Computer-aided design of new wide bandgap semiconductors with chalcopyrite structure

N. N. Kiselyova, A. V. Stolyarenko, T. Gu*, W. Lu*

A.A.Baikov Institute of Metallurgy and Materials Science of RAS,
49 Leninskii Prospect, 119991 GSP-1, Moscow; Russia. Tel. (495)-135-25-91, e-mail: kis@ultra.imet.ac.ru
**College of Sciences, Shanghai University, Shanghai 200444, China. Tel: +86-021-66132663 (O); e-mail: wclu@staff.shu.edu.cn*

The wide class of algorithms of computer training was used for design of new wide bandgap semiconductors with composition ABC_2 (A and B — various chemical elements; C — S, Se, Te, N, P, As, or Sb) with chalcopyrite crystal structure type. Application of methodology of collectives of algorithms in order to make a decision about belonging of substance to one or another class of phases has allowed considerably increase of reliability of the prediction of inorganic compound properties.

Introduction

Recent developments in optoelectronics and other fields of practical applications (in particular, solar power, equipments for detecting photons and charged particles) have stimulated an interest in wide bandgap semiconductors. Wide bandgap semiconductors are electronic materials in which the energy of the band-to-band electronic transitions exceeds approximately 2 eV [1]. Recently wide bandgap semiconductors with tetragonal chalcopyrite crystal structure type have attracted attention of the researchers [2]. Chalcopyrites have aroused interest because of their promising applications to nonlinear optics [3 – 5] and spintronics [6, 7] also.

Taking into account practical importance of chalcopyrites, the attempts of design of new compounds with this crystal structure and estimation of their some properties were made. The majority of researches were concerned with quantum-mechanical calculations of properties of already known compounds [8, 9]. For example, authors of [8] used full potential linearized augmented plane waves method within the local density approximation (LDA) for calculation of parameters of lattice, bulk modulus, bandgap and other properties of already well investigated chalcopyrites with composition $CuGaS_2$ and $AgGaS_2$. However theoretical estimation of bandgap of these compounds calculated in the paper [8], — 0,903 ($CuGaS_2$) and 0,9485 eV ($AgGaS_2$) differed sharply from experimental values — 2,44 ($CuGaS_2$) and 2,75 eV ($AgGaS_2$).

The arising at quantum-mechanical calculations of new semiconductors problems became the reason of development of empirical methods intended for an estimation of properties of compounds not yet synthesized [10 – 18]. These methods are based on the analysis of data about known semiconducting substances and the search for regularities connecting desired properties of compounds with properties of components. The used empirical methods can conditionally be divided into two groups: (1) methods which are intended for search for the simplest regularities including two or three properties, or algebraic functions of properties of components [10 – 13], and (2) methods which use multidimensional regularities including a wide set of properties of components [14 – 18]. The latter are based on application of various methods of the computer data analysis including multidimensional statistical analysis, pattern recognition, etc. [18 – 20]. Taking into account complexity of periodic dependences connecting property of compounds with properties of chemical elements which are included into their composition, computer analysis of the information and the search for multi-

dimensional regularities is more promising way for design of new inorganic compounds.

Statement of a problem and choice of methods of the computer data analysis

The task of design of new compounds with composition ABX_2 (A and B — various chemical elements; X — S, Se, Te, N, P, As or Sb) with chalcopyrite crystal structure type and with width of the forbidden zone (ΔE) more than 2 eV under normal conditions was stated. The information about properties of chemical elements and already known compounds of the above-mentioned composition for computer analysis was extracted from databases by Baikov Institute "Elements", "Phase", "Bandgap" and "Diagram" [14, www.imet-db.ru].

Let an each compound corresponds to a point in multidimensional space of properties of elements. Owing to periodicity of properties of chemical elements points, which correspond to combinations of close on properties elements, combining into compounds, form compact clusters. Thus, the task of a search for regularities connecting property of inorganic compounds with properties of chemical elements can be reduced to a problem of computer training in pattern recognition. In this case the analysis of the information about already known compounds, which are represented as a set of values of properties of chemical elements, allows discovery of classifying regularities. The latter allow separation of known compounds into predetermined classes. It is possible to predict new compounds and estimate their unknown parameters by substitution of the property values of the appropriate chemical elements into the found regularities.

After testing the programs the following software of pattern recognition were chosen:

— system RECOGNITION developed by A.A.Dorodnicyn Computer Center of Russian Academy of Sciences [20]. This multifunctional system of pattern recognition includes algorithms of *k*–nearest neighbors, Fisher's linear discriminant, linear machine, multi-level perceptron (neural networks), support vector machine, genetic algorithm, estimates calculation algorithms, LoReg (Logical Regularities), deadlock test algorithm, statistical weighted syndromes, etc.;

— system of concept formation ConFor developed by V.M.Glushkov Institute of Cybernetics of National Academy of Sciences of Ukraine [19]. The system is based on special data structure in a computer memory named as growing pyramidal networks;

— system of the computer data analysis MASTER developed by Shanghai University which includes the various pattern recognition algorithms, artificial neural networks and support vector machine [18].

Application of set of methods posed the problem of comparison of results obtained with use of various algorithms of computer training. The strategy of decision making on the basis of usage of collectives of algorithms has been advanced to apply to the solution of this task. The idea of this approach is as follows. The task is solved in two steps. Firstly all or part from available algorithms are used independently. Further the final "collective" decision is calculated using the obtained decisions of individual algorithms. Using this approach one would expect that mistakes of individual algorithms at synthesis of the collective decision could be compensated by the correct predictions of other algorithms. The simplest collective method is one of majority that refers the object to that class, for which the majority of algorithms has voted. However voting of the predictions not always allows improvement of reliability of prediction for very complex tasks, especially, in case of weak fulfillment of the basic hypothesis of pattern recognition — hypothesis of compactness of objects in elemental property space. More complex methods of collective decision making, such as the convex stabilizer, Bayesian correction, Wood's dynamic method, logic correction of sets of recognizing algorithms, decision templates, clustering and selection, etc. [20], are intended for the solution of very difficult tasks of prediction and were used in the present investigation.

The procedure of chalcopyrite design included three stages:

(I) Computer analysis of data about known compounds which were represented in a computer memory by a set of property values of chemical elements which are included into their composition (training set), and construction of classifying regularity with use of various algorithms of pattern recognition (computer training). The procedure of cross-validation [20], in which the given part (*f*) of objects was withdrawn from training set, computer training using residuary objects was carried out, and the withdrawn objects were used for examination recognition, was used for an estimation of quality of recognition. The number of mistakes was fixed. Then the "control" objects were came back into training set and other *f* objects were withdrawn from it, and the procedure was repeated so long as all objects of training set have not been as control examples. Such procedure allows more objective estimation of accuracy of algorithms of pattern recognition for certain training set, especially in the case of methods, for which the undesirable effect of "overfitting" is observed, when the algorithm is "adjusted" only to training set and subsequently gives bad results for new substances. After

end of training with cross-validation the average error of recognition is calculated, which further was used for a choice of the best algorithms of the collective decision making.

(II) Collective decision was made only for the algorithms which have shown the best results at cross-validation. As a rule, all algorithms of the collective decision making of system "RECOGNITION" [20] were used.

(III) The most exact algorithms of pattern recognition and collective decisions making, which were chosen on first two stages, were used further for the predicting new compounds and estimating their properties.

Such complex procedure has allowed an increase of accuracy of predicting.

Results of computer design of new chalcopyrites

The task of design of new wide bandgap semiconductors with crystal structure of chalcopyrite was divided into three subtasks:

(1) prediction of formation under normal conditions of new compounds with composition ABC_2 (A and B — various chemical elements; C — S, Se, Te, N, P, As or Sb);

(2) prediction of chalcopyrite crystal structure under normal conditions for the above mentioned new compounds;

(3) estimation of width of the forbidden zone and melting point of new chalcopyrites.

To illustrate potential of proposed methods of computer-aided design of inorganic compounds the

Table 1

Prediction of D*E* of known chalcopyrites

Compound	Exp. class	Δ*E* exp., eV	Prediction							
			BC	C&S	DT	WDM	CS	CM-MV	CM-A	LC
$CuAlS_2$	1	3.5	1	1	1	1	1	1	1	1
$CuGaS_2$	1	2.44	1	1	1	1	1	1	1	1
$CuInS_2$	2	1.5	2	2	2	2	2	2	2	2
$CuAlSe_2$	1	2.67	1	1	1	1	1	1	1	1
$CuGaSe_2$	2	1.63	2	2	2	2	2	2	2	2
$CuInSe_2$	2	0.95	2	2	2	2	2	2	2	2
$CuAlTe_2$	1	2.06	1	2	?	1	1	?	?	?
$CuGaTe_2$	2	1.18	2	2	2	2	2	2	2	2
$CuInTe_2$	2	0.88	2	2	2	2	2	2	2	2
$AgAlS_2$	1	3.13	1	1	1	1	1	1	1	1
$AgGaS_2$	1	2.75	1	1	1	1	1	1	1	1
$AgAlSe_2$	1	2.55	1	1	1	1	1	1	1	1
$AgGaSe_2$	2	1.65	2	2	2	2	2	2	2	2
$AgInSe_2$	2	1.24	2	2	2	2	2	2	2	2
$AgAlTe_2$	2	1.8	2	2	2	2	2	2	2	2
$AgGaTe_2$	2	1.1	2	2	2	2	2	2	2	2
$AgInTe_2$	2	0.96	2	2	2	2	2	2	2	2
$ZnSiP_2$	1	2.07	1	1	1	?	1	1	1	1
$ZnSiAs_2$	1	2.1	1	1	1	?	1	1	1	1
$ZnGeN_2$	1	2.9	1	1	1	?	1	1	1	1
$ZnGeP_2$	1	2.1	1	1	1	?	1	1	1	1
$ZnGeAs_2$	2	1.16	2	2	2	2	2	2	2	2
$ZnSnP_2$	2	1.45	2	2	2	2	2	2	2	2
$ZnSnAs_2$	2	0.74	2	2	2	2	2	2	2	2
$ZnSnSb_2$	2	0.4	2	2	2	?	2	2	2	2
$CdSiP_2$	1	2.2	1	1	1	?	1	1	1	1
$CdGeP_2$	2	1.8	2	2	2	?	2	2	2	2
$CdGeAs_2$	2	0.53	2	2	2	?	2	2	2	2
$CdSnP_2$	2	1.16	2	2	2	?	2	2	2	2
$CdSnAs_2$	2	0.3	2	2	2	?	2	2	2	2
$AgInS_2$	2	1.9	2	2	2	?	2	2	2	2
$CdSiAs_2$	2	1.51	2	2	2	?	2	2	2	2
$CuFeS_2$	2	0.53	2	2	2	?	2	2	2	2
$CuFeSe_2$	2	0.16	2	2	2	?	2	2	2	2
$CuFeTe_2$	2	0.1	2	2	2	?	2	2	2	2
$LiGaTe_2$	1	2.31	1	1	1	?	1	1	1	1
$LiInTe_2$	2	1.46	2	2	2	?	2	2	2	2
$AgFeSe_2$	2	0.23	2	2	2	?	2	2	2	2
$MgSiP_2$	1	2.35	1	1	1	?	1	1	1	1
$MnGeP_2$	2	0.24	2	2	2	?	2	2	2	2
$MnGeAs_2$	2	0.6	2	2	2	2	2	2	2	2

results of the predictions of bandgap (ΔE) for already investigated chalcopyrites are given in table 1. The following designations for the predictions were used: 1 — $\Delta E > 2$ eV; 2 — $\Delta E < 2$ eV; ? — uncertain prediction; for collective methods of decision making: BC — Bayesian correction, C&S — clustering and selection, DT — decision templates, WDM — Wood's dynamic method, CS — convex stabilizer, CM-MV — committee method (voting on the majority), CM-A — committee method (averaging), and LC — logic correction.

The compounds of composition ABC_2 were represented in a computer memory as a set of values of the following properties of elements A, B and C:

— electronegativity in the Martynov-Batsanov scale of values $\Delta\chi$, where $\Delta\chi = |2\chi_C - \chi_A - \chi_B|$;

— atomic electrovalent Z_A, Z_B, Z_C (for transition metals, the group number of element is used as Z);

— mean Born exponent $\overline{n} = \frac{n_A + n_B + 2n_C}{4}$;

— chemical scale χ of Pettifor [21];

— the proportion:

$$\left(\frac{I_z}{Z}\right)_{AC} = \left(\frac{I_z}{Z}\right)_A - \left[6 + 0.1\left(\frac{I_Z}{Z}\right)_C\right],$$

where I_z — final ionization potential;

— atomic radius.

In case of individual algorithms the best results at the cross-validation were obtained with use of algorithms of the linear machine and logic regularities [20]. The results of these methods were used for the collective decision making. The all correct predictions were obtained at use of Bayesian method and convex stabilizer. These collective methods of decision making, and also algorithms of the linear machine and logic regularities were used at the prediction of bandgap under normal conditions for new chalcopyrites (table 2).

Table 2

Prediction of ΔE for new chalcopyrites

Compound	BM	CS
$ZnAlS_2$	1	1
$ZnAlSe_2$	1	1
$ZnAlTe_2$	2	2
$AgFeS_2$	2	2
$AgFeTe_2$	2	2
$ZnGaTe_2$	2	2
$CdGaTe_2$	2	2
$HgGaTe_2$	2	2

According to our predictions, chalcopyrites $ZnAlS_2$ and $ZnAlSe_2$ belong to wide bandgap semiconductors promising for optoelectronic applications.

Conclusions

Results of computer design of new wide bandgap semiconductors with chalcopyrite structure has shown that application of computer training allows successful design of new inorganic compounds with the predefined properties.

Financial support from RFBR (grants №06-07-89120 and 05-03-39009) is gratefully acknowledged.

References

1. Vavilov V.S. Physics and applications of wide bandgap semiconductors. Physics – Uspekhi. 1994, v.37, p.269 – 277.
2. Siebentritt S. Wide gap chalcopyrites: material properties and solar cells. The solid films, 2002, v.403 – 404, p.1 – 8.
3. Verozubova G.A., Gribenyukov A.I., Korotkova V.V., *et al.* Synthesis and growth of $ZnGeP_2$ crystals for nonlinear optical applications. J.Crystal Growth., 2000, v.213, p.334 – 339.
4. Bai L., Lin Z., Wang Z., Chen C. Mechanism of linear and nonlinear optical effects of chalcopyrite $AgGaX_2$ (X = S, Se, and Te) crystals. J.Chem.Phys., 2004, v.120, p.8772 – 8778.
5. Isaenko L., Krinitsin P., Vedenyapin V., *et al.* $LiGaTe_2$: A new highly nonlinear Chalcopyrite Optical Crystal for the Mid-IR. Crystal Growth & Design, 2005, v.5, p.1325 – 1329.
6. Pearton S.J., Abernathy C.R., Norton D.P., *et al.* Advances in wide bandgap materials for semiconductor spintronics. Mater.Sci. and Eng., 2003, v.R40, p.137–168.
7. Zutic I., Fabian J., Erwin S.C. Bipolar spintronics: Fundamentals and applications IBM J. Res. & Dev., 2006, v.50, p.121 – 139.
8. Laksari S., Chahed A., Abbouni N., *et al.* First-principles calculations of the structural, electronic and optical properties of $CuGaS_2$ and $AgGaS_2$. Comput.Mater.Sci., 2006, v. 38, p. 223 – 230.
9. Freeman A.J. Materials by design and the exciting role of quantum computation/simulation. J.Comput.and Appl. Mathematics, 2002, v.149, p.27–56.
10. Goodman C.H.L. The prediction of semiconducting properties in inorganic compounds. Phys. and Chem. Solids., 1958, v.6, p.305 – 314.
11. Makino Y. Interpretation of band gap, heat of formation and structural mapping for sp-bonded binary compounds on the basis of bond orbital model and orbital electronegativity. Intermetallics, 1994, v.2, p.55 – 66.
12. Matsushita H., Katsui A. Materials design for Cu-based quaternary compounds derived from chalcopyrite-rule. J.Phys. and Chem.Solids., 2005, v.66, p.1933 – 1936.
13. Samanta L.K., Chatterjee S. On the energy band gap correlation and effective charges of some ternary

chalcopyrites and their mixed versions. Infrared Phys. & Technol., 2005, v.46, p. 370 – 378.
14. Kiselyova N.N. Computer-Aided Design of Inorganic Compounds. Application of Databases and Artificial Intelligence. Moscow: Nauka, 2005, 289 p.
15. Jackson A.G., LeClair S.R., Ohmer M.C. *et al.* Rough sets applied to materials data. Acta mater., 1996, v.44, no.11, p.4475 – 4484.
16. Jackson A.G., Pawlak Z., LeClair S.R. Rough sets applied to the discovery of materials knowledge. J.Alloys and Compounds, 1998, v.279, p.14 – 21.
17. Gu T., Lu W., Bao X., Chen N. Using support vector regression for the prediction of the band gap and melting point of binary and ternary compound semiconductors. Solid State.Sci., 2006, v.8, p.129 – 136.
18. Chen N.Y., Lu W.C., Yang J., Li G.Z. Support Vector Machine in Chemistry. Singapore: World Scientific Publishing Co. Pte. Ltd., 2004, 331 p.
19. Gladun V.P. Processes of Formation of New Knowledge. Sofia: SD "Pedagog 6". 1995, 192 p.
20. Zhuravlev Yu.I., Ryazanov V.V., Senko O.V. RECOGNITION. Mathematical Methods. Software System. Practical Solutions. Moscow: Phasis. 2006, 176 p.
21. Pettifor D.G. A chemical scale for crystal-structure maps. Solid State Commun. 1984, v.51, p.31 – 34.