



PERGAMON

Engineering Applications of Artificial Intelligence 13 (2000) 497–505

Engineering Applications of

**ARTIFICIAL
INTELLIGENCE**

www.elsevier.com/locate/engappai

Interplay of large materials databases, semi-empirical methods, neuro-computing and first principle calculations for ternary compound former/nonformer prediction

P. Villars^{a,*}, K. Brandenburg^b, M. Berndt^b, S. LeClair^c, A. Jackson^c, Y.-H. Pao^d,
B. Igel'nik^d, M. Oxley^e, B. Bakshi^f, P. Chen^g, S. Iwata^h

^aMaterial Phases Data System (MPDS), CH-6354, Vitznau, Switzerland

^bCrystal Impact, Bonn, Germany

^cMaterials and Manufacturing Directorate, Air Force Research Laboratory, Wright-Patterson AFB, USA

^dCase Western Reserve University, Cleveland, OH, USA

^eAir Force Institute of Technology, Wright-Patterson AFB, USA

^fOhio State University, Columbus, OH USA

^gWright State University, Fairborn, OH USA

^hUniversity of Tokyo, RACE, Tokyo, Japan

Abstract

A significant breakthrough has been achieved using materials databases, semi-empirical methods and neural networks to aid in the design of new materials. A collaborative, international, team discovered that a non-linear expression involving one elemental property parameter could be used to predict, with 99+ % accuracy, the occurrence of a compound for any ternary materials system. This elemental property parameter, referred to as the Mendelev Number, was conceived by D.G. Pettifor in 1984 to group binary compounds by structure type. The near term significance of this discovery is the obvious savings, in time and resources, relative to assessing the merits of future, yet-to-be-realized, materials systems. In longer term this breakthrough is the basis for both narrowing the search space for potentially beneficial new materials and enabling the prediction of even more specific materials information such as stoichiometries, crystal structures and intrinsic properties. © 2000 Elsevier Science Ltd. All rights reserved.

Keywords: Materials design; Neural networks; Compound prediction

1. Introduction

Materials design, which today is still more a process of discovery than design, is principally an investigation based upon known concepts and the acquired intuition of experimentalists. Analyzing the conditions which empower successful designs of new materials suggest that they do not rely on a particular technique, a unique experimental observation, or an abstruse theory to be proven or disproven. On the contrary, materials design efforts critically depend

upon amassing large volumes of experimentally determined data that permits an individual with deep insight to perceive an underlying pattern not previously apparent. Extending and automating this data-driven process is leading toward an emerging new area of materials design based upon four enabling technologies:

1. The creation and the use of huge, critically evaluated materials databases which comprehensively and continuously assimilate the published world literature (Materials databases).
2. Computer-aided reduction of the elemental property parameters and systematic combinations of them to find the salient feature sets which can link elemental

* Corresponding author.

property parameter expressions with materials system properties (Semi-empirical methods).

3. Refinement and optimization of the obtained results, from two (2), via computational approaches for obtaining high-accuracy predictions (Neuro-computing).
4. Computational prediction of the most promising materials systems with the objective of reducing the experimental work required for its verification, and to create a theory-based explanation for such quantitative results (First principle calculation).

1.1. Materials databases

The amount of critically evaluated materials data is beginning to reach an acceptable volume, but is still far from being considered comprehensive. Listed below are the six most significant materials databases in the world (available in electronic form) list:

Binary alloy phase diagrams CD-ROM. This CD-ROM is maintained by ASM International (editor-in-chief: T.B. Massalski).

CRYSTMET. This Intermetallic Structure Database was maintained until 1 April 1997 by National Research Council of Canada NRCC, CISTI. Hard copy versions are Pearson's Handbook of Crystallographic Data for Intermetallic Phases, by P. Villars and L. Calvert, ASM International, 1991 and Pearson's Desk Edition, by P. Villars, ASM International, 1997 which contain crystallographic data for intermetallics and alloys.

ICDD PDF2. This Powder Diffraction Patterns Database is maintained by the International Centre for Diffraction Data in Newton Square, PA, USA and contains mainly measured powder patterns.

ICSD. This Inorganic Structure Database is maintained by the Fachinformationszentrum in Karlsruhe, Germany and contains crystallographic data for inorganic compounds.

LPF (Linus Pauling File). A basic database for alloys, intermetallics and inorganics. This file is under construction by the Japan Science and Technology Corporation JST in Tokyo and MPDS in Switzerland, and is planned to enter the yearly update stage in 2007 which covers structure, diffraction, property, constitution and bibliographic data (editor-in-chief: P. Villars).

Ternary alloy phase diagrams CD-ROM. This CD-ROM is maintained by ASM International (editor-in-chief: P. Villars).

1.2. Semi-empirical methods

There exists in the literature an extensive repository

of 'high quality' correlations between experimentally determined materials properties and the chemical elements present. These correlations often involve features, defined as a relation between parameters, each of which may be the value or function of one or more elemental property parameters. Aside from the usual variation in the experimental data and resulting correlations among alternative overlapping sources, they all share common methods, herein referred to as semi-empirical methods, for deriving materials property values based upon varying, small to large, amounts of published data. These common methods range from simple two feature graphs to n -dimensional data reduced to either a two- or three-dimensional representation. The reduced data representations are intended to aid in identifying correlations or consistent relationships among the features of a set of materials systems. A comprehensive review of the available sources identifying such correlations is given in Villars (1995).

Unfortunately, while semi-empirical methods have been the paradigm of choice in materials design, they are typically *of little value in the prediction* of new, yet-to-be-realized, materials with novel or unusual properties. Among the literally thousands of physico-chemical representations, or formalisms in the literature, possibly the most famous is Mendeleev's law — the so-called Periodic Table. While the Periodic Table has been organized and/or reorganized (Pettifor, 1984) in many ways by various scientific communities, critically important to acknowledge here is that Mendeleev's Periodic Table has made a significant impact on materials science. More specifically, it is the basis for grouping elements by distinguishing features (e.g., alkali metals, transition metals, semiconductors, noble gases, etc.). This impact is due, in part, to the volume of information that is concisely captured in this two-dimensional (2D) representation. A testament to the value of such pictorial representations is the work of several other contributors in the scientific community that have utilized 2D methods such as crystal-structure maps, (Pettifor, 1984) and materials-selection-charts (Ashby, 1982) to emphasize features and clustering of the elements and/or compounds.

Within the context of materials design methods, Kiselyova (1993) has argued that, in principle, there are three ways to predict the existence of new inorganic compounds, and further, to forecast their intrinsic compound properties, based on a knowledge of their constituent component property parameters:

- quantum-mechanical calculations,
- two-dimensional criteria (classification rules) found by semi-empirical methods, and,
- multi-dimensional criteria found by computer learning techniques.

It is noteworthy that many, nearly all, semi-empirical methods involve 2D representations. This can be explained by a propensity to seek ‘linear’ relationships among features, which can be extended to 3D by involving a third dimension. To date, 3D representations have not been used because of the difficulty in perceiving a pattern as the density of points increases. To address this problem several researchers, including Pao(1998) and Chen (1998), exploit electronic representations as a means to enhance visualization and insight, while enabling a more dynamic, user-driven repository for depicting data. Of interest is the development of a fully automated discovery space to search for both salient features (combinations of elemental properties) and the identity of consistent relations, in either 2D or 3D feature sets which correlate with published, experimentally determined, intrinsic materials system properties.

1.3. Neuro-computing

Our neuro-computing work relates to the third approach, namely the learning of a functional model in multi-dimensional space of how materials property depends on features or equivalently, on the position of a system in that multi-dimensional space.

We know the task is very difficult when very large numbers of data items are involved and when the number of features for each data point may also be high.

Prior to using the Mendeleev Number as the sole feature of an element, we had followed a two step neuro-computing approach towards the building of the equivalent of ‘structure maps’. Initially five features were used to characterize each element. Procedures based on proximity in such pattern spaces could indeed predict, with about 90% accuracy, whether a newly proposed system would be former or nonformer. When local feedforward neural-nets were used in addition to self-organization, accuracy of over 99.5% was achieved. In other words, in that 15-dimensional space it was not proximity alone which determined whether a new system would be former or nonformer, the distribution of the data points was also important.

It was found that the Mendeleev Number could serve as a highly effective feature, more so, for example, than the Atomic Number. Nevertheless, the highest accuracy figures are obtained when Mendeleev Number is used in combination with some other features. The problem of multi-dimensional complexity is still with us and our neural net approach is still valid and effective. But all simplifications are welcome.

The neuro-computing approach as we have developed, it corresponds to an a priori approach where we proceed in an orderly manner to analyze data so as to enable inductive inference. The happy discovery of the

extraordinary effectiveness of the Mendeleev Number corresponds to the exploitation of an ‘a posteriori’ ordering. From a computational point of view, the question is how can we learn the next ‘a posteriori’ feature which is just right for the classification or estimation task being considered. In the meantime, we use both approaches, in combination. However, progress is also being made on the ‘a posteriori’ aspects of the approach, so that singularly effective features can be identified.

2. First principle calculations

Using semi-empirical methods, it is clear that the number of correlations needed to encompass 3+ element materials systems will be prohibitively expensive to discover. Consider, for example, the number of correlations one would expect to find among the potentially 161,700 ternary and 3,921,225 quaternary compounds. In addition, in order to establish structure and phase relationships, one has to prepare and investigate at least 10 times more samples per system for each additional element, e.g., from ternary to quaternary. Therefore, a need exists to focus first principle calculations towards selected groups of the most promising compound classes before starting with experimental verification. In the best case, one can achieve a theory-based explanation for such quantitative results from first principle calculations, which would be essential to find the direct processing parameters to optimize its production.

3. Ternary compound former/nonformer prediction from the materials databases and semi-empirical methods point of view

In this paper, we demonstrate the beneficial interplay of materials databases, semi-empirical methods and neuro-computing on the example of ternary materials information distinguishing between compound forming and nonforming systems.

3.1. Materials databases

In this work we used materials system data from the following databases:

- PEARSON’s DESK EDITION (electronic version)
- BINARY ALLOY PHASE DIAGRAMS CD-ROM
- TERNARY ALLOY PHASE DIAGRAMS CD-ROM

and we found published information (experimentally determined) concerning formers/nonformers for 6,998 distinct ternary systems, from which we prepared the following three data sets:

- Set 1** 455 test-data (383 nonformers/72formers) from TERNARY ALLOY PHASE DIAGRAMS CD-ROM and PEARSON'S DESK EDITION
- Set 2** 2279 training-data, all *nonformers* derived from 666 binary nonformers contained in the BINARY ALLOY PHASE DIAGRAMS CD-ROM
- Set 3** 4264 training-data, all *formers* from PEARSON'S DESK EDITION of systems having at least one ternary compound with a ternary structure type

We used the following important definitions to establish the data sets.

- Set 1** The difference between a former and a nonformer is that the compound-forming system possesses at least one ternary compound separated by three two-phase regions involving three adjacent binary and/or ternary compounds. In each case, when no phase diagram was known for a ternary crystal structure, the A-B-C compound was accepted as a compound former. For those systems where contradictory information existed between phase diagram and crystal structure data — the material system was not used.
- Set 2** Based upon the criteria that a ternary system is a nonformer when all its three binary boundary systems are also nonformers, we derived 2279 ternary nonformers based on the existence of 666 published binary nonformer phase diagrams.
- Set 3** Based upon the criteria that a ternary system is a former when at least one ternary compound with a ternary crystal structure is published, we found 4264 ternary formers. Note this simple selection criteria excludes the inclusion of pseudo-ternary compounds which are solid solutions of binary compounds.

3.2. Semi-empirical methods

When considering factors that govern compound formation, it is desirable to identify regularities, comprised of elemental property parameters of the constituent chemical elements, so that the existing experimental data (former/nonformer) can be efficiently systematized. Generally, we consider regularities comprised of a minimum, but most salient, number of elemental property parameters that have real predictive value to systematize a large group of data with an accuracy in the range of >99%. In predicting compound formation for not yet experimen-

tally investigated systems, perhaps the opposite is more the case, i.e., the more data considered, the more trustworthy is a prediction based on such regularities.

To make predictions, one starts from the chemical elements and the tabulated elemental property parameters of the chemical elements, and calculates the elemental property expressions for a system of interest. Therefore, elemental property parameters are only of practical interest where they are known for most chemical elements with adequate accuracy. For example, from the regularity and its constituent elemental property parameter expressions to be discussed, one can predict whether a yet-to-be-realized materials system will form a compound (with an accuracy of >99%).

In principle, it should be sufficient to use only the Atomic Numbers as elemental property parameter. Slater (1956) once made the following comment: "I don't understand why you metallurgists are so busy in working out experimentally the constitution (crystal structure and phase diagram) of multinary systems. We know the structure of the atoms (needing only the Atomic Number), we have the laws of quantum mechanics, and we have electronic calculation machines, which can solve the pertinent equation rather quickly."

Some 35 years later Chelikowsky (1991) wrote in an excellent review, the following: "Although the interactions in intermetallic compounds are well understood, it is not an easy task to evaluate the total energy of solids. Even at absolute zero, the energy of an isolated atom is in the order of about 10^6 eV, while the cohesive energy is only of the order of 1–10 eV/atom. Thus, one must have a method that is accurate to one part in one million, or better." The other fact that greatly complicates evaluating the cohesive energy by theoretical methods is the number of particles involved. Given that a macroscopic solid may contain 10^{23} nuclei and electrons, it is impossible to determine the total energy of the crystal structure without some approximations.

Within the last 20 years, two advances have made it possible to predict the cohesive energy of solids by numerical solution of the quantum-mechanical equations of motion (e.g., Schrödinger's equation), (1) the invention of high-speed computers and (2) density functional theories, which greatly simplify calculations involving many-electron interactions. The accuracy of these computations is impressive, but still not at the same level as the most careful experimental measurements. Nonetheless, it is now possible for chemical elements and simple intermetallic compounds to predict the relative stability of given crystal structures. The complexity of the above-discussed problem illustrates that one cannot, within the next decade, expect that the constitution (crystal structure and phase dia-

Table 1

The 73 elemental property parameters used in this work grouped according to the six (actually seven) factors

Size factor

Radii pseudo-potential Zunger (a.u.)
 Radii ionic Yagoda (nm)
 Radii covalent (nm)
 Radii metal Waber (nm)
 Distance valence electron Schubert (nm)
 Distance core electron Schubert (nm)
 Volume atom Villars-Daams (10^{-6}nm^3)
 Volume $V^{2/3}$ Miedema (cm^2)
 Number atomic environment Villars-Daams

Heat (Cohesion-energy) factor

Temperature melting (K)
 Temperature boiling (K)
 Enthalpy vaporization (kJ mol^{-1})
 Enthalpy melting (kJ mol^{-1})
 Enthalpy atomization (kJ mol^{-1})
 Enthalpy surface Miedema (kJ mol^{-1})
 Enthalpy vacancies Miedema (kJ mol^{-1})
 Energy cohesive Brewer (kJ mol^{-1})
 Modulus compression (GPa)
 Modulus bulk (GPa)
 Modulus rigidity (GPa)
 Modulus Young (GPa)

Electro-chemical factor

Electronegativity Martynov–Batsanov
 Electronegativity Pauling
 Electronegativity Alfred–Rochow
 Electronegativity absolute
 Energy ionization first (kJ mol^{-1})
 Energy ionization second (kJ mol^{-1})
 Energy ionization third (kJ mol^{-1})
 Potential chemical Miedema (a.u.)
 Work function (eV)
 $n^{W^{1/3}}$ Miedema (a.u. $^{-1/3}$)

Group Number (valence electron) factor

Number valence electron
 Number group

Atomic Number factor

Number Periodic Table start counting left top, left \rightarrow right sequence = number atomic
 Number Periodic Table start counting right top, right \rightarrow left sequence
 Number Periodic Table start counting left down, left \rightarrow right sequence
 Number Periodic Table start counting right down, right \rightarrow left sequence
 Number quantum
 Weight atomic (10^{-3} kg)
 Charge nuclear effective Clementi
 Charge nuclear effective Slater
 Coefficient mass attenuation for MoK α (cm^2g^{-1})
 Coefficient mass attenuation for CrK α (cm^2g^{-1})
 Coefficient mass attenuation for CuK α (cm^2g^{-1})
 Coefficient mass attenuation for FeK α (cm^2g^{-1})
 Factor atomic electron scattering at 0.5

Mendeleev Number factor

Number Periodic Table start counting left top, top \rightarrow down sequence = number Mendeleev
 Number Periodic Table start counting right top, top \rightarrow down sequence
 Number Periodic Table start counting left down, down \rightarrow top sequence
 Number Periodic Table start counting right down, down \rightarrow top sequence
 Number Mendeleev Pettifor sequence
 Number Mendeleev chemists' sequence

Factor with irregular behaviour as a function of the atomic number

Magnetic resonance (MHz)
 Magnetic frequency of nuclei
 Magnetic susceptibility (m^3kg^{-1})

(continued on next page)

Table 1 (continued)

Spin nuclei (h)
Density (kg m ³)
Resistivity electrical 298K (10 ⁻⁸ ohm m)
Conductivity electrical at 298K (10 ⁴ ohm ⁻¹ cm ⁻¹)
Conductivity thermal (J mol ⁻¹ K ⁻¹)
Velocity sound (m s ⁻¹)
Heat capacity molar (J mol ⁻¹ K ⁻¹)
Entropy of solid at 298 K, 100 MPa (J mol ⁻¹ K ⁻¹)
Electrochemical weight equivalent (mg K ⁻¹)
Gamma free electron to specific heat (mJ mol ⁻¹ K ⁻²)
Temperature Debye (K)
Thermal neutron capture cross section (bar/ns)
Hardness Brinell
Linear thermal expansion coefficient (10 ⁻⁶ K ⁻¹)
Poisson's ratio
Spectral lines no
Vapour pressure at melting temperature (N m ⁻²)
Surface tension at melting temperature (mJ m ⁻²)

gram) of multinary systems will be calculated from first principles. Based upon this awareness, it is sensible to adopt, in parallel, semi-empirical methods based on the experimentally determined data to search for the most reliable regularities.

It has been learned that it is impossible to start only with the Atomic Number, and thus one must attempt to find which other elemental property parameters of the chemical elements are needed to effect compound formation. In this work we conducted a survey of 73 different elemental property parameters as a function of the Atomic Number, and it was found that there exists six main groups, called factors, which are listed below:

Size factor
 Heat (cohesion-energy) factor
 Electro-chemical factor
 Group number (valence electron) factor
 Atomic Number factor
 Mendeleev Number factor

In Table 1 are the 73 elemental property parameters grouped according to the six (actually seven) factors. Taking these 73 elemental property parameters and five mathematical operators (*, /, -, +, maximum) into account, there are $5 * 73 = 365$ combinations (elemental property parameters expressions = features) resulting in $(365 * 364 * 363) / (3 * 2) = 8,038,030$ 3D feature sets, assuming the best separation is achieved using three different features. Because of the high number of to-be-investigated 3D feature sets to find the best separation between the 6988 published (experimentally determined) former/nonformer datasets we have developed a program called DISCOVERY based upon the following general idea.

We divided the DISCOVERY code into three tasks involving the development of:

1. An automatic generator for 3D feature sets resulting from combinations of elemental property parameters and mathematical operations (8,038,030 3D feature sets),
2. A program for the detection of those 3D feature sets with the "best" separation of the material property, distinguishing formers from nonformers,
3. A sophisticated 3D Graphics program for the interactive investigation of the best 3D feature sets.

The central problem is deciding what is a "good" separation and what is not in 3D feature space. What humans normally do, using their visual abilities, is to find some kind of border between areas and then counting how many points, with the same property, are on the same side of this border. However, finding such a border is a very complex task for a program. In 1D finding a border (point) which separates "left" and "right" points is relatively easy. In 2D you have to find a more or less complex border (line) between the areas and this is already a complex optimization problem. But in 3D you would have to find a more or less complex surface (one or more planes), and though algorithms to find them exist (e.g. neuro-computing), they would take too much calculation time to be applied on 8,038,030 3D feature sets.

So the detection algorithm must work on a much simpler basis. We must be aware that this strategy leads to different definitions of what is the quality of separation within one 3D feature set. However, our starting point is to identify that a separation is "good", then assess many points, involving nearest neighbor(s), to determine if they have the same system property. This can be detected with simple distance calculations. In principle, all distances from each point must be calculated to find the shortest distance, and

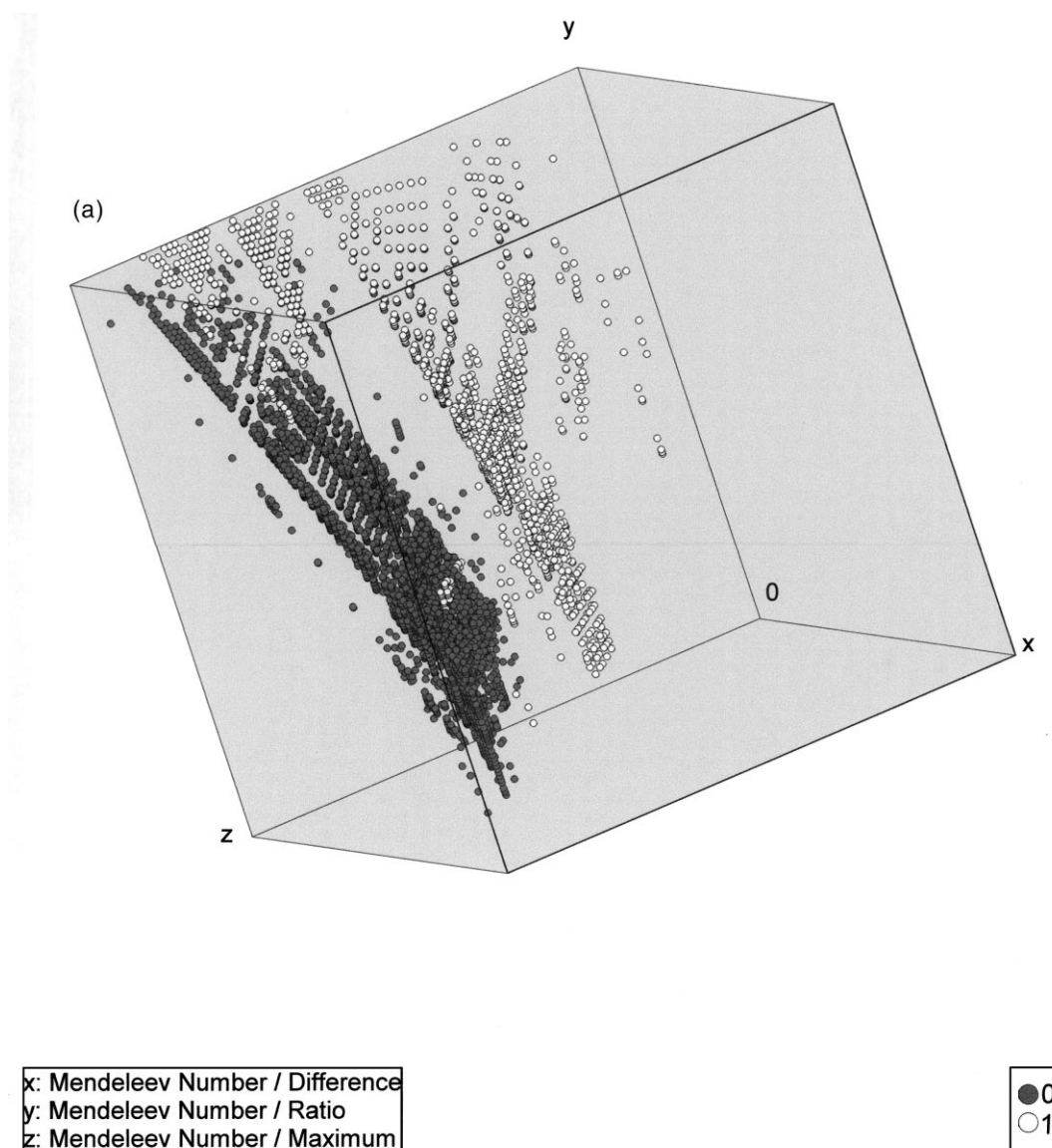


Fig. 1. (a) The 6,998 distinct different, experimentally determined, ternary systems plotted in the ‘Discovery Space’ (3D feature space). The red dots (0) are the ternary nonformers, the blue dots (1) are the ternary formers. (b) Distribution of the the potentially 161,700 ternary systems plotted in the ‘Discovery Space’ (3D feature space) x: Mendeleev number difference; y: Mendeleev number ratio; z: Mendeleev number maximum. The brown dots (13) represent each a potential ternary system, indicating the outline of a simple hyperplane

therefore, the next nearest neighbor. Given 6,998 points, the number of calculations required is $6,998 * 6,997 / 2 = 24,482,503$ distances — for each 3D feature set. We were able to reduce this number to approximately 100,000 calculations for a rough analysis and approximately one million calculations for a more precise analysis. Regardless of the coarseness of the analysis, the approach is limited by uncertainties — some nearest neighbors were not be found, and therefore, the analysis error was about 0.1% (for the more precise analysis).

Given this ‘distance’ approach, what does it mean to have “Separation of 99%”? It means that 99% of all points have a nearest neighbor with the same system

property. So what about the “other” neighbors? A really good separation means that points of the same property yield clusters of points, which are as big as possible. It is no surprise that the number of “hits” (nearest neighbors with the same system property) decreases with an increasing number of nearest neighbors because even large clusters will come to an end sooner or later. This indicates the real quality of separation, because a good separation starts with a high number of hits and decreases as slowly as possible.

Considering the 73 elemental property parameters, we list below the best separations taking into account the separation of the nearest neighbor and the separation of the nearest 50 neighbors ranking elemental

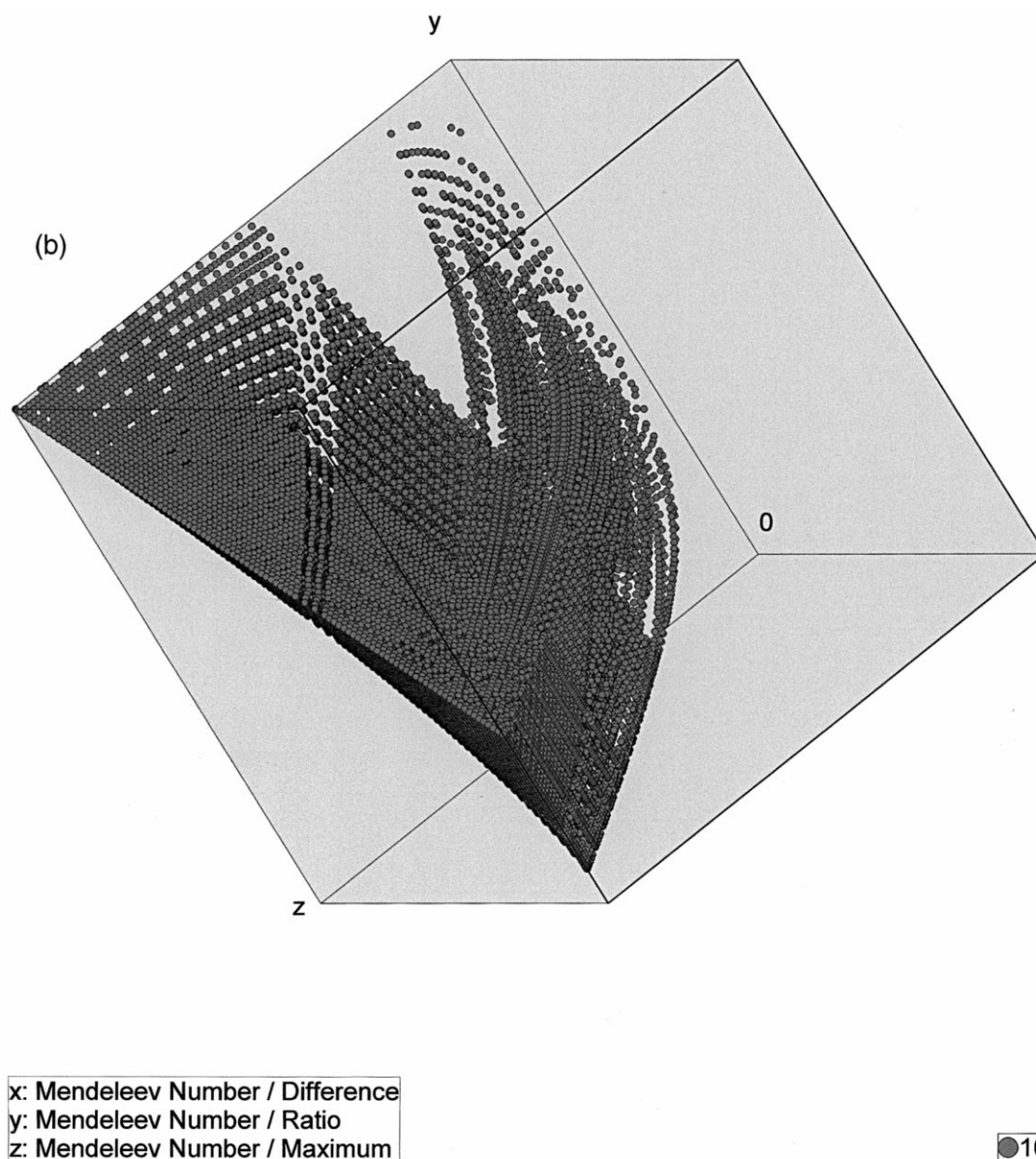


Fig. 1 (continued)

property parameters which have some relevance concerning our former/nonformer separation:

- No. 1:** Mendeleev Number factor
99(1) \Rightarrow 86(50)%
- No. 2:** Electro-chemical factor
96(1) \Rightarrow 71(50)%
- No. 3:** Size factor (only Zunger's pseudo-potential radii)
97(1) \Rightarrow 67(50)%
- No. 4:** Group Number (Valence Electron) factor
82(1) \Rightarrow 29(50)%
- No. 5:** Heat (Cohesion Energy) factor
95(1) \Rightarrow 27(50)%

- No. 6:** Atomic Number factor
91(1) \Rightarrow 25(50)%

It can be seen, that for each factor the best separation taking just one nearest neighbor into account is still quite good, but its separation power taking its 50 nearest neighbors into account decreases quickly, especially for the Atomic Number, Heat, and Group Number(valence electron) factors.

In conclusion, the ternary former/nonformer decision is strongly dominated by the Mendeleev Number of its chemical elements. The mathematical operators +, -, /, *, maximum are all about equally adequate. Concerning the correlation of the Mendeleev

Number and the Atomic Number there is a direct non-linear correlation looking at the Periodic Table: “One can string (in sequence) the chemical elements of the Periodic Table in eight different ways. One of the results of counting is the Atomic Number itself (start counting left top, left → right sequence), another the Mendeleev Number itself (start counting left top, top → down sequence). The Atomic Number stresses the Periods of the Periodic Table and the Mendeleev Number emphasizes the Groups of the Periodic Table.

The best results are achieved by using the Mendeleev Number using “Periodic Table number start counting left down, down → top sequence” with:

x: Mendeleev Number Difference vs.
y: Mendeleev Number Ratio vs.
z: Mendeleev Number Maximum

with 99.37% with nearest neighbour and 86.76% with 50 nearest neighbors with same materials property formers nonformers (see Fig. 1a).

This best result represents the most trustworthy 3D feature set as it is optimal in respect of:

- Using as few as possible elemental property parameters (just Mendeleev Number).
- Mathematical operators as simple as possible.
- Boundary surface between former/nonformer is simple in the ‘Discovery Space’ (3D feature space), see Fig. 1b.
- Distribution within the ‘Discovery Space’ of the available experimentally known data in respect to all potential data are ‘evenly’ spread and all potential data are distributed on a simple, well described hyperplane within the 3D feature space.

The separation can be slightly improved by replacing one of the features by a feature belonging to the electro-chemical factor, such as: Maximum of electronegativity (Martynov–Batsanov), ratio or maximum of electronegativity (Pauling), or maximum of third ionization potential.

4. Conclusions

Important for optimization of the method described here are the following important issues to be watched, as they might influence the results:

1. The quality of the starting data (materials property) which is to be correlated to the elemental property parameter(s) has to be as adequate as possible, otherwise one starts with a ‘too high’ noise. To select an ‘error-free starting data-set’ requires a very careful evaluation of the published, experimentally

determined data, by comparing crystal data and phase diagram data of each chemical system.

2. Include many known elemental property parameters, which are the starting point to derive the potential features.
3. Minimize the number of prediction steps by covering large materials groups (e.g. treat binary, ternary, quaternary systems together instead of simply treating binaries alone).
4. Investigate carefully the distribution of all potential systems in your 3D feature space, with respect to the experimentally known systems, and take this into account for deciding which is the ‘best’ solution.
5. Maximize the accuracy of the prediction of each prediction step to over 99% by using several different approaches and comparing prediction results (e.g. different neuro-computing approaches).
6. As already mentioned in the introduction to this publication, in principle, all materials properties should be derivable, starting from the Atomic Numbers of the constituent chemical elements. Our approach showed that the Atomic Number gives rather a limited separation in the 3D feature space, but an excellent separation is achieved using the Mendeleev Number.

Finally, it should be noted that the correlation between the Atomic Number and the Mendeleev Number is non-linear. We also hope that first principle calculations will reveal a theory-based understanding of this discovery.

References

- Villars, P., 1995. Factors governing crystal structures. In: Westbrook, J.H., Fleischer, R.L. (Eds.), *Intermetallic Compounds, Principles and Practice*, vol. 1. Wiley, New York (Chapter 11).
- Pettifor, D.G., 1984. A chemical scale for crystal-structure maps. *Journal of Solid State Communications* 51 (1), 31.
- Ashby, M.F., 1992. *Materials Selection in Mechanical Design*. Pergamon Press, New York.
- Kiselyova, N.N., 1993. Prediction of inorganic compounds: experiences and perspectives. *MRS Bulletin* 18 (2), 40.
- Pao, Y.H., et al., 1998. Materials discovery via topologically-correct display of reduced-dimension data. *Journal of Alloys and Compounds* 279 (1), 22.
- Chen, N., et al., 1998. KDPAG expert system applied to materials design and manufacture. *Engineering Applications of Artificial Intelligence Journal* 11 (5), 669.
- Slater, J.C., 1956. *Theory of Alloy Phases*. American Society of Metals, Cleveland, OH.
- Chelikowski, J.R., 1991. *Encyclopedia of Applied Physics*, vol. 3. VCH, Weinheim, p. 1.