Prediction of Inorganic Compounds: Experiences and Perspectives

N.N. Kiselyova

Introduction

There have been many achievements in the last decade in the development of materials science, chemistry, and physics (experimental and theoretical). However, the most difficult problem—calculating the intrinsic properties of multicomponent compounds starting from the knowledge of their constituent components' properties—still remains unsolved. Calculations or predictions based on only the properties of constituent components (or simply, properties) are called *a priori* calculations or predictions. These difficulties are due to the solution of mathematical problems arising in the quantum mechanical calculations of multi-electronic systems. As a result, scientists make use of many empirical prediction methods that use existing regularities from a variety of property data. Some of the empirical criteria for the formation of compounds with predefined properties use the rules of Hume-Rothery,[1] Laves,[2] Mathias,[3] Goldschmidt,[4] Villars,[5] and Darken-Gurry.[6]

The main problem of finding such criteria (classification rules) is the search for the appropriate parameters (expressions) of its constituent components with which it would be somehow possible to divide the physico-chemical systems into distinct domains. The advantage of such criteria consists in its simplicity and the ability to visualize the results with the help of two-dimensional plots. Often, however, the separation into distinct domains using these two-dimensional plots is deficient because other properties of the constituent components come into play. In addition, the two-dimensional criteria quickly lose their reliability when new data do not easily fit within the framework outlined by the classification rules. An ideal classification scheme must be adaptable to new phenomena, have a flexible structure, and be useful for recognizing new properties. Such a classification scheme will not restrict itself to the narrow boundaries of two-dimensional criteria (two-parameter planes). The search for multidimensional criteria (classification rules) could work with the aid of computer learning techniques in conjunction with databases, e.g., crystallographic, phase diagram, and physical properties.

In principle, there are three ways to predict new inorganic compounds and forecast their intrinsic compound properties, based on the knowledge of their constituent component properties:

- quantum-mechanical calculations,
- two-dimensional criteria (classification rules) found by semi-empirical approaches, and
- multidimensional criteria (classification rules) found by computer learning techniques (computational methods for acquiring new knowledge, new skills, and new ways to organize existing knowledge).

It should be emphasized that, until recently, the quantum mechanical calculations have failed to bring about a single prediction of new compounds that would be valuable to inorganic materials science. This situation will not change much in the near future. The two-dimensional criteria approach, however, is more efficient, but, in my opinion, not sufficiently flexible and comprehensive. The multidimensional criteria approach (cybernetic) is more suitable for *a priori* prediction of inorganic compounds. The first experiments using computer learning methods to search for multidimensional criteria for the formation of the binary phases[7] have been productive.

Concepts of the Cybernetic Approach

The advantages of using computer learning procedures to search for multidimensional criteria are apparent. They consist of (1) analyses of large databases and (2) computer learning methods to locate many and complex criteria.

Locating and analyzing such criteria is a computational task. For example, let a certain phenomenon be described by N properties $x_1, x_2 \ldots, x_N$, each of which has k discrete values (x_{ij}; $i = 1, 2 \ldots, N$; $j = 1, 2 \ldots, k$). If we assume that the formation of a certain type of the crystal structure of a given class of compounds depends on the properties of its constituent chemical elements, then the traditional way of representing the semi-empirical, two-dimensional criteria can be reduced to a Boolean expression that takes the form

$$(x_{12} \,\&\, x_{34}) \vee (x_{11} \,\&\, x_{23} \,\&\, x_{32}) \,\&\, \mathrm{NOT}\ (x_{11} \,\&\, x_{25} \,\&\, x_{37}) \vee (x_{2k} \,\&\, x_{39} \,\&\, x_{N4}), \quad (1)$$

where x_{ij} is the magnitude of the interval of the change of i properties or algebraic functions of these properties; and &, V, and NOT are respectively the symbols for the conjunction, disjunction, and negation operators. In the case of a semi-empirical analysis of the information, due to human limitations, the number of the simple conjunctions in Expression 1 does not exceed five or six, each conjunction having not more than three or four values. Even though the form of Expression 1 is unusual, all known semi-empirical rules can be well-described by it, e.g., Hume-Rothery, Laves, Mathias.

The cybernetical way of searching for the criteria consists of solving complicated criteria of the type

$$(x_{11} \,\&\, x_{26} \,\&\, x_{46} \ldots \&\, x_{159}) \vee (x_{29} \,\&\, x_{33} \,\&\, x_{48}) \vee \ldots \vee (x_{59} \,\&\, x_{68} \,\&\ldots\&\, x_{N4})\,(x_{2k} \,\&\, x_{37}) \vee (x_{18} \,\&\, x_{48} \,\&\ldots\&\, x_{Nk}) \vee (x_{17} \,\&\, x_{26} \,\&\ldots\&\, x_{79}) \ldots \&\, \mathrm{NOT}\ (x_{59}). \quad (2)$$

In this case, the complexity and the number of the conjunctions are limited only by the storage capacity of the computer. Computer learning strategies are used to search for the multidimensional criteria of Equation 2. As input, we use experimentally known compounds that possess the predefined property of interest, as

well as examples of the experimental knowledge of the absence of such compounds (i.e. knowledge of compounds *not* formed). All this information makes up the "learning step" (which includes the acquisition of knowledge, the development of cognitive skills through practice, the organization of new knowledge, and the discovery of new facts through observation and experimentation). Therefore, the method is an alternative learning method.

The principal difficulty of the direct-search multidimensional classification lies in the great number of searches in the selection of the simple conjunctions characteristic of a given class. In fact, if we have N properties of the constituent components, each of which can take k values, then the number of possible searches is equal to 2^{N^k}. So, looking at only ten such properties with ten discrete values, the number of searches would be much too large for practical applications. To reduce the number of searches, we use the Devingtal algorithm.[89] This algorithm, used for predicting the binary compounds, restricts the complexity of the searched conjunctions of the properties of its constituent components; this, naturally, affects their prediction capability. The Gladun algorithm,[10] which we use for predicting ternary and more complicated phases, appears more promising here. For reducing the number of searches, this algorithm uses a specially developed method of structuring computer memory in the form of growing pyramidal networks in which the lengths of the searched conjunctions are not reduced.

With the help of the Gladun algorithm we have predicted thousands of not-yet-synthesized compounds in ternary, quaternary, and more complicated systems.[11–27] Comparing our predictions with new experimental data shows that the average reliability of the prediction is about 90% (Table I).

Problems for Which the Cybernetic Approach Can Be Applied

The prediction of new compounds can be divided into three subproblems:

- Prediction of predefined compositions.
- Prediction of the crystal structure of compounds.
- Estimation of other properties (such as critical temperature of transition, superconducting state, melting or boiling points, formation heat, etc.).

Unless otherwise specified, the prediction is conducted for the physico-chemical systems kept at normal conditions, for example, the prediction of a phase at normal pressure and room temperature. In order to predict compounds that exist (e.g., high pressure), it is necessary to enter the data on known compounds that exist at high pressures into the "learning step" and add the pressure as a parameter into the description. Unfortunately, a majority of the experimental measurements of compound properties are conducted at incomplete equilibrium. In addition, the determination of the crystal structure of a given compound is not often undertaken in conjunction with its phase diagram investigation. Therefore, it is not always clear under what conditions a specific crystal modification of a compound (polymorphism) is stable. The standardization of the presentation of data for compound properties is a task for the future; meanwhile, in selecting examples for computer learning, we have to run a risk—which is not always rewarding.

The cybernetic approach based on computer learning for predicting new inorganic phases and estimating their properties is as follows:

1. Selection of the examples for computer learning and selection of the properties of the constituent components to be used for computer entry of these examples. The learning samples are given as input for the computer in the form of the matrix $M \times N$, where M is the number of examples for computer learning and N is the number of properties of the constituent components.

2. Computer learning consists of the construction of multidimensional criteria (classification rules).

3. Prediction of unknown phases is done by substitution into the multidimensional criteria using only properties of the constituent components of non-investigated systems.

For the selection of the properties of the constituent components, we provide some important examples:[15]

1. the distribution of electrons in the energy levels of separate atoms, the formal valences of elements in compounds (from electroneutrality considerations) and, for prediction of the type of crystal structure, the covalent or ionic (for oxide systems) radii of elements (cations in oxides);

2. the first four ionization potentials, the standard isobaric thermal capacities, the types of incomplete electronic shells, the covalent or ionic (for oxides) radii of the elements (cations), and the formal valence of the element in the compound;

3. the covalent (or ionic) radii, the enthalpies of the formation of simple oxides (sulphides, selenides, tellurides), the entropies of simple sulphides (selenides, tellurides) under standard conditions, or the thermal capacities of simple oxides and the formal valences of elements (in oxide compounds).

A fourth system of properties was used in predicting the availability and properties of complex chalcogenides:

4. the types of incomplete electronic shells, the electronegatives, the covalent radii, the formal valences exhibited in an AB_2X_4 compound, and the enthalpies of formation of corresponding simple sulphides (selenides, tellurides) under standard conditions.

Constraints of the Cybernetic Approach

The method of the cybernetic prediction of inorganic compounds offers the capability to predict new compounds with predefined properties. However, like any predictive technique, this method reflects the state of the art of the science of compounds and information science, and is therefore not free from defects. To improve the reliability of the predictions, we must describe the most important problems and suggest ways to solve them.

It is useful to bear in mind that this method of cybernetic prediction cannot predict a new physical phenomenon. The method is intended for the search of the physico-chemical systems similar to those already well-known.

The problems arising in making cybernetic predictions on the basis of the methods of computer learning are mainly as follows:

- Selection of the compounds for the "learning step" is the most difficult and least formalized task. The developed data bank containing ternary compound properties[28–31] does reduce the otherwise time-consuming search for information but it does not guarantee that these experimental values are correct. It must be stressed that incorrect experimental data, used in computer learning, constitute the major source of prediction errors. Often data indicating the absence of compounds may be incorrect. These occurrences lead to a decrease in the reliability for the prediction of the absence of compounds in other systems.
- Selection of the relevant properties of the constituent components for the description of the systems. The proper choice of the relevant properties is directly linked with the probability of the correct classification. Despite the fact that the Gladun algorithm is a powerful tool for the assessment of the informative combinations of the values of these properties, the "non-informative" combinations of the physico-chemical systems considerably weaken the results of the predictions. The procedures for the proper choice of the relevant component properties, addressed by Savitskii et al.[11] are:

a. An expert assessment of the proper-

ties using theoretical knowledge of modern physics and chemistry. This is not the best approach, since unwarranted weight is given to properties that the expert knows best. In addition, any property of the compounds, as a rule, contains a certain amount of information in only a narrow interval of change.

b. A computerized search, based on Gladun's algorithm, to find the relevant combinations of the given values of the properties of the components.

c. Automatic generation of a wide variety of algebraic functions from the initial component properties and selection of the most informative combinations for formulating criteria. The solution of the last problem requires high-performance computers with a great amount of storage.

■ From our experience the methods best suited for computer learning involve the use of logical expressions.[8–10] Their important feature is that they process only symbolic discrete values of the constituent component properties. On one hand, it allows the use of both numeric and symbolic properties and disregards absent (not-measured) values of properties in the description of the physico-chemical systems. On the other hand, unsatisfactory quantization of continuous constituent component properties (e.g., ionization potentials, radii, etc.) sharply weakens separability of the classes. To represent a continuous constituent component property by dividing it into a fixed number of intervals was often ineffective. To solve this problem we are developing a group of the sampling algorithms[32] specially as a subsystem of the preprocessing information for the computer learning system.[10]

Perspectives of Information-Predicting Systems

New perspectives of our studies in this field involve the development of an information-predicting computer system. This system is based on the cybernetic predicting subsystem, as discussed. In addition, this information-predicting system employs a data bank of inorganic com-

Table I: Characteristics of the Prediction Results for Ternary Phases

Compounds/Systems	Characteristics to be Predicted	Number of Predictions Tested for November 1, 1992	Reliability of Prediction	References
ABO_2	Possibility of formation of compounds	104	95	18
ABO_3	Possibility of formation of compounds	168	88	17
ABO_3	Perovskite structure type	115	77	17
ABO_4	Possibility of formation of compounds	125	98	17
AB_2O_4	Possibility of formation of compounds	130	95	17, 18
AB_2O_4	Spinel structure type	99	93	17, 19
$A_2B_2O_7$	Possibility of formation of compounds	83	77	20
$A_2B_2O_7$	Pyrochlore structure type	69	88	20
ABD_2 (D = Al, Co, Ni, Cu, Ga, Pd, In)	$MnCu_2Al$ structure type	45	91	21, 22
AB_2D_2 (D = Si, P, Ge, As, Sb)	$ThCr_2Si_2$ structure type	129	93	20–25
A(Hal)-B(Hal)$_3$	Systems with compounds	93	93	26
AB(Hal)$_3$	Possibility of formation of compounds	113	92	26
AB_2(Hal)$_4$	Possibility of formation of compounds	70	81	26
AB(Hal)$_4$ (Hal = F, Cl, Br, I)	Possibility of formation of compounds	195	93	27
A_2BX_2 (X = S, Se)	Possibility of formation of compounds	20	95	11
ABX_2	Possibility of formation of compounds	172	88	11, 28
ABX_3	Possibility of formation of compounds	55	91	11
AB_2X_4	Possibility of formation of compounds	350	81	11, 29, 30
AB_2X_4 (X = S, Se)	Structure types	165	94	11, 19, 30
$A_x(SO_4)_y \cdot B_zSO_4)_w$ $A(NO_3)_x \cdot B(NO_3)_y$	Possibility of formation of compounds	80	94	31

pounds properties (produced by the Institute of Metallurgy, Russian Academy of Sciences), a knowledge base, a conversational processor, and a monitor.[10,31,33] The database now contains information on about 35,000 ternary compounds taken from 9,000 publications.[28–31] The collection and evaluation of data on the properties of binary and quaternary compounds is in progress.

The Knowledge Base

The knowledge base stores the multidimensional criteria (classification rules) already obtained for various classes of inorganic compounds for their use in the prediction of phases and forecasting. The knowledge base data is represented in the form of semantic networks or of their equivalent conjunctive-disjunctive logical expressions.

Conversational Processor

The conversational processor manages the conversation of the user with the information-predicting system, as well as giving expert function support in the given application domain.

The Monitor

The monitor controls the computation process, provides the interface between the functional subsystems, and provides telecommunication access to the system. In addition, the monitor signals if new experimental data contradicts the predicted multidimensional criteria. Such contradictions will be eliminated by including the new data in the "learning step" and modifying the multidimensional criteria in the knowledge base.

Conclusions

The employment of the information-predicting system for the prediction of the formation of unknown inorganic compounds with predefined properties—with the help of multidimensional criteria and given the interdependence of some of these relationships—can be performed only by artificial intelligence systems. The development of such an information-predicting system opens new perspectives in the computerized design of substances with predefined properties.

References

1. W. Hume-Rothery in *Phase Stability in Metals and Alloys*, edited by P.S. Rudman, J. Stringer, and R.I. Jaffe (McGraw-Hill, New York, 1967).
2. F. Laves in *Phase Stability in Metals and Alloys*, edited by P.S. Rudman, J. Stringer, and R.I. Jaffe (McGraw-Hill, New York, 1967).
3. B.T. Mathias, *Phys. Rev.* **97** (1955) p. 74.
4. E.G. Fesenko, V.S. Filip'ev, and M.F. Kupriyanov, *Segnetoelektriki* (Rostov University, Rostov/Don, 1968) p. 63.
5. P. Villars, *J. Less-Common Met.* **119** (1986) p. 175.
6. L.S. Darken and R.W. Gurry, *Physical Chemistry of Metals* (McGraw-Hill, New York, 1953).
7. E.M. Savitskii, Y.V. Devingtal, and V.B. Gribulya, *Dok. Akad. Nauk SSSR [Sov. Phys. Dokl.]* **178** (1968) p. 79.
8. Y.V. Devingtal, *Izv. Akad. Nauk SSSR, Tech. Kibernetika* **1** (1968) p. 162.
9. Y.V. Devingtal, *Izv. Akad. Nauk SSSR, Tech. Kibernetika* **3** (1968) p. 139.
10. V.P. Gladun, *Heuristic Search in Complex Media* (Russian), (Naukova Dumka, Kiev, 1977).
11. E.M. Savitskii, V.B. Gribulya, N.N. Kiselyova, et al., *Prognozirovanie v Materialovedenii s Primeneniem EVM* (Nauka, Moscow, 1990).
12. N.N. Kiselyova and E.M. Savitskii, *Dok. Akad. Nauk SSSR [Sov. Phys. Dokl.]* **235** (1977) p. 1367.
13. N.N. Kiselyova, B.I. Pokrovskii, L.N. Komissarova, and N.D. Vashchenko, *Zh. Neorg. Khimii* **22** (1977) p. 883.
14. N.N. Kiselyova, V.I. Lutsyk, V.P. Vorob'eva, et al., *Dok. Akad. Nauk SSSR [Sov. Phys. Dokl.]* **304** (1989) p. 657.
15. E.M. Savitskii, V.B. Gribulya, and N.N. Kiselyova, *J. Less-Common Met.* **72** (1980) p. 307.
16. E.M. Savitskii, V.P. Gladun, and N.N. Kiselyova, *Dok. Akad. Nauk SSSR [Sov. Phys. Dokl.]* **233** (1977) p. 657.
17. N.N. Kiselyova, *Izv. Akad. Nauk SSSR, Metals* **2** (1987) p. 213.
18. N.N. Kiselyova and G.S. Burhanov, *Izv. Akad. Nauk SSSR, Metals* **1** (1989) p. 218.
19. N.N. Kiselyova and E.M. Savitskii, *Izv. Akad. Nauk SSSR Neorg. Mater., [Inorg. Mater. (USSR)]* **19** (1983) p. 489.
20. E.M. Savitskii and N.N. Kiselyova, *Izv. Akad. Nauk SSSR, Metals* **1** (1984) p. 191.
21. N.N. Kiselyova and G.S. Burhanov, *Izv. Akad. Nauk SSSR Neorg. Mater. [Inorg. Mater. (USSR)]* **23** (1987) p. 2006.
22. N.N. Kiselyova and E.M. Savitskii, *Zh. Neorg. Khimii* **29** (1984) p. 3104.
23. N.N. Kiselyova and E.M. Savitskii, *Zh. Neorg. Khimii* **27** (1982) p. 2880.
24. E.M. Savitskii and N.N. Kiselyova, *Izv. Akad. Nauk SSSR Neorg. Mater. [Inorg. Mater. (USSR)]* **15** (1979) p. 1101.
25. E.M. Savitskii, N.N. Kiselyova, and N.D. Vaschenko, *Dok. Akad. Nauk SSSR [Sov. Phys. Dokl.]* **239** (1978) p. 1154.
26. N.N. Kiselyova and E.M. Savitskii, *Zh. Neorg. Khimii* **24** (1979) p. 1427.
27. E.M. Savitskii and N.N. Kiselyova, *Dok. Akad. Nauk SSSR [Sov. Phys. Dokl.]* **272** (1983) p. 652.
28. E.M. Savitskii, N.N. Kiselyova, B.N. Pishchik, N.V. Kravchenko, and M.S. Golikova, *Dok. Akad. Nauk SSSR [Sov. Phys. Dokl.]* **279** (1984) p. 627.
29. M.S. Golikova, G.S. Burhanov, N.N. Kiselyova, et al., *Izv. Akad. Nauk SSSR Neorg. Mater. [Inorg. Mater. (USSR)]* **25** (1989) p. 700.
30. N.V. Kravchenko, G.S. Burhanov, N.N. Kiselyova, et al., *Izv. Akad. Nauk SSSR, Neorg. Mater. [Inorg. Mater. (USSR)* **27** (1991) p. 164.
31. N.N. Kiselyova and N.V. Kravchenko, *Zh. Neorg. Khimii* **37** (1992) p. 698.
32. O.Y. Markin, *Upravlyauyschie Systemi i Mashini* **2** (1988) p. 88.
33. N.N. Kiselyova, *Upravlyauyschie Systemi i Mashini* **5/6** (1992) p. 125. □