Accelerated Search for Materials via Adaptive Learning

Turab Lookman

Los Alamos National Laboratory

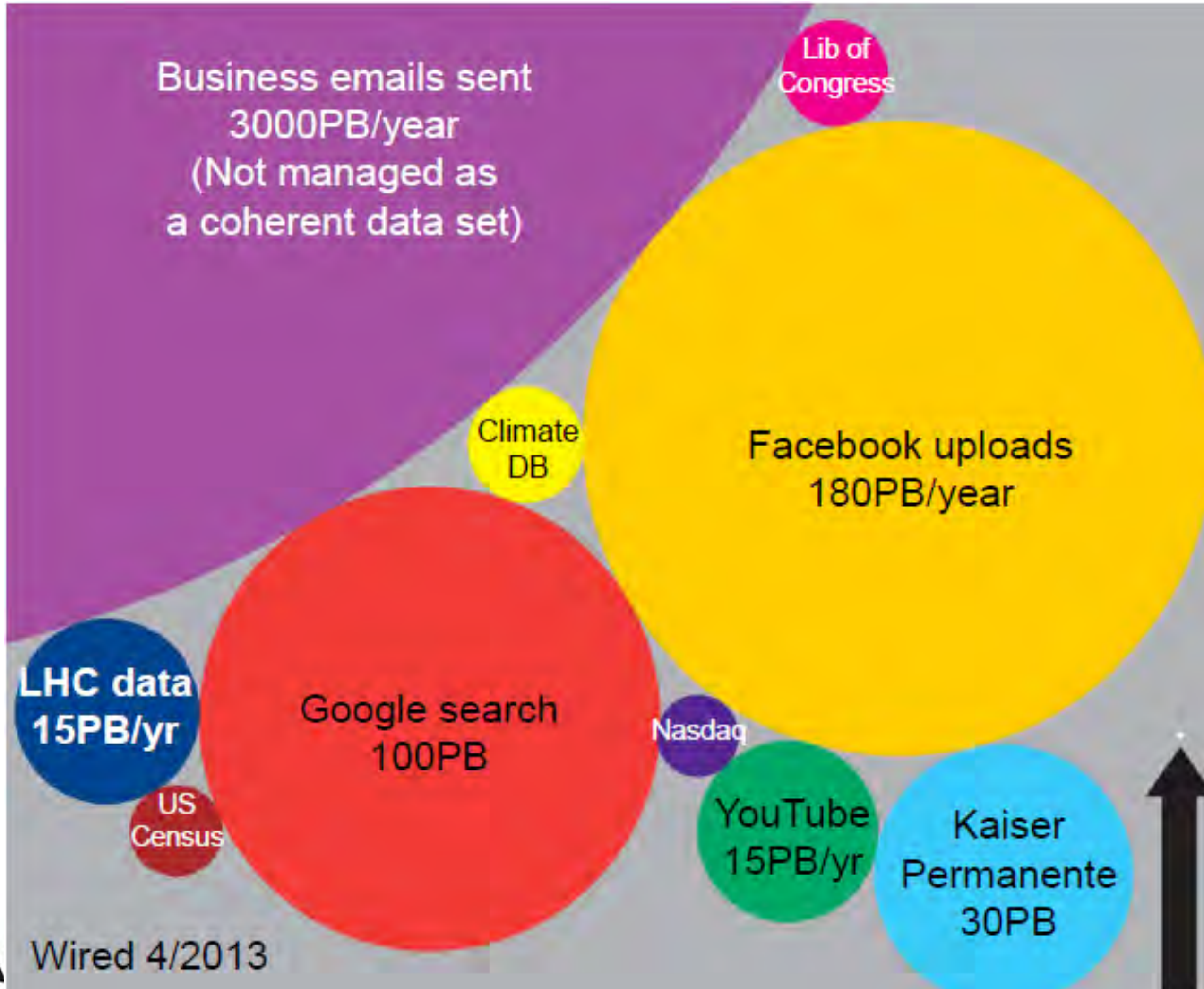
Collaborators: P. Balachandran, Z. Liu, J. Gubernatis, J. Theiler (LANL)
R. Yuan, D. Xue (Xian Jiatong)
E. Dougherty, *Center for Genomic Signalling* (Texas A & M)
I. Karaman

UNCLASSIFIED

OUTLINE

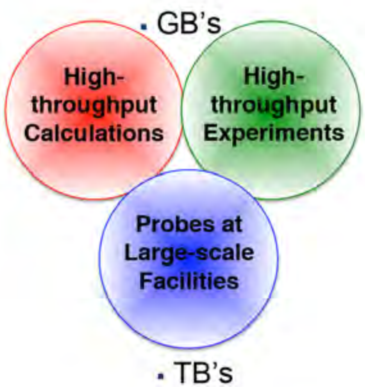
- Introduction: Approach to Learning
- Examples: NiTi alloys
  Pb- free piezoelectrics
- Connections to AM, Challenges

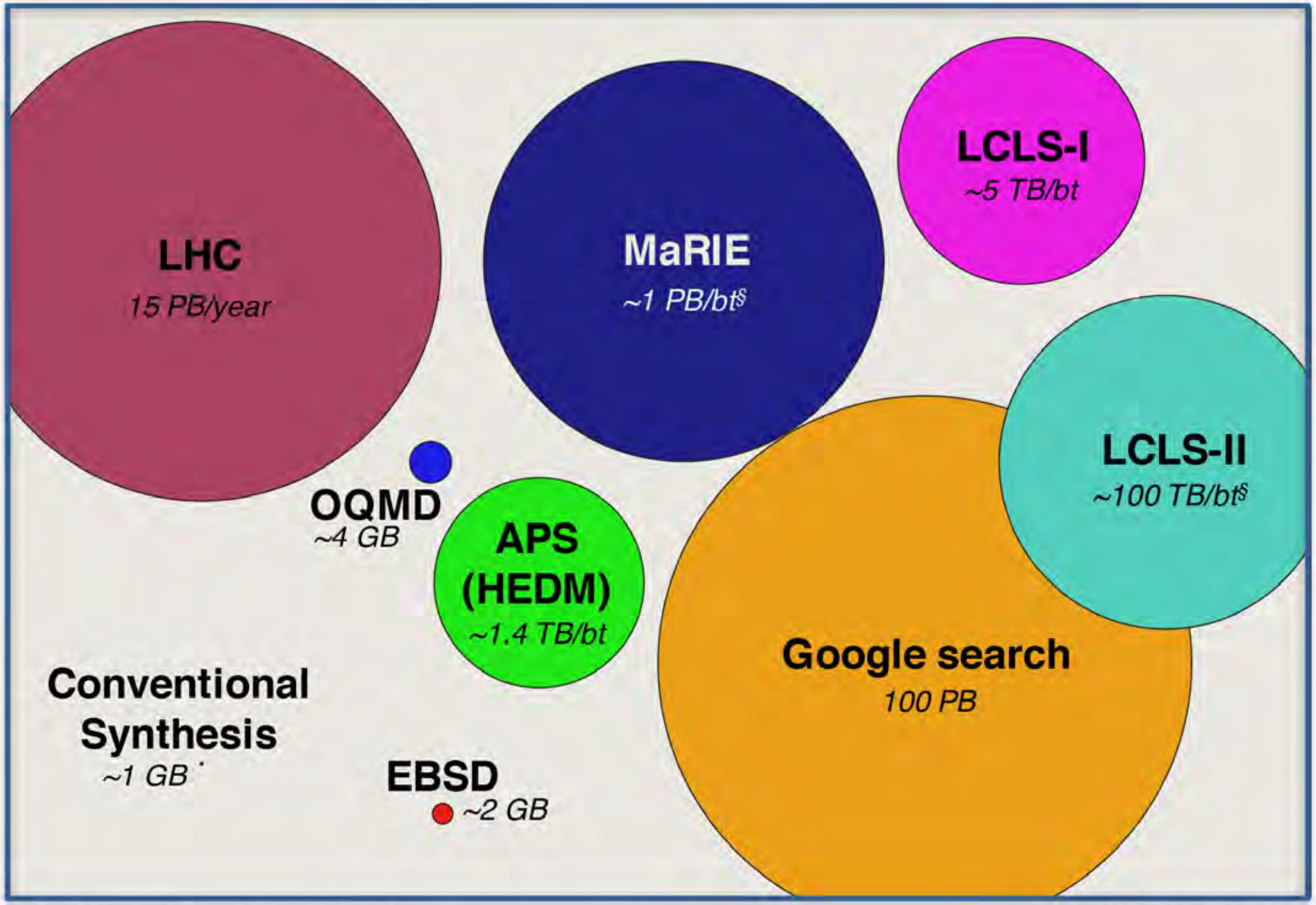
Does Materials Science have a BIG Data Problem ?



Operated by Los Alamos National Security, LLC for NNSA

Materials Data Landscape



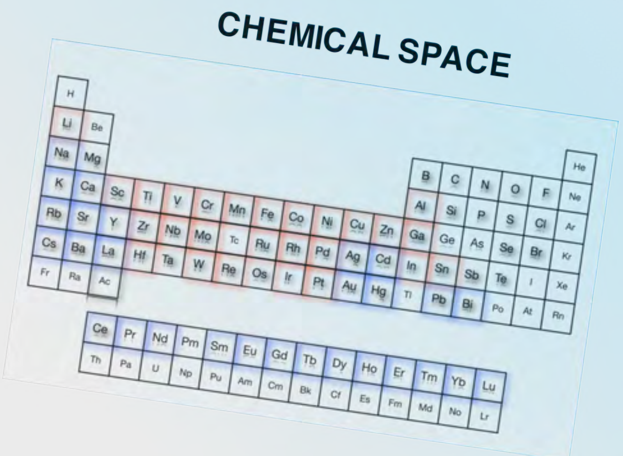


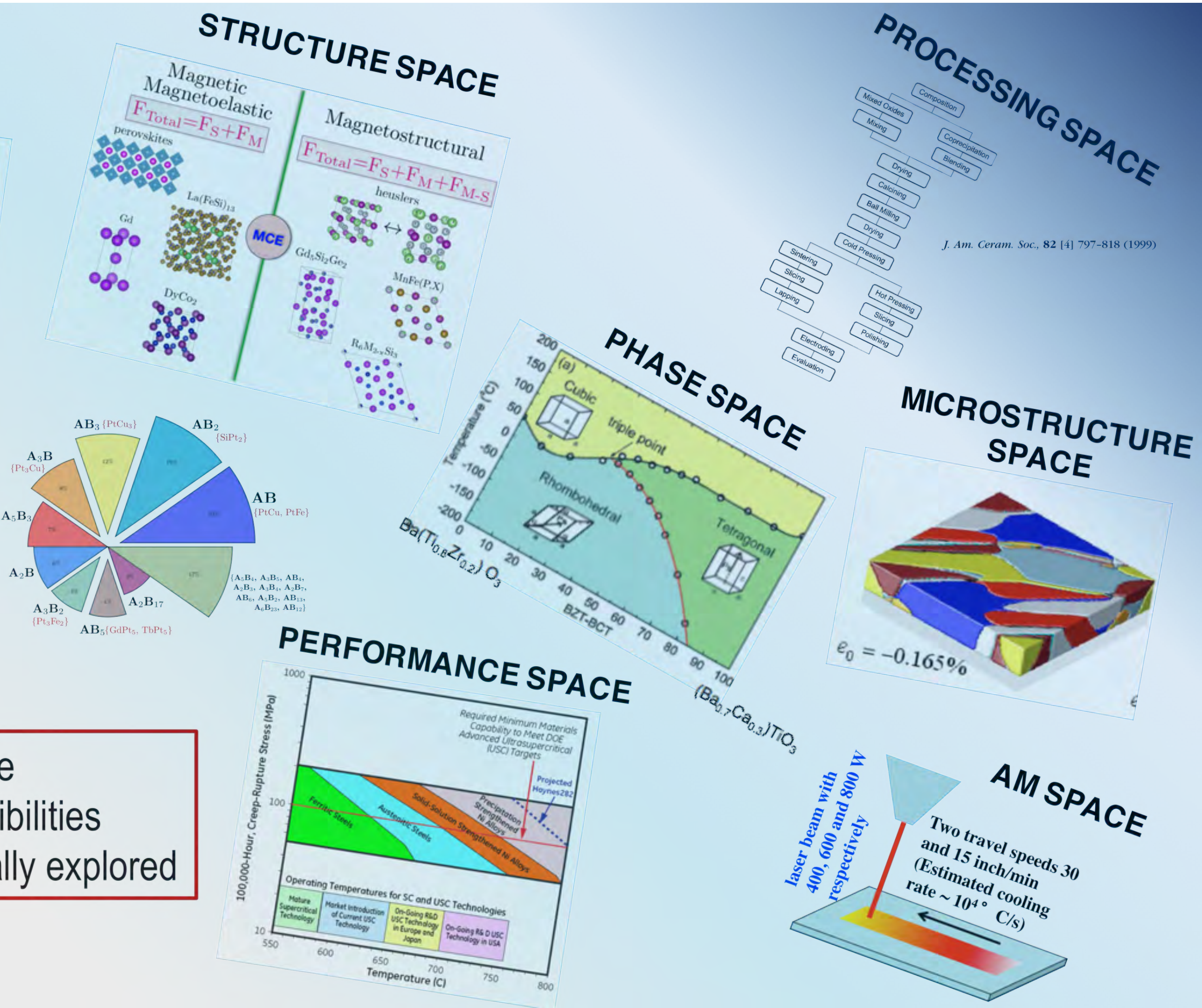
bt – Beam Time (one sample)
§ – Projected

GB – Gigabyte
TB – Terabyte
PB – Petabyte

UNCLASSIFIED

MATERIALS CHALLENGE





- High-dimensional discovery space
- Millions and 100's of Million Possibilities
- Only a small fraction experimentally explored

Learning from Data

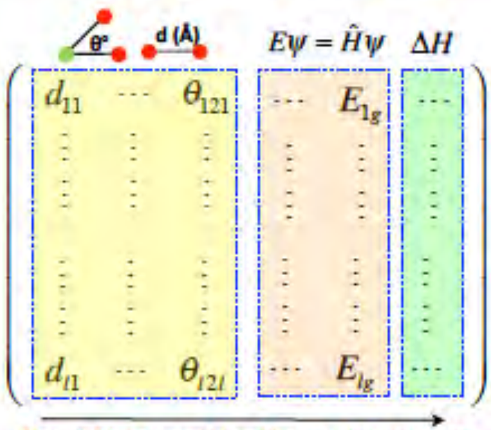
- How do we guide experiments towards finding materials with optimized target properties ?

Current status of Materials Informatics

- Generate lots of high throughput data, screen, make predictions
- Some studies: Use informatics (dimensionality reduction, inference for classification, regression) to make predictions
- Focus: feedback – from expts (small data sets but large search spaces) or calculations
- : use uncertainties in measurements, surrogate model predictions
- : Iteratively refine predictions ------ Adaptive Design

State-of-the-art materials informatics

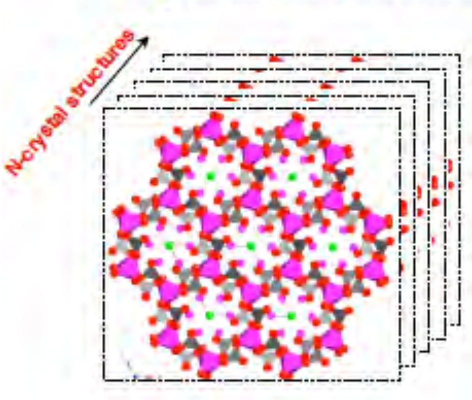
Domain knowledge physics models



Statistical model inference

- Classification
  Regression
  y=f(x)
  - *Dimensionality Reduction*
  - *Decision Tree*
  - *Clustering*
  - *Partial Least Squares*

Database

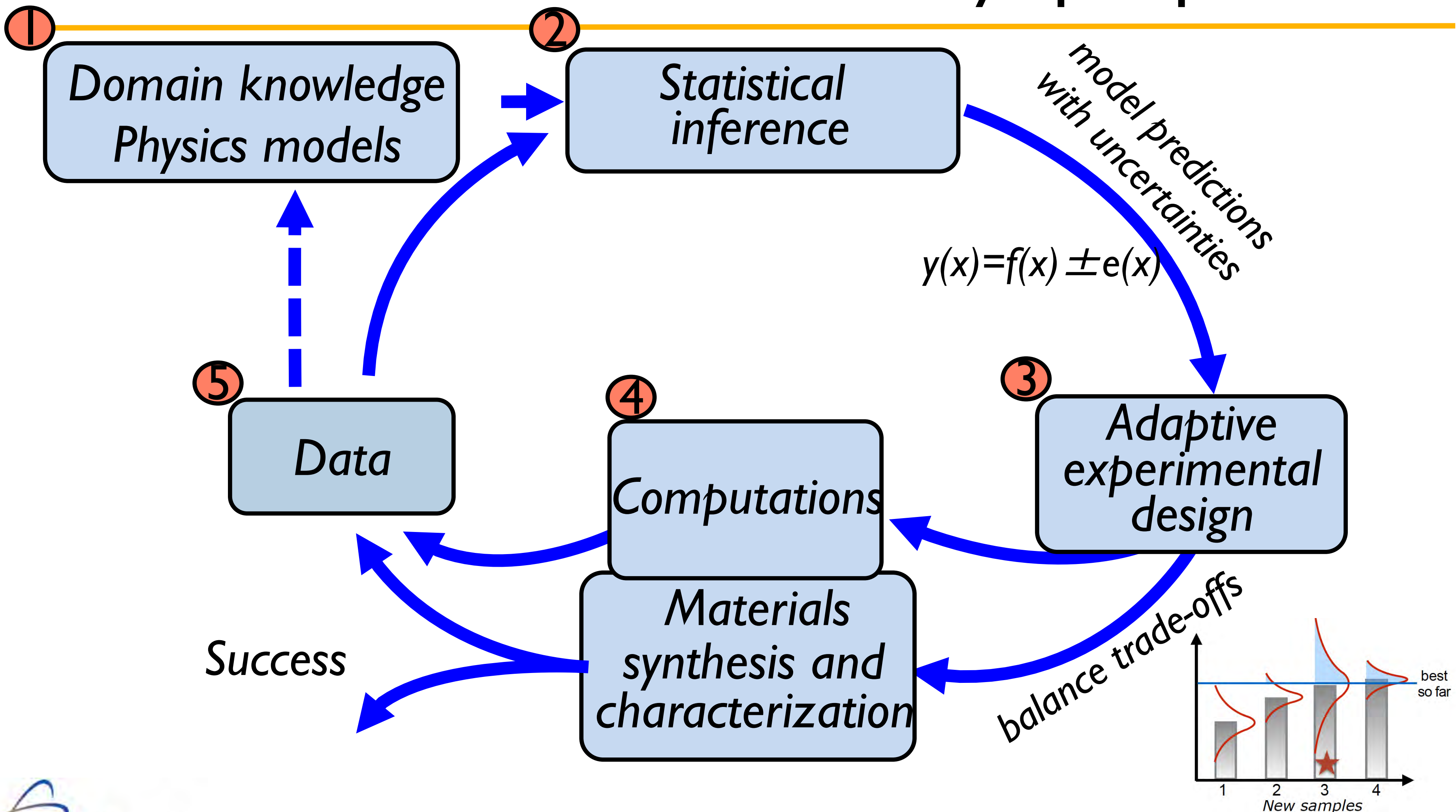


First principles calculations

Materials synthesis and characterization

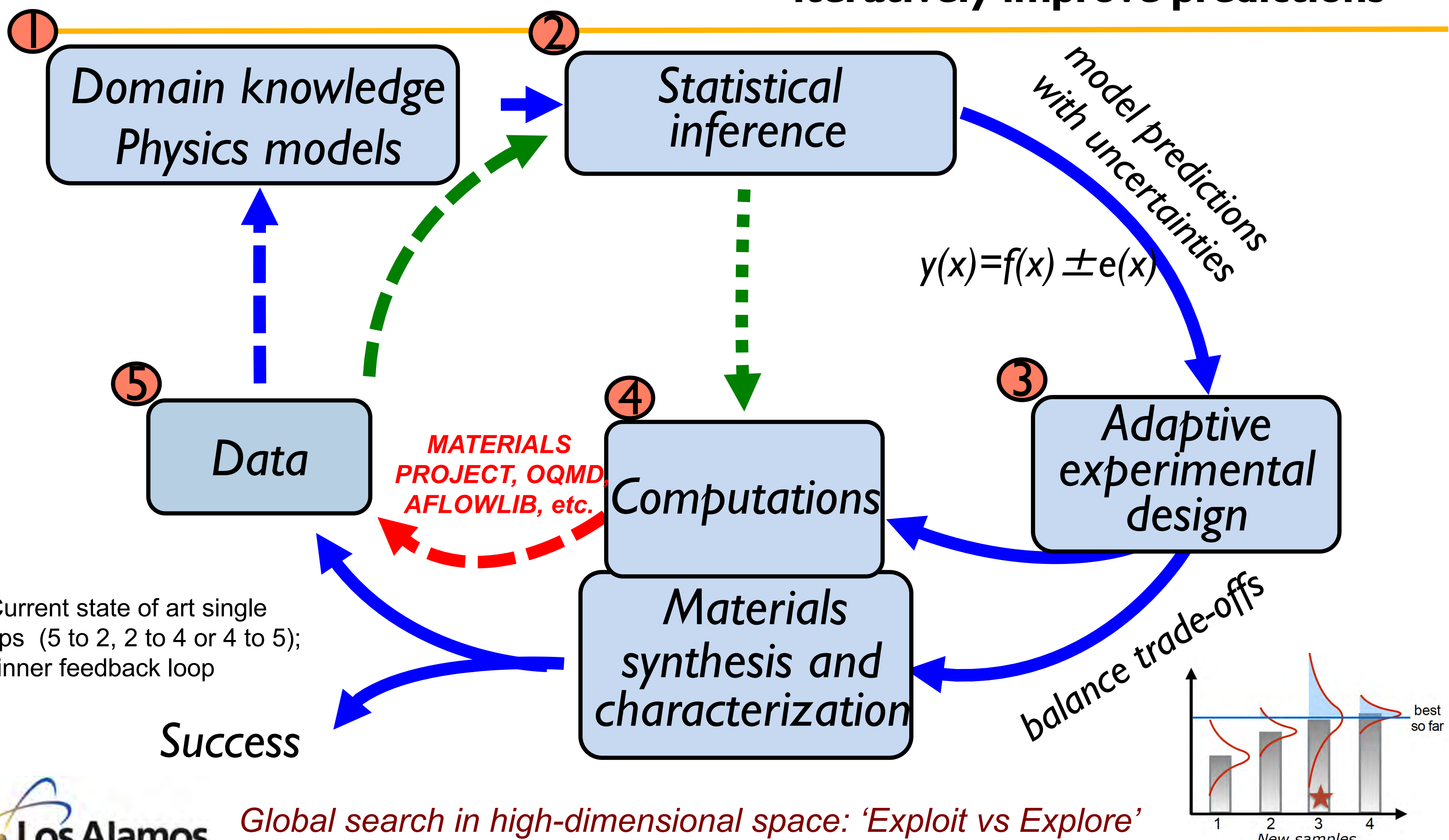
Codesign for materials design and discovery

- **iteratively improve predictions**



Codesign for materials design and discovery

- **iteratively improve predictions**



- Current state of art single steps (5 to 2, 2 to 4 or 4 to 5); No inner feedback loop

Global search in high-dimensional space: 'Exploit vs Explore'

Optimal Learning for guiding large scale computational tasks OR experiments

Preliminary Experiments
Variables
Initial Designs

Response surfaces

Surrogate model
Bayesian

$\hat{y}$

New designs
Infill points

x

y

Function Evaluation

High fidelity, Expensive Computer Calculation

Measurement Policies
Online: bandit problems
Offline :ranking and selection
: stochastic search

Marginal value of Information (Howard)
Sequential policies: Knowledge gradient (Powell)

Bayesian Optimization
P[I] Kushner'64
E[I] Mockus'78
$E[I^g]$ Jones'88

Probability Collectives (Wolpert'98)

Mean Objective Cost of Uncertainty (MOCU) (Dougherty'2013)

Constraints, Multiobjective, Multifidelity Parallel function evaluations

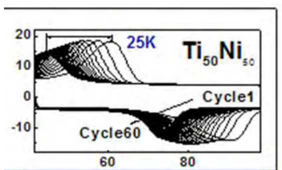
T.L. *et al, Curr Opin in Sol Stat and Mat Sci.'* 2016

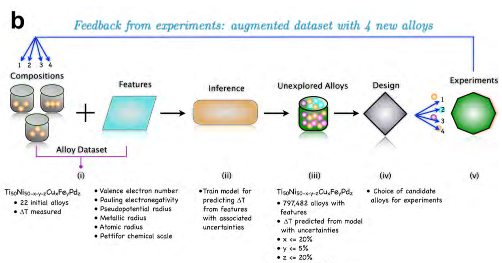
Examples of Adaptive Design for New Materials

Data-driven ⟷ *Experiments*

Use theory with data ⟷ *Experiments*

Strategy: $Ti_{50}\ Ni_{50-x-y-z}\ Cu_x\ Pd_y\ Fe_z$
Total possibilities: 800,000 (22 known)
Target: Minimize Thermal Hysteresis

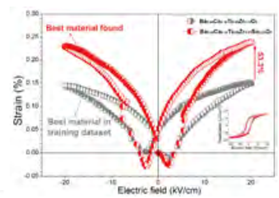


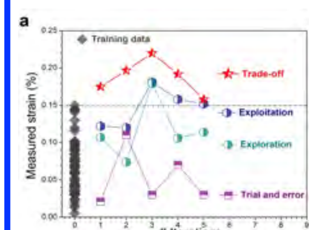


Guided 36 experiments
Found 14 with superior properties

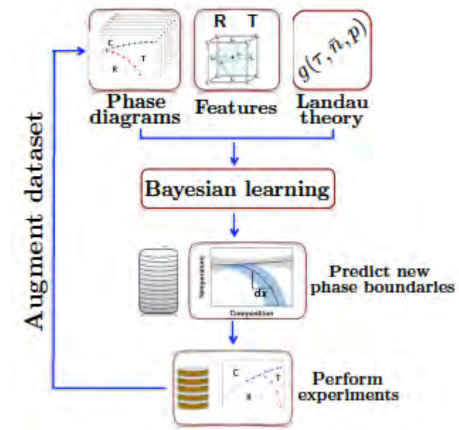
$(Ba_{100-x-y}Ca_xSr_y)(Ti_{100-u-v}Zr_uSn_v)O_3$

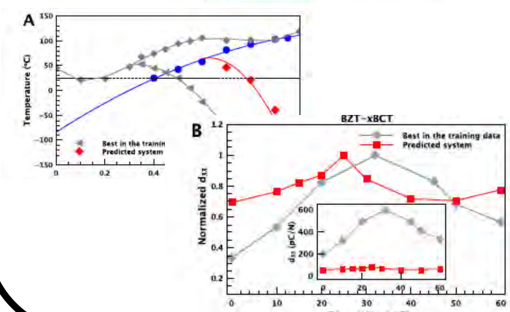
- Largest electrostrain





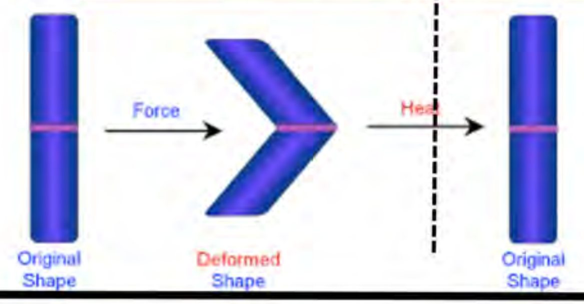
Validation of design

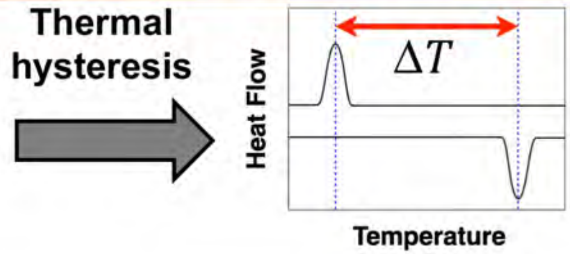


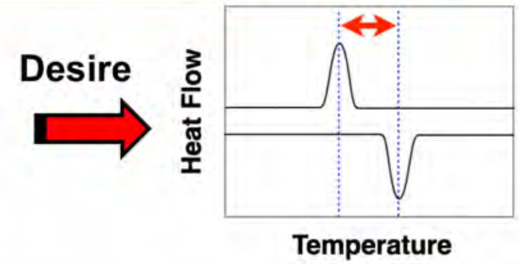


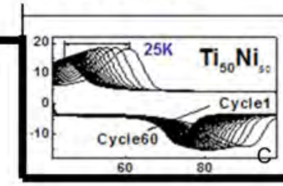
Beyond the capabilities of the state-of-the-art high-throughput DFT calculations

Example: Find NiTi alloy with lowest hysteresis









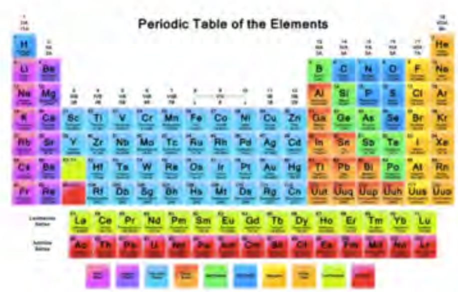
Strategy: $Ti_{50}\ Ni_{50\text{-}x\text{-}y\text{-}z}\ Cu_x\ Pd_y\ Fe_z$

Search space of multicomponent alloy ~800K
22 training samples (.003%) with measured ΔT

$$x \leq 20\%, y \leq 5\%, z \leq 20\%$$

$$50 - x - y - z \geq 30\%$$

Composition control 0.1%



Knowledge:

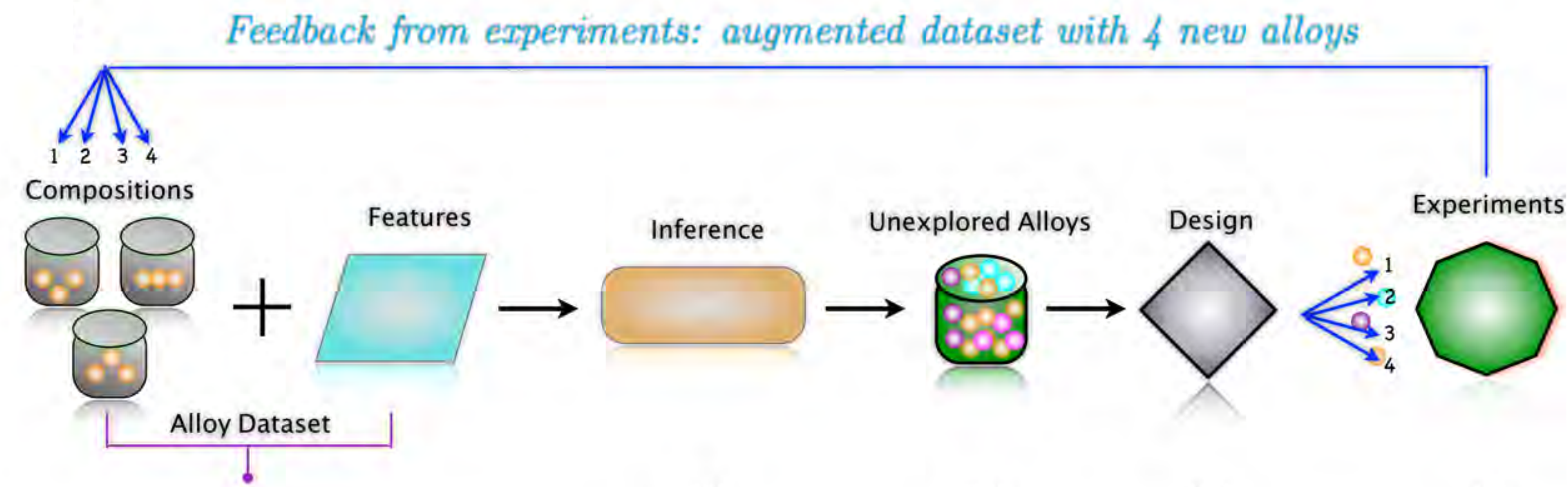
- Transition temperatures influenced by valence electron concentration
- Hysteresis influenced by atomic size
- Relative stability influenced by changes in electron number

Features:

Valence electron number
Radii: metallic, Clementi, Zunger
Electronegativity
Pettifor Chemical scale

Each $Ti_{50}\ Ni_{50\text{-}x\text{-}y\text{-}z}\ Cu_x\ Pd_y\ Fe_z$ weighted fraction of features

Adaptive Design for Alloy Discovery



(i)

$Ti_{50}Ni_{50-x-y-z}Cu_xFe_yPd_z$
- 22 initial alloys
- ΔT measured

- Valence electron number
- Pauling electronegativity
- Pseudopotential radius
- Metallic radius
- Atomic radius
- Pettifor chemical scale

(ii)

- Train model for predicting ΔT from features with associated uncertainties

(iii)

$Ti_{50}Ni_{50-x-y-z}Cu_xFe_yPd_z$
- 797,482 alloys with features
- ΔT predicted from model with uncertainties
- x <= 20%
- y <= 5%
- z <= 20%
- 50-x-y-z >= 30%

(iv)

- Choice of candidate alloys for experiments

(v)

nature COMMUNICATIONS

ARTICLE

Received 2 Sep 2015 | Accepted 4 Mar 2016 | Published 15 Apr 2016 | OPEN

Accelerated search for materials with targeted properties by adaptive design

Dezhen Xue[1,2], Prasanna V. Balachandran[1], John Hogden[3], James Theiler[4], Deqing Xue[2] & Turab Lookman[1]

Results from Design Loop on NiTi Alloy

NiTi-based alloy found with smallest dissipation

(1.84K)

$Ti_{50}Ni_{46.7}Cu_{0.8}Pd_{0.2}Fe_{2.3}$ (42% improvement)

TABLE I | Five new best alloys, which were found in iterations 6 and 7, amongst 14 with the lowest ΔT. From a total of 9 iterations, which resulted in 36 new alloys, 14 had a ΔT smaller than 3.15 K, the lowest in the original training set of 22. Transformation temperature is given by the endothermic peak in the DSC curve.

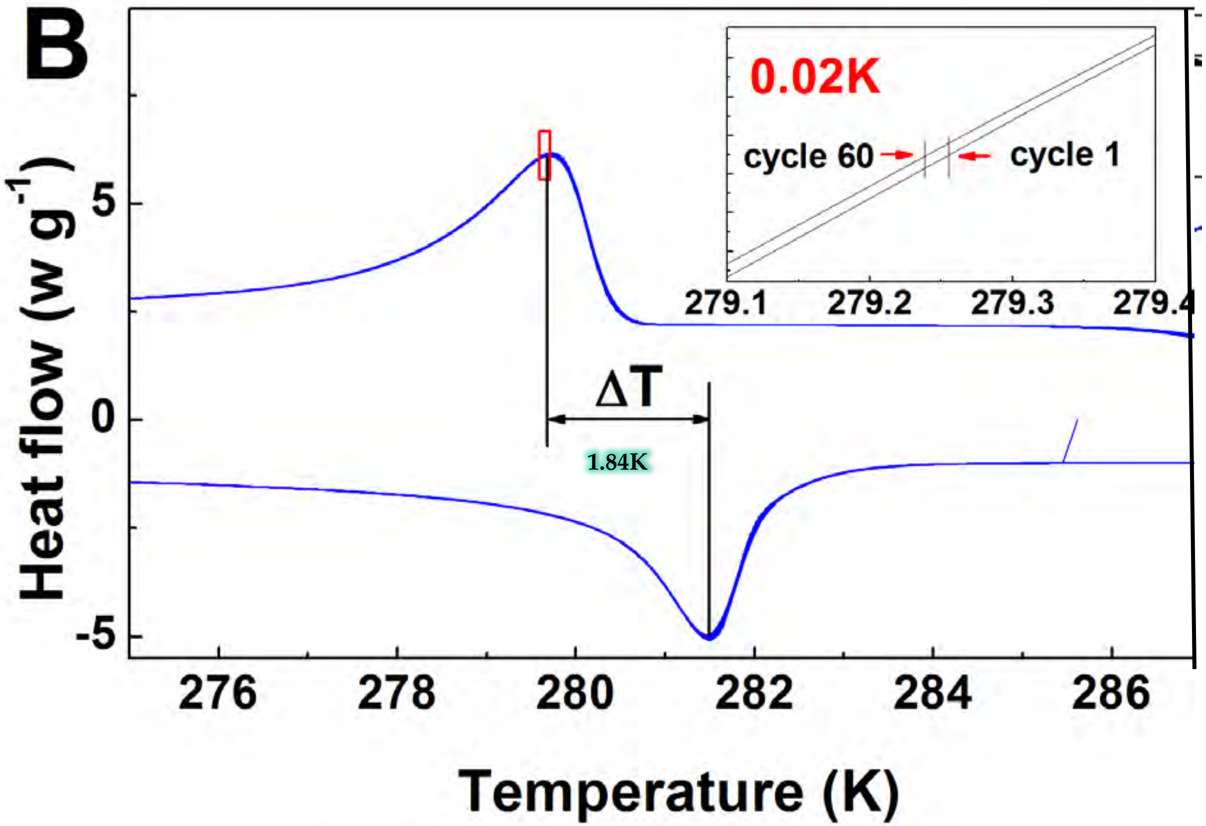
Iterations	Composition	ΔT (K)	Transformation temperature (K)
6	$Ti_{50.0}Ni_{46.8}Cu_{0.9}Fe_{2.0}Pd_{0.3}$	2.64	289.95
6	$Ti_{50.0}Ni_{44.2}Cu_{1.9}Fe_{3.8}Pd_{0.1}$	2.53	243.43
6	$Ti_{50.0}Ni_{46.7}Cu_{0.8}Fe_{2.3}Pd_{0.2}$	1.84	281.77
7	$Ti_{50.0}Ni_{48.1}Cu_{0.2}Fe_{1.5}Pd_{0.2}$	2.09	301.86
7	$Ti_{50.0}Ni_{46.5}Cu_{1.1}Fe_{2.2}Pd_{0.2}$	2.32	283.79

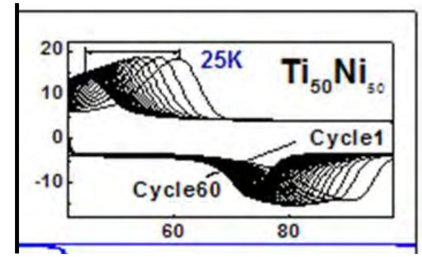
22 samples
3.15K – best in training set

9 loops, 36 predicted and synthesized

14 better than 3.15K, p value < .001

Material Performance for Synthesized Alloy

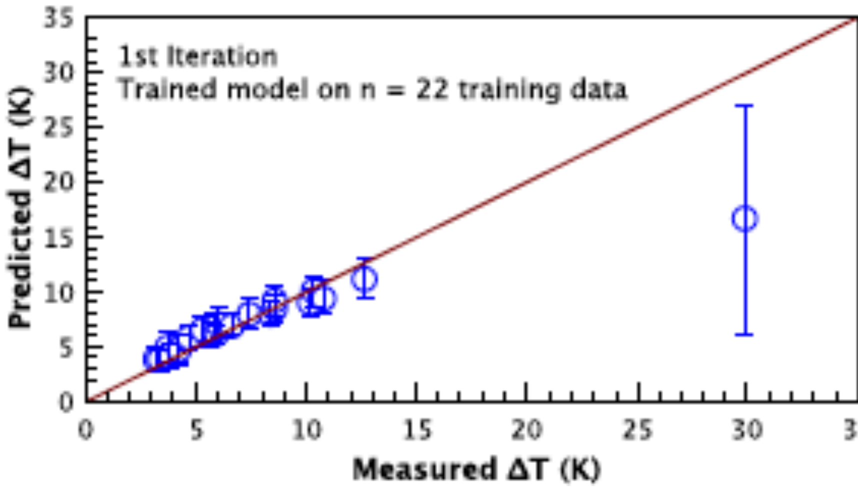


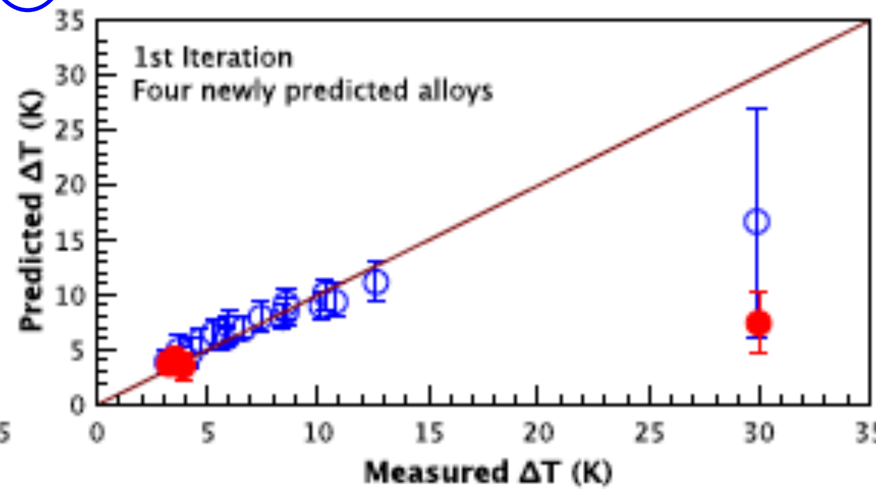


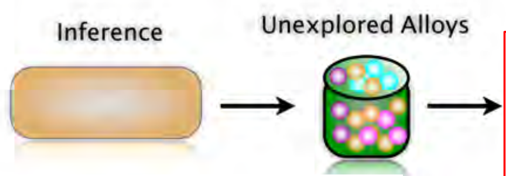
$Ti_{50}Ni_{46.7}Cu_{0.8}Pd_{0.2}Fe_{2.3}$

Inference not adequate, need to explore

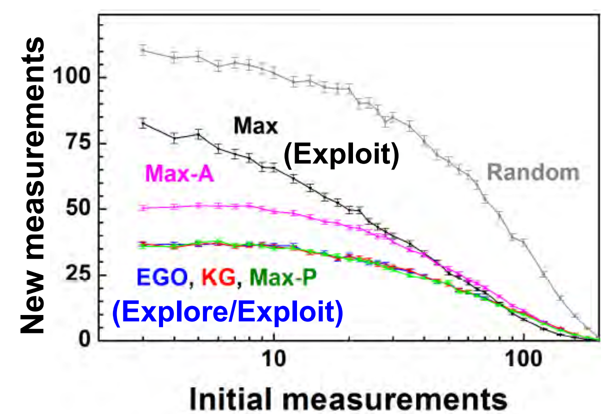
$$\Delta T \longrightarrow y(x) = f(x) \pm e(x)$$





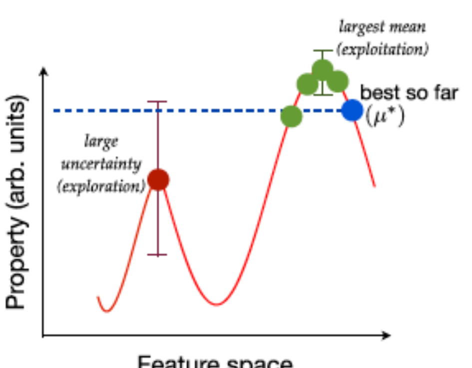


Exploitation is suboptimal



Largest modulus of M_2AX phases

223 data; features: r_s, r_p, r_d



Need design: Explore space using uncertainties

SCIENTIFIC REPORTS

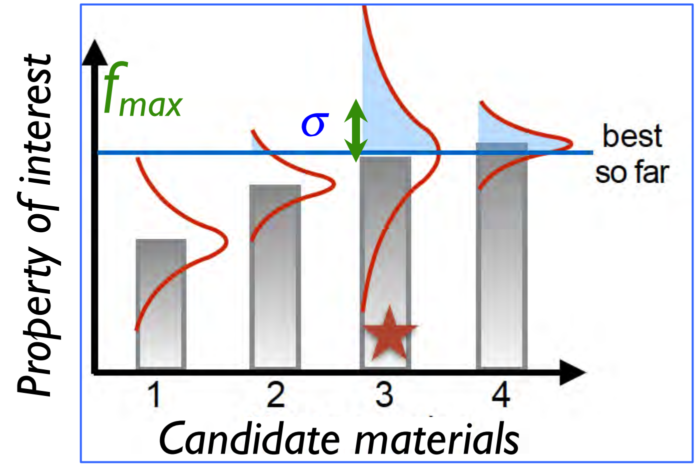
OPEN Adaptive Strategies for Materials Design using Uncertainties

Prasanna V. Balachandran, Dezhen Xue, James Theiler, John Hogden & Turab Lookman

Experimental Design

Exploitation
(local, utilize model)

vs. **Exploration**
(global, improve model)



Strategy 1 *(Exploit):*

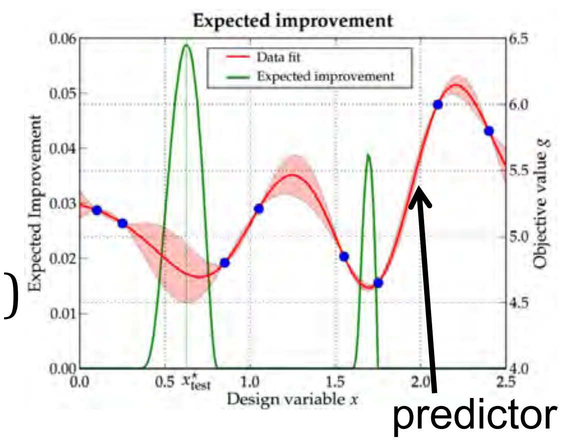
Next experiment x_i optimizes $y(x_i)$ by exploitation

Strategy 2 *(Explore):*

Choose next experiment via uncertainties

- Gaussian Process Model $y = f(x) + N(0,\sigma^2)$

$$\mathrm{cov}(f(x), f(x')) = e^{(-\theta|x-x'|^2)} + \delta(x,x')\sigma_n$$



Examples of Adaptive Design for New Materials

Data-driven ⟷ *Experiments*

Use theory with data

↕

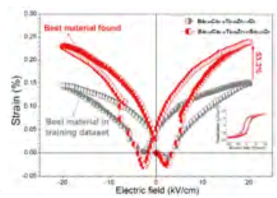
Experiments

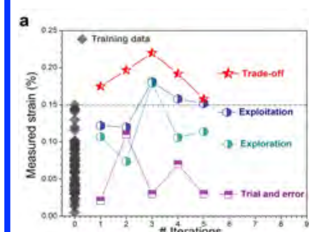
Strategy: $Ti_{50}\,Ni_{50-x-y-z}\,Cu_x\,Pd_y\,Fe_z$
Total possibilities: 800,000 (22 known)
Target: Minimize Thermal Hysteresis

Guided 36 experiments
Found 14 with superior properties

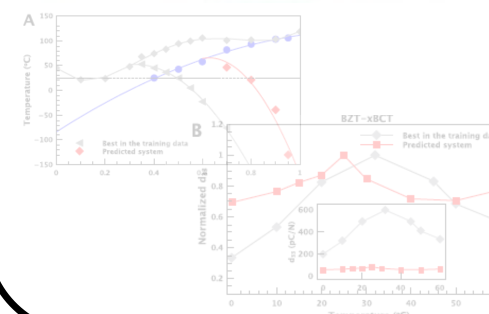
$(Ba_{100-x-y}Ca_xSr_y)(Ti_{100-u-v}Zr_uSn_v)O_3$

- Largest electrostrain



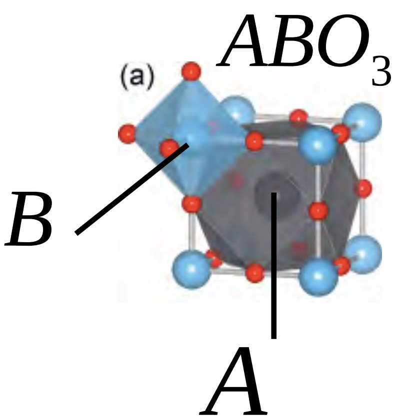


Validation of design



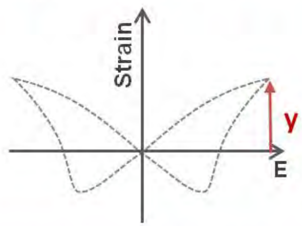
Search for BaTiO3 based solid solutions with large electrostrains

$$(Ba_{100-x-y}Ca_xSr_y)(Ti_{100-u-v}Zr_uSn_v)O_3 \begin{cases} 100-x-y \geq 60\% \; x \leq 40\% \; y \leq 30\% \\ 100-u-v \geq 60\% \; u \leq 30\% \; v \leq 30\% \end{cases}$$



Experimental Comparison of design strategies : Search for BaTiO3-based large electrostrains

1 *Input dataset*

Material features, x	Targeted property, y
• Electronegativity • Ionic radius • Volume • Ionic displacements • Polarization • Dopant effects on transitions	

2 *Statistical Inference and machine learning*

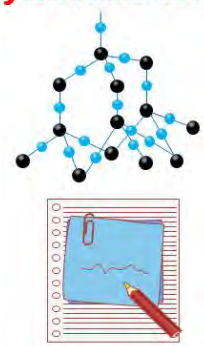
y = f(x) + uncertainties

3 *~ 605000 possible compounds*

4 *Design and optimization: choice of candidates for next experiment*

- Exploitation
- Exploration
- Trade-off between exploitation &exploration
- Random Selection

5 *Synthesis and characterization*

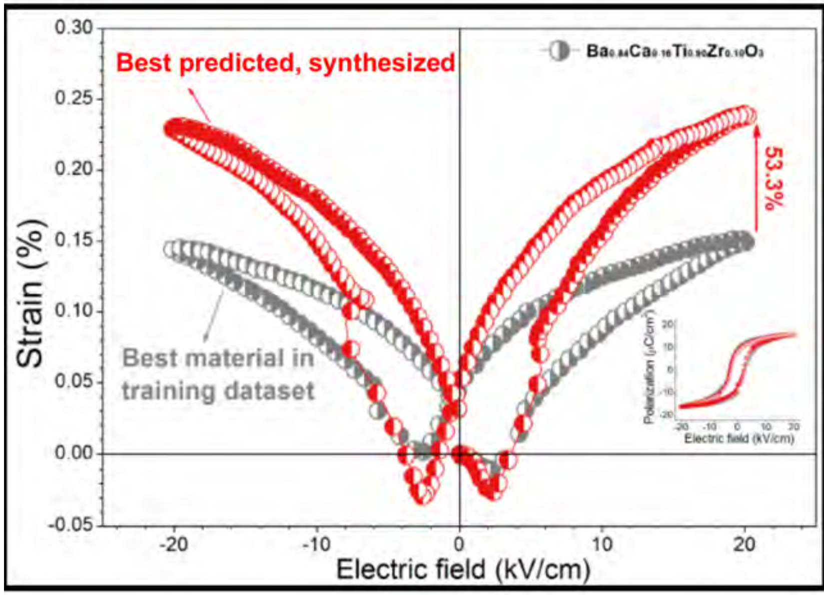


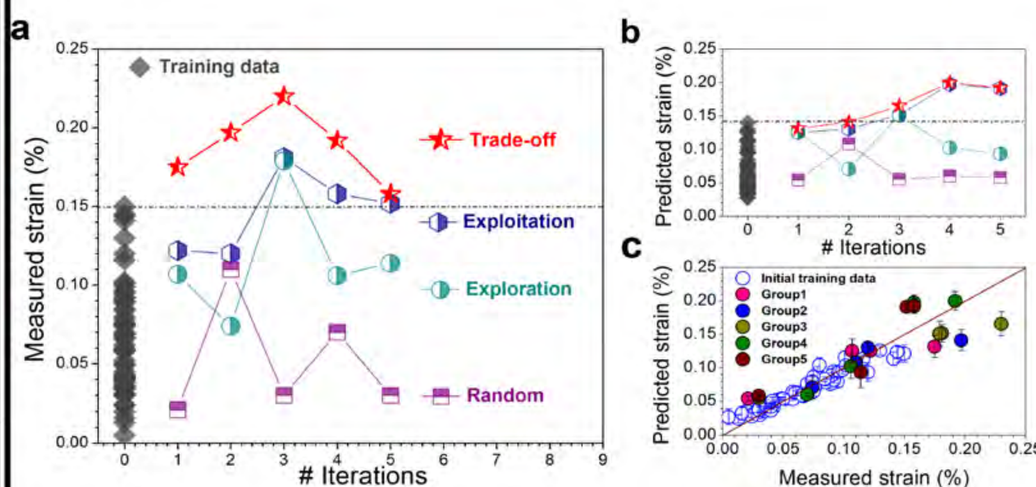
Four recommended samples for synthesis and characterization

6 *Feedback: Augmented dataset*

Success?

BCT-BZT based piezoelectric with largest electrostrain





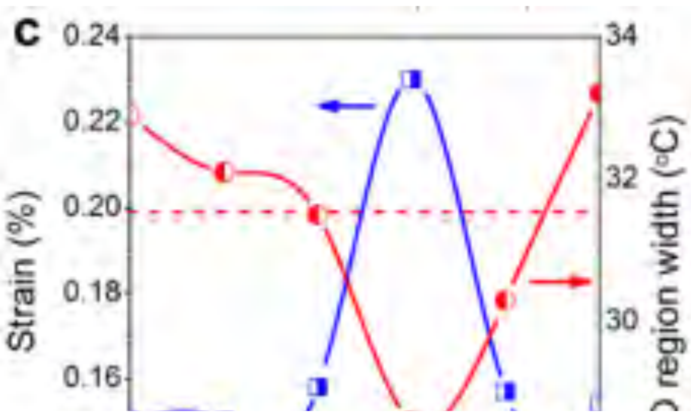
$$F = \alpha_1 \sum_i P_i^2 + \alpha_{11} \sum_i P_i^4 + \alpha_{12} \sum_{i>j} P_i^2 P_j^2 + \alpha_{111} \sum_i P_i^6 + \alpha_{112} \sum_{i>j} \left(P_i^4 P_j^2 + P_j^4 P_i^2 \right) + \alpha_{123} \prod_i P_i^2$$

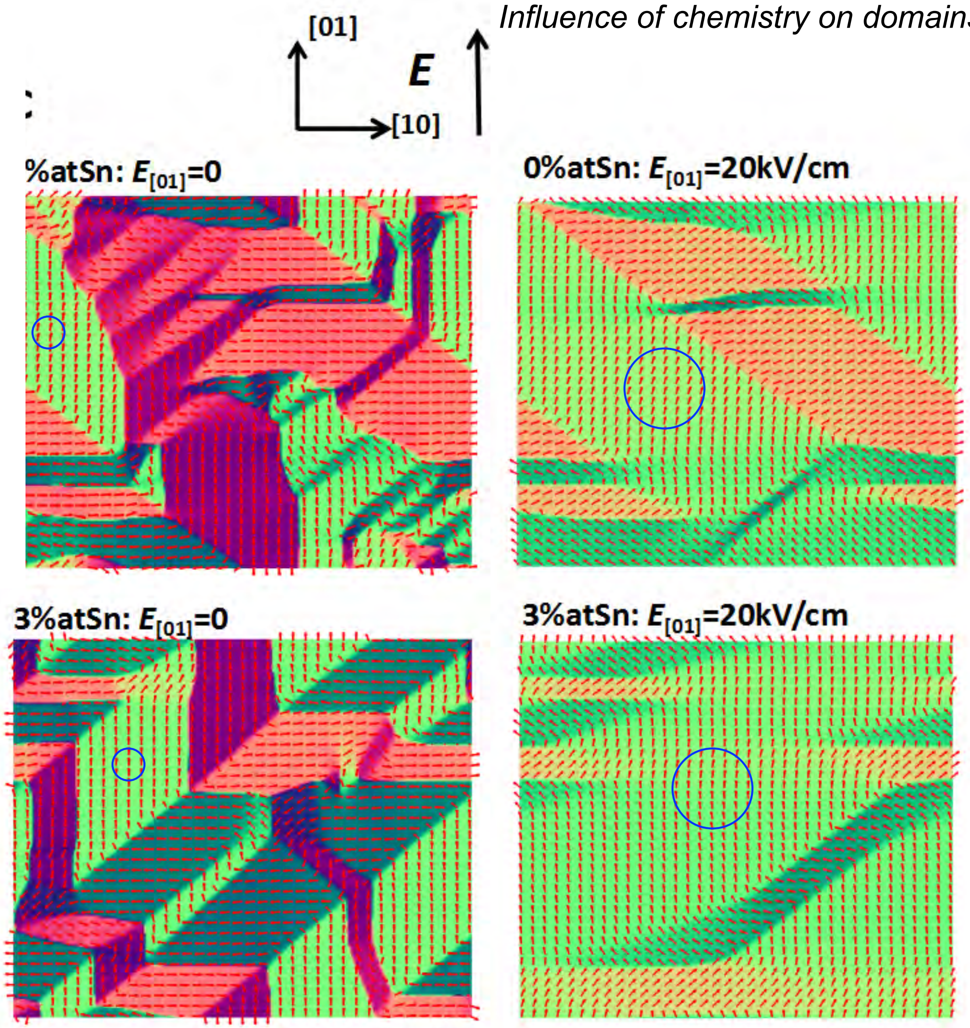
→ Switching of polarization domains

Yuan et al., 2017

Why Optimal ?: Ease of Domain Switching

Influence of chemistry on domains





Examples of Adaptive Design for New Materials

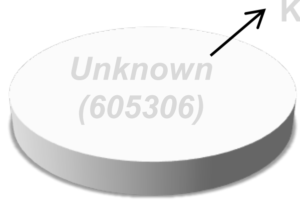
Data-driven ⟷ *Experiments*

Use theory with data ⟷ *Experiments*

Strategy: $Ti_{50}\,Ni_{50-x-y-z}\,Cu_x\,Pd_y\,Fe_z$
Total possibilities: 800,000 (22 known)
Target: Minimize Thermal Hysteresis

Guided 36 experiments
Found 14 with superior properties

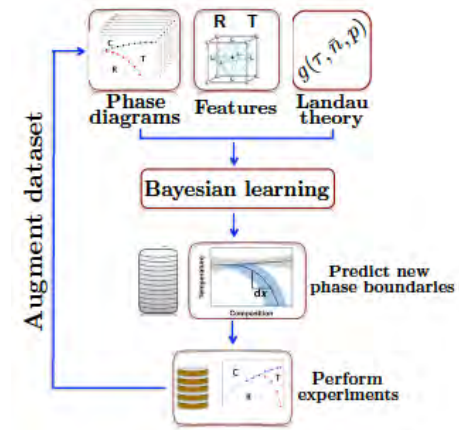
$(Ba_{100-x-y}Ca_xSr_y)(Ti_{100-u-v}Zr_uSn_v)O_3$

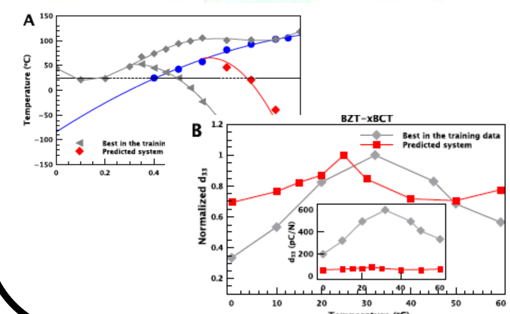


Known (61)

- Largest electrostrain

Validation of design



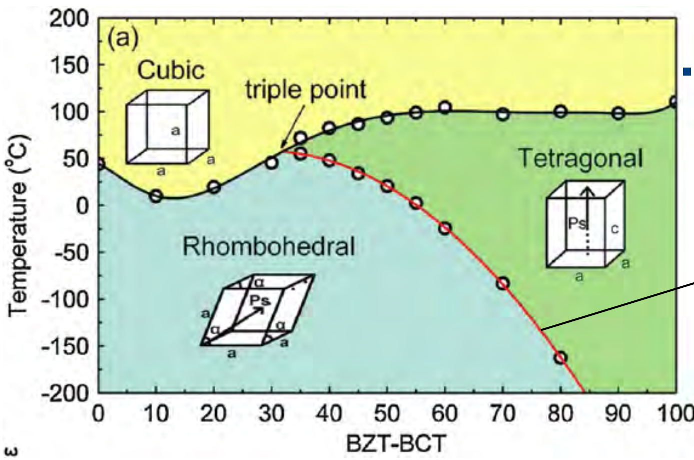


Example: Importance of knowledge

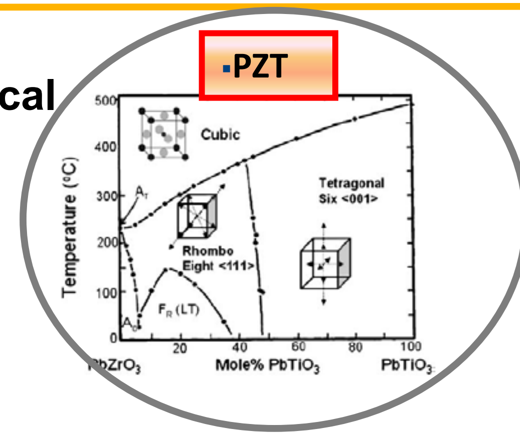
Accelerated search for $BaTiO_3$-based piezoelectrics with vertical morphotropic phase boundary using Bayesian learning

Dezhen Xue[a,b], Prasanna V. Balachandran[a], Ruihao Yuan[b], Tao Hu[c], Xiaoning Qian[c], Edward R. Dougherty[c], and Turab Lookman[a,1]

[a]Theoretical Division, Los Alamos National Laboratory, Los Alamos, NM 87545; [b]State Key Laboratory for Mechanical Behavior of Materials, Xi'an Jiaotong University, Xi'an 710049, China; and [c]Department of Electrical and Computer Engineering, Texas A&M University, College Station, TX 77843

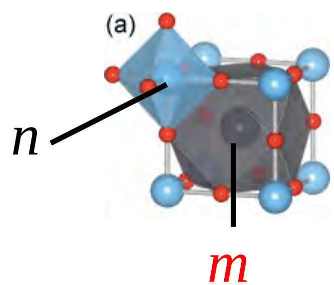


- Design criterion: Vertical

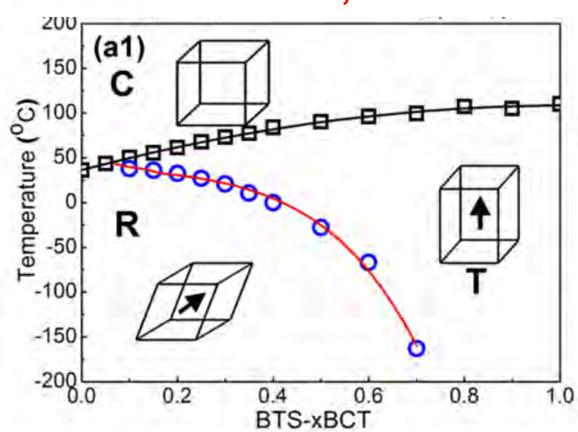


- Mix two ends:

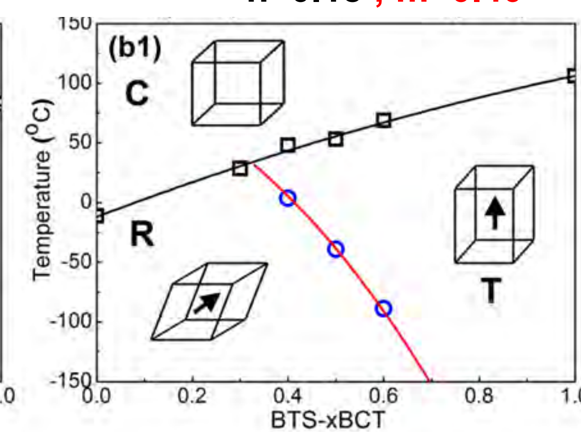
A_n - B_m



- n=0.12 ; m=0.30



- n=0.18 ; m=0.40



- Q: What combination of m,n and chemistries (Al, Li, …) will optimize phase boundaries, response?

Learning from theory + data

- $x(Ba_{1-m}Ca_m)TiO_3 - Ba(Zr_nTi_{1-n})O_3$
  - 18%< m < 50%; 15% < n < 30%

(1200 phase diagrams)

- Features:
- Order parameters: Polarization, Strain

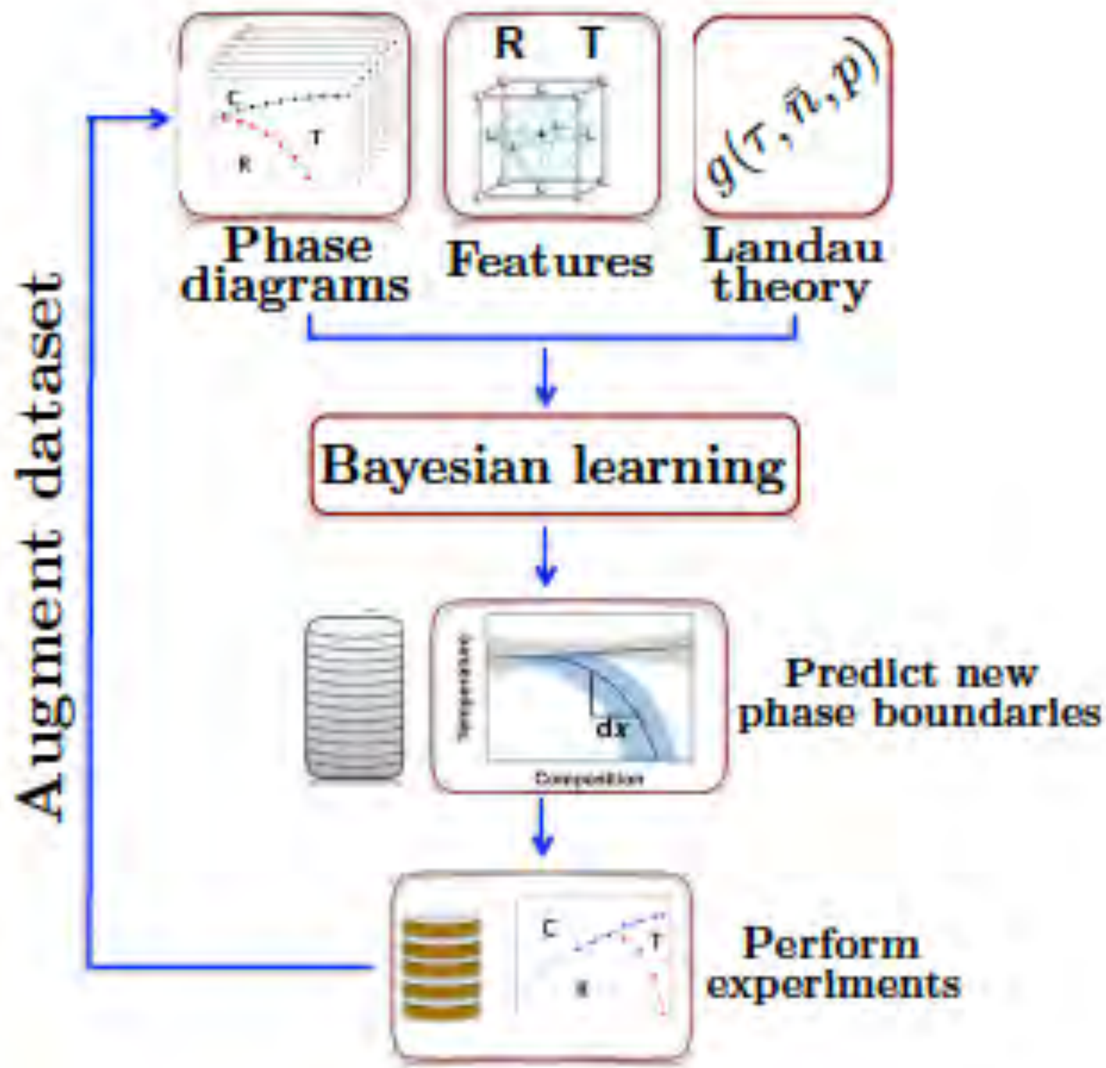
$$\Delta V = V_T - V_R \qquad u_T, u_R$$

$$t_f = \frac{R_A^T + R_o}{R_B^R + R_o} \qquad r_{eff\_nucl} = \frac{A_{enc_T}}{B_{enc_R}} \qquad r_{elec\_neg} = \frac{A_{en_T}}{B_{en_R}}$$

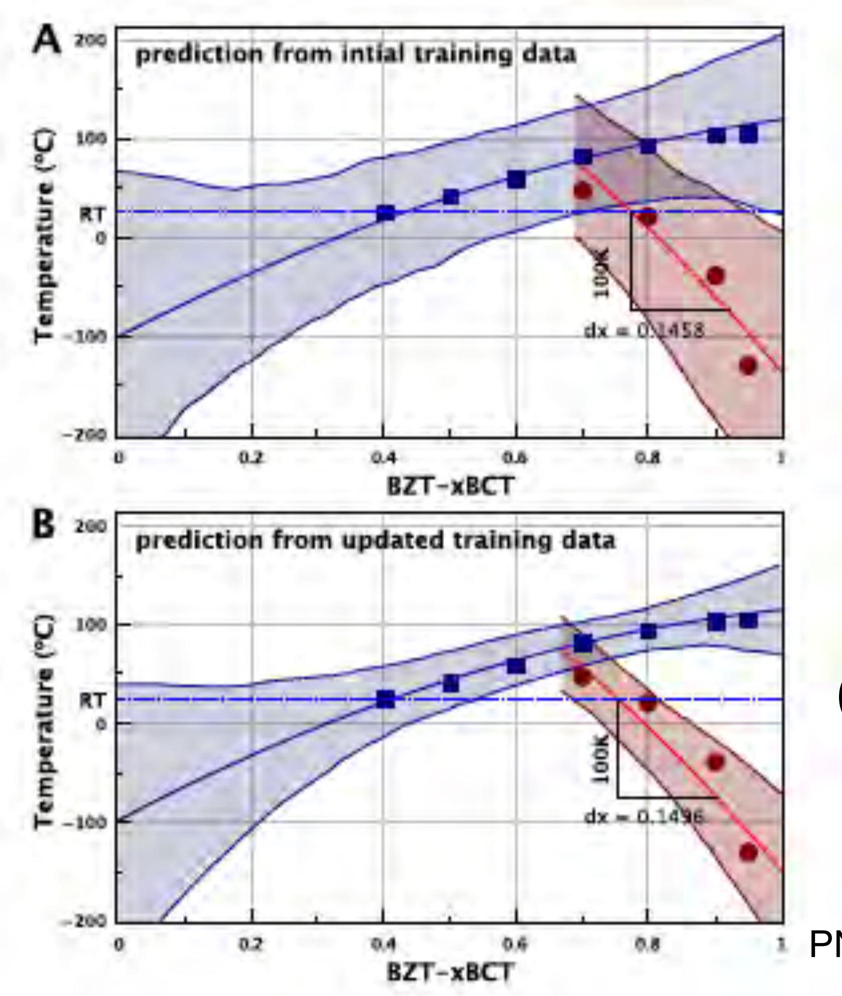
- Training data: 19 phase diagrams

$$\tau = f(\tau_C, a_2, a_6, .., p_R, p_T)$$

- Prior distribution subject to constraints
- Take samples from posterior



Predictions/ synthesis from model + data



$(Ba_{0.5}Ca_{0.5})TiO_3$ - $Ba(Ti_{0.7}Zr_{0.3})O_3$

PNAS, 2016

Other materials problems

- Calorics, Magnetic Shape Memory Alloys, High Entropy Alloys, …….

- Connections to Advanced Manufacturing

Chemical space explored: (Ba,Ca,Sr)(Ti,Zr,Sn,Hf)O_3: Mn, Fe

Pb-free based electrocaloric database

Range	Ba	Ca	Sr
Minimum	0.6	0	0
Maximum	1	0.21	0.4

Range	Ti	Zr	Sn	Hf
Minimum	0.82	0	0	0
Maximum	1	0.18	0.18	0.17

Mn: 0-0.003
Fe: 0-0.015

Number of unique chemical compositions explored in the literature: ~48
Potential chemical compositions: >10,000

Our database contains the following information:

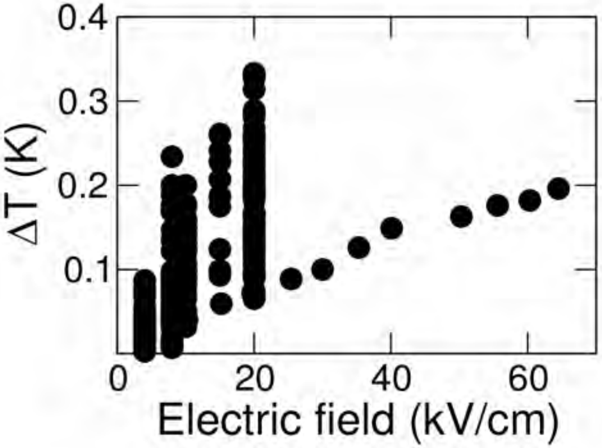
- Chemical composition
- Applied electric field
- Measurement temperature
- Type of Measurement (Direct or Indirect method)
- Measured △T and △S at each electric field and temperature

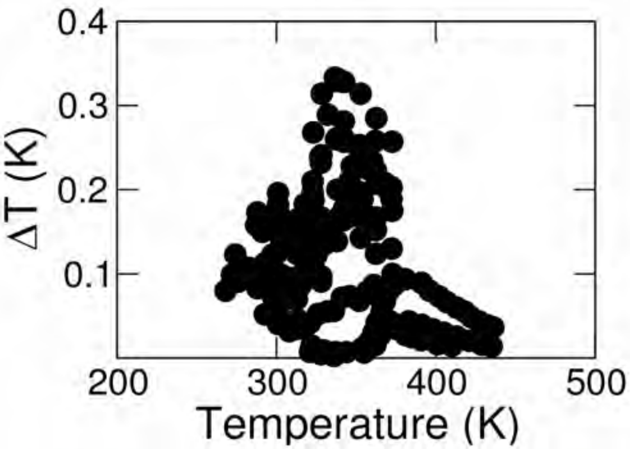
Total number of datapoints in our database: 1923
Number of datapoints with direct method: 200
Number of datapoints with indirect method: 1723

**Potential chemical space from crystal chemistry principles (isovalent constraints)*:
(Ba,Ca,Sr,*Cd,Sn*)(Ti,Zr,Sn,Hf)O_3: Mn, Fe

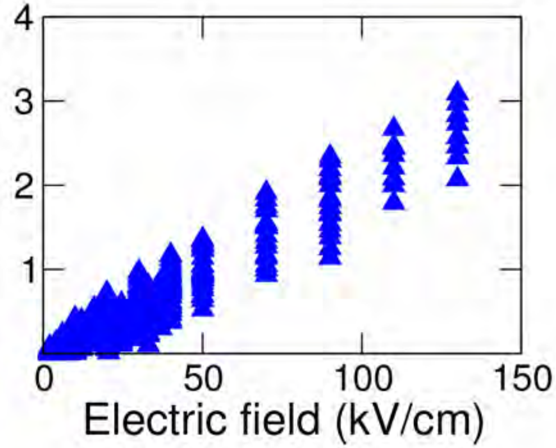
Cursory look at data

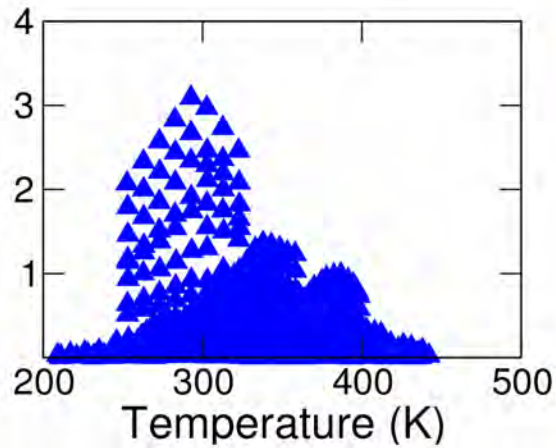
Direct method





Indirect method





Direct method: Best compound

APPLIED PHYSICS LETTERS **106**, 062901 (2015)

Strong electrocaloric effect in lead-free 0.65Ba($Zr_{0.2}Ti_{0.8}$)O_3-0.35($Ba_{0.7}Ca_{0.3}$)TiO_3 ceramics obtained by direct measurements

Mehmet Sanlialp,[1] Vladimir V. Shvartsman,[1] Matias Acosta,[2] Brahim Dkhil,[3] and Doru C. Lupascu[1]

T=~337 K
$Ba_{0.904}Ca_{0.096}Ti_{0.864}Zr_{0.136}O_3$
E=20 kV/cm
max ΔT=0.333 K

Indirect method: Best compound

Journal

Enhanced Electrocaloric Effects in Spark Plasma-Sintered $Ba_{0.65}Sr_{0.35}TiO_3$-Based Ceramics at Room Temperature

Xiao Qiang Liu,[†] Ting Ting Chen, Yong Jun Wu, and Xiang Ming Chen*

T=~293 K
$Ba_{0.65}Sr_{0.35}Ti_{0.997}Mn_{0.003}O_3$
E=130 kV/cm
max ΔT=3.09 K

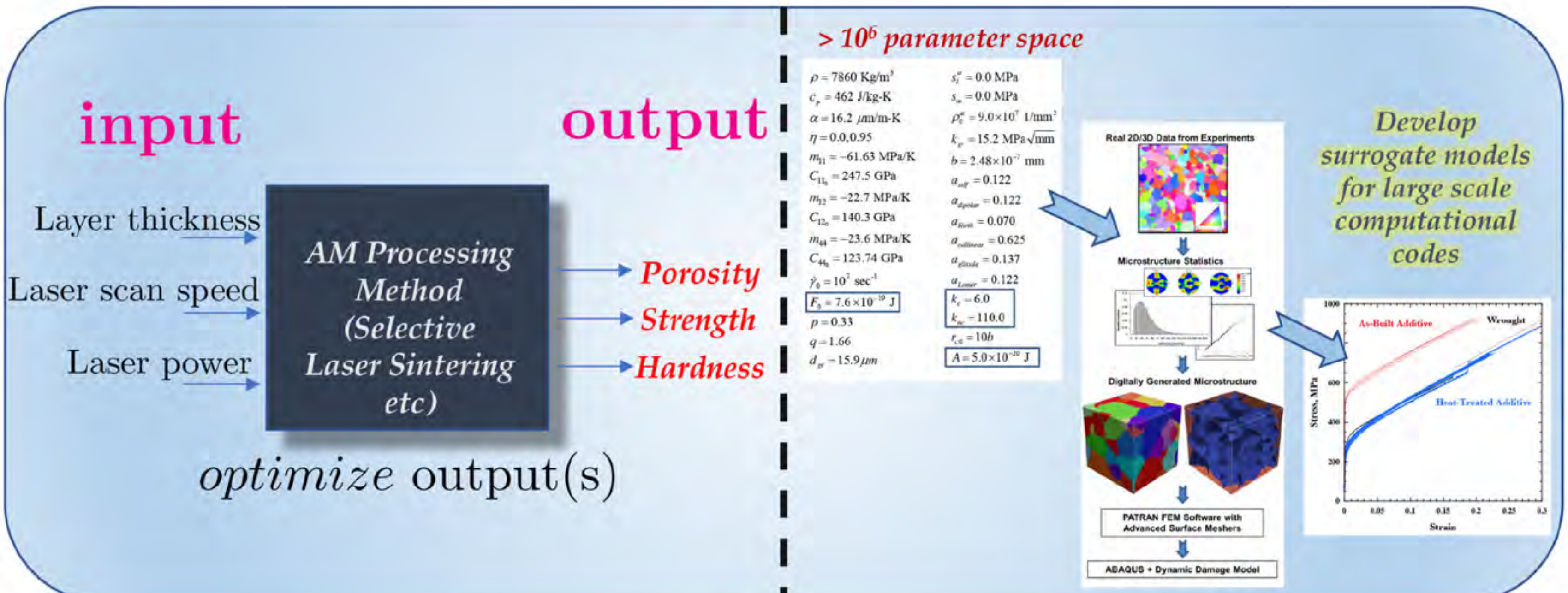
2016 Materials for the Future

Machine learning objective: *Predict* ΔT = *f*(Composition, Temperature, Electric field)

Impact of Design tools on AM

- *Guide AM processing conditions towards targeted objectives with* ***fewest experiments****.*

- *Rapidly* ***identify model*** *parameters (> 1M possibilities) in large scale computational codes to model AM response of microstructures.*



Accelerate optimization of AM processing conditions
(toy problem)

Int J Adv Manuf Technol (2015) 80:555–565

Objective: *maximize* Open porosity

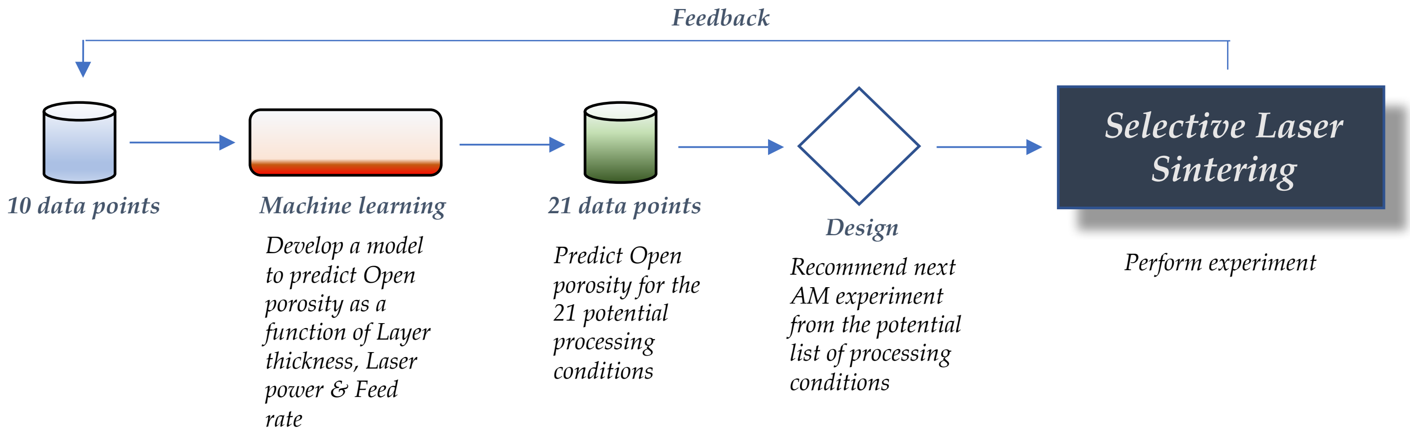
$$\text{Open porosity} = f(\text{Layer thickness, Laser Power, Feed rate})$$

$$\text{Layer thickness} = ??$$
$$\text{Laser power} = ??$$
$$\text{Feed rate} = ??$$

$$\text{Open porosity (\%)} = \frac{\text{Theoretical density - Apparent density}}{\text{Theoretical density}}$$

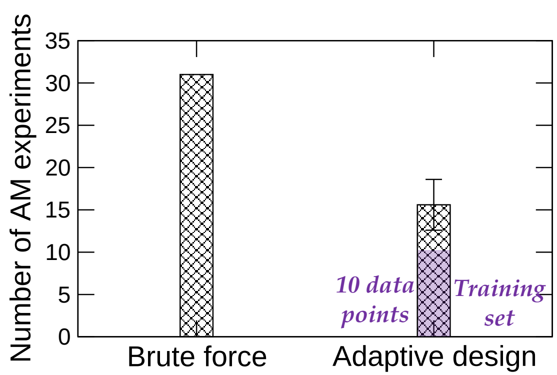
- *Number of AM SLS experiments reported in the literature: 31 data points*
- *Problem formulation to demonstrate the efficacy of machine learning and adaptive design:*
    - *Randomly choose 10 out of 31 data points.*
    - *Can we identify the optimal processing parameter within 21 additional iterations?*

WORKFLOW: Accelerate optimization of AM processing conditions (toy problem)



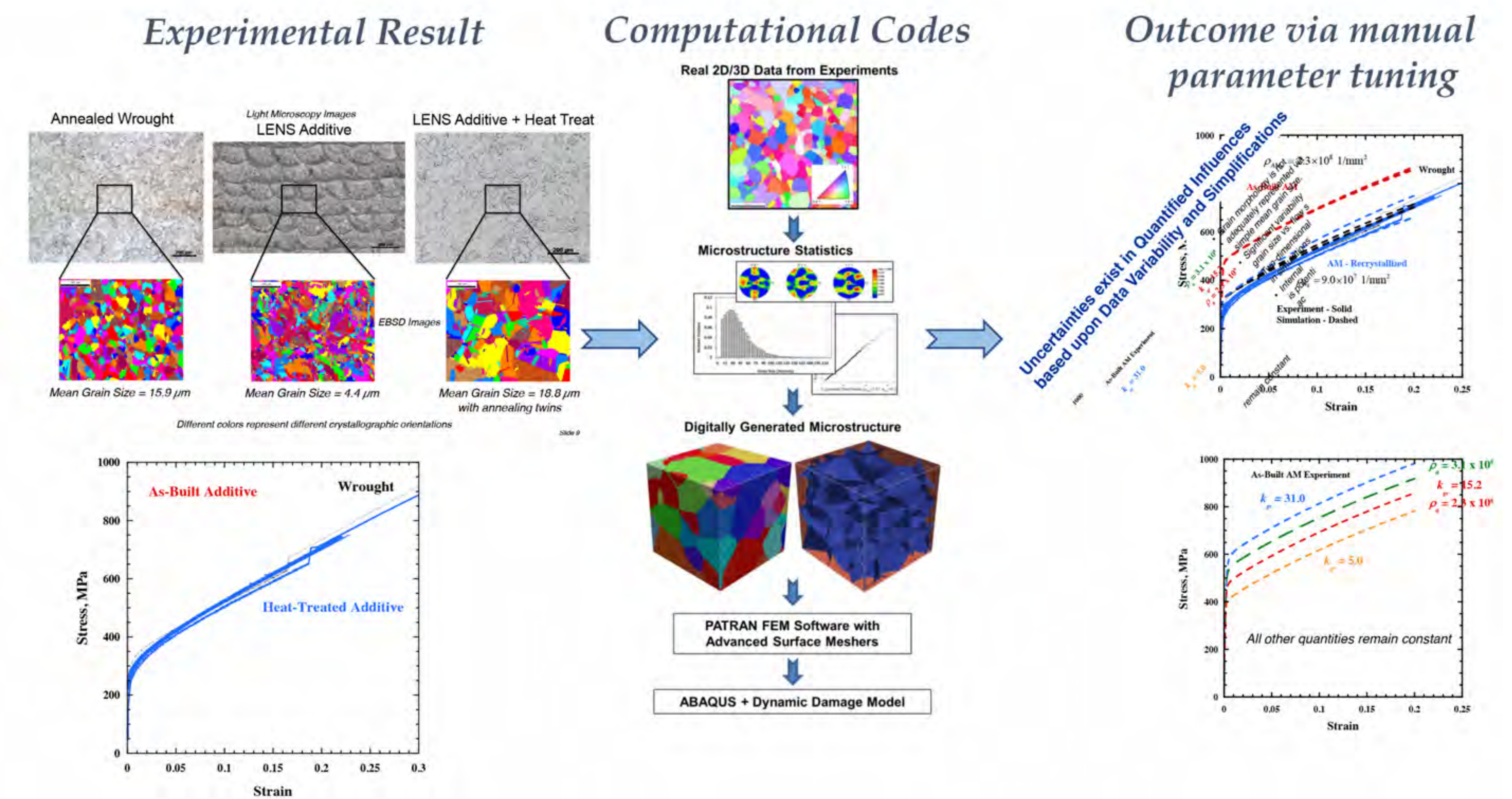
Efficacy of the above machine learning and optimal design capability was recently demonstrated for accelerating new materials discovery.

Balachandran *et al. Sci. Rep.* **6** 19660 (2016).
Xue *et al. Nat. Commun.* **7** 11241 (2016).
Lookman *et al. Curr. Opin. Sol. St. Mater.* (2016).
Xue *et al. PNAS* **113** 13301-13306 (2016)
Xue *et al. Acta Mater.* **125** 532-541 (2017).



*Adaptive design requires ∼**6±3** new AM experiments to find the optimal AM processing condition!*

Rapid Parameter Search for Large Scale Computational Codes



Parameters used in FEM models

elastic

$$\mathbf{L}^p = \dot{\mathbf{F}}^p \mathbf{F}^{p^{-1}} = \sum_\alpha \dot{\gamma}^\alpha \mathbf{S}_0^\alpha \qquad \mathbf{S}_0^\alpha \equiv \mathbf{m}_0^\alpha \otimes \mathbf{n}_0^\alpha$$

Slip System Direction Vector $\mathbf{m}_0^\alpha$

Slip System Normal Vector $\mathbf{n}_0^\alpha$

$$\dot{\gamma}^\alpha = \dot{\gamma}_0 \exp\left[-\frac{F_0}{k\theta}\left\langle 1-\left\langle \frac{|\tau^\alpha|}{\frac{\mu}{\mu_0}\left(s_\rho^\alpha + s_l^\alpha\right)}\right\rangle^p\right\rangle^q\right]\operatorname{sgn}\left(\tau^\alpha\right)$$

Thermally activated glide

Structural resistance @ 0K s_ρ^α

Lattice resistance @ 0K s_l^α

$$\tau^\alpha \equiv \left(\mathbf{C}^*\mathbf{T}^*\right)\cdot\mathbf{S}_0^\alpha \approx \mathbf{T}^*\cdot\mathbf{S}_0^\alpha \qquad \mathbf{C}^* = \mathbf{F}^{*T}\mathbf{F}^*$$

Resolved shear stress – small elastic strain

$$\mu(\theta) = \sqrt{C_{44}(\theta)\left(\frac{C_{11}(\theta)-C_{12}(\theta)}{2}\right)} \qquad C_{ij}(\theta) = C_{ij_0} + m_{ij}\theta$$

Shear modulus – temperature

plastic

Structural Evolution

$$s_\rho^\alpha = s_\infty + \frac{k_{gr}}{\sqrt{d_{gr}}} + \mu b\sqrt{\sum_\beta a^{\alpha\beta}\rho^\beta}$$

Long range resistance s_∞

Grain size d_{gr}

Dislocation interaction $a^{\alpha\beta}$

Dequiedt et al, JMPS 83, 2015
Kubin, Oxford Materials, 2013
Madec et al, Science 301, 2003
Devincre et al, Scripta 54, 2006
Devincre et al, Science 320, 2008
Kubin et al, Acta Mat 56, 2008
Hansen et al, IJP 44, 2013

$$\dot{\rho}^\alpha = \frac{1}{b}\left(\sqrt{\sum_\beta d^{\alpha\beta}\rho^\beta} - 2r_c\rho^\alpha\right)\left|\dot{\gamma}^\alpha\right|$$

Statistically stored dislocation evolution

$$d^{\alpha\beta} = \frac{a^{\alpha\beta}}{k_c^2}$$ *Intersecting slip systems*

$$d^{\alpha\beta} = \frac{a^{\alpha\beta}}{k_{nc}^2}$$ *Self interacting and coplanar systems*

$$r_c = r_{c0}\left(\frac{|\dot{\gamma}^\alpha|}{\dot{\gamma}_0}\right)^{\frac{k\theta}{A}}$$ *Annihilation capture radius*

$a^{\alpha\alpha} = a_{self}$ *Self Interaction*

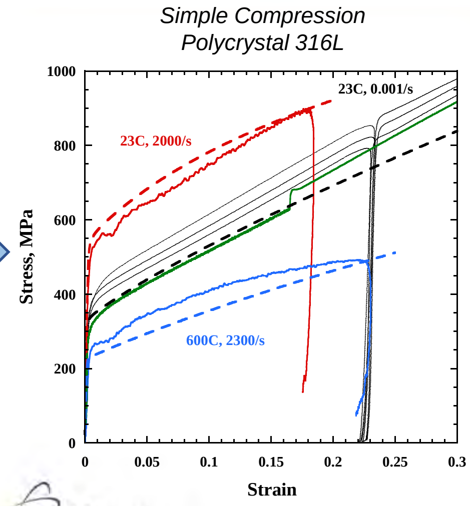
$a^{\alpha\beta} = a_{dipolar}$ *Dipolar Interaction*

a_{Hirth} *Hirth lock*

$a_{collinear}$ *Collinear Interaction*

$a_{glissile}$ *Glissile junction*

a_{Lomer} *Lomer lock*



$\rho = 7860$ Kg/m^3

$c_p = 462$ J/kg-K

$\alpha = 16.2\ \mu$m/m-K

$\eta = 0.0, 0.95$

$m_{11} = -61.63$ MPa/K

$C_{11_0} = 247.5$ GPa

$m_{12} = -22.7$ MPa/K

$C_{12_0} = 140.3$ GPa

$m_{44} = -23.6$ MPa/K

$C_{44_0} = 123.74$ GPa

$\dot{\gamma}_0 = 10^7$ sec^{-1}

$\boxed{F_0 = 7.6\times10^{-19}\ \text{J}}$

$p = 0.33$

$q = 1.66$

$d_{gr} = 15.9\,\mu m$

$s_l^\alpha = 0.0$ MPa

$s_\infty = 0.0$ MPa

$\rho_0^\alpha = 9.0\times10^7$ 1/mm^2

$k_{gr} = 15.2$ MPa$\sqrt{\text{mm}}$

$b = 2.48\times10^{-7}$ mm

$a_{self} = 0.122$

$a_{dipolar} = 0.122$

$a_{Hirth} = 0.070$

$a_{collinear} = 0.625$

$a_{glissile} = 0.137$

$a_{Lomer} = 0.122$

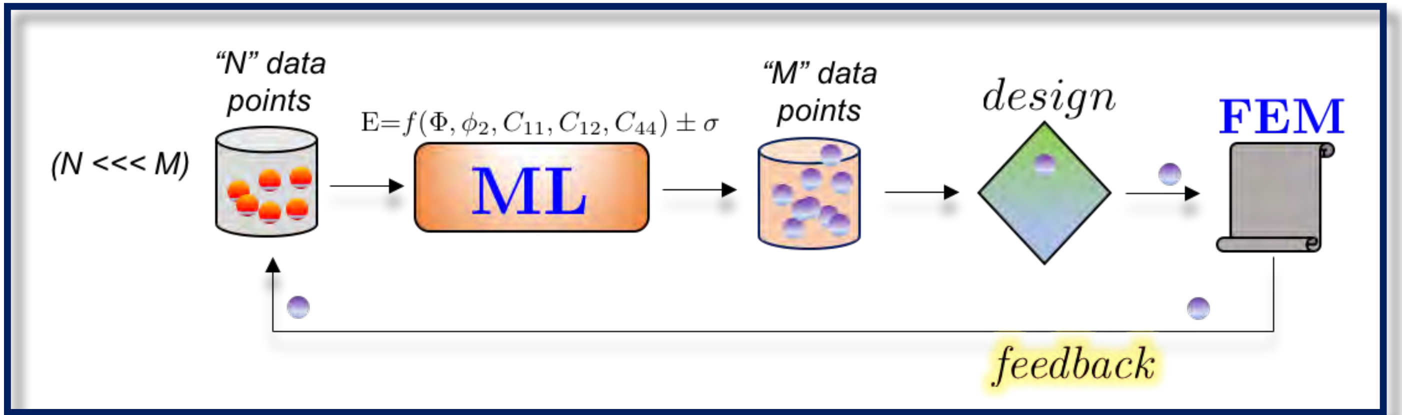
$\boxed{k_c = 6.0 \quad k_{nc} = 110.0}$

$r_{c0} = 10b$

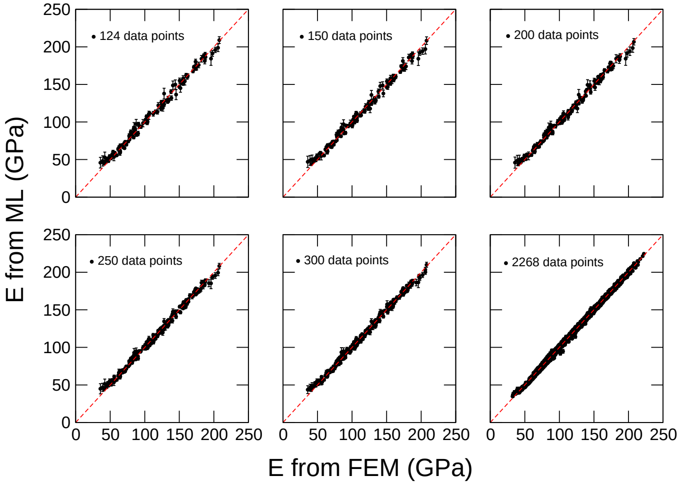
$\boxed{A = 5.0\times10^{-20}\ \text{J}}$

Rapid Parameter Search for Large Scale Computational Codes (elastic problem)

strategy



ML results



- *ML trained with 250 FEM calculations produces results with accuracy comparable to FEM .*
- *We saved ~90% of the computational cost involved in running FEM.*

Challenges – Path forward

- **Integrate** informatics tools with AM with **real time feedback and control** to navigate search space optimally

- Extensions to **multifidelity** and **multiobjectives**


ELSEVIER

Contents lists available at ScienceDirect

Computational Materials Science

journal homepage: www.elsevier.com

Editor's Choice

Multi-fidelity machine learning models for accurate bandgap predictions of solids

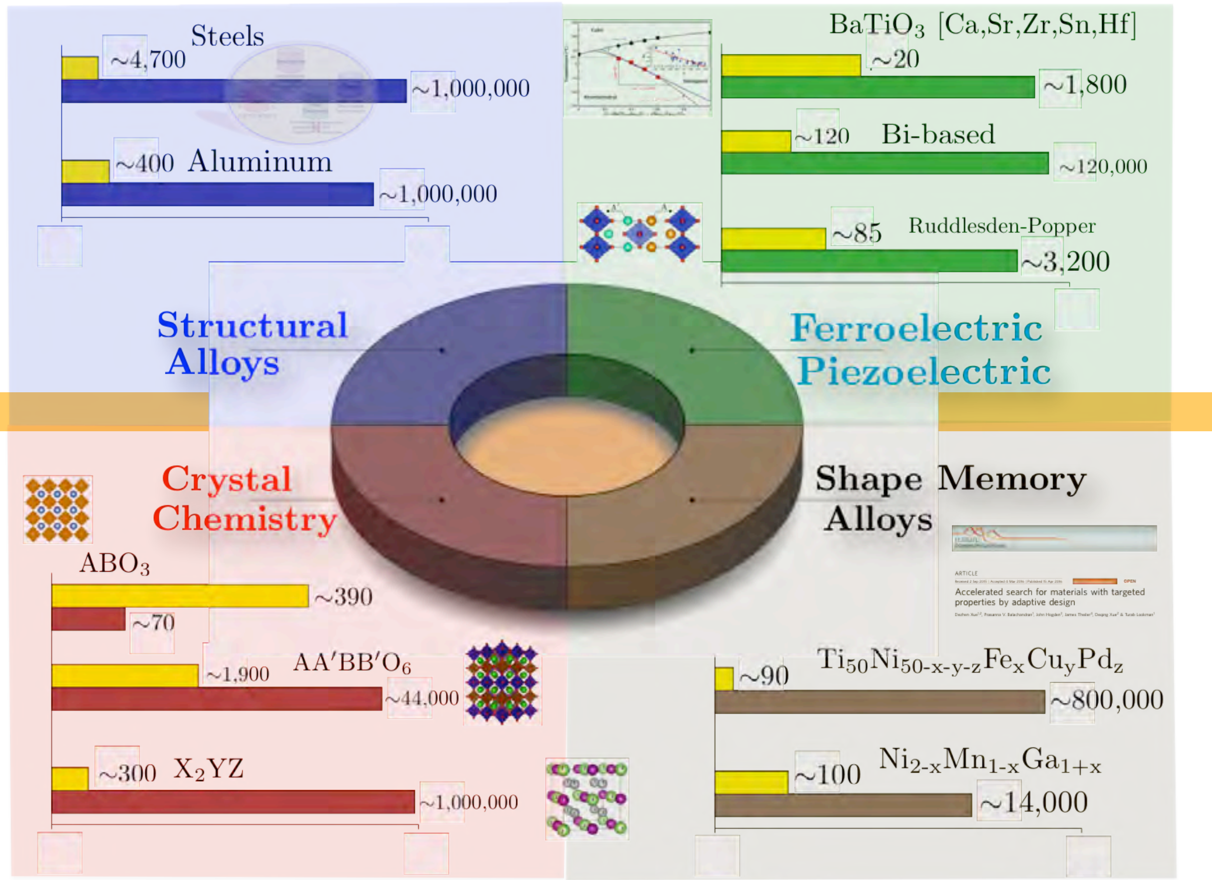
G. Pilania[a,*], J.E. Gubernatis[b], T. Lookman[b]

[a] Materials Science and Technology Division, Los Alamos National Laboratory, Los Alamos, NM 87545, USA
[b] Theoretical Division, Los Alamos National Laboratory, Los Alamos 87545 NM, USA

- Use of **uncertainties** for search

- **Data, data, data** - under controlled conditions!!!!!!!!!!!!!

Material Databases @ LANL



- *Databases integrate computational, phenomenological, processing **and** experimental data.*