estreicher is Paul Whitfield Horn Professor of Physics at Texas Tech University and a Fellow of the American Physical Society. He is author or co-author of some 130 papers, many in the area of defects in semiconductors and a few on the history of wine. He can be reached at Stefan.Estreicher@ttu.edu.