Materials Genome Initiative (USA) 2011-2020

Materials Informatics Initiative (Japan) 2014-2025

Materials Genome Institute (MGI-SHU, China)

MARVEL Initiative (Switzerland) 2014-2025

Materials Genome Engineering (USTB, China)

The Importance of the Mendeleev (Periodic) Number in Crystal Structure Prediction - a Data-Centric Approach

Pierre Villars
Material Phases Data System (MPDS)
Vitznau (Switzerland)

villars.mpds@bluewin.ch

http://paulingfile.com

https://mpds.io

Materials Genome Initiative (USA) 2011-2020

Materials Informatics Initiative (Japan) 2014-2025

Materials Genome Institute (MGI-SHU, China)

MARVEL Initiative (Switzerland) 2014-2025

Materials Genome Engineering (USTB, China)

A Data-Centric Approach
The PAULING FILE project

PAULING FILE

Inorganic Materials Database and Design System

PAULING FILE was started in 1994
as collaboration between JST and MPDS

Since January 2016 equally
copyrighted by NIMS and MPDS

During last 25 years:

500 fulltime academic manpower have been invested

Up to now used funds:
30 Mio. Euro

Present Status of LPF

397'000 crystal structure entries
59'000 phase diagrams
140'000 physical properties
(800'000 properties counts)
270'000 bibliographic entries

summarizing over 185'000 selected publications
(all done by professional experts)

12 products contain LPF data:

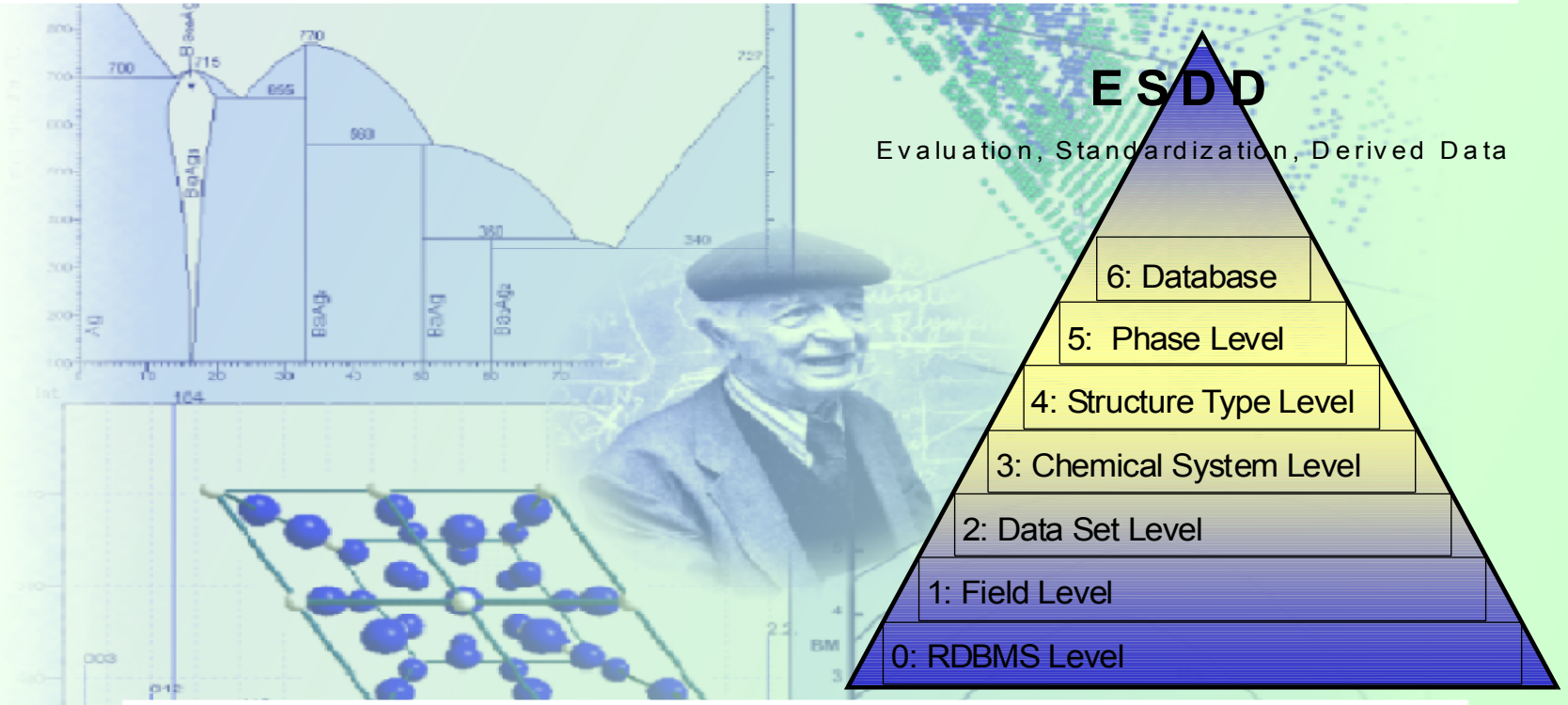
MPDS platform (MPDS/MD Inc.)
AtomWork/AtomWork-advanced (NIMS)
Alloy Phase Diagram Center (ASM)
LPF in SpringerMaterials (Springer)
Pearson's Crystal Data (ASM)
Coordinates in PDF4+ (ICDD)
...

http://paulingfile.com
https://mpds.io
https://crystdb.nims.go.jp

PAULING FILE

Critical Evaluation and Consequent Standardization is a **MUST**

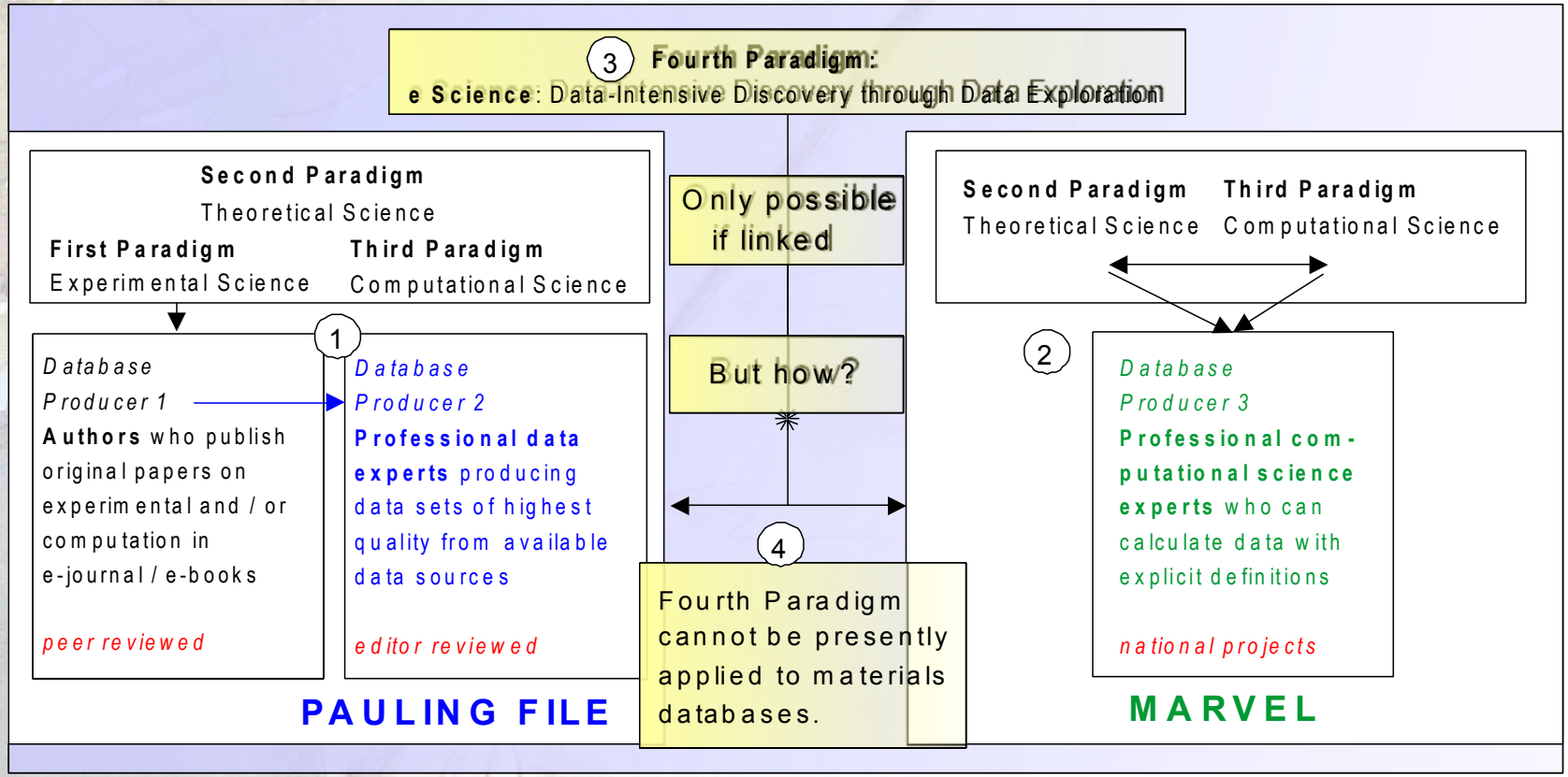
→ All done by professional experts (not by students)



For any Data Mining Technique the following is valid:

→ **Garbage in - Garbage out**

The Fourth Paradigm and the present materials database situation



LINK through

Chemical System and Chemical Formula (available in all materials databases)

IS NOT POSSIBLE AS

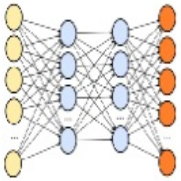
in about 50% of cases there exists 'Berthollide Phases' + 'Polymorphism' + Minerals

PAULING FILE

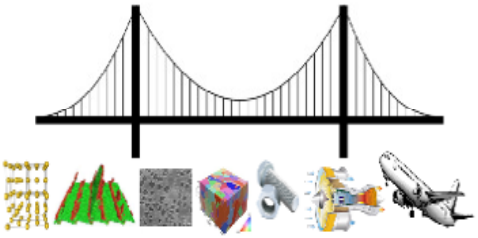
Inorganic Materials Database and Design System

CALPHAD

Machine learning and AI



MPDS platform



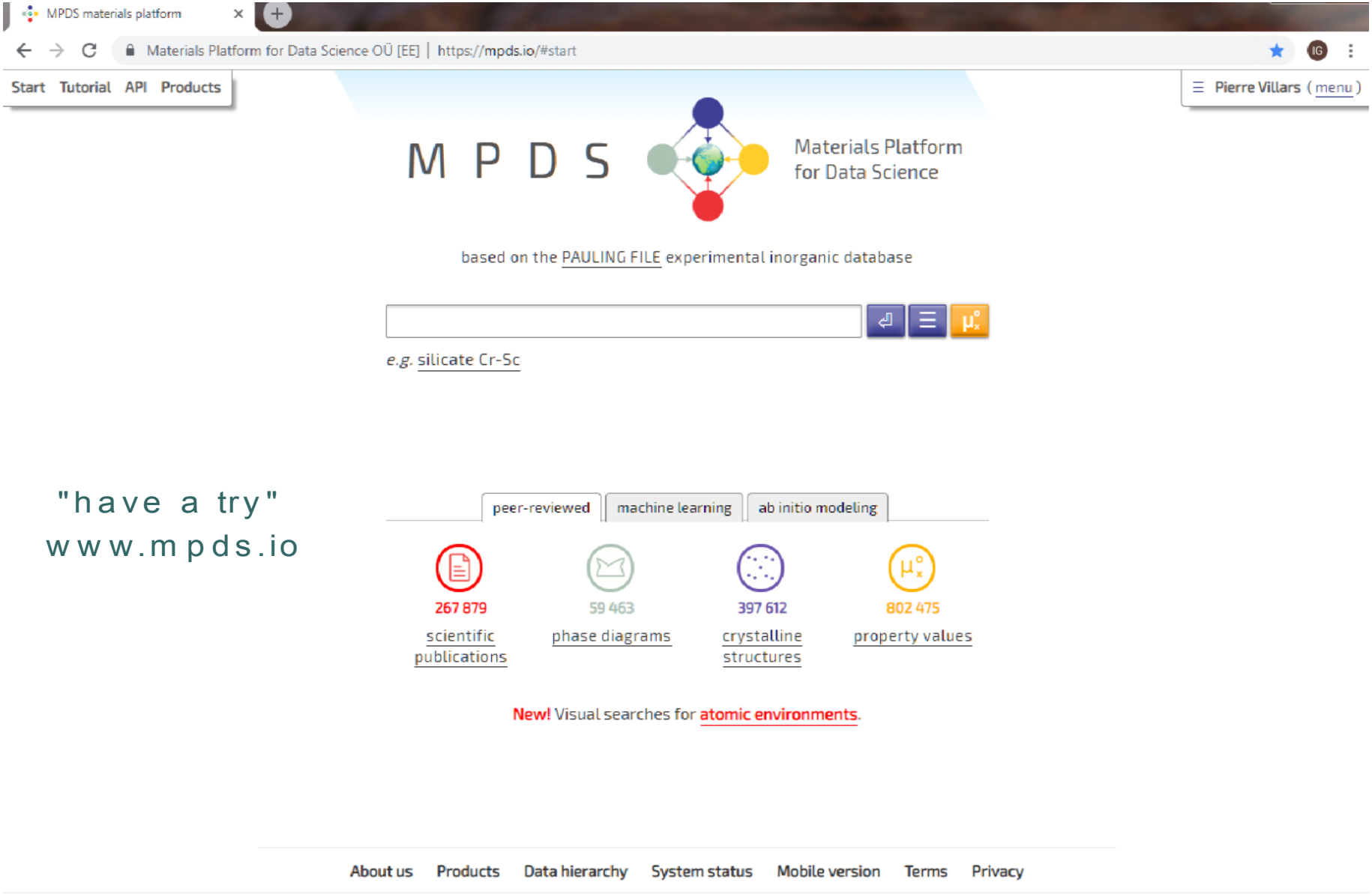
Single-multi-phase bridge

DFT

http://paulingfile.com
https://mpds.io

Future Development

"have a try"
www.mpds.io



Materials Genome Initiative (USA) 2011-2020

Materials Informatics Initiative (Japan) 2014-2025

Materials Genome Institute (MGI-SHU, China)

MARVEL Initiative (Switzerland) 2014-2025

Materials Genome Engineering (USTB, China)

Data Mining by using
Elemental Property Parameters (EPP)
(Governing Factors)
+
the PAULING FILE

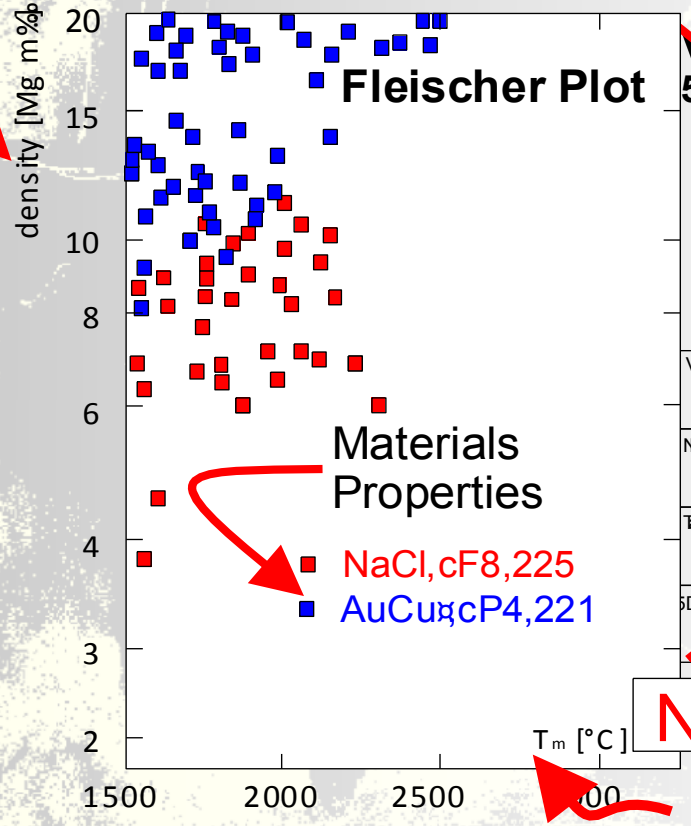
How many distinct Governing Factors exist?

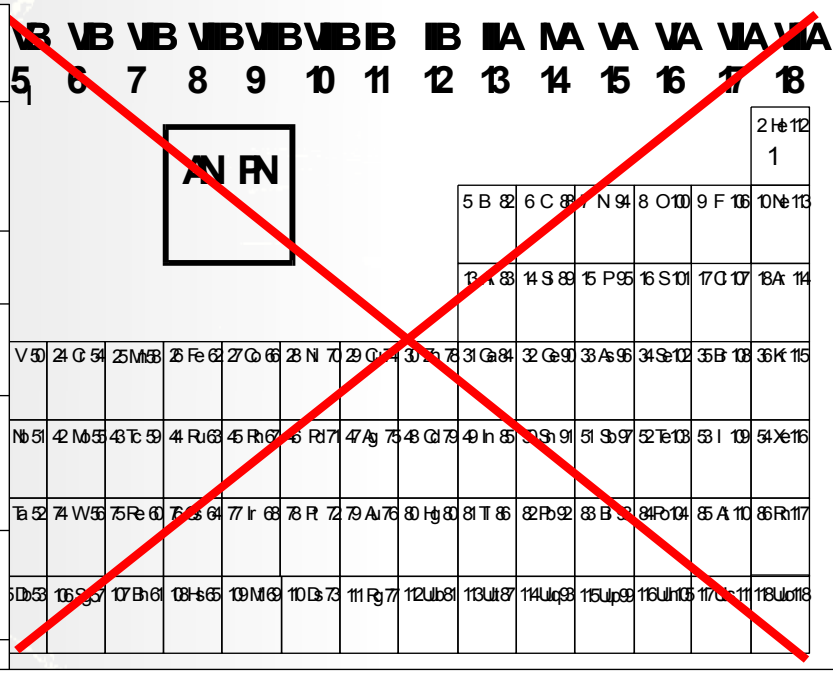
Data Mining
PAULING FILE

Two principal different ways of Data Mining:

I) Pattern within Materials Properties

Materials Properties





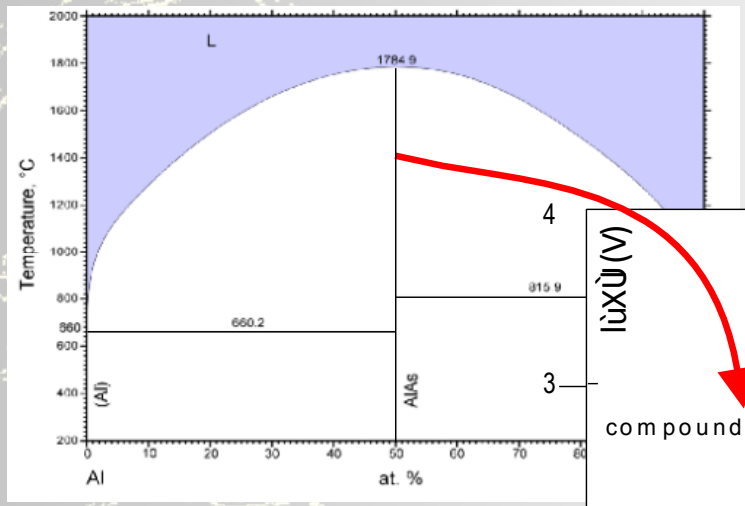
No Link to Periodic Table established

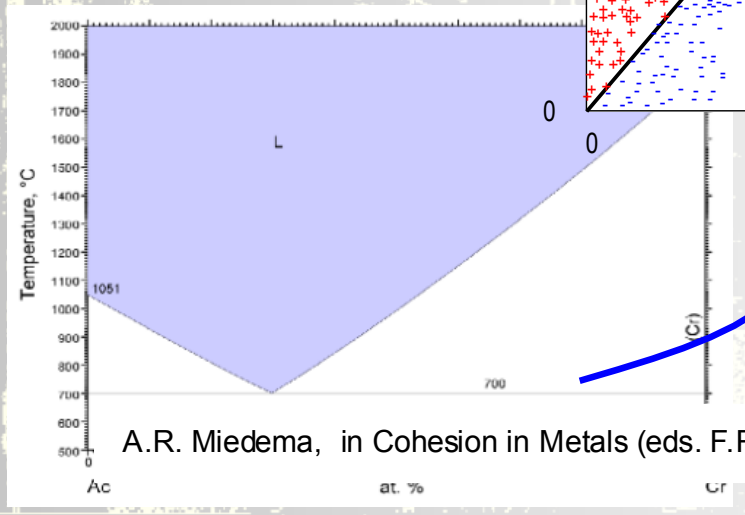
Materials Properties

R.L. Fleischer, J. Mater. Sci. (1987), 22, 2281

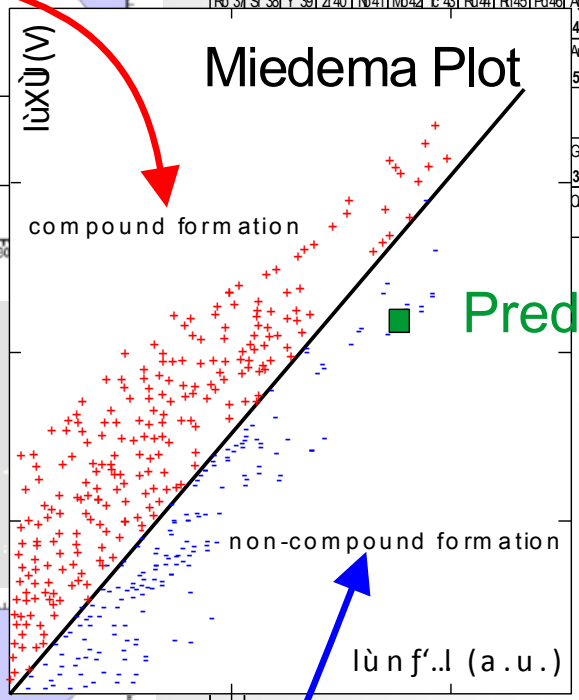
II) Linking Elemental Property Parameters of A, B and Materials Properties of A-B (A_xB_y)

Data Mining
PAULING FILE



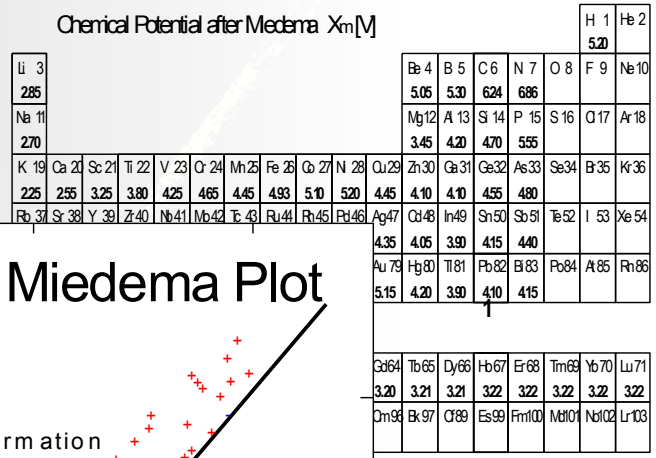


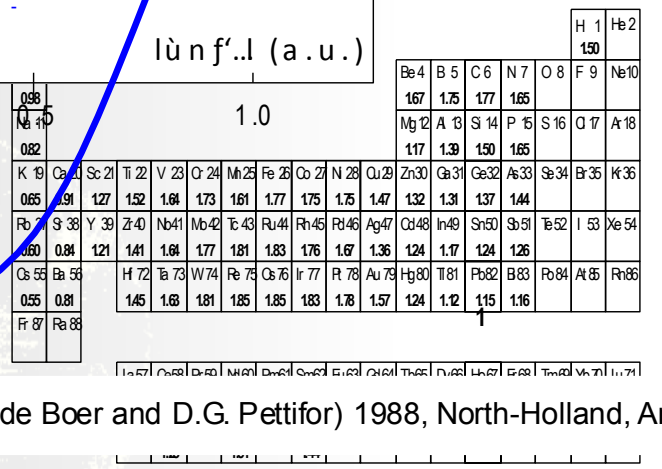
Materials Property of A-B resp. A_xB_y



Prediction Ability

Elemental Property Parameter of A resp. B



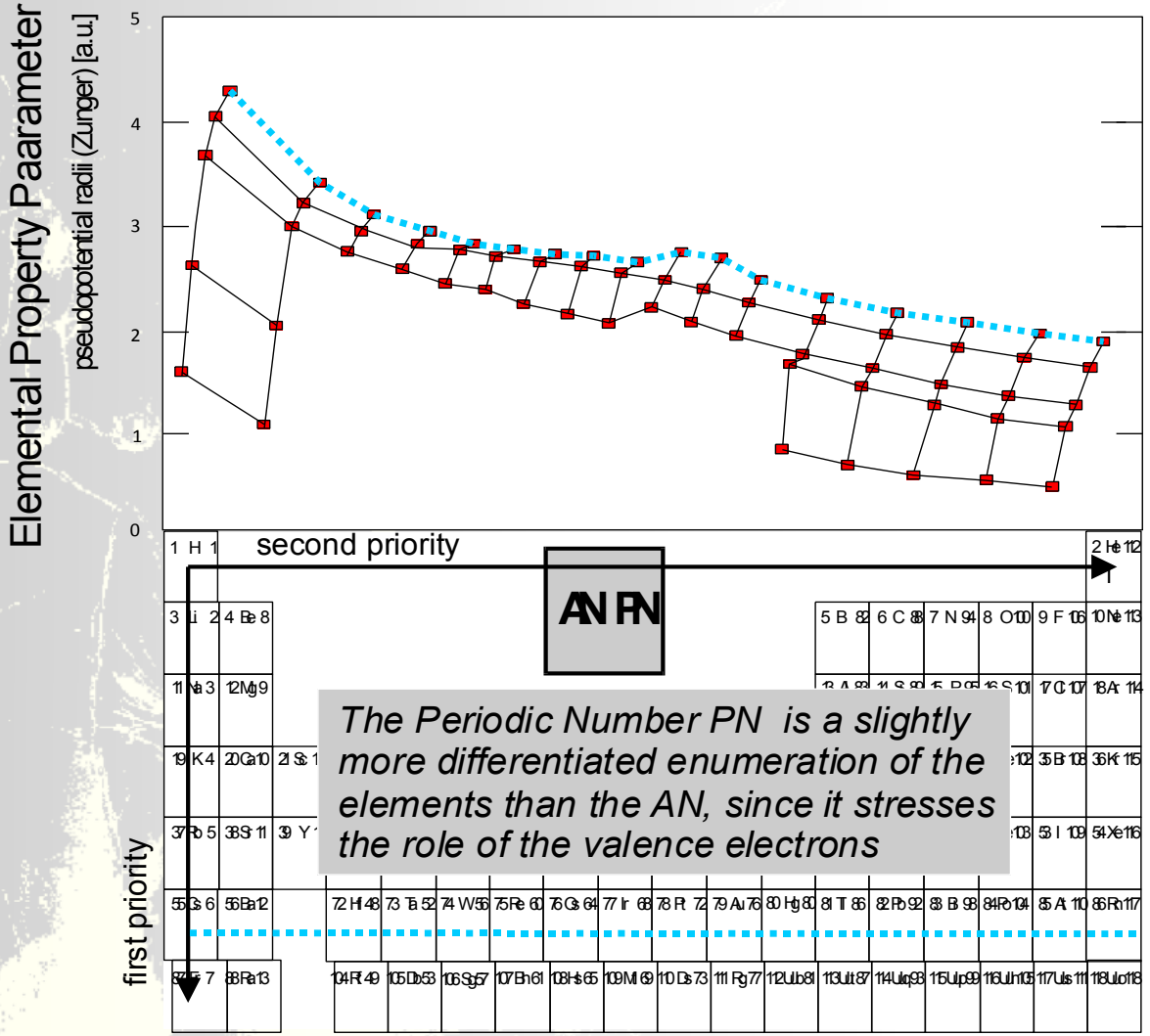


Link to Periodic Table established

A.R. Miedema, in Cohesion in Metals (eds. F.R. de Boer and D.G. Pettifor) 1988, North-Holland, Amsterdam

After Miedema: The EPP (A,B) have some Prediction Ability for A-B (A_xB_y) Materials Properties

Elemental Property Parameters (EPP) vs. Periodic Number (PN) Plots

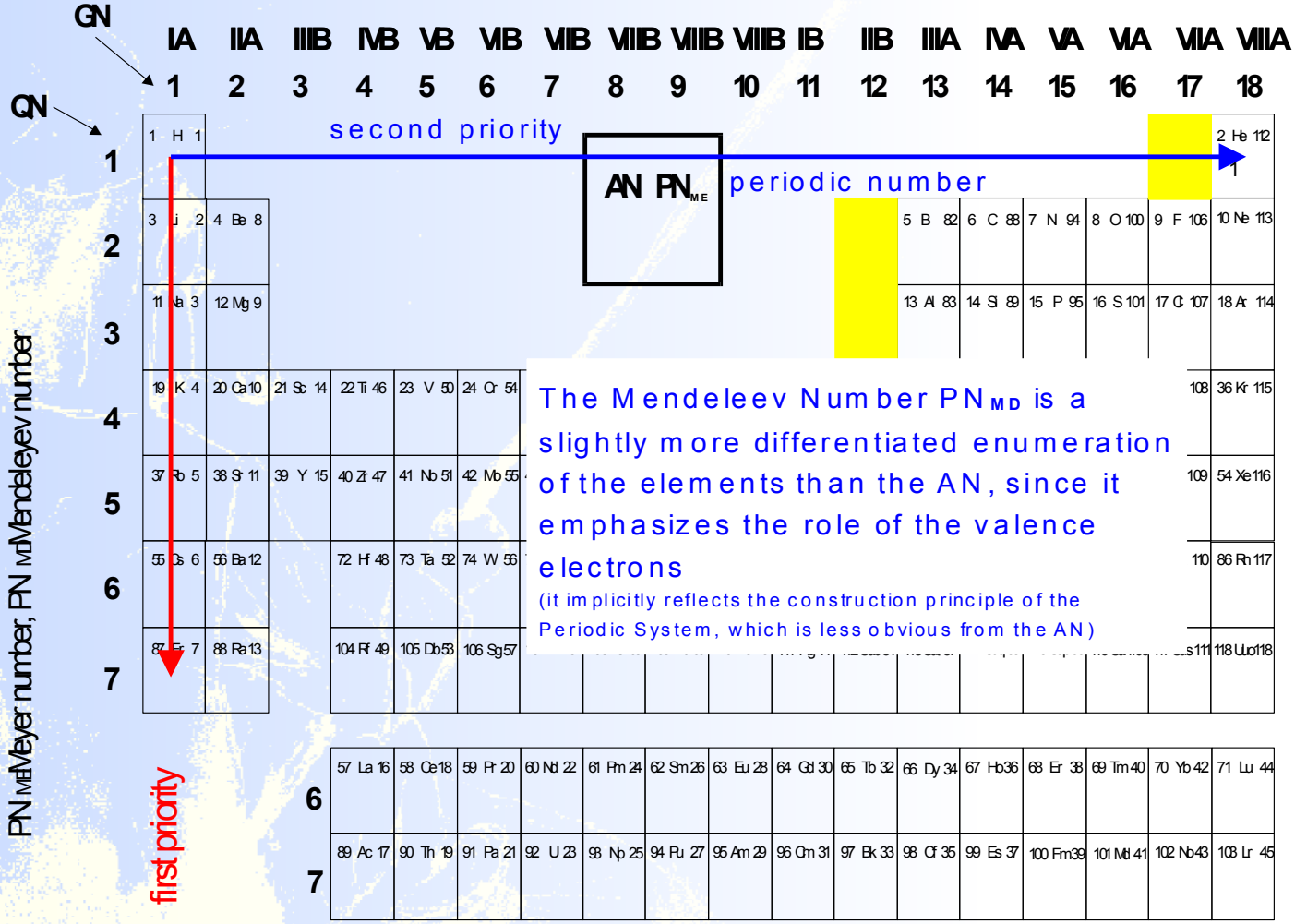


About 100 different EPPs' lead to 4 different patterns:

1) Atomic number group
2) atomic size group
3) atomic reactivity group
4) atomic density (atomic affinity) group

The 4 fundamental integer Elemental-Property Parameters (EPP):

AN, QN, PN (Mendeleev Number), GN



AN and PN (or expression of them) are optimal elemental-property parameters to be used for any kind of materials overview maps as the link to the Periodic Table is straight forward.

Chem. Met. Alloys 1 (2008) 1-23

Materials Genome Initiative (USA) 2011-2020

Materials Informatics Initiative (Japan) 2014-2025

Materials Genome Institute (MGI-SHU, China)

MARVEL Initiative (Switzerland) 2014-2025

Materials Genome Engineering (USTB, China)

$$SZ_a = k_{SZ}\,[\log(AN+1)]\,[k_{PN} - (\log PN)^3],$$

$$RE_a = k_{RE}\,\{[\log(AN+1)]\,[k_{PN} - (\log PN)^3]\}^{-1} = k_{SZ}\,k_{RE}\,(SZ_a)^{-1}$$

where AN is the Atomic Number

PN is the Mendeleev Number

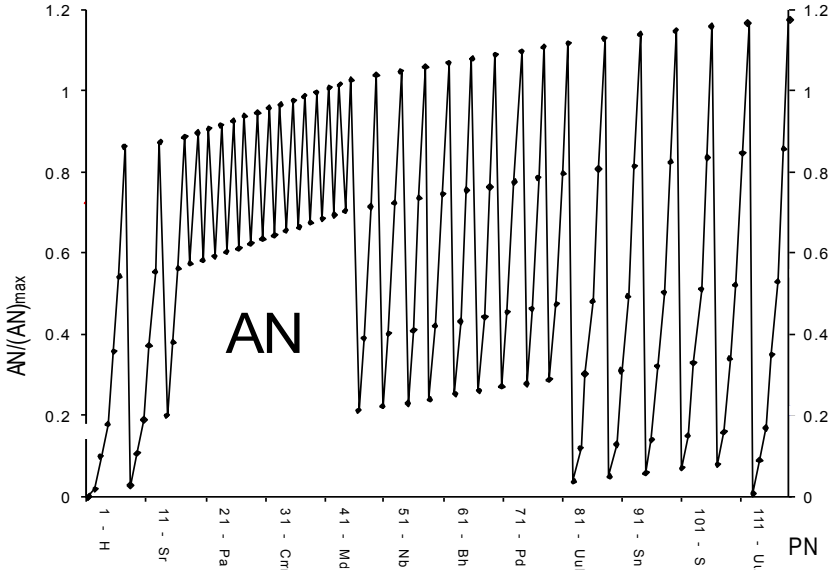
k_{SZ}, k_{RE} are fitting parameters to fit experimental data

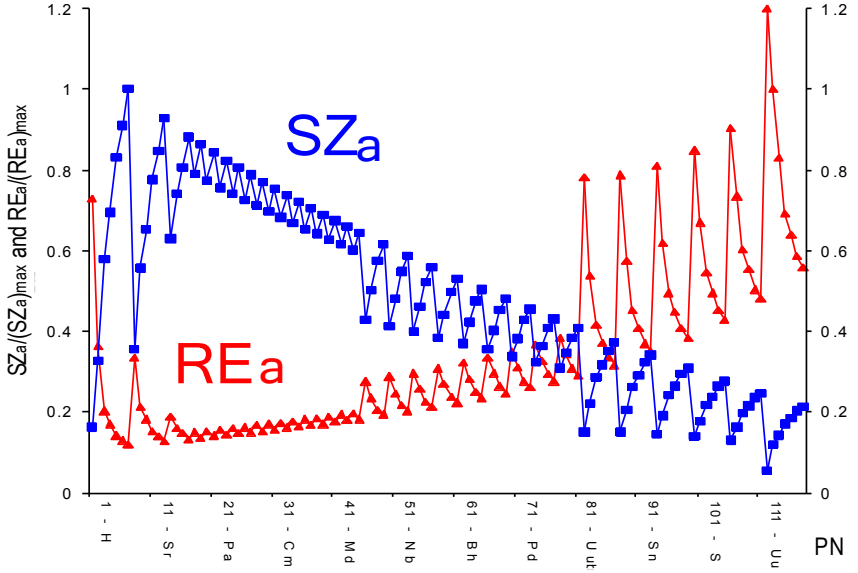
k_{PN} is a scaling factor

Chem. Met. Alloys 1 (2008) 1-23

Functions (Patterns) within Elemental Property Parameters (EPP) versus Periodic Numbers (PN) Plots

1) Atomic Number pattern 2) Atomic Size pattern 3) Atomic Reactivity pattern





AN = f(PN) **SZ_a = f(AN,PN)** **RA_a = f(AN,PN)**

P.Villars et al., Chem. Met. Alloys 1 (2008) 1-23 DEa = f(AN,PN)?

Do we have reached a new qualitative and quantitative Level of Data Mining, as EPP are simple functions of AN and PN?

Fundamental Approach:

Quantum Simulations
MARVEL

f(AN)

complementary

Pragmatic Approach:

Data Mining
PAULING FILE

f(AN, PN)

Materials Genome Initiative (USA) 2011-2020

Materials Informatics Initiative (Japan) 2014-2025

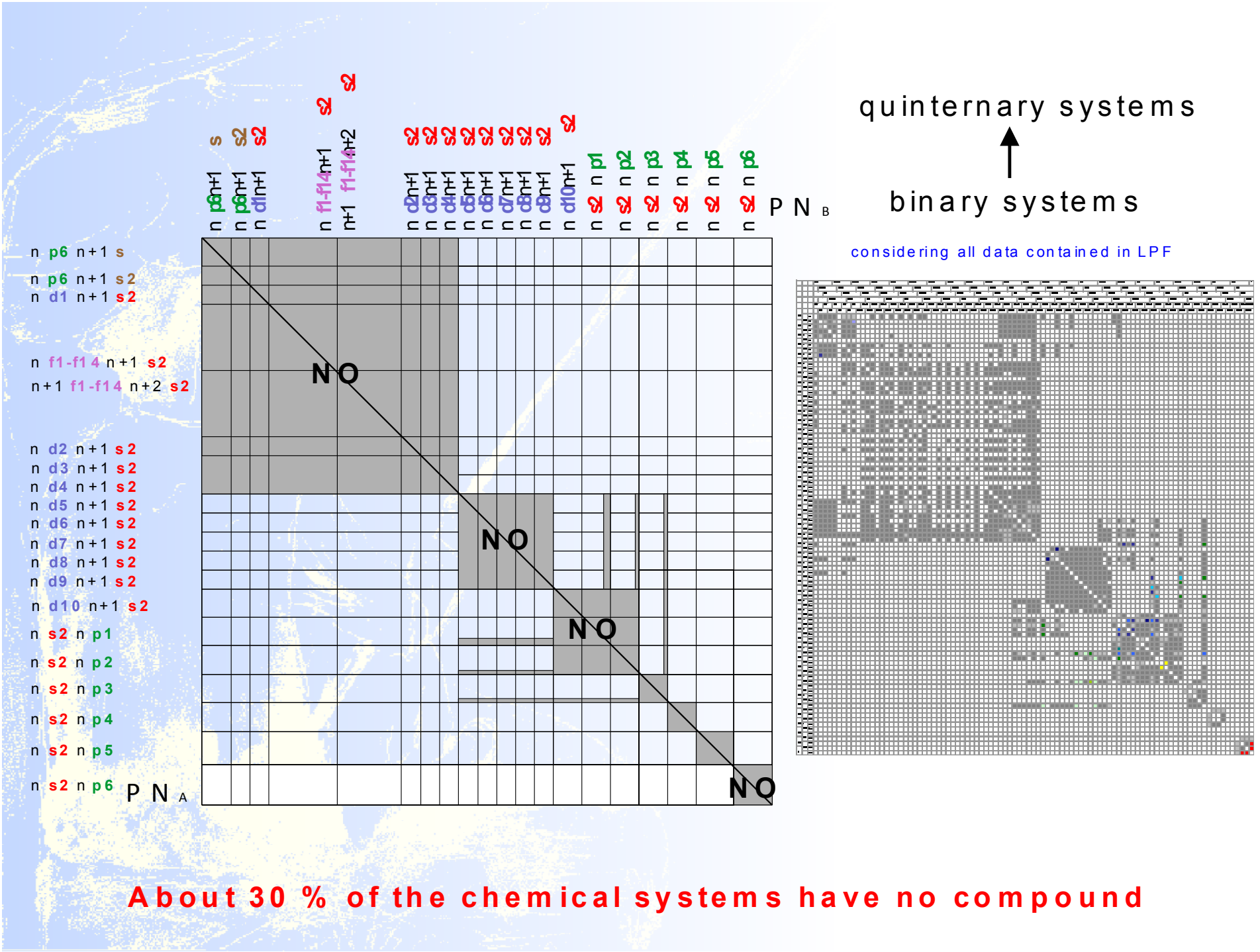
Materials Genome Institute (MGI-SHU, China)

MARVEL Initiative (Switzerland) 2014-2025

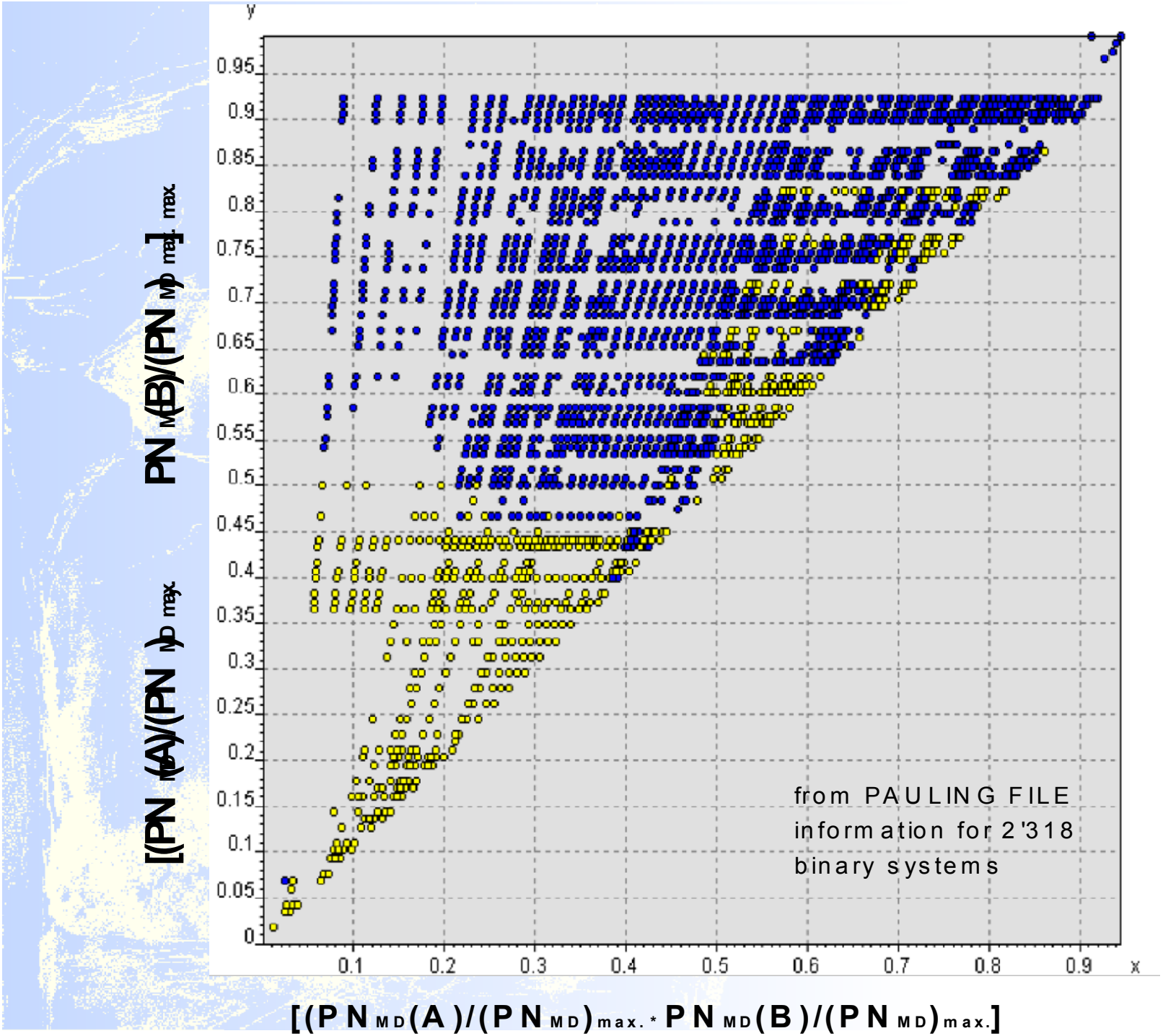
Materials Genome Engineering (USTB, China)

Miedema's Idea to
Link Materials Properties (e.g. Former/Nonformer)
to its Elemental Property Parameters (EPP)

solved with 1 Elemental Property Parameter:
PN_{MD} Mendeleev Number



About 30 % of the chemical systems have no compound



Chem. Met. Alloys 1 (2008) 1-23

Materials Genome Initiative (USA) 2011-2020

Materials Informatics Initiative (Japan) 2014-2025

Materials Genome Institute (MGI-SHU, China)

MARVEL Initiative (Switzerland) 2014-2025

Materials Genome Engineering (USTB, China)

PAULING FILE has two Crystal Structure Classifications:

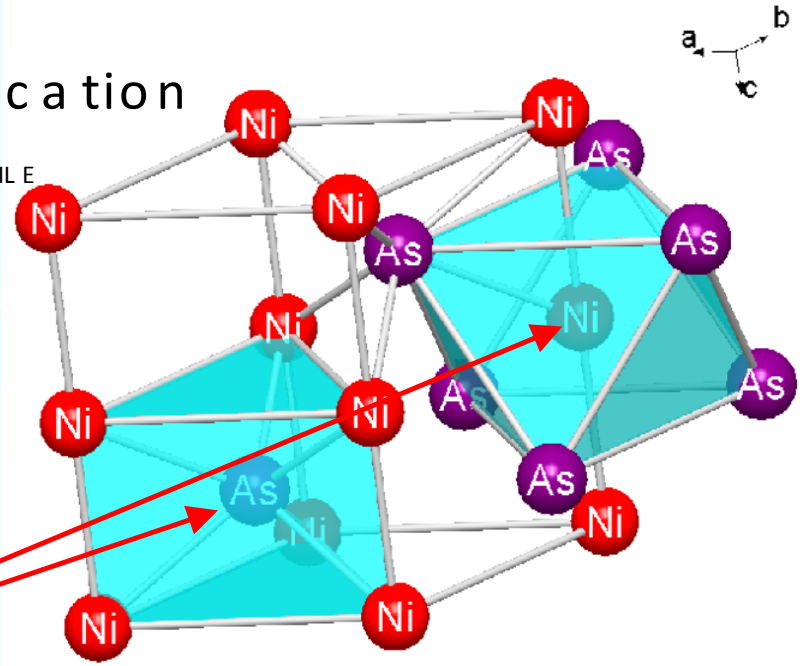
- **Structural Protype Classifications**
- **Atomic Environment Classification**

Structural Prototyping Classification

All data sets with published coordinates are in the PAULING FILE classified in Prototypes, following the criteria defined in TYPIX.

According to this definition isotypic compounds must crystallize in:

- same space group
- have similar cell parameter ratios
- occupy same Wyckoff positions (standardized cell)
- same or similar values of the atom coordinates



If all these criteria are fulfilled, the atomic environments should be similar

(No distinction is made between structures with fully and partly occupied atom sites)

Simplicity

Struktur Bericht Notation (IUCr):

e.g.: A1, B1, $B8_1$

Pearson's HB Notation:

Cu, ClNa, AsNi

Scientifically rigorous

PAULING FILE Notation:

Cu,cF4,225

NaCl,cF8,225

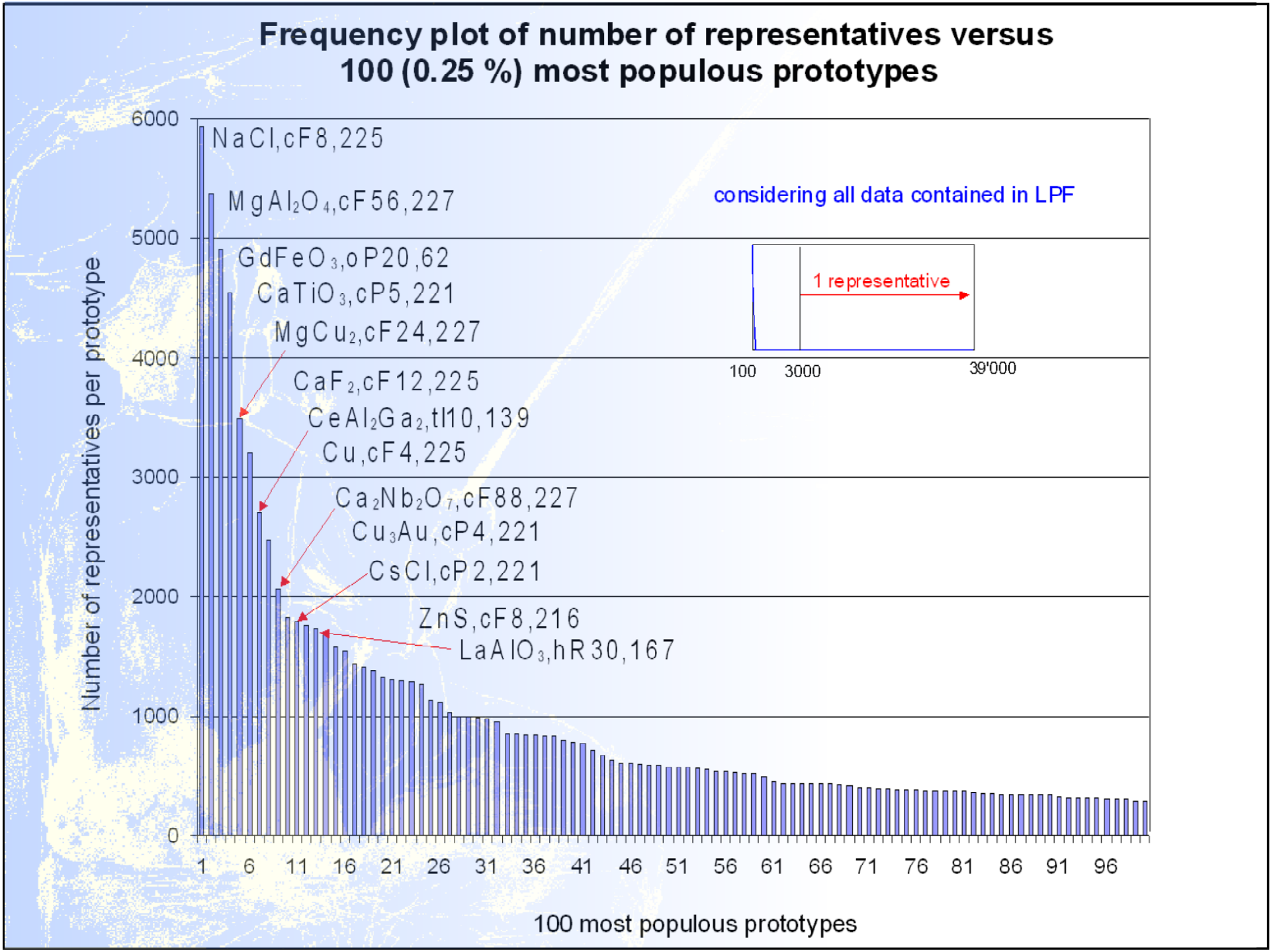
NiAs,hP4,194

Prototype

Pearson Symbol

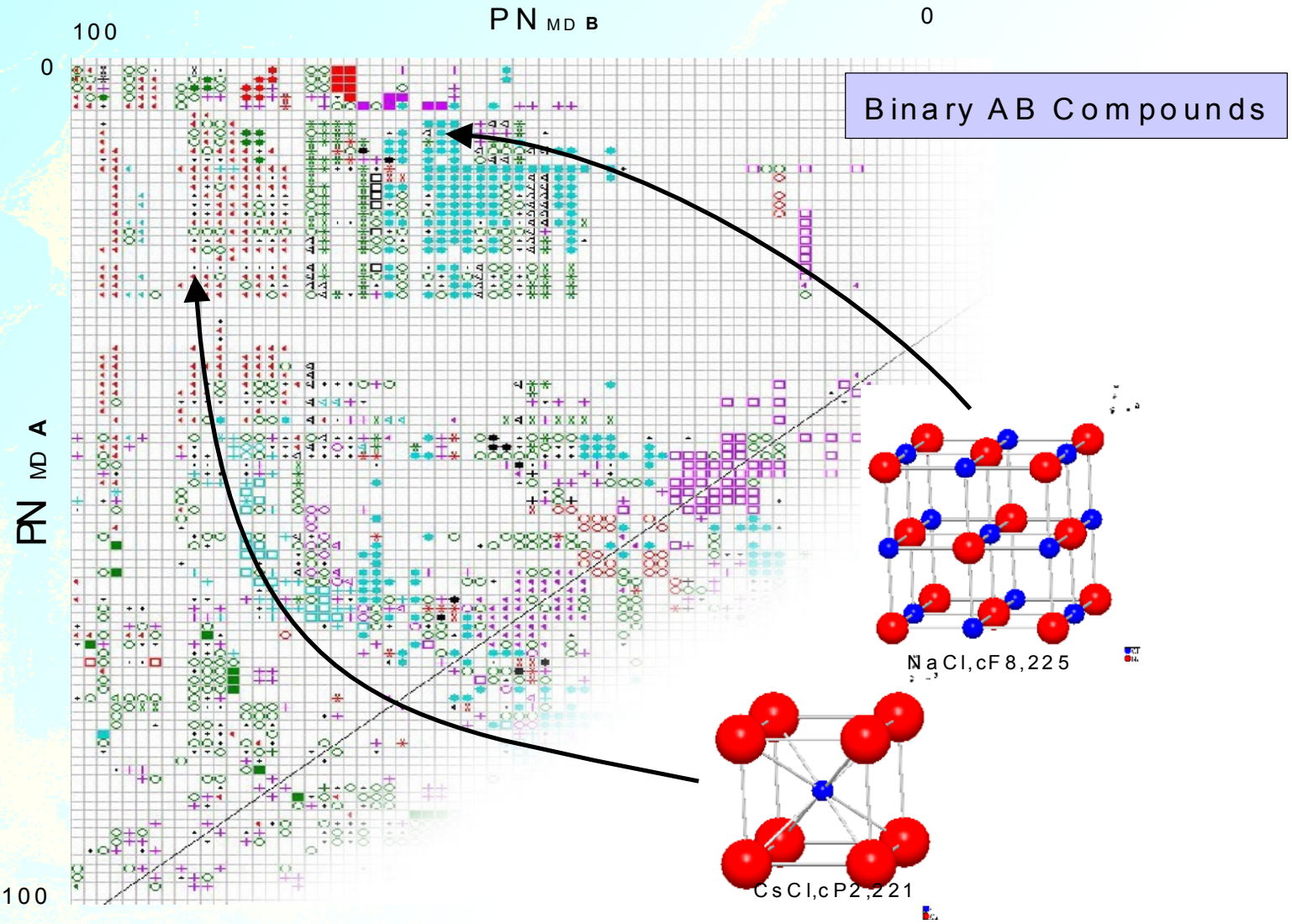
Space Group Number

E. Parthe, L. Gelato, B. Chabot, M. Penzo, K. Cenzual and R. Gladyshevskii, Gmelin Handbook of Inorganic and Organometallic Chemistry, 8th Ed. TYPIX - Standardized Data and Crystal Chemical Characterization of Inorganic Structure Types, 4 volumes,


Frequency plot of number of representatives versus 100 (0.25 %) most populous prototypes
Number of representatives per prototype
6000
5000
4000
3000
2000
1000
0
NaCl,cF8,225
MgAl2O4,cF56,227
GdFeO3,oP20,62
CaTiO3,cP5,221
MgCu2,cF24,227
CaF2,cF12,225
CeAl2Ga2,tI10,139
Cu,cF4,225
Ca2Nb2O7,cF88,227
Cu3Au,cP4,221
CsCl,cP2,221
ZnS,cF8,216
LaAlO3,hR30,167
considering all data contained in LPF
1 representative
100
3000
39'000
1 6 11 16 21 26 31 36 41 46 51 56 61 66 71 76 81 86 91 96
100 most populous prototypes

Classical Prototype Classification

Structure Map



P.Villars et al., J. Alloys Compd., 367 (2004) 167-175

Atomic Environment Classification

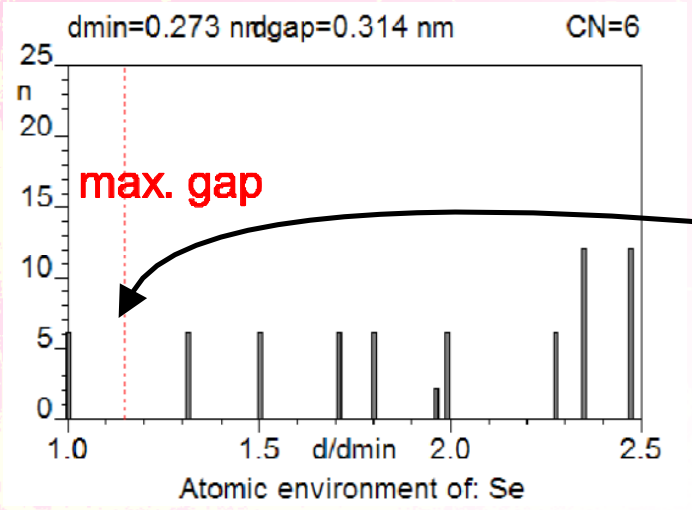
We defined an Atomic Environment Type (AET), also called coordination polyhedron, as follows:

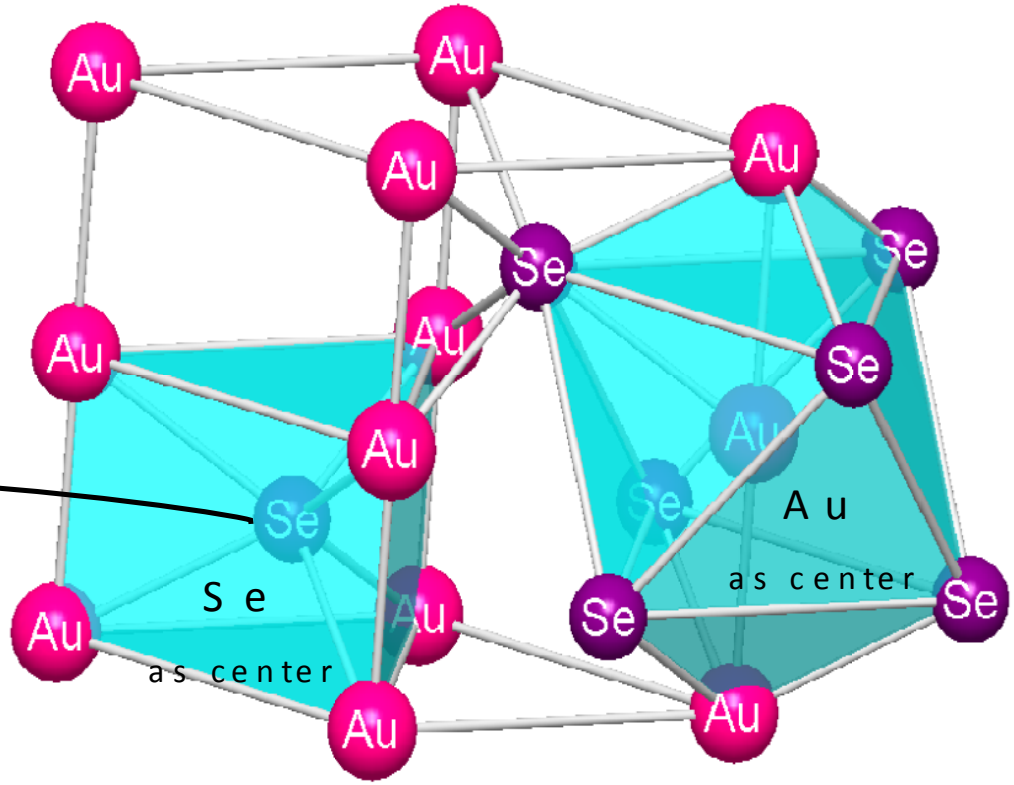
- maximum gap rule (Brunner and Schwarzenbach)
- maximum-convex-volume rule

We kept the number of different AETs in a prototype as small as possible.

Near Neighbor Histogram

n (no of neighbors) vs. d/dmin

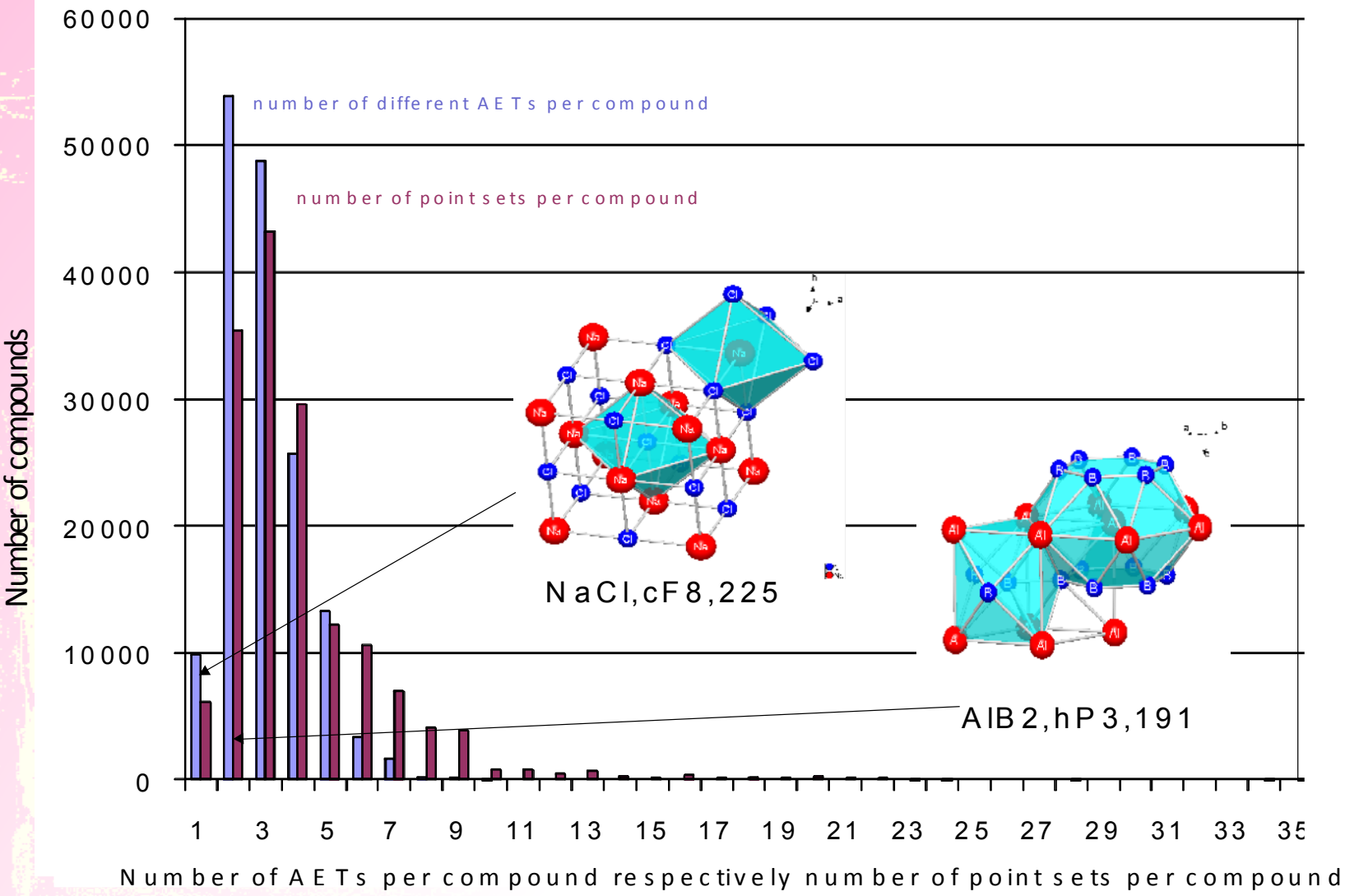




Atomic Environment Types AET are the same for all isotypic compounds

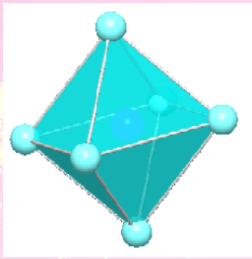
Atomic Environment Classification

far majority less than 5 AET per compound

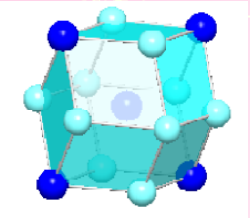


Atomic Environment Classification

Structure Map

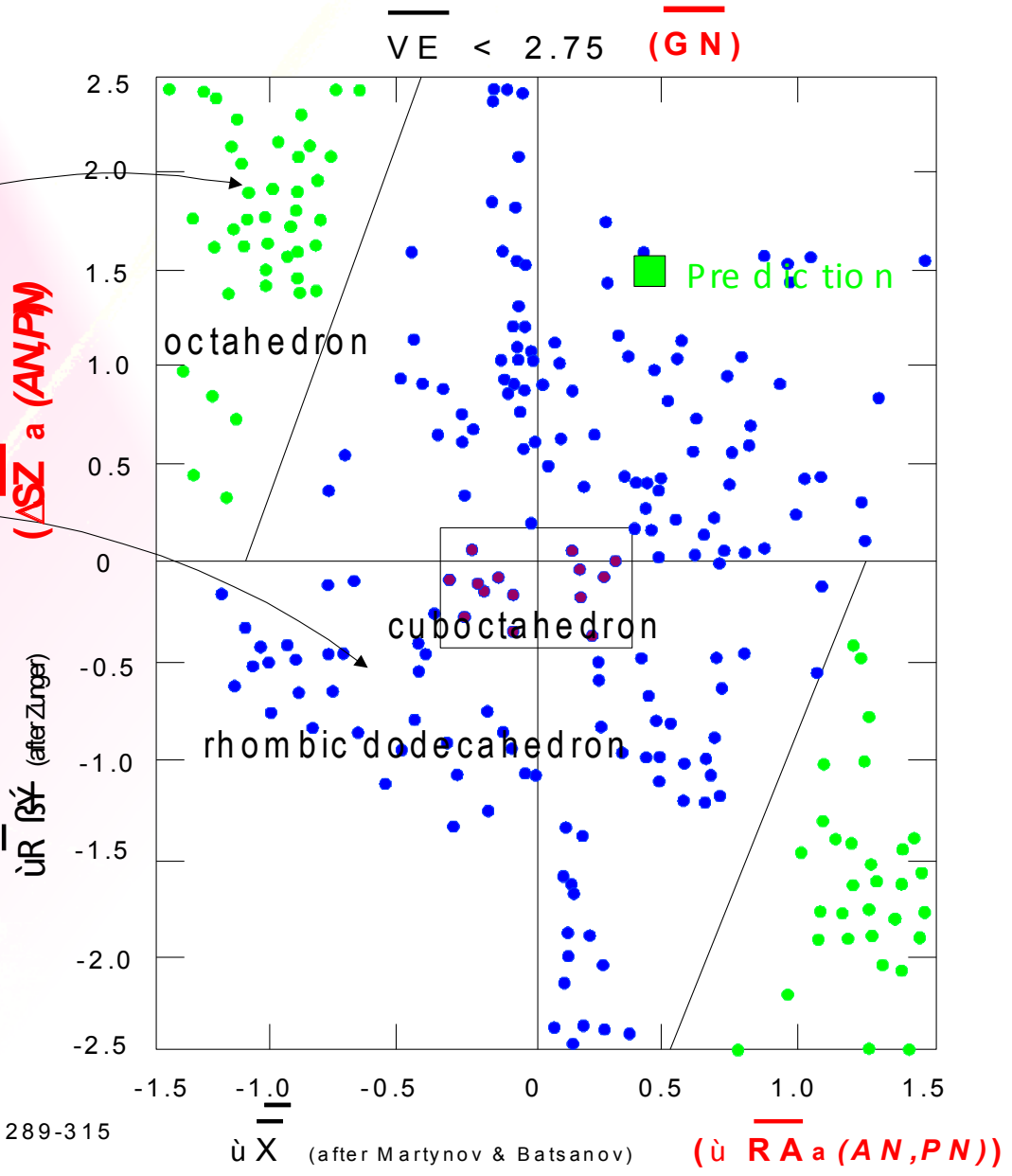


octahedron



rhombic dodecahedron

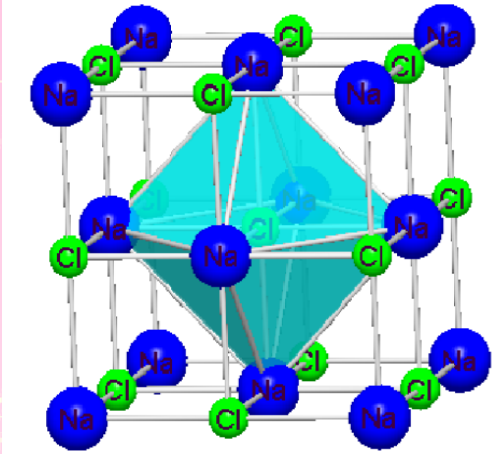
Materials Property:
AET Types (Single Environment)
Used Elemental Parameters: R, X, VE
Number of used Data-Sets:
2,486 single environment compounds
binary, ternary, quaternary daltonide
and berthollide compounds
Accuracy: 97%



P.Villars et al., J. Less-Com. Metals., 132 (1987) 289-315

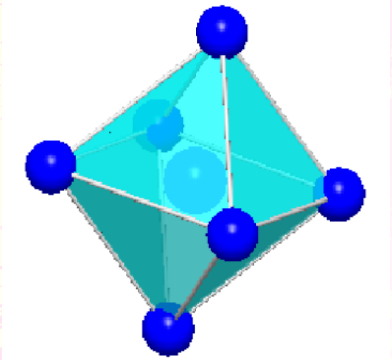
Classical Prototype vs. Atomic Environment Classifications

prototype classification



Prototype: NaCl,cF8,225

AET classification



AET: Octahedron

Classical View (Space Group Theory):

Isotypic compounds (same prototype) must crystallize in the same space group, having similar cell parameters occupy the same Wyckoff positions in the standardized description and have same or similar values of atom coordinates --> atomic environment should be similar.

Presently PAULING FILE contains:

over 39'000 different prototypes

with in total 397'000 entries!!

In average 10 compounds per prototype

Hopeless ?

Atomic Environment Classification:

Presently PAULING FILE contains:

over 1'000'000 fully occupied point-sets and 50 AET

In average 20'000 different pointsets per AET

Materials Genome Initiative (USA) 2011-2020

Materials Informatics Initiative (Japan) 2014-2025

Materials Genome Institute (MGI-SHU, China)

MARVEL Initiative (Switzerland) 2014-2025

Materials Genome Engineering (USTB, China)

Focusing on AB binary compounds
using the
Atomic Environment Classification

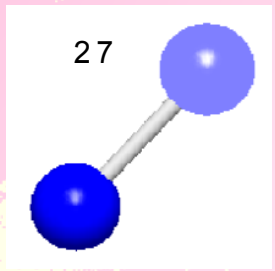
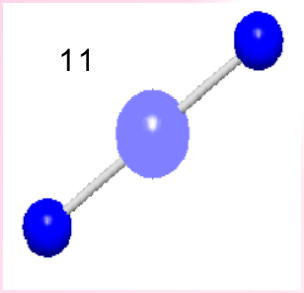
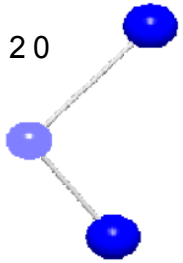
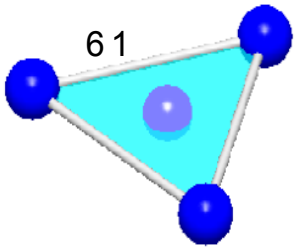
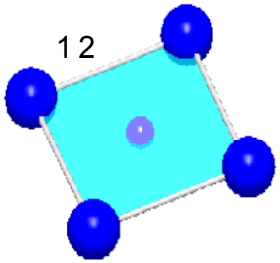
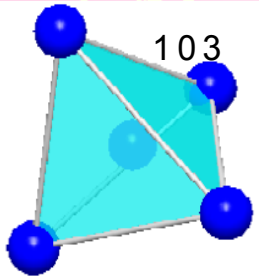
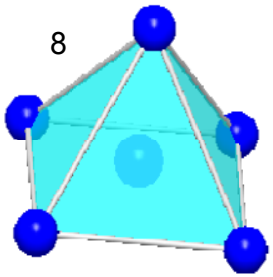
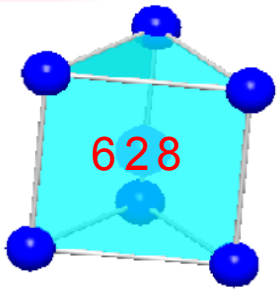
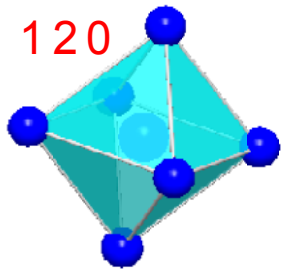
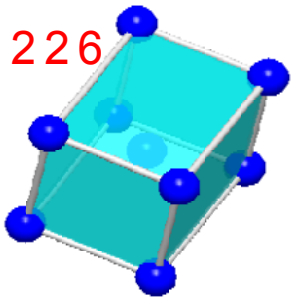
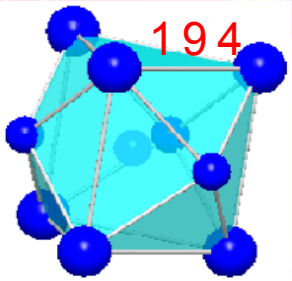
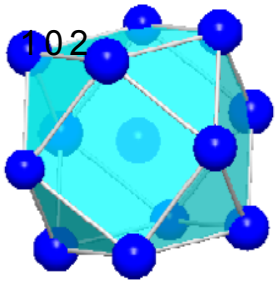
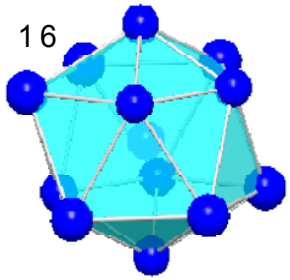
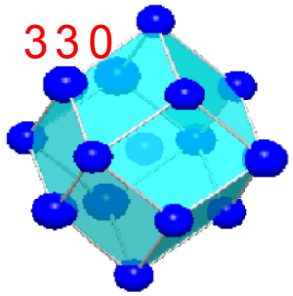
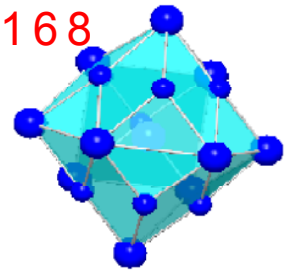
Used Database: PAULING FILE CD-ROM Binaries Edition

Chemical Systems	Limited to thermodyn. defined AB former Data set 1	Including chemical elements + extended s.s. Data set 2
not yet investigated	1480 (34.5%)	1481 (33.9%)
no compounds formed	760 (17.8%)	540 (12.3%)
no AB compounds formed	839 (19.6%)	794 (18.2%)
AB compounds (no crystal data)	73 (1.8%)	79 (1.8%)
AB compounds with crystal data	1126 (26.3%)	1384 (31.7%)
chemical elements		93 (2.1%)
Total	4278	4371

2002 AETs (15 most populous)

AET Dataset 1

Structure Prototyping Systems as Door to Fundamentals

				
single atom	linear	non-linear	(co-planar) triangle	(co-planar) square
				
tetrahedron	square pyramid	trigonal prism	octahedron	cube
				
tricapped trigonal prism	cuboctahedron	icosahedron	rhombic dodecahedron	heptacapped pentagonal prism

PN_{MD} (Center A) vs. PN_{MD} (Atomic Environment B)

for Equi-Atomic Binaries

Data set 1

	no compounds
	no 1:1 compound
?	unknown structure
C	complex
	single atom
	2 linear
	2 non-linear
	triangle
	square
	tetrahedron
	square pyramid
	trigonal prism
	octahedron
	7-vertex polyhedron
	cube
	tricapped trigonal prism
	9-11 others
	(anti)cuboctahedron
	icosahedron
	rhombic dodecahedron
	13-16 others
	pentacapped pentagonal prism

$PN_{MD\ A}$

1

100

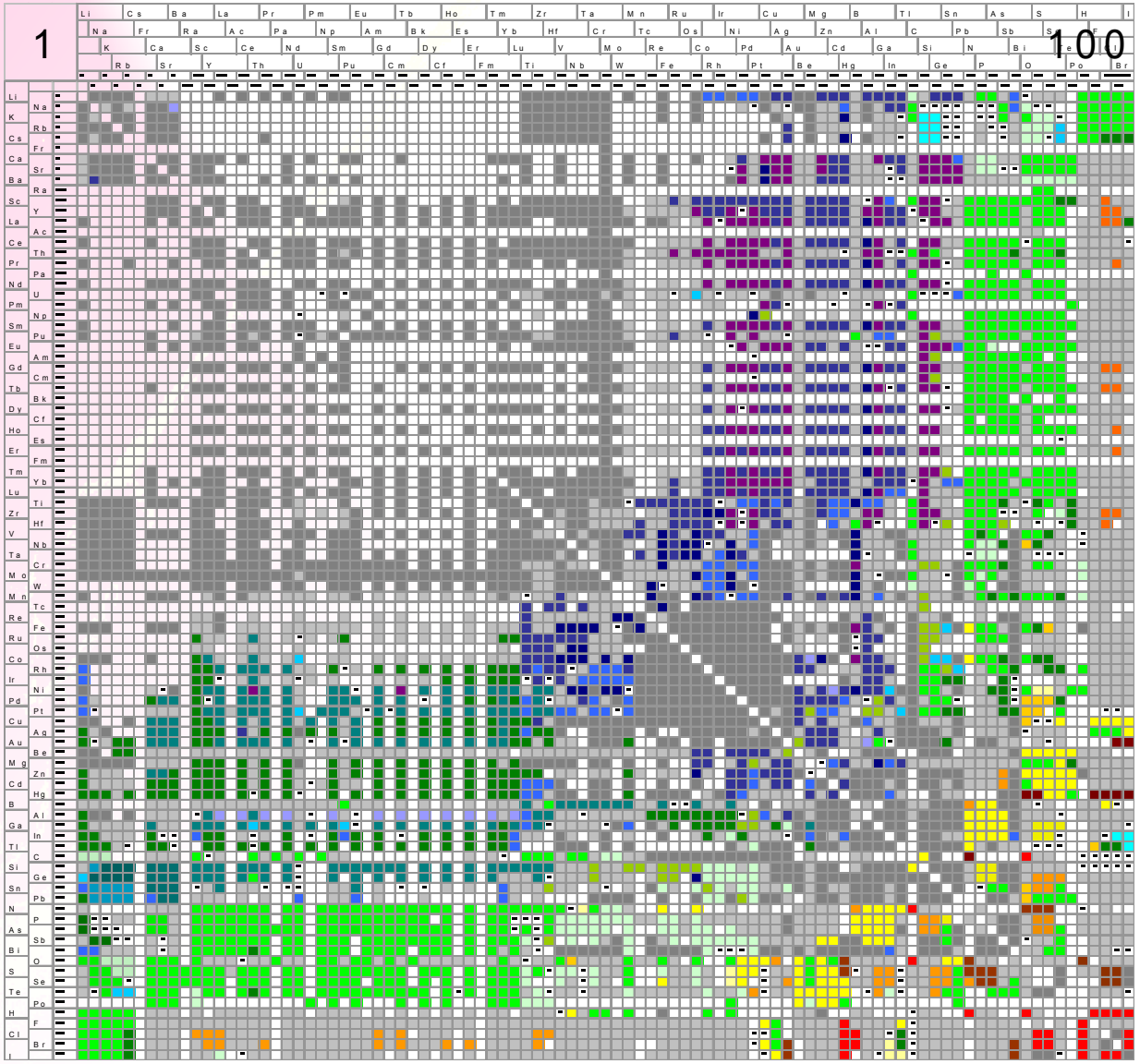
Central Atom

$PN_{MD\ B}$

Coordinating Atoms

1

100



PN_{MD} (Center A) vs. PN_{MD} (Atomic Environment B)

for Equi-Atomic Binaries

$PN_{MD\ B}$

Coordinating Atoms

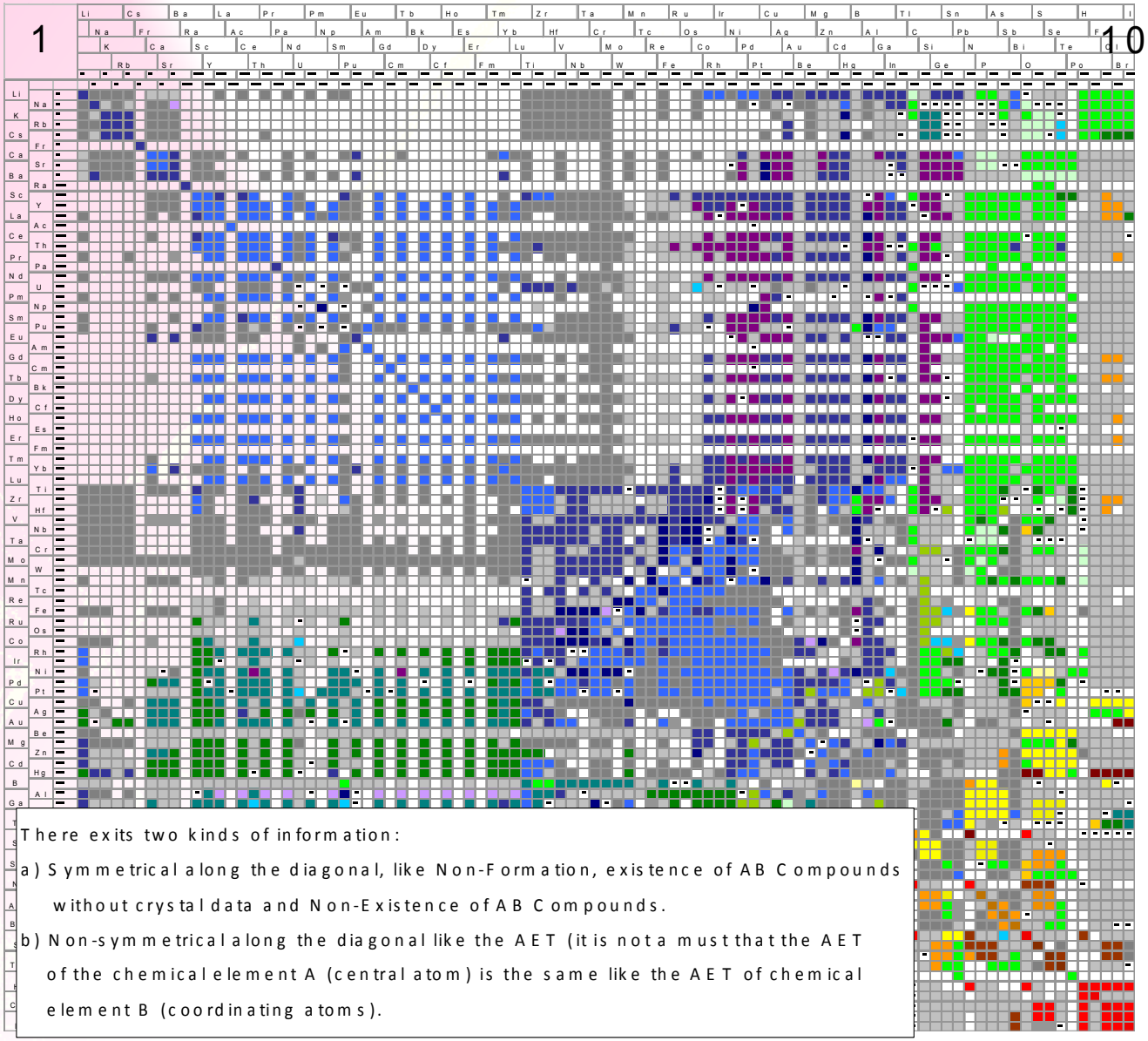
Data set 2

1

100

$PN_{MD\ A}$

	no compounds
	no 1:1 compound
?	unknown structure
C	complex
	single atom
	2 linear
	2 non-linear
	triangle
	square
	tetrahedron
	square pyramid
	trigonal prism
	octahedron
	7-vertex polyhedron
	cube
	tricapped trigonal prism
	9-11 others
	(anti)cuboctahedron
	icosahedron
	rhombic dodecahedron
	13-16 others
	pentacapped pentagonal prism



100

Central Atom

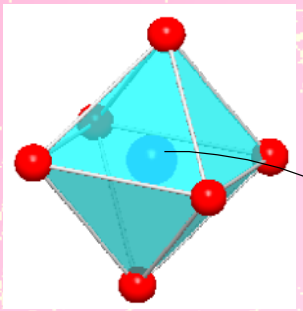
There exits two kinds of information:

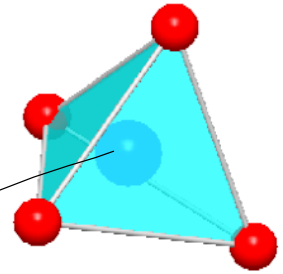
a) Symmetrical along the diagonal, like Non-Formation, existence of AB Compounds without crystal data and Non-Existence of AB Compounds.

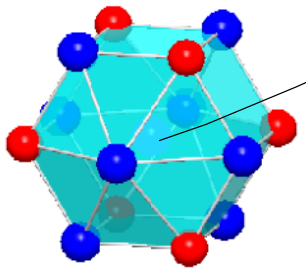
b) Non-symmetrical along the diagonal like the AET (it is not a must that the AET of the chemical element A (central atom) is the same like the AET of chemical element B (coordinating atoms).

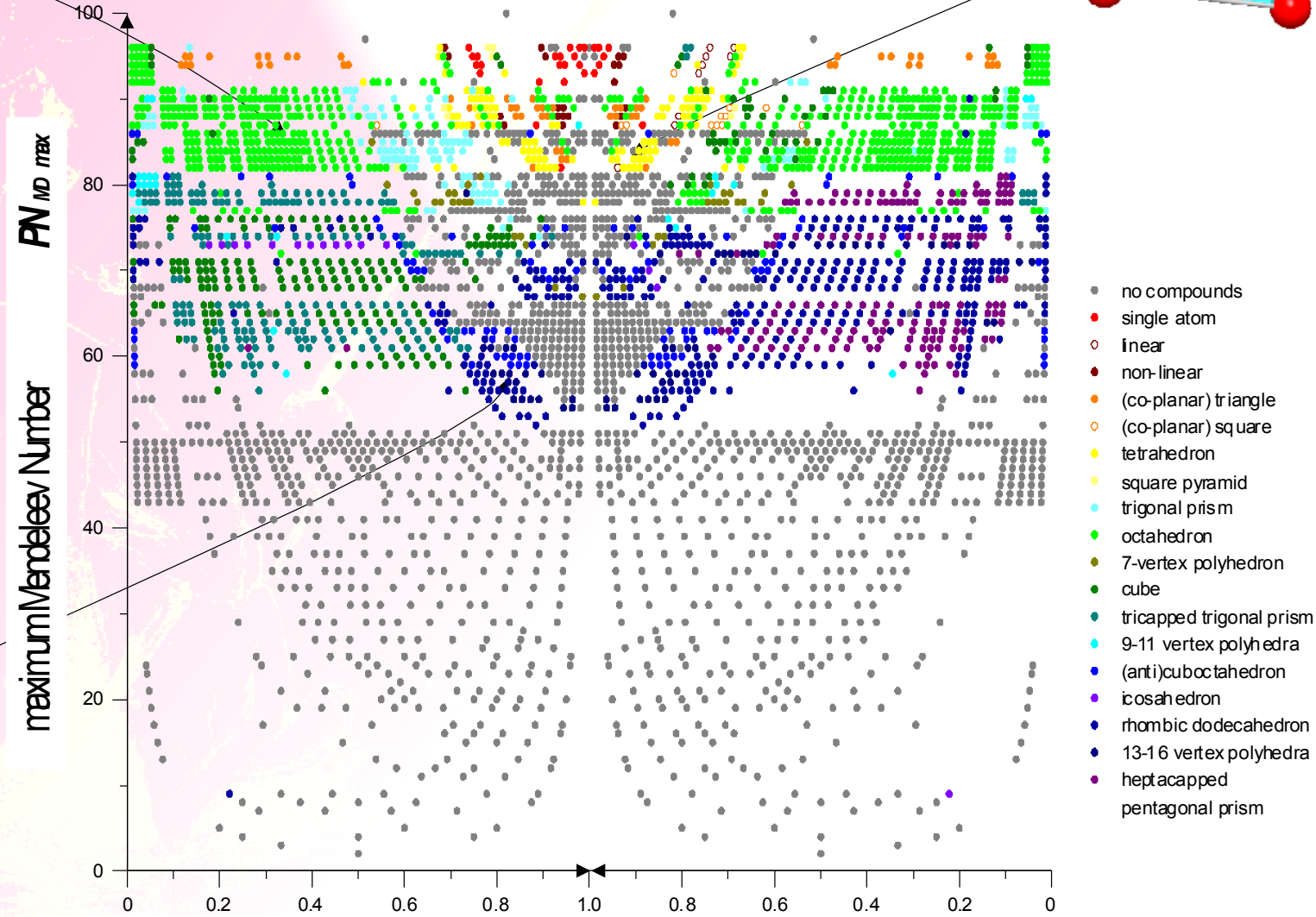
Atomic Environment Stability Maps for Equi-Atomic Binaries

Data set 1

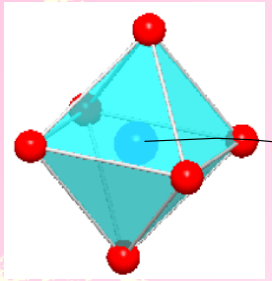








Atomic Environment Stability Maps for Equi-Atomic Binaries



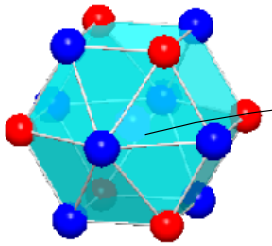
Results:

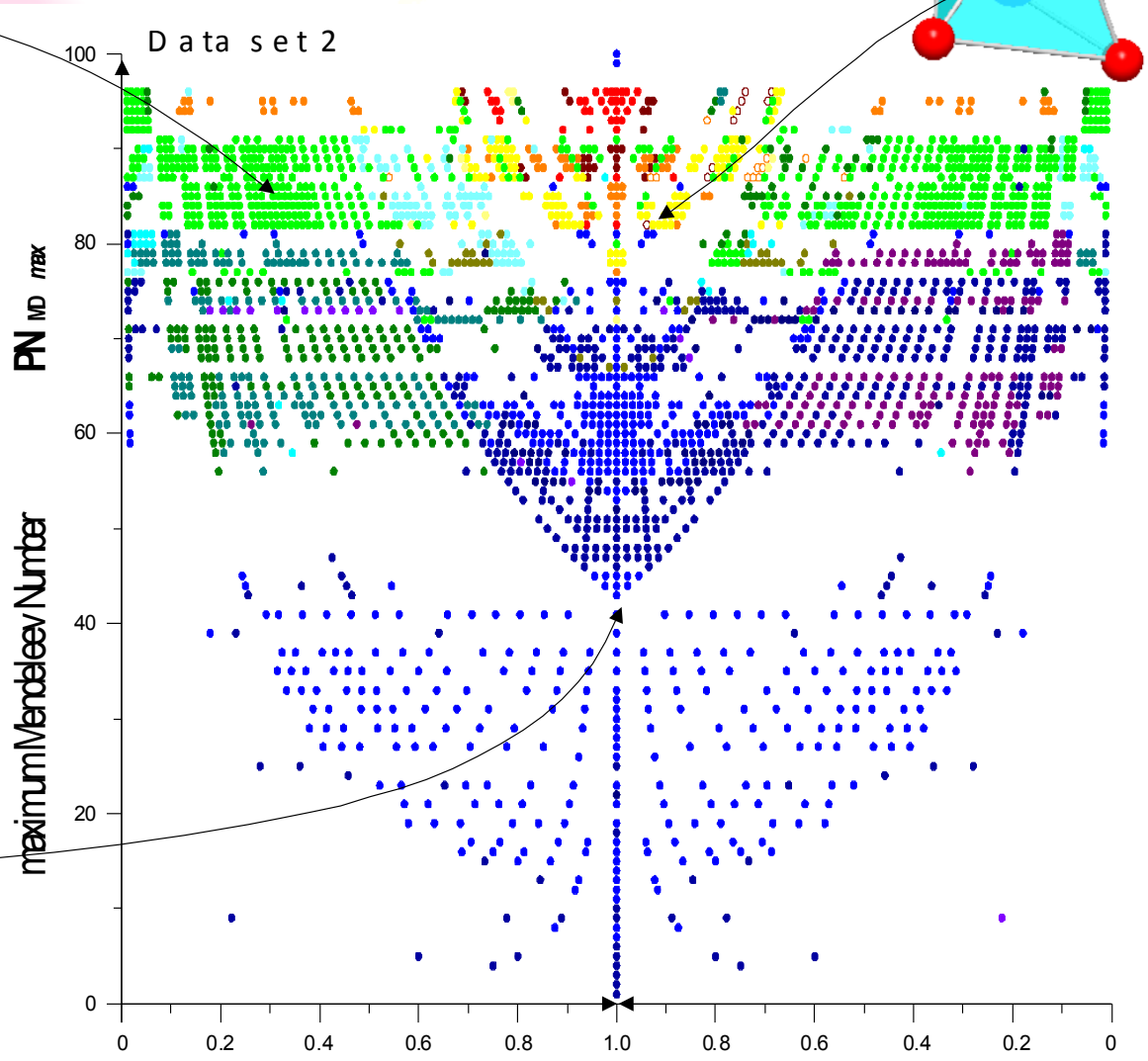
1) Only 6 of 15 AET's are populous.

2) There exist clearly defined AET domains.

3) Chemical elements prefer strongly to realize just one kind of AET with another chemical element.

4 In extremely many cases the AET of the both chemical elements A, B within the same AB compound are the same.





$PN_{MD\ min}/PN_{MD\ max}$, central atom $PN_{MD\ max}$ $MN_{MD\ min}/PN_{MD\ max}$, central atom $PN_{MD\ min}$

Materials Genome Initiative (USA) 2011-2020

Materials Informatics Initiative (Japan) 2014-2025

Materials Genome Institute (MGI-SHU, China)

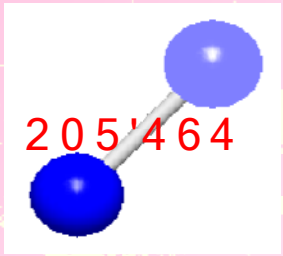
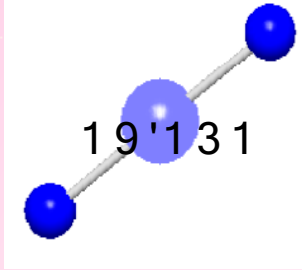
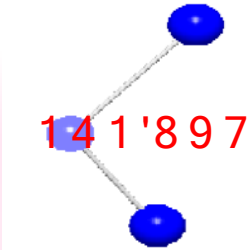
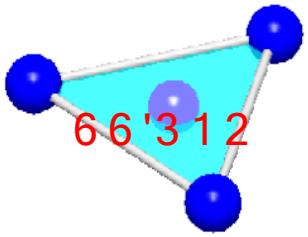
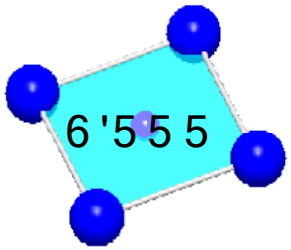
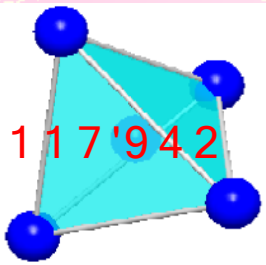
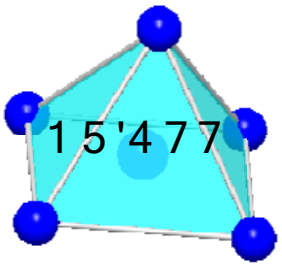
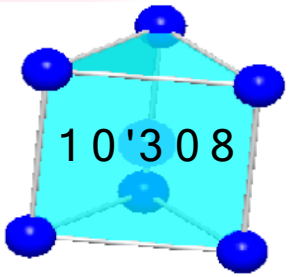
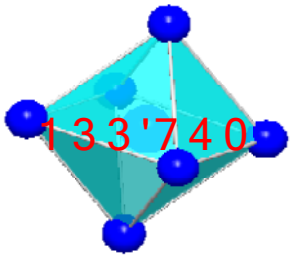
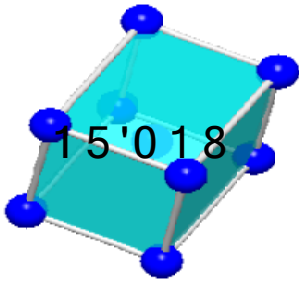
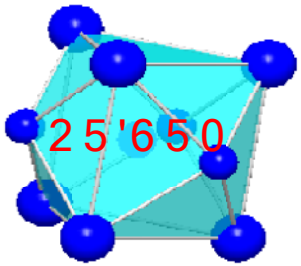
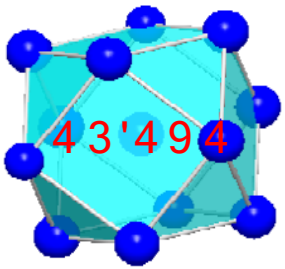
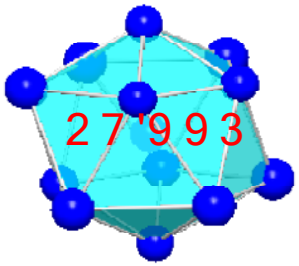
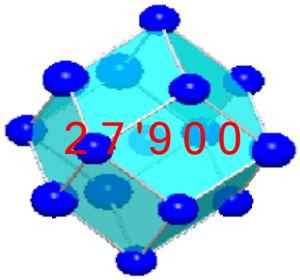
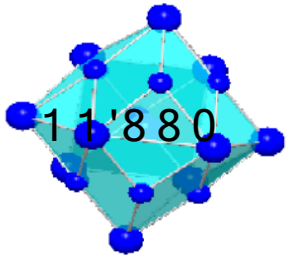
MARVEL Initiative (Switzerland) 2014-2025

Materials Genome Engineering (USTB, China)

Focusing on binary to multinary compounds using the Atomic Environment Classification

Atomic Environment Types Classification

36 most populous AET represent over 96 % of about 1'000'000 considered atoms (point-sets) equal to 320'000 data sets (37'000 different prototypes)

 205'464	 19'131	 141'897	 66'312	 6'555
single atom	collinear	non-collinear	coplanar triangle	coplanar square
 117'942	 15'477	 10'308	 133'740	 15'018
tetrahedron	square pyramid	trigonal prism	octahedron	cube
 25'650	 43'494	 27'993	 27'900	 11'880
tricapped trigonal prism	cuboctahedron	icosahedron	rhombic dodecahedron	Frank-Kasper 16-vertex

In average 28'000 different point-sets per AET

Materials Genome Initiative (USA) 2011-2020

Materials Informatics Initiative (Japan) 2014-2025

Materials Genome Institute (MGI-SHU, China)

MARVEL Initiative (Switzerland) 2014-2025

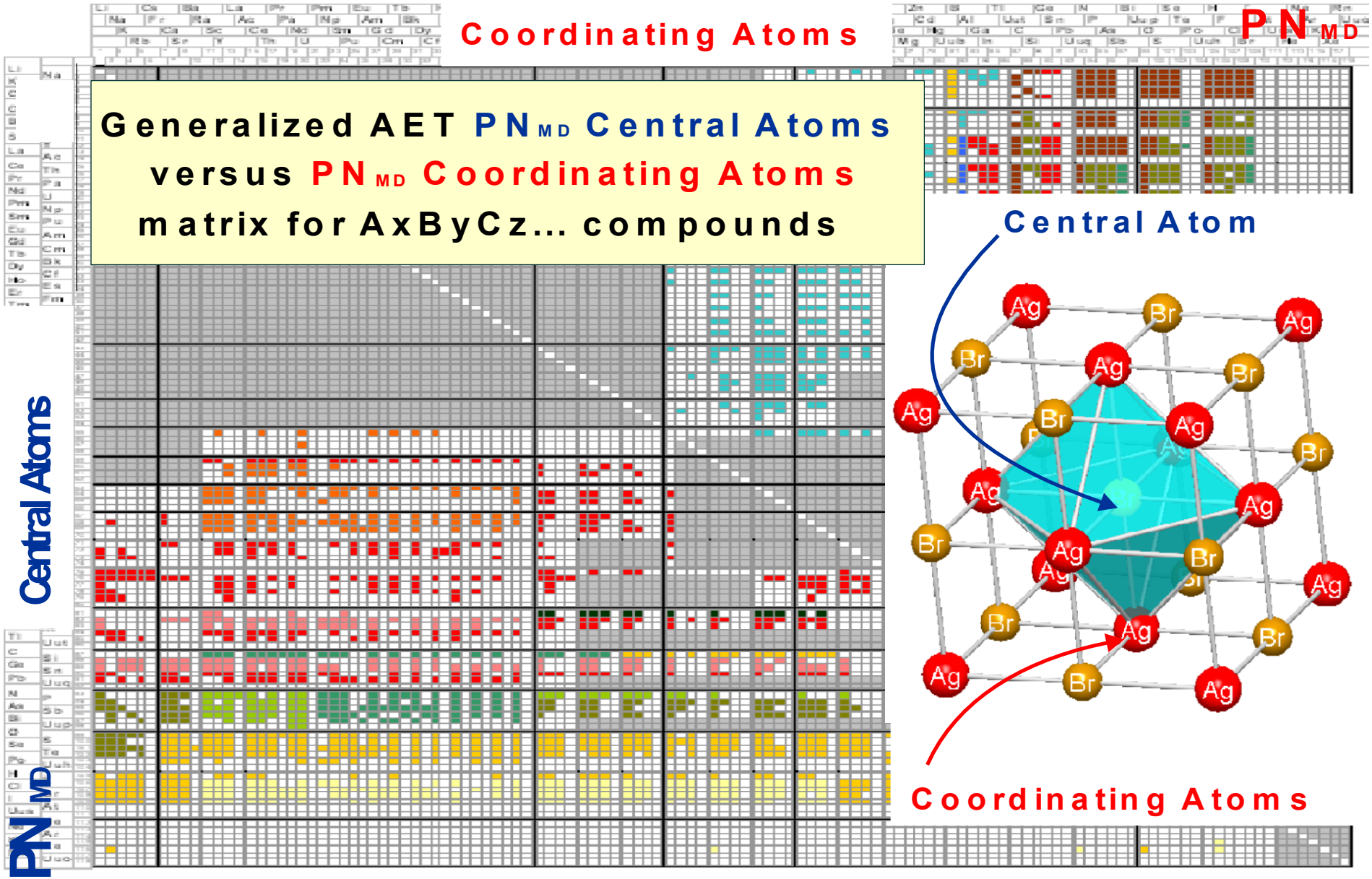
Materials Genome Engineering (USTB, China)

Even having in average 28'000 different point-sets per AET

we had to simplify the Atomic Environments Types (AET) to its Coordination Numbers (CNs) resp. sometimes its CNs ranges

(bonus)

Independent of the number of involved chemical elements, as well as of its stoichiometric ratios



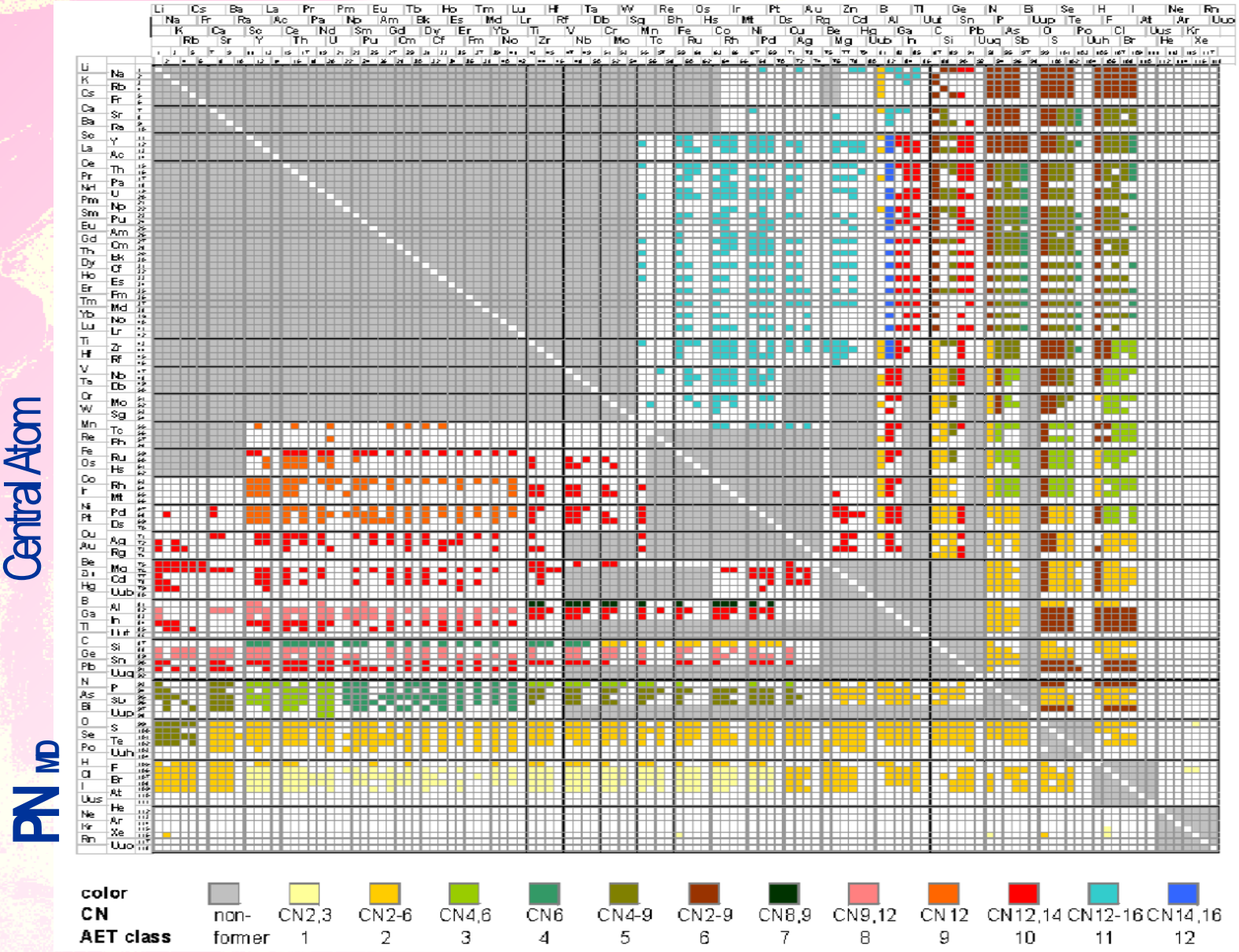
P.Villars et al., Chem. Met. Alloys 1 (2008) 210-226

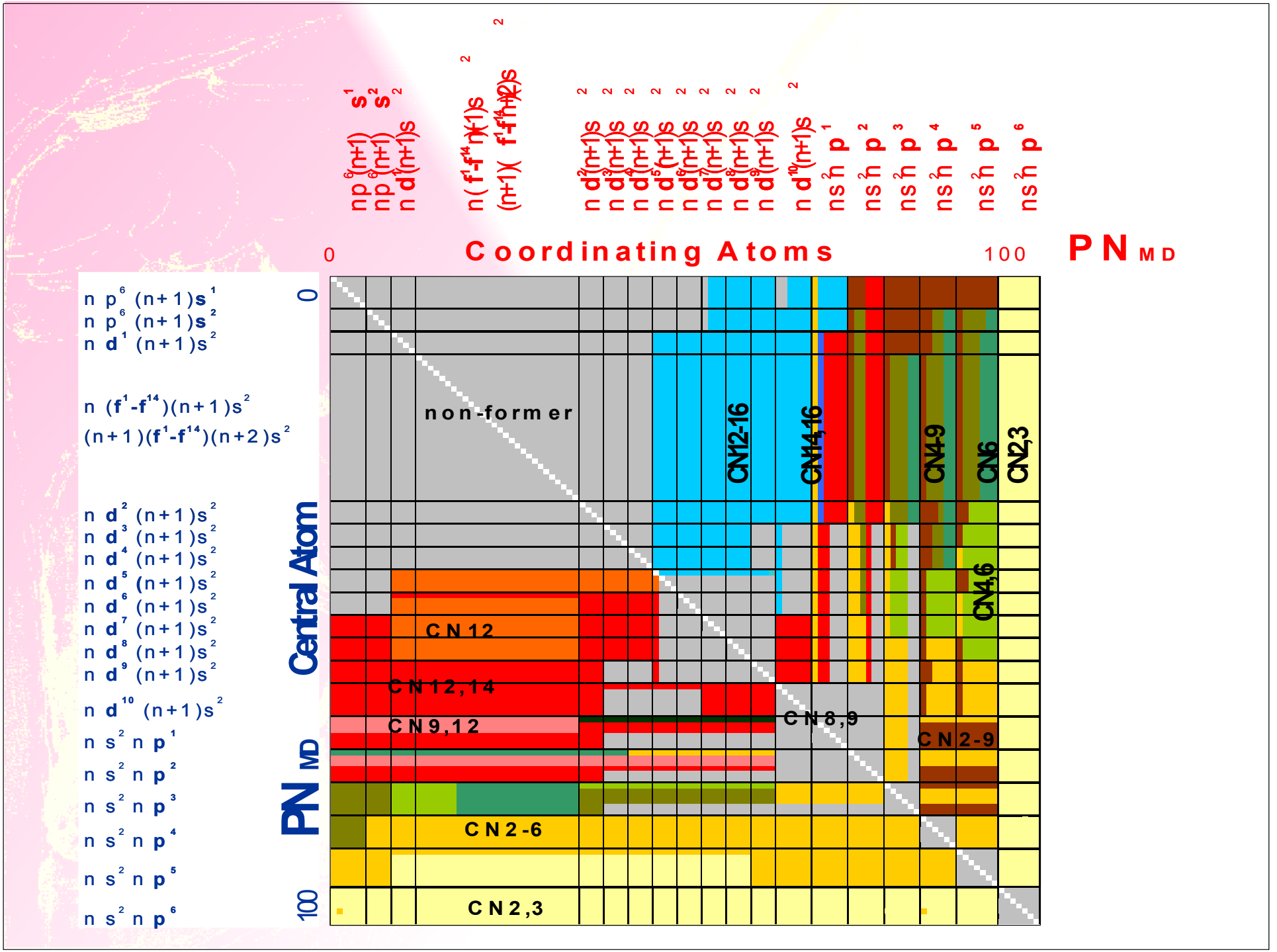
color		non-former	CN2,3	CN2-6	CN4,6	CN5	CN4-9	CN2-9	CN8,9	CN9,12	CN12	CN12,14	CN12-16	CN14,16
CN			1	2	3	4	5	6	7	8	9	10	11	12
AET class														

Experimental facts

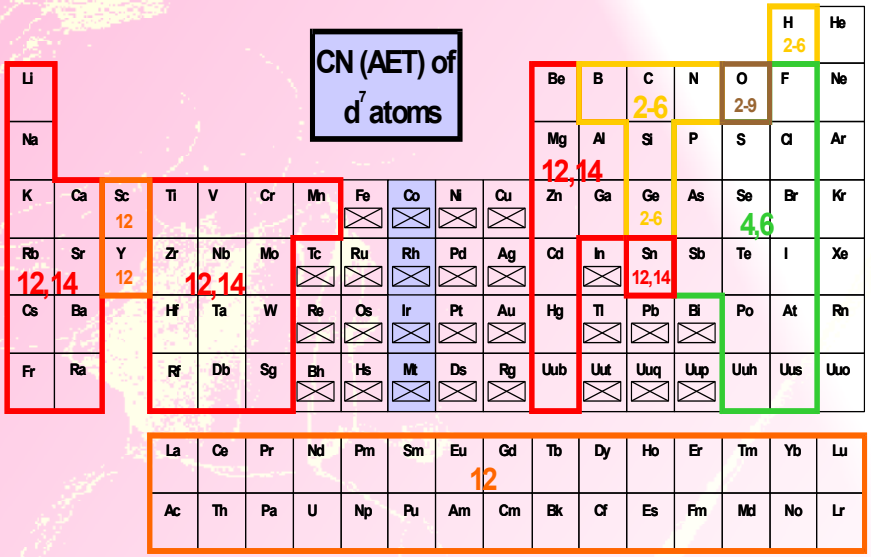
Coordinating Atoms

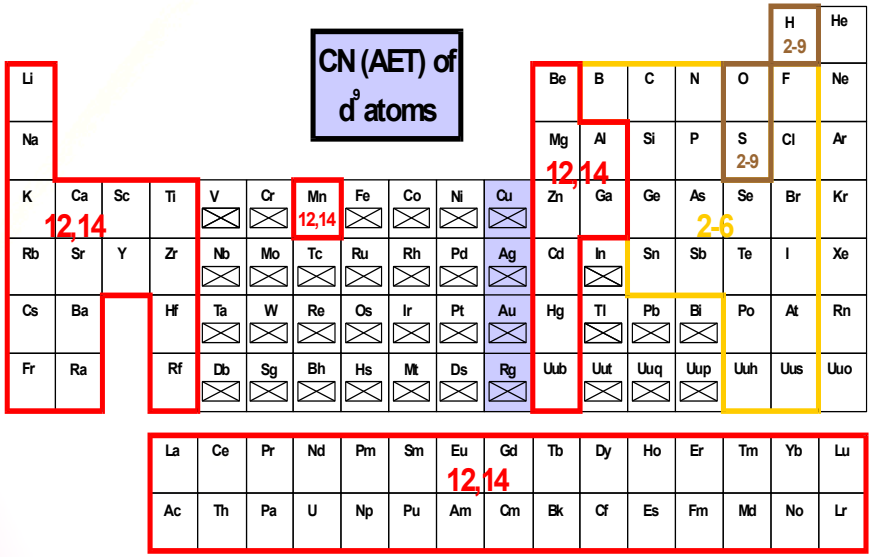
PN MD

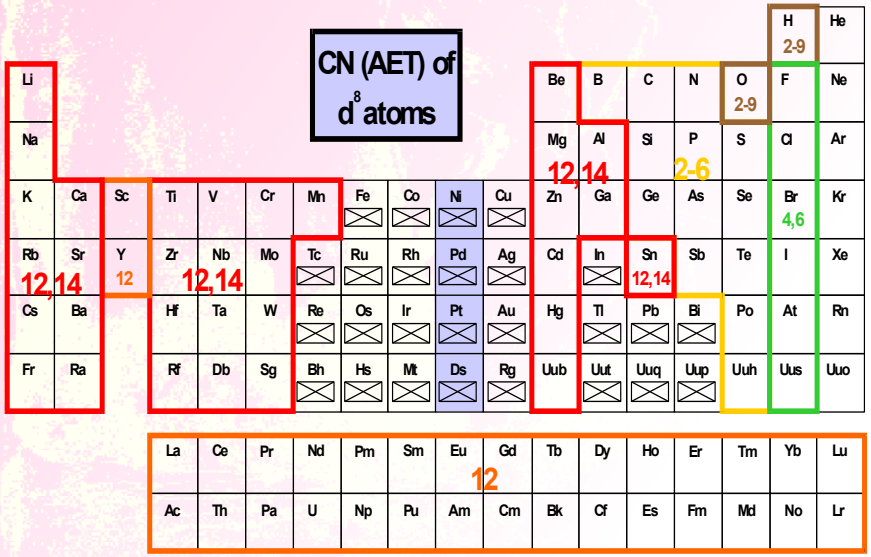


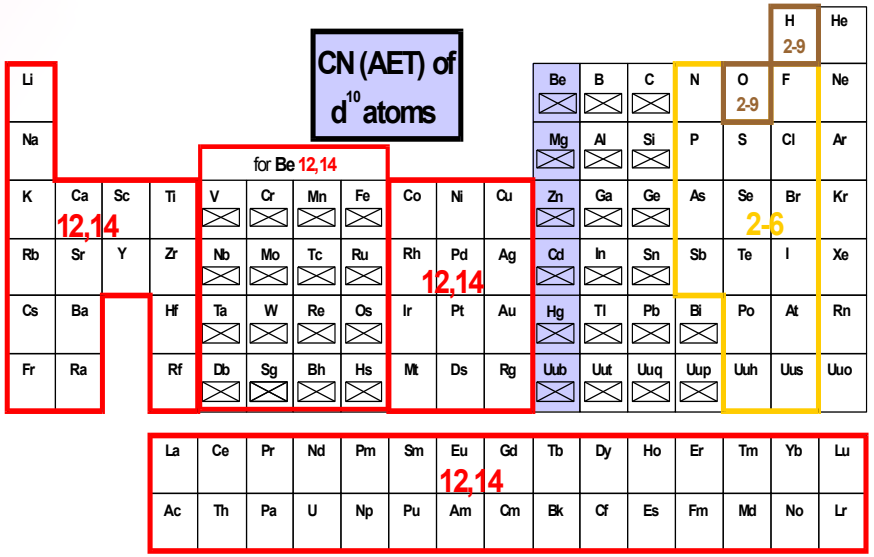

Coordinating Atoms
0
100
PN MD
Central Atom
PN MD
0
100
non-former
CN12-16
CN14,16
CN4-9
CN6
CN2,3
CN4,6
CN 12
CN 12,14
CN 9,12
CN 8,9
CN 2-9
CN 2-6
CN 2,3
n p6 (n+1) s1
n p6 (n+1) s2
n d1 (n+1) s2
n (f1-f14) (n+1) s2
(n+1) (f1-f14) (n+2) s2
n d2 (n+1) s2
n d3 (n+1) s2
n d4 (n+1) s2
n d5 (n+1) s2
n d6 (n+1) s2
n d7 (n+1) s2
n d8 (n+1) s2
n d9 (n+1) s2
n d10 (n+1) s2
ns2 np1
ns2 np2
ns2 np3
ns2 np4
ns2 np5
ns2 np6

Generalized Mendeleev Table AET graph









The following observations have been made:

General Valid:

i) The CN are independent of the stoichiometry, as well as of the number of chemical elements in the inorganic compound: they depend only on the [Central Atom A - Coordinating Atoms B] combinations.

ii) There exist 21 different groups of chemical elements acting as Central Atom A, which prefer particular CNs for different Coordinating Elements B, or do not form [Central Atom - Coordinating Atoms] combinations with B.

iii) Chemical elements with $PN_{MD} > 54$ define the CNs they are part of, regardless of wether they act as Central or as Coordinating Atoms.

iv) The observation that the non-existence of [Central Atom - Coordinating Atoms] combinations is consistent with the non-existence of a binary equi-atomic compound formed by the same chemical elements