Informatics for/ from Electronic Structure to Microstructure

Krishna Rajan
Department of Materials Design and Innovation
University at Buffalo- The State University of New York
Outline

- Data for Design
- Learning from Databases
- Mapping Data Pathways
- Informatics / MDA
INFORMATICS

Data management

Data generation

Data visualization

Data dissemination

Data uncertainty

Data mining
"Optics" of Data
Poisson’s ratio and modern materials

G. N. Greaves, A. L. Greer, R. S. Lakes and T. Rouxel

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

Materials Informatics: The Materials “Gene” and Big Data
Annual Review of Materials Research
Vol. 45: 153-169 (July 2015)
1. Identification of sequence of modeling and experiments:
   - Sequence not always known

2. Definition of clear pathways between each design steps
   - Intermediate steps outside pre-defined pathways exist

3. Data needs and corresponding modeling tools defined for each design step
   - Data may be sparse and uneven

Need to be able to integrate information and discover with alternate pathways
<table>
<thead>
<tr>
<th><strong>What?</strong></th>
<th><strong>Why?</strong></th>
<th><strong>How?</strong></th>
</tr>
</thead>
</table>
| Searching for patterns of behavior among multivariate data sets | • Establish new correlations  
• Identify outliers  
• Enlarge database / virtual libraries  
• Evaluate databases  
• Establish predictions | **Data + Correlations + Theory = Knowledge-base** |
| Can pattern recognition lead to predictions? | ![Data compression](image)  
• Pattern recognition  
• Machine learning / Automation | ![atomistic based calculations](image)  
• Continuum based theories |
| Enhancing information resolution at sub nano-scale levels: meeting the challenge of information resolution vs spatial resolution | ![Simulations](image)  
• “Combinatorial” derived datasets  
• Digital libraries  
• Spectral and imaging libraries  
• Focused experiments  
• Heuristics and phenomenological data | ![atomistic based calculations](image)  
• Continuum based theories |
MDA for “Knowledge Bases” vs Data Bases

- Zetabytes
- Terabytes
- Volume
- Velocity
- Veracity
- Variety
- Multiscale
- Reference Data
- High levels of uncertainty
- Single scale

Materials Informatics: The Materials “Gene” and Big Data: K. Rajan
Annual Review of Materials Research
Vol. 45: 153-169 (July 2015)
"Genomic" / data intensive framework for materials functionality

Data Driven Modeling - INFORMATICS

Data

"Hard" Modeling / Experiments

Statistical Learning

Constitutive Equations

PREDICTIVE MODELS /CLASSIFICATIONS /ONTOLOGIES

"Materials Informatics: The Materials "Gene" and Big Data – K. Rajan
Annual Review of Materials Research
Vol. 45: 153-169 (July 2015)
MDA: discovering need for new data


Statistical removing of outliers:
- better models from same data
- new design rationale for metal oxide frameworks
MDA: Discovery of reference data

Accelerated reference data discovery

Outlier detection

Classification detection

http://iusti.polytech.univ-mrs.fr/MOLTEN_SALTS/
High Dimensional Data Imaging:

Specimen Parameters
(Tip Radius, Shank Angle, Ellipticity)

Instrument Parameters
(Pulse Energy, Pulse Fraction, Temperature etc.)

Depth and Lateral Resolution

Noise

Trajectory Aberrations

Missing Data (Detector Efficiency)

Structural Information

Atom Probe Data
MDA: Optimizing data volume

Predator-Prey models
Ranking similarity of nanoparticle adjuvants to pathogen and identified chemical mechanism contributing to pathogen-like response
Discovering Systematics in Databases

Accelerated Discovery - tracking statistical impact of descriptors

Informatics-aided bandgap engineering for solar materials

(a) Number of New Chalcopryite Compounds with Known Band Gaps

(b) Periodic Table with Chalcogenides

(c) Chalcopyrite CIGS

(d) Wurtzite CZTS

(e) Kesterite CZTS

(f) Stannite CZTS

Number of new compounds identified based on QSAR type models

Materials Data Analytics: A Path-Finding Workshop
October 8-9, 2015
Designer Surfaces

J. Chem. Phys. 140, 094705 (2014);

Informeatics guided discovery of surface structure-chemistry relationships in catalytic nanoparticles
REVIEW • OPEN ACCESS
Informatics derived materials databases for multifunctional properties
Scott Broderick and Krishna Rajan
Published 13 January 2015 • Science and Technology of Advanced Materials, Volume 16, Number 1

Krishna Rajan
Erich Bloch Chair & Empire Innovation Professor

Identification of the 'inorganic gene' for high-temperature piezoelectric perovskites through statistical learning
Prasanna V. Balachandran, Scott R. Broderick and Krishna Rajan
Proc. R. Soc. A published online 2 March 2011
doi: 10.1098/rspa.2010.0543
Microstructural Design Pathway of ICME

Aerospace ICME Toolset

8. Develop or Obtain Models for Predicting Processing Outcomes
   - V&V'd by Experimental Data

9. Use Verified and Validated Models to Predict Processing Outcomes
   - V&V'd by Experimental Data

10. Use Verified and Validated Models to Predict Processing Outcomes
    - Model examples include:
      - MeltFlow-VAR
      - COMPRO 3D
      - FLUENT
      - DEFORM
    - Experiment examples include:
      - VAR/ESR melting
      - Heat treatment, forging, machining etc. followed by characterization
      - Optical pyrometry

11. Linking Tools

12. Develop or Obtain Models for Predicting Microstructure
    - Yes
    - No

13. V&V'd by Experimental Data
    - Yes
    - No

14. Utilize Verified and Validated Models (Tools) to Predict Microstructure
    - Model examples include:
      - VASP (Vienna Ab Initio Simulation Package)
      - LAMMPS (Large Scale Atomic/Molecular Massively Parallel Simulator)
    - Experiment examples include:
      - TEM (crystal structure/interfaces)
      - 3D atom probe
      - X-ray tomography
      - SEM

15. Linking Tools

16. Develop or Obtain Models for Predicting Materials/Component Properties
    - V&V'd by Experimental Data
    - Yes
    - No

17. Linking Tools

18. Utilize Verified and Validated Models (Tools) to Predict Materials/Component Properties
    - Model examples include:
      - Thermocal/PANDAT
      - HSC Chemistry
      - ANSYS
      - JMatPro
    - Experiment examples include:
      - Heat treatments
      - Neutron diffraction
      - X-ray diffraction
      - Creep testing
      - Corrosion testing

KEY:
- Suits of models are identified or developed (8,12,16)
- Models are verified and validated [V&V'd] in iterative processes (9,13,17)
- Linking Tools transmit data between models (11,15,19)
- Suits of V&V'd computational tools are applied to the specific product development (10,14,18)
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.
Available data
Target design assessment function(s)
Data integration pathways
Data vs Knowledge
Building the Workforce

Materials Design and Innovation
School of Engineering and Applied Sciences &
College of Arts and Sciences

1\textsuperscript{st} department dedicated to \textbf{Data Driven Knowledge Discovery} in materials science
Informatics: data fusion for Knowledge Discovery