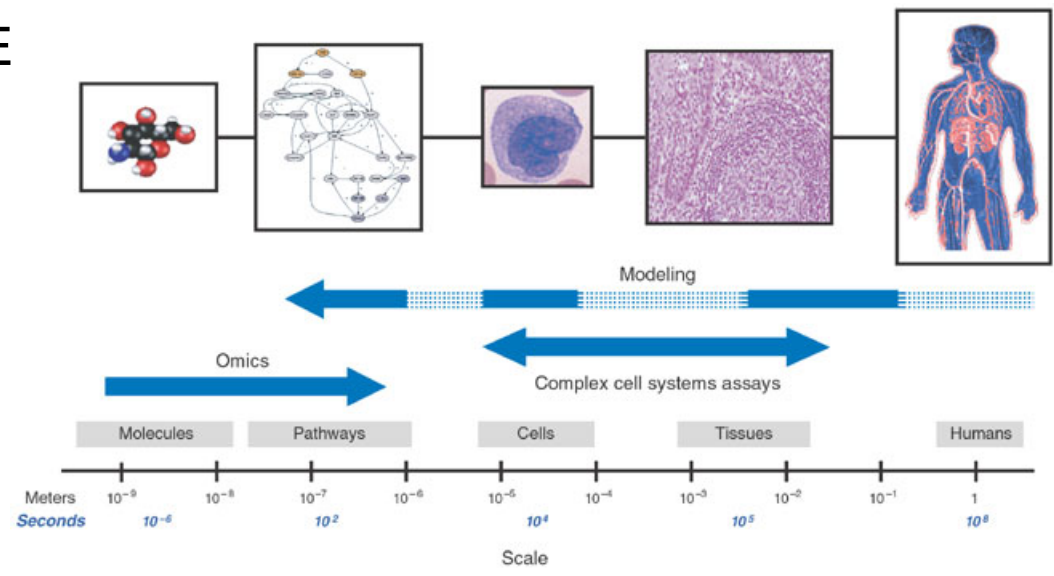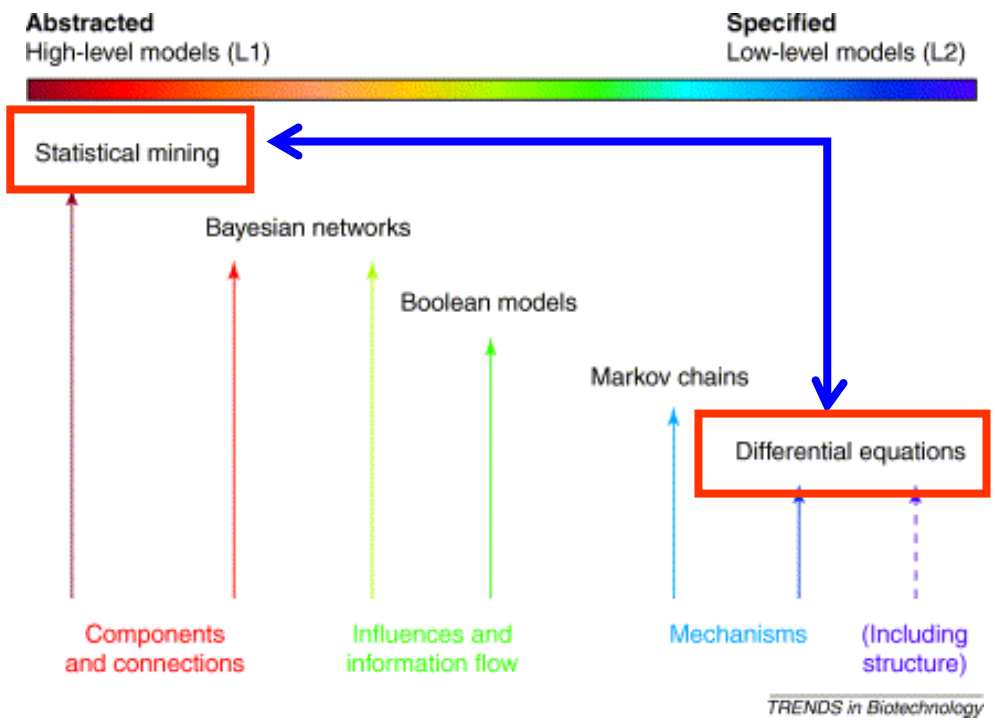# Materials Informatics: An "omics" approach to materials based design for battery technology

Krishna Rajan
Iowa State University

"OMICS" IN MATERIALS SCIENCE

Modeling

Omics

Complex cell systems assays

| Molecules | Pathways | Cells | Tissues | Humans |

| Meters | 10^{-9} | 10^{-8} | 10^{-7} | 10^{-6} | 10^{-5} | 10^{-4} | 10^{-3} | 10^{-2} | 10^{-1} | 1 |

Seconds 10^{-6} 10^{2} 10^{4} 10^{5} 10^{8}

Scale

Butcher, Berg & Kunkel Systems biology in drug discovery Nature Biotechnology (2004)

Abstracted
High-level models (L1)

Specified
Low-level models (L2)

Statistical mining

Bayesian networks

Boolean models

Markov chains

Differential equations

Components and connections

Influences and information flow

Mechanisms

(Including structure)

TRENDS in Biotechnology

Ideker and Lauffenburger: Trends in Biotechnology (2003)

NSF/MIT – EFRI Workshop: Krishna Rajan

What does it take to achieve a breakthrough?

- Discovery of mechanisms within a given length scale : fundamental materials behavior
- Linking behavior across length and time scales : materials behavior embedded in engineering system

•Both take time….acceleration is USUALLY achieved via :…

- •unexpected discovery : eg. superconducting ceramics , conducting polymers, quasicrystals, fullerenes …etc
- • failure analysis---engineering discovery: eg. ductile/ brittle transition

•Acceleration via INFORMATICS- data driven discovery : two examples

Chemical crystallography:
- • What information is important?
  - •Ranking the quality and relevance of information
    - •Classification of information
  - • Establishing a taxonomy for materials … a core principle in materials design
    - •Mining materials attributes to find associations : eg. descriptions of structure-property relationships
- • Prediction of new information
  - • Materials equivalent of "soft modeling" in bioinformatics

Microarrays for combinatorial materials chemistry
- •Role of visualization …another informatics tool

NSF-IMI: International Materials Institute
CoSMIC
Combinatorial Sciences and Materials Informatics Collaboratory
NSF
NSF/MIT – EFRI Workshop: Krishna Rajan
IOWA STATE UNIVERSITY
OF SCIENCE AND TECHNOLOGY

DATA INPUTS

- Crystallographic descriptors / semi-empirical electronic structure parameters / property data
- Descriptors for each element in compound…weighted by stoichiometry?

DATA MINING

Structure classification
- PCA bi –plots, SVM, Frequency Association Mining & Clustering analysis
: **New structure maps/ classifications:**
- PCA + GA + NN hybrid techniques : **Classification structural distortions**

- **Establish design rules:**
- **Recursive partioning using Shannon entropy criterion for classification: "if-then rules" / Recursive partioning for prediction**
- PLS: **Discovery of new descriptors**

Structure-property prediction
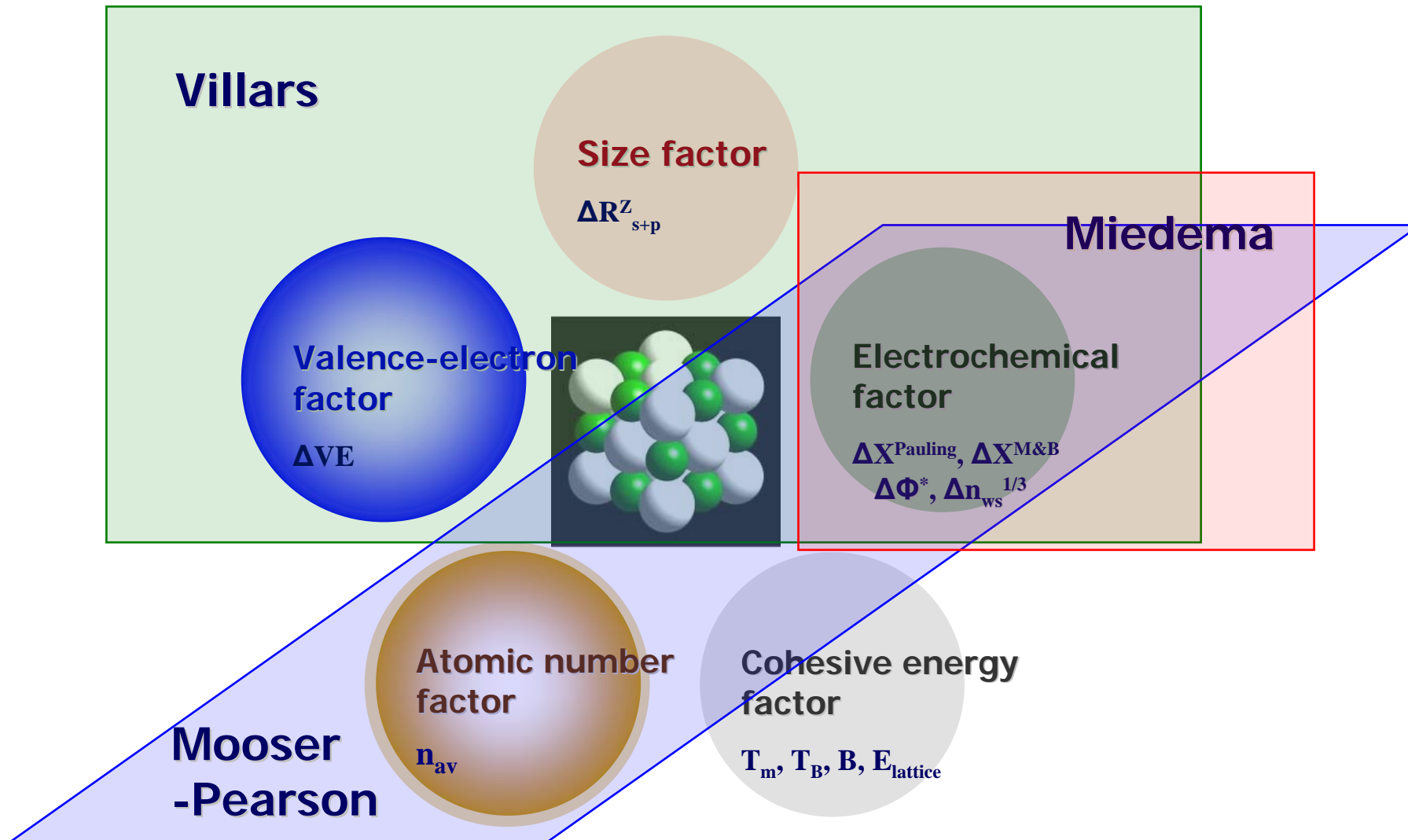- PLS & SVM- **QSAR for inorganic crystals**

Structure prediction:
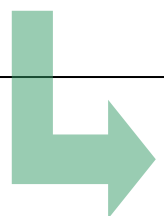- **Screen** clusters of compounds from above procedures- down select those for **energy** calculations

Reduce dimensionality : **PCA**

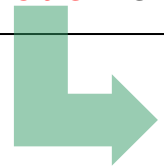Rank & reduce descriptors**: loading plots &entropy calculations**

NSF-IMI: International Materials Institute

CoSMIC

Combinatorial Sciences and Materials Informatics Collaboratory

NSF/MIT – EFRI Workshop: Krishna Rajan

IOWA STATE UNIVERSITY
OF SCIENCE AND TECHNOLOGY

DESCRIPTOR DEVELOPMENT: Building on Empirical Design Rules

Villars

Size factor

ΔR^Z_{s+p}

Miedema

Valence-electron factor

ΔVE

Electrochemical factor

$\Delta X^{Pauling}, \Delta X^{M\&B}$
$\Delta \Phi^*, \Delta n_{ws}^{1/3}$

Atomic number factor

n_{av}

Cohesive energy factor

$T_m, T_B, B, E_{lattice}$

Mooser -Pearson

Information entropy (S) is a criterion to determine the position of hyperplanes to maximize the information on the content of each compartment.

$$S = -k \sum p_i \log p_i$$

Information-entropy change
(Information gain, ΔS)

$$\Delta S = | S_{ascendant} - S_{descendant} |$$
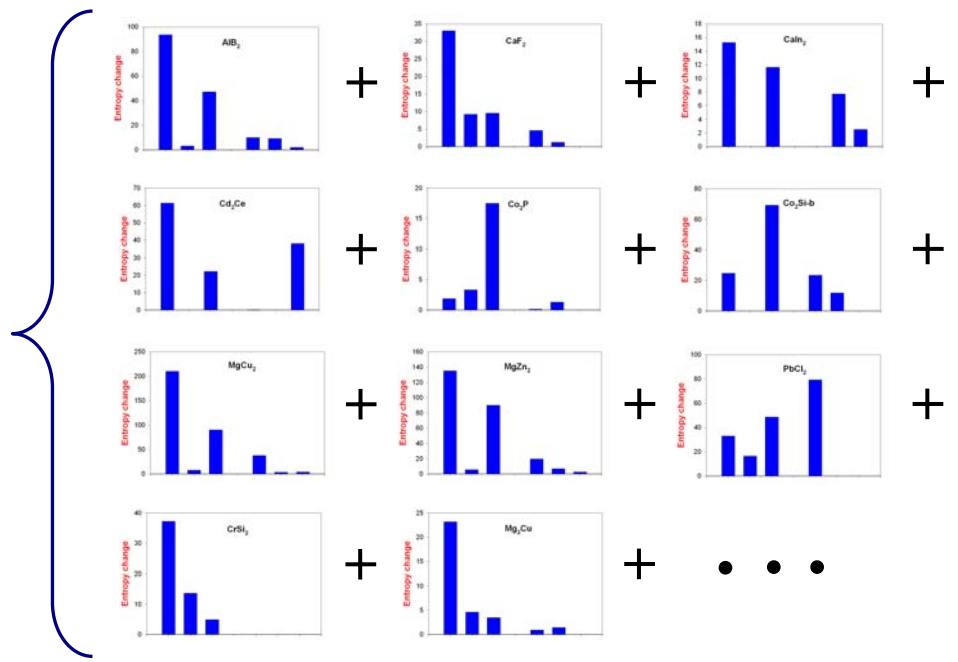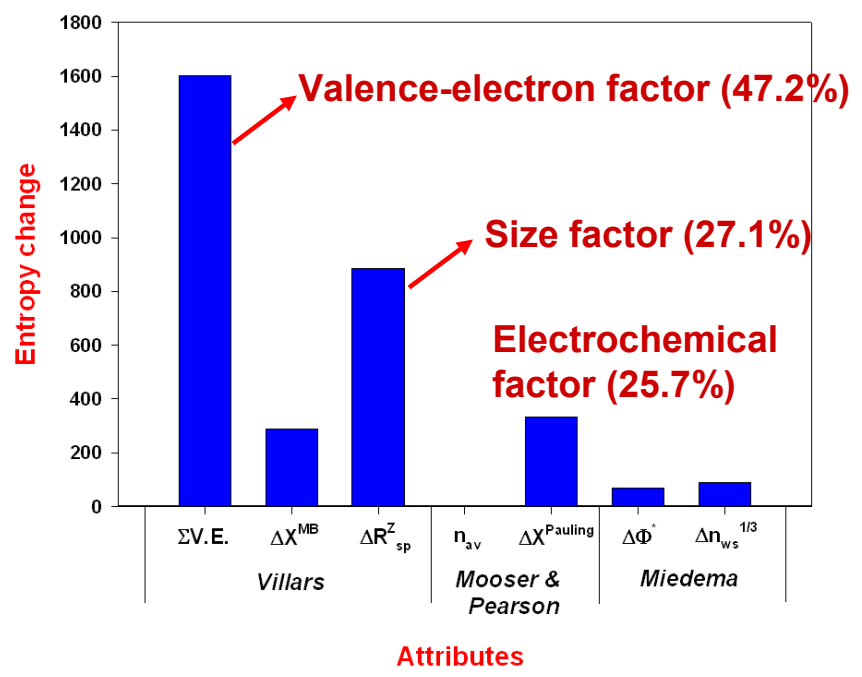
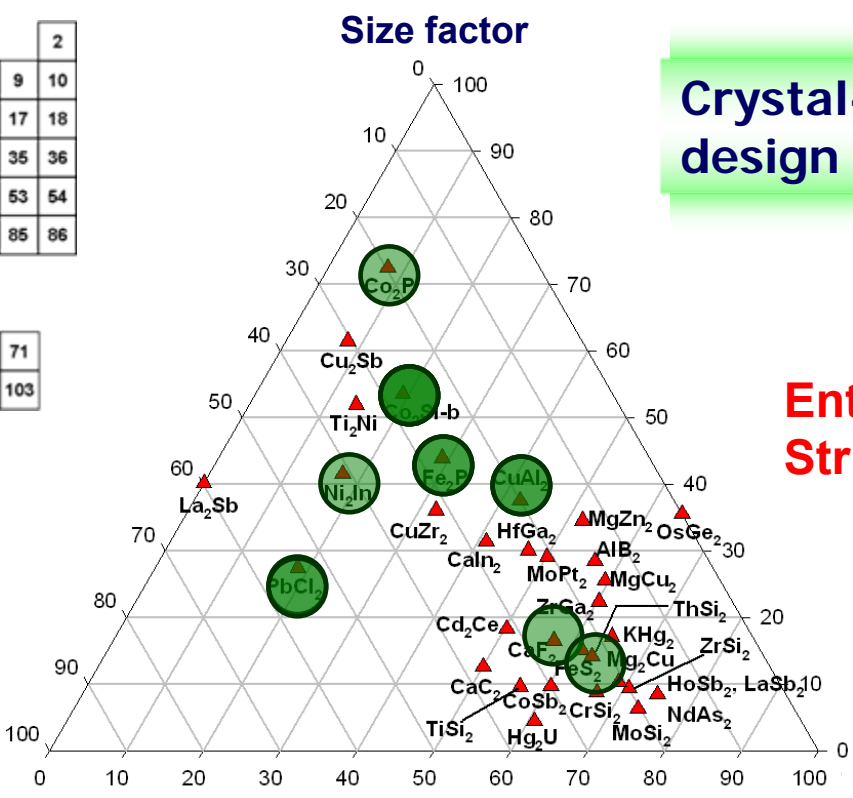**Valence-electron factor (47.2%)**

Size factor (27.1%)

Electrochemical factor (25.7%)

Entropy change

ΣV.E. ΔX^{MB} ΔR^{Z}_{sp} n_{av} $\Delta X^{Pauling}$ $\Delta \Phi^{*}$ $\Delta n_{ws}^{1/3}$

Villars   *Mooser & Pearson*   *Miedema*

Attributes

ΔS= ∑(34 Structure types)

NSF/MIT – EFRI Workshop: Krishna Rajan

NSF-IMI: International Materials Institute

CoSMIC
Combinatorial Sciences and Materials Informatics Collaboratory

IOWA STATE UNIVERSITY
OF SCIENCE AND TECHNOLOGY

Relative contribution of three governing factors was quantitatively evaluated for the respective structure types in terms of the change of information entropy.

GeX$_2$
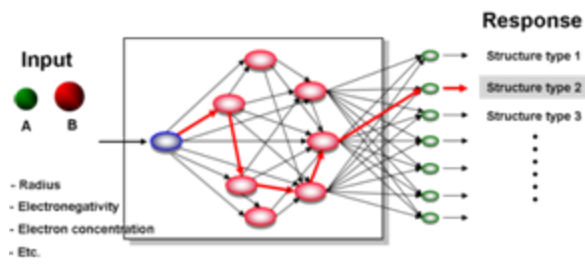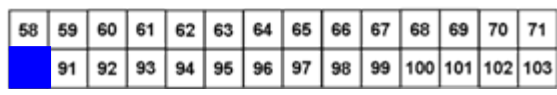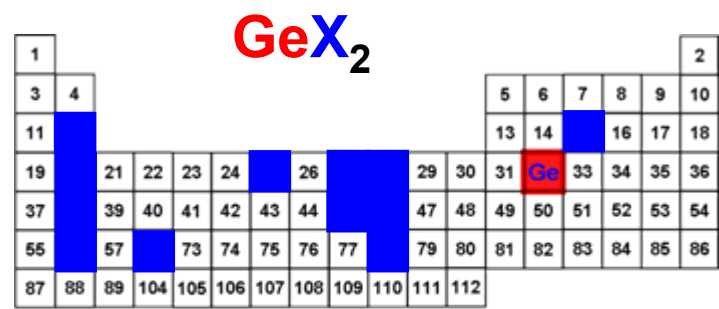
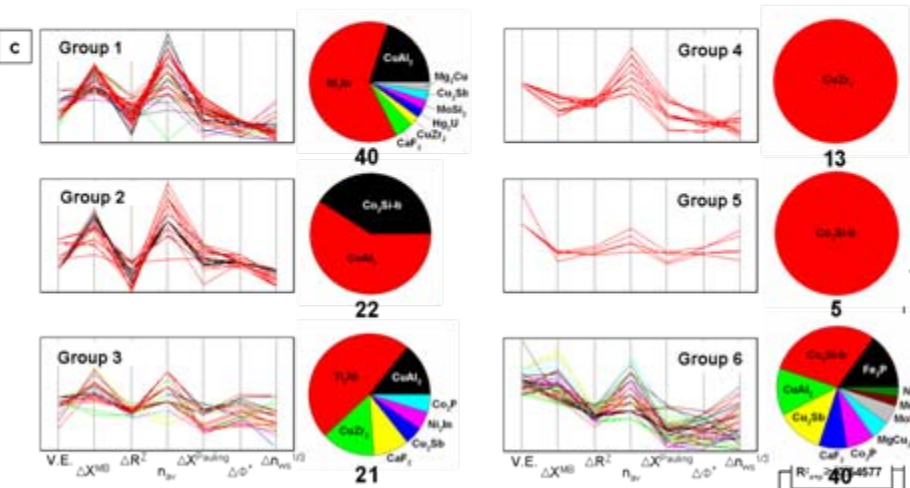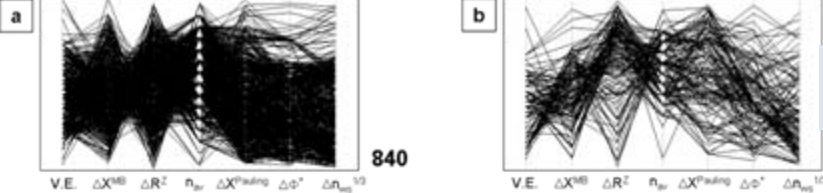**Size factor**

Crystal-structure design rules

Entropy scaled Structure map

Electrochemical factor

Valence-electron factor

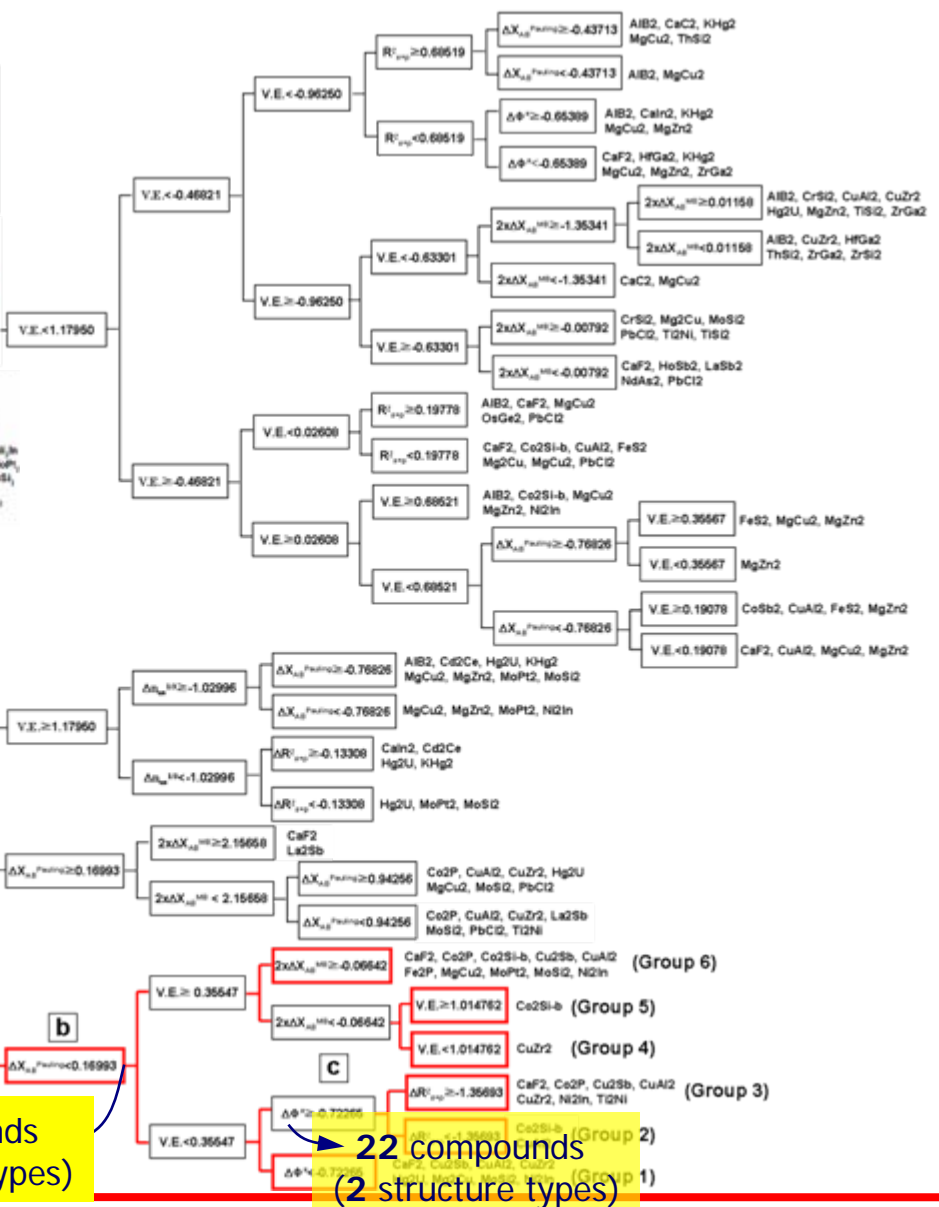DATA MINING : tracking structural correlation

Tracking crystallographic Statistics in databases

Recursive partioning to track Evolution of design rules

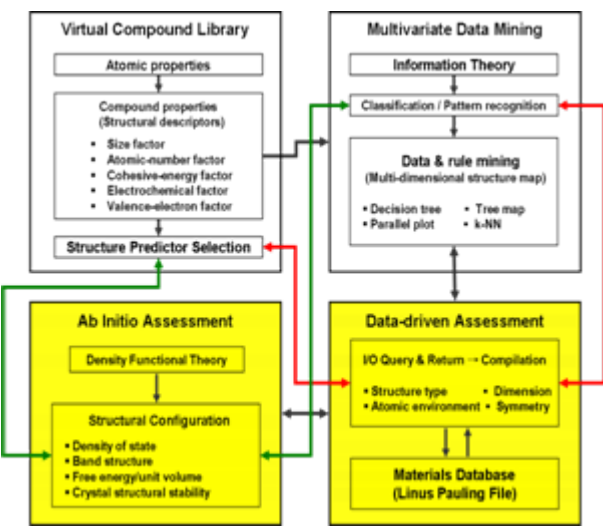**840** compounds (**34** structure types)

140 compounds (**14** structure types)
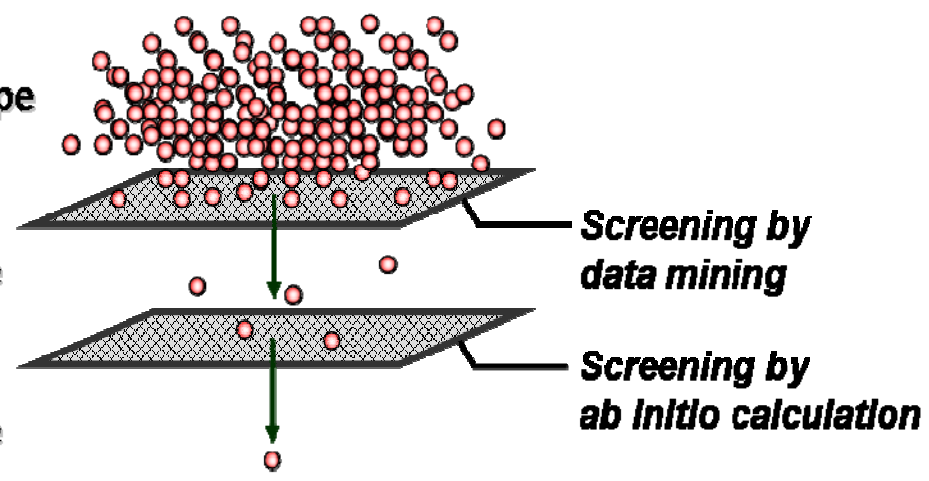
22 compounds (**2** structure types)

NSF-IMI: International Materials Institute
CoSMIC
Combinatorial Sciences and Materials Informatics Collaboratory

NSF/MIT – EFRI Workshop: Krishna Rajan

IOWA STATE UNIVERSITY
OF SCIENCE AND TECHNOLOGY

- **Two-stage Screening process**

Virtual Compound Library
- Atomic properties
- Compound properties (Structural descriptors)
  - Size factor
  - Atomic-number factor
  - Cohesive-energy factor
  - Electrochemical factor
  - Valence-electron factor
- Structure Predictor Selection

Multivariate Data Mining
- Information Theory
- Classification / Pattern recognition
- Data & rule mining (Multi-dimensional structure map)
  - Decision tree
  - Tree map
  - Parallel plot
  - k-NN

Ab Initio Assessment
- Density Functional Theory
- Structural Configuration
  - Density of state
  - Band structure
  - Free energy/unit volume
  - Crystal structural stability

Data-driven Assessment
- I/O Query & Return → Compilation
  - Structure type
  - Dimension
  - Atomic environment
  - Symmetry
- Materials Database (Linus Pauling File)

109 structure-type candidates

5 structure-type candidates

1 structure-type candidate

Screening by data mining

Screening by ab initio calculation

INPUT:

AB_2, A=Au, B=Be

$\Delta X^{Pauling} = 0.85978$, $\Sigma VE = -0.30361$
$\Delta R^z_{s+p} = 0.61403$, $\Delta n_{av} = -1.61573$
$\Delta n_{ws}^{1/3} = -0.93887$, $\Delta X^{M\&B} = -0.14441$
$\Delta \Phi^* = -1.68529$

Start

$R^z_{s+p} \geq -0.54577$

$R^z_{s+p} < -0.54577$

V.E. < 1.17950

V.E. ≥ 1.17950

V.E. < -0.46821

V.E. ≥ -0.46821

V.E. < 0.02608

V.E. ≥ 0.02608

$R^z_{s+p} \geq 0.19778$

$R^z_{s+p} < 0.19778$

AlB_2
CaF_2
$MgCu_2$
$OsGe_2$
$PbCl_2$
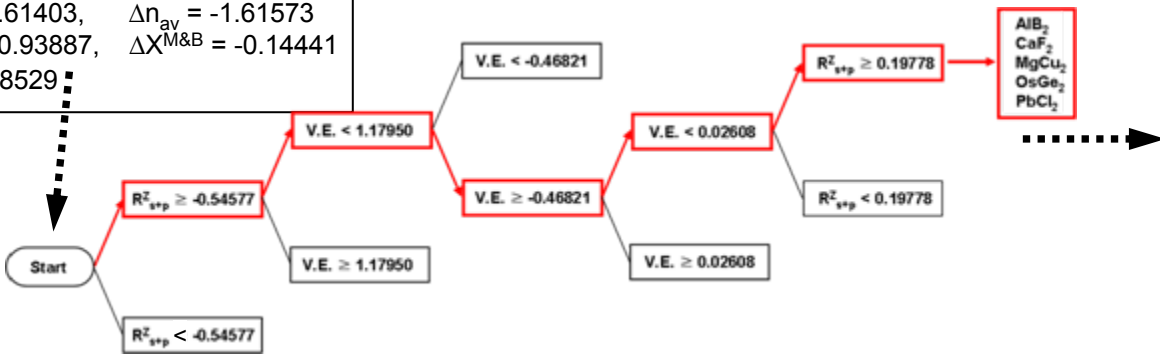
OUTPUT: Structure type candidates list

1. $MgCu_2$ (-3.65757 eV)
2. $PbCl_2$ (-3.60992 eV)
3. $OsGe_2$ (-3.58157 eV)
4. CaF_2 (-3.46498 eV)
5. AlB_2 (-3.46430 eV)

NSF-IMI: International Materials Institute

CoSMIC
Combinatorial Sciences and Materials Informatics Collaboratory

NSF

NSF/MIT – EFRI Workshop: Krishna Rajan

IOWA STATE UNIVERSITY
OF SCIENCE AND TECHNOLOGY

INFORMATICS STRATEGY:
QSAR…following the biologists

$$\text{Functionality} = \mathcal{F} \, (\, X_1 \, , \, X_2 \, , \, X_3 \, , \, X_4 \, , \, X_5 \, , \, X_6 \, , \, X_7 \, , \, X_8 \, \ldots\ldots$$

Issues:

• how many variables?
• which variables are important?
• classify behavior among variables
• making quantitative predictions …relate functionality to variables …
    • traditionally we describe them by empirical equations:
    •Quantitative Structure Activity Relationships (QSARs) are derived from
       data mining techniques not assuming a priori which physics is the most important
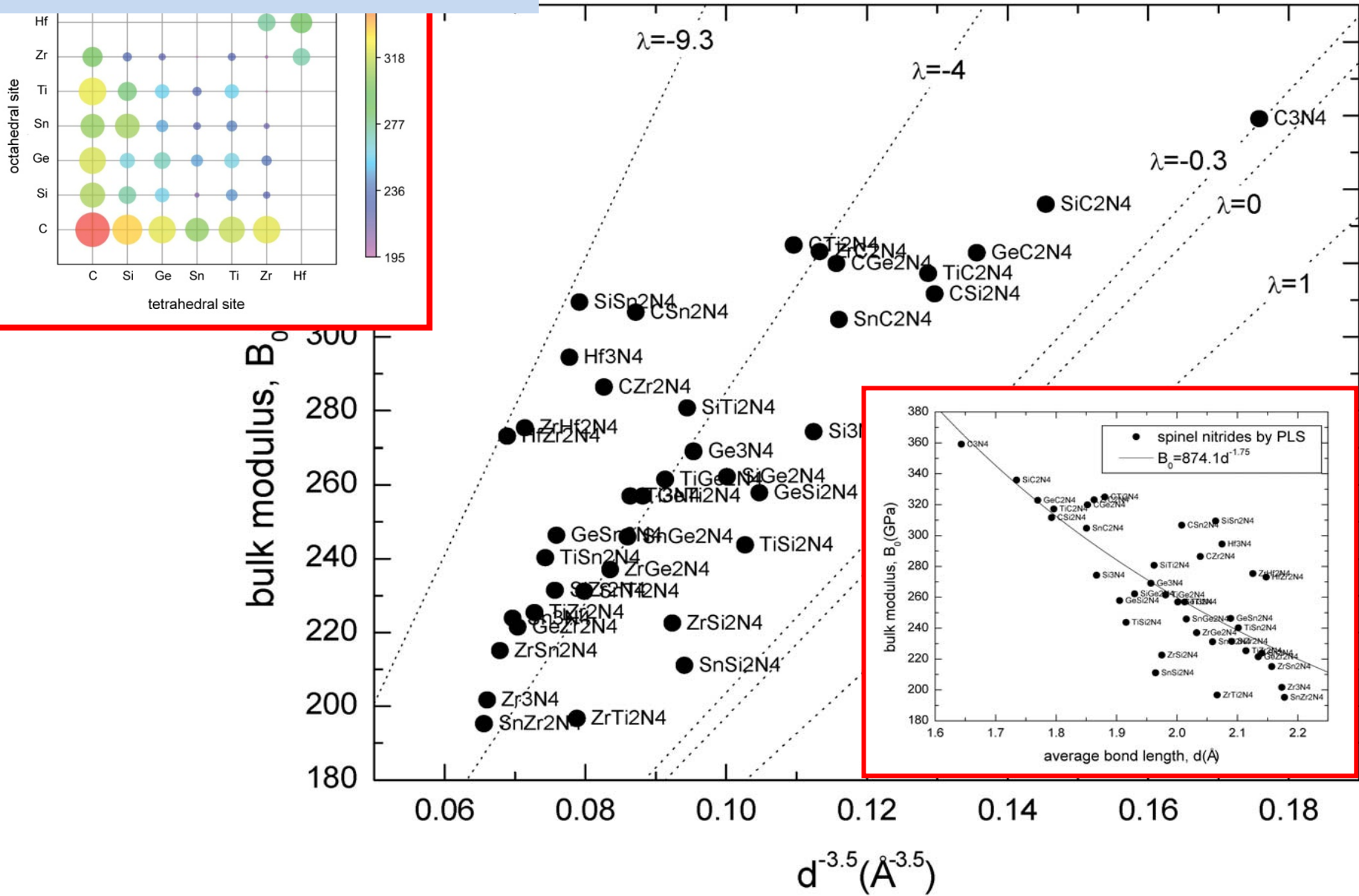
Need to build database with these variables

NSF-IMI: International Materials Institute
CoSMIC
Combinatorial Sciences and Materials Informatics Collaboratory
NSF

IOWA STATE UNIVERSITY
OF SCIENCE AND TECHNOLOGY

"VIRTUAL LIBRARIES" via INFORMATICS

Bulk Modulus(GPa) ▼

Virtual library:
via informatics

Refractory metals

"Real" library:
via first principles calculations
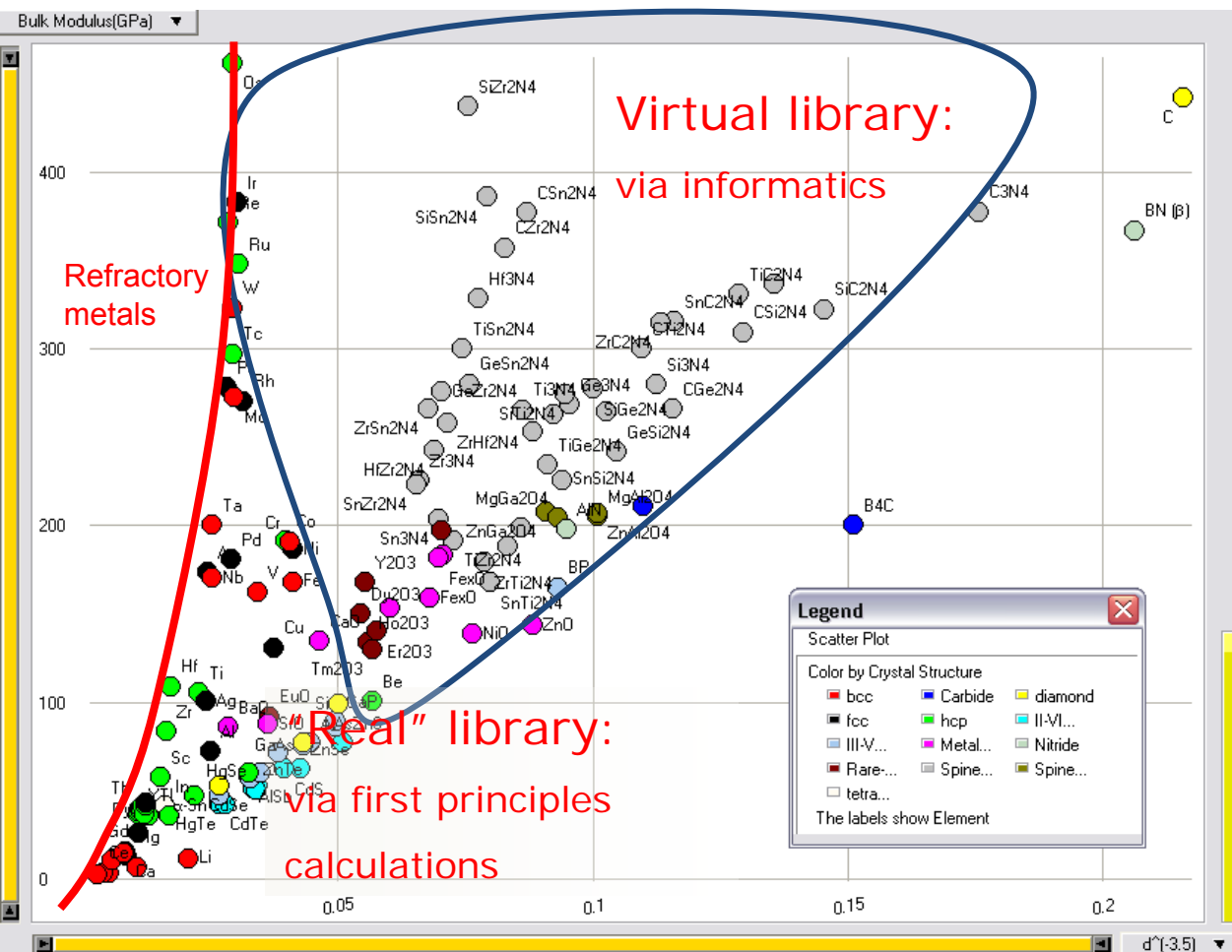
Suh and Rajan (2005, 2006)

Broderick & Rajan (2008)

Legend

Scatter Plot

Color by Crystal Structure

bcc	Carbide	diamond
fcc	hcp	II-VI...
III-V...	Metal...	Nitride
Rare-...	Spine...	Spine...
tetra...		

The labels show Element

covalent

nitrides

carbides   oxides

borides   Silicides

metallic

ionic

Stoichiometric Effects

Degree of tetragonal distortion
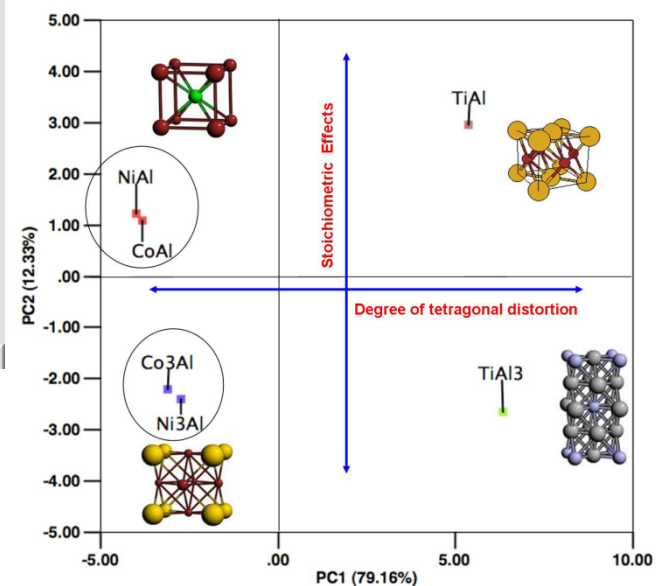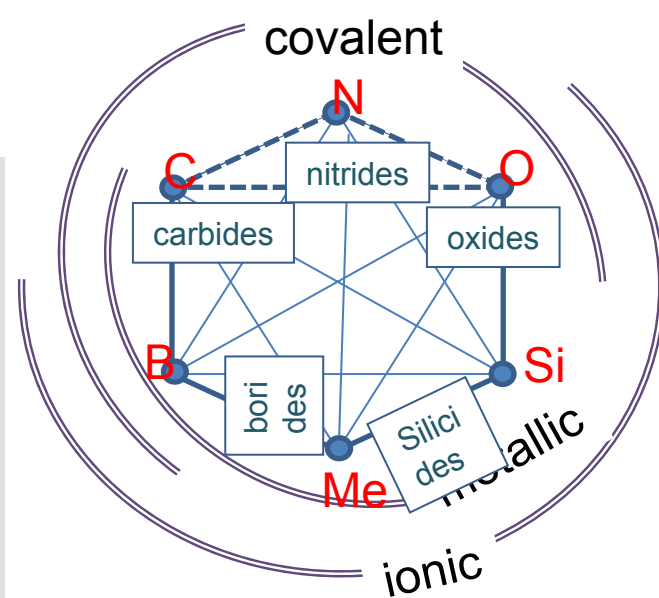
CRYSTAL CHEMISTRY DESIGN

INFORMATICS DERIVED SCALING LAWS

Cohen	I-VII rocksalt compounds	$B=550d^{-3}$		
Cohen	Tetrahedral semiconductors	$B=(N_c/4)(1972-220\lambda)d^{-3.5}$		
Al-Douri *et al.*	Tetrahedral semiconductors	$B=(3000-100\lambda)(a/2)^{-3.5}$		
Sung *et al.*	Diamond like semiconductors	$B=9.75P^{-0.0448}C^{0.0462}d^{0.423}$		
Makino *et al.*	Elements	$B=Cd^{-m}$		
		sp^3 bonding	$C=2062$	$m=-3.57$
		spd bonding	$C=3702.6$	$m=-4.33$
		$3d$ bonding	$C=23012$	$m=-5.27$
		$4d$ bonding	$C=163120$	$m=-6.64$
		$5d/4f$ bonding	$C=73170$	$m=-7.81$
Ching *et al.*	18 spinel nitrides	$B=919d^{-1.86}$		
Informatics Aided Design	**39 spinel nitrides**	$B=812.9d^{-1.68}$ QSAR		

DEVELOPING the SEARCH SPACE: cathode/anode materials

General formula
$A_{10}(BO_y)_6X_2$

A site cations: larger
1. Divalent: Ca^{2+}, Sr^{2+}, Pb^{2+}, Ba^{2+}, Eu^{2+}, La^{2+} etc.
2. Monovalent: Na^+, Li^+ etc.
3. Trivalent: Y^{3+}, La^{3+}, Ce^{3+}, Nd^{3+}, Sm^{3+}, Dy^{3+} etc.

B site cations: smaller
1. Metals: Re, V, Mn, Cr, Ge etc.
2. Metalloids: P, As, B, Si etc.

X site anions:
1. Halides: F^-, Cl^-, Br^-, I^-.
2. Others: OH^-, O_2, N_2, CO_3 etc.
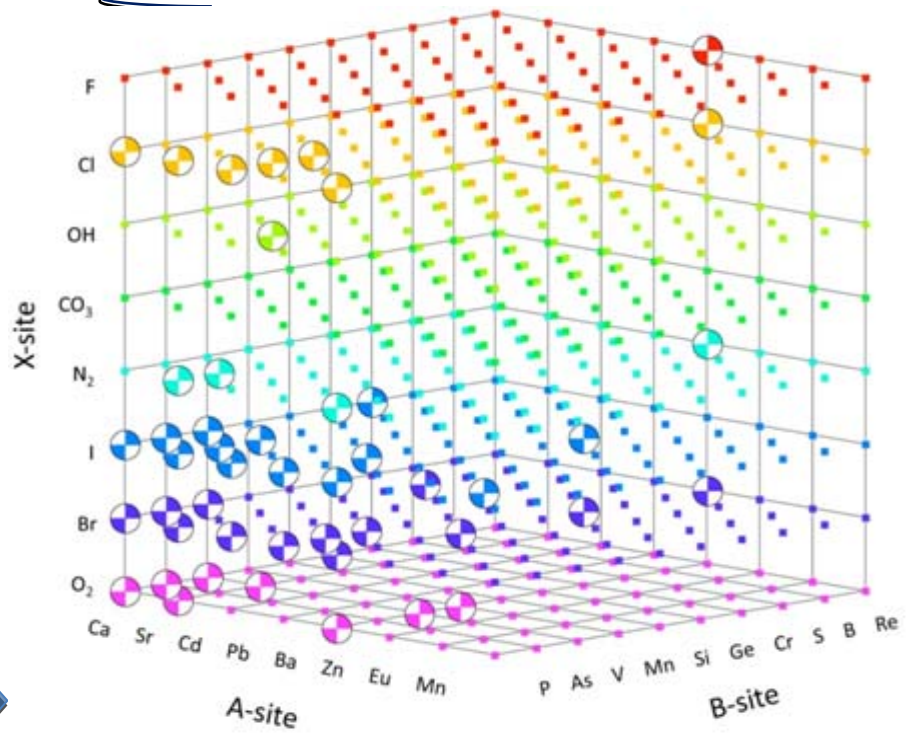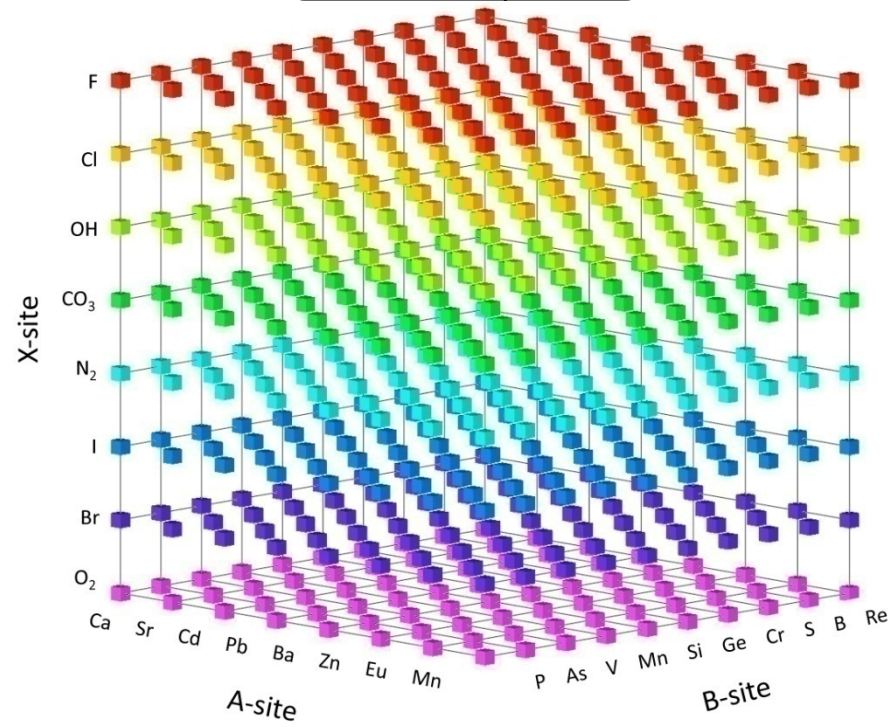
Approximately **640 possible** simple apatites ($A_{10}(BO_4)_6X_2$) and of that some **commonly known** apatites (~47)

NARROWING THE SEARCH SPACE: structural building units

Search Space
$A_{10}(BO_4)_6X_2$

Channel structure
ex. $Ca_{10}(PO_4)_6F_2$

$A(I)O_6$ polyhedra

BO_4 tetrahedra
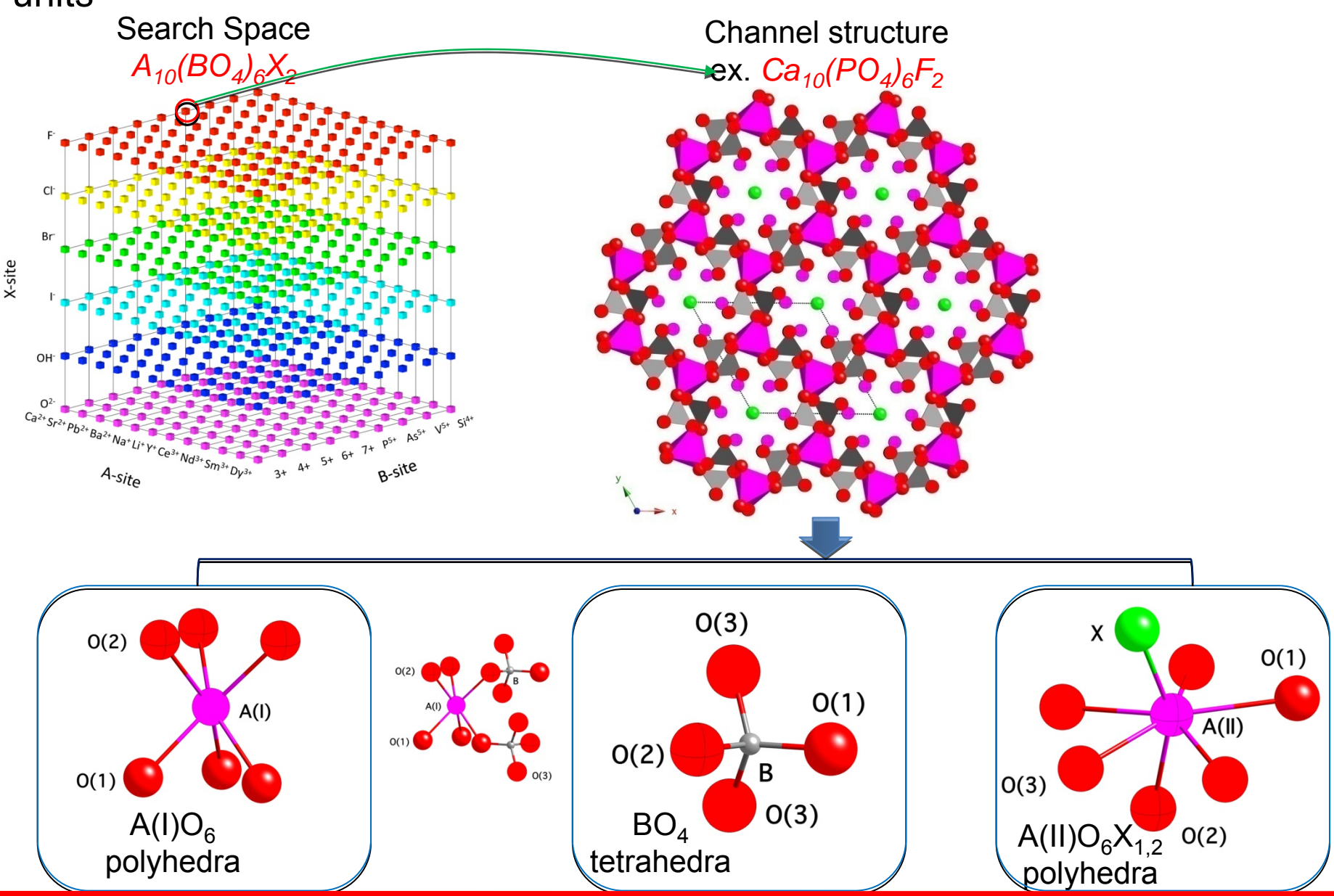
$A(II)O_6X_{1,2}$ polyhedra

2 PRIMARY crystal parameters & Wyckoff positions

16 SECONDARY crystal parameters (bond lengths and angles)

Unit cell parameters
2. a (Å)
3. c (Å)

Atomic coordinates
4. $A(I)_z$ 5. $A(II)_x$ 6. $A(II)_y$ 7. B_x 8. B_y 9. $O(1)_x$ 10. $O(1)_y$ 11. $O(2)_x$ 12. $O(2)_y$ 13. $O(3)_x$ 14. $O(3)_y$, 15. $O(3)_z$ 16. X_z

Energy of apatite
1. E_{total} (eV/unit cell)

17. $d_{A(I)-O(1)}$ (Å)
18. $d_{A(I)-O(1)}^{A(I)z=0}$ (Å)
19. $\Delta_{A(I)-O}$ (Å)
20. $\Delta^{A(I)z=0}_{A(I)-O}$ (Å)
21. $\angle\psi_{A(I)-O(1)}$ ($^\circ$)
22. $\angle\psi^{A(I)Z=0}_{A(I)-O(1)}$ ($^\circ$)
23. $\angle\delta_{A(I)}$ ($^\circ$)
24. $\angle\varphi_{A(I)}$ ($^\circ$)
25. $\angle\alpha_{A(I)}$ ($^\circ$)

$A(I)O_6$ polyhedra

26. $d_{<B-O>}$ (Å)
27. $\angle<\tau_{O-B-O}>$ ($^\circ$)
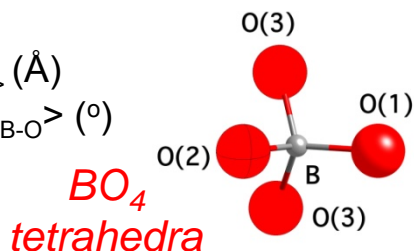
BO_4 tetrahedra

28. $\rho_{A(II)}$ (Å)
29. $d_{A(II)-X}$ (Å)
30. $\angle\alpha_{A(II)}$ ($^\circ$)
31. $d_{A(II)-O(3)}$ (Å)
32. $\angle\phi_{O(3)-A(II)-O(3)}$ ($^\circ$)

$A(II)O_6X_{1,2}$ polyhedra

NSF-IMI: International Materials Institute
CoSMIC
Combinatorial Sciences and Materials Informatics Collaboratory
NSF/MIT – EFRI Workshop: Krishna Rajan
IOWA STATE UNIVERSITY
OF SCIENCE AND TECHNOLOGY
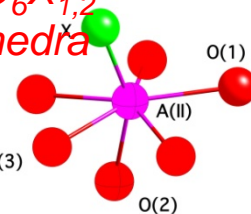
DESCRIPTOR REDUCTION: Decomposition map" of apatite crystal structure

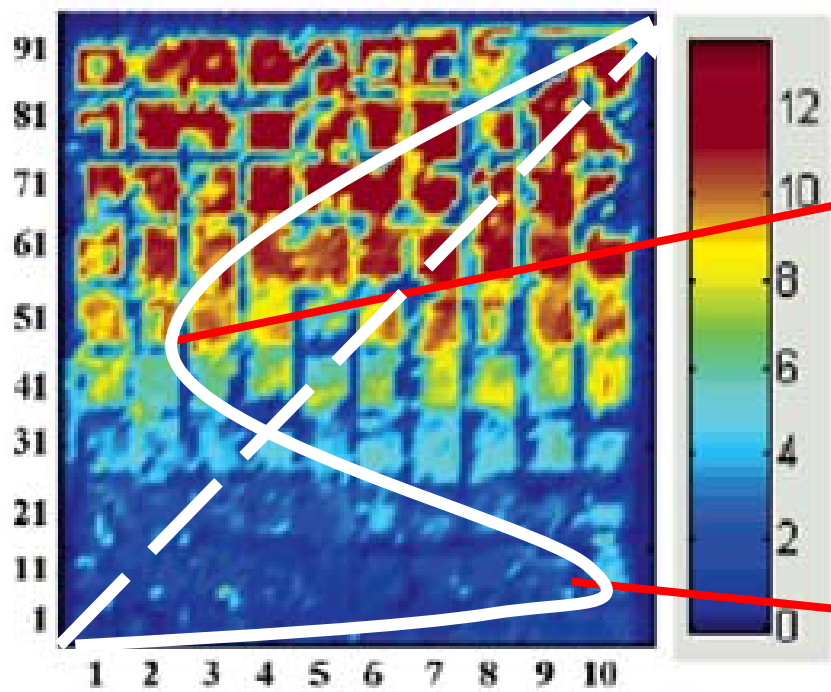**- Analysis based on variable-by-variable in terms of building blocks (for *unit cell parameters-c*)**

O(3)-B-O(3)-A(II)-O(3) chains
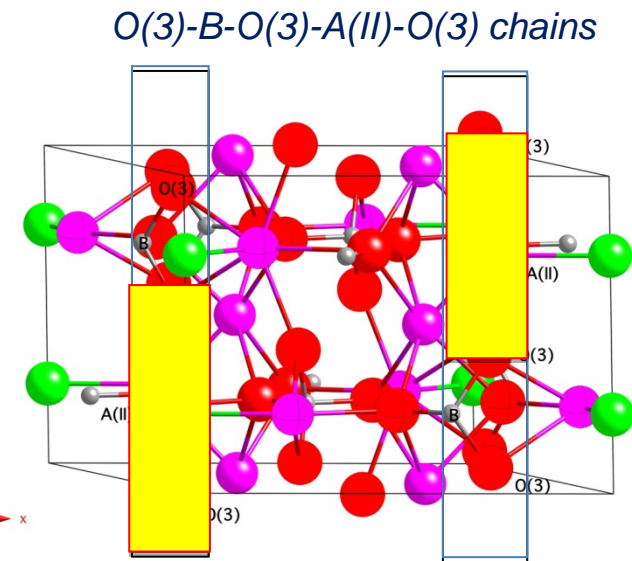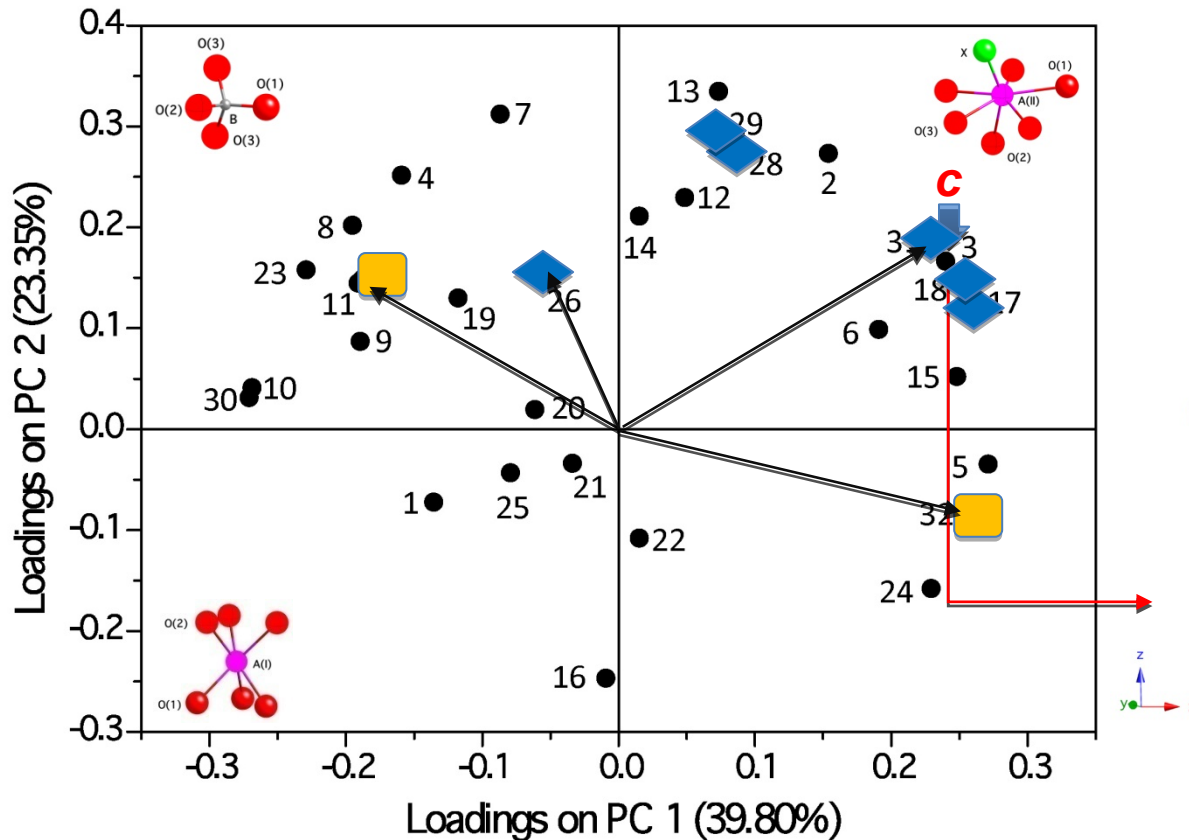
- *The linear(bond length) parameters (#17, #18, #28, and #29) were found to affect "c" than others.*
- *Then, "c" is more closely related with two variables, $d_{A(II)-O(3)}$ (#32) and $\angle\phi_{O(3)-A(II)-O(3)}$ (#32), than other two descriptors, $d_{<B-O>}$ (#26) and $\angle<\tau_{O-B-O}>$ (#27). (marked with yellow box in right figure)*
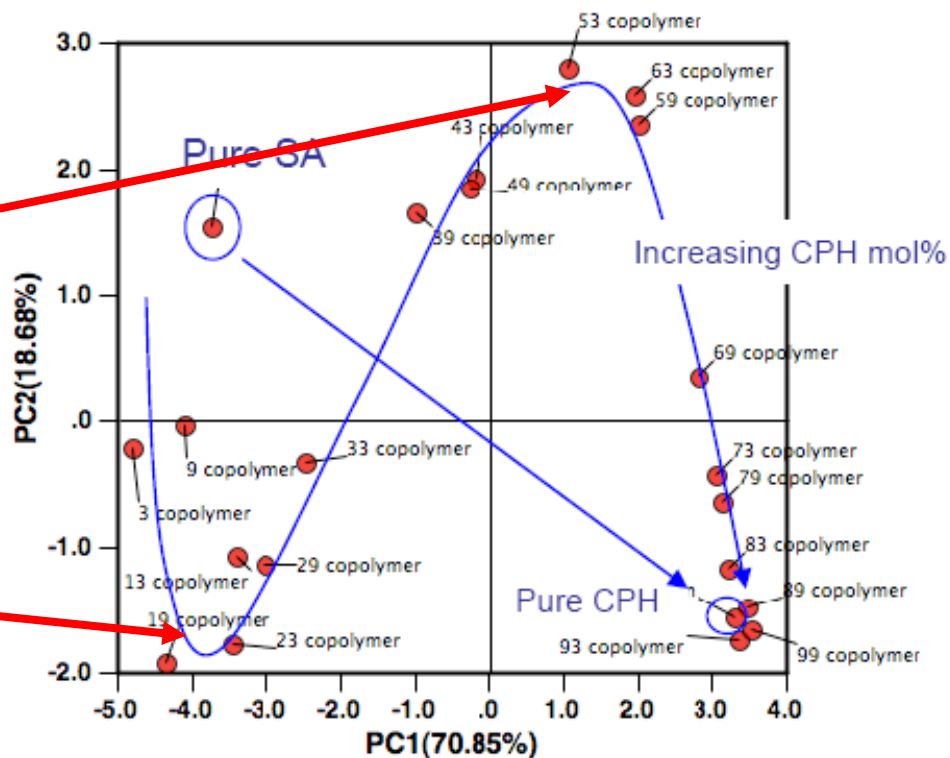
*As shown in right figure, "c" is characterized by O(3)-B-O(3)-A(II)-O(3) chains. **Interesting to note is that $A(II)O_6X_{1,2}$ polyhedra seems to have an effect on "c" but, the B-site (BO_4 tetrahedra) has less effect.***

SEARCH SPACE IN COMBINATORIAL EXPERIMENTS : What you see is not necessarily what you get!

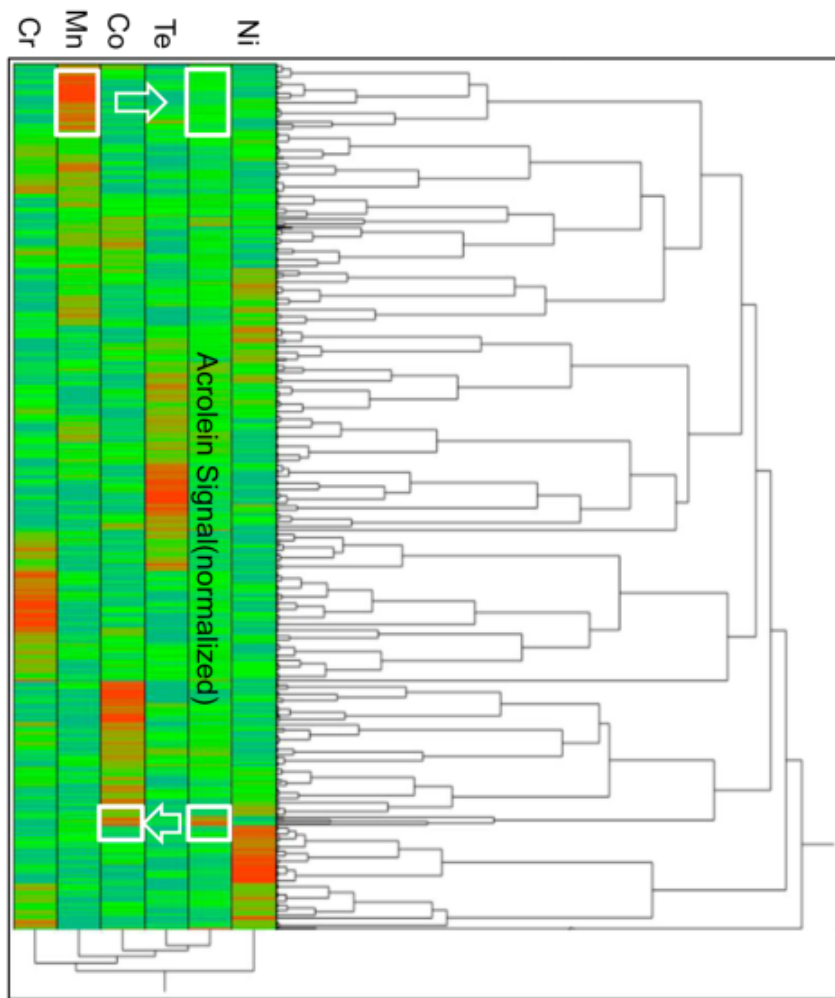Identifying multidimensional trajectories for targeted properties
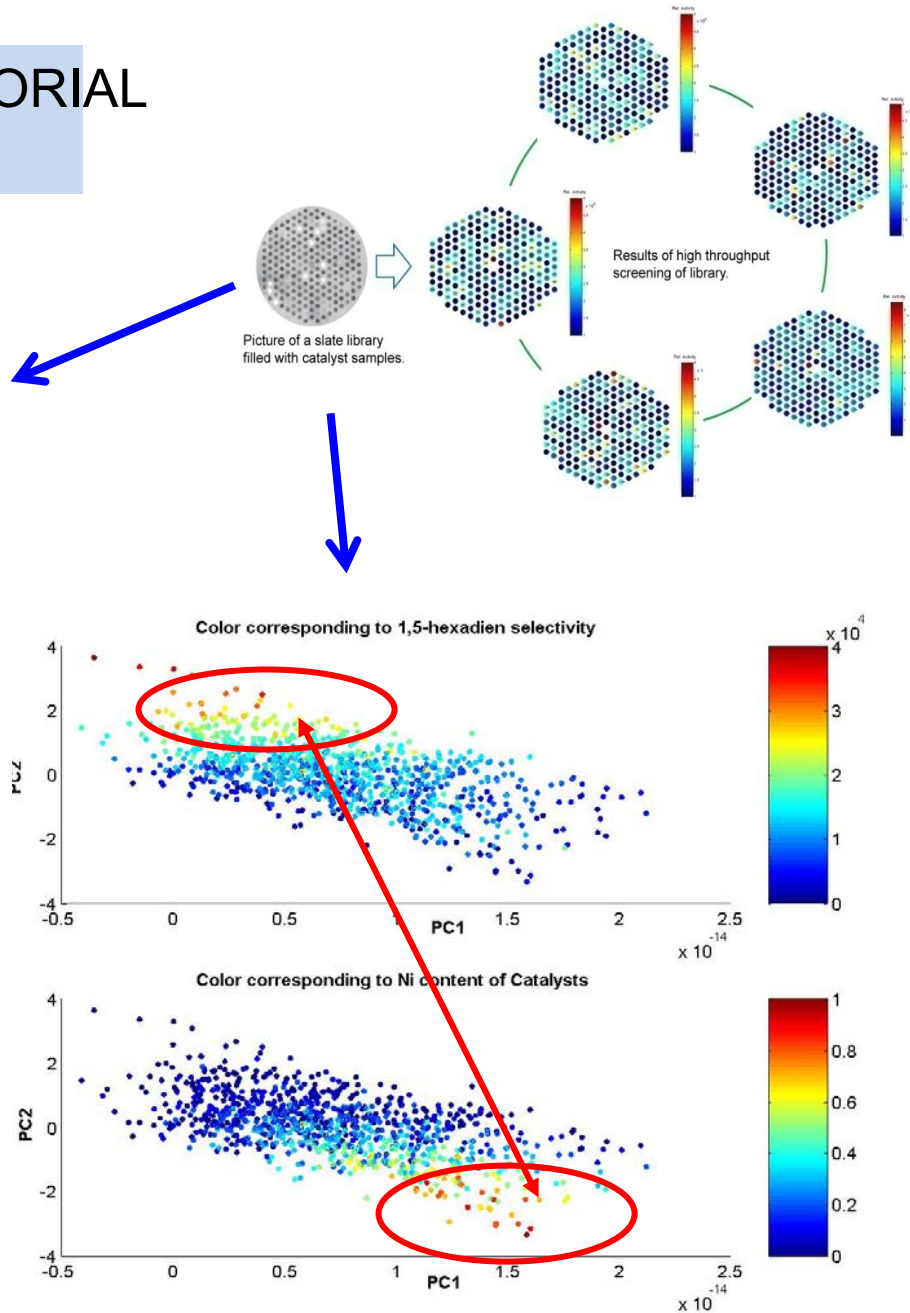

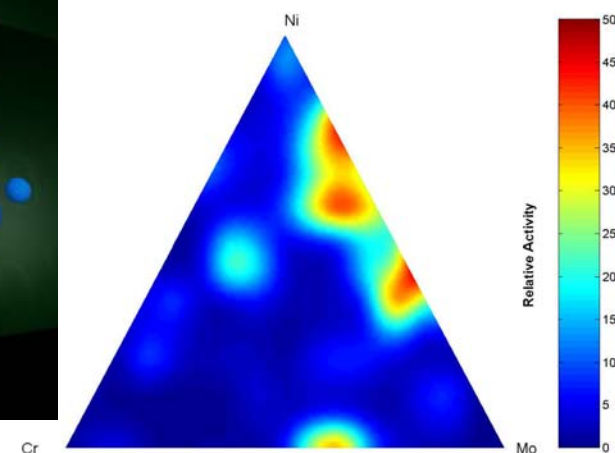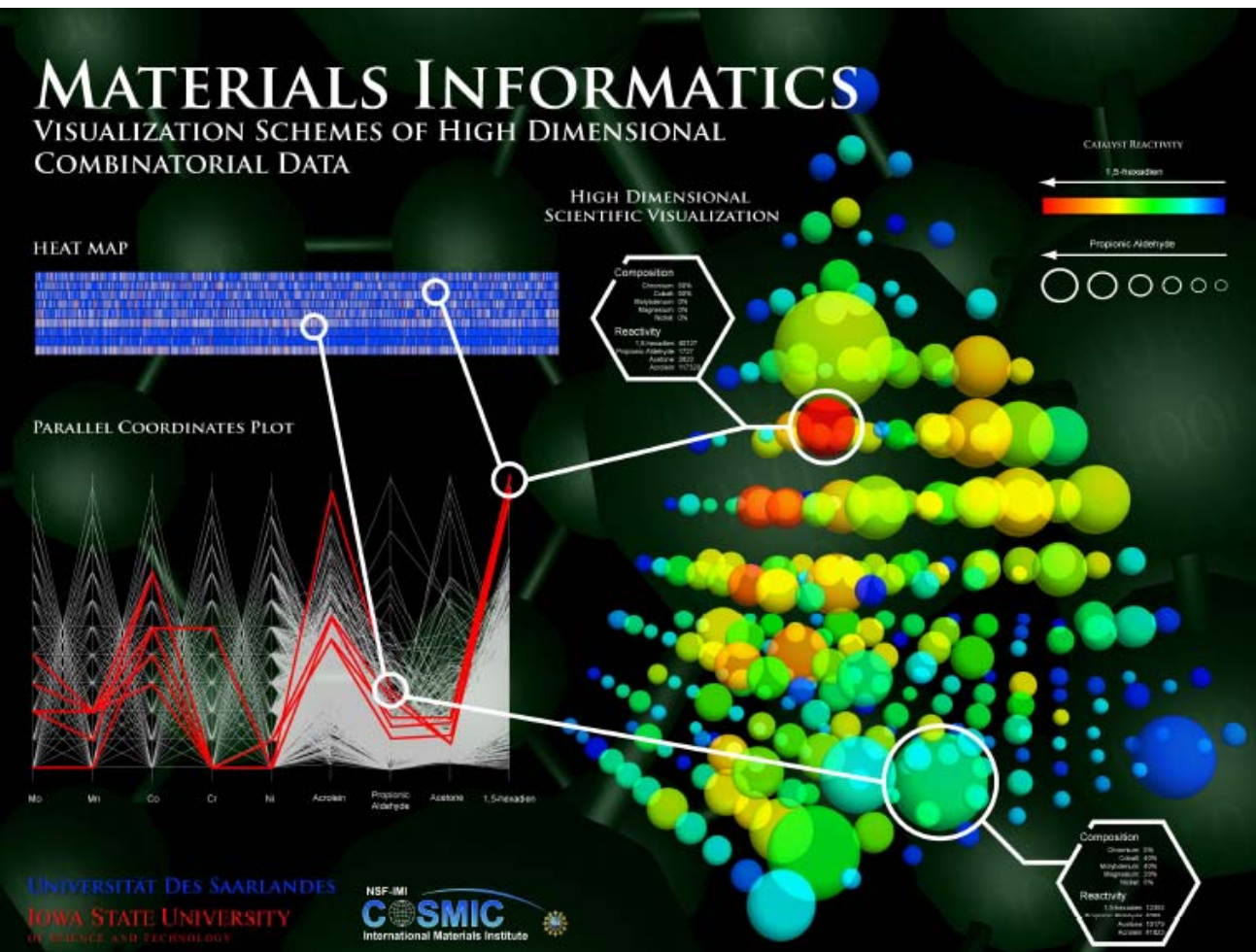Narsimihan and Mallapragada

Broderick et.al 2007

INFORMATICS GUIDED COMBINATORIAL EXPERIMENTATION

Picture of a slate library filled with catalyst samples.

Results of high throughput screening of library.

Color corresponding to 1,5-hexadien selectivity

Color corresponding to Ni content of Catalysts

Seig, Suh, Maier, Rajan

Seig, Suh, Maier

Simulated combinatorial array

DATA DRIVEN MATERIALS SCIENCE

Data + **Correlations** + **Theory** = **Knowledge Discovery**

| • Combinatorial experimentation

• Digital libraries & data bases | + | •Data mining

•Dimensionality reduction | + | • Atomistic based calculations

• Continuum based theories | = | •Materials discovery
•Structure-property-processing relationships
• Hidden data trends |
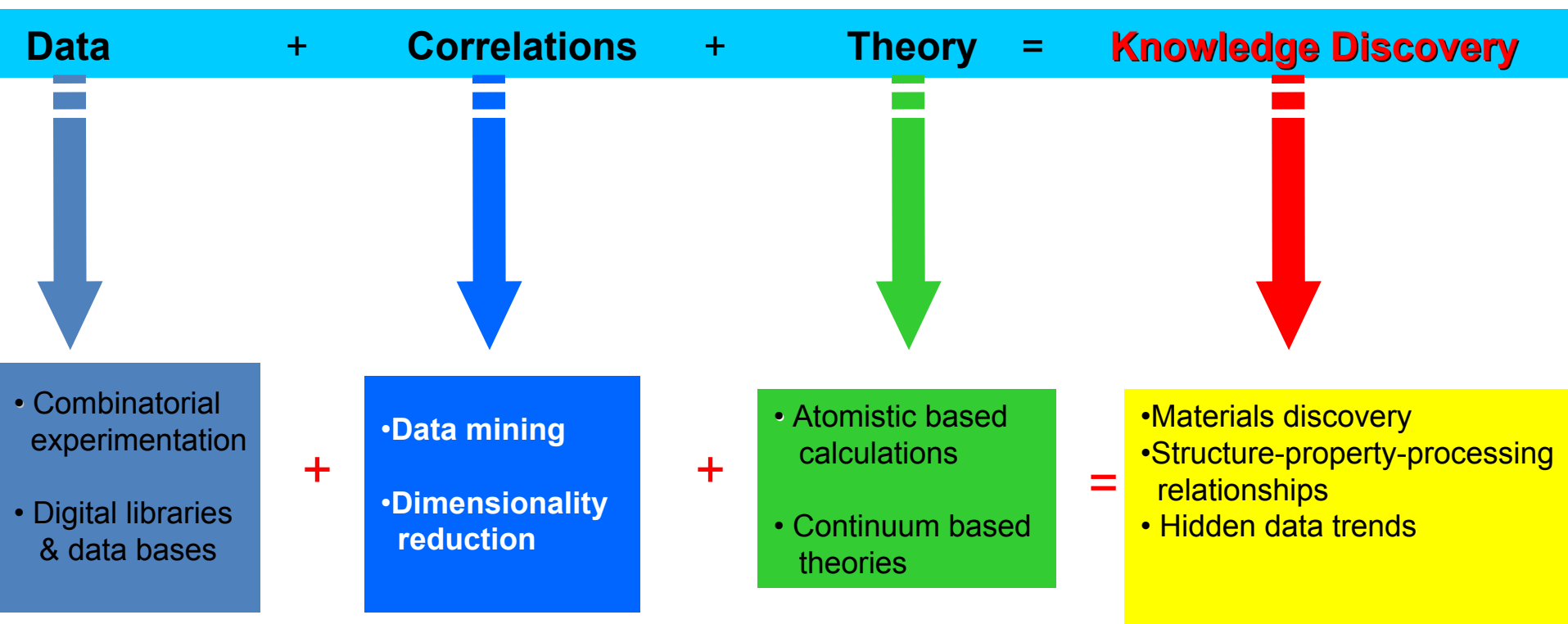
Information is multivariate, diverse , very large and access / expertise is globally distributed

From Cyberinfrastructure to Cyberdiscovery in Materials Science:

Enhancing outcomes in materials research, education and outreach through cyberinfrastructure
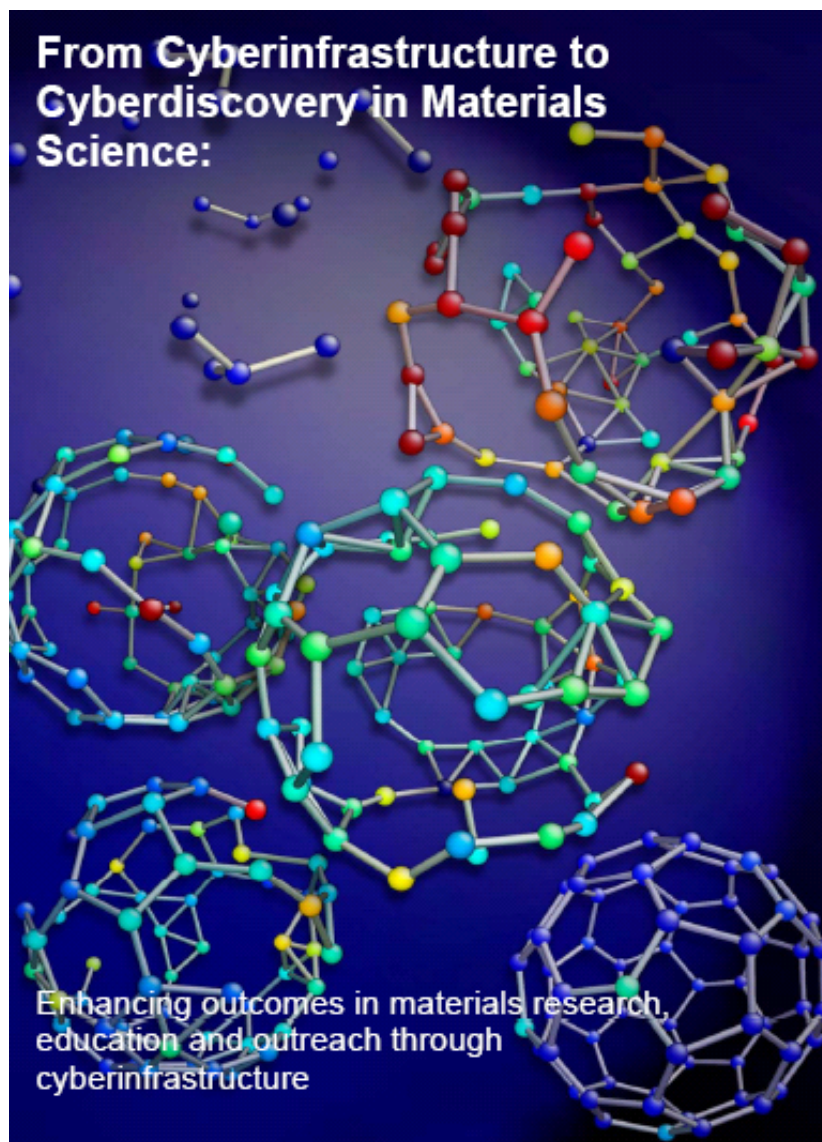
Report from a workshop held in Arlington, Virginia
August 3rd - 5th, 2006

Sponsored by the National Science Foundation

Professor Simon J. L. Billinge
Department of Physics and Astronomy, Michigan State University

Professor Krishna Rajan
Department of Materials Science and Engineering, Iowa State University

Professor Susan B. Sinnott
Department of Materials Science and Engineering, University of Florida