Accelerated Discovery of Materials for a Sustainable Future

Committee on Anticipatory Research for EPA's Research and Development Enterprise to Inform Future Environmental Protection: The Road Ahead

Applying Data Science for Environment and Health Assessment

Daniel P. Sanders, Ph.D.
Senior Manager – Materials Discovery
Future of Computing
IBM Research - Almaden

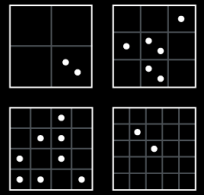
IBM **Research**

Accelerated Discovery

1,000x faster ingestion

Deep Search

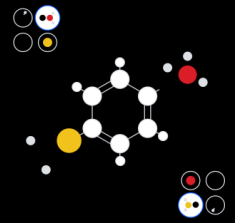
Make connections from unstructured data.



10–100x faster screening

AI-enriched Simulation

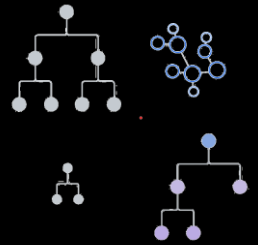
Get more value from less compute.



10x faster designs

Generative Models

Expand creativity in the molecular design process.



100x faster synthesis

Cloud-based AI-driven Autonomous Labs

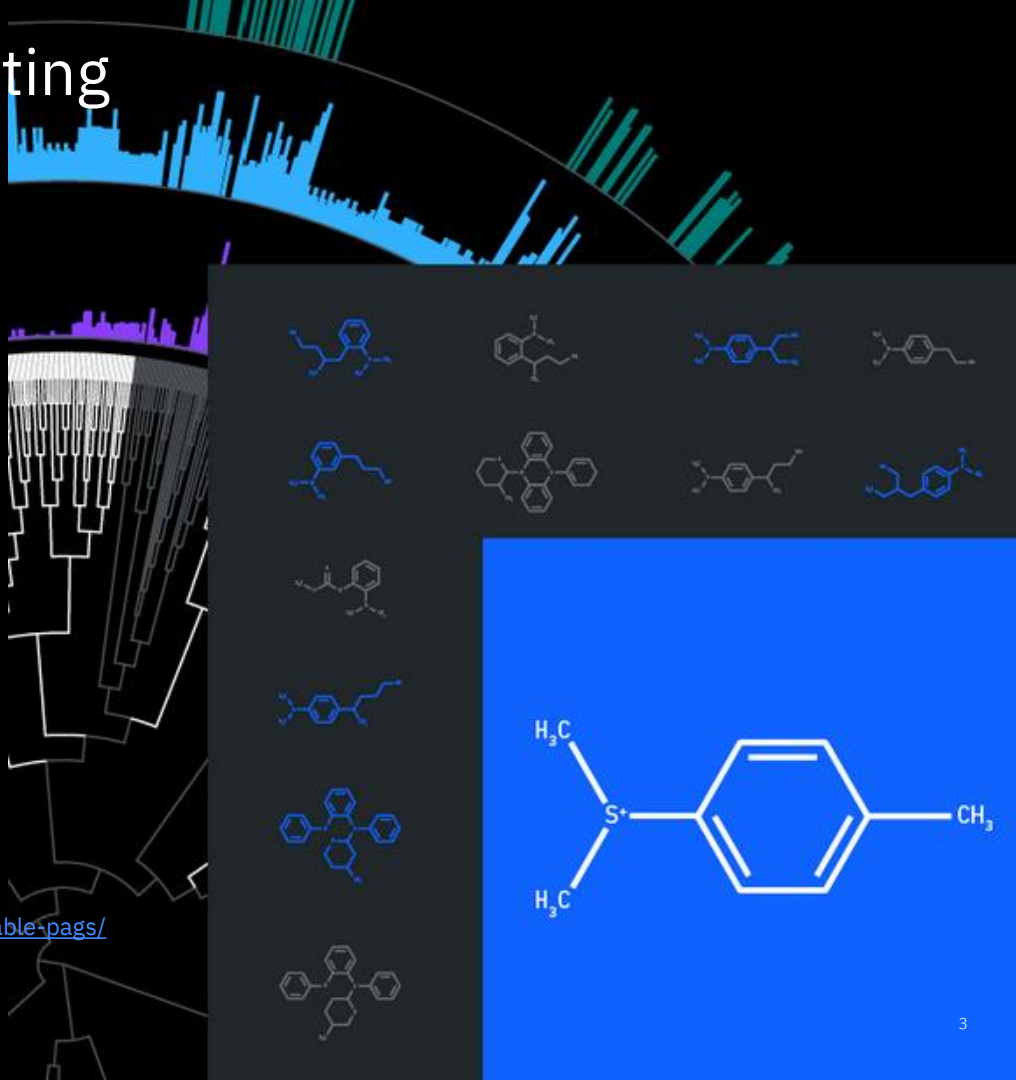
End to end material synthesis.

Can we integrate sustainability into chemical and materials discovery?

A sustainable future of computing

Sustained growth of computing platforms is contingent on sustainable development of semiconductor technology

Can we build the computational tools, technologies, platforms, and workflows necessary to allow scientists to design more sustainable materials for our industry and for the world?



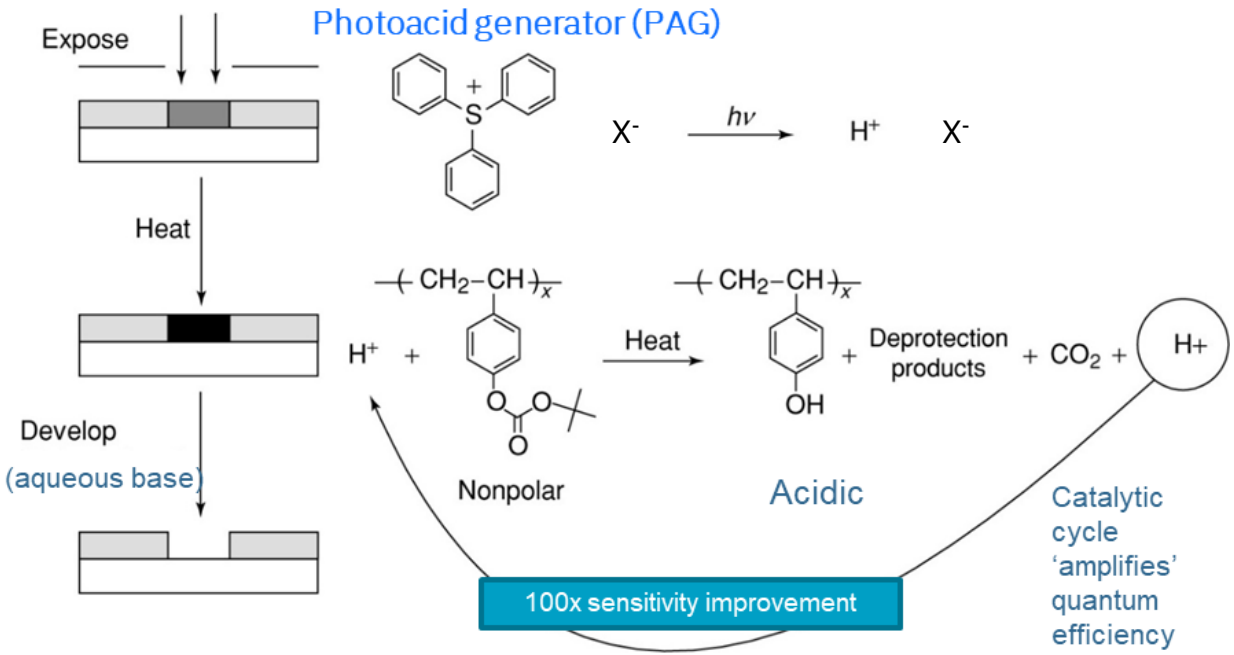
https://www.research.ibm.com/science/photoresist/
https://youtu.be/A0mvSAGMtoY
https://www.ibm.com/blogs/research/2021/02/accelerated-discovery-sustainable-pags/
https://www.research.ibm.com/5-in-5/photoresists/

Chemically Amplified Photoresists
Enabling the past, present, and future

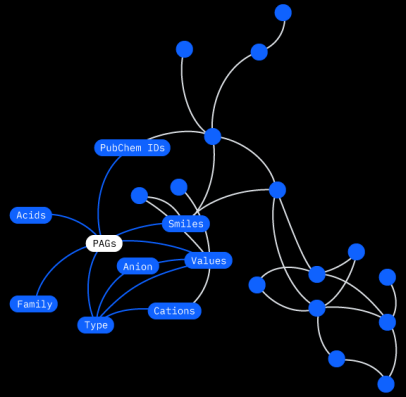
C. Grant Willson and J. Frechet
The 2020 Charles Stark Draper Prize
National Academy of Engineering

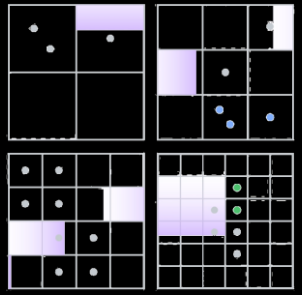
https://youtu.be/IWO1byfQctQ

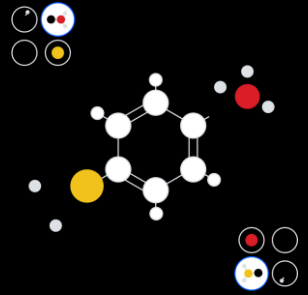
Willson, Frechet, and Ito



Project Photoresist Summary







Knowledge Graph
(P. Starr –IBM Zurich Lab)

6000 patents and articles

Knowledge Graph with 2.2M nodes and 38M edges

5000 structure dataset

Intelligent Simulation
(E. Pyzer-Knapp – IBM UK Lab)

Simulation of relevant properties

QSAR models for environmental, health, and safety predictions

Structure-property dataset sufficient to train AI models

AI Candidate Generation
(S. Takeda – IBM Tokyo Lab)

1000's of new candidates with targeted EHS properties produced by generative AI model

AI methods to intelligently screen and prioritize for *in silico* or experimental evaluation.

Automated synthesis
(T. Laino – IBM Zurich Lab)

AI retrosynthesis to identify optional reaction route

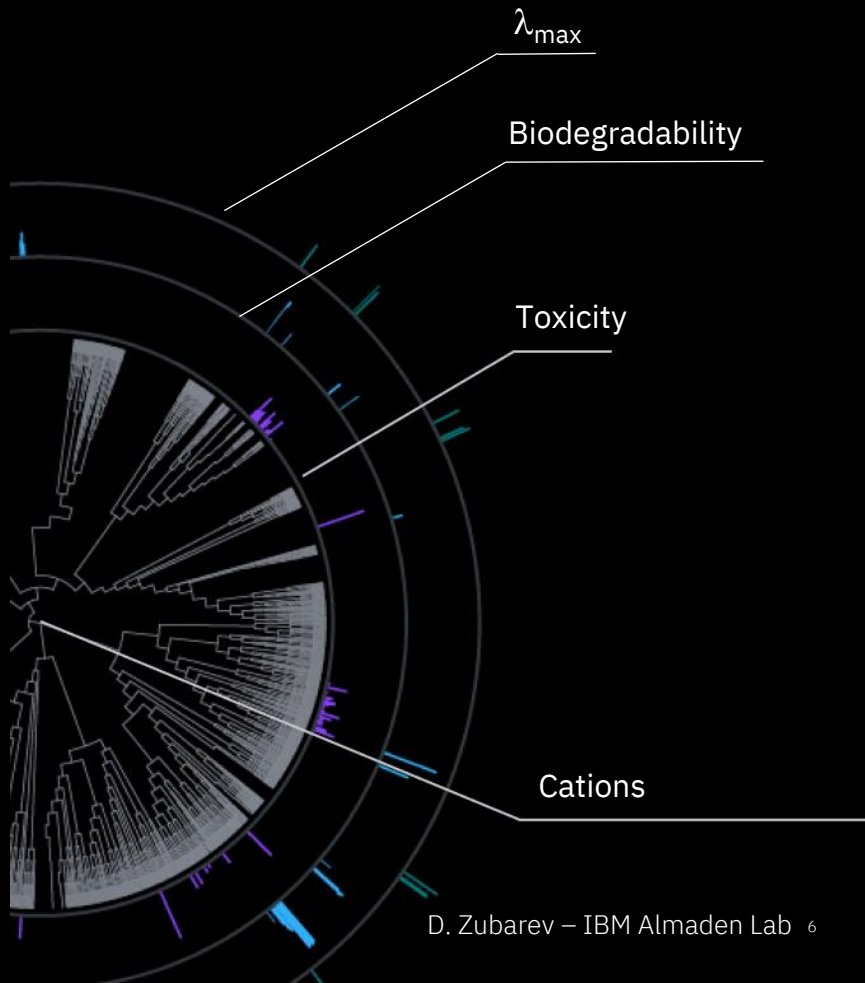
Forward reaction generation

Automated synthesis by robotic chemical reactor to produce and characterize desired product

Data visualization

Material structures and some sparse properties extracted via Deep Search.

Organize cations into families by chemical similarity and plot data set for visualization.



Data augmentation and AI candidate generation

Use simulation to populate important property values.

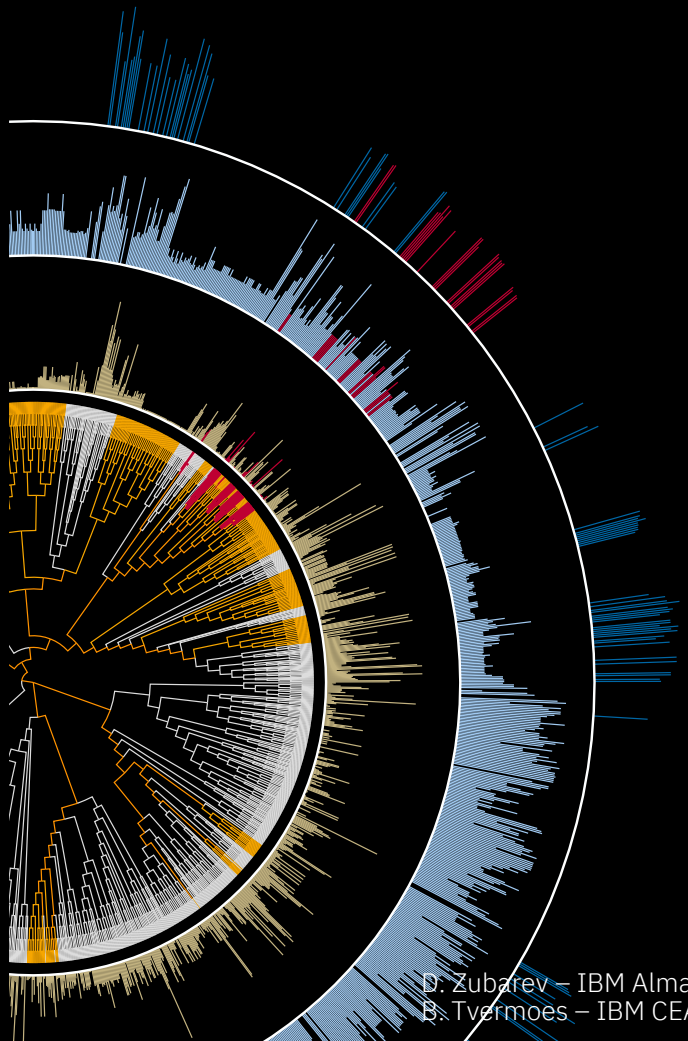
Use structure-property dataset to train generative AI model

AI model predicted several thousand new cation candidates

Downselection

Expert-in-the Loop and AI tools were used to filter, refine, and prioritize candidates for *in silico* evaluation.

Reduced number of candidates for final review by human expert by 100x.



D. Zubarev – IBM Almaden Lab
B. Tvermoes – IBM CEA

Opportunities

Communities of Discovery

Open data and open tools

New platforms for collaboration while maintaining data privacy

Informatics

Open and quantitative methods to determine:

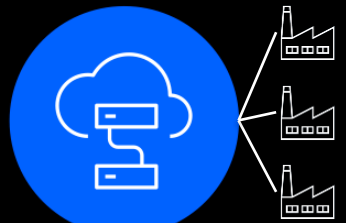
- Representativeness
- Class and family definitions
- Beyond structural similarity

AI

Rapid advances in AI models
- Explainable
- Safe
- Trustworthy

Impart super-human ability for scientists to discover

Autonomous Discovery



Automation

On-demand, distributed, remote synthesis

Standardization and Expansion

Characterization and Testing!