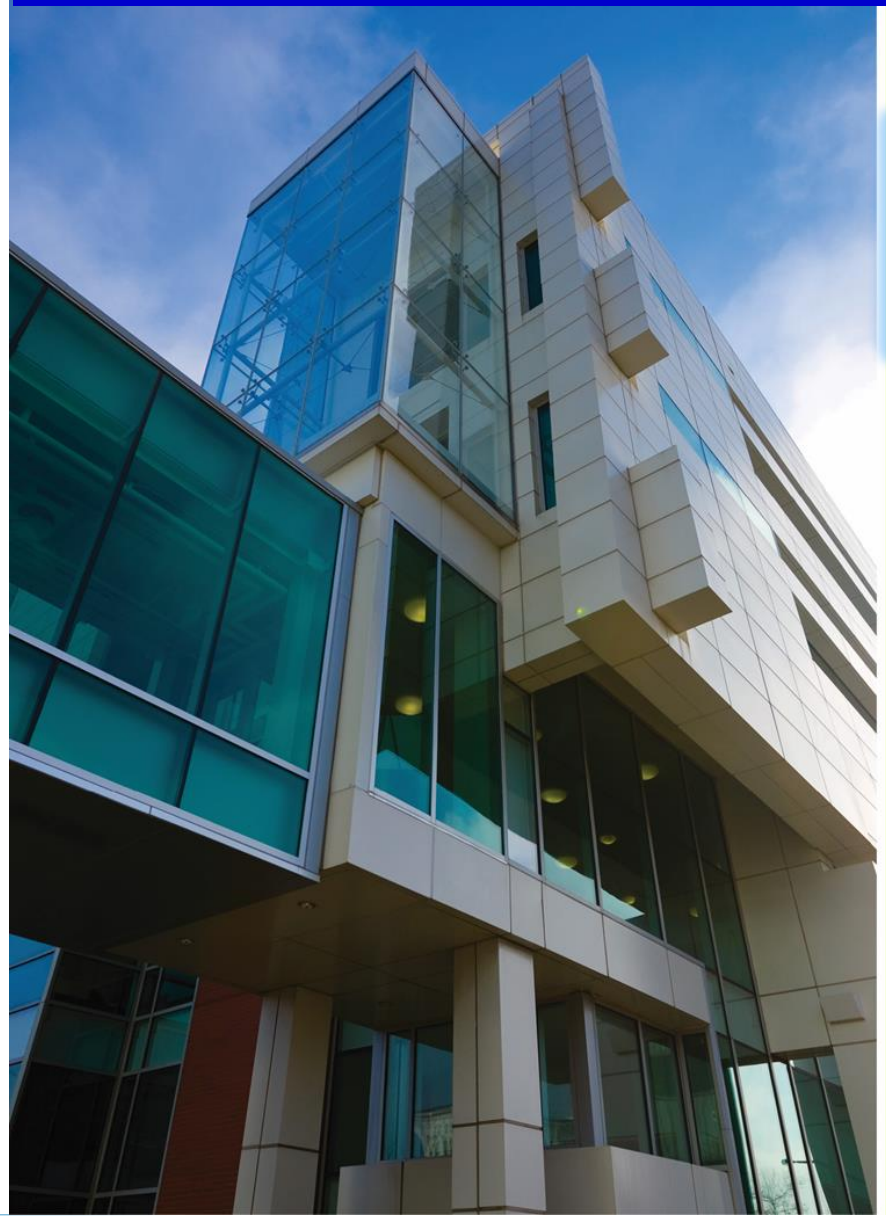
Materials Informatics: Mining and Learning from Data for Accelerated Design and Discovery

Krishna Rajan

Department of Materials Design and Innovation &
New York Center of Excellence in Materials Informatics
University at Buffalo- The State University of New York

Dept. Materials Design and Innovation



NEW YORK STATE CENTER OF EXCELLENCE
MATERIALS INFORMATICS

University at Buffalo
The State University of New York

http://engineering.buffalo.edu/materials-design-innovation.html

http://www.cas.buffalo.edu/departments-programs/

krajan3@buffalo.edu

A new paradigm - computational and experimental materials science + Big Data

NEW YORK STATE CENTER OF EXCELLENCE
MATERIALS INFORMATICS

University at Buffalo
The State University of New York

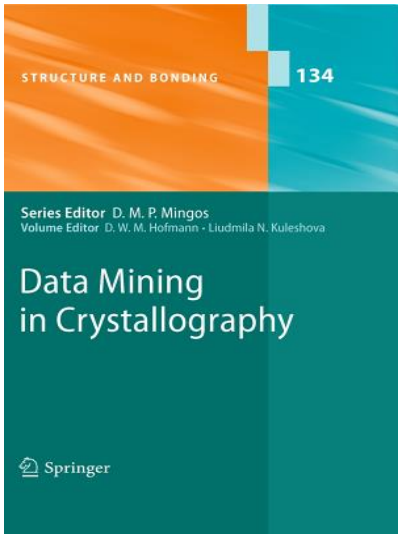
CENTER FOR COMPUTATIONAL RESEARCH

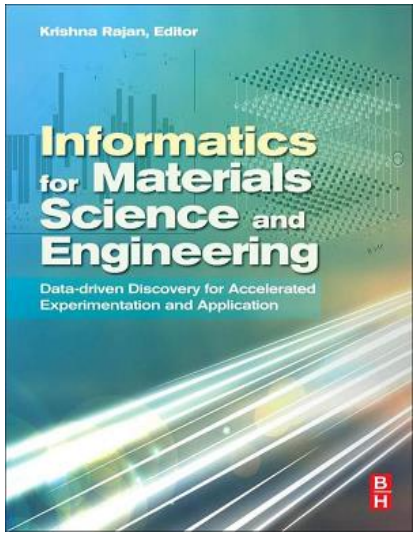
University at Buffalo
The State University of New York

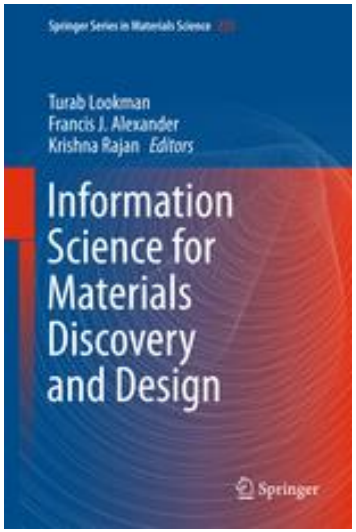
BUFFALO
MANUFACTURING
WORKS

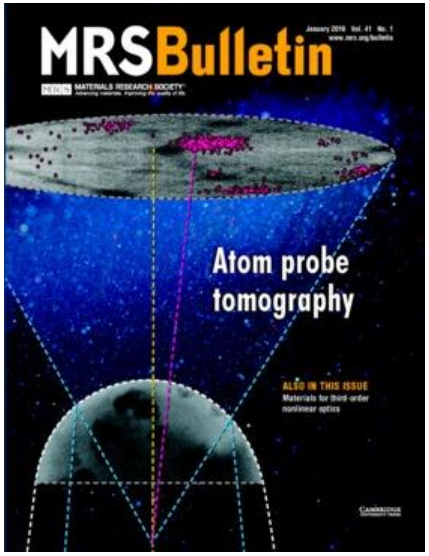
“Optics” of Data

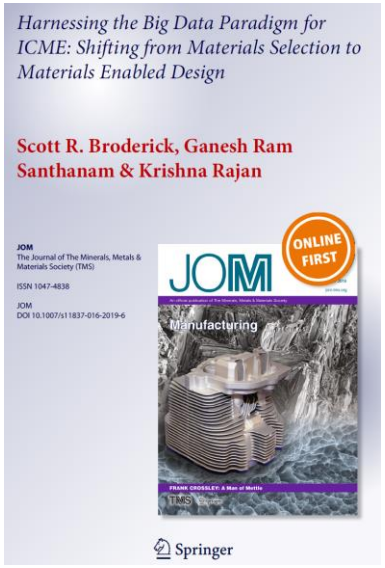
Bibliography

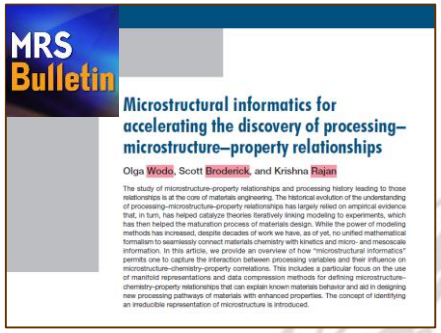


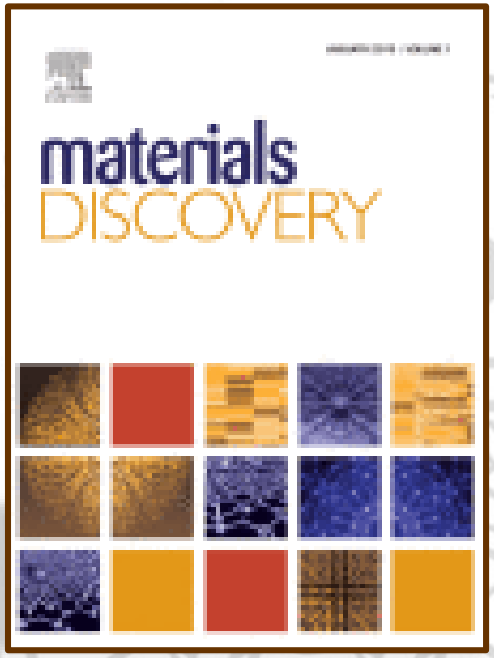




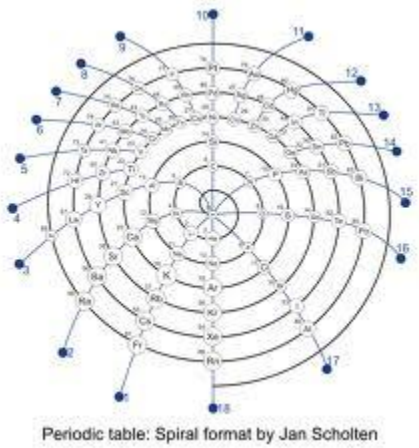




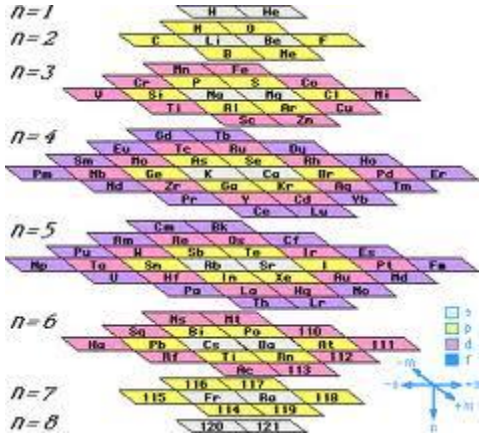


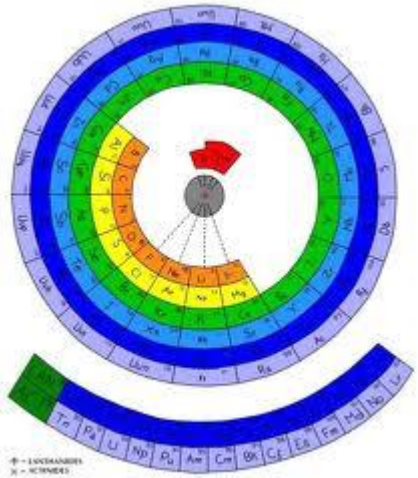


Cartography for Materials : *mapping elemental characteristics*



Periodic table: Spiral format by Jan Scholten



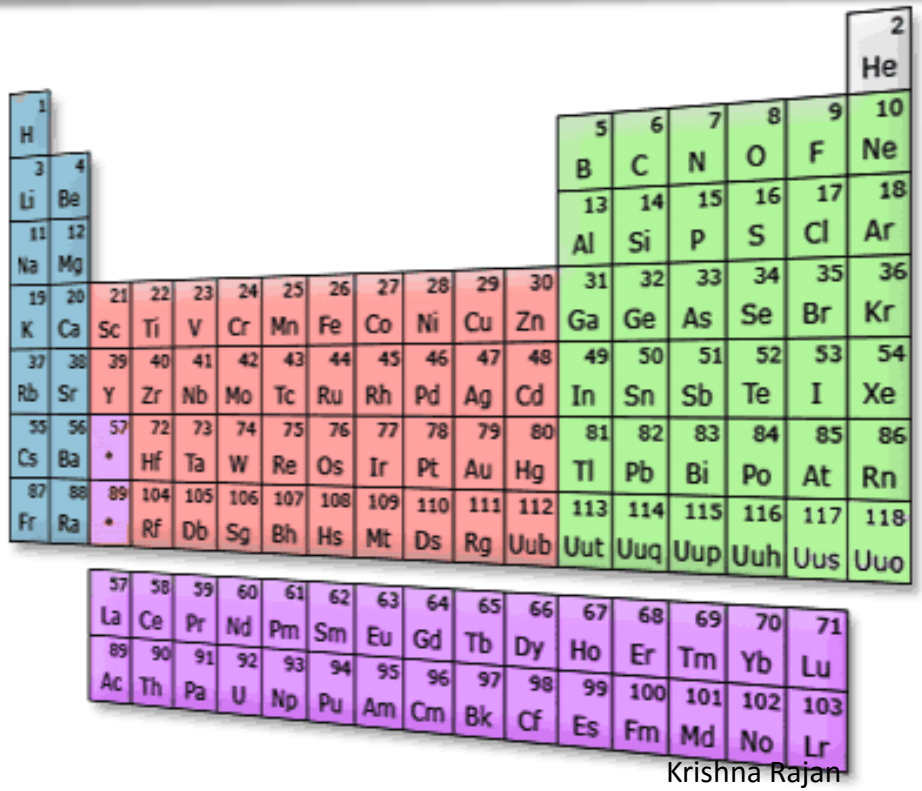


ОПЫТЪ СИСТЕМЫ ЭЛЕМЕНТОВЪ.

ОСНОВАННОЙ НА ИХЪ АТОМНОМЪ ВѢСѢ И ХИМИЧЕСКОМЪ СХОДСТВѢ.

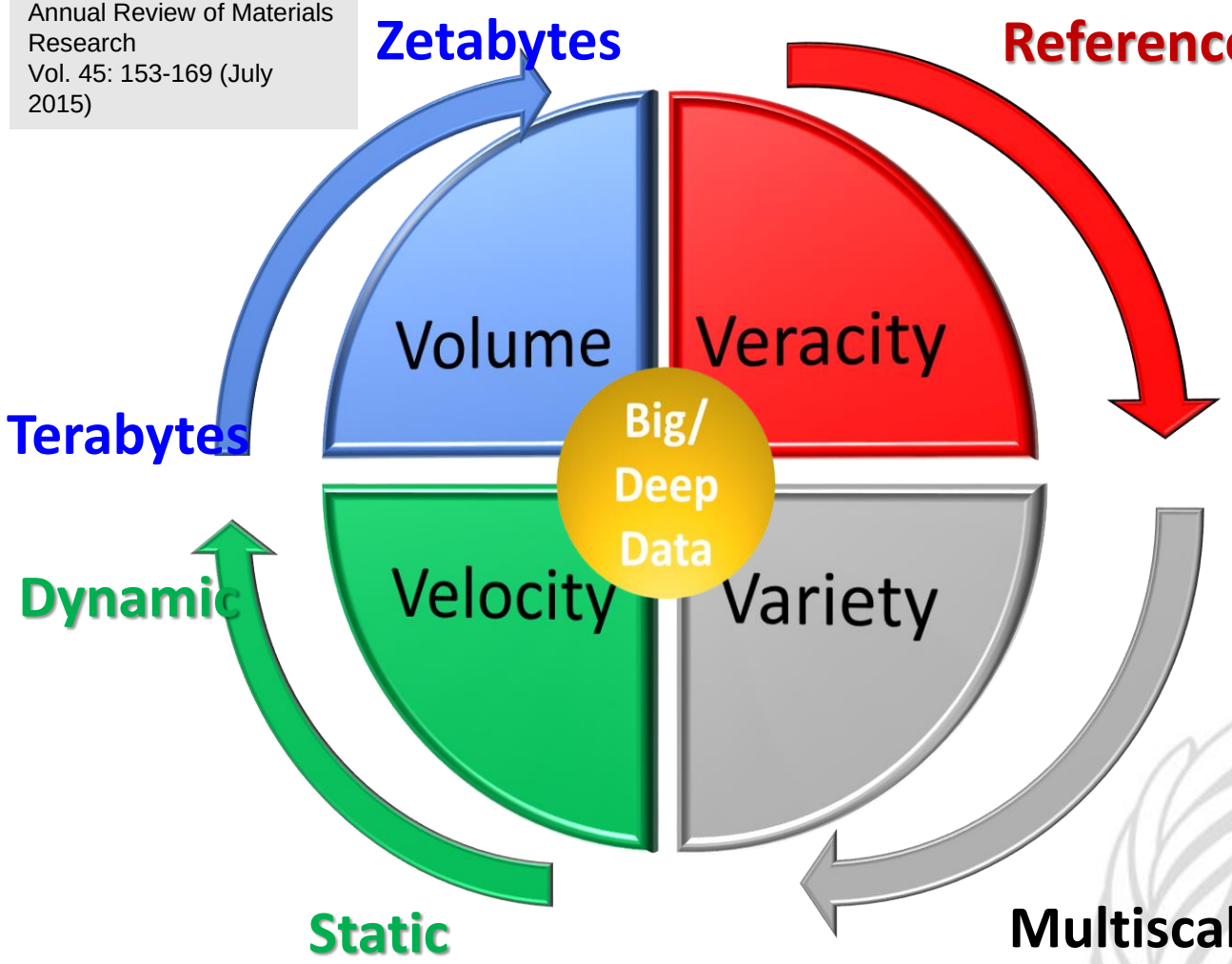
			Ti = 50	Zr = 90	? = 180.
			V = 51	Nb = 94	Ta = 182.
			Cr = 52	Mo = 96	W = 186.
			Mn = 55	Rh = 104,4	Pt = 197,1.
			Fe = 56	Ru = 104,4	Ir = 198.
			Ni = Co = 59	Pl = 106,6	O = 199.
H = 1			Cu = 63,4	Ag = 108	Hg = 200.
	Be = 9,4	Mg = 24	Zn = 65,2	Cd = 112	
	B = 11	Al = 27,4	? = 68	Ur = 116	Au = 197?
	C = 12	Si = 28	? = 70	Sn = 118	
	N = 14	P = 31	As = 75	Sb = 122	Bi = 210?
	O = 16	S = 32	Se = 79,4	Te = 128?	
	F = 19	Cl = 35,6	Br = 80	I = 127	
Li = 7	Na = 23	K = 39	Rb = 85,4	Cs = 133	Tl = 204.
		Ca = 40	Sr = 87,6	Ba = 137	Pb = 207.
		? = 45	Ce = 92		
		?Er = 56	La = 94		
		?Yt = 60	Di = 95		
		?In = 75,6	Th = 118?		

Д. Менделѣевъ



MDI for "Knowledge Bases" vs Data Bases

Materials Informatics: The Materials "Gene" and Big Data : K. Rajan
Annual Review of Materials Research
Vol. 45: 153-169 (July 2015)



- *Crystallography*
- *Metrology/ Sensing*
- *Biopolymer design*
- *Chemical design of nanostructures*
- *Catalysis*
- *Alloy discovery and design*
- *Composites*
- *Decision systems in materials engineering*

Outline

- Formatting the inverse design problem as a data mining problem
- Defining the concept of "data" in materials science- mathematical framework for materials design via statistical learning: materials informatics
- Applications in crystal chemistry design

Inverse Design: an informatics/ data mining perspective

- **Discovering classifiers:**
  - **Can we trace and identify key characteristics of a material that distinguishes key differences in a materials behavior**

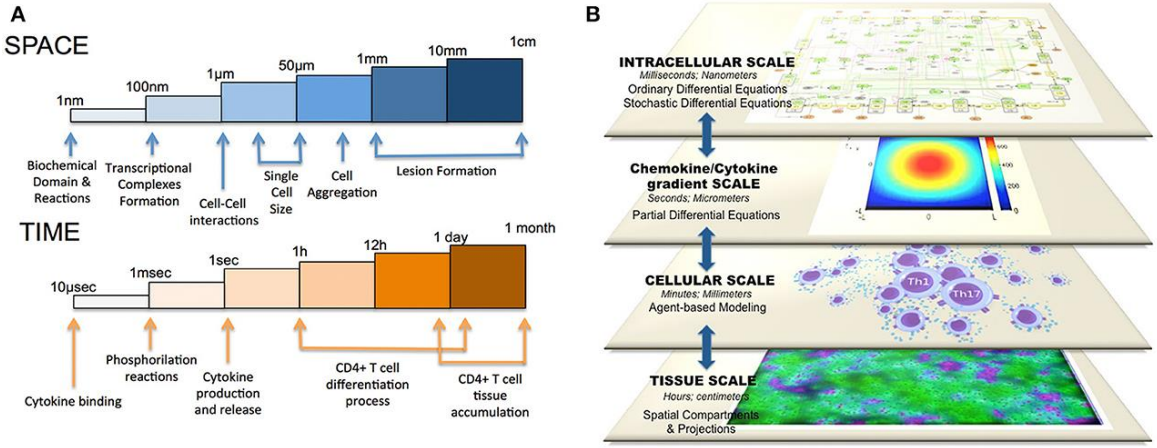
- **Ranking descriptors:**
  - **Can we trace back and discover key parameters or correlations of parameters without explicitly knowing the forward design problem?**

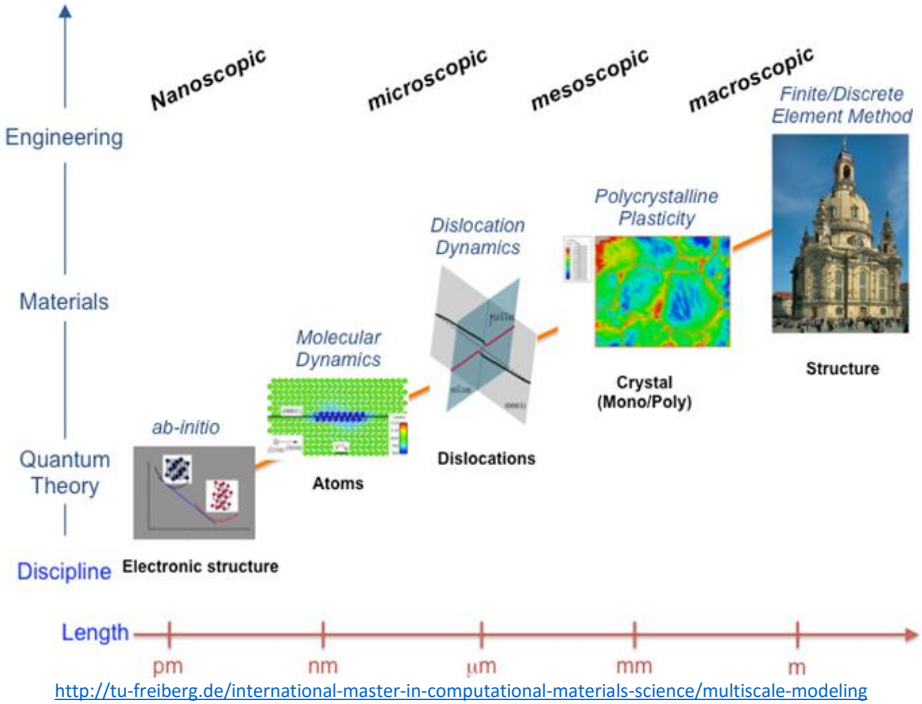
Informatics for materials discovery

- Data driven discovery with limited or uncertain information
- Data from different sources---
  - How can we deal with training and test data from different distributions?
  - What can be learned when a classifier (or predictor) is used outside of the domain of training?
  - Can we build classifiers?
- Establishing predictive materials science through data science

Multiscale Perspective

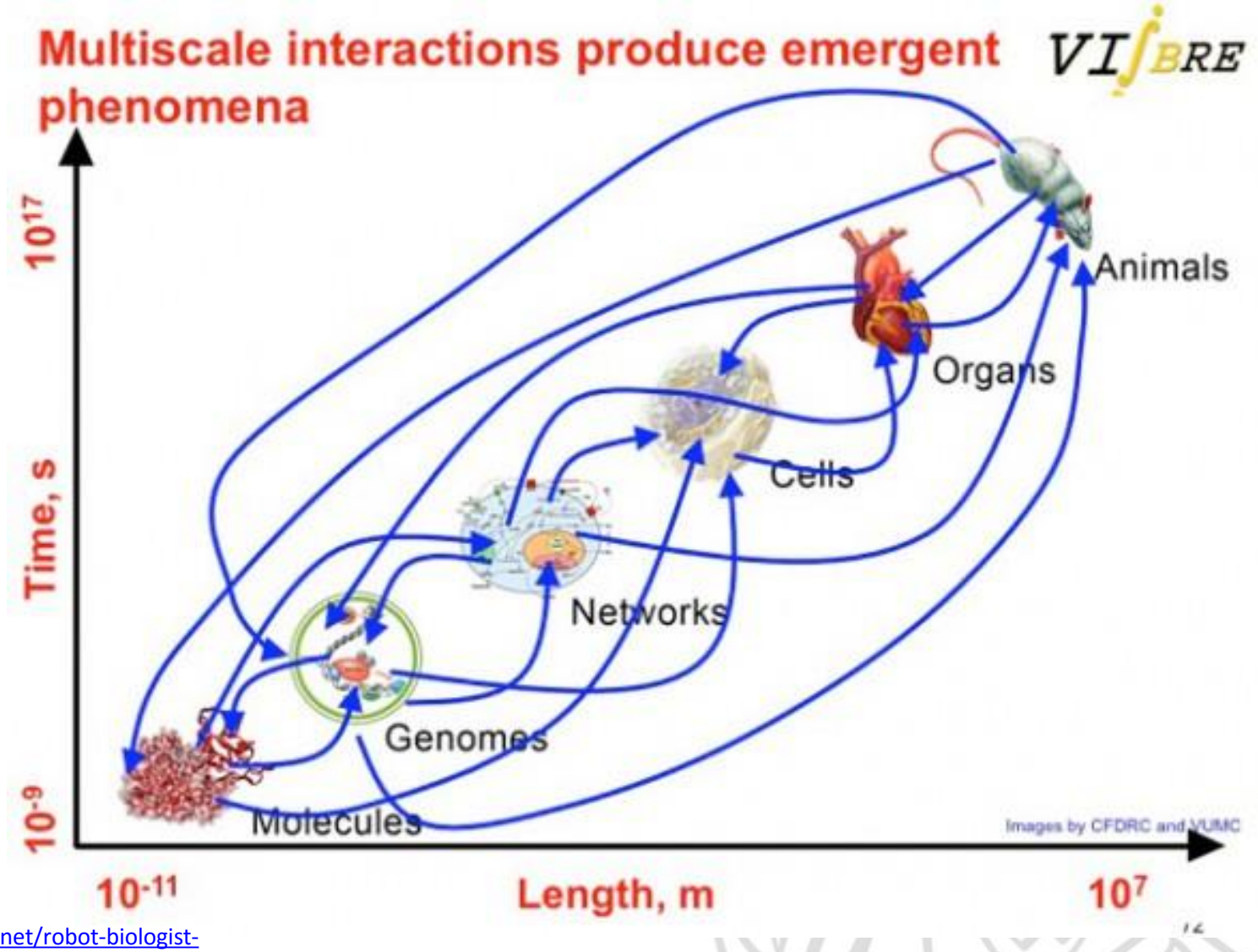
http://www.frontiersin.org/files/Articles/101961/fcell-02-00031-HTML/image_m/fcell-02-00031-g003.jpg





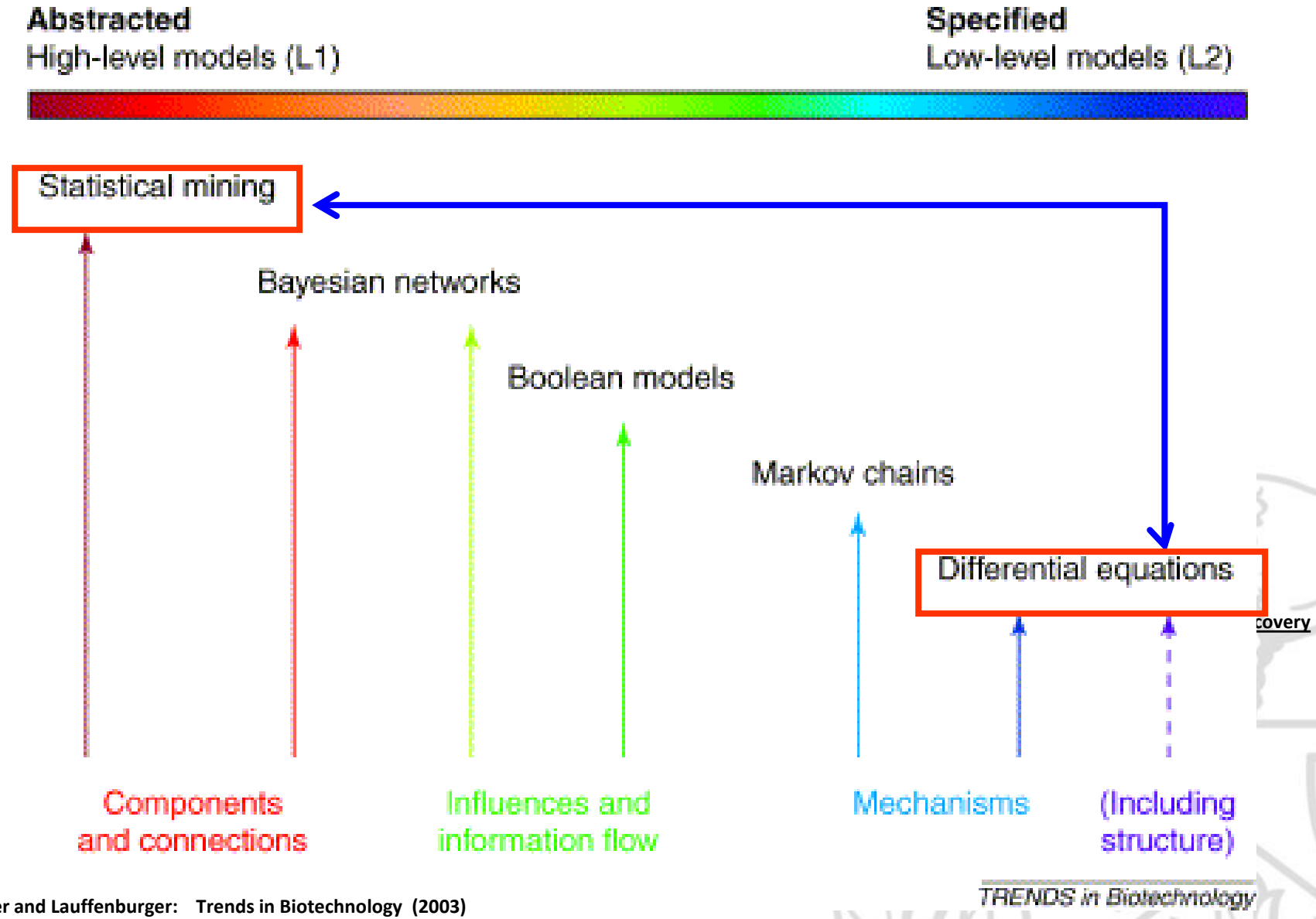
http://tu-freiberg.de/international-master-in-computational-materials-science/multiscale-modeling

Multiscale Perspective



http://www.kurzweilai.net/robot-biologist-solves-complex-problem-from-scratch

“Omics in Materials Science



Ideker and Lauffenburger: Trends in Biotechnology (2003)

$$\text{Functionality} = \mathcal{F}\,(x_1, x_2, x_3, x_4, x_5, x_6, x_7, x_8 \ldots\ldots)$$

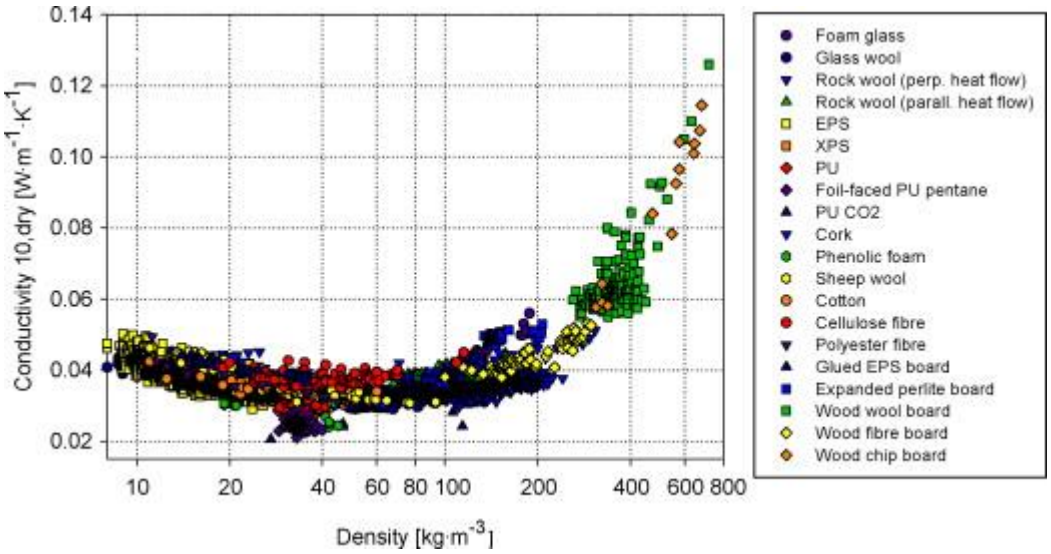
Issues:

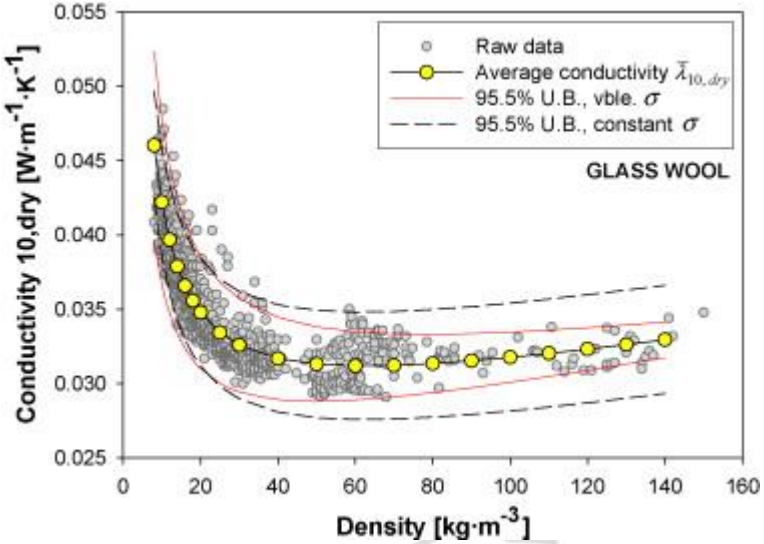
- how many variables?
- which variables are important?
- classify behavior among variables
- making quantitative predictions …relate functionality to variables …
  - traditionally we describe them by empirical equations:
  - Quantitative Structure Activity Relationships (QSARs) are derived from data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

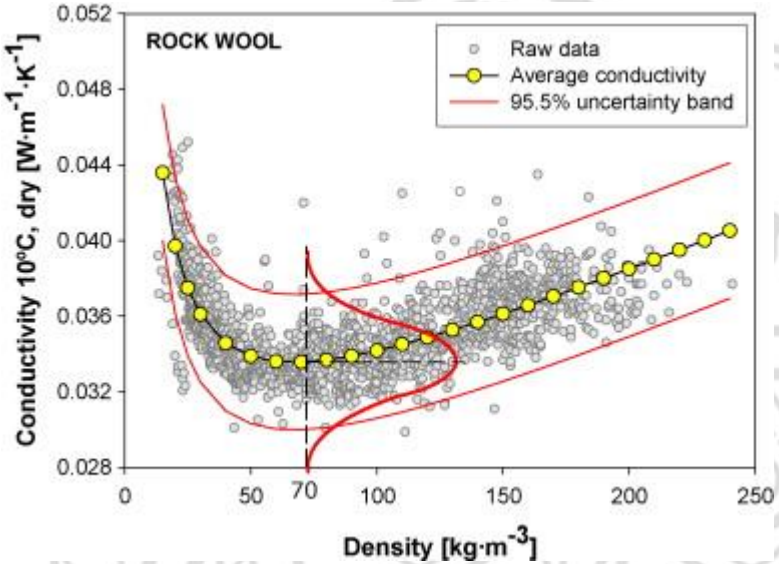
Need to build database with these variables

- *Need to establish a genomics-proteomics-metabolomics etc. :Systems Biology approach*
- *What are the data management and sharing issues for a systems approach?*

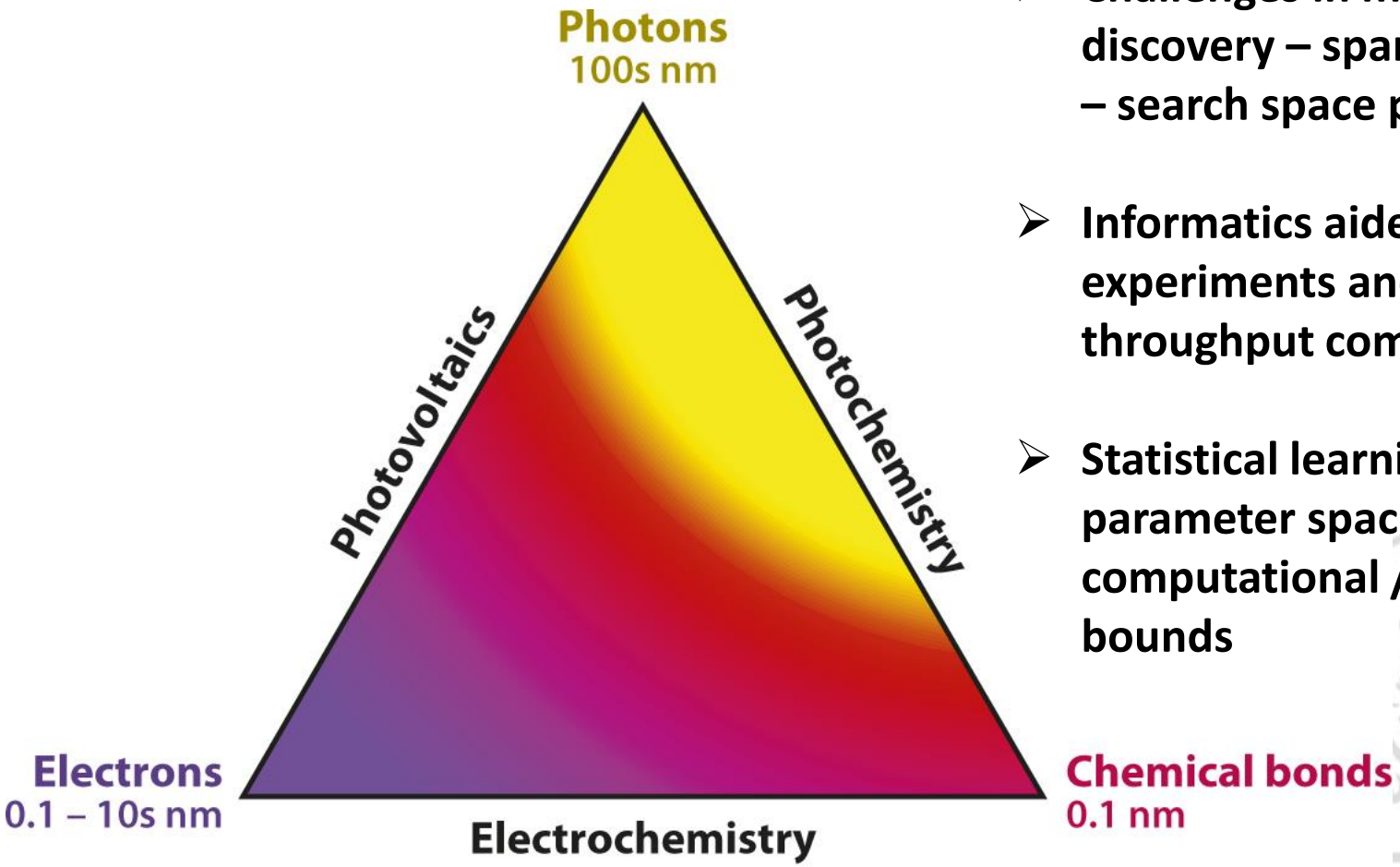
Uncertainty in the thermal conductivity of insulation materials Original Research Article
Energy and Buildings, Volume 42, Issue 11, November 2010, Pages 2159-2168
Fernando Domínguez-Muñoz, Brian Anderson, José M. Cejudo-López, Antonio Carrillo-Andrés







Discovery Challenge

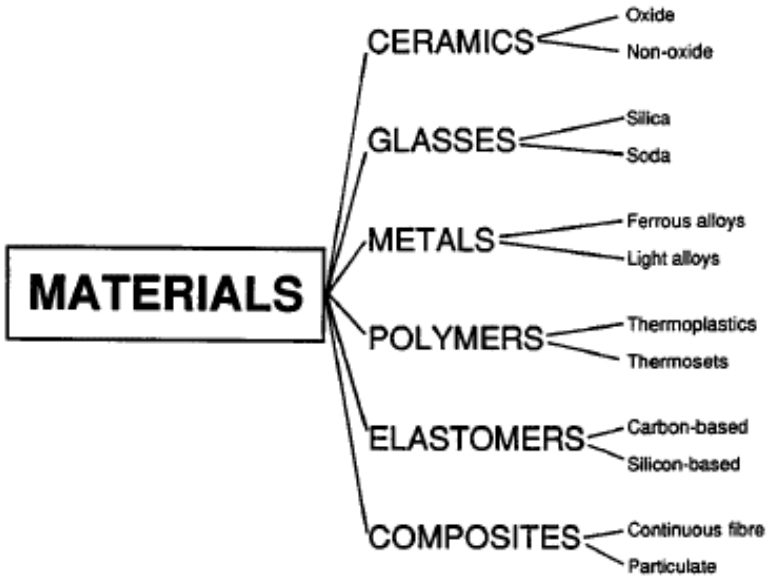


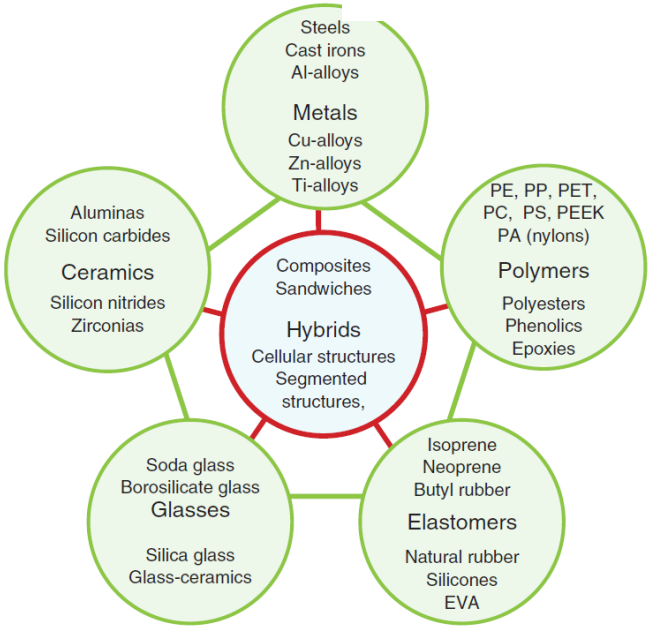
- Challenges in materials discovery – sparse information – search space prohibitive
- Informatics aided combinatorial experiments and high throughput computation
- Statistical learning- explore parameter space outside computational / experimental bounds

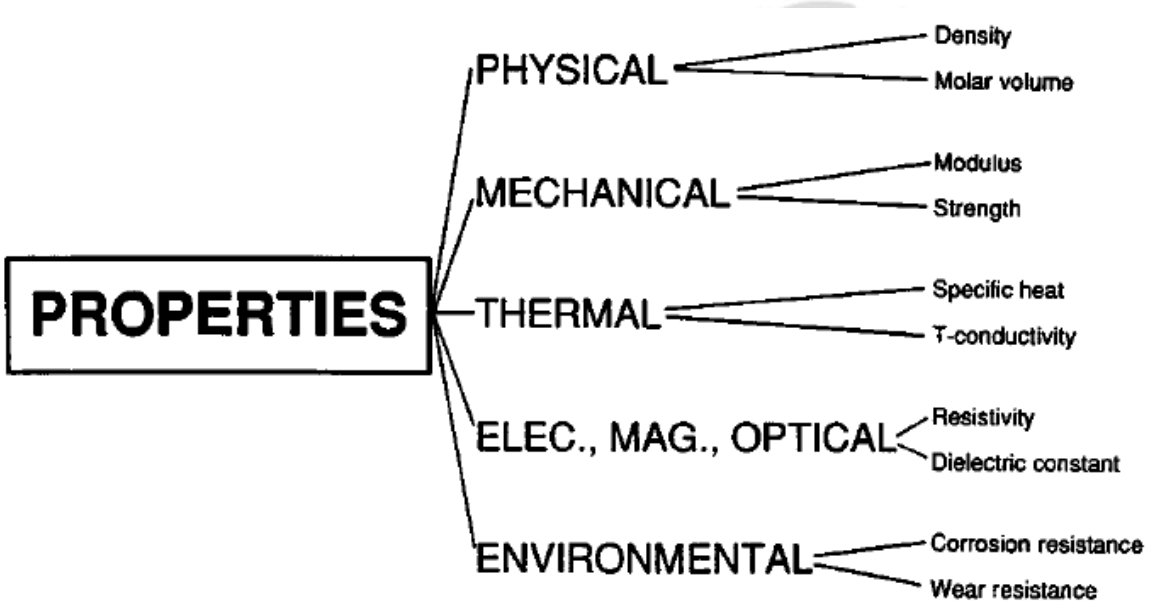
Crabtree G, Sarrao J. 2011.

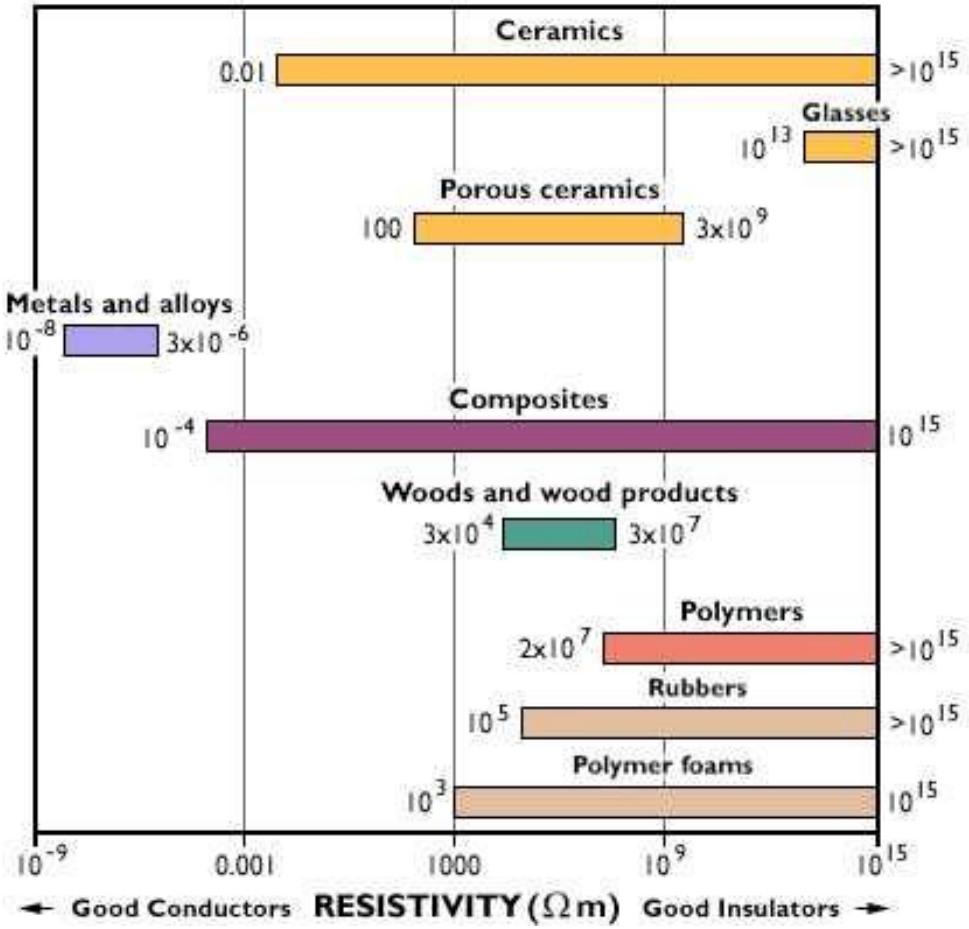
Annu. Rev. Condens. Matter Phys. 2:287–301

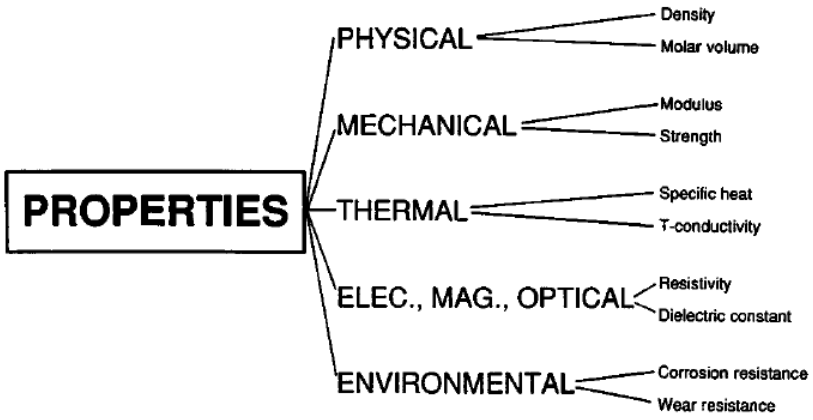
CLASSIFICATION:

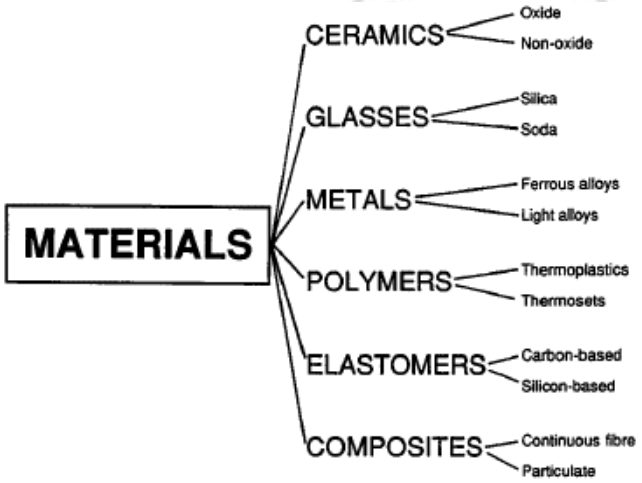


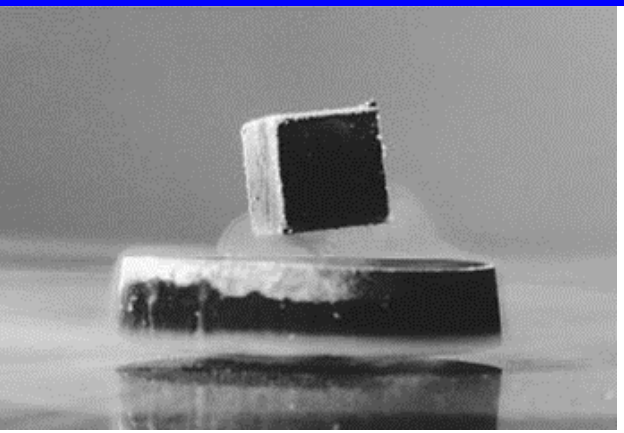


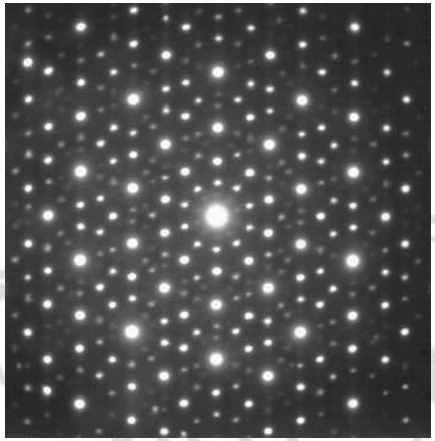



Ceramics
0.01
>10^15
Glasses
10^13
>10^15
Porous ceramics
100
3x10^9
Metals and alloys
10^-8
3x10^-6
Composites
10^-4
10^15
Woods and wood products
3x10^4
3x10^7
Polymers
2x10^7
>10^15
Rubbers
10^5
>10^15
Polymer foams
10^3
10^15
10^-9
0.001
1000
10^9
10^15
Good Conductors
RESISTIVITY (Ω m)
Good Insulators

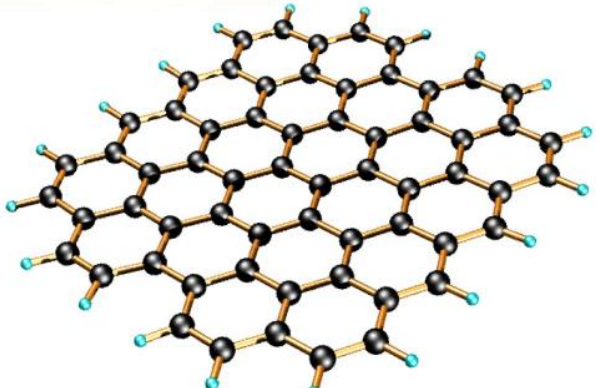

PROPERTIES
PHYSICAL
Density
Molar volume
MECHANICAL
Modulus
Strength
THERMAL
Specific heat
T-conductivity
ELEC., MAG., OPTICAL
Resistivity
Dielectric constant
ENVIRONMENTAL
Corrosion resistance
Wear resistance

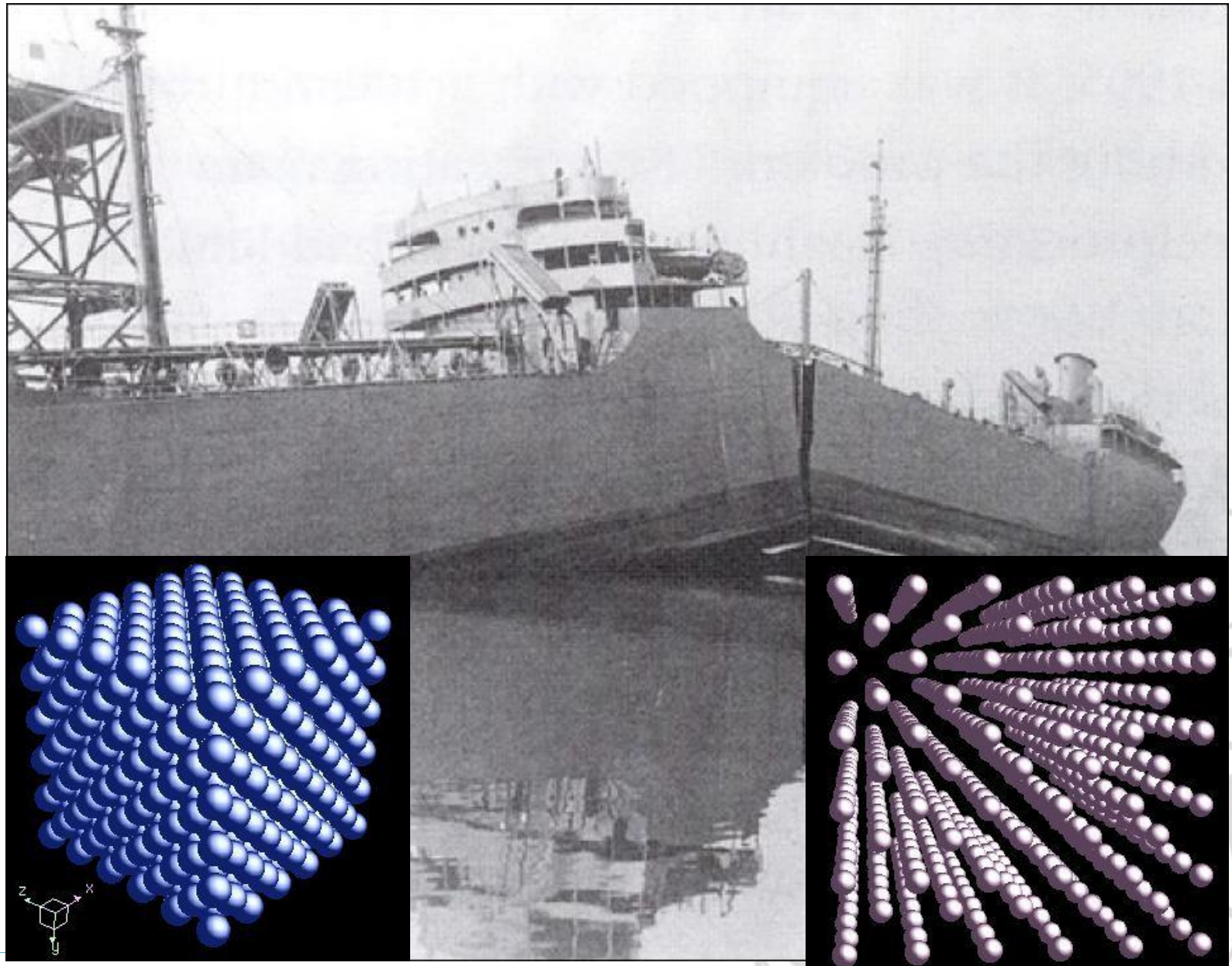

MATERIALS
CERAMICS
Oxide
Non-oxide
GLASSES
Silica
Soda
METALS
Ferrous alloys
Light alloys
POLYMERS
Thermoplastics
Thermosets
ELASTOMERS
Carbon-based
Silicon-based
COMPOSITES
Continuous fibre
Particulate

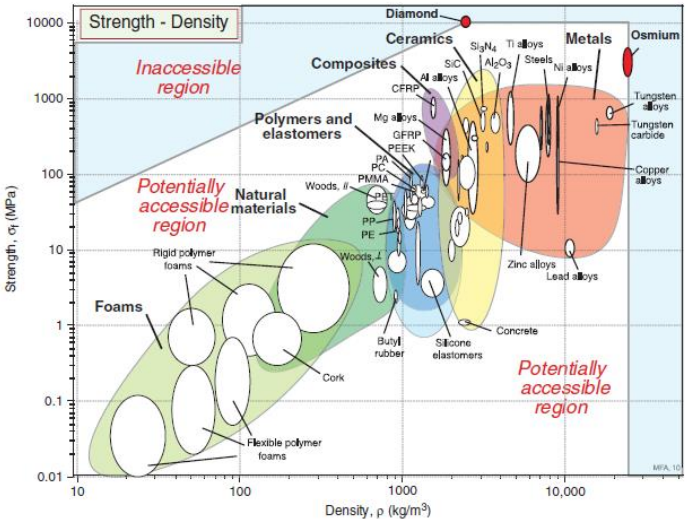




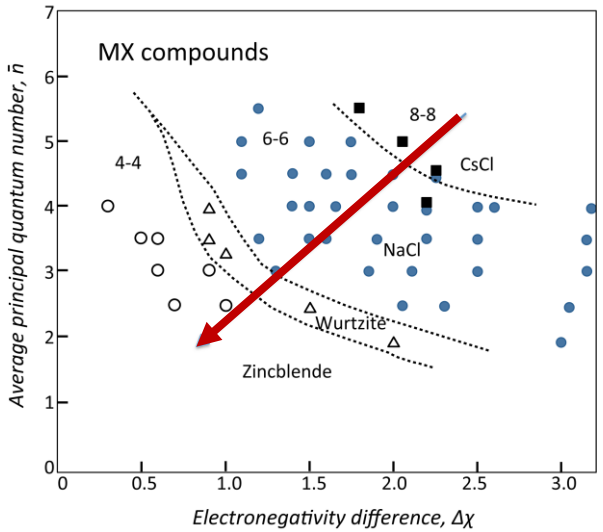

LED
OLED





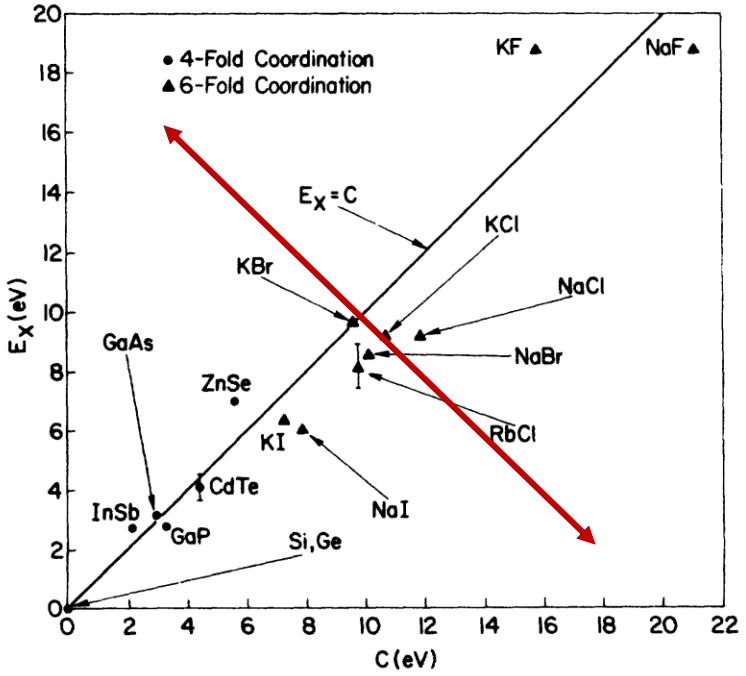


Nested clusters



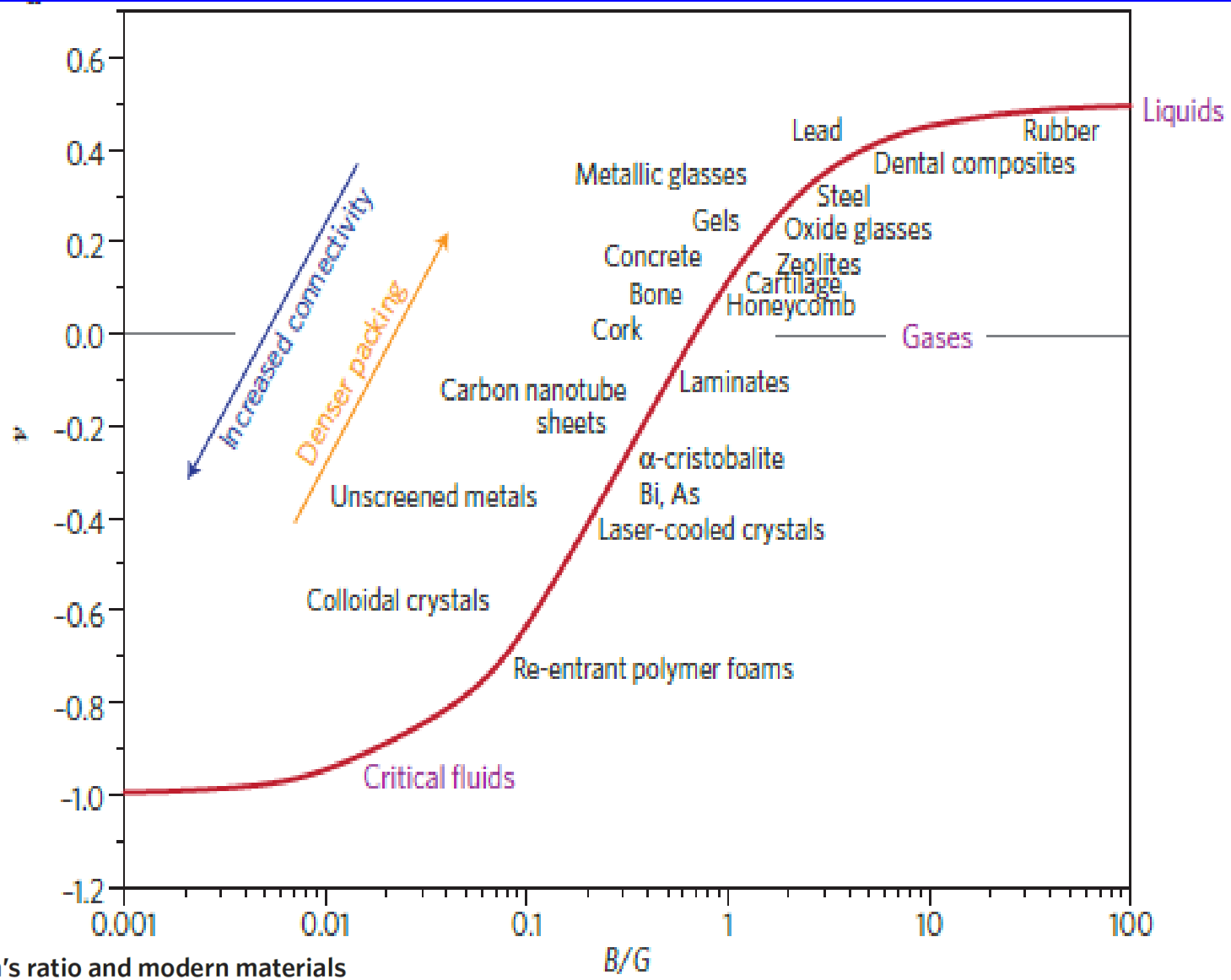
https://commons.wikimedia.org/wiki/File:Mooser-pearson.png

Decision Boundary



PHYSICAL REVIEW B VOLUME 8, 15 DEC. 1973
Atomic Pseudopotentials and the Ionicity Parameter of Phillips and Van Vechten
D. J. Chadi~ and Marvin L. Cohen* and D. Grobman

Data for Design

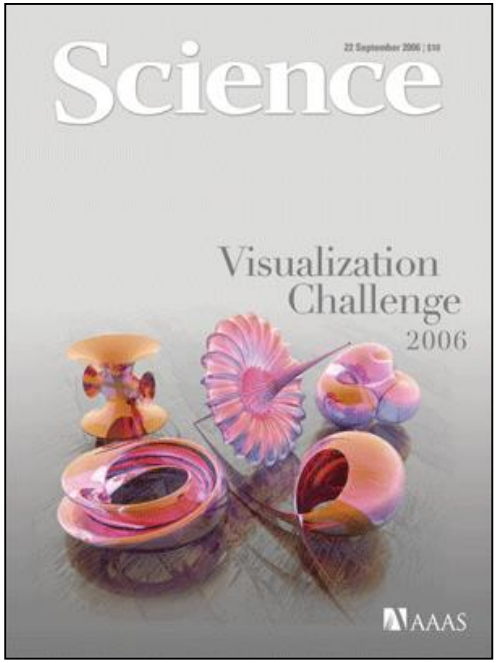


Poisson's ratio and modern materials

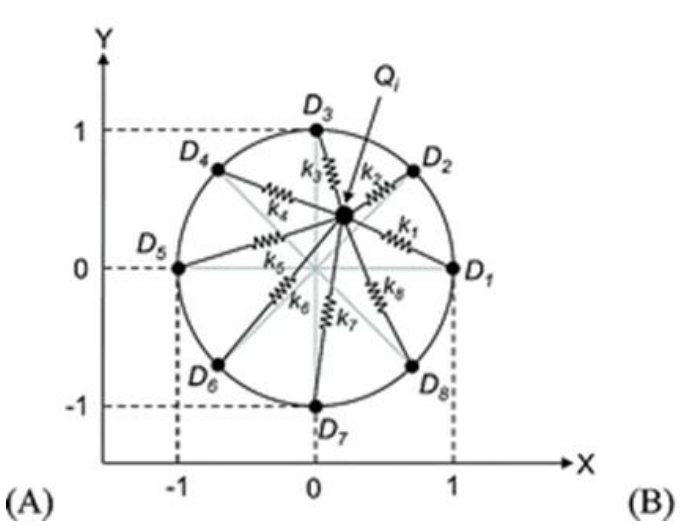
G. N. Greaves[1,2*], A. L. Greer[1], R. S. Lakes[3] and T. Rouxel[4]

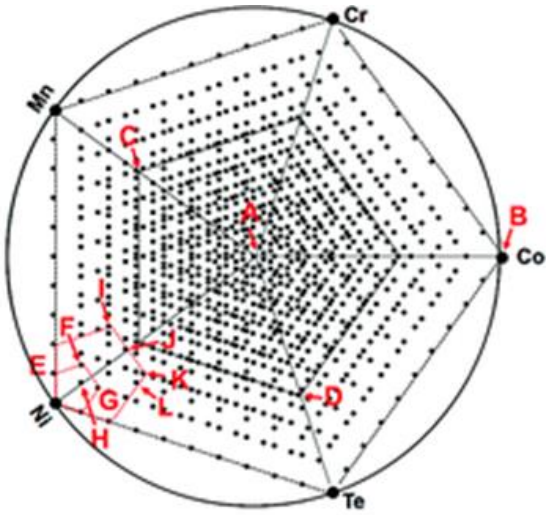
NATURE MATERIALS | VOL 10 | NOVEMBER 2011 | www.nature.com/naturematerials

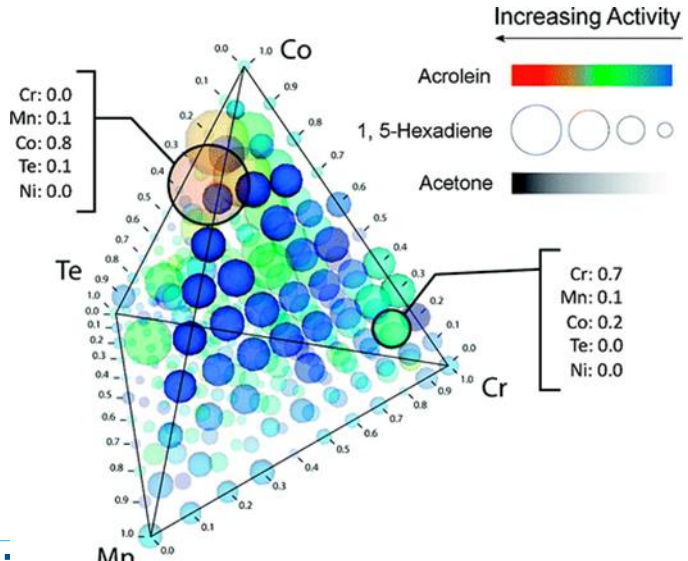
Data Velocity: *discovering catalyst chemistries*

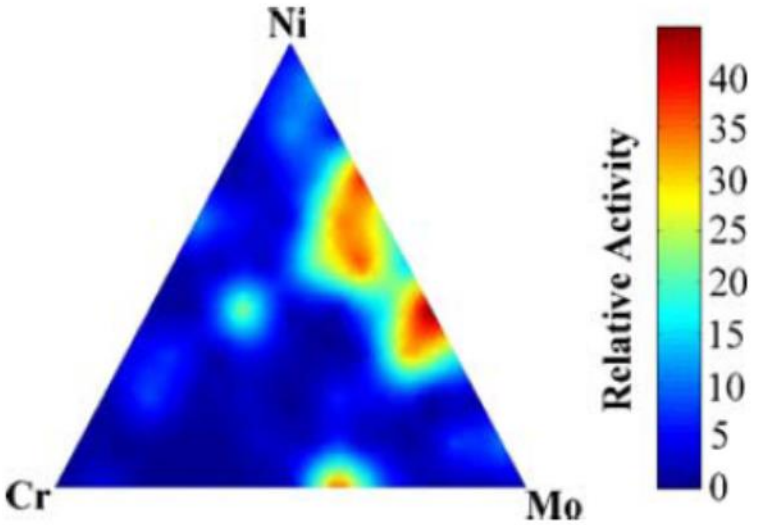


J. Comb. Chem., 2009, 11 (3), pp 385–392









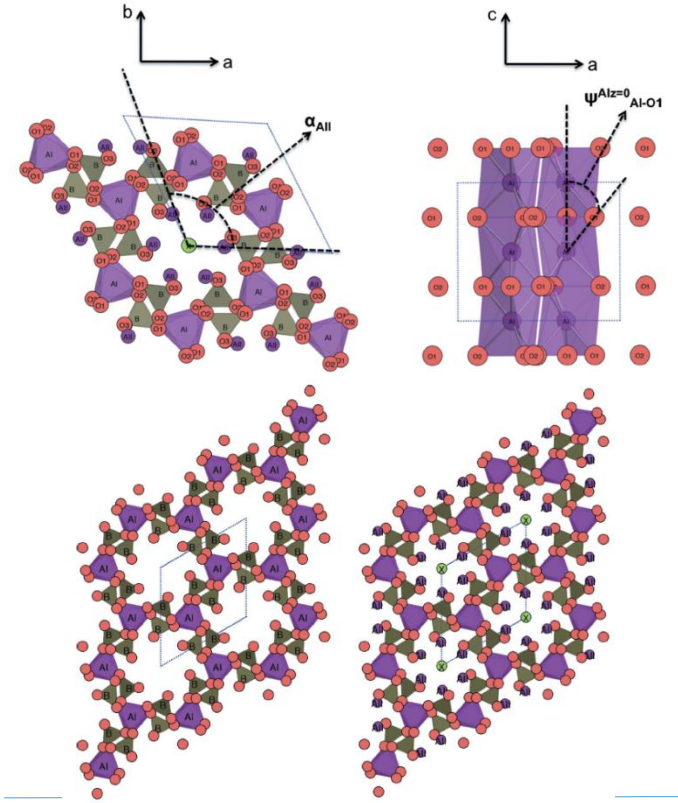
Discovering Classifiers

research papers

Acta Crystallographica Section B
Structural
Science
ISSN 0108-7681

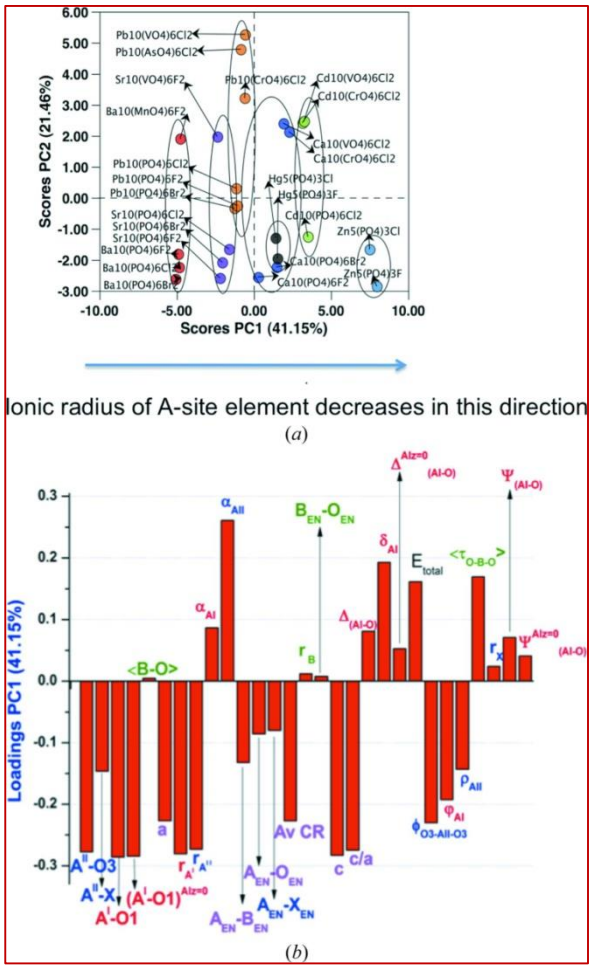
Structure maps for $A^{I}_{4}A^{II}_{6}(BO_4)_6X_2$ apatite compounds *via* data mining

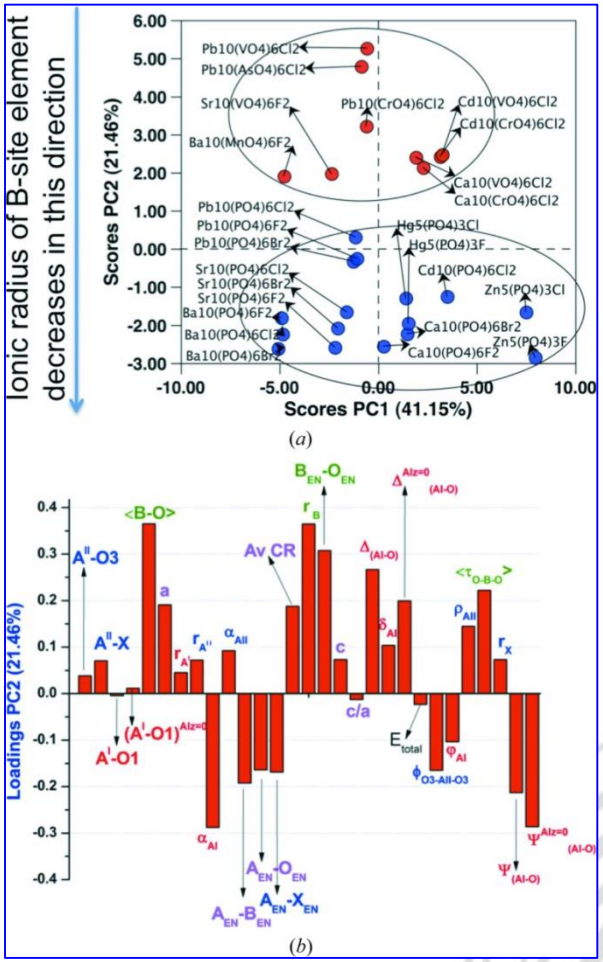
Prasanna V. Balachandran and
Krishna Rajan*

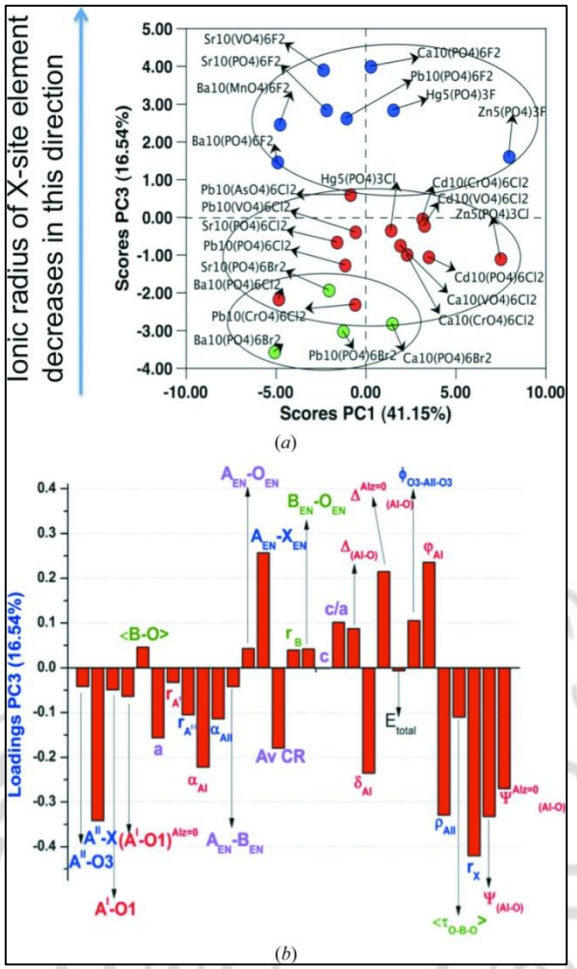


Descriptor	Brief description
a (Å)	Lattice constant of the hexagonal unit cell
c (Å)	Lattice constant of the hexagonal unit cell
c/a	Variable axial ratio (no unit)
r_{AI} (Å)	Shannon's ionic radii of A^{I}-site ion (nine-coordination)
r_B (Å)	Shannon's ionic radii of B-site ion
r_{AII} (Å)	Shannon's ionic radii of A^{II}-site ion (seven-coordination for F^- and eight-coordination for Cl^- and Br^-; Đordević *et al.*, 2008)
r_X (Å)	Shannon's ionic radii of X-site ion
Av CR (Å)	Average crystal radius = $[(r_{AI}x4) + (r_{AII}x6) + (r_Bx6) + (r_Ox24) + r_Xx2)]/42$
$A_{EN} - O_{EN}$	Electronegativity difference A atom and O atom
$B_{EN} - O_{EN}$	Electronegativity difference B atom and O atom
$A_{EN} - X_{EN}$	Electronegativity difference A atom at A^{II} site and X atom
$A_{EN} - B_{EN}$	Electronegativity difference A atom at A^{I} site and B atom
$A^{I}-O1$ (Å)	Distance between A^{I} and O1 atom
$A^{I}-O1^{AIz=0}$ (Å)	Distance between A^{I} and O1 atom with the constraint $z = 0$ at A^{I}
Δ_{AI-O} (Å)	Difference in the lengths $A^{I}-O1$ and $A^{I}-O2$
$\Delta^{AIz=0}_{AI-O}$ (Å)	Difference in the lengths $A^{I}-O1$ and $A^{I}-O2$ with the constraint $z = 0$ at A^{I}
ψ_{AI-O} (°)	The angle that the $A^{I}-O1$ bond makes with respect to c
$\psi^{AIz=0}_{AI-O}$ (°)	The angle that the $A^{I}-O1$ bond makes with respect to c with the constraint $z = 0$ at A^{I}
δ_{AI} (°)	Counter-rotation angle of $A^{I}O_6$ structural unit
φ_{AI} (°)	Metaprism twist angle ($\pi/3 - 2\delta_{AI}$)
α_{AI} (°)	Orientation of $A^{I}O_6$ unit with respect to a
$\langle B-O\rangle$ (Å)	Average $B-O$ bond length
$\langle\tau_{O-B-O}\rangle$ (°)	Average $O-B-O$ bond-bending angle
ρ_{AII} (Å)	$A^{II}-A^{II}$ triangular side length
$A^{II}-X$ (Å)	Distance between A^{II} and X atom
α_{AII} (°)	Orientation of $A^{II}-A^{II}-A^{II}$ triangles with respect to a
$A^{II}-O3$ (Å)	Distance between A^{II} and O3 atom
$\Phi_{O3-AII-O3}$ (°)	$O3-A^{II}-O3$ angle
E_{total} (eV)	Total energy calculated from *ab initio* calculations

Ranking Descriptors

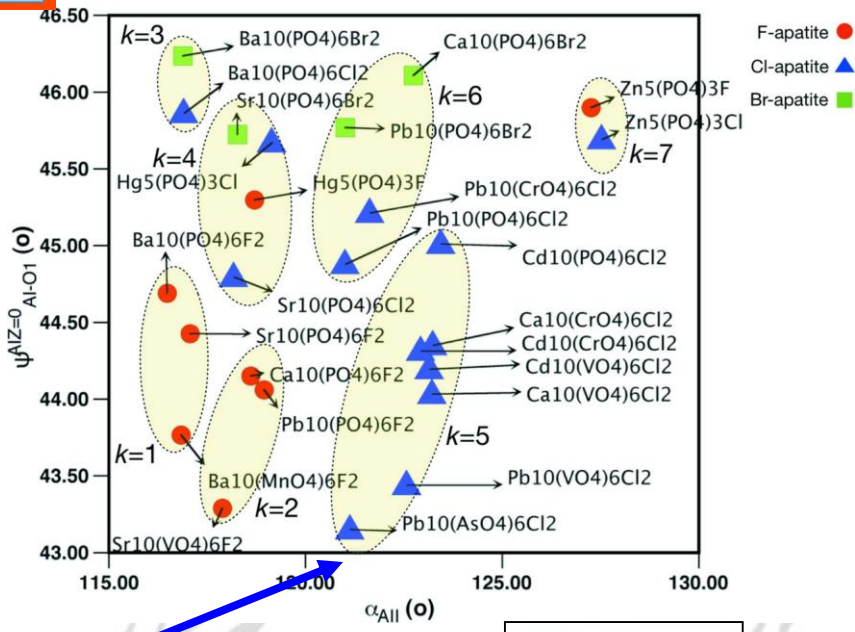




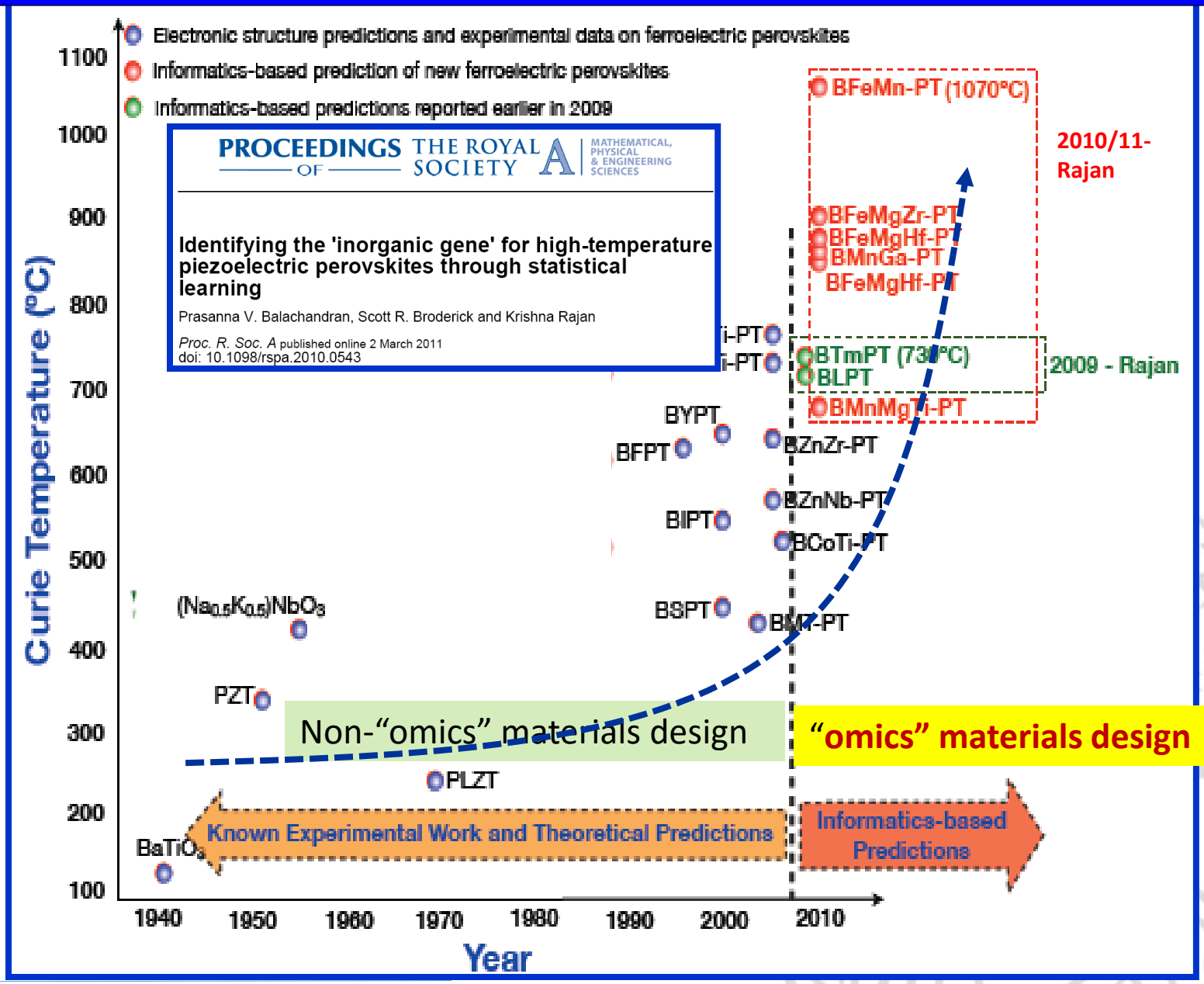


Structure Classification Maps from Data Mining

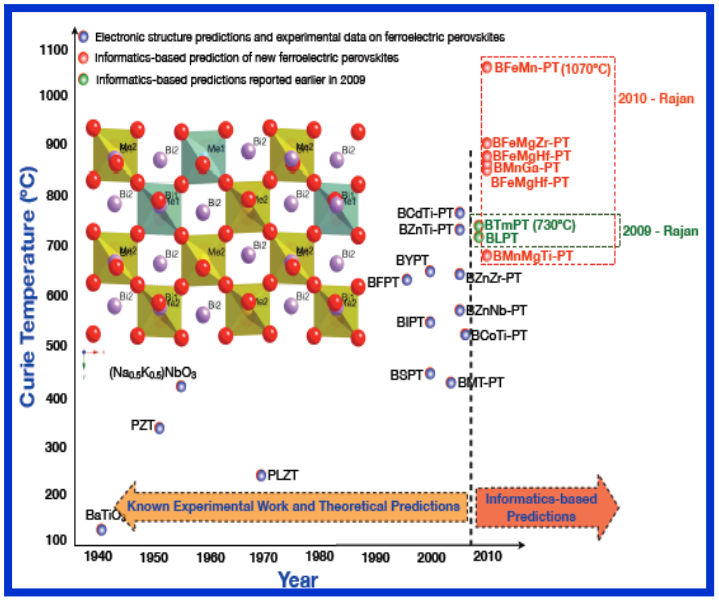
Descriptor	Brief description
a (Å)	Lattice constant of the hexagonal unit cell
c (Å)	Lattice constant of the hexagonal unit cell
c/a	Variable axial ratio (no unit)
r_{AI} (Å)	Shannon's ionic radii of A^{I}-site ion (nine-coordination)
r_B (Å)	Shannon's ionic radii of B-site ion
r_{AII} (Å)	Shannon's ionic radii of A^{II}-site ion (seven-coordination for F^- and eight-coordination for Cl^- and Br^-; Đordević *et al.*, 2008)
r_X (Å)	Shannon's ionic radii of X-site ion
Av CR (Å)	Average crystal radius = $[(r_{AI}x4) + (r_{AII}x6) + (r_Bx6) + (r_Ox24) + r_Xx2)]/42$
$A_{EN} - O_{EN}$	Electronegativity difference A atom and O atom
$B_{EN} - O_{EN}$	Electronegativity difference B atom and O atom
$A_{EN} - X_{EN}$	Electronegativity difference A atom at A^{II} site and X atom
$A_{EN} - B_{EN}$	Electronegativity difference A atom at A^{I} site and B atom
A^I-O1 (Å)	Distance between A^{I} and O1 atom
$A^I-O1^{AIz=0}$ (Å)	Distance between A^{I} and O1 atom with the constraint $z = 0$ at A^{I}
Δ_{AI-O} (Å)	Difference in the lengths A^I-O1 and A^I-O2
$\Delta_{AI-O}^{AIz=0}$ (Å)	Difference in the lengths A^I-O1 and A^I-O2 with the constraint $z = 0$ at A^{I}
ψ_{AI-O} (°)	The angle that the A^I-O1 bond makes with respect to c
$\psi_{AI-O}^{AIz=0}$ (°)	The angle that the A^I-O1 bond makes with respect to c with the constraint $z = 0$ at A^{I}
δ_{AI} (°)	Counter-rotation angle of A^IO_6 structural unit
φ_{AI} (°)	Metaprism twist angle ($\pi/3 - 2\delta_{AI}$)
α_{AI} (°)	Orientation of A^IO_6 unit with respect to a
$\langle B-O \rangle$ (Å)	Average $B-O$ bond length
$\langle \tau_{O-B-O} \rangle$ (°)	Average $O-B-O$ bond-bending angle
ρ_{AII} (Å)	$A^{II}-A^{II}$ triangular side length
$A^{II}-X$ (Å)	Distance between A^{II} and X atom
α_{AII} (°)	Orientation of $A^{II}-A^{II}-A^{II}$ triangles with respect to a
$A^{II}-O3$ (Å)	Distance between A^{II} and O3 atom
$\Phi_{O3-AII-O3}$ (°)	$O3-A^{II}-O3$ angle
E_{total} (eV)	Total energy calculated from *ab initio* calculations

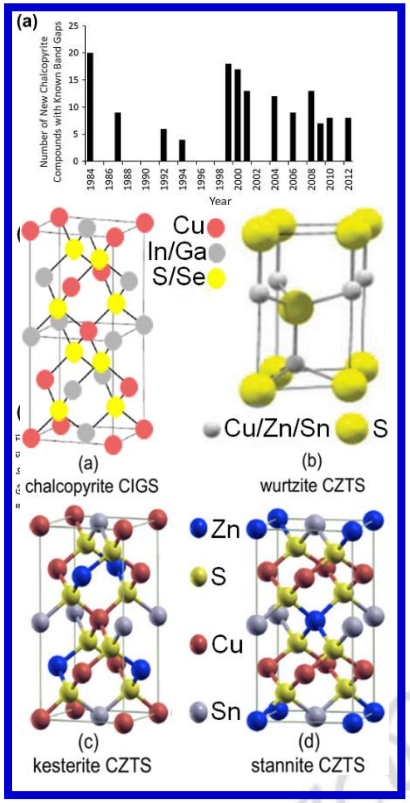


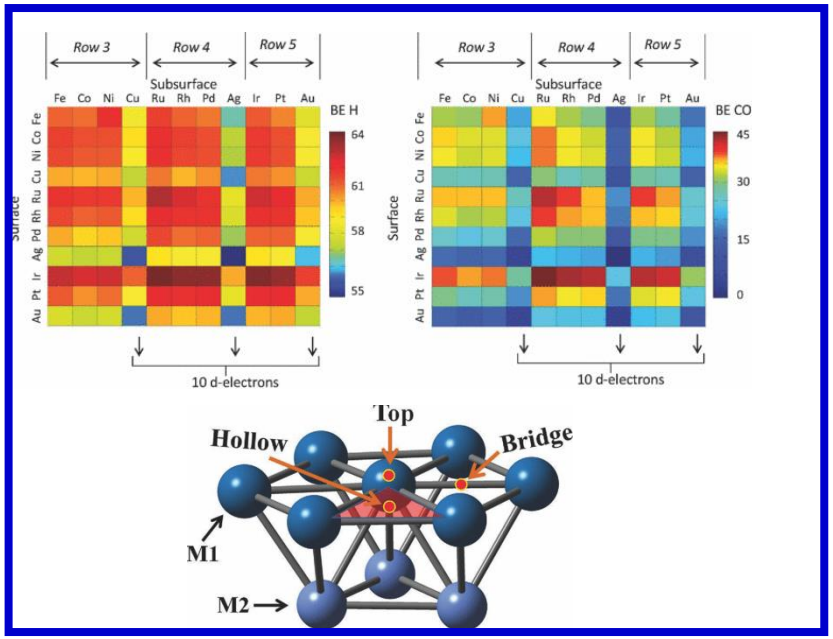
Accelerated Design- value of “omics” design

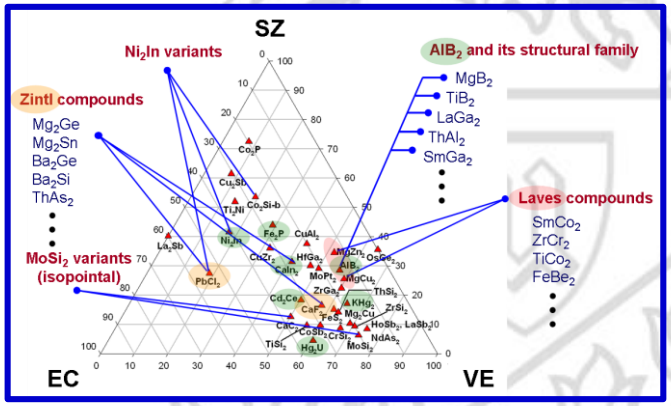


Informatics Driven Materials Discovery





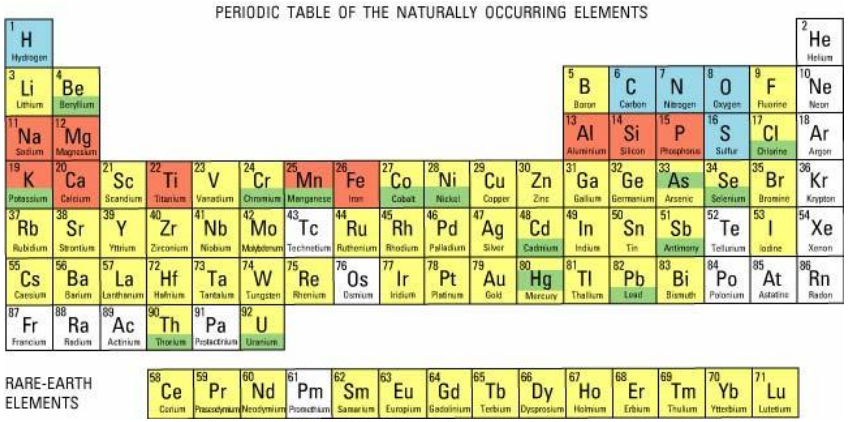


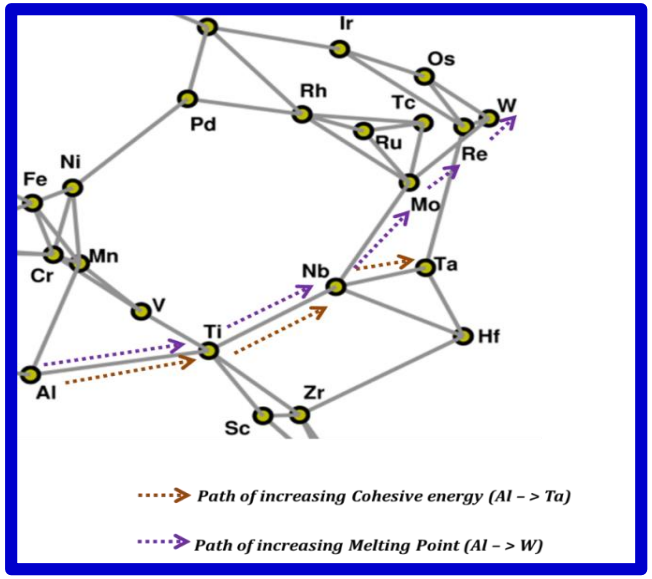


Discovering Chemical Pathways for Microstructure Sensitive Properties – Alloy Design

SCIENTIFIC REPORTS

S.Srinivasan et.al. **5**,17960 doi:10.1038/srep17960(2015)





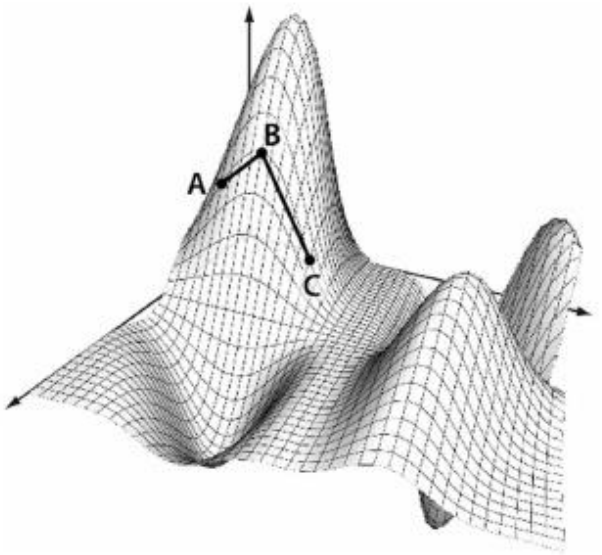
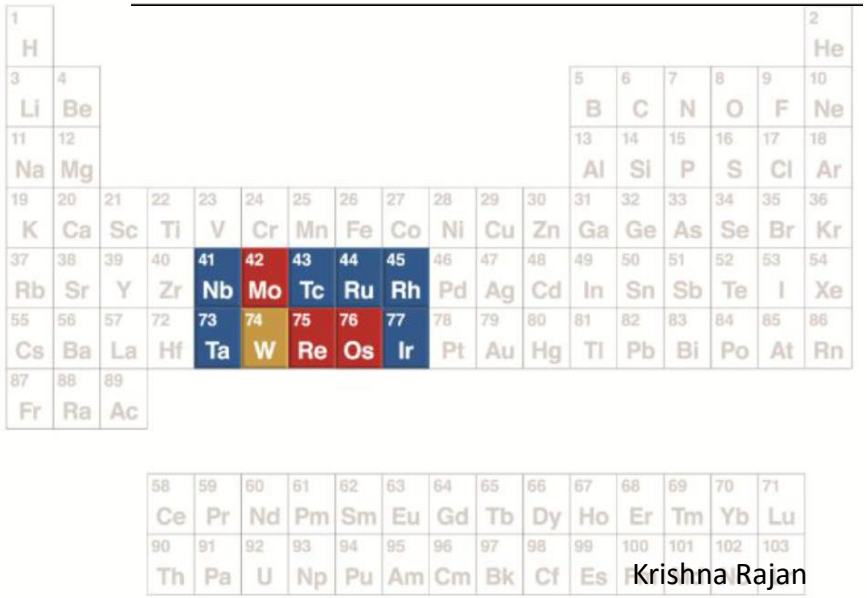


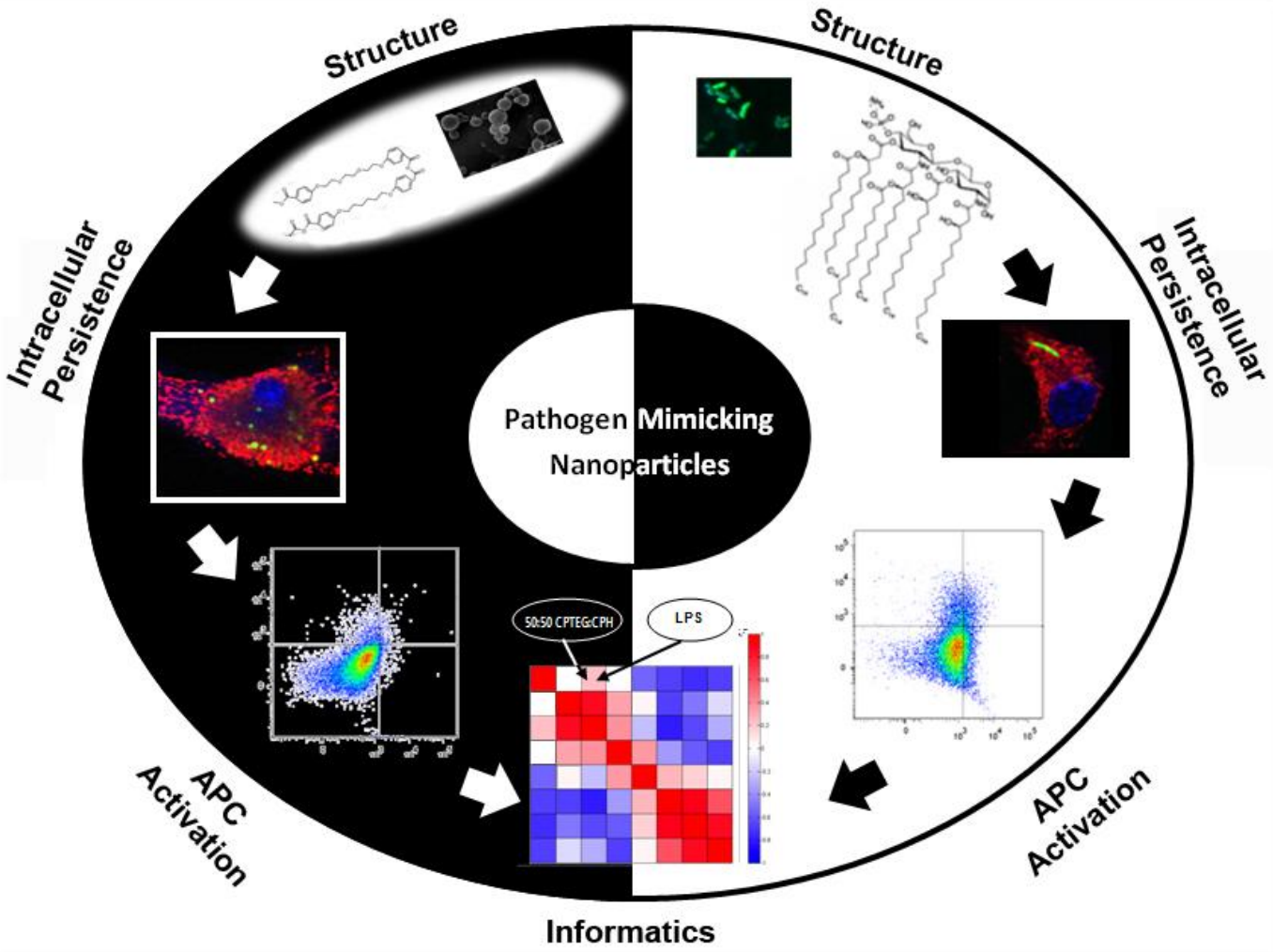
FIGURE 5 | Geodesic distance calculation. The distance between A and C is calculated as the manifold path distance from A to B to C, instead of the direct path from A to C. This eliminates the assumption that the points occupy a linear space when using a Euclidean distance.



SCIENTIFIC REPORTS

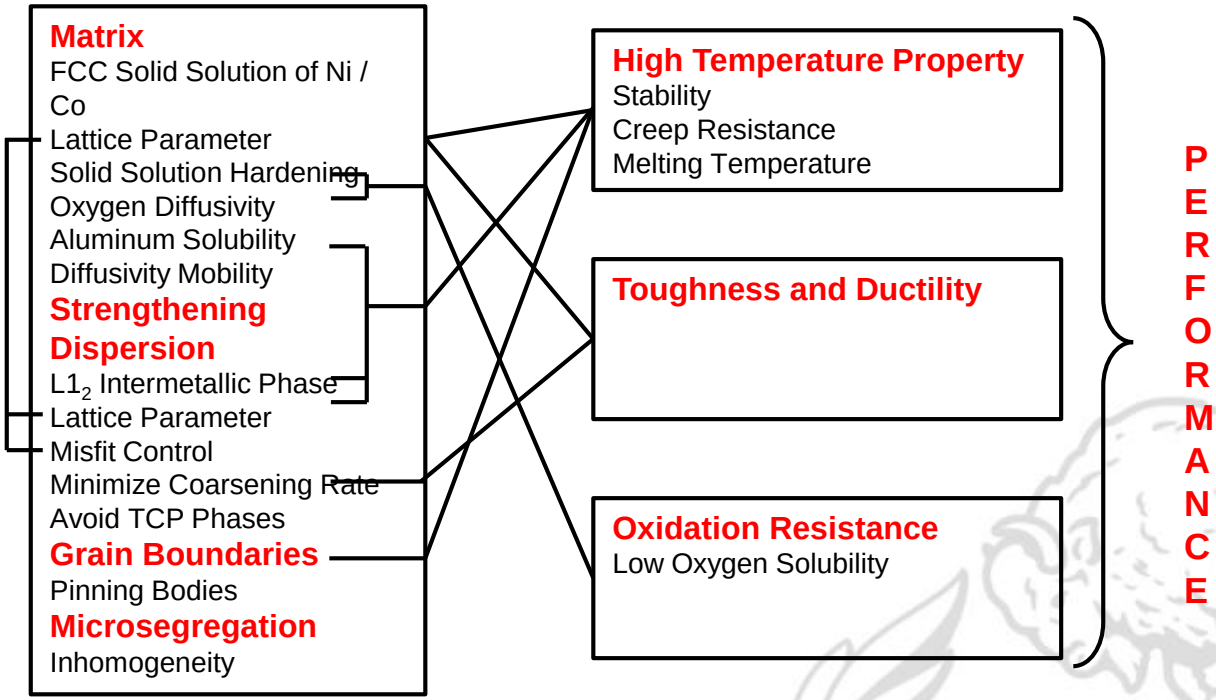
Rational Design of Pathogen-Mimicking Amphiphilic Materials as Nanoadjuvants

Ideal vaccine will mimic the way in which a naturally occurring infection induces a robust immune response yet avoid the undesirable effects of disease



Ulery et.al (2012)

Traditional Design Approach

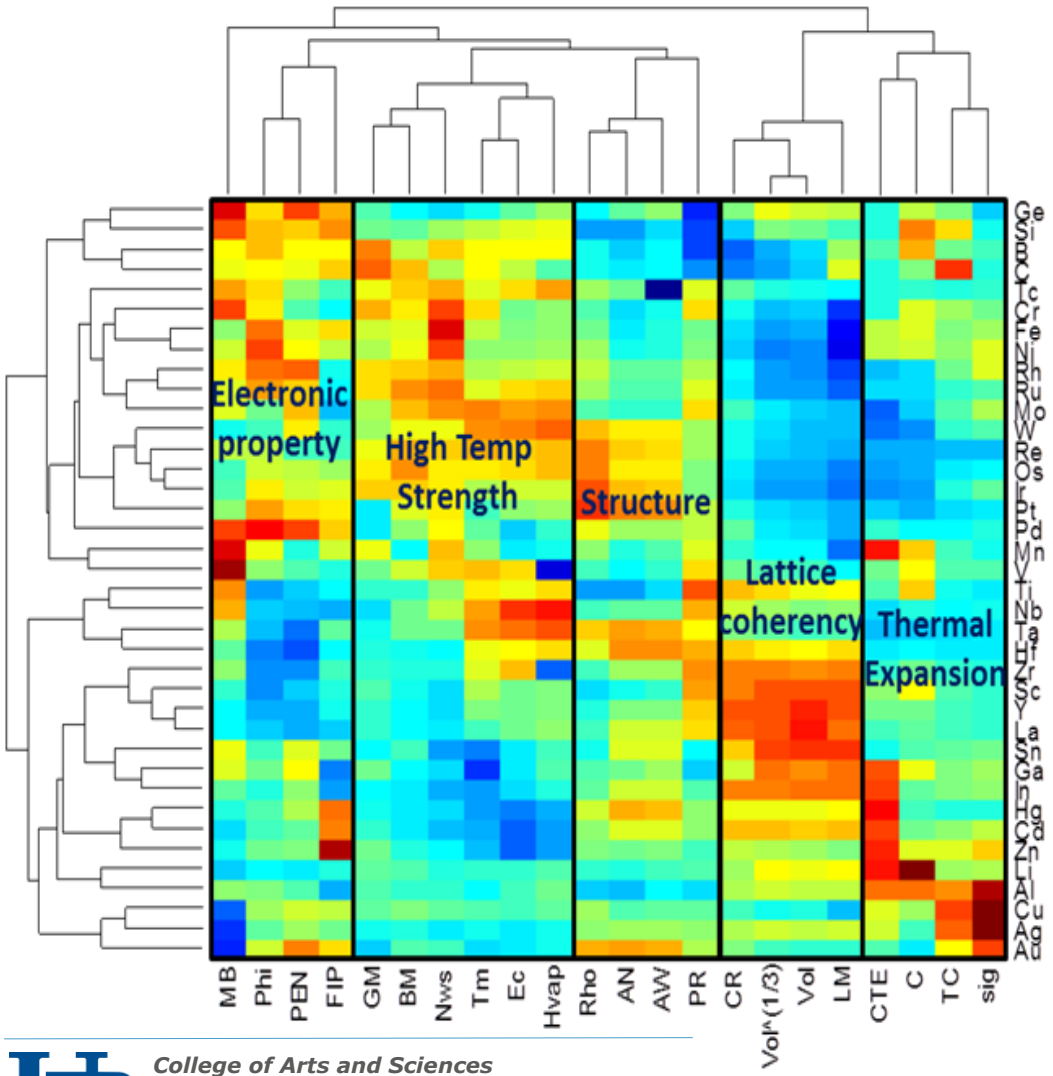


By integrating all the properties into our "alloy design periodic table", we capture the design minutia and make all of these connections without having to do all possible calculations and experiments we would otherwise sequentially do.

Chemical similarity / dissimilarity and materials design

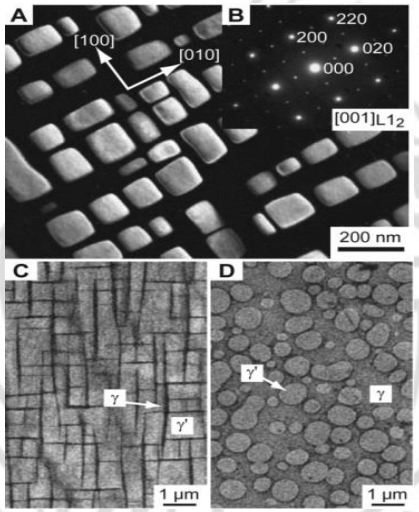
SCIENTIFIC REPORTS

S.Srinivasan et.al. **5**,17960 doi:10.1038/srep17960(2015)



Design constraint – intermetallic chemistry that impacts microstructural level properties :

- L12
- Lattice misfit-
- High temperature strength – modulus / cohesive energy- "Dissimilarity metric"

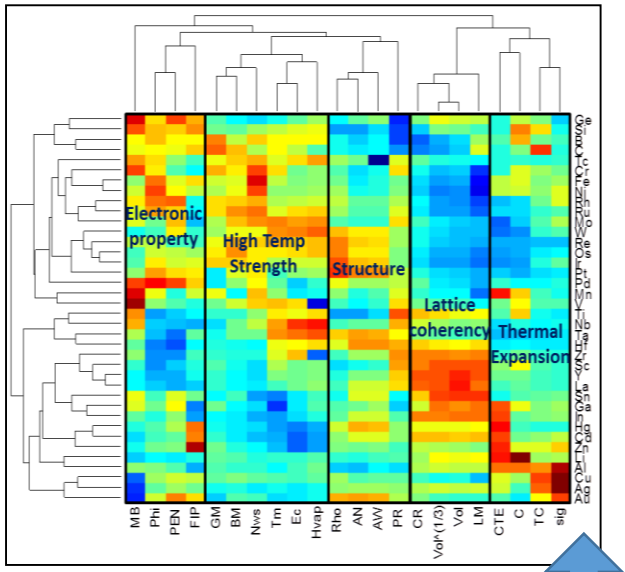


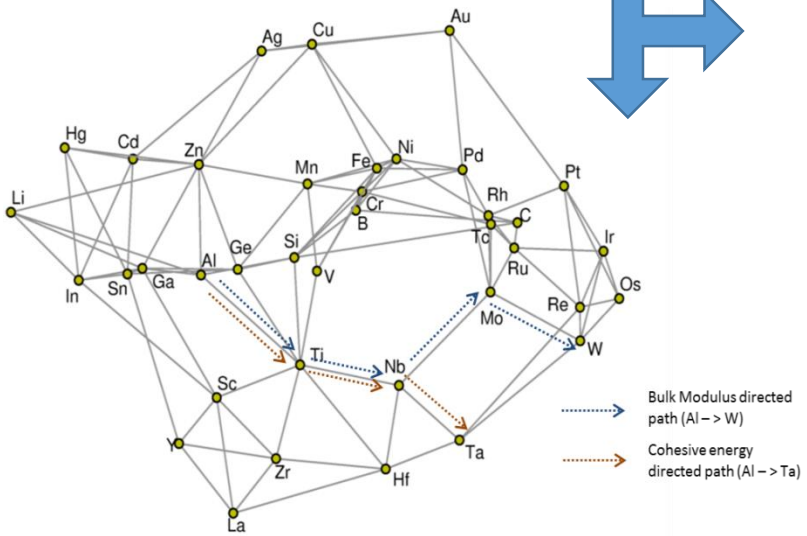
College of Arts and Sciences
School of Engineering and Applied Sciences

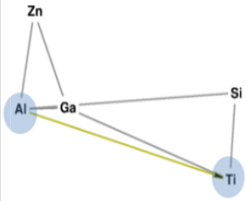
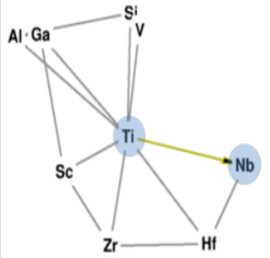
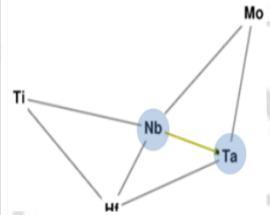
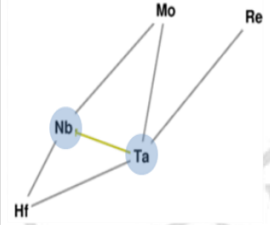
Krishna Rajan

Advances in Methodology

Graph traversal



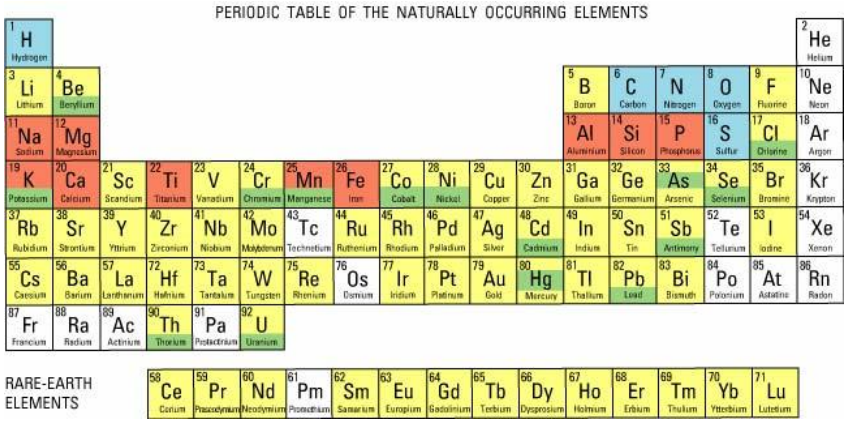


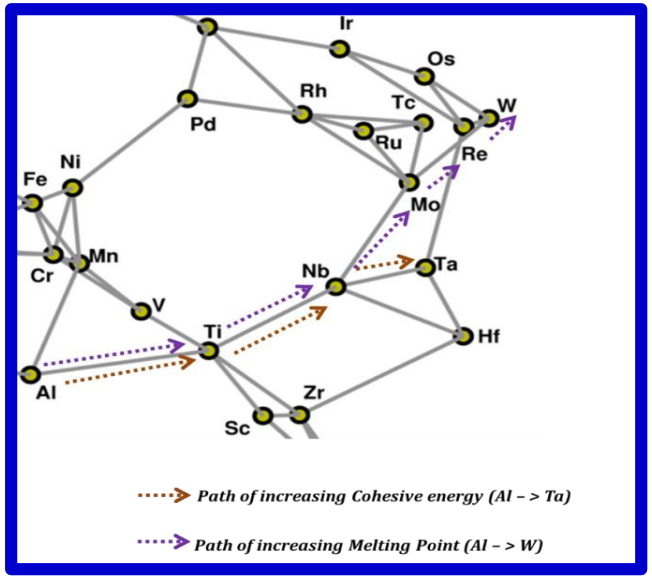
Start	Nearest Neighbor	Enthalpy of Formation	Cohesive Energy	Nearest Neighbor Map
Al	Ga	-0.24	4.23	
	Si	-0.25	4.70	
	Ti	-0.31	4.82	
	Zn	-0.12	3.75	
Ti	Al	-0.31	4.45	
	Ga	-0.24	4.23	
	Hf	-0.40	5.31	
	Nb	-0.27	5.45	
	Sc	-0.34	4.61	
	Si	-0.25	4.70	
	V	-0.15	4.77	
Nb	Hf	-0.40	5.31	
	Mo	-0.05	5.05	
	Ta	-0.27	5.59	
	Ti	-0.31	4.82	
Ta	Hf	-0.40	5.31	
	Mo	-0.05	5.05	
	Nb	-0.27	5.45	
	Re	0.02	5.28	

Discovering Chemical Pathways for Microstructure Sensitive Properties – Alloy Design

SCIENTIFIC REPORTS

S.Srinivasan et.al. **5**,17960 doi:10.1038/srep17960(2015)





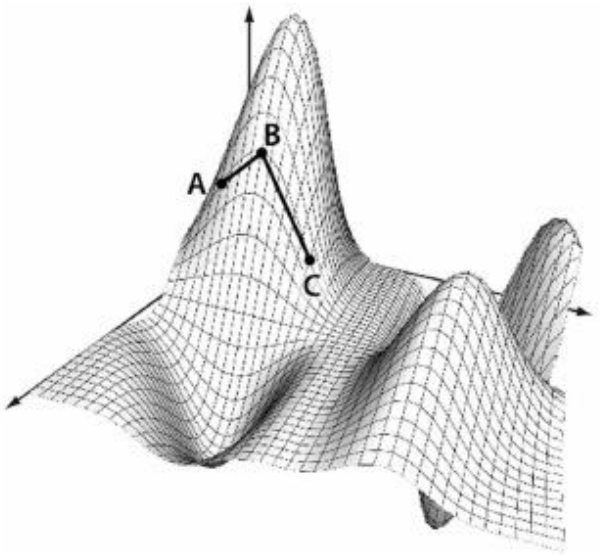


FIGURE 5 | Geodesic distance calculation. The distance between A and C is calculated as the manifold path distance from A to B to C, instead of the direct path from A to C. This eliminates the assumption that the points occupy a linear space when using a Euclidean distance.

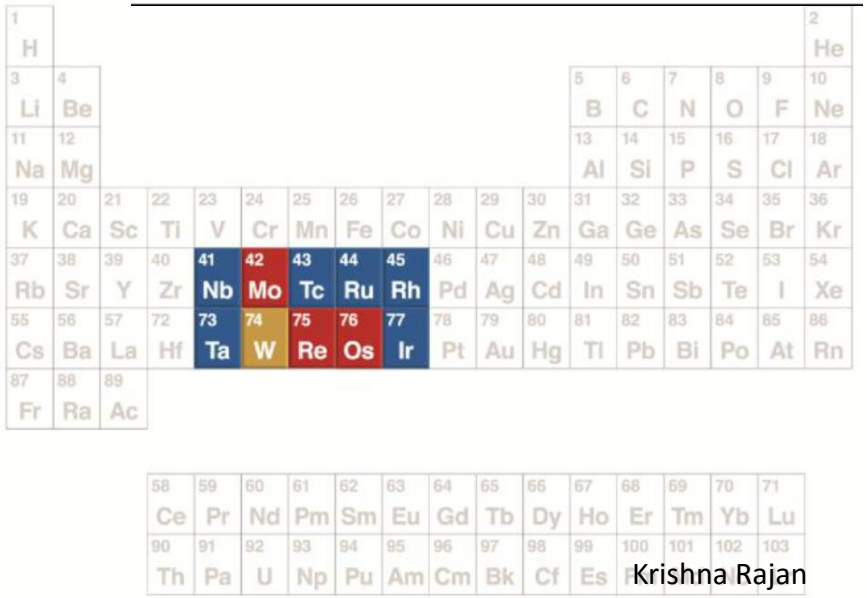


Image interpretation : “inverse problem”…mapping the gene

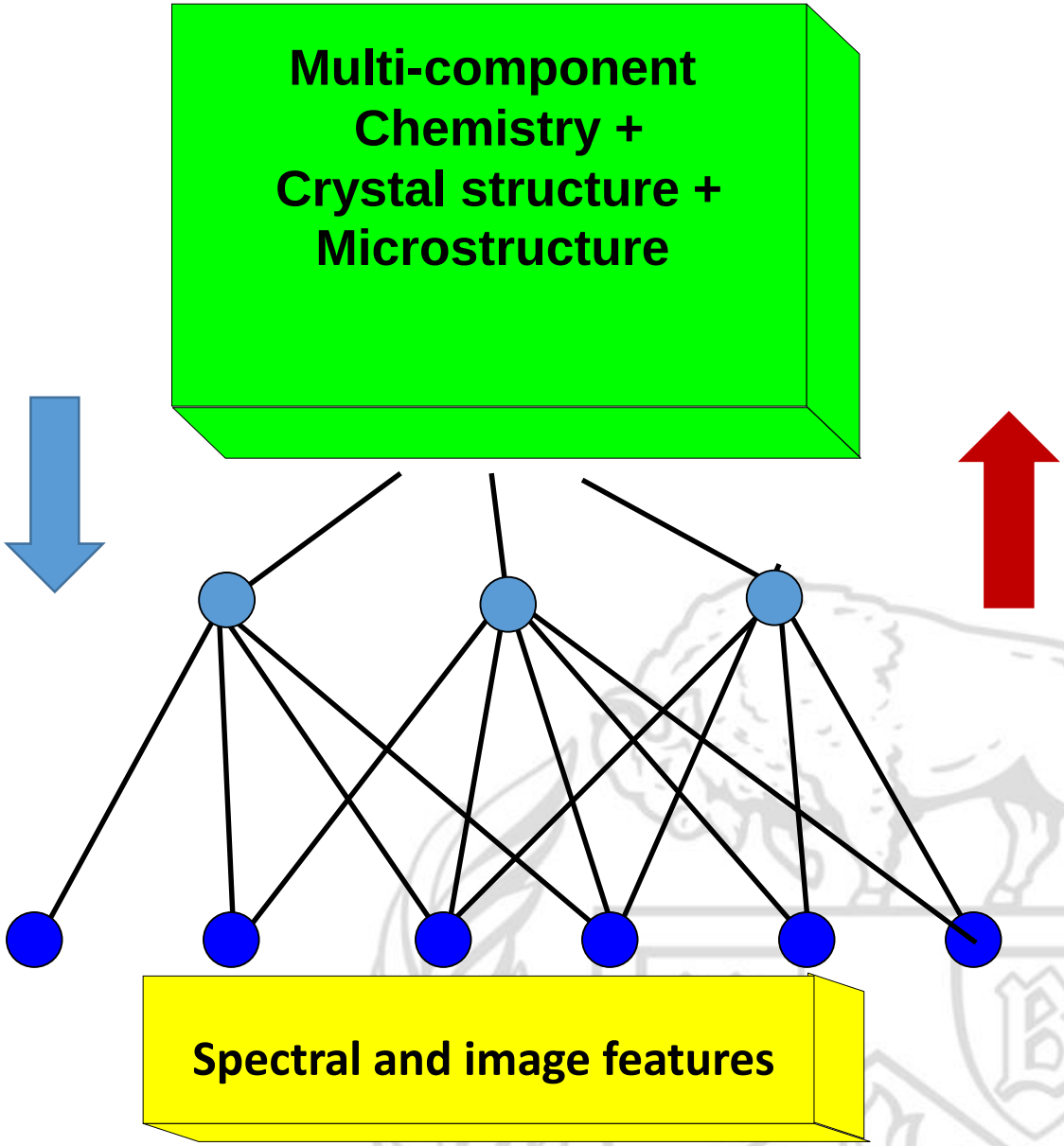
Challenge:

To construct Robust Correlations
between chemistry features and
image features

Methods:

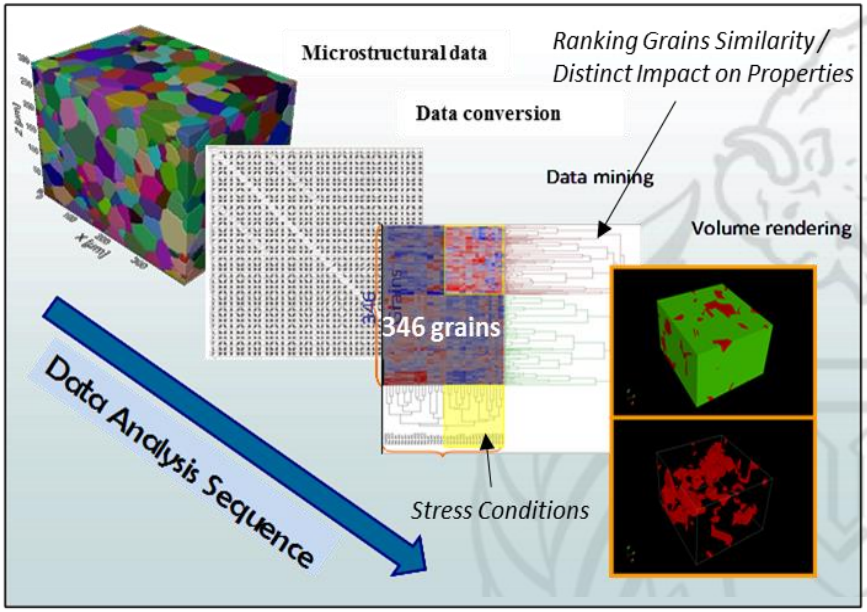
Data analysis
- Statistical learning

Mechanisms of image /spectral formation
- First principles modeling
- Molecular dynamics



Data Driven Image Based Modeling

Data forms: (i) voxel intensity corresponding to different modalities of imaging (orientation – EBSP, chemical – EDS); (ii) distribution functions of stereological metrics of microstructure (grain size distribution); (iii) discrete data and bounds in discrete data (materials properties); (iv) qualitative data (classes of materials, qualitative processing methods)



Atom Probe Tomography : the "Hubble" telescope in materials science

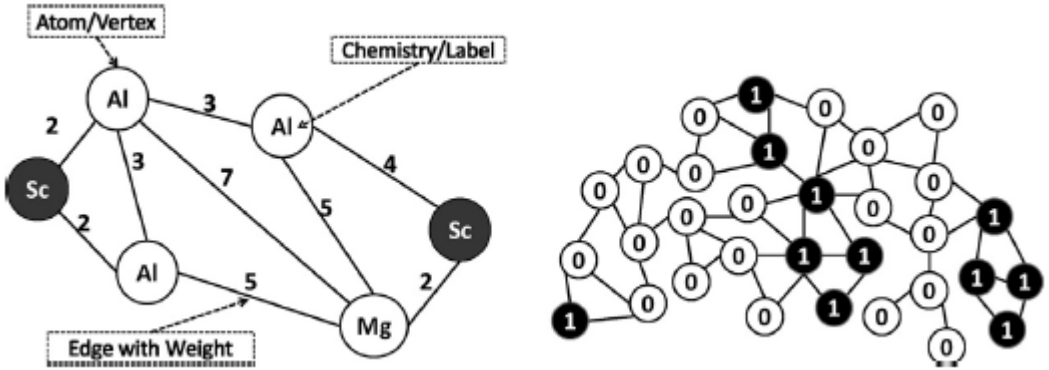
Contents lists available at SciVerse ScienceDirect

Computational Materials Science

journal homepage: www.elsevier.com/locate/commatsci

A graph-theoretic approach for characterization of precipitates from atom probe tomography data

S. Samudrala [a], O. Wodo [a], S.K. Suram [b], S. Broderick [b], K. Rajan [b], B. Ganapathysubramanian [a,*]



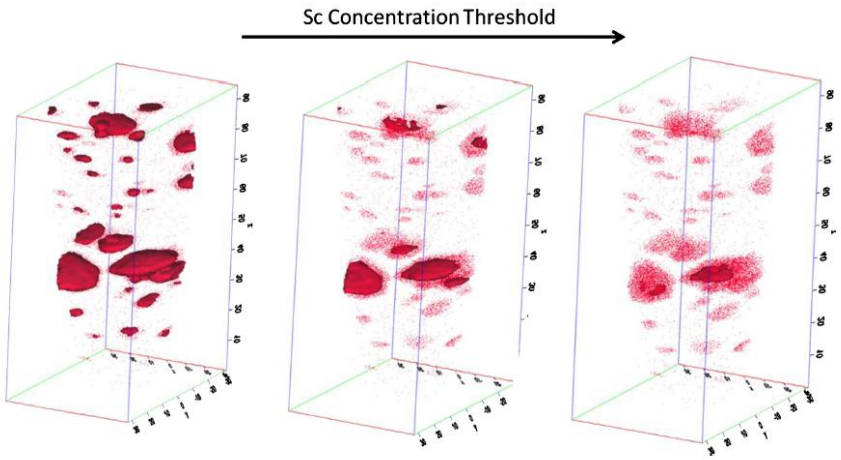
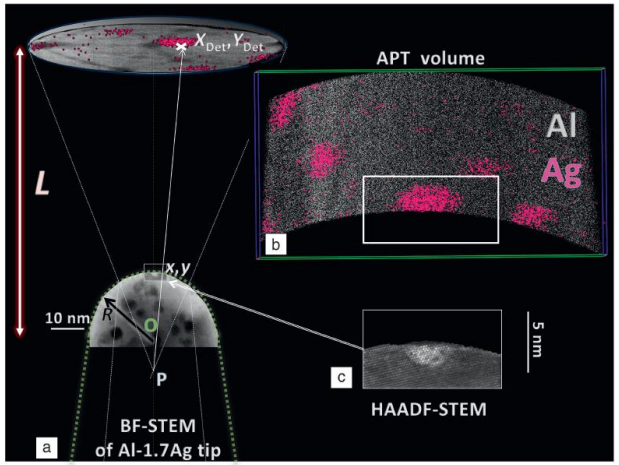


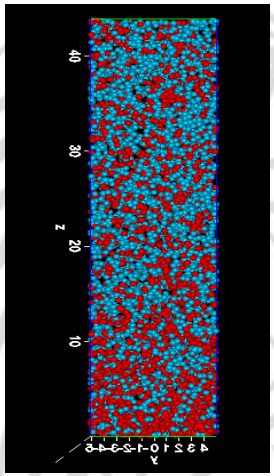
Fig. 13. Concentration isosurfaces of precipitates as a function of Sc concentration threshold value.

Advanced volume reconstruction and data mining methods in atom probe tomography

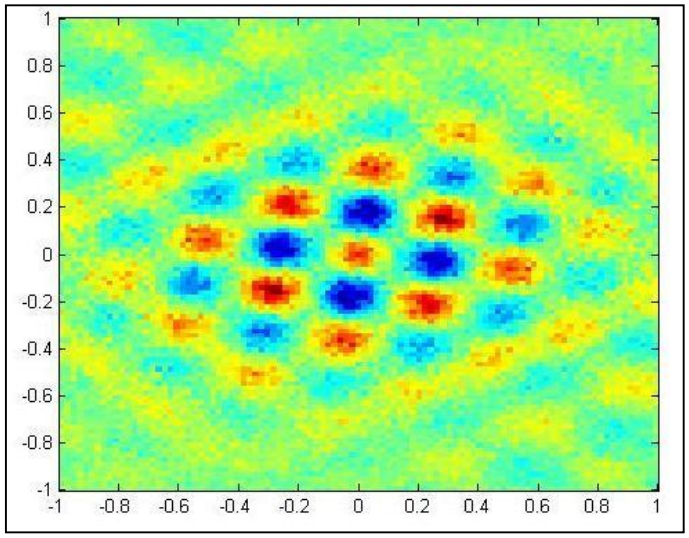
F. Vurpillot, W. Lefebvre, J.M. Cairney, C. Oberdorfer, B.P. Geiser, and K. Rajan

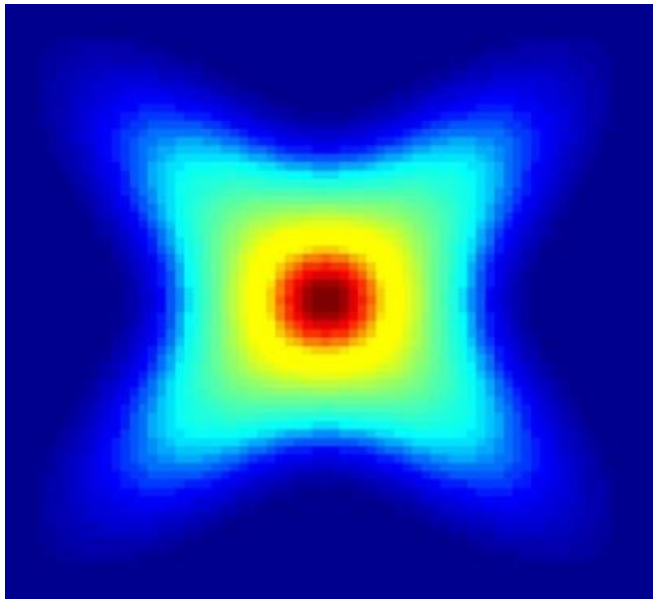
MRS BULLETIN • VOLUME 41 • JANUARY 2016





3 D imaging : reaching the limits of imaging





Isolating structural noise via informatics

Microsc. Microanal. 18, 941–952, 2012
doi:10.1017/S1431927612001171

Microscopy AND Microanalysis

Refining Spatial Distribution Maps for Atom Probe Tomography via Data Dimensionality Reduction Methods

Santosh K. Suram and Krishna Rajan*

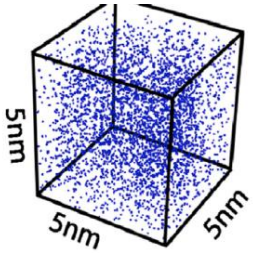
Contents lists available at ScienceDirect

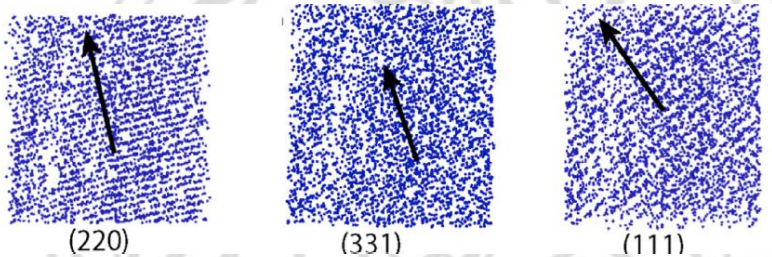
Ultramicroscopy

journal homepage: www.elsevier.com/locate/ultramic

Mining information from atom probe data

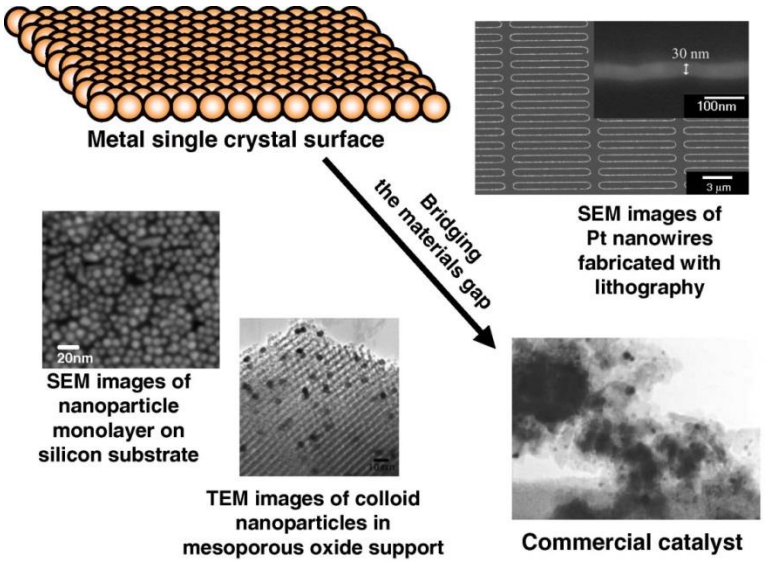
Julie M. Cairney [a,b,*], Krishna Rajan [c], Daniel Haley [d,e], Baptiste Gault [d], Paul A.J. Bagot [d], Pyuck-Pa Choi [e], Peter J. Felfer [a,b], Simon P. Ringer [a,b], Ross K.W. Marceau [f], Michael P. Moody [d]

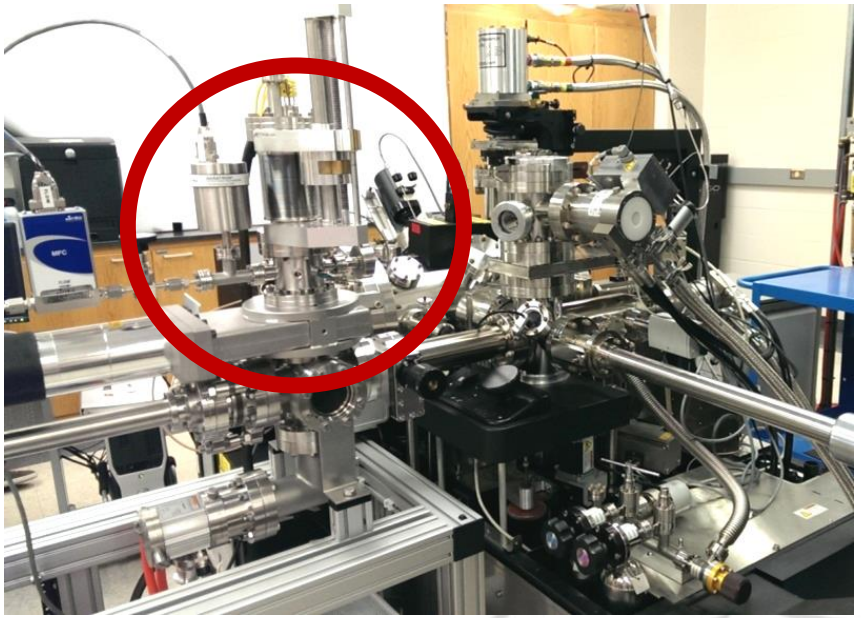


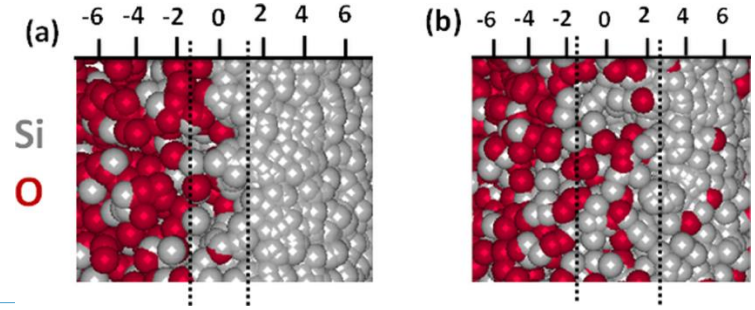


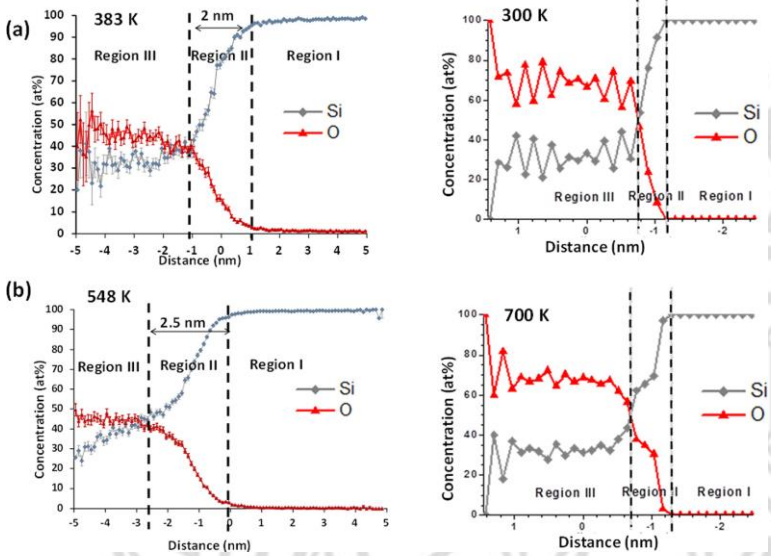
Bridging the Pressure-Materials Gap

G A. Somorjai and J Y. Park Concepts, instruments, and model systems that enabled the rapid evolution of surface science *Surface Science* 603 , 1293-1300 (2009)









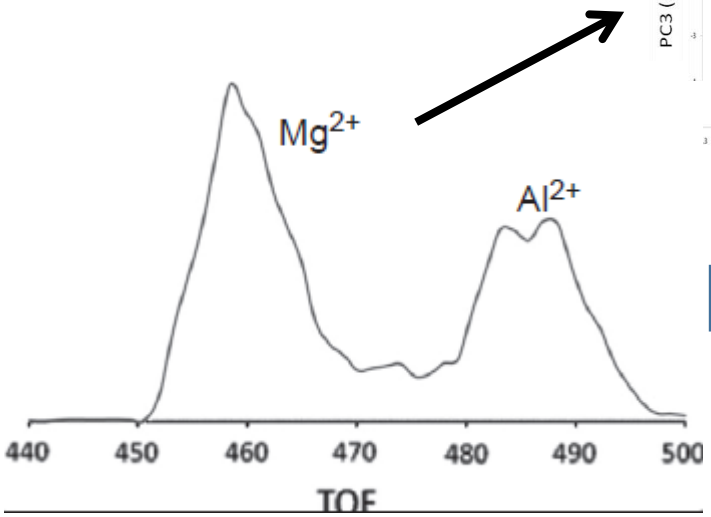
College of Arts and Sciences
School of Engineering and Applied Sciences

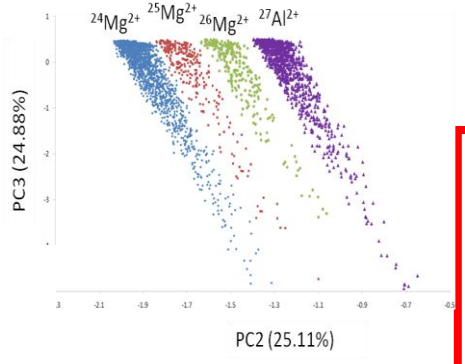
Krishna Rajan

3 D chemical imaging : reaching the limits

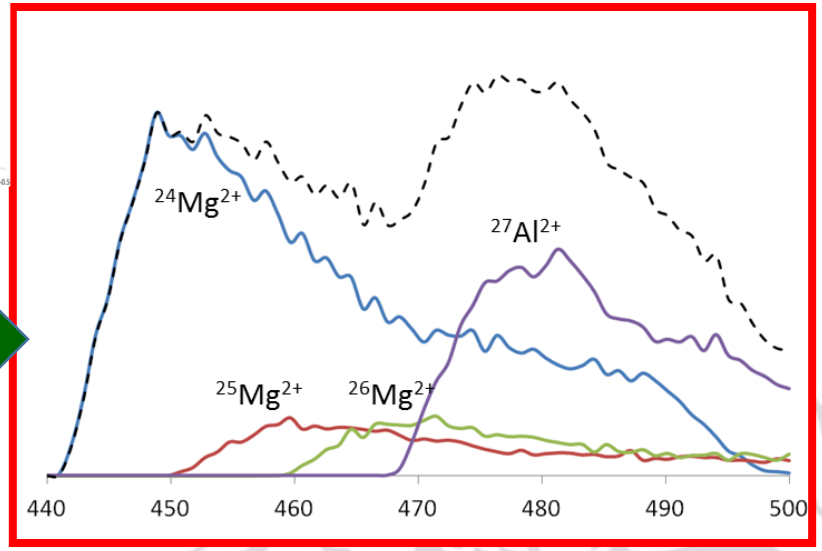
Isolating chemical noise via informatics

Kinetic Energy Discrimination



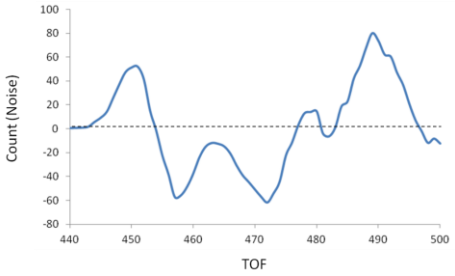


INFORMATICS



Raw Data

Noise



Model Free
Peak Deconvolution

Materials Informatics

Informatics: fusion of data