Using Density Functional Theory to Engineer Direct Gap Germanium-Tin Alloy

Chris Darmody, D. P. Ettisserry, Neil Goldsman
Department of Electrical and Computer Engineering
University of Maryland
College Park, MD 20742

Nibir K. Dhar
Night Vision and Electronic Sensors Directorate
CERDEC
Fort Belvoir, VA

***Abstract*— Germanium is a group IV semiconductor commonly used in Short Wave Infrared (SWIR) optical devices due to its relatively small band gap of 0.66eV. Like silicon in the period above it, the conduction band minimum of germanium does not lie at the same point in k space as the valence band maximum, making it an indirect-gap material and thus reducing its absorption efficiency. Unlike silicon however, the direct-gap of germanium is only slightly larger than its indirect-gap energy, giving it the possibility of possibly transitioning to a direct-gap material with clever band structure engineering. One such method showing promise involves alloying germanium with tin in various ratios. Using density functional theory (DFT), we can calculate the effects the alloy has on the band structure for different percentages of tin and thus predict the percentage needed to transition germanium into a direct-gap material.**

Keywords—germanium; tin; GeSn; direct; indirect; gap; DFT

I. INTRODUCTION

As with most semiconductors and their alloys, the valence band maximum of germanium is located at the gamma point in k space. The conduction band minimum for germanium is located at the L point for the conventional fcc lattice structure making its gap nature indirect. Fortunately, the indirect-gap energy is only 140meV less than the direct-gap energy [1]. This relatively small difference has been the motivation for band structure research and engineering to achieve a direct-gap form of the material for use in optical devices. The most promising techniques applied to achieve such a structure include applying strain and alloying with various other elements [2]. The obvious choice of alloy material has been tin due to its location in the period immediately below germanium. The alpha allotrope variant of tin has the same diamond crystal structure as germanium but acts like a semimetal with a negative band gap at the gamma point [2]. An elementary application of the simple linear form of Vegard's law gives the indication that the transition from direct to indirect-gap should occur at approximately a 21% uniform tin concentration [3]. Experimental and calculated results both show the presence of a bowing parameter which is needed to fit the non-linear experimental data for how the band gaps change with varying tin concentration [4]. In general, it is known that with increasing tin concentration, both the direct and indirect-gaps of $Ge_{1-x}Sn_x$ shrink but the direct-gap does so at a faster rate. If the exact concentration of tin needed to cause this transition can be determined, direct-gap devices can be fabricated while still maintaining as much of the gap as possible.

Band structure calculations have predominately been performed using single electron empirical pseudopotential methods with extrapolation to fit data to GeSn alloys [5]. Some calculations have also been applied to specific compositional percentages of GeSn using Density-Functional Theory (DFT). The theory of DFT is used to solve the many body Schrodinger equation for a desired atomic system, while accounting for electron-electron interactions and the exclusion principle [2]. Figure 1 shows the cells on which we have used to perform calculations, containing 12.5, 6.26, and 3.125% tin respectively.

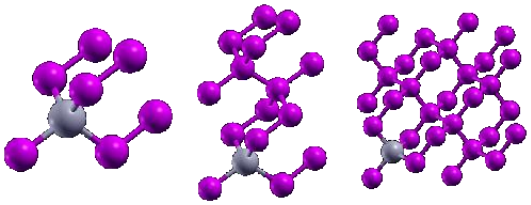

Fig. 1. Atomic supercells of $Ge_{1-x}Sn_x$ (12.5%, 6.25%, 3.125%) constructed using repeated 8 atom cubic unit cells.

II. DENSITY FUNCTIONAL THEORY

To obtain the band structure information for the system we would ideally need to solve the multi-body Schrodinger wave equation, with the many-body wave function ψ.

$$\left(\sum_i -\frac{\hbar^2}{2m}\nabla_{r_i}^2 - \sum_i\sum_j \frac{Z_j}{r_i - r_j} + \frac{1}{2}\sum_i\sum_{i'}\frac{q^2}{r_i - r_{i'}}\right)\psi = E\psi \quad (1)$$

This equation can be replaced by solving the self-consistent Kohn-Sham equation for a periodic solid, decreasing the difficulty of the problem from an N-body problem to N, single body problems coupled by an effective potential.

$$\left(-\frac{\hbar^2}{2m}\nabla^2 + v_{eff}(r)\right)\phi_i(r) = \varepsilon_i\phi_i(r) \quad (2)$$

$$\rho(r) = \sum_i^N |\phi_i(r)|^2 \quad (3)$$

This equation represents an effective system of non-interacting particles each with Kohn-Sham orbital ϕ_i and energy ε_i, moving in an effective potential (v_{eff}), which generate the same density ρ(r) as the actual system of interacting particles. In the DFT framework, the total energy of the system is written as a functional of the charge density.

978-1-4673-7860-4/15/$31.00 ©2015 IEEE

$$E[\rho] = T[\rho] + \int V_{ext}(r)\rho(r)dr + E_H[\rho] + E_{XC}[\rho] \quad (4)$$

Here, T[ρ] is the kinetic energy, V_{ext} is the external potential (electron-nuclei interaction), E_H is the Hartree energy due to Coulombic interaction , and E_{XC} is the exchange-correlation energy which accounts for electron-electron correlated motion and the exclusion principle. To solve the system in a self-consistent manner, orbitals are expanded with a finite plane wave basis set and the nuclear core potentials are replaced by slowly varying pseudopotentials which reduce the number of plane waves needed to reconstruct them. The expansion into plane waves allows the equation to be changed from a set of partial differential equations to an algebraic eigenvalue problem that can be solved iteratively [6].

The accuracy of DFT results are highly dependent on the combination of the system being studied, the pseudopotential used, and the functional applied to the calculation [7]. Pseudopotentials act as an effective potential that each valence electron interacts with, and acts to stabilize calculations by treating the high frequency components of the core electron potentials as a smoother approximation within some defined radius. By making this approximation, larger systems of atoms become solvable by reducing the number of plane waves needed in the calculation. Different methods exist for generating pseudopotentials and accuracy is generally dependent on the configuration of the system being studied. After testing numerous pseudopotentials generated using different functionals and valence occupancies, we were able to use a pseudopotential generated using Perdew-Burke-Ernzerhof (PBE) [8] exchange-correlation to obtain an accurate band structure of pure germanium crystal. The calculation was performed using the hybrid PBE0 functional [9] which allowed us to tune the indirect and direct bandgap energies to their experimental values by applying the appropriate mixing fraction of Hartree-Fock exchange energy. Calculations in DFT transform the set of Kohn-Sham equations for non-interacting particles by expanding the potentials and wave functions onto a finite set of N_{pw} plane waves thus turning the set of differential equations into an equivalent N_{pw}-dimensional eigenvalue equation which can be solved iteratively and self-consistently. A kinetic energy cutoff is established to provide an upper limit for the number of plane waves used in the expansion. By adjusting this cutoff energy the number of plane waves in the expansion can be adjusted. The cutoff should be set sufficiently large to ensure good representation of the wave functions but this increase comes at the cost of increased computation time and memory usage. In addition to requiring a sufficient cutoff energy, calculations also should be performed on enough k points to adequately sample the Brillouin zone (BZ). During the calculation, integrals over the entire BZ are needed to calculate the charge density at each iteration. This integration is approximated by taking a weighted sum over a finite set of special points in the irreducible wedge of the BZ with weights corresponding to the number of equivalent points in the entire BZ [6]. As with increasing the cutoff energy, increasing the number of k points increases the computation time. Ideally, we would use the Monkhorst-Pack k point grid as a way to select points in an unbiased manner and perform a relatively computation-heavy self-consistent field (SCF) calculation to obtain an accurate representation of the system potential [10] [11]. This potential can then be used as an input to an inexpensive non-self-consistent field (NSCF) calculation with k points selected along the path of the high symmetry points of the desired band structure plot. However, Quantum Espresso does not allow NSCF calculations using hybrid functional exchange-correlations, so our calculations had to be performed with the desired k point path directly in the SCF calculations. To minimize integration errors, we used a long path length as well as a high k point density to attempt to cover the majority of the irreducible wedge.

While trying to tune the parameters used to obtain the experimental band gaps for the pure germanium crystal cell, we noticed that the indirect-gap energy varied linearly with the mixing fraction as shown in Figure 2. By simply adjusting the mixing fraction linearly, the direct-gap linearly varied as well but at a different rate to the indirect-gap. By adjusting the lattice constant of the cell to values slightly deviated from experimental values, we found that it linearly changed the energy difference between the direct and indirect-gap values as shown in Figure 3. In addition, this change was almost entirely independent from the absolute indirect-gap energy value which facilitated the adjustment of both indirect and direct-gaps.

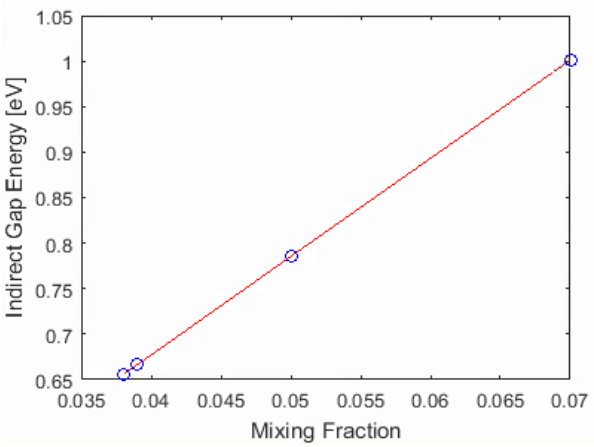

Fig. 2. Indirect-gap energy linearly tunable with hybrid functional mixing fraction

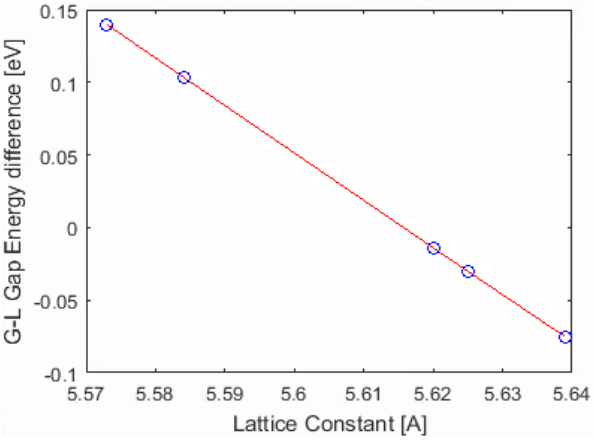

Fig. 3. Gap energy difference tunable by adjusting lattice constant slightly

The final band structure obtained of the pure germanium crystal is shown in Figure 4, with an indirect-gap of 0.66eV and a direct-gap of 0.80eV achieved using a lattice constant of 5.573Å, only differing 1% from the quoted 5.646Å [2]. This calculation was performed on a simple 2 atom cell with a face-centered cubic (FCC) Bravais lattice and the results give confidence in using this pseudopotential and functional for further, larger calculations on germanium. Similar results were obtained with the 8 atom cell.

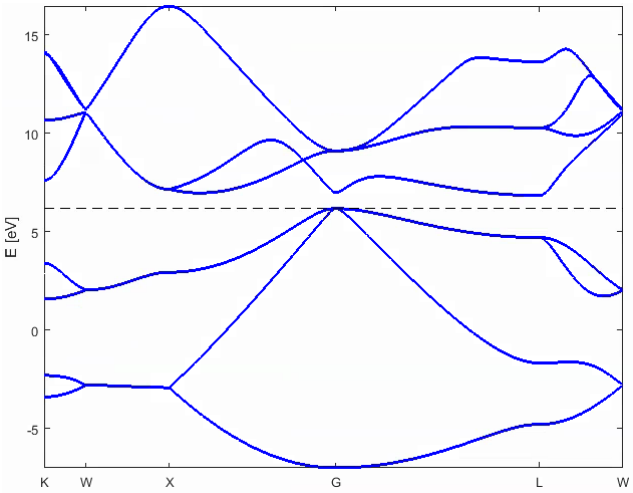

Fig. 4. Pure germanium band structure for 2 atom cell

III. Germanium-Tin Alloy Calculations

A. Supercells

To generate supercells with variable percentage of tin, we change the cell from an FCC Bravais lattice with a 2 atom basis to a simple cubic lattice with an 8 atom basis or a tetragonal lattice with a 16 or 32 atom basis, each with a single tin atom. As with the 2 atom cell, we tune the lattice constant and mixing fraction of an 8 atom cell which can be repeated to generate the other cells. The mixing fraction used to tune the gap values of the pure germanium 8 atom cell is kept constant when tin is added and when the larger cells with tin are used.

Performing computations using larger supercells comes with the cost of significantly increased computation time. The Virtual Crystal Approximation (VCA) is a work-around to this problem, wherein the characteristics from the pseudopotentials of both germanium and tin are merged into a hybrid pseudopotential representing a non-existent atom which approximates a percent concentration of each atom, allowing for small atomic cells but introducing non-physical atoms [3][5]. Because of the potentially limiting transferability inherent to certain kinds of pseudopotentials, care must be taken to ensure that pseudopotentials maintain accuracy. The effects of these approximations are evident in published results on the subject, where the various transition percentage predictions range greatly from 6-21% or more depending on the method used [2]. With access to the High-Performance Computing Cluster Deepthought2 at the University of Maryland, the computationally expensive, large supercell DFT calculations can be run massively parallelized with Quantum ESPRESSO [12] to obtain accurate band structures for many different cell sizes and fractions of tin. Instead of using the VCA approach, we perform ab initio calculations, where the effect of the potential due to tin is localized at a particular lattice site, instead of being averaged over the entire cell.

B. Numerics, Parallelization and Results

Calculation time is system dependent and increases greatly with an increase in the number of k points used, the number of bands present, and the number of plane waves employed. As an approximation, the time to complete a calculation in order or big O notation is given by [6]:

$$T_{CPU} \approx N_{iter} N_k \times (\mathrm{O}(N_b N_{pw}^2) + \mathrm{O}(N_b N_{pw} \log(N_{pw})) + \mathrm{O}(N_b N_{pw}^2)) \quad (5)$$

Where N_{iter} is the number of iterations required to achieve self-consistency, N_k is the number of k points specified, N_b is the number of bands, and N_{pw} is the number of plane waves in the expansion. The number of k points required to achieve convergence generally decreases with an increase in the supercell size. In contrast, the number of bands required will increase with cell size. The number required is given by the number of atoms in the supercell multiplied by the number of valence electrons per atom. The number of plane waves required to achieve convergence will also generally increase with an increasing supercell size. The number of iterations to achieve self-consistency is more difficult to predict but can be assumed to fall within 5 to 20. To complete calculations in a reasonable amount of time due to this highly-nonlinear time scaling, we use the multiple parallelization levels available in the Quantum Espresso PWscf program. Quantum Espresso is set up with different levels of parallelism forming a hierarchical structure. The top level divides the processors into pools, each of which takes care of the calculation at a group of k points. The next level, known as plane wave parallelization, distributes the wave function coefficients across the processors in each pool, offering one of the biggest calculation speedups. Once the speedup for this level saturates, the final level of parallelization can extend the processor scaling. This level divides each group of plane wave processors into task groups, each of which perform the calculation on a group of electronic states. The plane wave groups can also be partitioned into linear algebra groups which parallelize diagonalization and matrix multiplication by distributing across groups of a square number of processors [6].

Through various trials, we have been able to find optimal distributions of processors for each parallelization level and bring calculation times down significantly. For the two atom cell using hybrid functional DFT with 1500 k points and a cutoff energy of 100Ry, the calculation time was able to be reduced from multiple hours in a serial hybrid functional calculation to under 5 minutes in parallel using 144 processors.

Similar to the two atom cell, the 8, 16, and 32 atom cell calculations were parallelized to reduce their computation time while maintaining accuracy with respect to cutoff energy and number of k points. The band structures were obtained from these calculations and the direct and indirect-gap energies were extracted for the various compositional percentages of the GeSn

alloy. Plotting these energies, we extract the tin percentage needed to transition from an indirect to a direct-gap material shown in Figure 5. The transition at 8.5% tin is in agreement within the range predicted by various other methods [2].

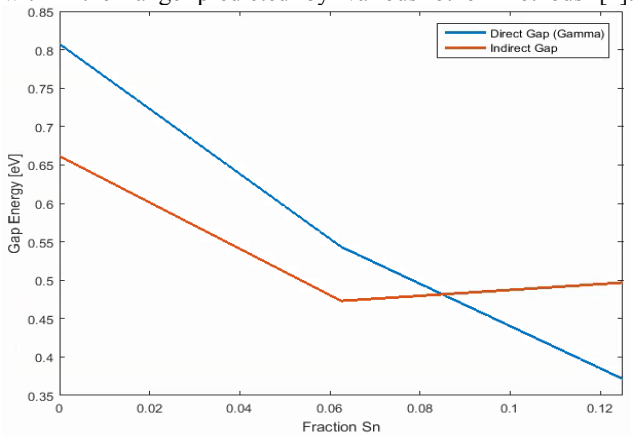

Fig. 5. Indirect to direct-gap transition occurs at approximately 8.5% tin concentration.

IV. Conclusion

In performing these calculations, we have ensured their convergence by increasing the number of k points and the cutoff energy used until the total energy of each system studied varied less than 0.01 Ry. The systems studied were representative of uniformly distributed tin in a GeSn crystal, which we take to approximate a real world uniform distribution. From these calculations, we obtained the band structure for different compositional percentages of tin and have extracted their various direct and indirect-gap energies. By plotting these, we found the crossing point which indicates the transition from an indirect-gap to occur at 8.5% tin. At this tin concentration, and presumably all percentages higher than this, the material has undergone the transition into a direct-gap material. By manufacturing the alloy directly at this transition percent, the band structure is predicted to be direct. In addition, it will also have the added benefit of maintaining as much of the initial gap value as possible. This is due to the trend which indicates that adding higher and higher concentrations of tin will continue to shrink the value of the direct-gap.

Acknowledgment

The authors would like to acknowledge the University of Maryland supercomputing resources made available for conducting the research reported in this paper (http://www.it.umd.edu/hpcc).

References

[1] D.S. Sukhdeo et al., "Direct bandgap germanium-on-silicon inferred from 5.7% ⟨100⟩ uniaxial tensile strain," Photon. Res. 2, A8-A13 (2014)

[2] S. Gupta. "Germanium-Tin (GeSn) Technology," Dissertation, Stanford University, California (2013).

[3] S. Gupta et al., "Achieving direct band gap in germanium through integration of Sn alloying and external strain," J. Appl. Phys. vol. 113, issue 7, no. 073707 (2013).

[4] P. Moontragoon et al., "Band structure calculations of Si-Ge-Sn alloys: achieving direct band gap materials," Semicond. Sci. Technol. 22 742 (2007).

[5] S. Gupta et al., "Band Structure and Ballistic Electron Transport Simulations in GeSn Alloys," SISPAD 2012, 5-7 Sept

[6] A. Corso, "A pseudopotential plane waves program (pwscf) and some case studies," Lecture Notes in Chemisty, Vol. 67, C. Pisani editor, Springer Verlag, Berlin (1996).

[7] P. Haas et al., "Calculation of the lattice constant of solids with semilocal functionals," Phys. Review B 79, 085104 (2009).

[8] J.P. Perdew, K. Burke, M. Ernzerhof, "Generalized Gradient Approximation Made Simple," Phys. Rev. Lett. 77, 3865 (1996).

[9] J.P. Perdew, K. Burke, M. Ernzerhof, "Rationale for mixing exact exchange with density functional approximations," J. Chem. Phys. 105 22 (1996).

[10] M.C. Gibson, "Implimentation and Application of Advanced Density Functionals," Thesis, Department of Physics, University of Durham (2006).

[11] H.J. Monkhorst and J.D. Pack, "Special points for Brillouin-zone integrations," Phys. Rev. B 13, 5188 (1976).

[12] P. Giannozzi et al., "QUANTUM ESPRESSO: a modular and open-source software project for quantum simulations of materials," J. Phys.: Condens. Matter, 21, 395502 (2009).