ELSEVIER

Journal of Alloys and Compounds 317–318 (2001) 26–38

Journal of ALLOYS AND COMPOUNDS

www.elsevier.com/locate/jallcom

Binary, ternary and quaternary compound former/nonformer prediction via Mendeleev number

P. Villars[a,*], K. Brandenburg[b], M. Berndt[b], S. LeClair[c], A. Jackson[c], Y.-H. Pao[d], B. Igelnik[d], M. Oxley[e], B. Bakshi[f], P. Chen[g], S. Iwata[h]

[a]*Material Phases Data System (MPDS), CH-6354 Vitznau, Switzerland*
[b]*Crystal Impact, Bonn, Germany*
[c]*Materials and Manufacturing Directorate, Air Force Research Laboratory, Wright-Patterson AFB, USA*
[d]*Case Western Reserve University, Cleveland, OH, USA*
[e]*Air Force Institute of Technology, Wright-Patterson AFB, USA*
[f]*Ohio State University, Columbus, OH, USA*
[g]*Wright State University, Fairborn, OH, USA*
[h]*University of Tokyo, RACE, Tokyo, Japan*

Abstract

A significant breakthrough has been achieved in the design of new materials by using materials databases, semiempirical approaches and neural networks. It was found in the present work that a nonlinear expression involving one elemental property parameter can be used to predict, with an overall accuracy exceeding 99%, the occurrence of a compound for any binary, ternary or quaternary system. This elemental property parameter, referred to as the Mendeleev number, was conceived by D.G. Pettifor in 1983 to group binary compounds by crystal structures. The immediate profit of this discovery is the obvious savings, in time and resources, relative to the investigation of yet-to-be-studied, materials systems. In the longer term the relation found here will make it possible to better define the search space for the development of new materials and encourage attempts to predict more specific information such as stoichiometries, crystal structures and physical properties.

Keywords: Materials design; Neural networks; Compound prediction; Materials databases

1. Introduction

Materials design is at present mainly based on a certain number of known concepts and the intuition of the experimenters. An analysis of the conditions that made it possible to discover the concepts known in materials science shows that it was not a new technique, a unique experimental observation, or an abstruse theory that formed the take-off point. It was rather the amassing of a large volume of experimentally determined data that permitted individuals with deep insight to perceive an underlying, not previously apparent pattern (see [1] for a review).

Extending the concept of pattern recognition to the area of materials design relies on the following three key-points

- The creation and use of large, critically evaluated materials databases, which comprehensively cover the world literature (materials databases).
- Computer-aided reduction of the elemental property parameters and systematic combinations of them to find salient features sets, which can qualitatively/quantitatively link materials properties with the chemical elements present (semiempirical approaches).
- Refinement of the obtained results with the help of neurocomputing to obtain optimized quantitative results (neurocomputing).

To increase the efficiency of the search for new compounds, major efforts should go towards creating an internationally accessible information–knowledge system incorporating all experimentally determined values, generally valid principles, and 'highest-quality' regularities, but also pattern recognition methods, such as e.g. neural networks.

*Corresponding author.

0925-8388/01/$ – see front matter
PII: S0925-8388(00)01410-9

The combination of the experience and intuition of the experimenter and the easy access to such an information–knowledge system would very much help to coordinate world research activities. Furthermore, it would reduce the number of unwanted duplications, as well as increase the probability of investigating directly the most promising chemical systems.

We make here the following postulate: materials properties are quantitatively contained in elemental property parameters of its constituent chemical elements. If this postulate is correct one should be able to deduce quantitative dependences from a sufficiently large set of high-quality materials property data (known from experiments), under the condition to have access to complete, high-accuracy data-sets for different elemental property parameters.

Fundamental materials properties considered in this context are for example

- compound formation within a given chemical system (binary, ternary, quaternary)
- stoichiometry of stable compounds within a compound-forming system
- crystal structure of a given compound
- melting point of a given compound

As elemental property parameters we use for example

- atomic number of the chemical elements
- group number of the chemical elements
- Mendeleev number of the chemical elements
- pseudo-potential radii of the chemical elements
- first ionization energy of the chemical elements
- melting point of the chemical elements

An elemental property parameter is of practical interest only if numerical values are known for all, or most chemical elements. In addition, the higher the accuracy of these values is, the better. As an example, the atomic number is known with 100% accuracy, in contrast to the melting points for which some chemical elements show inaccuracies in the range of 2–5%. The elemental property parameter(s) of the constituent chemical elements are combined by means of mathematical operators, such as addition, subtraction or multiplication.

An optimal solution would use as few as possible different elemental property parameters, require no or little subdivision of the materials property data-set, use simple mathematical expressions to link the elemental property parameters of the constituent chemical elements.

2. Materials databases

Most important for the successful discovery of such correlations is the comprehensiveness and especially the quality of the data (critically evaluated experimental facts), as they represent the starting point. In this work we used the below listed five materials databases. To increase the quality of our data-sets we checked the consistency between the different materials databases, especially the consistency between crystal structure data information and phase diagram information. Chemical systems where we found contradictions were excluded from our work.

2.1. ICSD [2]

This Inorganic Crystal Structure Database is maintained by the Fachinformationszentrum in Karlsruhe, Germany and contains crystallographic data for Inorganic Compounds.

2.2. Pearson's Handbooks [3,4]

Electronic version available by MPDS, its hard-copy versions are Pearson's Handbook of Crystallographic Data for Intermetallic Phases, second edition, ASM International, 1991 and Pearson's Desk Edition, ASM International, 1997 which contain crystallographic data for intermetallics and alloys.

2.3. Binary alloy phase diagrams CD-ROM [5]

This CD-ROM is maintained by ASM International.

2.4. Ternary alloy phase diagrams CD-ROM [6]

This CD-ROM is maintained by ASM International.

2.5. LPF, Linus Pauling file [7]

One of our ongoing efforts to create comprehensive materials databases is represented by our most recent database creation activity [7], the LPF project, which is shortly outlined below.

The LPF consists of a data part, as well as a smart software part. The data part covers all non-organic (alloys, intermetallics, inorganics, ceramics and minerals) ordered solid state materials (systems) and consists of structure, diffraction, constitution, physical (intrinsic) property and bibliographic information. To have these four groups of materials property data as numerical, factual and image data under the same computer environment is world unique. Experimentally determined and calculated data published in over 100 000 relevant publications will be included, which will result in at least 200 000 structure, diffraction and physical (intrinsic) property data entries as well as about 35 000 constitution data entries (image) covering the world literature from 1900 to date. The LPF data are processed by an international group of highly experienced editors, assisted by an evaluation and derived-data creation software package containing over 100 differ-

ent modules, ESDD [11]. The as-published data are accompanied by value-added information, such as calculated powder patterns (Lazy Pulverix [8]) and fully standardized structure data (Structure Tidy [9], Compare [10]) etc.). All data are extracted from the original publications, using as starting point the references of the following materials databases: JOIS-F Crystal Structure Data [12] (Japan Science and Technology, Tokyo), parts of Crystmet [13] (Materials Phases Data System (MPDS) formerly CISTI/NRCC), Pearson's Handbooks [3,4], Structure Type Atlas [14], Typix [15] and Ternary Alloy Phase Diagram Handbook [6]. The LPF project is a long-term project mainly funded by Japan Science and Technology, Tokyo (JST) and is a collaboration between JST, MPDS and the University of Tokyo (RACE). It started in 1996 and is planned to enter the yearly update stage in 2007. The LPF will be available worldwide through online (internet) as well as offline (CD-ROM/DVD/hard-copy) products.

One has the impression by reviewing the world literature in materials science, that materials scientists have experimentally investigated a major part of all possible binary, ternary and quaternary systems. This is by no means the case, not even for the binary systems, and the situation gets worse by looking at the ternary and quaternary systems. Below we made some estimates based on the available crystallographic and phase diagram compilations

- Binary:<50%
- Ternary: <5%
- Quaternary: <0.5%

What is not seen in those numbers is that in most systems counted above as 'investigated', are only partially investigated. The positive part of this is that nature has for our society still an enormous basin of yet not investigated systems/compounds available.

From the five above-listed material databases we got the following numbers of distinct binary, ternary and quaternary systems former, respectively nonformer information

- 2016 distinct binary systems
- 6382 distinct ternary systems
- 7021 distinct quaternary systems

These data were divided in the following data-sets

- binary formers, 1382 systems (B1) from Pearson's desk edition [4] of systems having at least one binary compound with a binary structure type
- binary nonformers, 634 systems (B2) from the binary alloy phase diagrams CD-ROM [5]
- binary training set, 1327 systems (B3)
- binary test set, 689 systems (B4)
- ternary formers, 4264 systems (T1) from Pearson's desk edition [4] of systems having at least one ternary compound with a ternary structure type
- ternary nonformers, 2118 systems (T2) derived from 634 binary nonformers contained in the binary alloy phase diagrams CD-ROM [5]
- ternary training set, 4240 systems (T3)
- ternary test set, 2142 systems (T4)
- quaternary formers, 2747 systems (Q1) from the ICSD file [2]
- quaternary nonformers, 4274 systems (Q2) derived from 634 binary nonformers contained in the binary alloy phase diagrams CD-ROM [5]
- quaternary training set, 4668 systems (Q3)
- quaternary test set, 2353 systems (Q4)

Below we have given the three different definitions which we used to decide whether a binary, ternary and quaternary system belongs to a former or nonformer system. These three definitions (here explicitly explained for the ternary case) for separating formers from nonformers are based on the following fundamentals

- Description of crystal structure within the frame of space group theory
- Gibb's phase rule

Definition 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 chemical element(s) and/or binary compound(s) and/or ternary compound(s). In each case, where no phase diagram was known for a A–B–C system, but a ternary compound with a ternary crystal structure is known, the A–B–C compound was accepted as a compound former.

Definition 2: Based upon the criteria that a ternary system is a nonformer when all its three binary boundary systems are also nonformers, we derived 2118 ternary and 4274 quaternary nonformers based on the existence of 634 published binary nonformer phase diagrams [24]. This criteria has been confirmed, so far, by all published ternary isothermal sections.

Definition 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.

A very crucial point in our work was to separate former from nonformer by making a 'clear cut' separation between a 'real' ternary compound and a 'pseudoternary' compound which is a ternary solid solution of a binary boundary compound of that specific ternary system retaining its binary structure type. With our three definitions we were able to establish a clear assignment for all chemical systems to former or nonformer. For the group of former very efficient in doing so was to focus on binary, ternary and quaternary structure types, as it is general practice in

all Pearson's handbooks [3,4]. For the group of nonformer the criteria that a ternary, respectively a quaternary system is a nonformer when its 3, respectively 6 binary boundary systems are experimentally established nonformers was most helpful. It showed that to establish ternary and quaternary systems as nonformers based on experiments requires several very well established investigated isothermal sections with many investigated compositions. In very few cases nonformers have been experimentally well established as most scientists are not too much interested in nonformers, they rather search for formers. In addition it showed that for many of the not so thoroughly investigated nonformer systems there exists contradictions with the published crystal structure data.

3. Semiempirical approaches

Within the context of materials design methods there are three ways to predict the existence of new compounds based on the knowledge of their constituent elemental property parameters

- two(three)-dimensional criteria (classification rules) found by semiempirical approaches [1]
- multidimensional criteria found by computer learning techniques [16] (e.g. neural networks)
- quantum-mechanical calculations [17,18].

To prove the correctness of our postulate that 'Materials properties are quantitatively contained in elemental properties of its constituent chemical elements', we focused on the fundamental problem to separate formers from nonformers based on the published binary, ternary and quaternary experimentally determined facts and to use in addition only elemental property parameters of its constituent chemical elements. To tackle this problem it seemed most adequate to combine semiempirical approaches and neural networks. First principle calculations seemed not to be adequate looking at the high number of to be considered experimental investigated systems (15 000).

It is noteworthy that many, nearly all, semiempirical approaches involve 2D representations. This can be explained by a propensity to seek 'linear' relationships among features. To date, 3D representations have rarely been used because of the difficulty in perceiving a pattern as the density of points increases. Therefore important is also the ability to visualize the results in a way that scientists can easily see the separation between the different groups of materials properties, e.g. compound formers from nonformers. Here we show the results of the by us developed program Discovery [19] to search and visualize systematically for the salient 3D features sets and to correlate qualitatively/quantitatively materials properties of the chemical system (e.g. formers/nonformers) with elemental property parameters of the chemical elements present (tabulated for most chemical elements for most elemental properties).

Our search was subdivided into three steps.

(1) Collection of published elemental property parameters: in this work we included 56 different elemental property parameter sets. These 56 sets can in first approximation be grouped into six groups, here called factors. Each elemental property belonging to a specific factor, e.g. atomic number factor is in first approximation linear dependent to each others. In addition the 6 different factors (elemental properties belong to such factors) show distinct different dependence in elemental property parameter vs. Mendeleev number number plots. They show different behaviors of the chemical elements along the periods with increasing atomic number, as well as along a group number with increasing main quantum number. In other words one can say the six different factors can represent a chemical elements most significantly as being distinct different. Our work showed also that elemental property parameters belonging to the same factor can be interchanged and the result do not principally change, one can only achieve an optimization of the result for a given starting materials data-set.

From past work we knew the existence of five different factors [1]. In the context with this work we discovered a sixth factor, the Mendeleev number factor, which showed to be extremely powerful in context with our former/nonformer problem. The 56 elemental property parameters can be grouped as follows

- Size factor (9 elemental property data-sets)
- Cohesion-energy factor (12 elemental property data-sets)
- Electrochemical factor (10 elemental property data-sets)
- Group number factor (2 elemental property data-sets)
- Atomic number factor (13 elemental property data-sets)
- Mendeleev number factor (10 elemental property data-sets)

(2) Building an automatic generator for 2D- and 3D-features sets resulting from combinations of elemental property parameters and mathematical operators

We introduce operators (+, −, *, / and maximum value) to link the elemental property parameters (EP) of the different constituting elements A, B, C,.. to form a global elemental property parameters (EP1(tot)=EP1(A) op EP1(B) op EP1(C)..). DISCOVERY [19] generates automatically for selected elemental property parameter data-sets and selected mathematical operators all combinations using two-, and three different features sets. As a feature we mean the combination of an elemental property parameter and a mathematical operator. Taking 56 elemental property parameters and five mathematical operators into account, there are $5\times56=280$ combinations (elemental property parameter expressions=features) resulting in 39 060 2D- and 3 619 560 3D-features sets, assuming the

best separation is achieved using 2 respectively 3 different features.

(3) Automatic high-quality separation detection and its visualization: the number of to be investigated 2D- and 3D-features sets can become astronomical high for cases where many elemental property parameters sets and mathematical operators are chosen. In this work the most extreme case 56 different elemental property parameter sets and 5 different mathematical operators, leads to over 3 650 000 different 2D- and 3D-features sets. We therefore developed an automatic high-quality separation detector and its visualization, so that e.g. one can visualize e.g. the 100 best separations.

The general idea is very simple, assuming we consider in a selected 3D-features set plot 2000 data-points (e.g. 1000 ternary formers and 1000 ternary nonformers) we investigate for each data-point if it nearest neighbor is of the same class, here former or nonformer and make a statistical analysis for all 2000 data-points. The best results we achieved with optimal 3D-features sets were in the range of 99.0(1) –99.8(1)%. The 1 in the bracket indicates that only the nearest neighbor is taken into account. Actual optimal separation is achieved when the separation accuracy is as high as possible considering e.g. the nearest 50 neighbors. This means in a separation accuracy versus nearest neighbor plot the steepness of the function would be as small as possible. In the best cases we got 85(50)–88(50)%, which means that 85–88% of all 2000 data-points are surrounded by 50 nearest neighbors of the same class, e.g. formers by formers. It visualization showed to be very important, as we got thousands high-quality separations focusing just on our accuracy numbers (99(1)–85(5)), even the spatial distribution of the 2000 data-points can be very different. The two extreme cases are: (a) relatively large formers clusters within the nonformers evenly distributed (b) formers respectively nonformers separated in a relatively complex but well-defined non-linear hyper-plane ‘sheet’ in a 3D-features set space.

4. Neurocomputing

Our neurocomputing work relates to the learning of a functional model in multidimensional space of how materials property depends on features or equivalently, on the position of a system in that multidimensional space [20,21].

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 use of the Mendeleev number as the sole feature of an element, we had followed a two-step neurocomputing 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 neurocomputing 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.

5. Results

We first focus just on the ternary 6382 systems [22]. In Table 1 56 different elemental property parameters are listed grouped according the six factors. First we ran DISCOVERY on all elemental property parameters belonging to the same factor using all above-mentioned mathematical operators and got the following results. We got for each run hundreds to thousands ‘best’ 2D- and 3D-features sets and below we give the accuracy numbers looking at the accuracy numbers for the first neighbors and the accuracy numbers for the 50 nearest neighbors.

1 Mendeleev number factor
99(1)→86(50)%
2 Electrochemical factor
96(1)→71(50)%
3 Size factor (only Zunger’s pseudo-potential radii)
97(1)→67(50)%
4 Group number factor
82(1)→29(50)%
5 Cohesion energy factor
95(1)→27(50)%
6 Atomic number factor
91(1)→25(50)%

It can be clearly seen that the most outstanding factor for the former/nonformer problem is the Mendeleev number factor.

Table 1
The 56 in this work used elemental property parameters grouped according to the six 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)
Electrochemical 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)
$n^{WS\ 1/3}$ 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→right sequence=number atomic (/)
Number Periodic Table start counting right top, right→left sequence (/)
Number Periodic Table start counting left down, left→right sequence (/)
Number Periodic Table start counting right down, right→left sequence (/)
Number quantum (/)
Weight atomic (10^{-3} kg)
Charge nuclear effective Clementi (/)
Charge nuclear effective Slater (/)
Coefficient mass attenuation for MoKα (cm^2 g^{-1})
Coefficient mass attenuation for CrKα (cm^2 g^{-1})
Coefficient mass attenuation for CuKα (cm^2 g^{-1})
Coefficient mass attenuation for FeKα (cm^2 g^{-1})
Factor atomic electron scattering at 0.5 (/)
Mendeleev number factor number Periodic Table start counting left top, top→down sequence (H placed above F)=number Mendeleev (/)
Number Periodic Table start counting right top, top→down sequence (H placed above F) (/)
Number Periodic Table start counting left down, down→top sequence (H placed above F) (/)
Number Periodic Table start counting right down, down→top sequence (H placed above F) (/)
Number Periodic Table start counting left top, top→down sequence (H placed above Li) (/)
Number Periodic Table start counting right top, top→down sequence (H placed above Li) (/)
Number Periodic Table start counting left down, down→top sequence (H placed above Li) (/)
Number Periodic Table start counting right down, down→top sequence (H placed above Li) (/)
Number Mendeleev Pettifor sequence (/)
Number Mendeleev chemists' sequence (/)

The successful separation of two groups of information in 2D- and/or 3D-features plots creates a link between the constituent chemical elements A, B, C.... and its materials properties of A–B–C– . . . The better the separation the more quantitative is its link.

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 a rather limited separation in the 3D-features space, but an excellent separation is achieved using the Mendeleev number. Below the difference between the atomic number and the Mendeleev number is shown

	Atomic number= Number of electrons	Mendeleev number= Sequence number of active part of valence electrons
1st priority	Main quantum number (Number of shells)	Group number (Number of electrons within shell)
2nd priority	Group number (Number of electrons within shell)	Main quantum number (Number of shells)

In the next step we were interested in discovering whether any 2D-, 3D-features sets exist including elemental property parameters belonging to the other five factors which are significantly better compared to the best solutions using just MNs. In conclusion it can be stated that no 2D-, 3D-features sets, including other elemental property parameters which can significantly increase the separation of ternary formers from ternary nonformers, exist. Slight improvement can be achieved by inclusion of Zunger's pseudo-potential radii and Pauling's and Rochow's electronegativity scales, but the improvements are small and they change by using different starting materials data-sets (test sets), therefore they are not significant.

As the number of investigated systems in relation to the potential possible systems gets less and less the more one goes to multi-nary systems it would be optimal being able to include binary, ternary and quaternary systems in the same 2D-, 3D-features set plot. To achieve this we introduced the number of chemical elements as variable in our mathematical expressions, see below the formula for the ternary case

Sum	$(EP(A)+EP(B)+EP(C))/n$
Difference	$(\lvert EP(A)-EP(B)\rvert+\lvert EP(A)-EP(C)\rvert+\lvert EP(B)-EP(C)\rvert)/(n\cdot(n-1)/2)$
Ratio	$((EP(A)/EP(B))+(EP(A)/EP(C))+(EP(B)/EP(C)))/(n\cdot(n-1)/2)$
	with $EP(A)<EP(B)<EP(C)$
Product	$(EP(A)\cdot EP(B)\cdot EP(C))^{1/n}$
Maximum	Max (EP(A), EP(B), EP(C))
	where n=number of elements
	EP(A)=elemental property parameter of the chemical element A
	EP(B)=elemental property parameter of the chemical element B
	EP(C)=elemental property parameter of the chemical element C

Encouraged by the good results for the ternary systems, we used the 'very best' features sets also for the binary and quaternary systems, each separately and also 'all together' and got very similar separation behaviors (less high accuracy for binaries, slight better accuracy for quaternaries). The 'all together' does not just mean taking the average values from the binary, ternary and quaternary cases. It means by introducing the mathematical expressions shown above we were able to treat all about 15 000 binary, ternary and quaternary systems in the same 2D- respectively 3D-plots and achieving similar high accuracies.

Table 1 shows that we used ten different Mendeleev number scales for our investigation. When the Mendeleev number was introduced by Pettifor [23] in context with structure maps, he used a quite distorted string going through the periodic table starting at the top left and ending at the bottom right. To reproduce the periodicity of the periodic table in a regular manner, it was obvious to introduce a whole range of different Mendeleev numbers, trying to be as regular as possible. The most obvious case is the Mendeleev number which is shown in Fig. 1a. One starts to count from the left top corner down to the bottom left corner going through from the first group to the last group of the periodic table (MN1).

Depending where we chose the starting point of the counting we get four different Mendeleev scales (see Fig. 1a–d for MN1–MN4). It proved to be beneficial to put the hydrogen close to the chemical element fluorine (without changing the atomic numbers, MN1–MN4). The analog scales are received by placing the hydrogen above Li (MN5–MN8).

Most surprising was that almost all 2D-, 3D-features sets generated from the elemental property parameters MN1–MN10 were very high. The choice of which Mendeleev number to use out of the ten different scales, as well as which mathematical expression to use for generating the different features sets to give the 'very best' solution was based on items (a–g) under conclusions. The least outstanding results were obtained with Pettifor's Mendeleev number (MN9).

The 'very best' 3D-features set using different MN scales is:

MN2(max) vs. MN3(ratio) vs. MN3(max) plot

Binary (Fig. 2)	Ternary (Fig. 3)	Quaternary (Fig. 4)	All together (Fig. 5)
91.5(1)→32(50)	99.6(1)→86.9(50)	100(1)→98.0(50)	99.5(1)→86.5(50)

The 'very best' 2D features set using just one MN scale is

MN3(difference) vs. MN3(max) plot

Binary (Fig. 6)	Ternary (Fig. 7)	Quaternary (Fig. 8)	All together (Fig. 9)
83.1(1)→29(50)	99.3(1)→85.3(50)	100(1)→99.0(50)	96.0(1)→78.2(50)

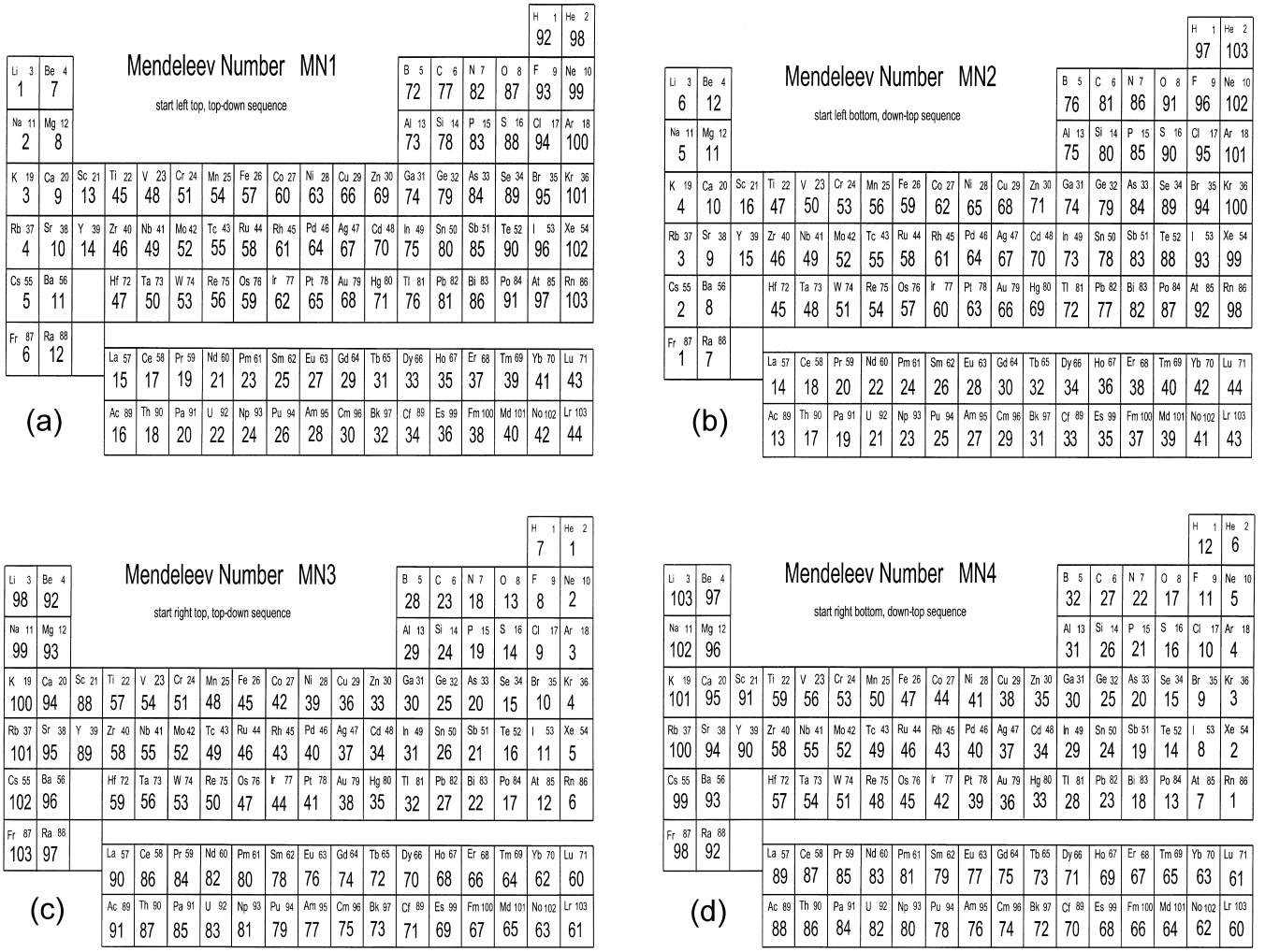

Fig. 1. Mendeleev number (MN1) given as periodic table.

Figs. 2–5 show the same 3D features sets plots for just binary, just ternary, just quaternary and finely binary, ternary and quaternary formers/nonformers (over 15 000 systems) in just one plot.

With the 'very best' 3D-features set: MN2(max) vs. MN3(ratio) vs. MN3(max) we achieved the following results for the prediction of the three given test sets.

Binary systems
Training set B3 training, T1+2, Q1+2 (13 277)
Test file: 221 nonformers 468 formers (689)
violations: 2 nonformers, 26 formers (96.0% accuracy)
no decision: 7 nonformers, 40 formers

Ternary systems
Training set B1+2, T3, Q1+2 (13 277)
test file 713 nonformers 1429 formers (2142)
violations: 4 nonformers, 5 formers (99.6% accuracy)
no decision: 18 nonformers, 24 formers

Quartenary systems
Training set B1+2, T1+2, Q3 training (13 066)
test file 1431 nonformers 922 formers (2353)
violations: 1 nonformers, 0 formers (99.9% accuracy)
no decision: 15 nonformers, 0 formers

- Binaries wrongly predicted: formers instead nonformers (2):
  Al–K; Ag–U
- Binaries wrongly predicted: nonformers instead formers (25):
  Mg–Th; Mg–Pr; Mg–Nd; Ce–Mg; Gd–Mg; Bi–In; Eu–Mg; Pr–Re; La–Mg; Mo–Zn; B–Si; Al–C; Hf–V; Mo–Ni; Be–Mo; Au–V; Be–Ti; Ir–Li; Cd–Sb; Ho–Mn; W–Zr; As–B; Ir–Sr; Re–W; Mn–U
- Binaries no decision: formers versus nonformers (7):
  Be–Ga; Ni–Tl; Bi–Ta; Ru–V; Fe–In; Fe–Pb; Bi–Os
- Binaries no decision: nonformers versus formers (41):
  B–P; Cr–Rh; Co–Mo; Cu–Sb; P–Si; Ag–Sb; Nb–Ni; Au–Sn; As–Cu; C–Cr; Ga–Nb; Au–Hg; Os–U; Ag–

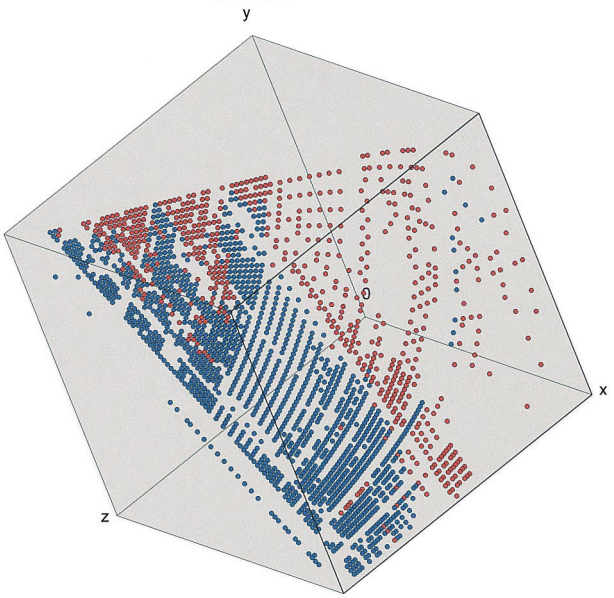

Fig. 2. 2016 distinct different, experimentally determined, binary systems plotted in the 'Discovery Space' (3D features set space) *x*: Mendeleev Number (MN2) Maximum versus *y*: Mendeleev number (MN3) Ratio versus *z*: Mendeleev number (MN3) Maximum. The red dots are the binary nonformers, the blue dots are the binary formers.

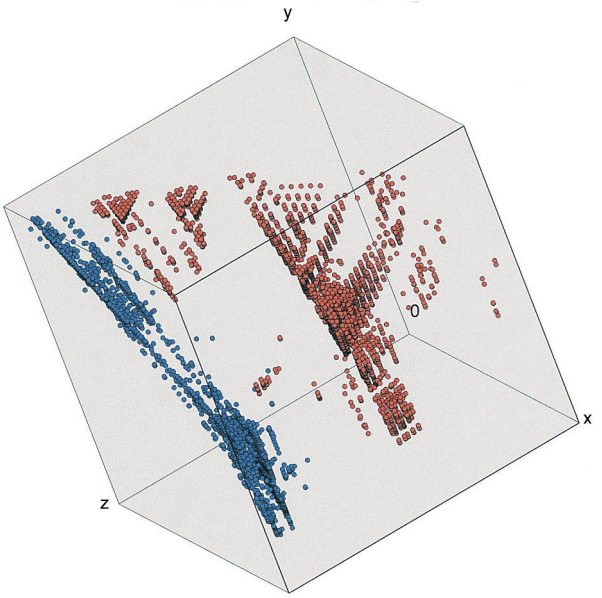

Fig. 4. 7021 distinct different, experimentally determined, quaternary systems plotted in the 'Discovery Space' (3D features set space) *x*: Mendeleev number (MN2) Maximum versus *y*: Mendeleev Number (MN3) Ratio versus *z*: Mendeleev number (MN3) Maximum. The red dots are the quaternary nonformers, the blue dots are the quaternary formers.

Hg; Be–Fe; Ga–V; Mo–Sb; La–Os; Ni–Sn; Au–Pb; Cd–Zr; Ir–Mg; As–Zn; Ge–Ni; In–Mn; Au–Cd; Ba–Pd; Ir–Sb; In–Ru; Au–Rb; Ga–Os; Ba–Pt; Cr–Zr; Cr–Ta; Mo–Re; B–P; Ta–Zn; P–Sn; Cr–Os; Hg–Pd; Mo–Tc

- Ternaries wrongly predicted: formers instead nonformer (4):
  Al–Be–Na; Ag–Cr–U; Ag–U–V; Ag–U–W
- Ternary wrongly predicted: nonformers instead formers (5):

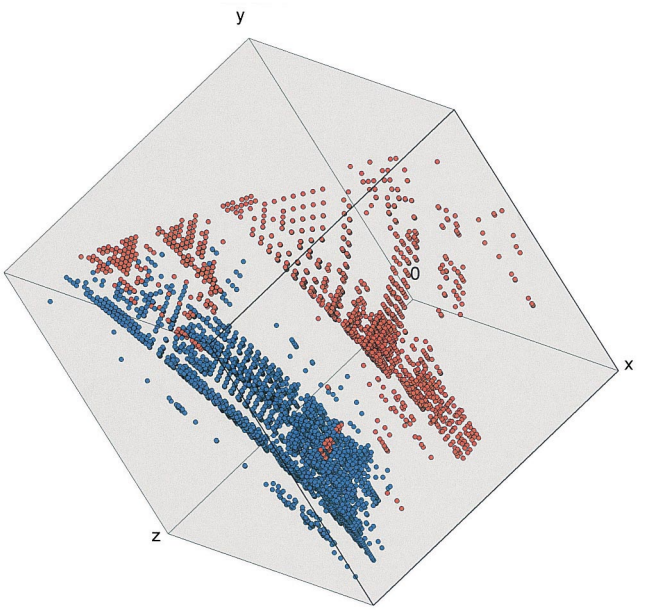

Fig. 3. 6382 distinct different, experimentally determined, ternary systems plotted in the 'Discovery Space' (3D features set space) *x*: Mendeleev Number (MN2) Maximum versus *y*: Mendeleev number (MN3) Ratio versus *z*: Mendeleev number (MN3) Maximum. The red dots are the ternary nonformers, the blue dots are the ternary formers.

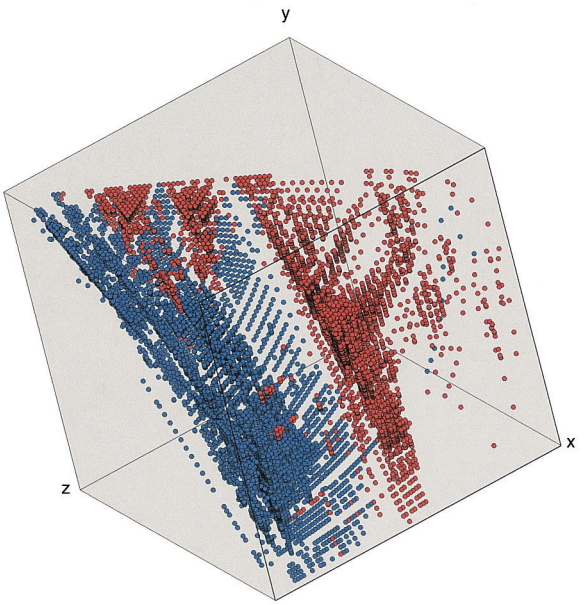

Fig. 5. 15 419 distinct different, experimentally determined, binary, ternary and quaternary systems plotted in the 'Discovery Space' (3D features set space) *x*: Mendeleev Number (MN2); Maximum versus *y*: Mendeleev number (MN3) Ratio versus *z*: Mendeleev number (MN3) Maximum. The red dots are the binary, ternary and quaternary nonformers, the blue dots are the binary, ternary and quaternary formers.

Al–B–C; Al–C–Si; Sb–Sn–Zn; Cd–Cu–Sb; Cr–Pt–S

- Ternary no decision: formers versus nonformers (18): Fe–Ru–V; Cu–Ni–Tl; Bi–Co–Cu; Cu–Li–Mn; Cu–Li–Nb; Cu–Li–V; Ba–Fe–Ni; Cr–Li–Ni; Ag–Mn–Pb; C–Co–Ir; Mn–Pb–Tl; C–Ir–Ni; Ag–Cu–Tl; Ag–Bi–Os; Cd–Fe–Pb; Co–Fe–Pb; Ag–U–W; Be–In–Zn
- Ternaries no decision: nonformers versus formers (24): C–Cr–Ga; As–Cd–Ge; As–Cd–Sn; As–Sn–Zn; C–Mg–Ni; C–Co–Mg; Ni–Sb–V; Bi–Mn–Pd; Mn–Pt–Sb; Mn–Pd–Sb; As–Fe–Re; As–Co–Re; As–Ni–Re; Cr–Rh–Sn; Mn–Ni–Sb; C–Ru–V; C–Nb–Zn; C–Nb–Rh; C–Ga–V; C–Fe–Mo; C–In–Nb; C–Co–Mo; Al–C–Mo; Al–Fe–Si
- Quaternaries wrongly predicted: formers instead nonformers (1): Ag–Cr–U–V
- Quaternaries wrongly predicted: nonformers instead formers (0):
- Quaternaries no decision: formers versus nonformers (15): Ag–Mn–Pb–Tl; Ag–Co–Mn–Pb; Bi–Cu–Fe–Os; Bi–Co–Cu–Os; Bi–Co–Fe–Os; Ag–Fe–Ni–Pb; Co–Cu–Ni–Pb; Cu–Li–Nb–Ta; Cu–Fe–Li–Ni; Cd–Mo–Pb–Tl; In–Mo–Pb–Tl; Al–Be–Ga–Zn; Bi–Co–Cu–Pb
- Quaternaries no decision: nonformers versus formers (0)

6. Conclusions

The most outstanding key-point of the different Mendeleev number scales is that the periodicity of the chemical elements within the periodic group and its sequence depending on the main-quantum number is retained for all of them. It is very interesting that the atomic number, the group number and the main quantum numbers themselves loose this information and therefore are not able to describe the former/nonformer behaviors.

The 'very best' results published here represent the most trustworthy 2D-/3D-features sets as they are optimal in respect of

(a) use just one elemental property parameter
(b) accuracy of the used elemental property parameters is 100%
(c) mathematical expressions link the elemental parameters of the different constituting elements to form a global elemental parameter (EP1(tot)=EP1(A) op EP1(B) op EP1(C) . . .) are very simple
(d) boundary line/surface between former/nonformer is simple in the 2D-/3D-features set plots
(e) distribution of the available experimentally known data in respect to all potential data are 'evenly' spread
(f) all potential data are distributed on a simple, well described hyper-plane 'sheet' within the 3D-features set space
(g) separation of binary, ternary and quaternary systems in one 2D-/3D-features set plot.

Important for the optimization of this semiempirical approach described here are the following issues to be watched, as they might influence the results

1. The quality of the starting data (materials property) which one likes to correlate to its elemental property parameter(s) has to be as high as possible, otherwise one starts with a 'too high' noise. To select an 'error-poor' starting materials data-set requires a very careful evaluation of the published, experimentally determined, data by comparing crystal structure data and phase diagram data for each chemical system.
2. Include many known elemental property parameters which are the starting point to derive the potential features, and therefore its 2D-, 3D-features sets.
3. Minimize the number of prediction steps by covering large materials groups (e.g. treat binary, ternary, quaternary systems together instead of e.g. simply treating binaries alone).
4. Investigate carefully the distribution of all potential systems in your 3-D features space, with respect to the experimentally known systems, and take this into account for deciding which is the 'very best' solution.
5. Optimize the accuracy of the prediction of each prediction step to over 99% by using several different approaches and comparing its prediction results.

Finally giving a closer look to our results and its violations (see Figs. 2–9) and under results). It is very obvious that the separation is the least optimal for the binary case (Figs. 2 and 6). Most violations belong to the systems containing Mg and Be, as well as some other s-elements. They show some kind of 'reverse' behavior.

The big advantage of using just one 2D-and/or 3D-features set plot for binary, ternary and quaternary systems is that the training set can be made much larger. Therefore one gets the ability to visualize the results in one 2D-and/or 3D-plot and to locate any binary–quaternary system in context with all experimental knowledge of over 15 000 experimentally investigated systems.

With the results of this work we can prove that our postulate is correct: Materials properties are quantitatively contained in elemental property parameters of its constituent chemical elements. That encourages us to extent our approach to even more specific information relative to new yet-to-be-realized materials systems such as stoichiometries, crystal structures and physical properties.

Acknowledgements

The authors are grateful to Dr. K. Cenzual for her interest in the work and for the critical reading of the manuscript.

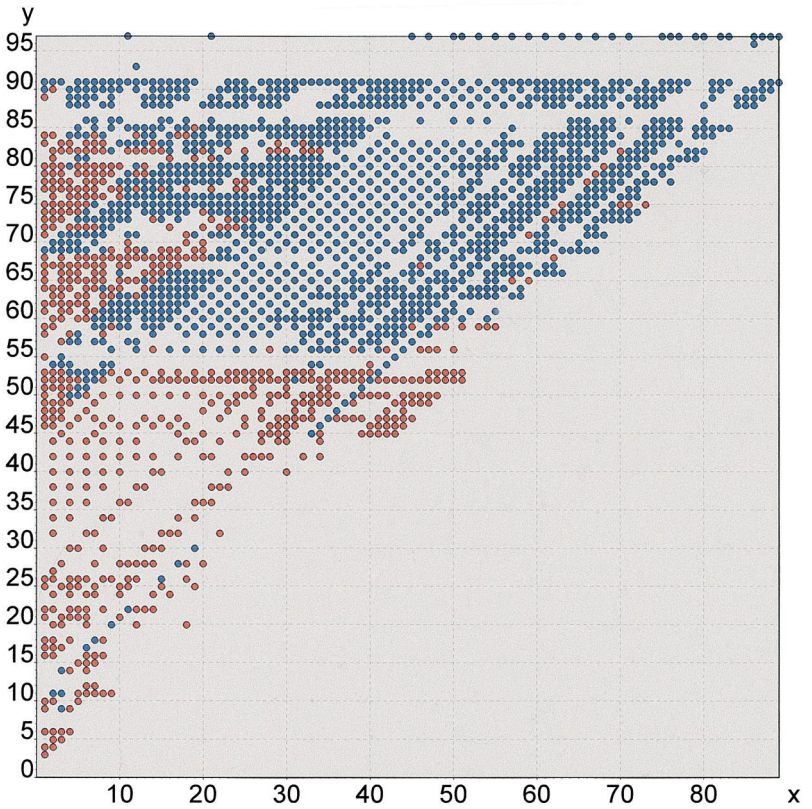

Fig. 6. 2016 distinct different, experimentally determined, binary systems plotted in the 'Discovery Field' (2D features set plot) *x*: Mendeleev number (MN3) Difference versus *y*: Mendeleev number (MN3) Maximum. The red dots are the binary nonformers, the blue dots are the binary formers.

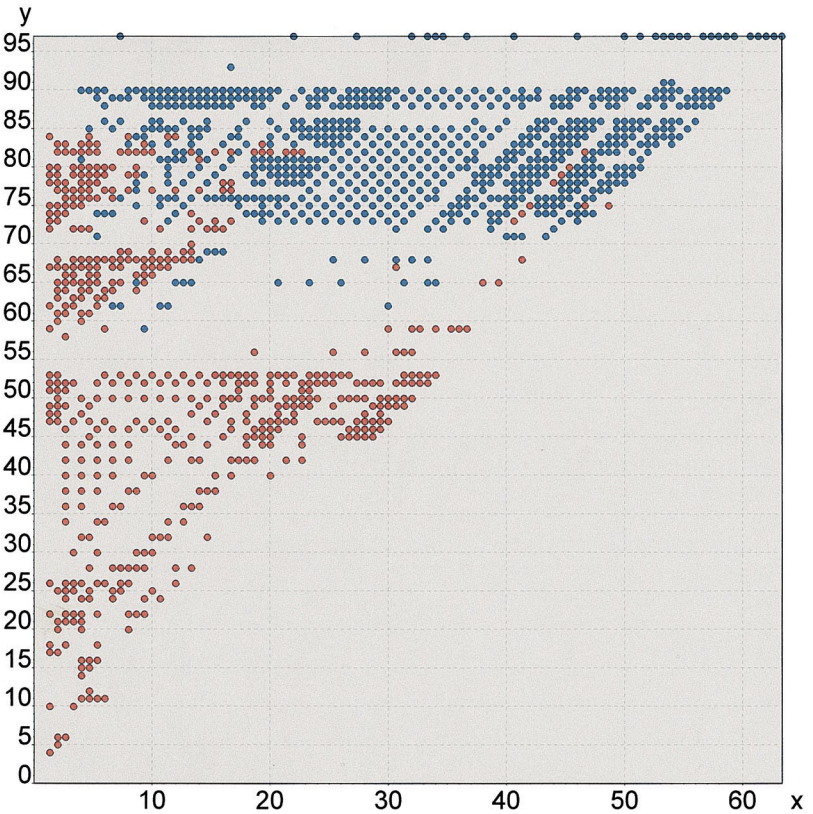

Fig. 7. 6382 distinct different, experimentally determined, ternary systems plotted in the 'Discovery Field' (2D features set plot) *x*: Mendeleev number (MN3) Difference versus *y*: Mendeleev number (MN3) Maximum. The red dots are the ternary nonformers, the blue dots are the ternary formers.

Fig. 8. 7021 distinct different, experimentally determined, quaternary systems plotted in the 'Discovery Field' (2D features set plot) *x*: Mendeleev number (MN3) Difference versus *y*: Mendeleev number (MN3) Maximum. The red dots are the quaternary nonformers, the blue dots are the quaternary formers.

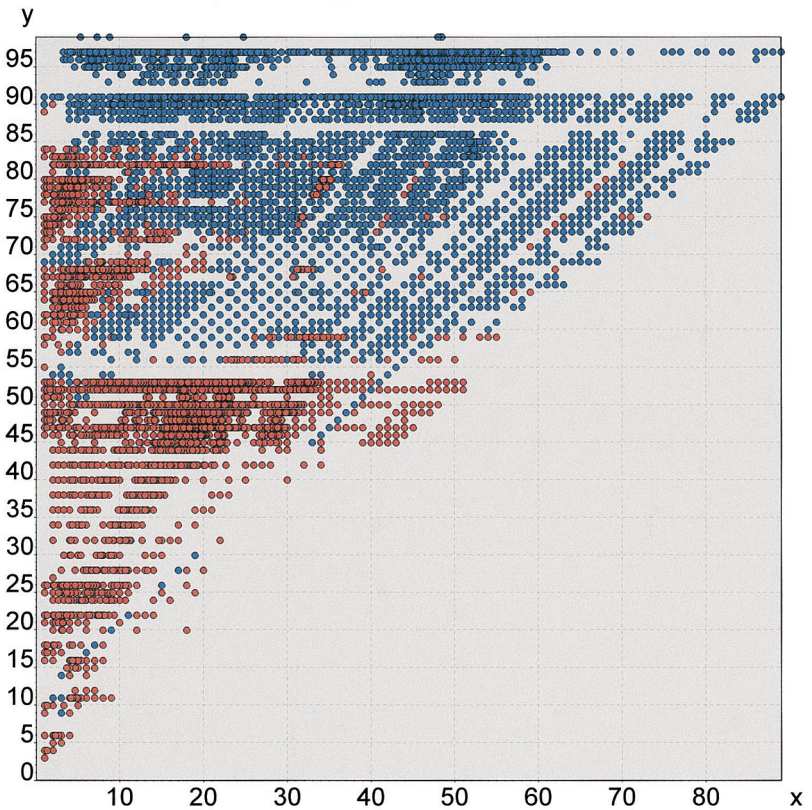

Fig. 9. 15 419 distinct different, experimentally determined, binary, ternary and quaternary systems plotted in the 'Discovery Field' (2D features set plot) *x*: Mendeleev Number (MN3) Difference versus *y*: Mendeleev number (MN3) Maximum. The red dots are the binary, ternary and quaternary nonformers, the blue dots are the binary, ternary and quaternary formers.

References

[1] P. Villars, in: J.H. Westbrook, R.L. Fleischer (Eds.), Intermetallic Compounds, Principles and Practice, Vol. 1, Wiley, 1995, Chapter 11.
[2] ICSD, FIZ/Gmelin, Fachinformations Zentrum Karlsruhe, Germany, 1999.
[3] P. Villars, L.D. Calvert, in: 2nd Edition, Pearson's Handbook of Crystallographic Data For Intermetallic Phases, Vols. 1–4, ASM International, Materials Park, OH, 1991.
[4] P. Villars, in: Pearson's Desk Edition, Vols. 1 and 2, ASM International, Materials Park, OH, 1997.
[5] T.B. Massalski, H. Okamoto, in: Binary Alloy Phase Diagrams, 2nd Edition, ASM International, Materials Park, OH, 1994, and on CD-ROM, 1998.
[6] P. Villars, A. Prince, H. Okamoto, in: Handbook of Ternary Alloy Phase Diagrams, Vols. 1–8, ASM International, Materials Park, OH, 1994, and on CD-ROM, 1996.
[7] Pauling File, LPF, 2000, Japan Science and Technology Corporation (JST), Tokyo and Materials Phases Data System (MPDS), Vitznau, editor-in-chief: P. Villars, editors: K. Cenzual, F. Hulliger, H. Okamoto, J. Daams, K. Osaki, A. Prince, software developer: M. Berndt and K. Brandenburg.
[8] K. Yvon, W. Jeitschko, E. Parthé, J. Appl. Cryst. 10 (1977) 73.
[9] L.M. Gelato, E. Parthé, J. Appl. Cryst. 20 (1987) 139.
[10] COMPARE, M. Berndt, Crystal Impact, Bonn, Germany.
[11] ESDD, 2000, Japan Science and Technology Corporation (JST).
[12] JOIS-F Crystal data, JST, Japan Science and Technology Corporation, Tokyo, Japan
[13] CRYSTMET, CAN/SND, Canadian Institute for Scientific and Technical Information, NRCC, Ottawa, Canada, 1994.
[14] J.L.C. Daams, P. Villars, J.H.N. vanVucht, in: Atlas of Crystal Structure Types For Intermetallic Phases, Vols. 1–4, ASM International, Materials Park, OH, 1991.
[15] E. Parthé, L. Gelato, B. Chabot, M. Penzo, K. Cenzual, R. Gladyshevskii, Gmelin Handbook of Inorganic and Organometallic Chemistry, 8th Ed.: TYPIX, Standardized Data and Crystal Chemical Characterization of Inorganic Structure Types, Vols. 1–4, Springer, Heidelberg, 1993, 1994.
[16] N.N. Kiselyova, MRS Bull. 18 (1993) 40.
[17] J.C. Slater, Theory of Alloy Phases, American Society of Metals, Cleveland, OH, 1956.
[18] J.R. Chelikowski, in: Encyclopedia of Applied Physics, Vol. 3, VCH, Weinheim, 1991, p. 1.
[19] Discovery, K. Brandenburg, M. Berndt, P. Villars, Crystal Impact, Bonn, Germany, 2000.
[20] Y.H. Pao et al., J. Alloys Comp. 279 (1998) 22.
[21] N. Chen et al., Eng. Appl. Artific. Intell. J. 11 (1998) 669.
[22] P. Villars, K. Brandenburg, M. Berndt, S. LeClair, A. Jackson, Y.-H. Pao, B. Igelnik, M. Oxley, B. Bakshi, P. Chen, S. Iwata, Eng. Appl. Artificial Intell. J., in press, 2000.
[23] D.G. Pettifor, A chemical scale for crystal-structure maps, J. Solid State Commun. 51 (1984) 31.
[24] A.R. Miedema, personal communication, 1986.