MATHEMATICAL FRONTIERS

The National Academies of SCIENCES ENGINEERING MEDICINE

nas.edu/MathFrontiers

Board on Mathematical Sciences & Analytics

MATHEMATICAL FRONTIERS
2019 Monthly Webinar Series, 2-3pm ET

Feb 12: *Machine Learning for Materials Science*

Mar 12: *Mathematics of Privacy*

Apr 9: *Mathematics of Gravitational Waves*

May 14: *Algebraic Geometry*

June 11: *Mathematics of Transportation*

July 9: *Cryptography and Cybersecurity*

Aug 13: *Machine Learning in Medicine*

Sept 10: *Logic and Foundations*

Oct 8: *Mathematics of Quantum Physics*

Nov 12: *Quantum Encryption*

Dec 10: *Machine Learning for Text*

This webinar series is made possible by support for BMSA from the

National Science Foundation
Division of Mathematical Sciences

and the

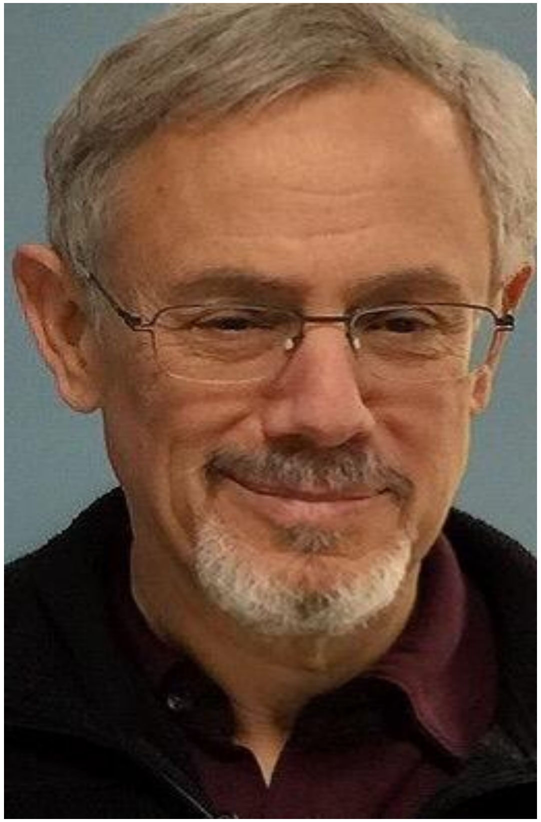
Department of Energy
Advanced Scientific Computing Research

View webinar videos and learn more about BMSA at www.nas.edu/MathFrontiers

MATHEMATICAL FRONTIERS
Machine Learning for Materials Science

Elizabeth Holm,
Carnegie Mellon University

Rampi Ramprasad,
Georgia Institute of Technology



Mark Green,
UCLA (moderator)

MATHEMATICAL FRONTIERS
Machine Learning for Materials Science

Elizabeth Holm,
Carnegie Mellon University

Professor of
Materials Science and Engineering

Incorporating machine intelligence in materials science and engineering

Demystifying machine intelligence (AI)

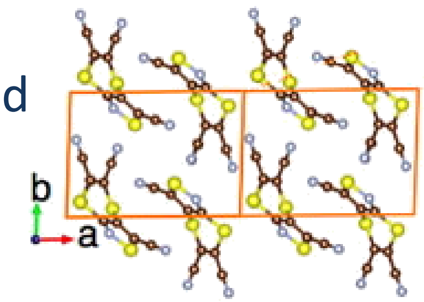
https://xkcd.com/1838/

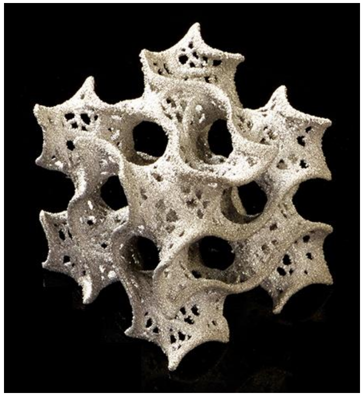
- Machine intelligence is comprised of three basic building blocks:
  - Data science = statistics of multi-dimensional data
  - Machine learning = optimization via feedback loops
  - Convolutional neural networks = signal processing
- These three elements combine to give

Deep learning

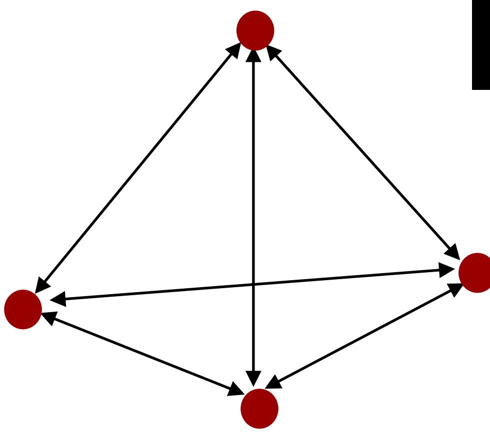
Opportunities for AI in materials science and engineering

- Discovery and design of materials
- Optimization of materials structure and properties
- Material selection
- Autonomous experiments and analysis, including adaptive experiments
- Managing experimental data streams, including forward modeling
- Extracting information from microstructural images
- Property prediction
- Failure analysis
- Process and quality control
- …



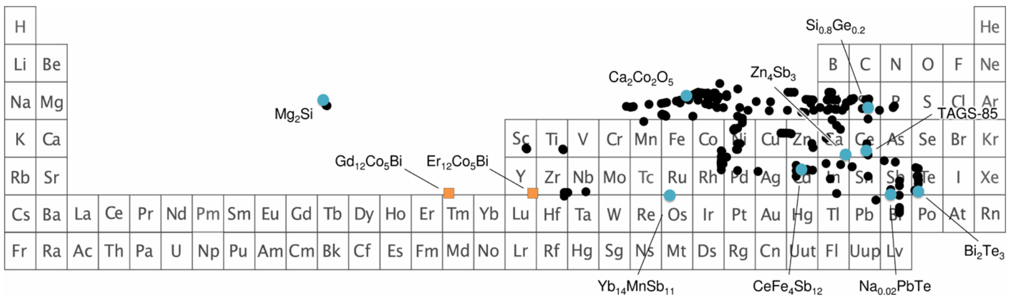


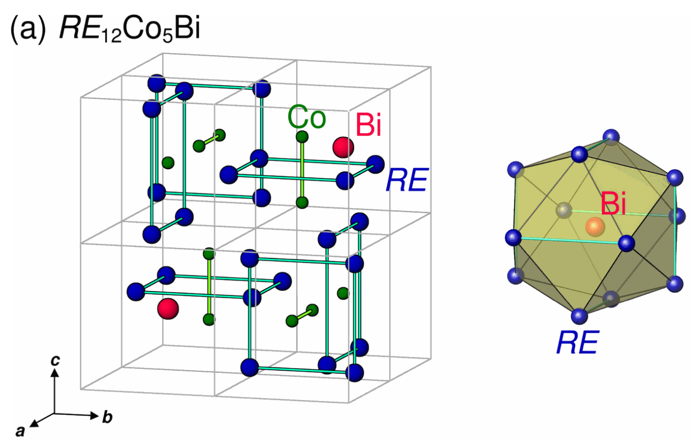
Microstructure

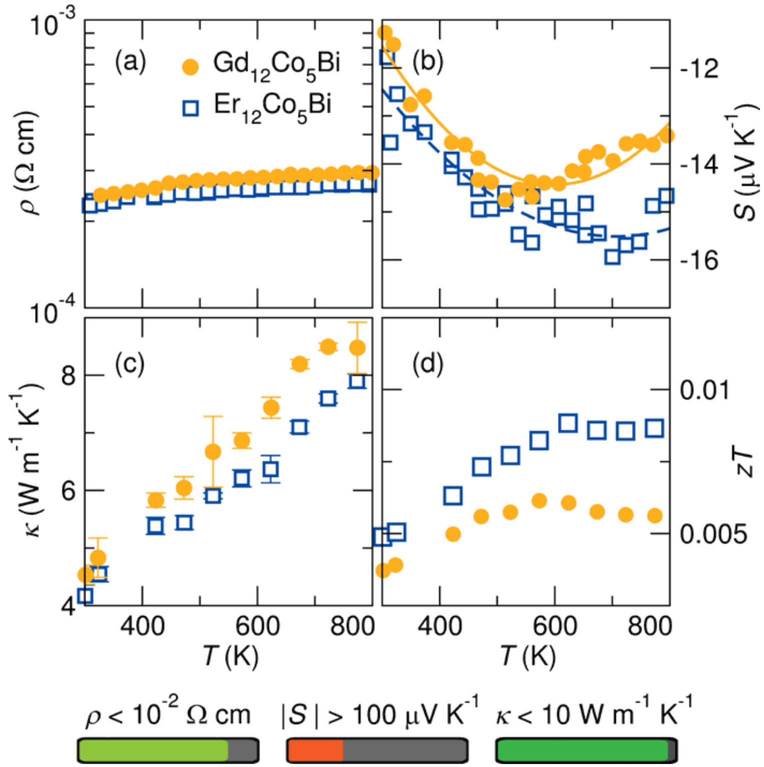


Search for new materials using data science and machine learning

- Big data: Discovery of new thermoelectric material







M. W. Gaultois, A. O. Oliynyk, A. Mar, T. D. Sparks, G. J. Mulholland, B. Meredig, APL Mater., 4 053213 (2016)

What is holding physical scientists back?

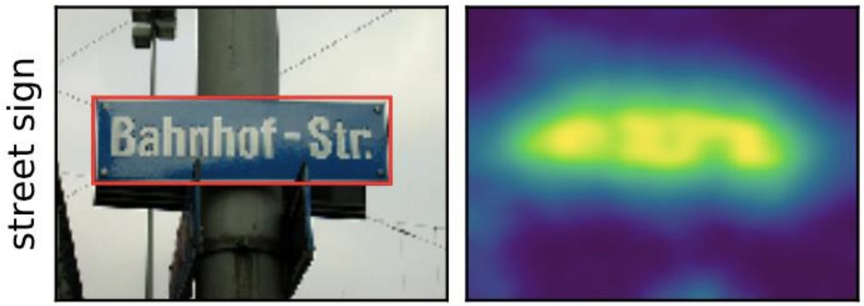
- Informal survey of 14 physical scientists / engineers who attended a CMU workshop on Machine Learning in Science:

  What are the biggest barriers to implementing data science in your research?

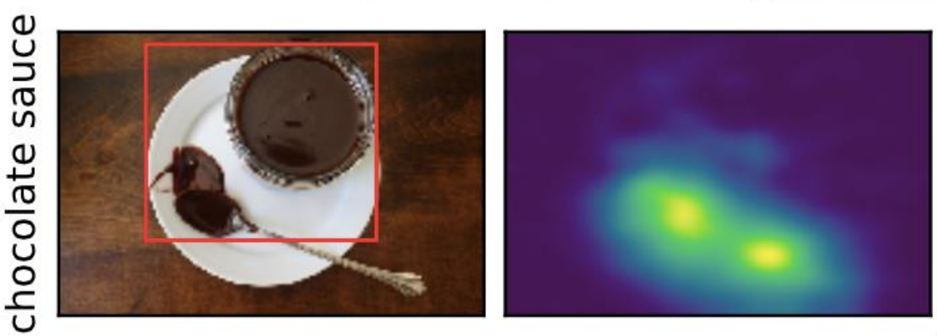
  - Big data (28%)
  - Small data (43%)
  - Rare events (43%)
  - Rich/multimodal data (57%)
  - Data representation (71%)
  - Interpretability (79%)

Why interpretability matters:
The machine doesn't always learn the right things

- A CNN-based deep learning system was trained to identify classes of objects in photographs.
- Masking was used to evaluate critical features that the computer associates with an object.
- Some masks made sense:



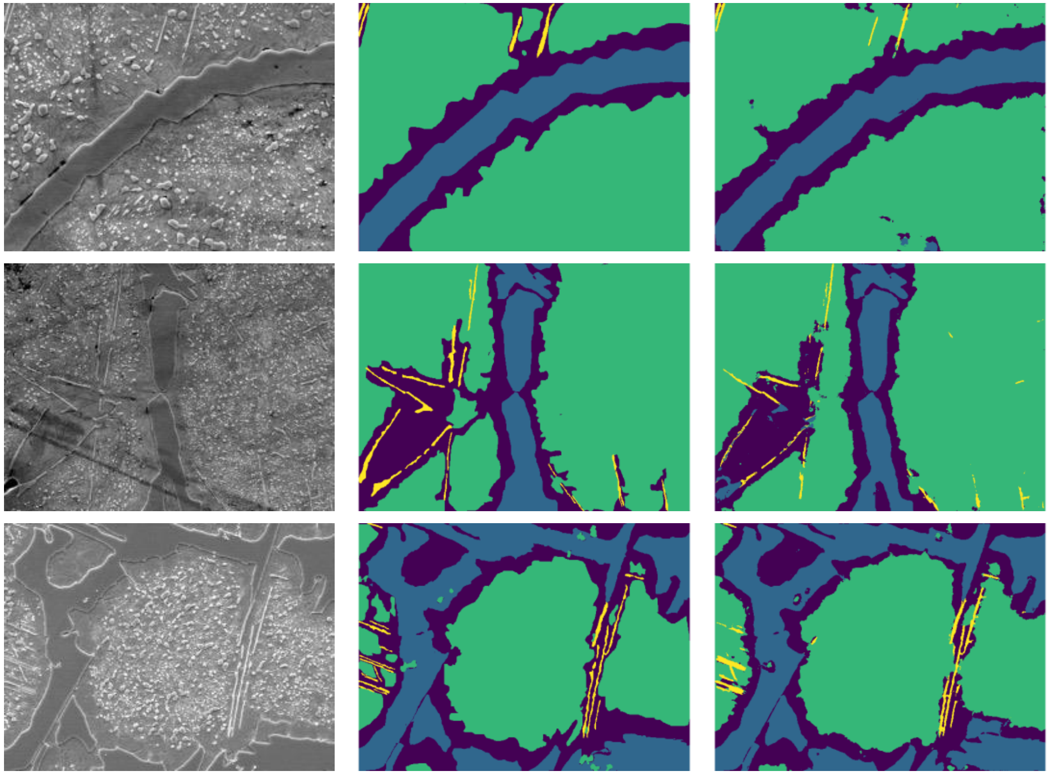
- Some did not:



Fong, et. al arXiv:1704.03296v1

Black box for tedious tasks: Autonomous microstructural segmentation

- Segmenting complex, multi-component microstructures


Original Image Grad student Pixel-Net

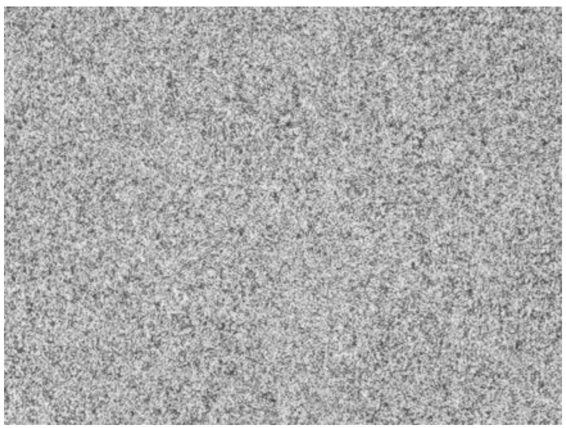
- Accurate (93 ± 3%)
- Objective
- Repeatable
- Indefatigable
- Permanent
- Exactly as interpretable as a graduate student

B. DeCost et al., arXiv:1805.08693

Black box for repetitive workflow: Quality control via computer vision

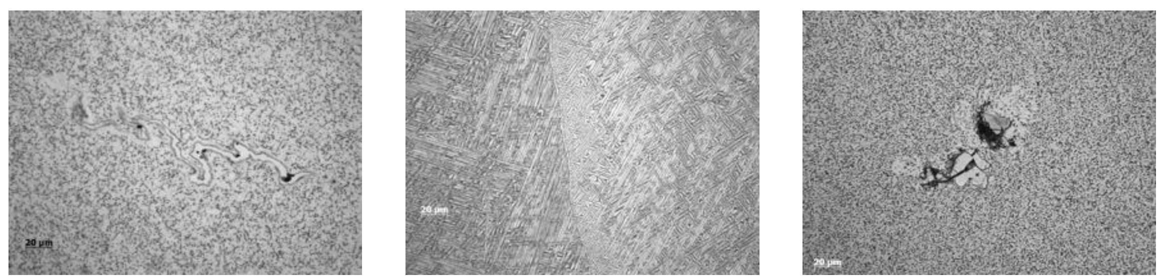
- Autonomous evaluation: identify “out of spec” microstructures

specified microstructure:



Meet specifications:

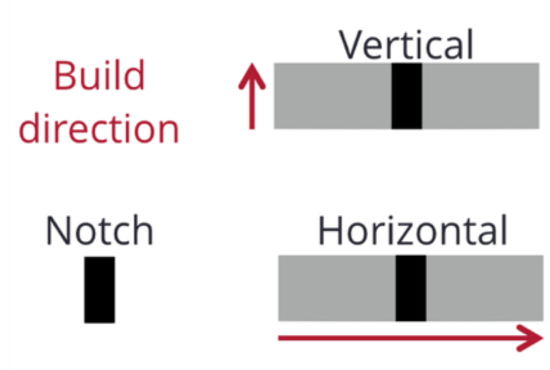
Out of spec:

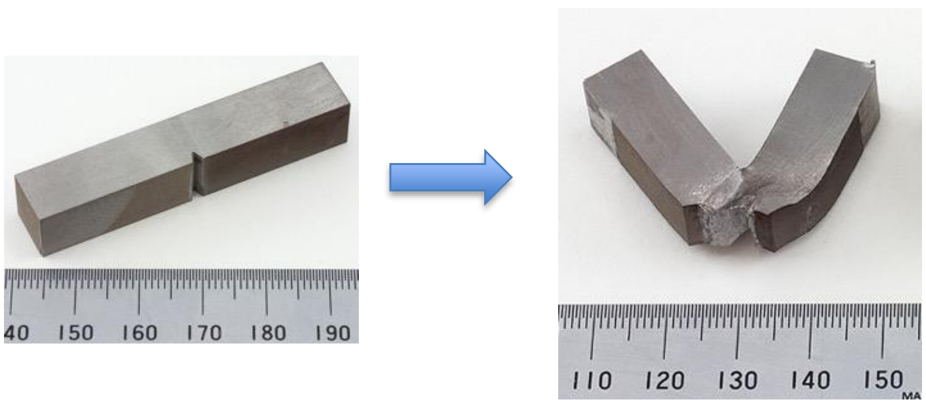


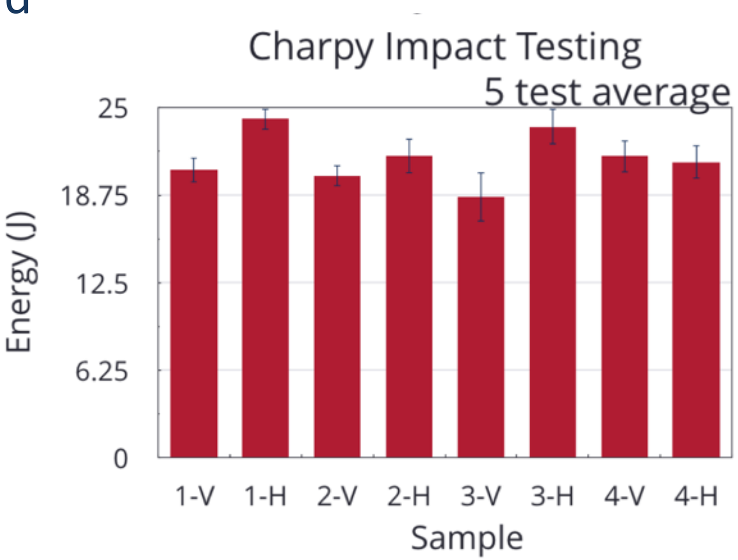
B. L. DeCost, E. A. Holm, *Comp. Mater. Sci.* 110 126-133 (2015).

Where the black box fails: Extracting knowledge from information

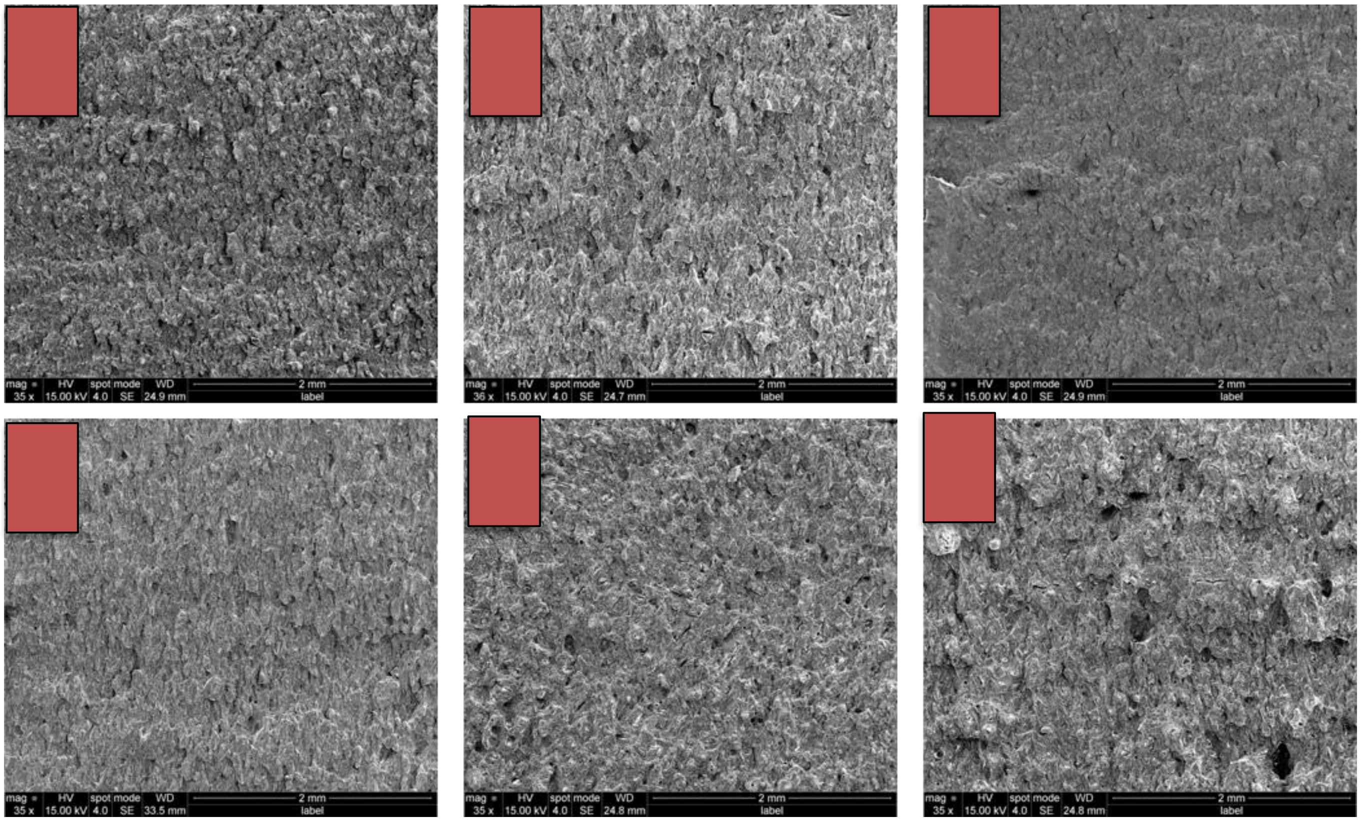
- Extracting information from microstructural images:
  - Inconel 718 Charpy impact specimens built using additive manufacturing.
  - Two build orientations, horizontal and vertical.
  - Charpy impact energy depends on build orientation.



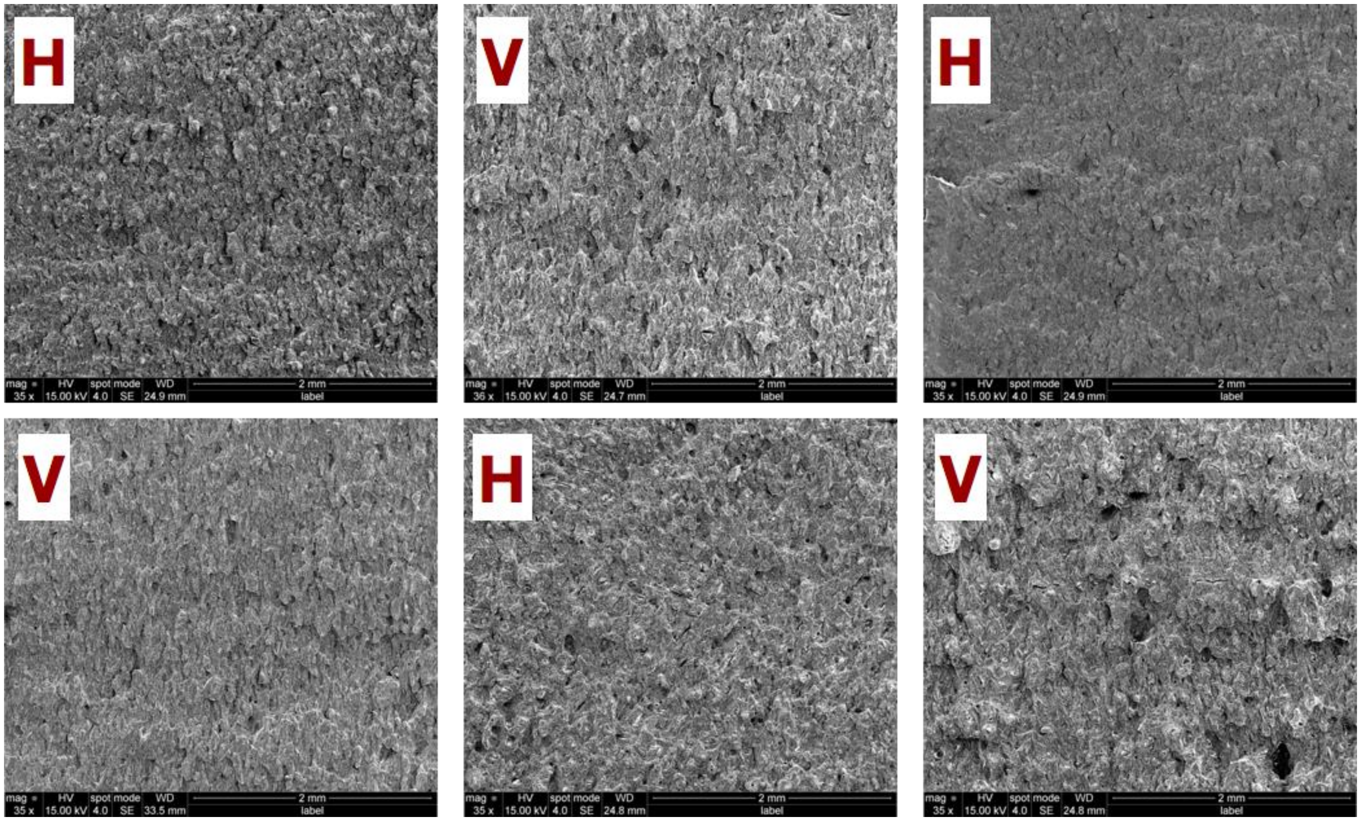




Can you see the difference in the fracture surfaces?

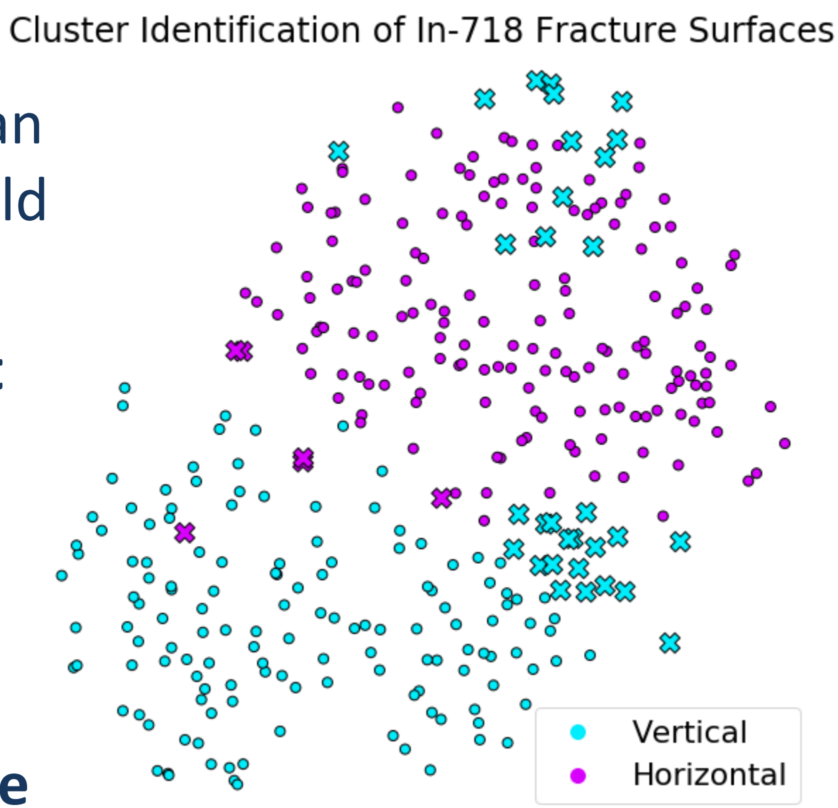


Can you see the difference in the fracture surfaces?



What is the computer learning?

- Using unsupervised ML with k-means clustering, the computer can identify horizontal and vertical build fractures with **88 ± 3 % accuracy**.
- **What does the computer see that we cannot?**
- **Does the distinguishing visual information provide physical insight?**
- **Has the computer learned fracture mechanics?**



A. Kitahara, E. Holm , *IMMI* 7[3] 148 (2018)

Conclusion: Philosophical musings

- When is a black box OK?
  - The overall cost of wrong answers is low.
  - The method is better than all alternatives within its domain.
- When does the black box fail?
  - The goal is not what, but why.

"All right," said Deep Thought. "The Answer to the Great Question..."
"Yes..!"
"Of Life, the Universe and Everything..." said Deep Thought.
"Yes...!"
"Is..." said Deep Thought, and paused.
"Yes...!"
"Is..."
"Yes...!!!...?"
"Forty-two," said Deep Thought, with infinite majesty and calm.

—Douglas Adams, The Hitchhiker's Guide to the Galaxy

Acknowlegements

- Current and former students:
  - Andrew Kitahara
  - Nan Gao
  - Bo Lei
  - Srujana Yarasi
  - Keith Kozlosky
  - Ankita Mangal (P&G)
  - Toby Francis (UCSB)
  - Anna Smith (Merck)
  - Brian DeCost (NIST)

National Science Foundation
DMR-1507830 and DMR-1826218

MATHEMATICAL FRONTIERS
Machine Learning for Materials Science

Rampi Ramprasad,
Georgia Institute of Technology

Michael E. Tennenbaum Family Chair and GRA Eminent Scholar, School of Materials Science and Engineering

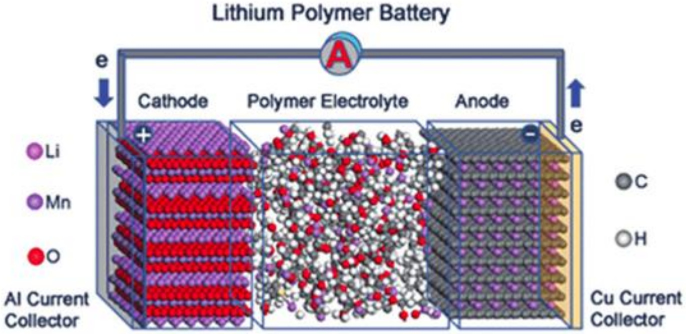
Polymer Genome:
An Informatics Platform for Prediction & Design

POLYMERS ARE UBIQUITOUS

High Energy Density Capacitors

Need: high band gap, high dielectric constant

Solid-State Battery Electrolytes



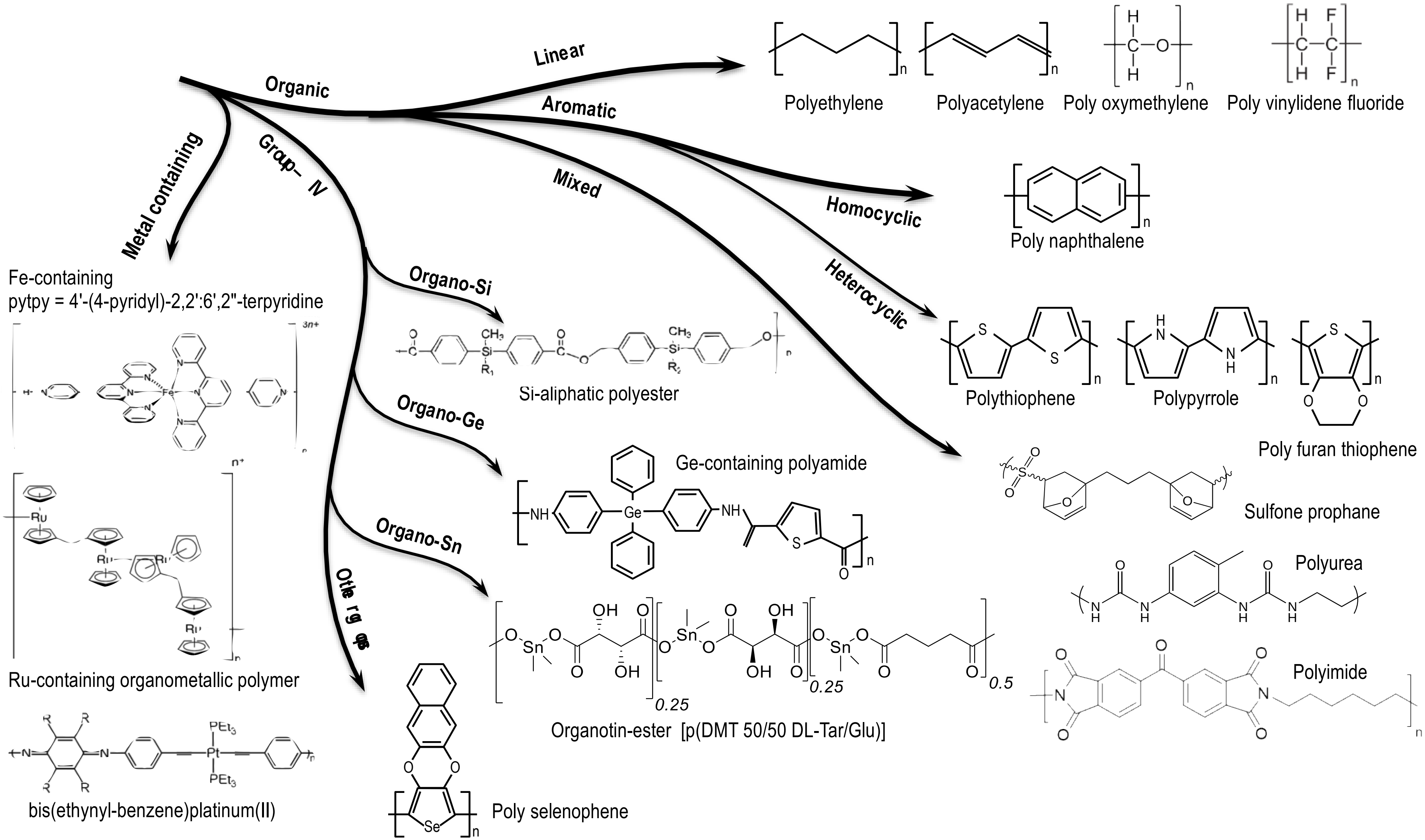
Need: low T_g, high mechanical strength

Organic / Flexible Electronics

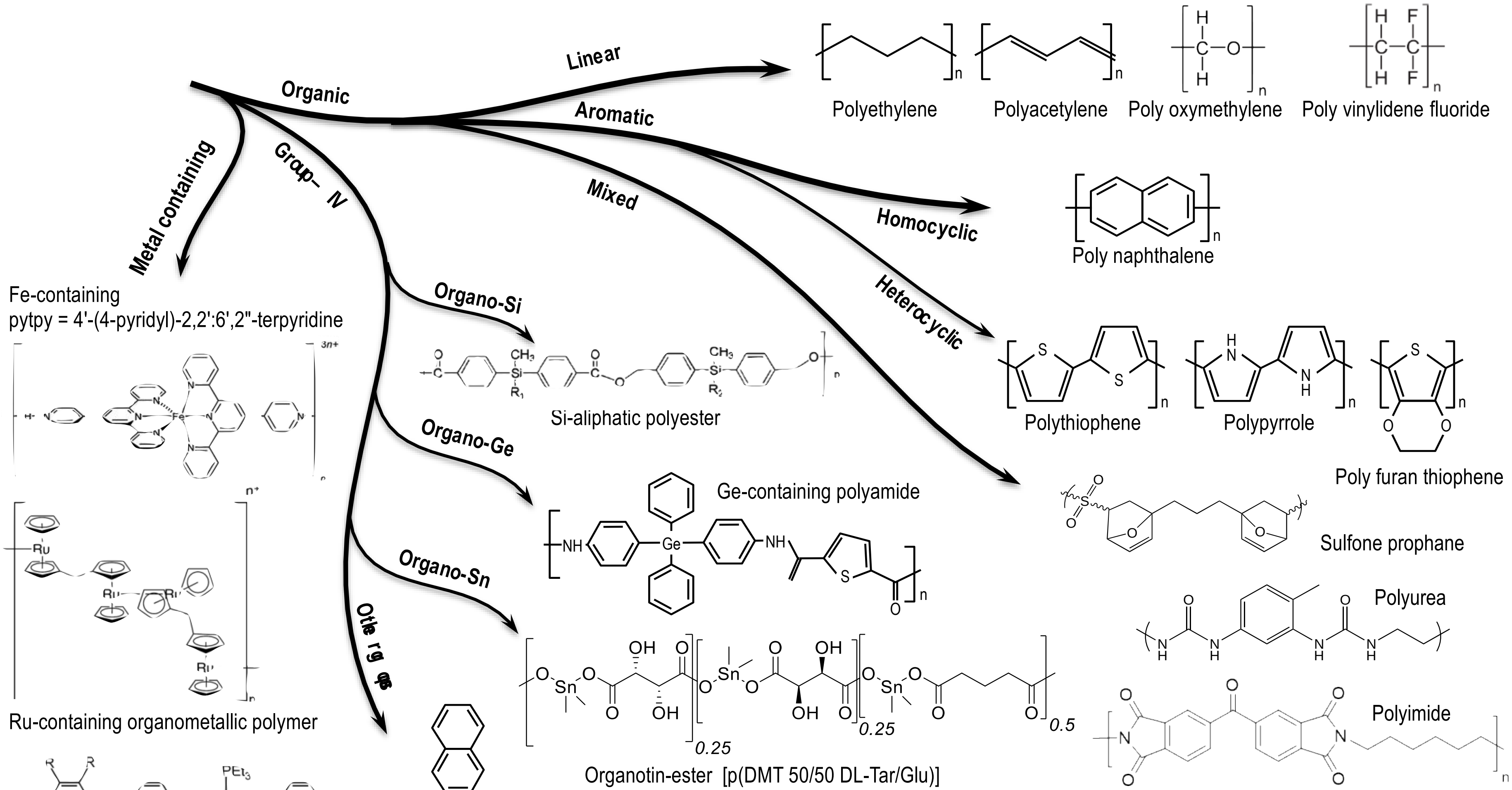
Need: low band gap, low carrier recombination

Different applications have different property requirements
(Optimal materials selection is non-trivial)

POLYMER CHEMICAL UNIVERSE



POLYMER CHEMICAL UNIVERSE



Staggering! Where do we start?!
And how do we search efficiently?

BENCHMARK DATASET

Data sources

Computational data via high-throughput DFT

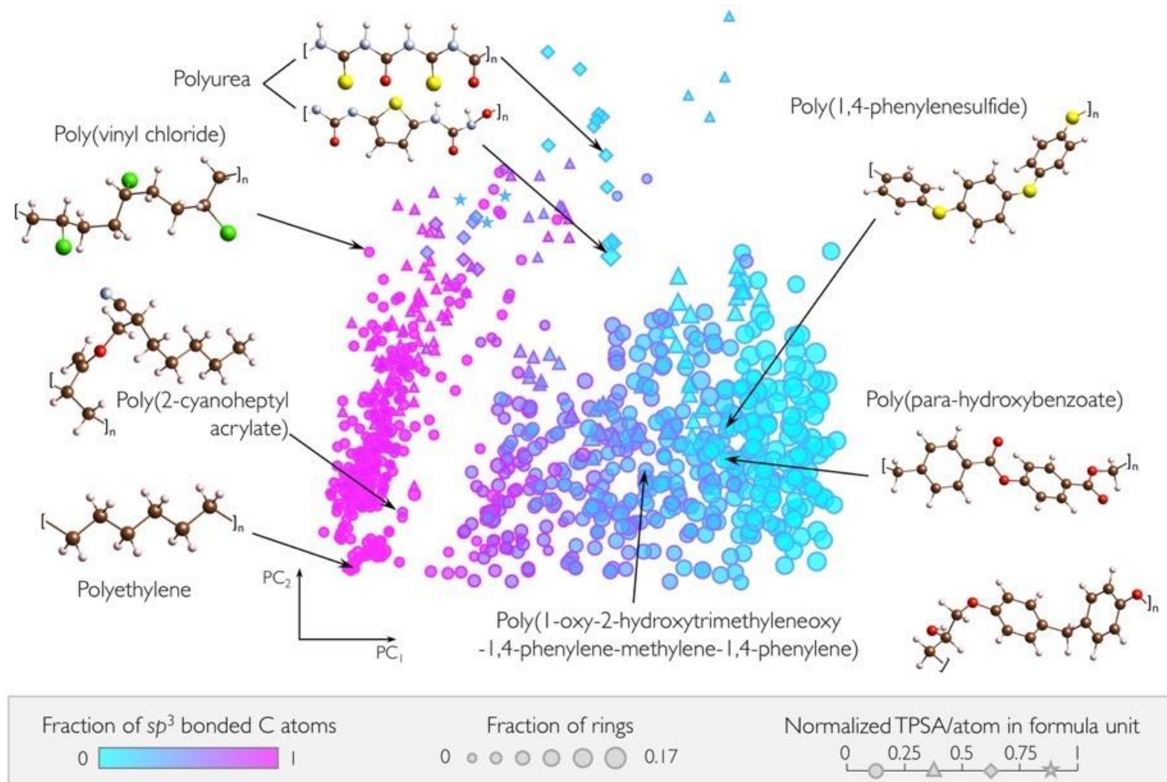
+

Experimental data from collaborators, literature & data collections

Property space

Band gap
Dielectric constant
Glass transition temperature
Atomization energy,
etc.

Chemical space (~900 organic polymers)



MACHINE LEARNING IN MSE

a Example dataset

Material	Property Value
Material 1	P_1
Material 2	P_2
⋮	⋮
Material N	P_N

b The learning problem

Material	Property Value
Material X	?

c Fingerprinting, learning and prediction

Material	Fingerprint	Property Value
Material 1 →	$F_{11}, F_{12}, \ldots F_{1M}$ →	P_1
Material 2 →	$F_{21}, F_{22}, \ldots F_{2M}$ →	P_2
⋮	⋮	⋮
Material N →	$F_{N1}, F_{N2}, \ldots F_{NM}$ →	P_N

Fingerprinting

Learning

Prediction Model

$f(F_{i1}, F_{i2}, \ldots, F_{iN}) = P_i$

T. Mueller, A. G. Kusne, R. Ramprasad “Machine Learning in Materials Science: Recent Progress and Emerging Applications”, Reviews in Computational Chemistry, John Wiley & Sons, Inc., Volume 29, (2016).

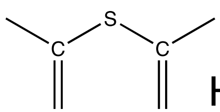
R. Ramprasad, R. Batra, G. Pilania, A. Mannodi-Kanakkithodi, C. Kim, “Machine Learning and Materials Informatics: Recent Applications and Prospects”, npj Computational Materials 3, 54 (2017).

POLYMER FINGERPRINTS (GENOME)

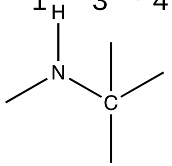
We represent polymers numerically at three length-scales

Atomic-level

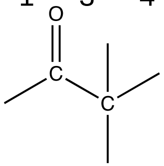
C_3-S_2-C_3



H_1-N_3-C_4



O_1-C_3-C_4

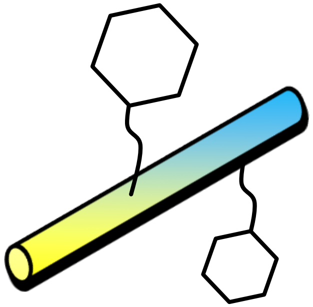


Block-level

C_6H_4, CH_2, etc., blocks;
Van der Waals volume/surface
, etc.

Chain-level

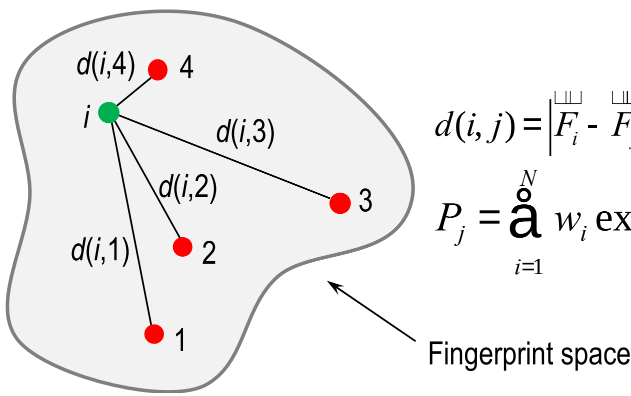
Distance between rings
Length of sidechain
Aromatics in sidechain
...



"Polymer Genome: A Data-Powered Polymer Informatics Platform for Property Predictions"
C. Kim, A. Chandrasekaran, T. D. Huan, D. Das and R. Ramprasad, J. Phys. Chem. C (2018)

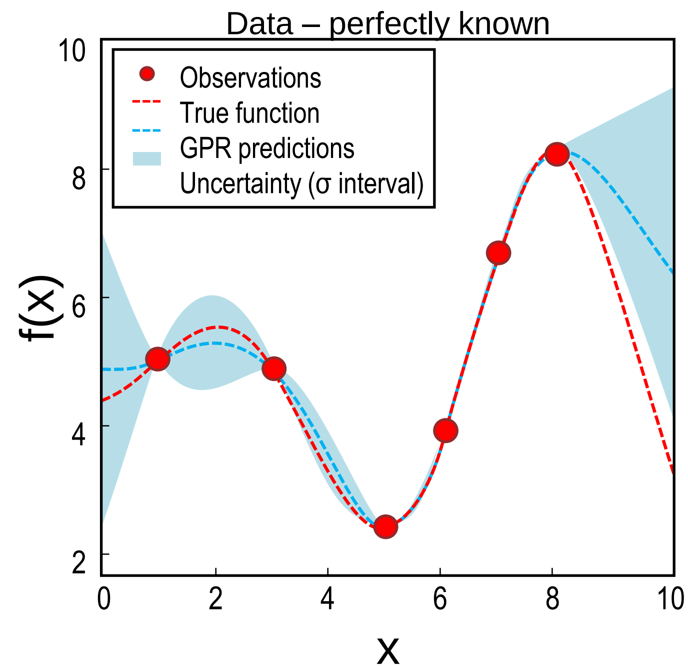
LEARNING FROM DATA

Gaussian process regression (GPR)



$$d(i,j) = \left| \vec{F}_i - \vec{F}_j \right|$$

$$P_j = \sum_{i=1}^{N} w_i \exp\frac{-d(i,j)^2}{2\sigma^2}$$

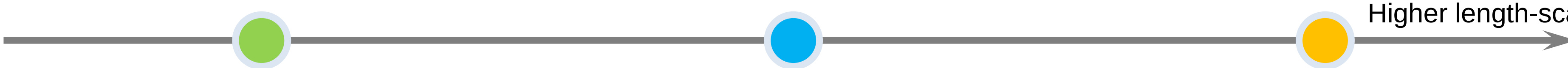


"Polymer Genome: A Data-Powered Polymer Informatics Platform for Property Predictions"
C. Kim, A. Chandrasekaran, T. D. Huan, D. Das and R. Ramprasad, J. Phys. Chem. C (2018)

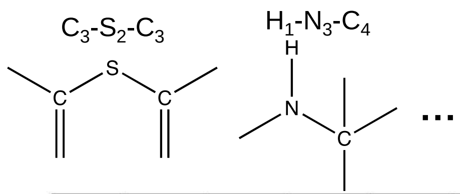
HIERARCHICAL FINGERPRINTING

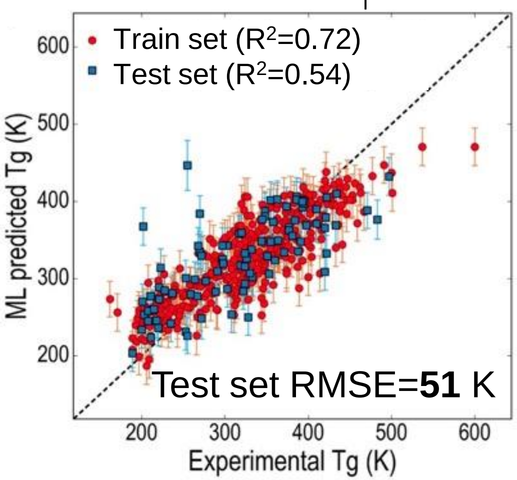
Impact on glass transition

Atomic-level descriptors | + Block-level descriptors | + Chain-level descriptors



Atom-triples



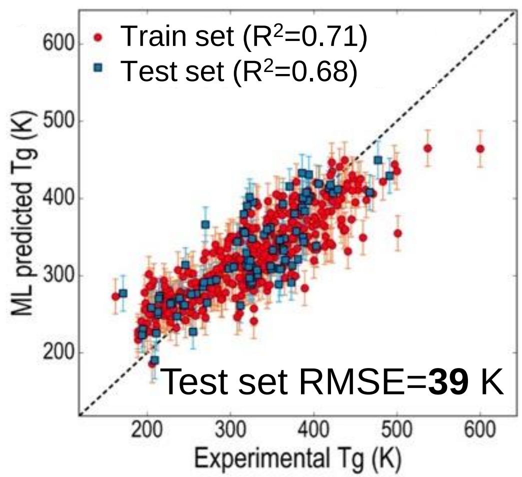


Van der Waals volume

Types of blocks

Fraction of rotatable bonds

...

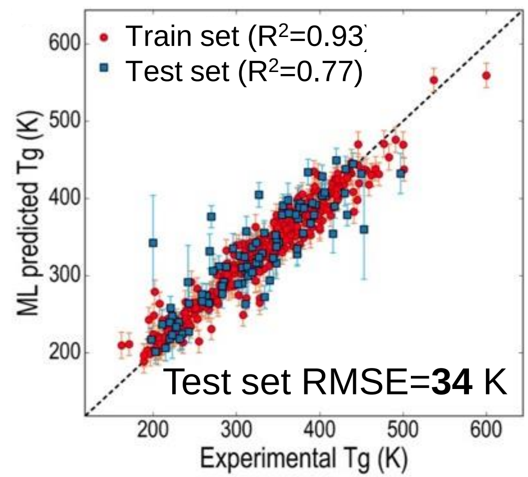


Distance between rings

Length of sidechain

Length of main chain

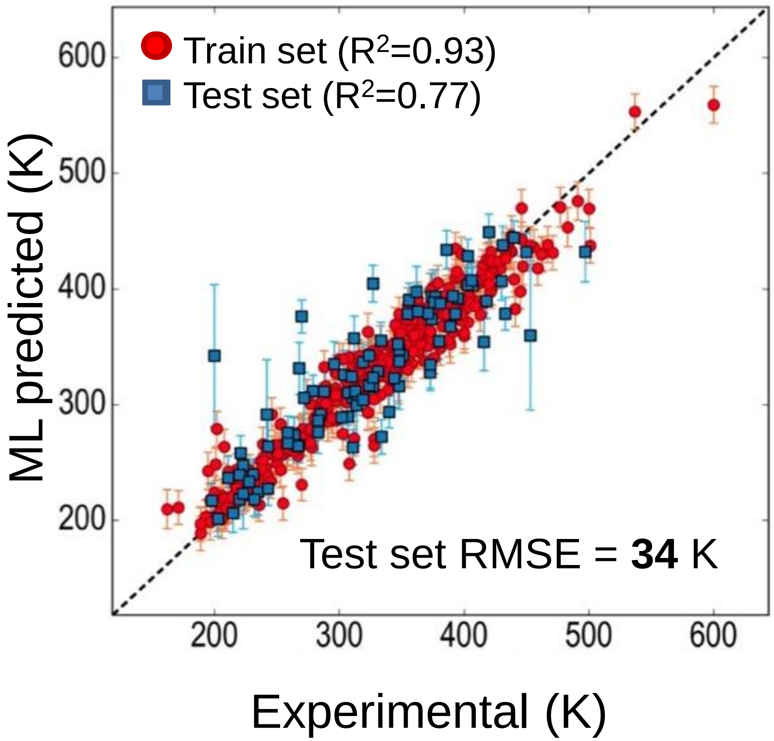
...



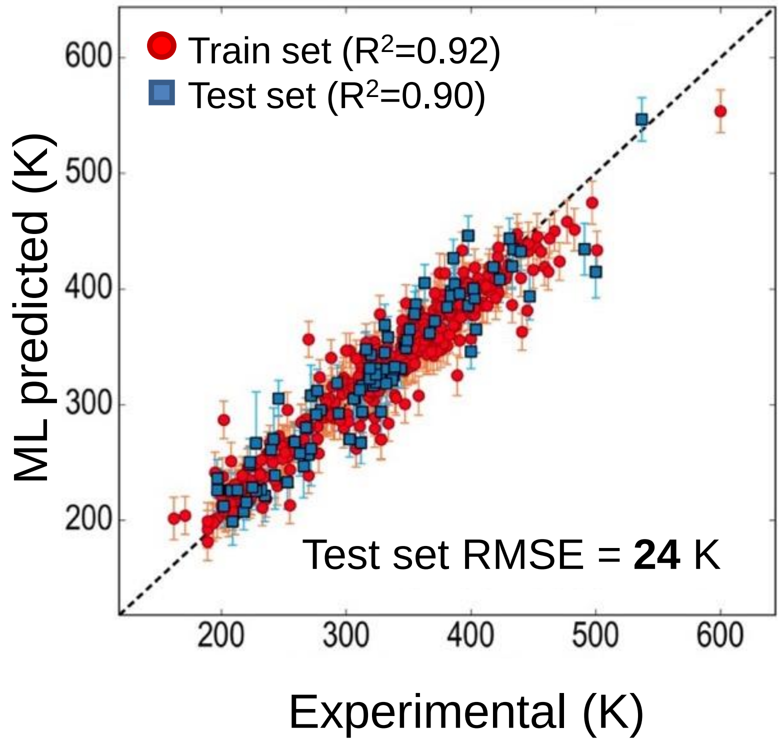
FINGERPRINT-DIMENSION REDUCTION

Glass transition temperature

All the fingerprint without RFE

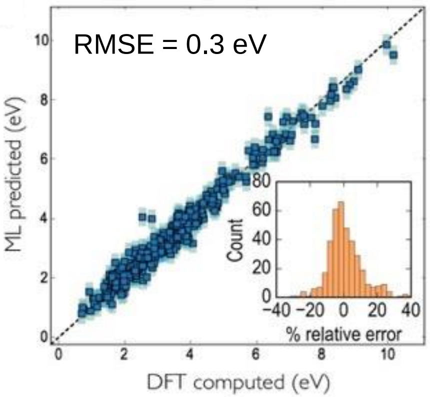


With recursive feature elimination (RFE)

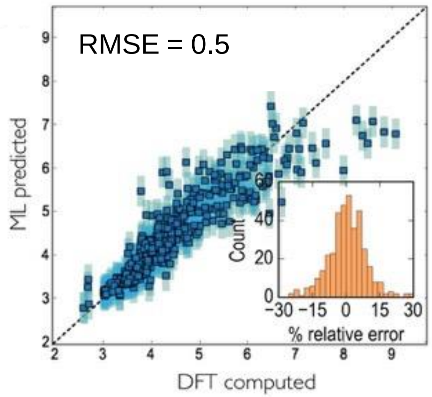


PROPERTY PREDICTION MODELS

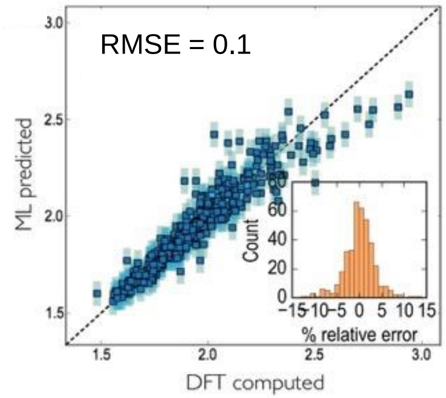
Band gap



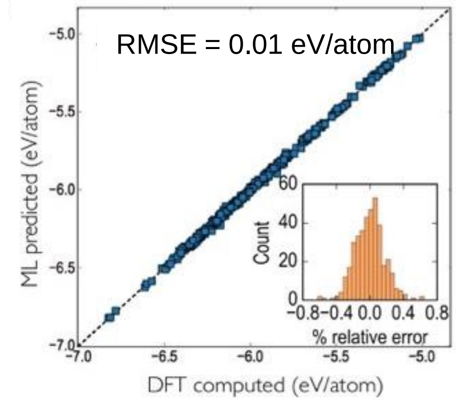
Dielectric constant



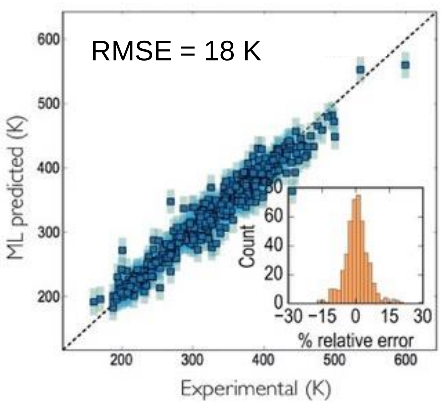
Refractive index



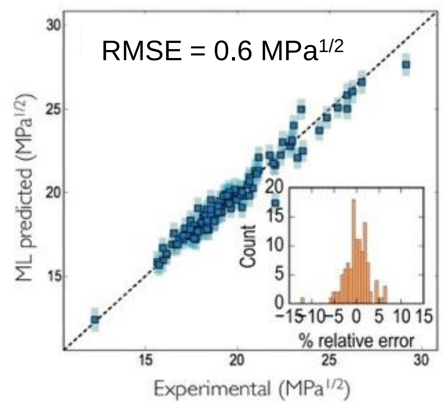
Atomization energy



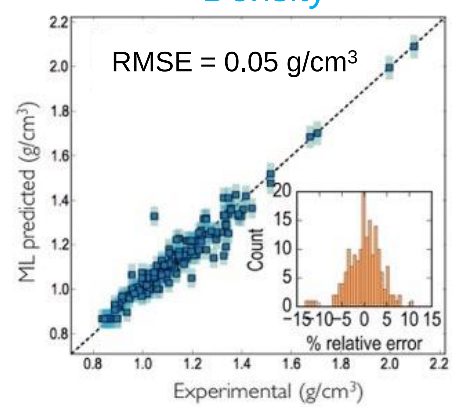
Glass transition temperature



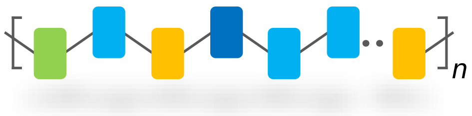
Solubility parameter



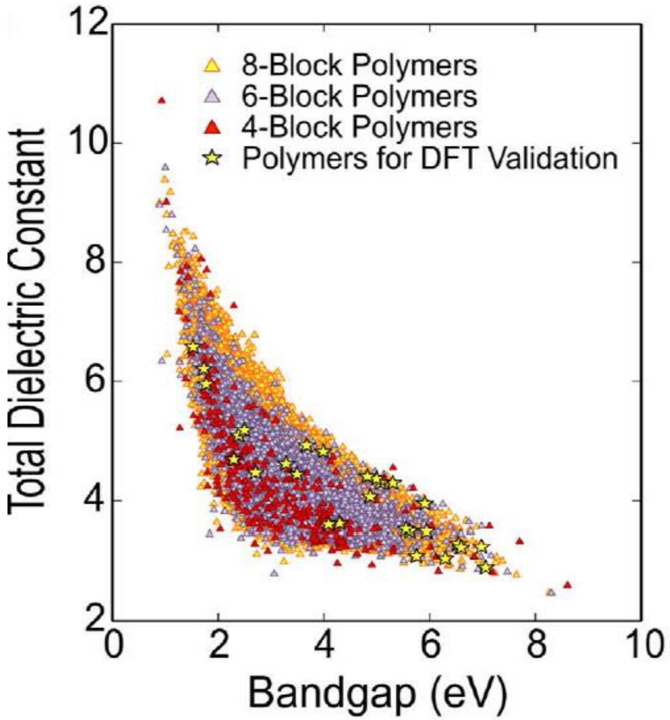
Density



OUTCOME: INSIGHTS

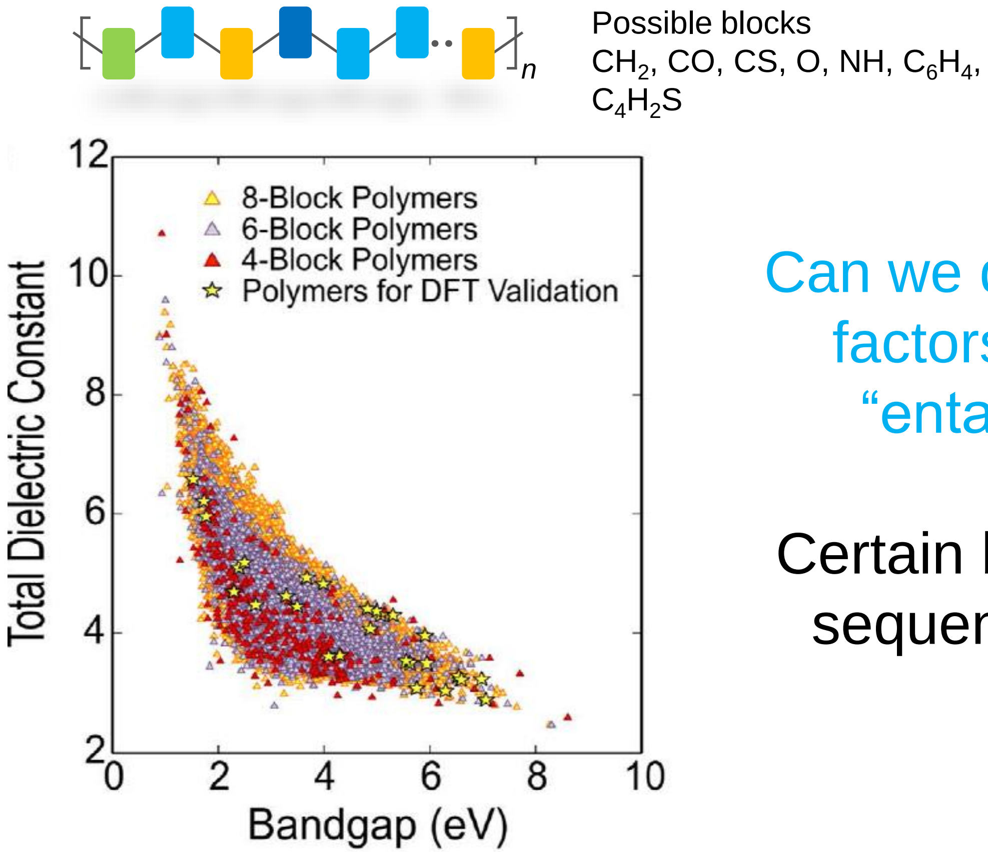


Possible blocks
CH_2, CO, CS, O, NH, C_6H_4,
C_4H_2S



"Mining materials design rules from data: The example of polymer dielectrics"
A. Mannodi-Kanakkithodi, et al, Chemistry of Materials (2017).

OUTCOME: INSIGHTS



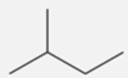
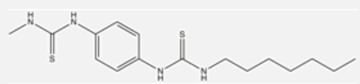
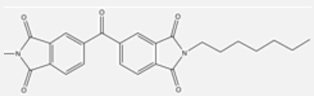
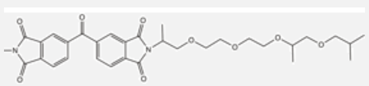
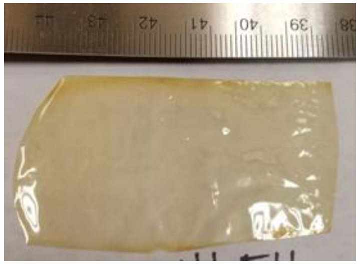
Can we determine underlying factors that govern such “entangled” behavior?

Certain blocks in a particular sequence provide optimal behavior

“Mining materials design rules from data: The example of polymer dielectrics”
A. Mannodi-Kanakkithodi, et al, Chemistry of Materials (2017).

OUTCOME: MATERIAL DISCOVERIES

Example: High energy density capacitors
New materials discovered with performance up to 3.5x of BOPP, the current standard material!

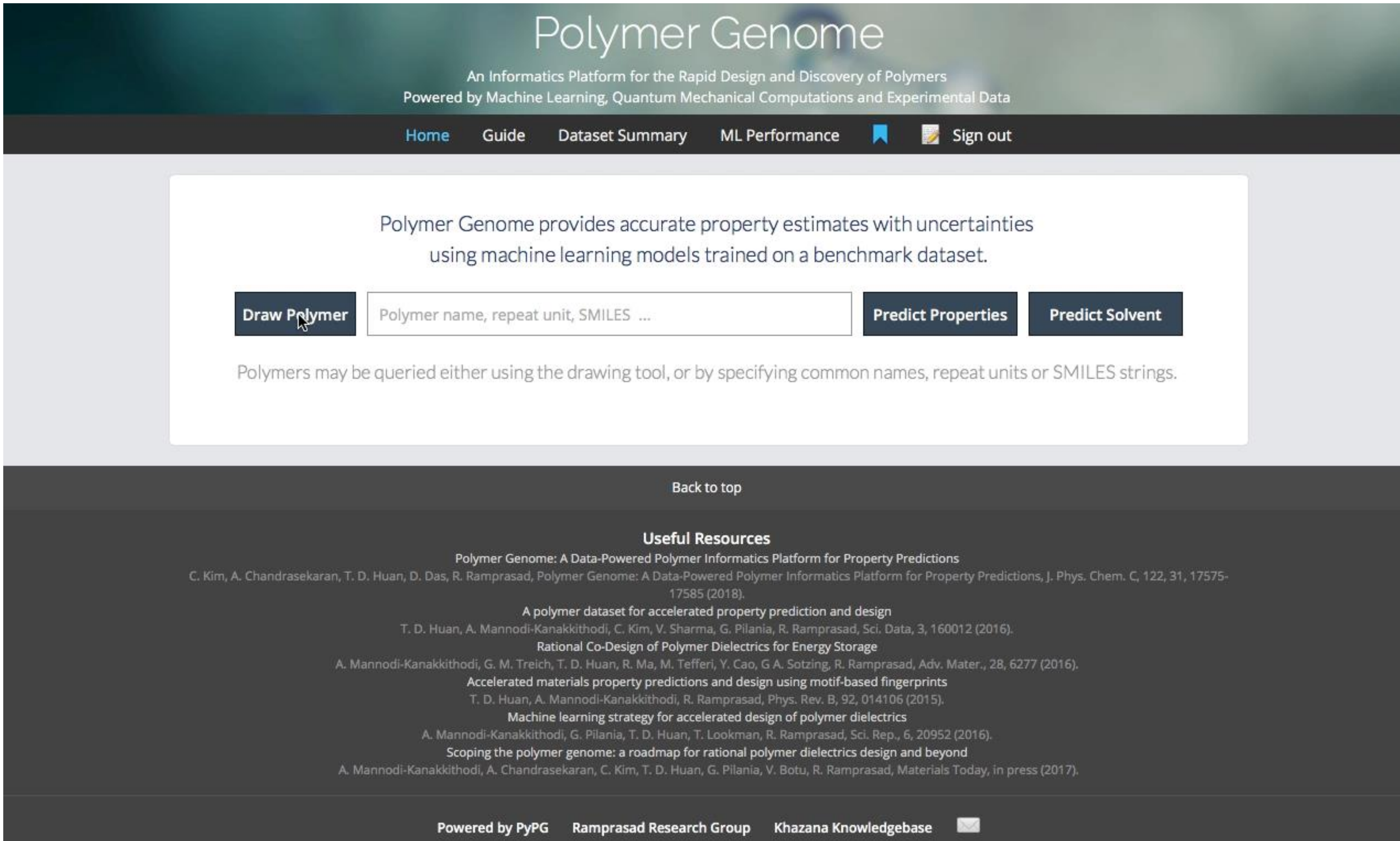
Polymer name	BOPP	PDTC-HDA (Polythiourea)	BTDA-HDA (Polyimide)	BTDA-HK511 (Polyimide)
Repeat unit				
Synthesized polymer				
Energy density (J/cm^3)	5	9	10	16

"Scoping the Polymer Genome: A Roadmap for Rational Polymer Dielectrics Design and Beyond,"

A. Mannodi-Kanakkithodi, et al, Materials Today 21, 7, 785-796 (2018)

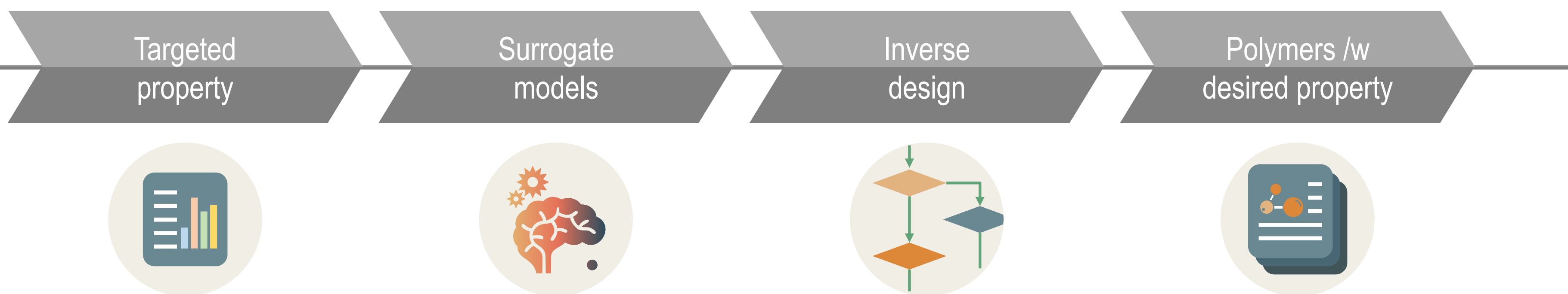
OUTCOME: ONLINE APP

https://www.polymergenome.org



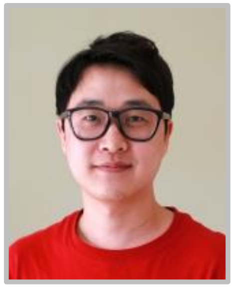
CHALLENGES AND NEXT STEPS

- Experimental data capture, and data uncertainty
- Other applications / properties
- Handling morphological complexity
- "Inverse" design (properties to polymers)



ACKNOWLEDGEMENT


Dr. Arun Mannodi Kanakkithodi (@ Argonne)


Dr. Chiho Kim

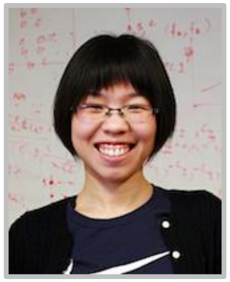

Dr. Anand Chandrasekaran


Dr. Tran Doan Huan


Dr. Deya Das (@ Mahindra)


Anurag Jha


Shruti Venkatram


Dr. Lihua Chen


Dr. Rohit Batra


Dr. Abhirup Patra


Deepak Kamal

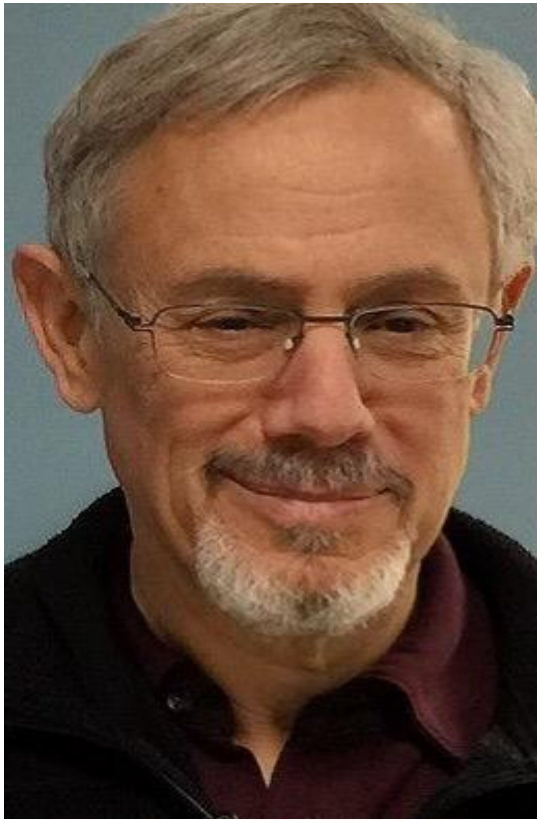

Pranav Shetty

MATHEMATICAL FRONTIERS

Machine Learning for Materials Science

Elizabeth Holm,
Carnegie Mellon University

Rampi Ramprasad,
Georgia Institute of Technology



Mark Green,
UCLA (moderator)

View webinar videos and learn more about BMSA at www.nas.edu/MathFrontiers

MATHEMATICAL FRONTIERS
2019 Monthly Webinar Series, 2-3pm ET

Feb 12: *Machine Learning for Materials Science*

Mar 12: *Mathematics of Privacy*

Apr 9: *Mathematics of Gravitational Waves*

May 14: *Algebraic Geometry*

June 11: *Mathematics of Transportation*

July 9: *Cryptography and Cybersecurity*

Aug 13: *Machine Learning in Medicine*

Sept 10: *Logic and Foundations*

Oct 8: *Mathematics of Quantum Physics*

Nov 12: *Quantum Encryption*

Dec 10: *Machine Learning for Text*

This webinar series is made possible by support for BMSA from the

National Science Foundation
Division of Mathematical Sciences

and the

Department of Energy
Advanced Scientific Computing Research

View webinar videos and learn more about BMSA at www.nas.edu/MathFrontiers

#ElevatingMath

VIDEO COMPETITION

The National Academies of SCIENCES ENGINEERING MEDICINE

nas.edu/ElevatingMath