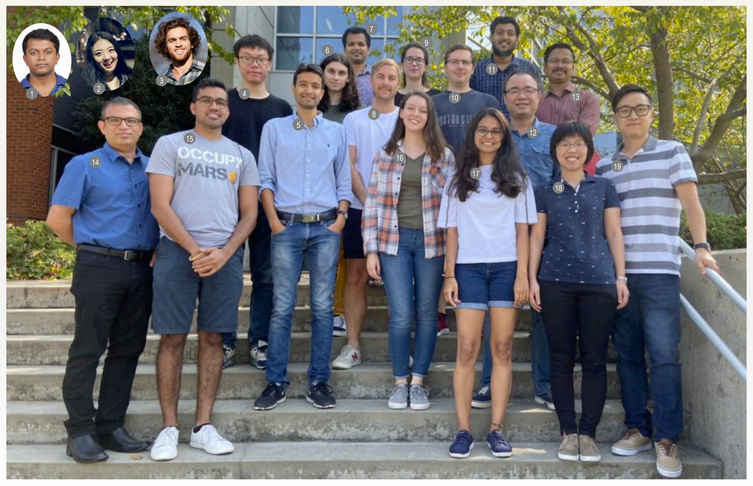
Materials Discovery & Informatics at

RAMPRASAD RESEARCH GROUP

School of Materials Science and Engineering
Georgia Institute of Technology

We develop and utilize computational and data-driven tools to aid materials design



1. Arunkumar Chitteth Rajan, 2. Yifan Liu, 3. Jordan P. Lightstone, 4. Yujie Zhu, 5. Rohit Batra (@ Argonne National Lab.), 6. Vivian Bond, 7. Pranav Shetty, 8. Christopher Kuenneth, 9. Beatriz Gonzalez del Rio, 10. James Chapman, 11. Deepak Kamal, 12. Huan Tran, 13. Harikrishna Sahu 14. Rampi Ramprasad, 15. Rishi P Gurnani, 16. Keara Frawley, 17. Shruti Venkatram, 18. Lihua Chen, 19. Chiho Kim

PI - Dr. Rampi Ramprasad (Michael E. Tennenbaum Family Chair and GRA Eminent Scholar)
E-mail : rampi.ramprasad@mse.gatech.edu / Website : http://ramprasad.mse.gatech.edu

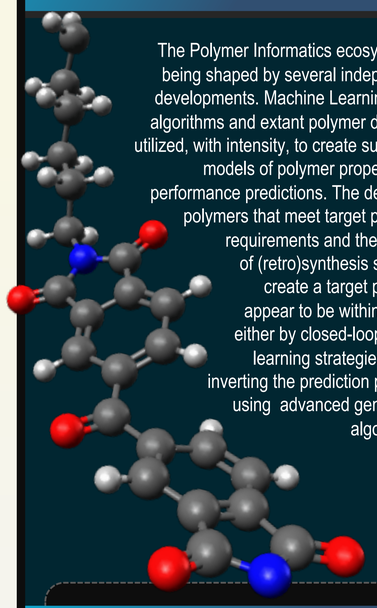
Georgia Institute of Technology · Office of Naval Research · National Science Foundation · US Department of Energy · Toyota Research Institute · Kolon Industries · Exxon Mobil

Khazana - A portal for data and tools from the Ramprasad Research Group

khazana 1.0 — https://khazana.gatech.edu

Data Repository · Polymer Genome · AGNI (DFT Emulator) · ML-Maidan

Polymer Informatics Ecosystem



The Polymer Informatics ecosystem is being shaped by several independent developments. Machine Learning (ML) algorithms and extant polymer data are utilized, with intensity, to create surrogate models of polymer property and performance predictions. The design of polymers that meet target property requirements and the design of (retro)synthesis steps to create a target polymer appear to be within reach, either by closed-loop active learning strategies, or by inverting the prediction pipeline using advanced generative algorithms.

Autonomous Computations

AGNI: ML-Assisted QM Simulations

Semi-Automatic Data Capture

Data integration & management
- Computational data
- Experimental data

Artificial intelligence
- Deep learning, regression, multi-fidelity information-fusion, multi-task learning …
- VAE, genetic algorithm active learning …

Design platform
- New polymers
- *Instant property prediction*
- Targeted property
- *Accelerated polymer design*

Synthesis / Computation
AI-guided & automated data generation
(AI for QM / AI for experiments)

Feedforward new information for adaptive data augmentation

ML-assisted expansion of polymer chemical / property space

Materials Discovery and Design

Forward Problem: Polymer → Hierarchical fingerprinting → ML prediction model → *Instant property prediction*

Inverse Problem: Target property → Variational autoencoder (V A E) / Genetic algorithm / Active learning → Polymer designs

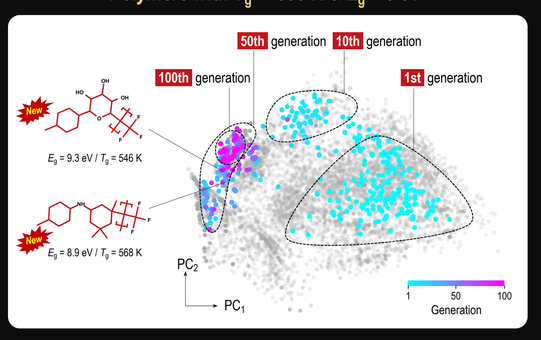
Battery electrolytes · *Membranes* · *Conducting polymers* …

Example discovery
High energy density polymer

New **PDTC-HK511**
Energy density ~16 J/cm³

Reference polymer
BOPP … Energy density ~5 J/cm³

*Example design – using genetic algorithm**
Polymers with T_g^* > 500 K & E_g^* >6 eV



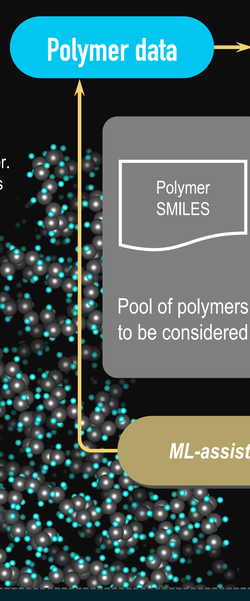
8,453 polymers projected on 2D principal component (PC) space (PC generated using the polymer fingerprints). All polymers created during 100 generations are represented by gray points. Area of polymers created at the generation #1, 10, 50, and 100 are selected to visualize the convergence in chemical diversity with evolution.

**Genetic algorithm* … A powerful method for solving materials design problems based on natural selection, the process that drives biological evolution.
*T_g … glass transition temperature), E_g … bandgap

- Rational co-design of polymer dielectrics for energy storage (Advanced Materials, 28, 6277, 2016)
- Polymers for extreme conditions designed using syntax-directed variational autoencoders (submitted, 2019)
- Active-learning and materials design: the example of high glass transition temperature polymers (MRS Comm 9, 860, 2019)

Autonomous Computations

ML-guided design of computations (and experiments) expands and diversifies the polymer data.



Functional polymers are being discovered and, at the same time, high-quality computational data is being created in a targeted and autonomous manner. Starting from the available polymer data, ML models are developed to select candidates from a big dataset of polymers, balancing between exploitation and exploration. 3D models of the candidates are then predicted, from which polymers with targeted (computational) properties are identified and data are progressively curated/updated. Requiring minimal human intervention, this autonomous workflow will be the primary source of computational data for polymers informatics.

- A multi-fidelity information-fusion approach to machine learn and predict polymer bandgap (Comput. Mater. Sci., 172, 109286 (2019)

Polymer Genome

www.polymergenome.org

User input
SMILES, name, abbreviation, sketch

General information
Polymer class, IUPAC name, similar polymers

"A Machine Learning Platform for New Polymer Discovery"

Predicted properties
Polymer properties predicted by ML models

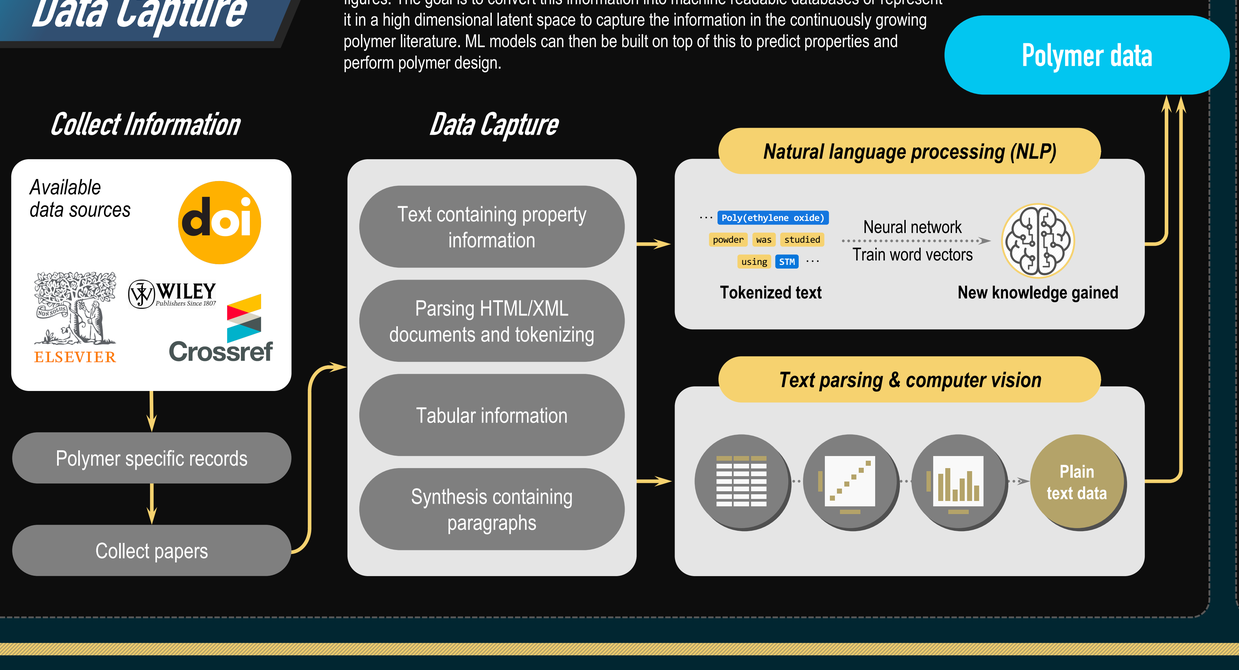
- Electronic properties - Bandgap, ionization energy, electron affinity
- Dielectric & optical properties - Dielectric constant, refractive index
- Thermal properties - Glass transition temperature, melting temperature
- Solubility properties - Solvent & non-solvent, solubility parameter
- Permeability properties - Gas permeability
- Mechanical properties - Tensile strength, Young's modulus
- Physical & thermodynamic properties - Density, atomization energy, specific heat
- Structure - 2D & 3D structure of monomer

- Polymer Genome: a data-powered polymer informatics platform for property predictions (Journal of Physical Chemistry C 122, 31, 17575-17585 2018)

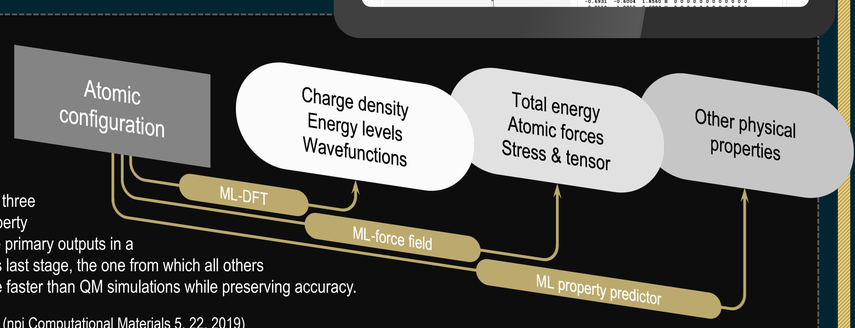
Structure visualizer
3D structure of monomer, 'mol' file

Semi-Automatic Data Capture

Natural language processing (NLP) enables automated extraction of information & data curation. Most polymer domain knowledge like polymer properties and synthesis recipes are locked up in journal papers in plain text, tables and figures. The goal is to convert this information into machine readable databases or represent it in a high dimensional latent space to capture the information in the continuously growing polymer literature. ML models can then be built on top of this to predict properties and perform polymer design.



AGNI: ML-Assisted QM Simulations



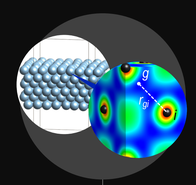
Deep learning based-QM simulations can be performed at three stages, each with a different objective in mind: 1) as a property predictor; 2) as a force-field generator; and 3) to obtain the primary outputs in a DFT calculation, namely the charge density and DOS. This last stage, the one from which all others can be reached (or derived) is several orders of magnitude faster than QM simulations while preserving accuracy.

- Solving the electronic structure problem with machine learning (npj Computational Materials 5, 22, 2019)
- General atomic neighborhood fingerprint for machine learning based methods (Journal of Physical Chemistry C 123, 15859, 2019)
- A universal strategy for the creation of machine learning-based atomistic force fields (npj Computational Materials 3, 27, 2017)
- Learning scheme to predict atomic forces and accelerate materials simulations (Physical Review B 92, 094306 2015)

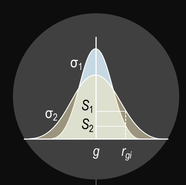
Ultrafast DFT emulation

Computational time & scaling ▶

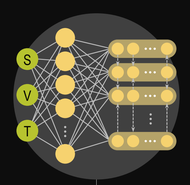
Step 1 … Data generation → *Step 2 … Fingerprinting* → *Step 3 … Machine learning* → *Step 4 … Validation*



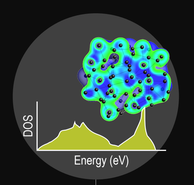
Reference atomic configurations & electronic structures



Grid point based & atom position based numerical features



Neural network



Unseen atomic configuration & electronic structures

AGNI framework for ML-QM