Prediction of Semiconductor Material Properties by the Properties of Their Constituent Chemical Elements

Kok Liang Heng,[†] Soo Jin Chua,[‡] and Ping Wu*,[†]

Institute of High Performance Computing, 89B Science Park Drive, #01-05/08 The Rutherford, Singapore Science Park I, Singapore 118261, and Department of Electrical Engineering, National University of Singapore, Lower Kent Ridge Road, Singapore 119260

Received October 5, 1999. Revised Manuscript Received March 22, 2000

We demonstrate that it is possible to predict the semiconductor material properties, like lattice constant and band gap energy, from their chemical stoichiometries and fundamental element properties of the constituents. We use a correlation technique to show that band gap energy is strongly correlated to electronegativity and pseudopotential radii whereas the lattice constant is strongly correlated to melting temperature and atomic number. Prediction of band gap energy and lattice constant of some III−V and II−VI semiconductors is made on the basis of the developed correlation model. The technique can also be applied to predict bulk properties of ternary or quaternary semiconductors.

Introduction

In the literature, about 100 unitary, 4000 binary, 8000 ternary, and 1000 quaternary systems are reported, while the maximum number of possible systems is 100 for unitary, 4950 for binary, 161 700 for ternary, and 3 921 225 for quaternary.[1] Only 5% and less than 1% of the possible ternary and quaternary systems are experimentally known. There are millions of other possible candidates that are ignored by the Edisonian approach to materials design. A more scientific and systematic approach based on correlation techniques will be needed for a more effective and efficient design of new materials.

Chen et al. used such a correlation approach, a self-architecting neural network, to model the relation between materials structure and properties for the purpose of predicting new materials.[2] The mechanism of atomic behaviors that control the structure and properties of semiconductors is, however, not revealed by such neural network research. In this study, a correlation technique that has been successfully applied either in the prediction of new additives for galvanizing process[3] or in the material design of new nickel metal hydride electrodes[4] will be used for the binary system of semiconductor alloys based on available literature data. It will provide insightful physics that determines the structure and properties of semiconductors. The technique can also be applied and extended to ternary or quaternary systems.

Experimental Section

Figures 1 and 2 show band gap energy plotted as a function of lattice constant for some common semiconductors.[5] Dashed lines are compounds of indirect band gap materials, and solid lines represent compounds with a direct band gap. Chemical composition of an alloy can be read from these band gap energy−lattice constant plots. For instance, a combination of two binaries such as GaSb and GaAs will give the composition GaAs(i)Sb(1−i) where i is the mole fraction of As. When $i = 0$, the point is at GaSb, and when $i = 1$, it is at GaAs. When i varies from 0 to 1, the point varies continuously along the curve joining the two end points. The exact point on the curve is obtained from the fact that its lattice constant is proportionately contributed in the ratio of the mole fractions of the binary constituents. For simplicity, our binary system only consider the compositions $i = 0$, 0.5, and 1.

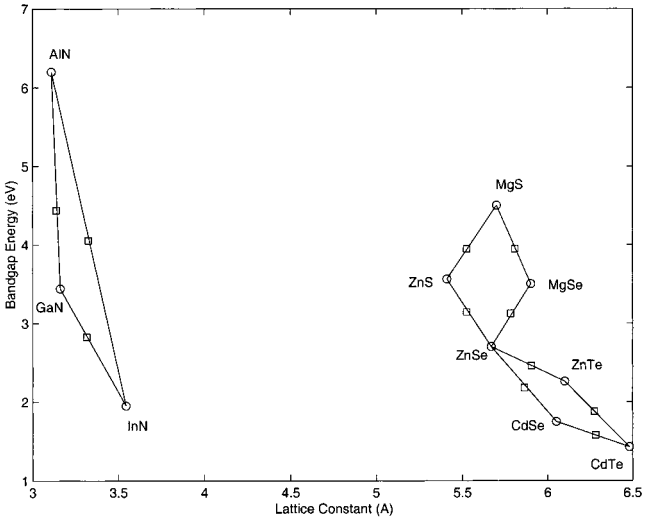

Figure 1. Band gap energy and lattice constant of some important semiconductors.

To correlate materials properties to their constituent element properties, classification of element properties will be discussed first. Chemical element properties are generally

[†] Institute of High Performance Computing.
[‡] National University of Singapore.
* Corresponding author. E-mail wuping@ihpc.nus.edu.sg; telephone (65) 770-9212; Fax (65) 778-0522.

(1) Rodgers, J. R.; Villars, P. *J. Alloys Compd.* **1993**, *197*, 167.
(2) Chen, P. C. L.; Cao, Y.; LeClair, S. R. *J. Alloys Compd.* **1998**, *279*, 30.
(3) Jin, H. M.; Li, Y.; Wu, P. *J. Mater. Res.* **1999**, *14*, 1791.
(4) Wu, P.; Heng, K. L. *Chem. Mater.* **1999**, *11*, 858.
(5) Madelung, O. *Numerical Data and Functional Relationships in Science and Technology. New Series. Group 3: Crystal and Solid State Physics;* Springer-Verlag: Berlin, 1982.

10.1021/cm9906194 CCC: $19.00
Published on Web 05/16/2000

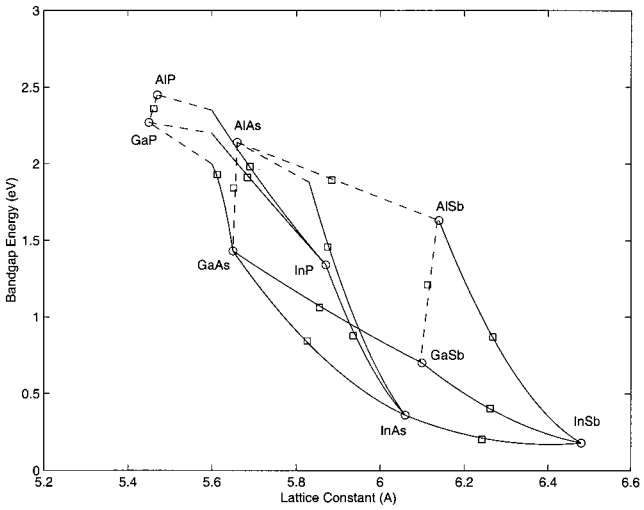

Figure 2. Band gap energy and lattice constant of some important semiconductors.

classified in six groups: size factor, atomic number factor, electrochemical factor, valence-electron factor, cohesive energy factor, and angular valence-orbital factor. The angular valence-orbital factor is not available in the literature for some of the elements and is, therefore, not included in this study. The data listed in Table 1 are used to correlate chemical element properties such as pseudopotential radii, atomic number, melting point, electronegativity, and valence electron from each of the five factors to two semiconductor material properties, lattice constant and band gap energy. This table has a total of 44 records; each record has 14 predictor variables (mole fractions of Al, Cd, Ga, In, Mg, Zn, As, N, P, S, Sb, Se, Te) and 2 response variables (LC = lattice constant, BE = band gap energy). The first 19 records represent i equals 0 or 1 of the binary system, and the remaining 25 records represent i equals 0.5. These 25 records are represented by squares in Figures 1 and 2. We apply a 2 sublattice model[6] for all 44 A_i-B_j semiconductors. The first six variables of Table 1 belong to A lattice (group II or III) whereas the next seven variables belong to B lattice (group V or VI). For each record, the values of the variables of A lattice are summed up to unity and similarly for B lattice. The lattice fraction is defined as

$$Y_{A_i} = \frac{n_{A_i}}{\sum_{i=1}^{6} n_{A_i}} \quad i = 1-6 \tag{1}$$

$$Y_{B_j} = \frac{n_{B_j}}{\sum_{j=1}^{7} n_{B_j}} \quad j = 1-7 \tag{2}$$

where n_{A_i} and n_{B_j} are the number of moles of A_i and B_j in lattices A and B, respectively.

To build a property map, its coordinates, which will be functions of element properties and element concentrations in the alloy, have to be defined. For an A_i-B_j atomic pair, the difference between atom A_i and atom B_j in terms of element property P (i.e., P_{A_i} for A_i and P_{B_j} for B_j) is

$$Q_{A_iB_j} = |P_{A_i} - P_{B_j}| \quad i = 1-6, j = 1-7 \tag{3}$$

Subsequently, in a multicomponent and 2 sublattice alloy, a property map coordinate, O_p (with respect to element property P), was defined as follows:

$$O_p = \sum_{A_i}\sum_{B_j} Y_{A_i} Y_{B_j} Q_{A_iB_j} \tag{4}$$

where $Y_{A_i}Y_{B_j}$ is proportional to the total number of A_i-B_j pairs and O_p is, therefore, the overall element property difference of all nearest-neighbor atomic pairs in the alloy.

Some properties of A_i and B_j elements are listed in Table 2.[7] In the present work, a few element properties such as electronegativity (E, $eV^{1/2}$), atomic number (AN), melting point (T, K), pseudopotential radii (R, au), and valence electron number (V) are selected to define the property map coordinates based on eqs 1−4. Coordinates O_1, O_2, O_3, O_4, and O_5 based on element properties E, AN, T, R, and V, respectively, are generated as shown in Table 3 together with LC and BE. All records are divided into two classes based on the values of LC and BE. Class label C_1 is assigned to be 1 if LC > 5.8 or LC < 3.6 and 2 if 3.6 < LC < 5.8. Similarly, class label C_2 is assigned value 1 if BE > 2.5 and 2 if BE < 2.5.

Results and Discussion

A simple correlation between lattice constant, band gap energy of the binary alloy, and the chemical element properties is established by an in-house artificial intelligence system APEX (Advanced Process Expert).[8] In particular, a set of artificial intelligence techniques[8] is employed by APEX to search among all the possible element properties, the most significant properties that may govern lattice constant, and band gap energy of a binary semiconductor system. In this paper, band gap energy and lattice constant are identified by APEX to be strongly correlated to O_1/O_4 and O_3/O_2, respectively. To view the pattern on a two-dimensional space, shifting is done to the factors O_1/O_4 and O_3/O_2, and the record index is used as an artificial axis. In Figure 3, the x-axis is the shifted ratio O_1/O_4 and the y-axis represents the record index. It is observed that the two classes are well separated. In Figure 4, the record index is plotted against the shifted ratio O_3/O_2. It can be seen from Figure 4 that data with 3.6 < LC < 5.8 are distributed in a bandlike region between two regions of the data with LC > 5.8 or LC < 3.6.

To predict LC and BE from the chemical properties, a neural network model is developed for each of the bulk properties. For this purpose, neural network model builder[8] of APEX is used. Noise is removed from the sample pool before the neural network training. In particular, records with lattice constant less than 4.0 are removed from the original data since the model focuses more on the high lattice constant side. In our case study, a random validation set containing 10% of the samples (four records) is held out to implement early stopping[9] to help prevent overfitting and to improve generalization for both the bulk properties. The correlation of predicted values to actual values for training set and validation set of lattice constant are 0.95 and 0.99, respectively. The aforementioned values for band gap energy are 0.97 and 0.98, respectively. The values of lattice constant and band gap energy from experiments versus those predicted by the neural networks are shown in Figures 5−8. It is observed from these figures

(6) Pelton, A. D. *Calphad J.* **1988**, *12*, 127.

(7) Samsonov, G. V. *Handbook of the Physiochemical Properties of the Elements*; IFI/Plenum: New York, 1968; pp 10−312.

(8) Heng, K. L.; Jin, H. M.; Li, Y.; Wu, P. *J. Mater. Chem.* **1999**, *9*, 837.

(9) Nelson, M. C.; Illingworth, W. T. *A Practical Guide to Neural Nets;* Addison-Wesley: Reading, MA, 1991; p 165.

Table 1. Some Basic Semiconductor Data from Literature[5]

no.	Al	Ga	In	Zn	Cd	Mg	N	P	As	Sb	S	Se	Te	LC (Å)	BE (eV)
1	1	0	0	0	0	0	1	0	0	0	0	0	0	3.11	6.20
2	1	0	0	0	0	0	0	1	0	0	0	0	0	5.47	2.45
3	1	0	0	0	0	0	0	0	1	0	0	0	0	5.66	2.14
4	1	0	0	0	0	0	0	0	0	1	0	0	0	6.14	1.63
5	0	1	0	0	0	0	1	0	0	0	0	0	0	3.16	3.44
6	0	1	0	0	0	0	0	1	0	0	0	0	0	5.45	2.27
7	0	1	0	0	0	0	0	0	1	0	0	0	0	5.65	1.43
8	0	1	0	0	0	0	0	0	0	1	0	0	0	6.10	0.70
9	0	0	1	0	0	0	1	0	0	0	0	0	0	3.54	1.95
10	0	0	1	0	0	0	0	1	0	0	0	0	0	5.87	1.34
11	0	0	1	0	0	0	0	0	1	0	0	0	0	6.06	0.36
12	0	0	1	0	0	0	0	0	0	1	0	0	0	6.48	0.18
13	0	0	0	1	0	0	0	0	0	0	1	0	0	5.41	3.56
14	0	0	0	1	0	0	0	0	0	0	0	1	0	5.67	2.70
15	0	0	0	1	0	0	0	0	0	0	0	0	1	6.10	2.26
16	0	0	0	0	1	0	0	0	0	0	0	1	0	6.05	1.75
17	0	0	0	0	1	0	0	0	0	0	0	0	1	6.48	1.43
18	0	0	0	0	0	1	0	0	0	0	1	0	0	5.70	4.50
19	0	0	0	0	0	1	0	0	0	0	0	1	0	5.90	3.50
20	0.5	0.5	0	0	0	0	1	0	0	0	0	0	0	3.14	4.44
21	0.5	0	0.5	0	0	0	1	0	0	0	0	0	0	3.32	4.05
22	0	0.5	0.5	0	0	0	1	0	0	0	0	0	0	3.31	2.82
23	0	0	0	0.5	0	0.5	0	0	0	0	1	0	0	5.52	3.95
24	0	0	0	0	0	1	0	0	0	0	0.5	0.5	0	5.81	3.95
25	0	0	0	1	0	0	0	0	0	0	0.5	0.5	0	5.52	3.14
26	0	0	0	0.5	0	0.5	0	0	0	0	0	1	0	5.78	3.13
27	0	0	0	0.5	0.5	0	0	0	0	0	0	1	0	5.86	2.18
28	0	0	0	1	0	0	0	0	0	0	0	0.5	0.5	5.90	2.46
29	0	0	0	0	1	0	0	0	0	0	0	0.5	0.5	6.28	1.58
30	0	0	0	0.5	0.5	0	0	0	0	0	0	0	1	6.27	1.88
31	0.5	0.5	0	0	0	0	0	1	0	0	0	0	0	5.46	2.36
32	0	1	0	0	0	0	0	0.5	0.5	0	0	0	0	5.61	1.93
33	0.5	0.5	0	0	0	0	0	0	1	0	0	0	0	5.65	1.84
34	0	0.5	0.5	0	0	0	0	1	0	0	0	0	0	5.68	1.91
35	0.5	0	0.5	0	0	0	0	1	0	0	0	0	0	5.69	1.98
36	1	0	0	0	0	0	0	0	0.5	0.5	0	0	0	5.88	1.89
37	0.5	0	0.5	0	0	0	0	0	1	0	0	0	0	5.87	1.46
38	0	1	0	0	0	0	0	0	0.5	0.5	0	0	0	5.85	1.06
39	0.5	0.5	0	0	0	0	0	0	0	1	0	0	0	6.11	1.21
40	0	0.5	0.5	0	0	0	0	0	1	0	0	0	0	5.82	0.84
41	0	0	1	0	0	0	0	0.5	0.5	0	0	0	0	5.94	0.88
42	0.5	0	0.5	0	0	0	0	0	0	1	0	0	0	6.27	0.87
43	0	0.5	0.5	0	0	0	0	0	0	1	0	0	0	6.26	0.40
44	0	0	1	0	0	0	0	0	0.5	0.5	0	0	0	6.24	0.20

Table 2. Some Basic Chemical Element Properties

element	electro-negativity (E, $eV^{1/2}$)	atomic no. (AN)	melting temp (T, K)	pseudo-potential radii (R, au)	valence electron (V)
Al	1.61	13	933	1.675	3
Cd	1.69	48	594	2.215	12
Ga	1.81	31	302	1.695	3
In	1.78	49	429	2.05	3
Mg	1.31	12	922	2.03	2
Zn	1.65	30	692	1.88	12
As	2.18	33	1090	1.415	5
N	3.04	7	63	0.54	5
P	2.19	15	317	1.24	5
S	2.58	16	388	1.1	6
Sb	2.05	51	904	1.765	5
Se	2.55	34	494	1.285	6
Te	2.1	52	722	1.67	6

that the predicted values are in good agreement with the experimental values.

We further test the correlation model by screening all possible binary III−V compounds. The following 25 compounds are examined assuming they do exist: BN, BP, BAs, BSb, Bbi, AlN, AlP, AlAs, AlSb, AlBi, GaN, GaP, GaAs, GaSb, GaBi, InN, InP, InAs, InSb, InBi, TlN, TlP, TlAs, TlSb, and TlBi. Of the 25 compounds, only BN, BP, BAs, BSb, BBi, AlBi, GaBi, InBi, TlN, TlP, TlAs, TlSb, and TlBi will be selected for testing since the other 12 compounds have been used before in the development of the correlation model.

For the band gap energy correlation model, InBi and TlBi are found to be out of the range of the training set; i.e., they have either large or small O_1/O_4 values. As a result, these two compounds will not be used in the testing. For the other 11 compounds, BN and TlN are located in the right region of Figure 9 (BE > 2.5); BP, BAs, BSb, BBi, AlBi, GaBi, TlP, TlAs, and TlSb are located in the left region of Figure 9 (BE < 2.5). Figure 10 shows the prediction of TlSb, InBi, and TlBi using the developed lattice constant correlation model. These three compounds have O_3/O_2 values within the range of the training set. The other 10 compounds are discarded. It is obvious from Figure 10 that InBi is located in the bandlike region (3.6 < LC < 5.8). For TlSb and TlBi, they have either LC < 3.6 or LC > 5.8.

Similarly, group II and VI elements from the periodic table can also be screened and tested using the developed correlation models. Since MgS and MgSe have been used in the models, they are excluded from the test set. The following 28 compounds from the periodic table are tested assuming they do exist: BeO, BeS, BeSe, BeTe, BePo, MgO, MgTe, MgPo, CaO, CaS, CaSe, CaTe, CaPo, SrO, SrS, SrSe, SrTe, SrPo, BaO, BaS, BaSe,

Table 3. Property Map Coordinates, Response Variables, and Class Labels

O_1	O_2	O_3	O_4	O_5	LC (Å)	BE (eV)	C_1	C_2
1.43	6	859	1.135	2	3.11	6.20	1	1
0.58	2	605	0.435	2	5.47	2.45	2	2
0.57	20	168	0.260	2	5.66	2.14	2	2
0.44	38	18	0.090	2	6.14	1.63	1	2
1.23	24	239	1.155	2	3.16	3.44	1	1
0.38	16	15	0.455	2	5.45	2.27	2	2
0.37	2	788	0.280	2	5.65	1.43	2	2
0.24	20	602	0.070	2	6.10	0.70	1	2
1.26	42	366	1.510	2	3.54	1.95	1	2
0.41	34	112	0.810	2	5.87	1.34	1	2
0.40	16	661	0.635	2	6.06	0.36	1	2
0.27	2	475	0.285	2	6.48	0.18	1	2
0.93	14	304	0.780	6	5.41	3.56	2	1
0.90	4	198	0.595	6	5.67	2.70	2	1
0.45	22	30	0.210	6	6.10	2.26	1	2
0.86	14	100	0.930	6	6.05	1.75	1	2
0.41	4	128	0.545	6	6.48	1.43	1	2
1.27	4	534	0.930	4	5.70	4.50	2	1
1.24	22	428	0.745	4	5.90	3.50	1	1
1.33	15	549	1.145	2	3.14	4.44	1	1
1.345	24	612.5	1.323	2	3.32	4.05	1	1
1.245	33	302.5	1.333	2	3.31	2.82	1	1
1.10	9	419	0.855	5	5.52	3.95	2	1
1.255	13	481	0.838	4	5.81	3.95	1	1
0.915	9	251	0.688	6	5.52	3.14	2	1
1.07	13	313	0.670	5	5.78	3.13	2	1
0.88	9	149	0.763	6	5.86	2.18	1	2
0.675	13	114	0.403	6	5.90	2.46	1	2
0.635	9	114	0.738	6	6.28	1.58	1	2
0.430	13	79	0.378	6	6.27	1.88	1	2
0.48	9	310	0.445	2	5.46	2.36	2	2
0.375	9	401.5	0.368	2	5.61	1.93	2	2
0.47	11	478	0.270	2	5.65	1.84	2	2
0.395	25	63.5	0.633	2	5.68	1.91	2	2
0.495	18	358.5	0.623	2	5.69	1.98	2	2
0.505	29	93	0.175	2	5.88	1.89	1	2
0.485	18	414.5	0.448	2	5.87	1.46	1	2
0.305	11	695	0.175	2	5.85	1.06	1	2
0.34	29	310	0.080	2	6.11	1.21	1	2
0.385	9	724.5	0.458	2	5.82	0.84	1	2
0.405	25	386.5	0.723	2	5.94	0.88	1	2
0.355	20	246.5	0.188	2	6.27	0.87	1	2
0.255	11	538.5	0.178	2	6.26	0.40	1	2
0.335	9	568	0.46	2	6.24	0.20	1	2

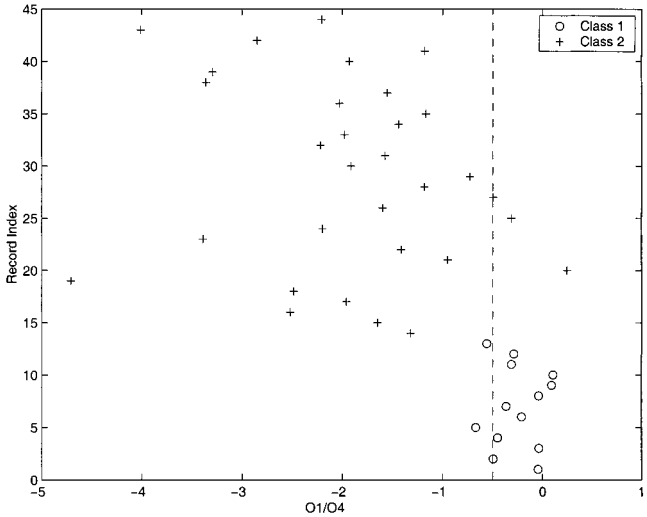

Figure 3. Classification of band gap energy using electronegativity (O_1) and pseudopotential radii (O_4).

BaTe, BaPo, RaO, RaS, RaSe, RaTe, and RaPo. Out of these 28 compounds, only the following 12 compounds are within the range of the developed correlation models: BeS, MgO, CaO, CaS, CaSe, SrS, SrSe, SrTe,

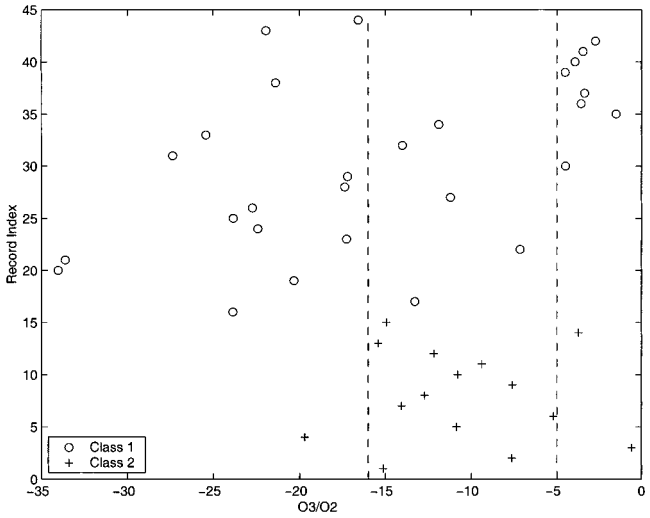

Figure 4. Classification of lattice constant using atomic number (O_2) and melting temperature (O_3).

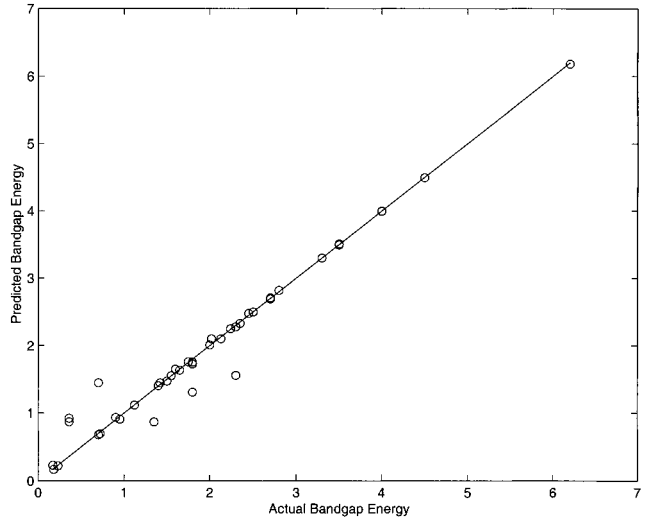

Figure 5. Comparison of actual and predicted band gap energy (training set).

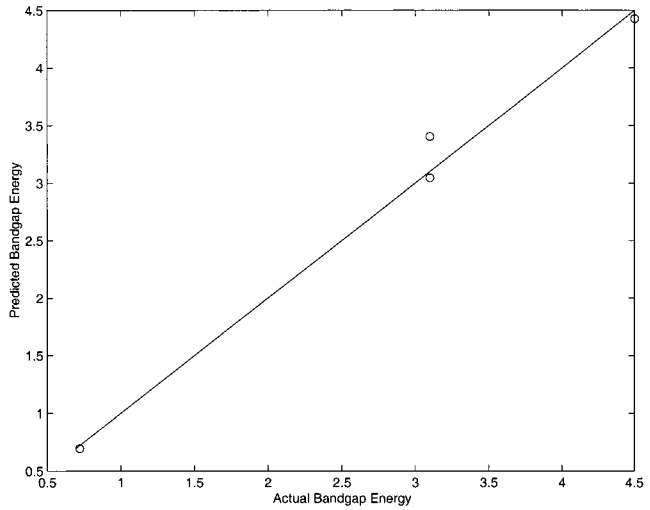

Figure 6. Comparison of actual and predicted band gap energy (validation set).

BaSe, BaTe, BaPo, and RaPo. All these 12 compounds are located in the right region of Figure 9 (BE > 2.5). It is obvious from Figure 10 that the four compounds SrSe,

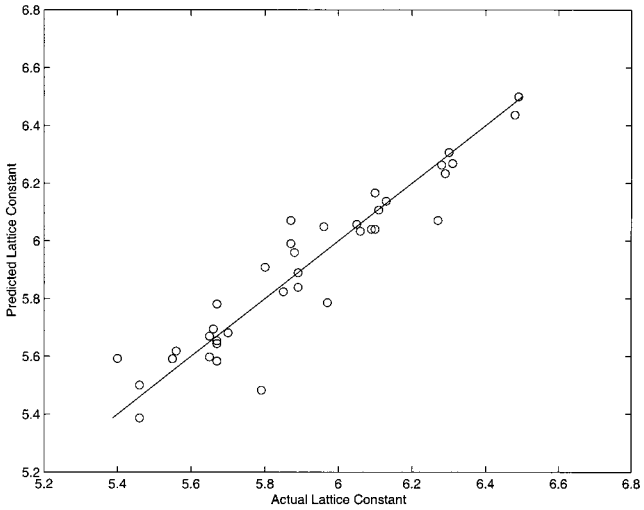

Figure 7. Comparison of actual and predicted lattice constant (training set).

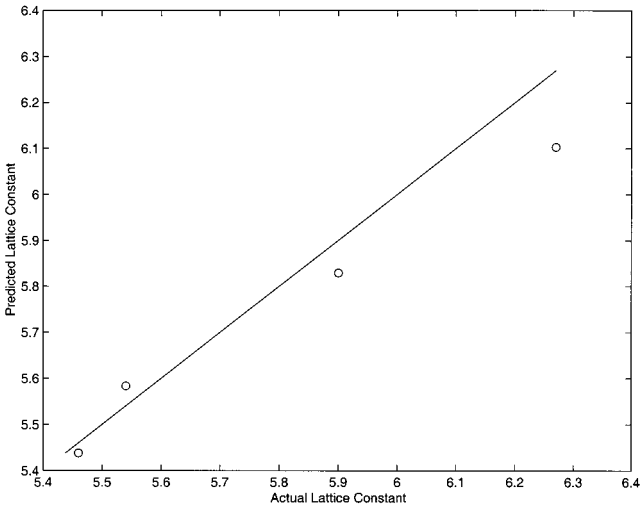

Figure 8. Comparison of actual and predicted lattice constant (validation set).

SrTe, BaTe, and RaPo are located in the bandlike region ($3.6 < LC < 5.8$). For the other eight compounds, they have either $LC < 3.6$ or $LC > 5.8$.

Band gap energy is the energy needed to remove an electron across a bond in the solid. In this paper, the factor O_1/O_4 is identified to be closely related to the band gap energy. O_1 represents electronegativity, which is a measure of the competing attraction of different atoms for the available valence electrons in compound or chemical bond formation. It is related to the difference of valence energy levels of its constituent elements. O_4 represents pseudopotential radii, which is the atomic size difference. Daams[10] showed that the average atomic radius of a compound, $(R_A + R_B)/2$, was strongly correlated to the shortest interatomic distance d (or bond length) for a few material systems:

$$d = k_1(R_A + R_B)/2 + k_2 \qquad (5)$$

In eq 5, R_A and R_B are the atomic radius of A and B atoms in the A_i-B_j compound, respectively; k_1 and k_2

(10) Daams, Jo L. C. *Intermetallic Compounds Principles and Practice;* John Wiley & Sons: New York, 1995; Vol. 1, p 363.

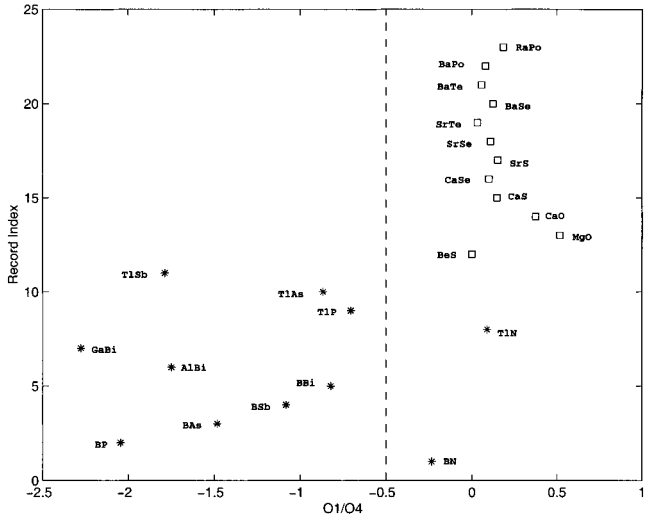

Figure 9. Prediction of band gap energy of some III−V and II−VI semiconductor compounds (assuming they exist) using electronegativity (O_1) and pseudopotential radii (O_4).

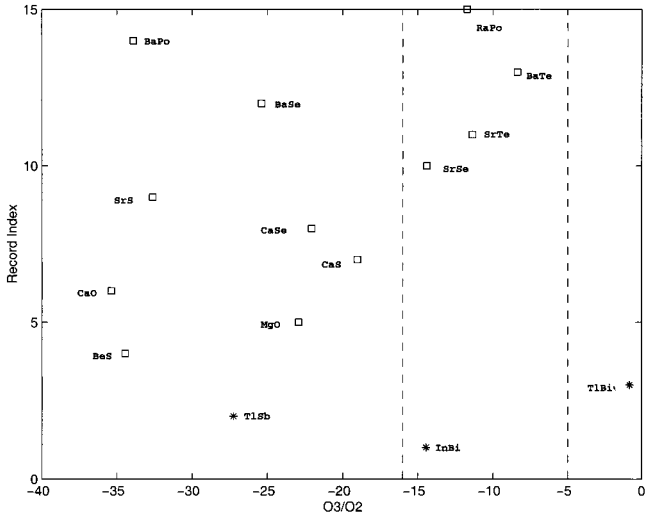

Figure 10. Prediction of lattice constant of some III−V and II−VI semiconductor compounds (assuming they exist) using atomic number (O_2) and melting temperature (O_3).

are constants that are independent of atomic sizes. From eqs 4 and 5, it is obvious that O_4 is proportional to the shortest interatomic distance d, which is a measure of the bond strength. In summary, O_1 is a measure of the intrinsic capability of the constituent elements in attracting valence electrons to form a chemical bond, and O_4 (through d) is a measure of the short-range atomic arrangement in the compound, which may represent repulsive force between the A−B atomic pair.[11] The term O_1/O_4 can therefore be regarded as a quantity to measure the competition between the attractive and repulsive forces when the compound is formed. In other words, the band gap energy of a binary semiconductor system may be governed by the measure of competition between the attractive and repulsive forces (O_1/O_4).

The melting temperature and atomic number are identified to be significant in response to the lattice

(11) Chen, Y.; Iwata, S.; Liu, J. N.; Villars, P.; Rodgers, J. *Model. Simul. Mater. Sci. Eng.* **1996**, *4*, 335.
(12) Rice, S. A.; Nachtrieb, N. H. *J. Chem. Phys.* **1959**, *31*, 139.
(13) Nachtrieb, N. H.; Handler, G. S. *Acta Metall.* **1954**, *2*, 797.
(14) Tiwari, G. P. *Z. Metallkd.* **1981**, *72*, 211.

constant in the form of O_3/O_2. A close relationship between the process of diffusion and melting of metals has been shown by Rice and Nachtrieb.[12] A relation between the activation energy for self-diffusion (D) and the melting point (T_f) was shown[13]

$$D = KT_f$$

where K is a constant, and its value is expected to be vary from one system to another. Furthermore, Tiwari[14] showed that this correlation between the parameters of self-diffusion and melting is valid for normal metals as well as the metals showing anomalies in their diffusion behaviors. Therefore, O_3 is proportional to the diffusion behaviors of the element. Atomic number is proportional to the mass of the element. Since the element melting point and atomic number are strongly related to its self-diffusion parameter and mass (m), respectively, the term O_3/O_2 is proportional to D/m. The term D/m may be regarded as a measure of atomic diffusion capability normalized by its mass or, alternatively, a measure of atomic speed since D/m has the same unit as the square of speed. The strong correlation observed between LC and D/m may suggest that the lattice constant of a binary semiconductor relies on how fast atoms redistribute themselves through diffusion during its formation process.

Conclusions

Two bulk semiconductor properties, band gap energy and lattice constant, are found strongly correlated to a few fundamental constituent element properties, which are electronegativity and psuedopotential radii for the former and melting temperature and atomic number for the latter. It further suggests that the band gap energy and lattice constant rely on the result of the competition between attractive and repulsive forces and on how easily the atoms may diffuse taking into account their mass, respectively, in the compound formation process. New semiconductors may be designed on the basis of the developed correlation.

CM9906194