Data Infrastructure to Meet the Prediction Challenge

Anne Chaka

Physics Laboratory

National Institute of Standards and Technology

Materials by Design and EHS Depend On Chemical and Physical Property Data

- Published evaluated chemical and physical property data doubles every 10 years.
- Woefully inadequate to keep up with demand. For nano it's even worse.
- Requires meticulous experimental measurements and/or thorough evaluations of related data from multiple sources
- Data acquisition processes are time- and resource-consuming ➲ must be initiated well prior to an anticipated need within an industrial or scientific application
- A significant part of the existing data infrastructure is not directly used in any meaningful application, because data requirements often shift between the initiation and completion of a data project
- Analysis and fitting, such as for EOS, QSAR models, must be reinitiated when significant new data become available

Simulations have the potential to fill the gap for both performance and EHS.

Data Infrastructure to Address the Prediction Challenge

The Problem: Computer models and simulations can yield inconsistent and often contradictory results

- *Range of validity of models is unknown. Models developed for point solutions are not transferable.*
- *Academic researchers and software companies don't have experimental capability to validate simulation tools, and are not rewarded for establishing where methods fail.*
- *Only the industrial elite and large national programs have been able to run expensive experiments to validate models over a narrow range of applicability.*
- *Physics-based models are predictive, but are so computationally expensive that without HPC and extensive experimental data they can only be run on toy problems of limited relevance.*
- *Obtaining systematic experimental data to validate simulations is not glamorous. Very difficult to fund.*

Need: Experimental data infrastructure for validation of models and simulation to drive the science forward

- *New physical measurements needed to delineate underlying science, test limits of models, and support applications of practical relevance*
- *Methodology, tools for verification, validation, and uncertainty assessment of models and simulations*

"The prediction challenge is now the most serious limiting factor for computational science…New methods of verifying and validating complex codes are mandatory if computational science is to fulfill its promise for science and society." **–D.E. Post (LANL) and L.G. Votta (Sun Microsystems), Physics Today, January 2005.**

"…further development [of verification and validation and uncertainty quantification] will have a profound impact on the reliability and utility of simulation methods in the future." **NSF Blue Ribbon Panel, 2006.**

"The issue of model validation is crucial to regulatory approval of HPC models as an alternative to physical testing." **–Council on Competitiveness, 2004.**

The Problem in Materials Design

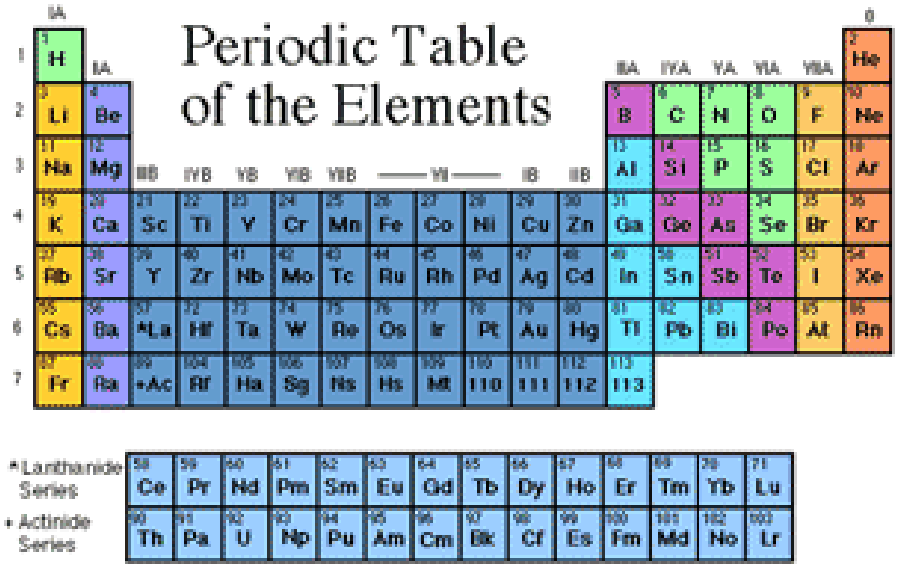
If we are interested in perhaps 50 elements on the periodic table, there are:

2,500 binary combinations, 125,000 ternary combinations, etc. . . .

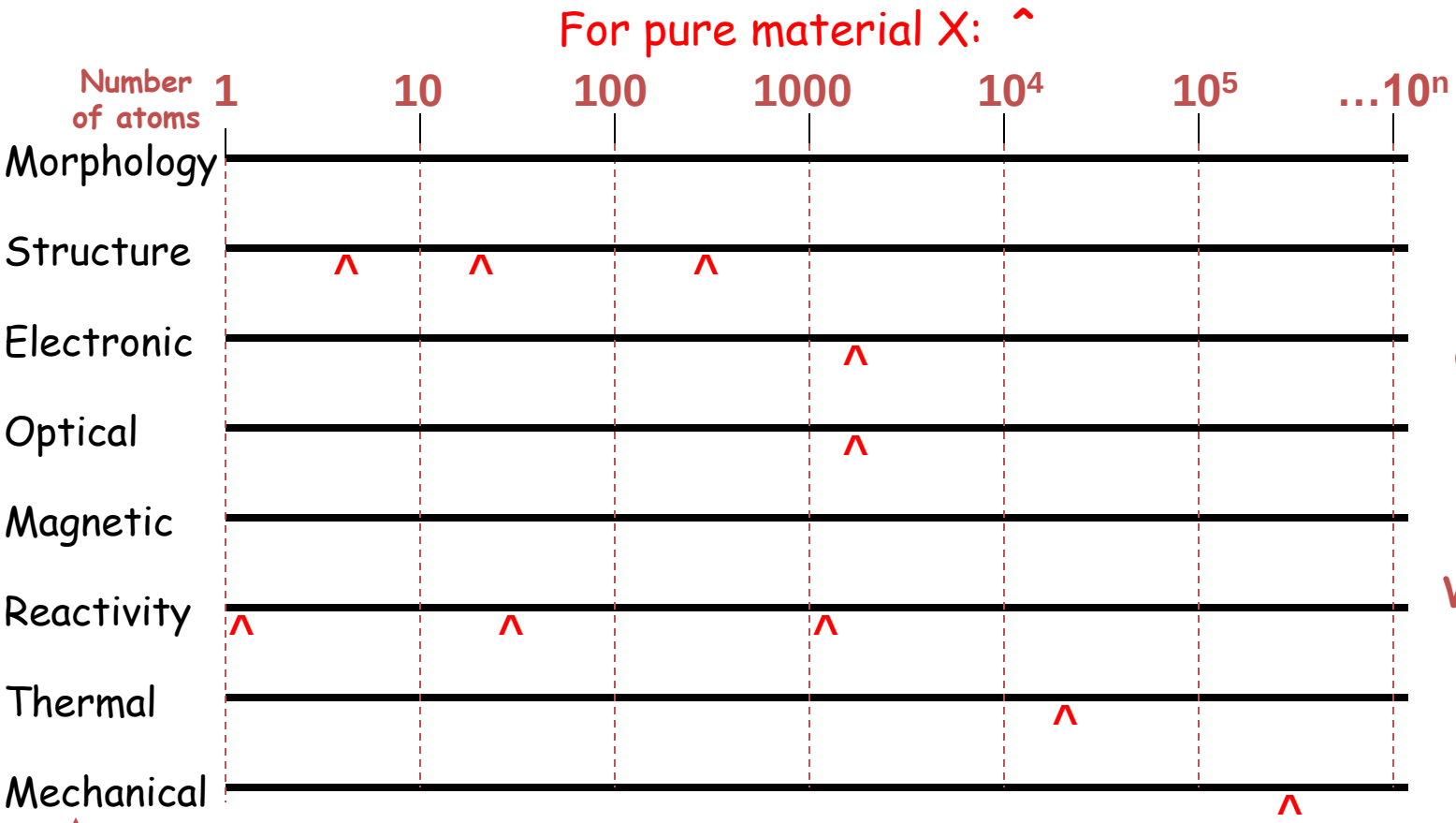
How to vary concentration and distribution? Materials with identical chemical compositions can have totally different properties, depending upon processing and dimensions (nano vs. macro)

The ever-increasing pace of science and technology has already outstripped our ability to produce needed data.

Screening classes of materials prior to a synthesis and testing cycle is essential.



The Problem for Design of Nanoproperties: As size decreases, at what point do qualitatively different properties emerge? How do these points change with material and property? With temperature, pressure, and environment?



What can we currently measure to provide data and guidance?

What can we calculate?

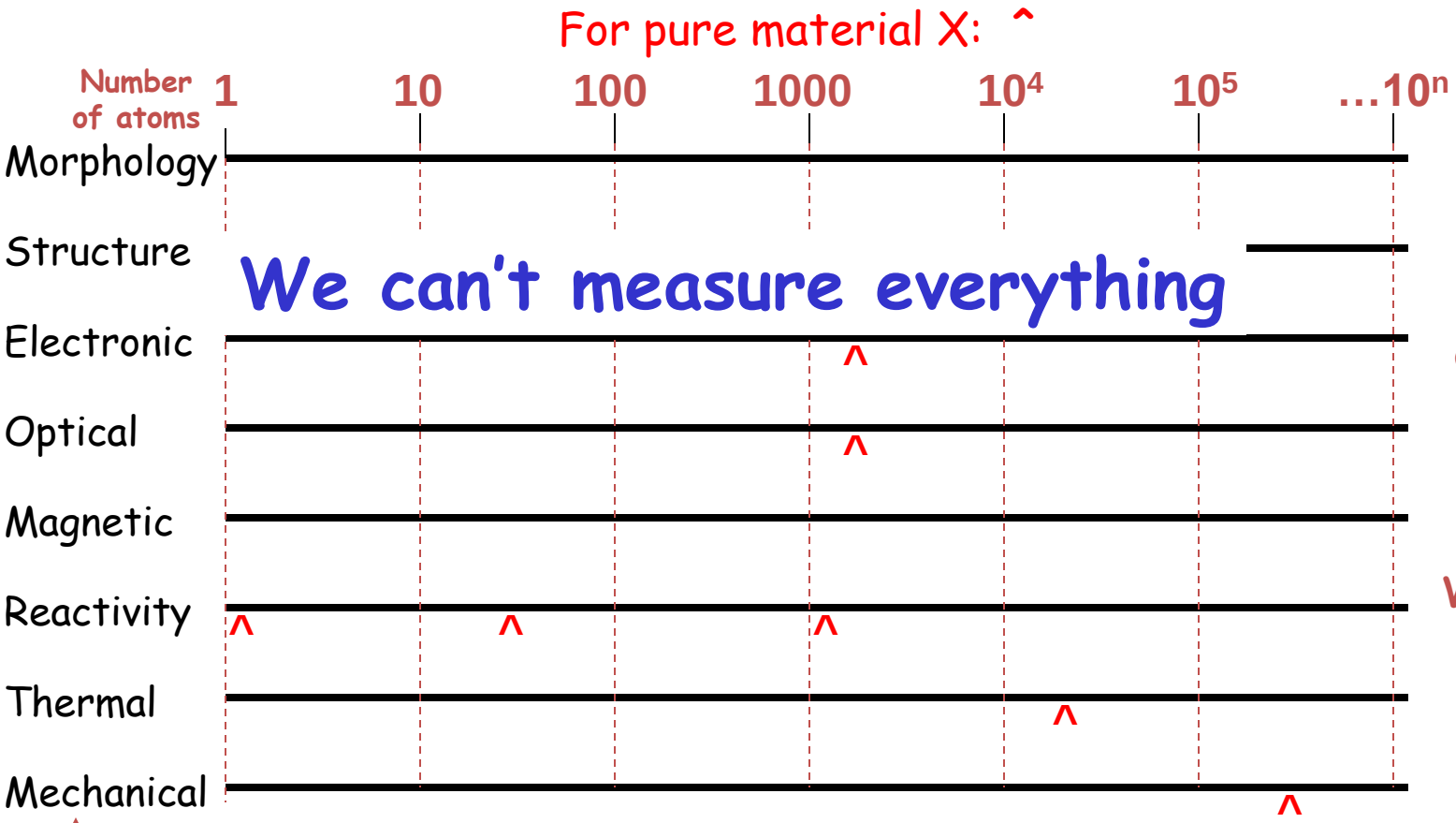
Where are the gaps?

All these properties may be important for device or particle performance

How do these points change as X is doped with Y?
For other AB binaries? For XYA ternary compounds....?

What is reasonable to try next? (For industrial design)
What is reasonable to measure next? (For NIST to develop SR data)

The Problem for Design of Nanoproperties: As size decreases, at what point do qualitatively different properties emerge? How do these points change with material and property? With temperature, pressure, and environment?



What can we currently measure to provide data and guidance?

What can we calculate?

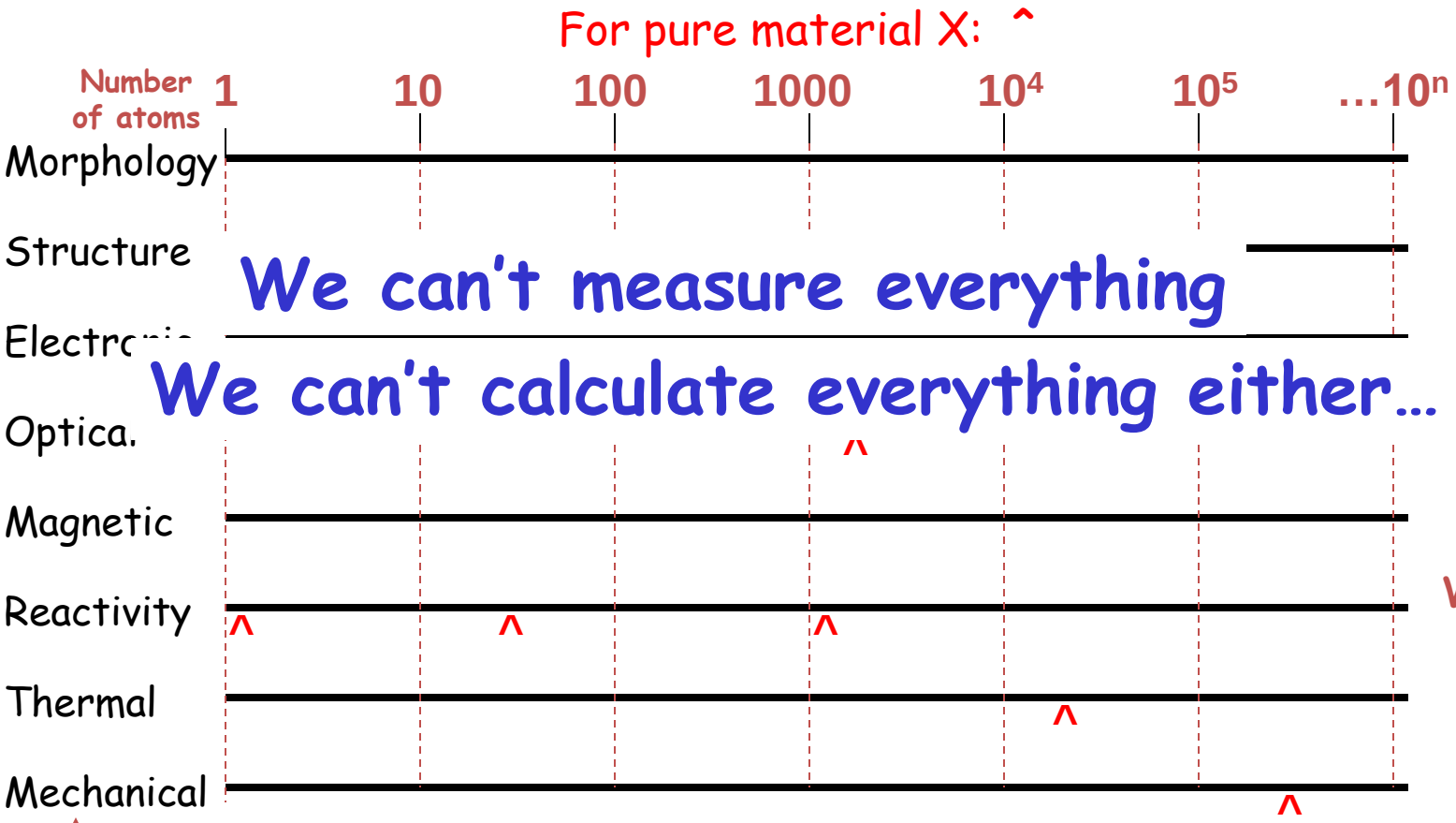
Where are the gaps?

All these properties may be important for device or particle performance

How do these points change as X is doped with Y?
For other AB binaries? For XYA ternary compounds....?

What is reasonable to try next? (For industrial design)
What is reasonable to measure next? (For NIST to develop SR data)

The Problem for Design of Nanoproperties: As size decreases, at what point do qualitatively different properties emerge? How do these points change with material and property? With temperature, pressure, and environment?



What can we currently measure to provide data and guidance?

What can we calculate?

Where are the gaps?

All these properties may be important for device or particle performance

How do these points change as X is doped with Y?
For other AB binaries? For XYA ternary compounds….?

What is reasonable to try next? (For industrial design)
What is reasonable to measure next? (For NIST to develop SR data)

Industrial Fluid Properties Simulation Collective

How good is that molecular simulation for calculating a physical property for organic or biological molecules?

NIST - in conjunction with industry - administers a blind simulation challenge every year.
The results were shocking.

What we thought was easy to do, could be very wrong.

Example: Determine the heat of mixing of
n-butylamine and water

CH_3-CH_2-CH_2-CH_2-NH_2

3M DUPONT DOW ExxonMobil

http://FluidProperties.org

Butylamine in water results underscore the greatest challenge for reliable predictions

For reliable molecular simulations and property predictions, the balance has to be right between *all* terms:

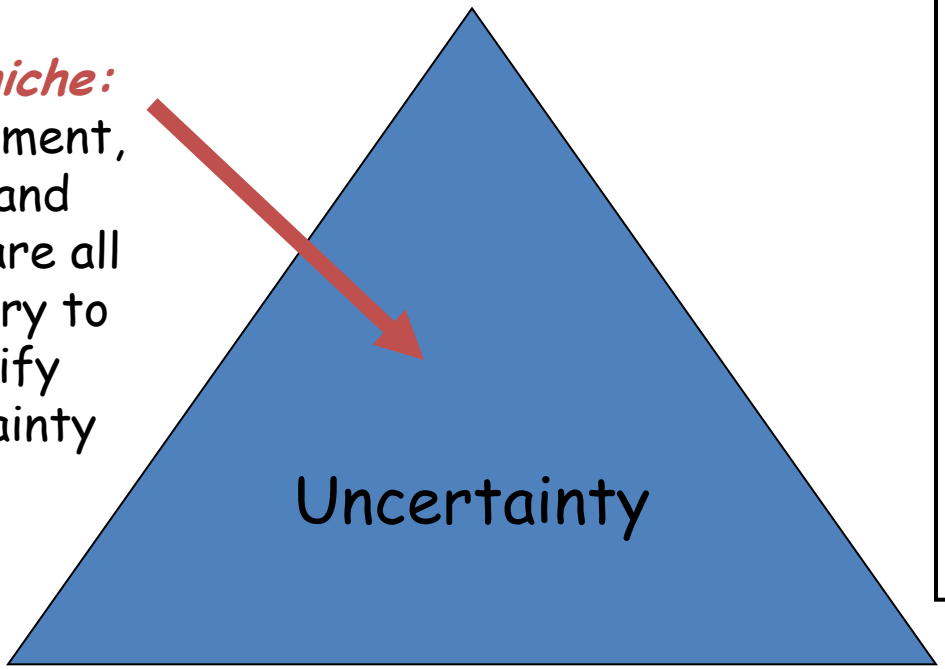
- *Bonded and Non-bonded interactions*
- *Hydrophobic & Hydrophilic interactions*
- *H - bonding*
- *Torsions*
- *Enthalpy & Entropy*

Extensive experimental data are needed to ensure we have the balances correct.

The drug discovery and life science simulation community has observed the same problem.*

*Report on NIST Workshop 2006 "Validating Modeling and Experimental Methods to Enable Drug Discovery"

What is needed to succeed?



NIST niche:
Measurement, data, and models are all necessary to quantify uncertainty

Goal:

- *To develop a synergy between experiment, theory, and modeling to delineate the essential physics required for quantitative predictions and transferability of models to novel systems.*
- *To develop guidelines and standards to achieve quantitative simulations*

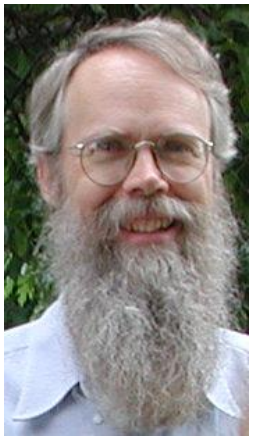
For each model:

Quantum

↓

Macroscopic

Assumptions, advantages, limitations, applicability, cost, data quality required for validation?



Computational Chemistry Comparison and Benchmark Database

"How good is that ab initio *calculation?"*

PURPOSE: Expand the applicability of computational thermochemistry by providing benchmark data for evaluating theoretical methods and assigning uncertainties to computational predictions.

Contents

- Computed, accurate experimental thermochemical quantities for >650 species, >85,000 calculations
- Dipole moments, polarizabilities, transition states, barriers to internal rotation, atomic charges, group additivity methods

Tutorials

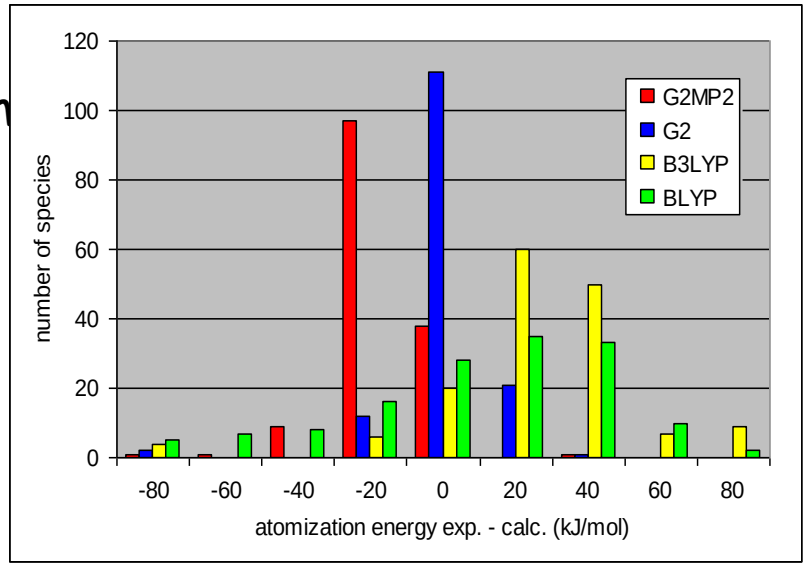
User Interface

Graphical and statistical analysis of theory vs. experiment

Web Access

>25,000 webpages/month ave., 100,000 peak

http://srdata.nist.gov/cccbdb

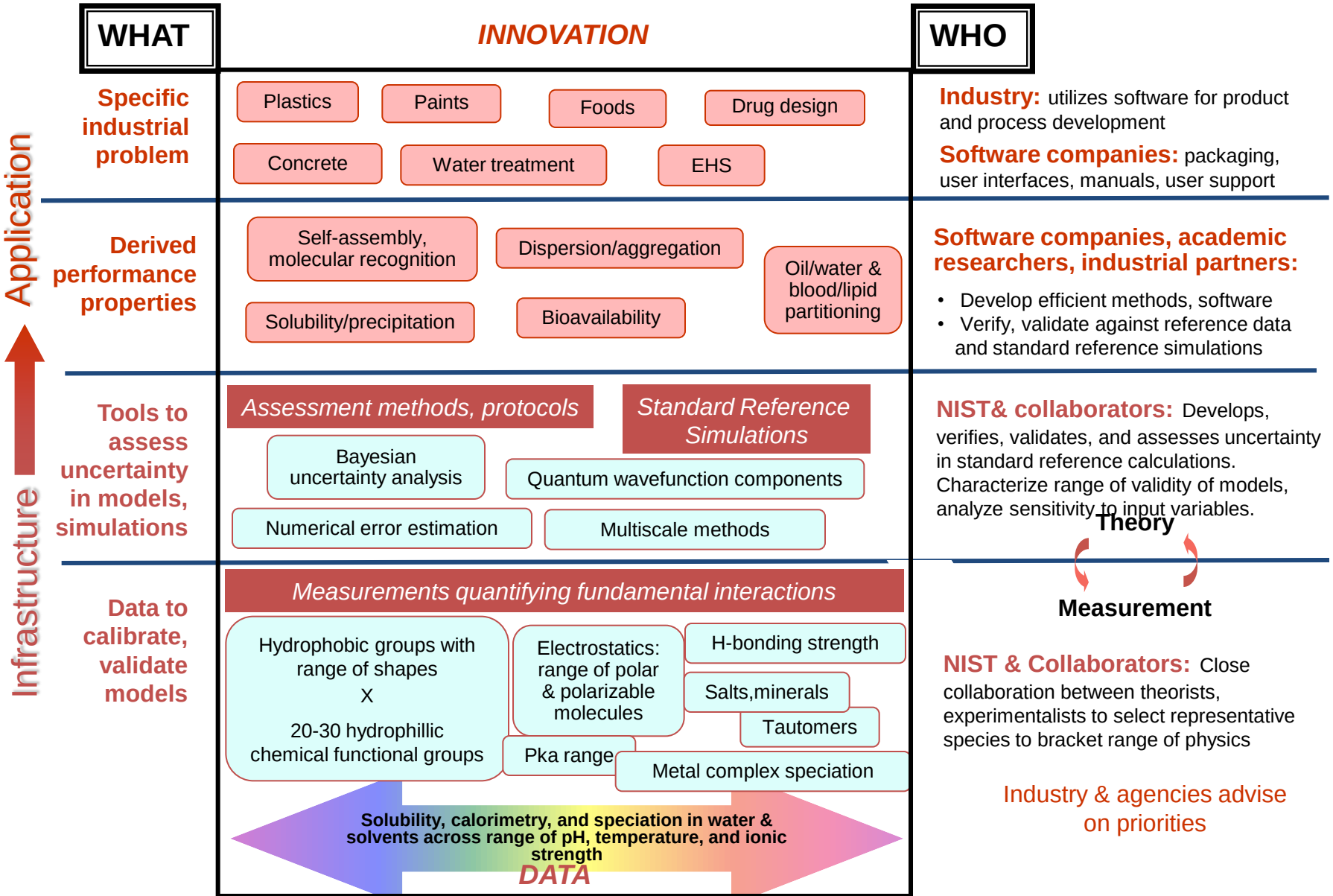


Errors in computed atomization energies at 0 K for 150 carbon-containing species and four theoretical methods.

russell.johnson@nist.gov

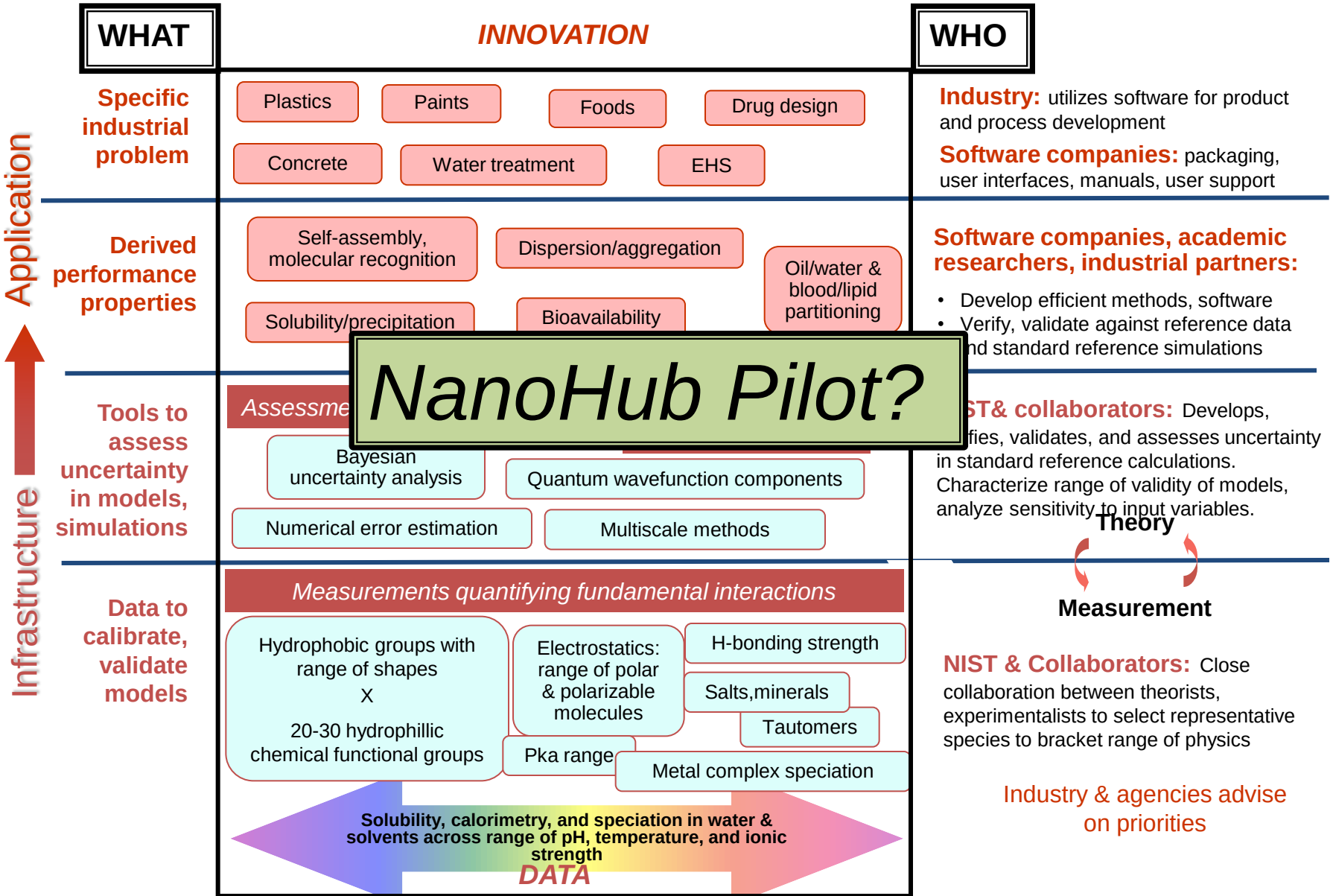
From Data Infrastructure to Innovation –

Quantitative Simulations for Nonbonded Interacitons



From Data Infrastructure to Innovation –

Quantitative Simulations for Nonbonded Interacitons

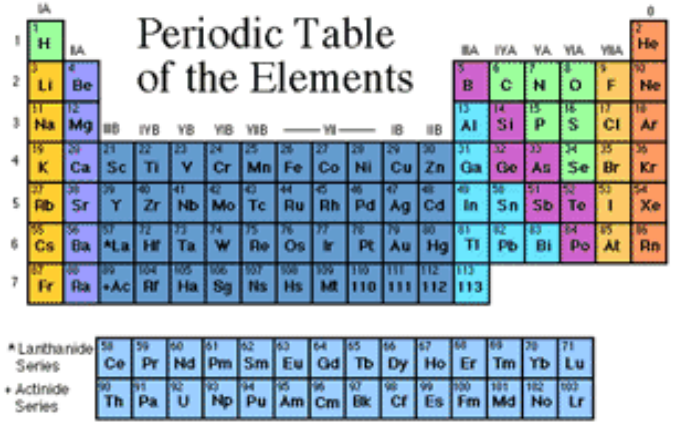


We can't measure everything

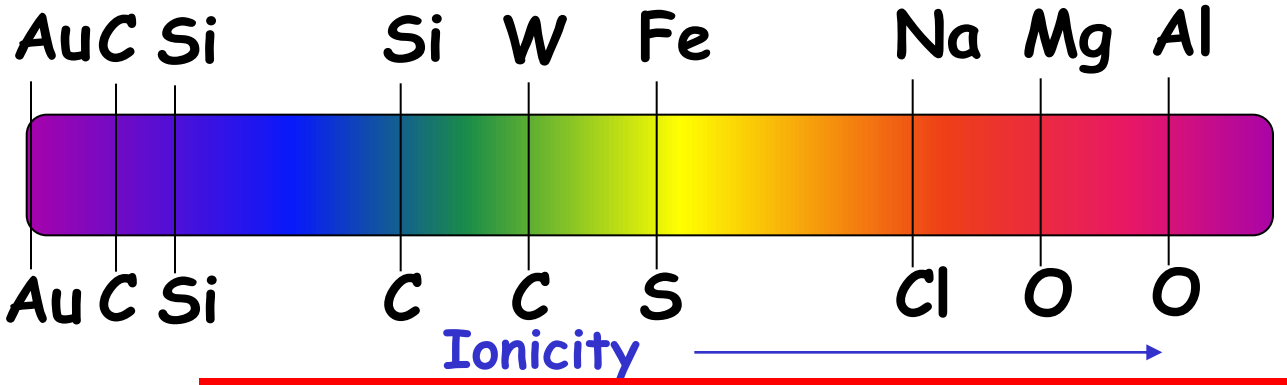
We can't calculate everything either…

Aim:

- ***Develop a synergy between experiment, theory, and modeling to delineate the essential physics and develop a rational strategic approach to design.***
- ***Perform benchmark calculations and experiments***
- ***Determine extent of "neighborhood" in periodic table and composition: Same neighborhood => same model can be used with reasonable certainty to determine properties.***



Example:
AB Charge Transfer (EA/IP) & band gap



What is the effect on brittleness vs ductility across classes of materials?

Standard Reference Simulation: contains physics-based rules for deciding what class of model to be used. For example, if band gap is more than 2 eV or EA/IP ratio is greater than y, then maintain constant charge on atoms during classical simulation. If small band gap, allow charge to redistribute, which requires QM.