

# Bridging computation and experiment for energy materials discovery



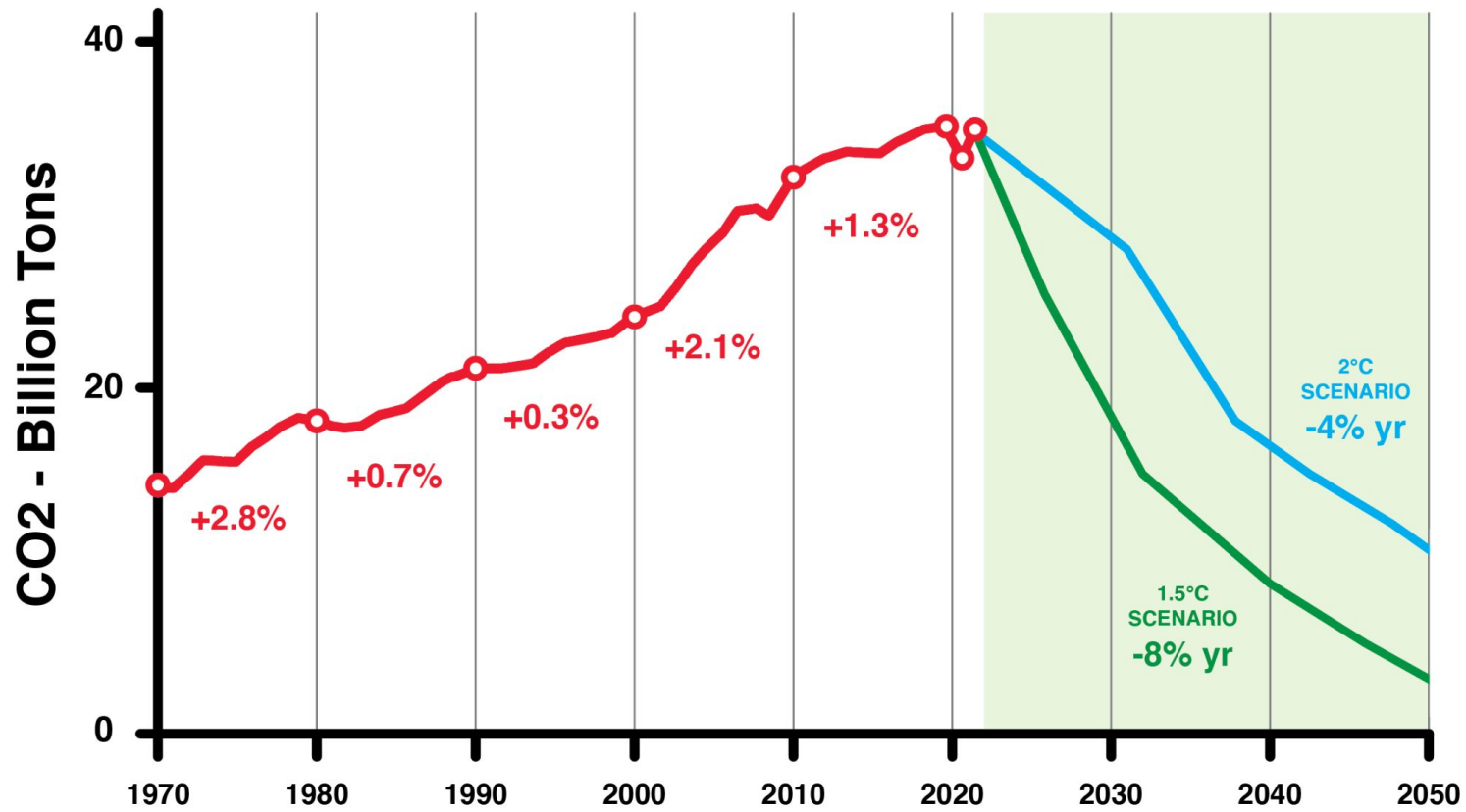
**TOYOTA**  
RESEARCH INSTITUTE

- TRI overview
- Battery lifetime
- Novel materials

Linda Hung  
Manager, Energy & Materials Division, TRI

Complex Scientific Workflows at Extreme Computational Scales  
IPAM, May 2, 2023

# Challenge - Global carbon emissions

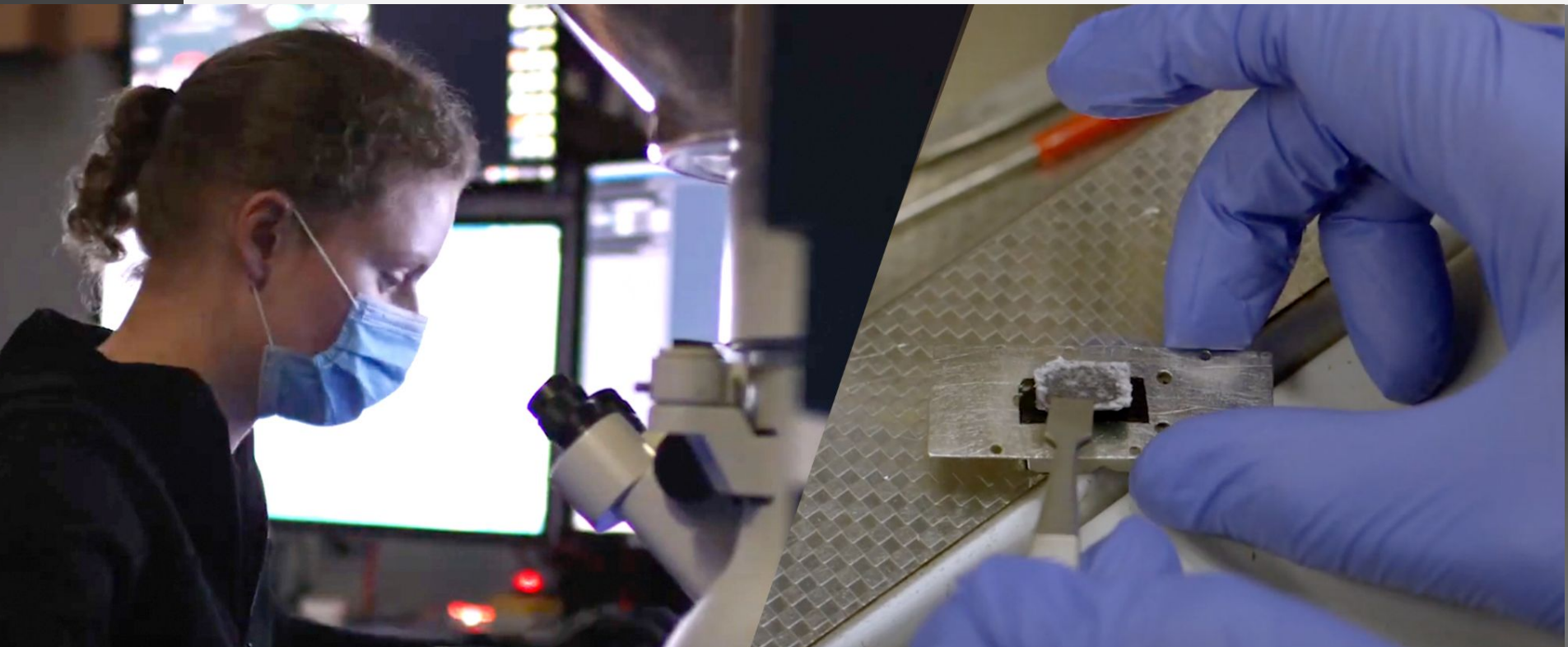


# Challenge - Resources and sustainability





# The world needs: **Better materials**







# The world needs: **More batteries**





# The world needs: **Diverse solutions**



# Energy and Materials Division

## Materials Discovery

**Better materials** are needed for breakthrough performance



## Battery Manufacturing

To make **more batteries**, we must perfect battery manufacturing.



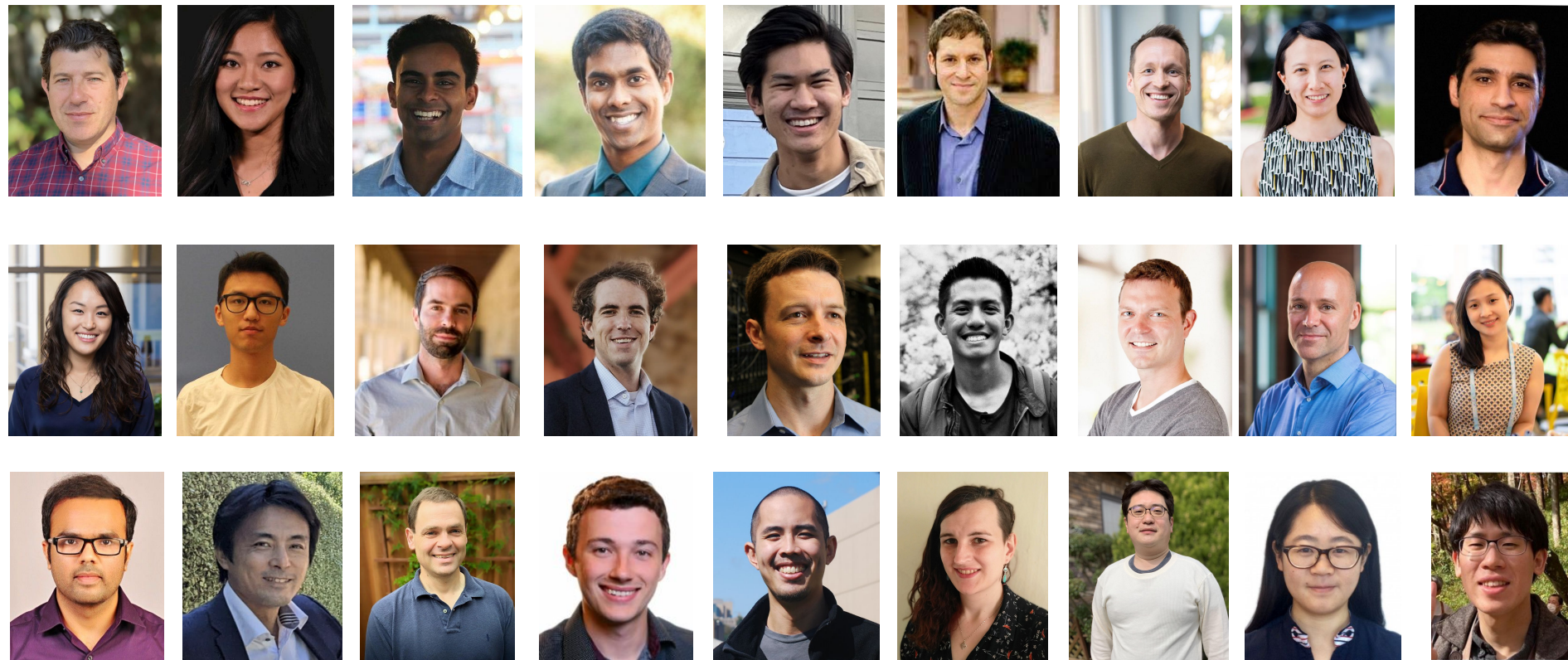
## Carbon Neutrality

Which **diverse solutions** are viable depends on economics, society, and policy.





# E&M team members







The [unclear] Center  
212 Northern Avenue



# MATERIALS DISCOVERY

# A typical project

Unique

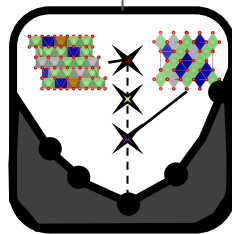
1 *Data Generating Process*



At universities, national labs,  
or internal to TRI

Tailored

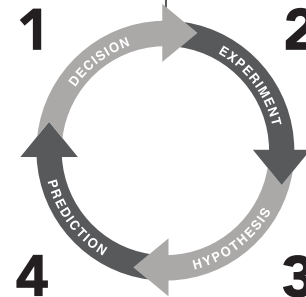
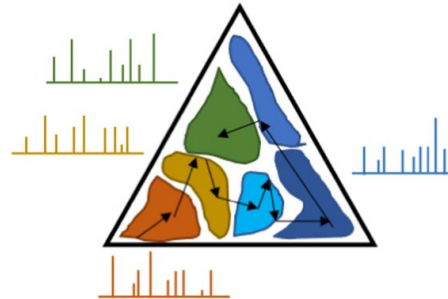
2 *Data*



AI-powered

3 *Analysis*

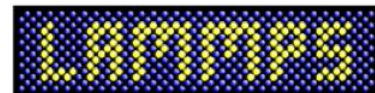
4 *Feedback*



**Applications: batteries, fuel cells, fundamental materials**



# Our tools for compute



python™



Open-Source Cheminformatics and Machine Learning



PostgreSQL



pymatgen



MATERIALS PROJECT

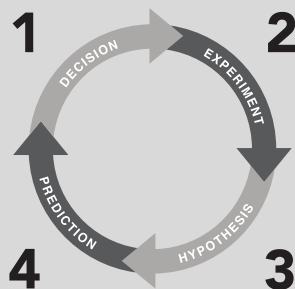


# Challenges with accelerating materials discovery

## Efficiency is too low

### Our approach

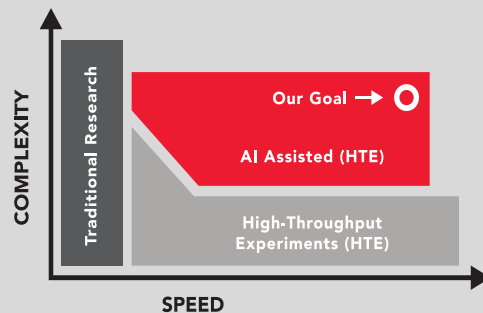
Automation & closed loop discovery.



## Experimental feedback is very difficult

### Our approach

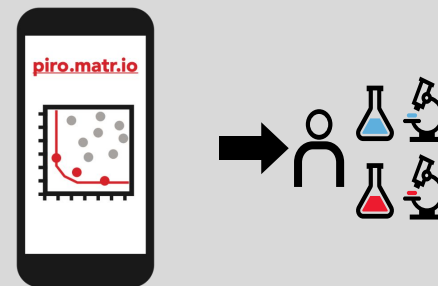
AI powered data analysis.



## Lack of theory to enable design

### Our approach

Actionable input to experiments

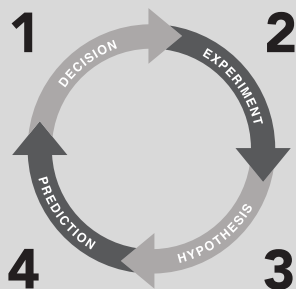


# Challenges with accelerating materials discovery

**Efficiency is too low**

Our approach

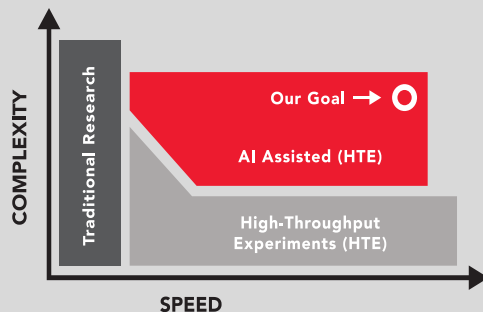
Automation & closed loop discovery.



**Experimental feedback  
is very difficult**

Our approach

AI powered data analysis.

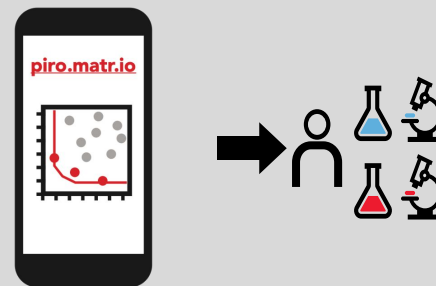


**Battery lifetime**

**Lack of theory to  
enable design**

Our approach

Actionable input to experiments



**Novel materials**



**TOYOTA**  
RESEARCH INSTITUTE



Shijing  
Sun



Amalie  
Trewartha

# Battery state of health

Data-driven and physics-driven approaches

# How long do we expect a battery to last?



Chemistry



History

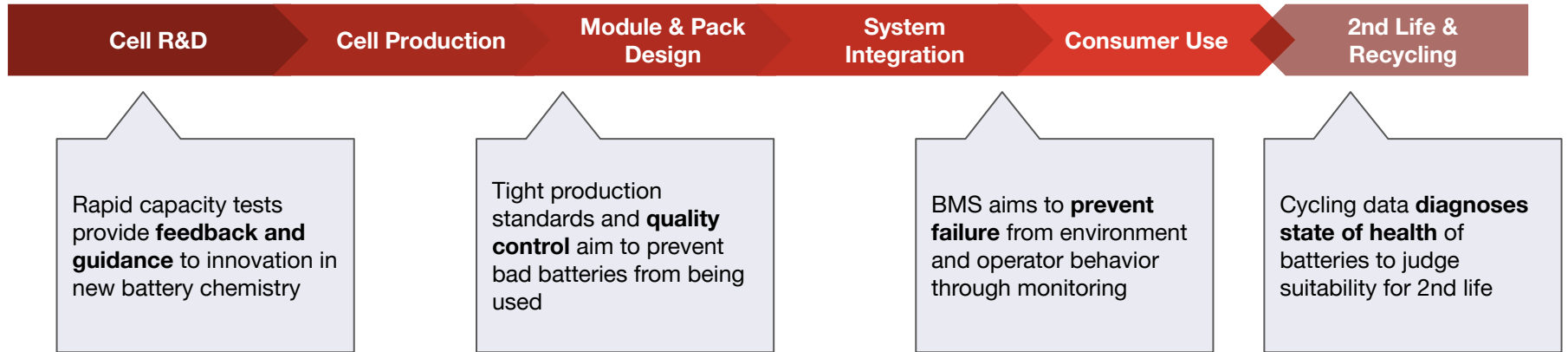


Environment

... infinite ways to charge and discharge (use), leading to **diverse** lifespans



# Electrical cycling testing provides insight at all stages of the battery life cycle



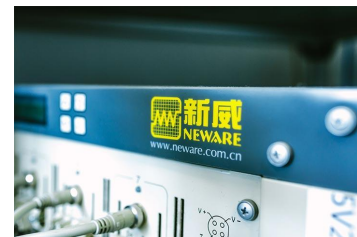
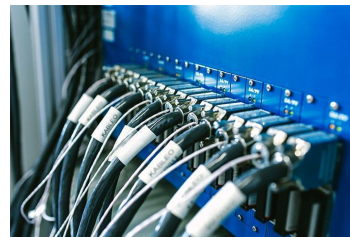
... traditional multi-factor systems optimization requires testing **until end of life**,  
This process takes **years**.

# Challenges of battery cycling data

Testing systems are designed for *low-throughput*, *manual* testing

- Each testing system has its own format
- Metadata is vitally important but not stored
- They may not distinguish cycle types - e.g. Diagnostics vs accelerated aging
- Time-series is not necessarily the most helpful format for every problem
- Traditional battery cycling takes **months per battery**
- Feedback and optimization takes **years**

	test_time	date_time	step_time	step_index	cycle_index	current	voltage	charge_capacity	c
1	0.000100	1.494646e+09	-1780.003052	0	0	0.0	3.301788	0.0	
2	9.998300	1.494646e+09	-1770.004883	0	0	0.0	3.301751	0.0	
3	20.002001	1.494646e+09	-1760.001221	0	0	0.0	3.301784	0.0	
4	30.000601	1.494646e+09	-1750.002563	0	0	0.0	3.301805	0.0	
5	30.000700	1.494646e+09	-1750.002441	0	0	0.0	3.301805	0.0	
6	40.001900	1.494646e+09	-1740.001343	0	0	0.0	3.301790	0.0	



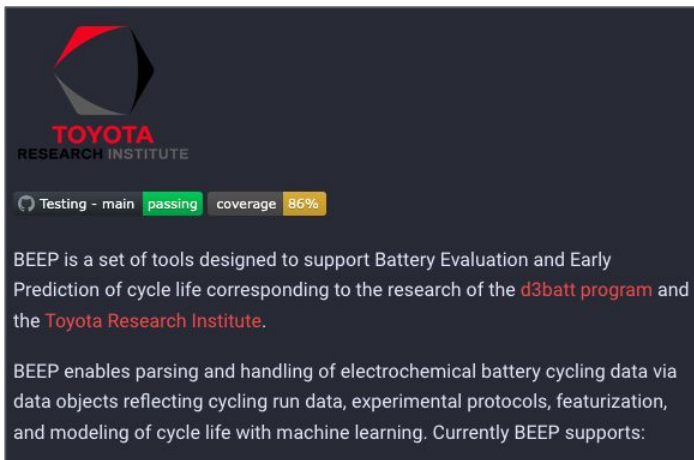
# BEEP: Battery Evaluation and Early Prediction

```
pip install beep
```

<https://github.com/TRI-AMDD/beep>

Support multiple cyclers manufacturers including:

- Maccor
- Arbin
- BioLogic
- Neware
- Indigo
- Novonix (NEW)

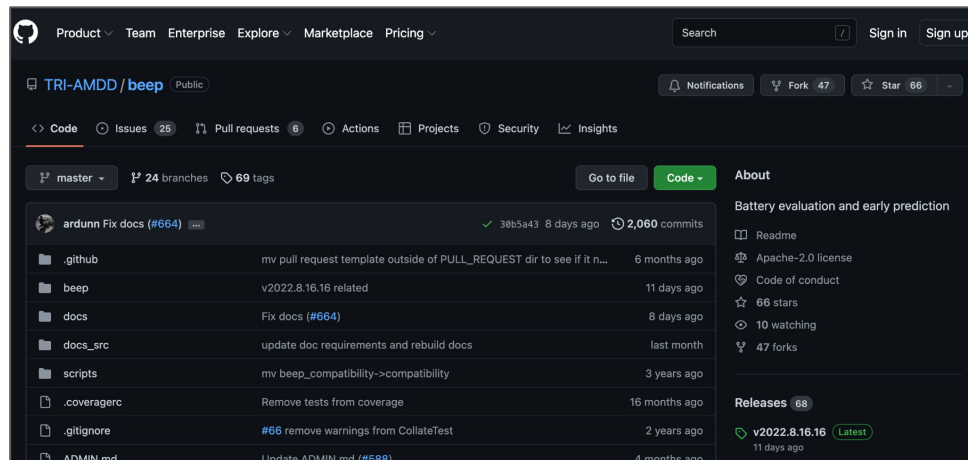


**TOYOTA RESEARCH INSTITUTE**

Testing - main passing coverage 86%

BEEP is a set of tools designed to support Battery Evaluation and Early Prediction of cycle life corresponding to the research of the **d3batt program** and the **Toyota Research Institute**.

BEEP enables parsing and handling of electrochemical battery cycling data via data objects reflecting cycling run data, experimental protocols, featurization, and modeling of cycle life with machine learning. Currently BEEP supports:



Product Team Enterprise Explore Marketplace Pricing

Search Sign in Sign up

TRI-AMDD / beep Public

Code Issues (25) Pull requests (6) Actions Projects Security Insights

master 24 branches 69 tags

Go to file Code

About

Battery evaluation and early prediction

- Readme
- Apache-2.0 license
- Code of conduct
- 66 stars
- 10 watching
- 47 forks

Releases (68)

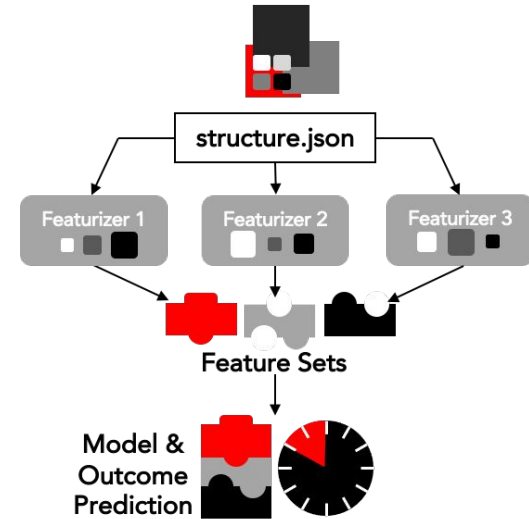
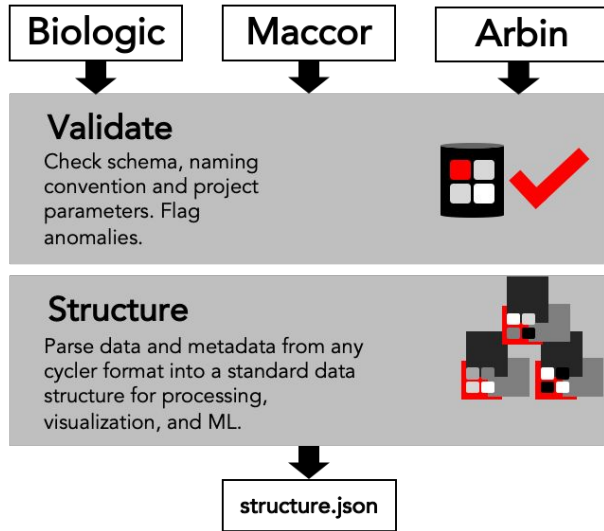
v2022.8.16.16 Latest 11 days ago

File	Commit	Time
.github	mv pull request template outside of PULL_REQUEST dir to see if it n...	6 months ago
beep	v2022.8.16.16 related	11 days ago
docs	Fix docs (#664)	8 days ago
docs_src	update doc requirements and rebuild docs	last month
scripts	mv beep_compatibility->compatibility	3 years ago
.coveragerc	Remove tests from coverage	16 months ago
.gitignore	#66 remove warnings from CollateTest	2 years ago
ADMIN.md	Update ADMIN.md (#588)	4 months ago

“BEEP: A Python Library for Battery Evaluation and Early Prediction,” P.K. Herring *et al.*, SoftwareX **11**, 100506 (2020).

# BEEP makes battery cycling data ML ready

*Highlight #1 - structuring:* process data from multiple hardware vendors into one common data structure.

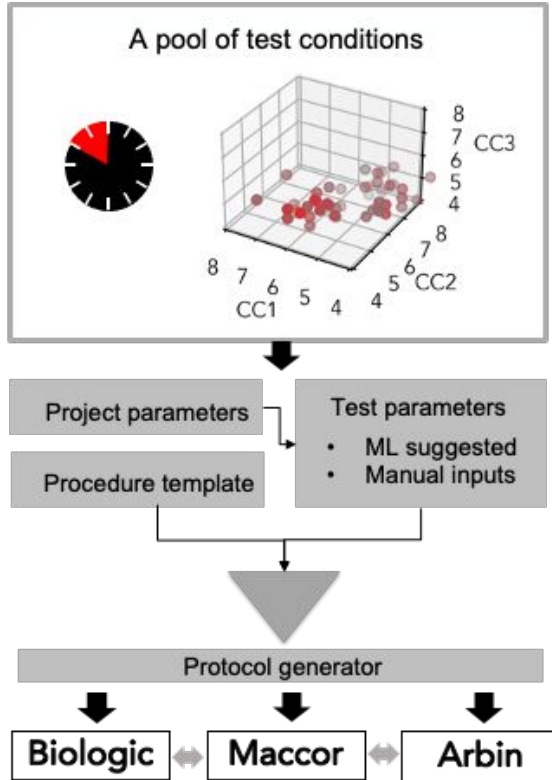


*Highlight # 2 – feature extraction:* transform structured data into ML-ready feature objects, alongside corresponding metadata.

“BEEP: A Python Library for Battery Evaluation and Early Prediction,” P.K. Herring *et al.*, *SoftwareX* 11, 100506 (2020).



# BEEP provides programmatic scheduling of testing



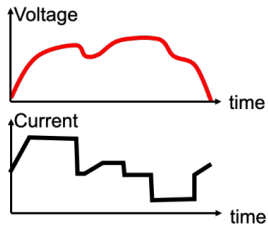
## Highlight #3 - scheduling:

- Communicate with hardware to launch high-throughput experiments
- Reduce manual errors and expedite scheduling
- Protocols convertible between cyclers

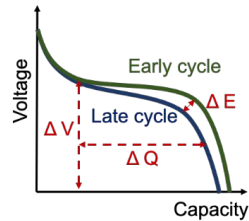
“BEEP: A Python Library for Battery Evaluation and Early Prediction,” P.K. Herring *et al.*, SoftwareX 11, 100506 (2020).

# BEEP automation enables data-driven degradation prediction

Raw Device Cycling Data



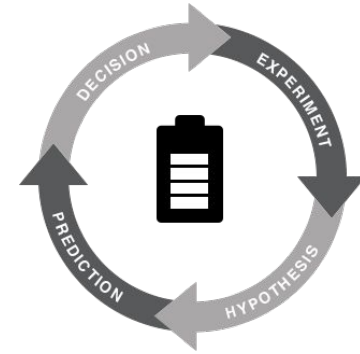
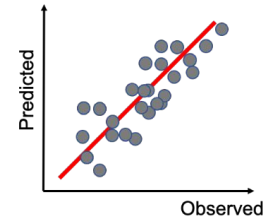
Physics-based Feature Extraction & Selection



Feature Matrix

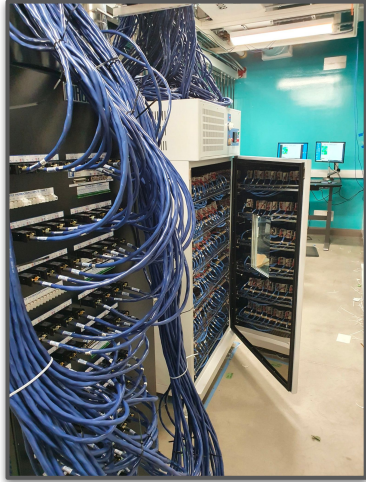
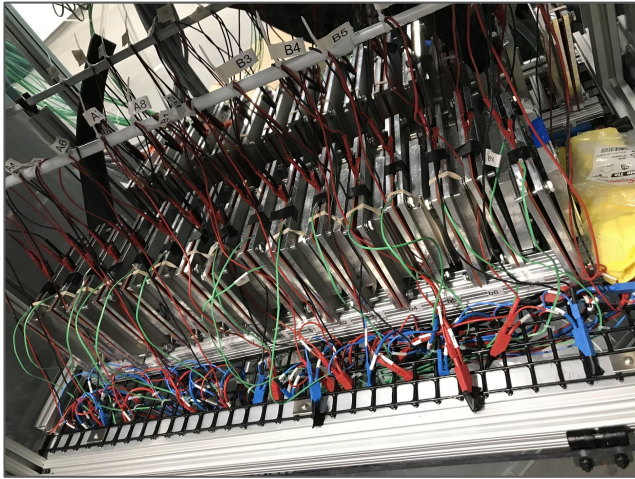
$$\begin{pmatrix} \Delta V_1, \Delta V_2, \dots, \Delta V_n \\ \Delta Q_1, \Delta Q_2, \dots, \Delta Q_n \\ \Delta E_1, \Delta E_2, \dots, \Delta E_n \end{pmatrix}$$

ML Degradation Prediction

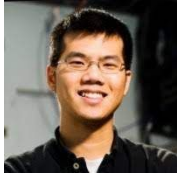


- Large, diverse cycling dataset collected and processed automatically
- Machine learning based analysis: extracting subtle signals from battery cycling data to offer **actionable, interpretable** insights into cell internal state

# High-throughput testing at SLAC

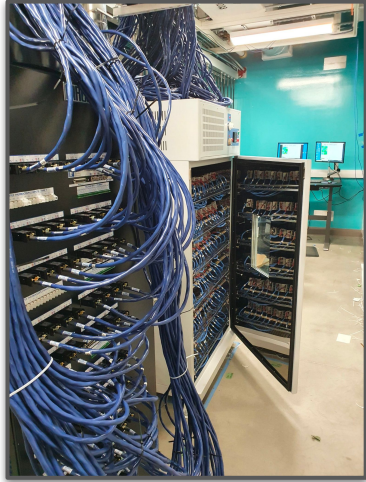
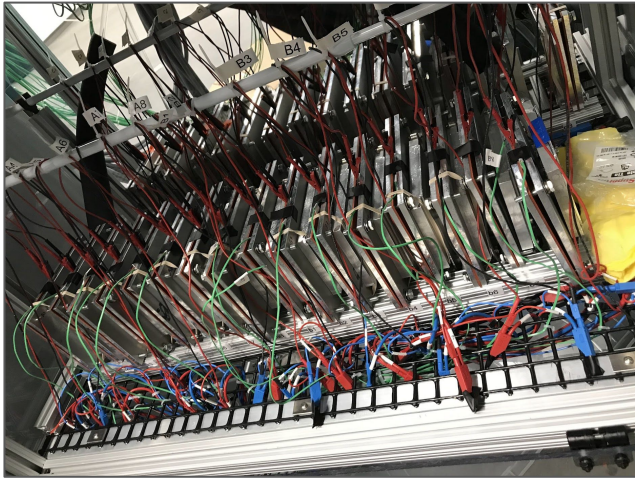


**D3BATT**  
Data-Driven Design of Li-ion Batteries

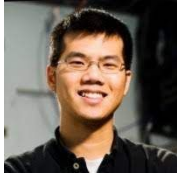


Professor Will Chueh

# High-throughput testing at SLAC



**D3BATT**  
Data-Driven Design of Li-ion Batteries



Professor Will Chueh



Airflow

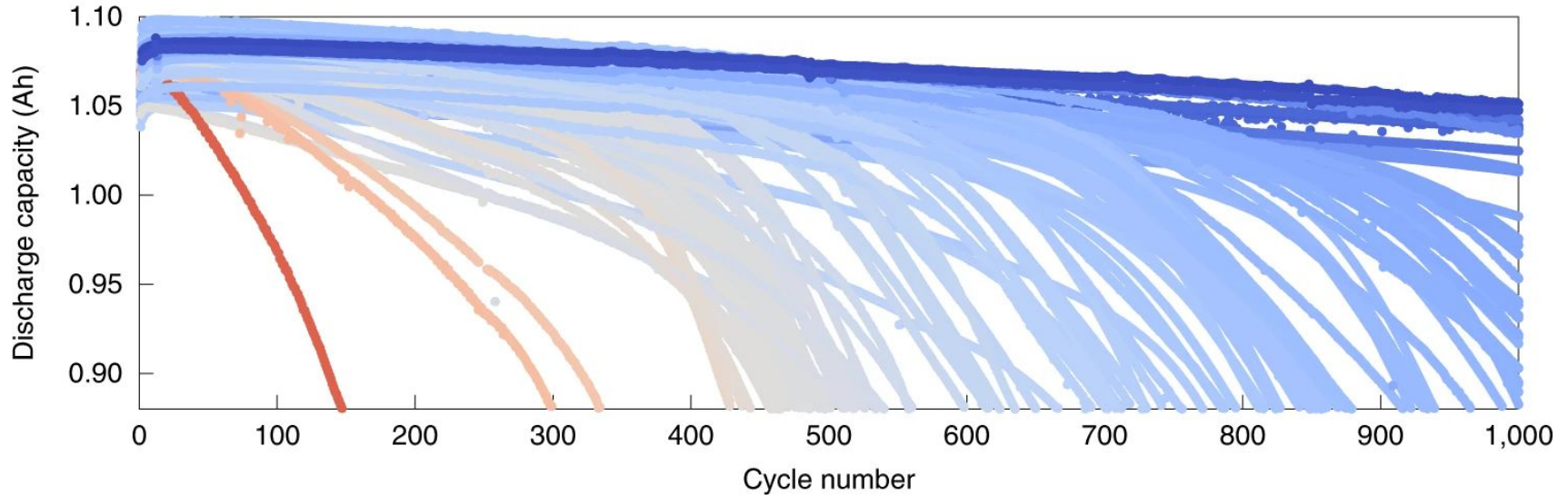


S3



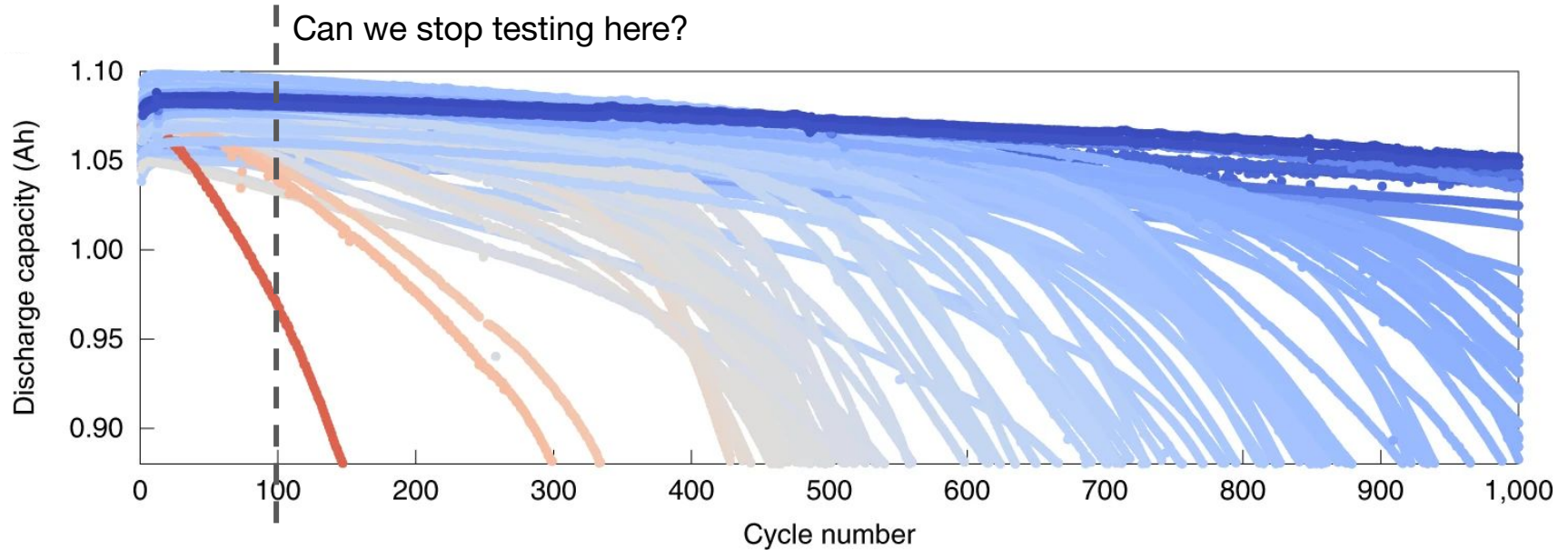
Sagemaker

# Challenge: full aging tests are **time-consuming**





# Challenge: full aging tests are **time-consuming**

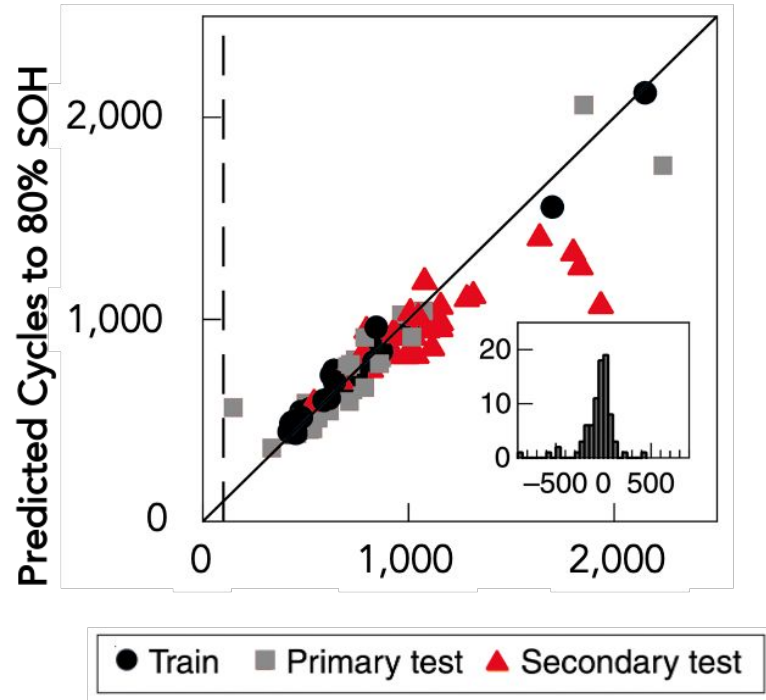


“Data-driven prediction of battery cycle life before capacity degradation,” K.A. Severson *et al.*, *Nat Energy* **4**, 383 (2019).



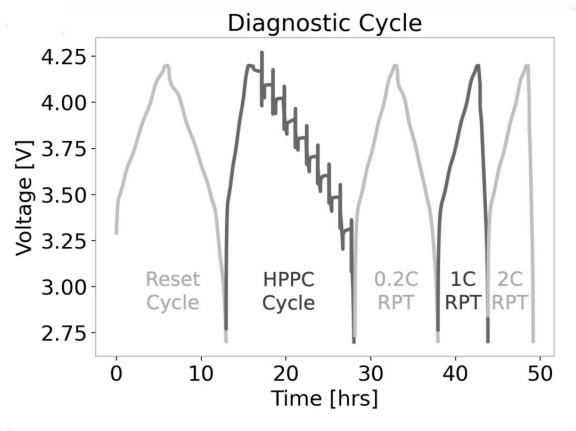
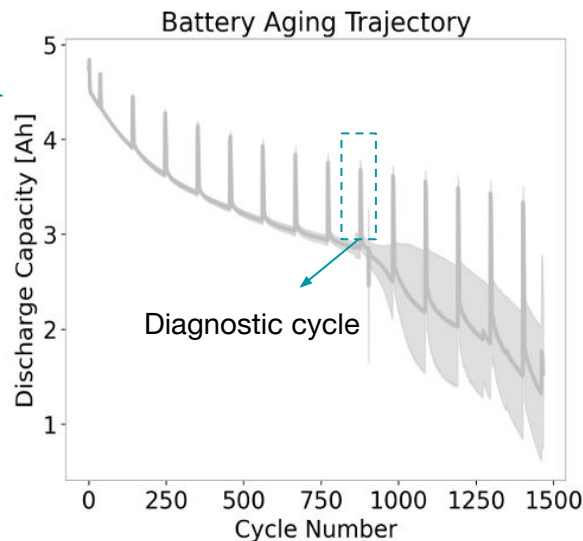
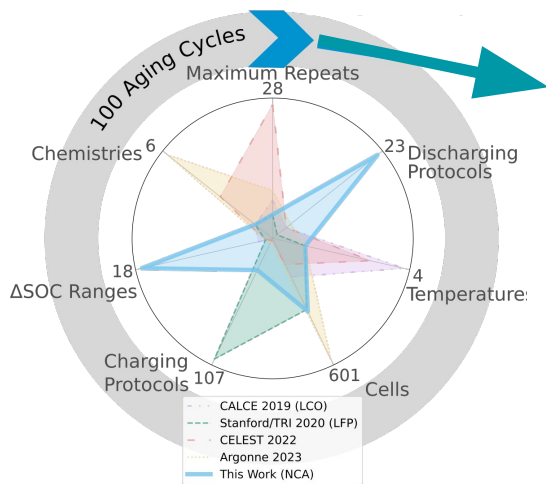
# Previously at TRI: **early-prediction** reduces test time by 10x

- Features from early cycle predict cycle life
- ML predictions made before observed decrease in capacity
- > 90% accuracy using first 100 aging cycles on LFP cells



“Data-driven prediction of battery cycle life before capacity degradation,” K.A. Severson *et al.*, *Nat Energy* 4, 383 (2019).

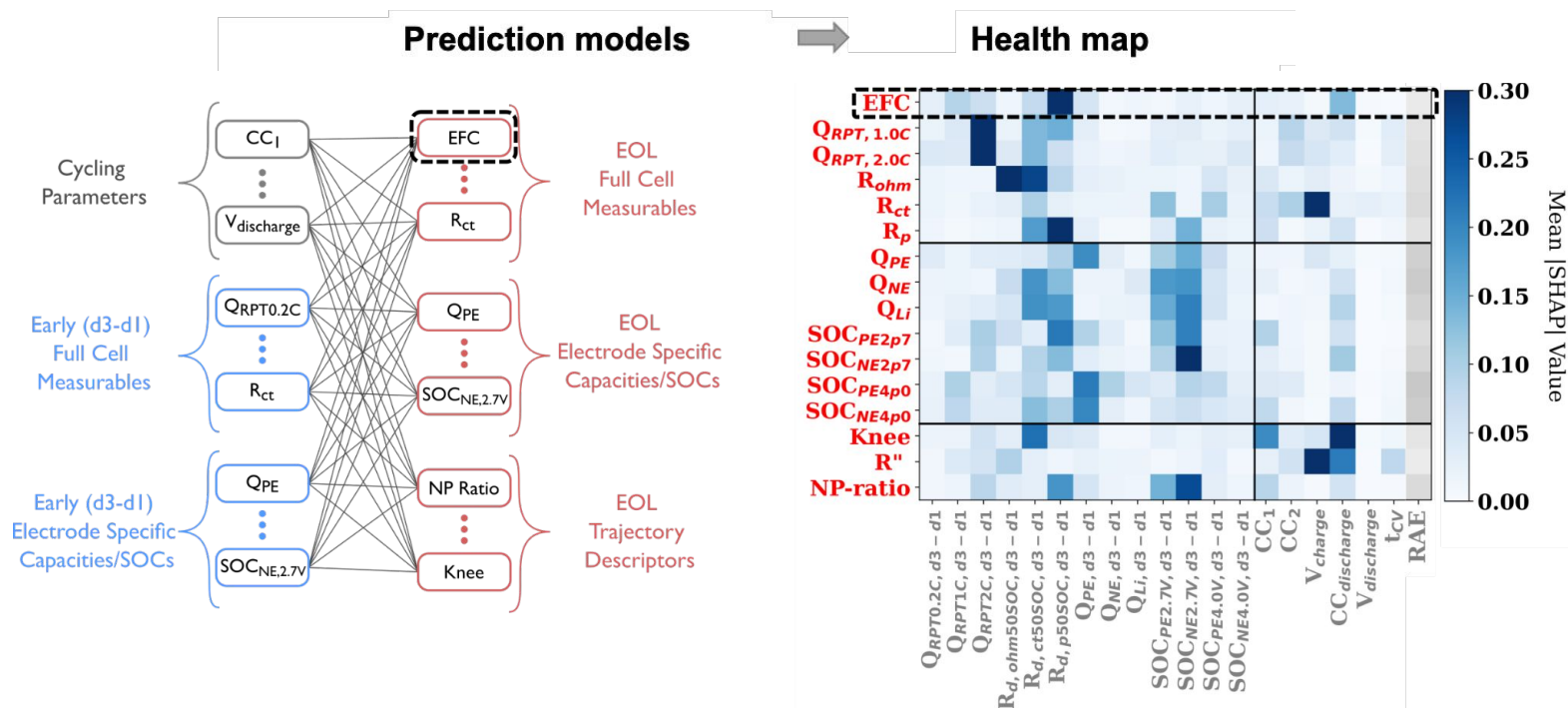
# New dataset: diverse usage with **standardized** diagnostics



- **363 cells** tested until the end of 1st lifetime on NCA/Gr + SiOx cells<sup>1</sup>
- **218 cycling conditions** with diverse discharge profiles unexplored previously
- **Standardized** diagnostics enable comprehensive health evaluation and comparison

“Interpretable Data-Driven Modeling Reveals Complexity of Battery Aging”, Bruijs van Vlijmen *et al.*, DOI: [10.26434/chemrxiv-2023-zdl2n](https://doi.org/10.26434/chemrxiv-2023-zdl2n) (2023).

# Early-prediction using physically meaningful features



Health map generated from early prediction models provide actionable insights to battery design and optimization

“Interpretable Data-Driven Modeling Reveals Complexity of Battery Aging”, Bruijs van Vlijmen *et al.*, DOI: [10.26434/chemrxiv-2023-zdl2n](https://doi.org/10.26434/chemrxiv-2023-zdl2n) (2023).

# Battery state of health - continuing challenges

## Efficiency is too low

- Early prediction for cycling
- Efficient / fast simulation

## Experimental feedback is very difficult

- Datasets: other points of battery lifecycle
- Datasets: diverse battery chemistries
- Characterization

## Lack of theory to enable design

- Models that accounting for relevant processes in physics-based models



**TOYOTA**  
RESEARCH INSTITUTE



Joey  
Montoya

# Materials discovery

From simulation to synthesis

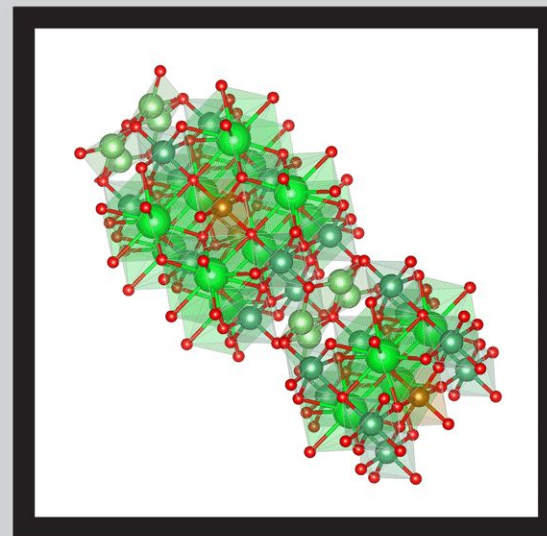
# The infinite search space

## COMPOSITION

1 H Hydrogen 1.008																	2 He Helium no data				
3 Li Lithium 0.98	4 Be Beryllium 1.57															5 B Boron 2.04	6 C Carbon 2.55	7 N Nitrogen 3.04	8 O Oxygen 2.44	9 F Fluorine 3.98	10 Ne Neon no data
11 Na Sodium 0.93	12 Mg Magnesium 1.31															13 Al Aluminum 1.61	14 Si Silicon 1.90	15 P Phosphorus 2.19	16 S Sulfur 2.58	17 Cl Chlorine 3.16	18 Ar Argon no data
19 K Potassium 0.82	20 Ca Calcium 1.00	21 Sc Scandium 1.36	22 Ti Titanium 1.54	23 V Vanadium 1.63	24 Cr Chromium 1.66	25 Mn Manganese 1.55	26 Fe Iron 1.63	27 Co Cobalt 1.88	28 Ni Nickel 1.91	29 Cu Copper 1.90	30 Zn Zinc 1.65	31 Ga Gallium 1.81	32 Ge Germanium 2.61	33 As Arsenic 2.18	34 Se Selenium 2.55	35 Br Bromine 2.96	36 Kr Krypton 3.00				
37 Rb Rubidium 0.82	38 Sr Strontium 0.95	39 Y Yttrium 1.22	40 Zr Zirconium 1.33	41 Nb Niobium 1.6	42 Mo Molybdenum 2.16	43 Tc Technetium 1.9	44 Ru Ruthenium 2.2	45 Rh Rhodium 2.28	46 Pd Palladium 2.20	47 Ag Silver 1.93	48 Cd Cadmium 1.69	49 In Indium 1.78	50 Sn Tin 1.96	51 Sb Antimony 2.05	52 Te Tellurium 2.1	53 I Iodine 2.66	54 Xe Xenon 2.6				
55 Cs Cesium 0.79	56 Ba Barium 0.69	57-71 Lanthanides	72 Hf Hafnium 1.3	73 Ta Tantalum 1.5	74 W Tungsten 2.36	75 Re Rhenium 1.9	76 Os Osmium 2.2	77 Ir Iridium 2.2	78 Pt Platinum 2.28	79 Au Gold 2.54	80 Hg Mercury 2.00	81 Tl Thallium 1.62	82 Pb Lead 2.33	83 Bi Bismuth 2.02	84 Po Polonium 2.0	85 At Astatine 2.2	86 Rn Radon no data				
87 Fr Francium 0.7	88 Ra Radium 0.69	89-103 Actinides	104 Rf Rutherfordium no data	105 Db Dubnium no data	106 Sg Seaborgium no data	107 Bh Bohrium no data	108 Hs Hassium no data	109 Mt Meitnerium no data	110 Ds Darmstadtium no data	111 Rg Roentgenium no data	112 Cn Copernicium no data	113 Nh Nihonium no data	114 Fl Flerovium no data	115 Mc Moscovium no data	116 Lv Livermorium no data	117 Ts Tennessine no data	118 Og Oganesson no data				

57 La Lanthanum 1.10	58 Ce Cerium 1.12	59 Pr Praseodymium 1.13	60 Nd Neodymium 1.14	61 Pm Promethium 1.13	62 Sm Samarium 1.17	63 Eu Europium 1.2	64 Gd Gadolinium 1.2	65 Tb Terbium 1.22	66 Dy Dysprosium 1.23	67 Ho Holmium 1.24	68 Er Erbium 1.24	69 Tm Thulium 1.25	70 Yb Ytterbium 1.1	71 Lu Lutetium 1.27
89 Ac Actinium 1.1	90 Th Thorium 1.3	91 Pa Protactinium 1.5	92 U Uranium 1.38	93 Np Neptunium 1.36	94 Pu Plutonium 1.28	95 Am Americium 1.3	96 Cm Curium 1.3	97 Bk Berkelium 1.3	98 Cf Californium 1.3	99 Es Einsteinium 1.3	100 Fm Fermium 1.3	101 Md Mendelevium 1.3	102 No Nobelium 1.3	103 Lr Lawrencium no data

## STRUCTURE





# CAMD: Computational Autonomy for Materials Discovery

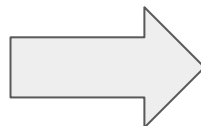
User chooses chemical system, e.g. Ag-Ba-O

Enter a hyphen-separated list of elements (e.g. "Fe-V" or "Mn-S-P")

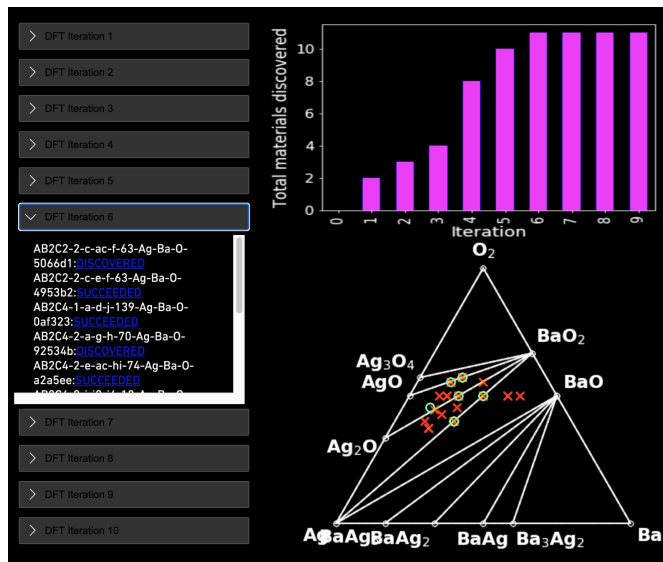
Ag-Ba-O

Check Synthesis Discovery

Start Synthesis Discovery

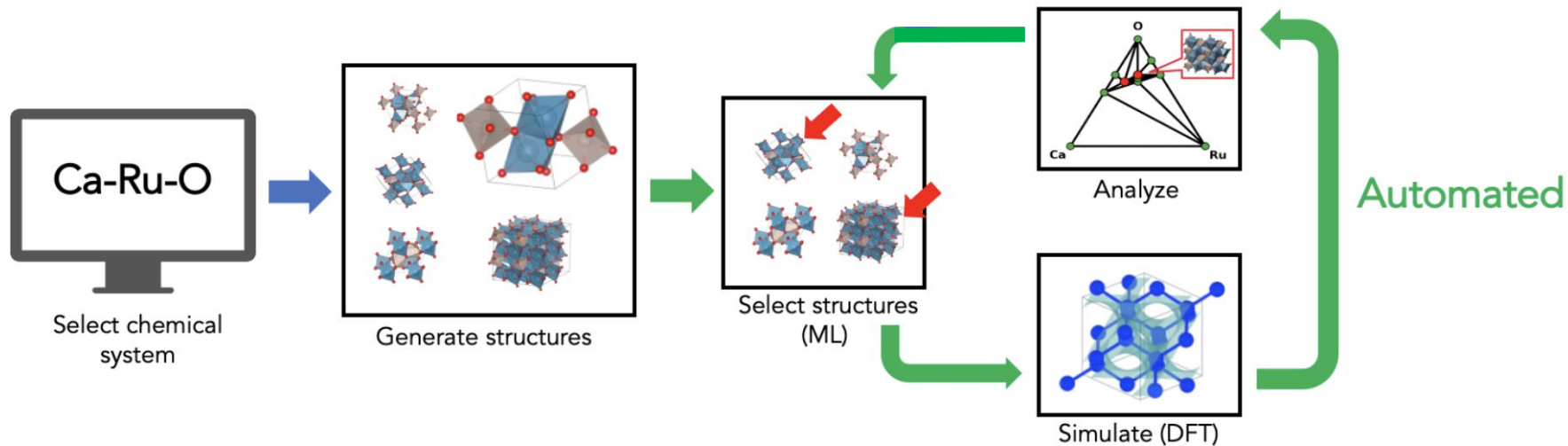


CAMD displays real-time results on frontend



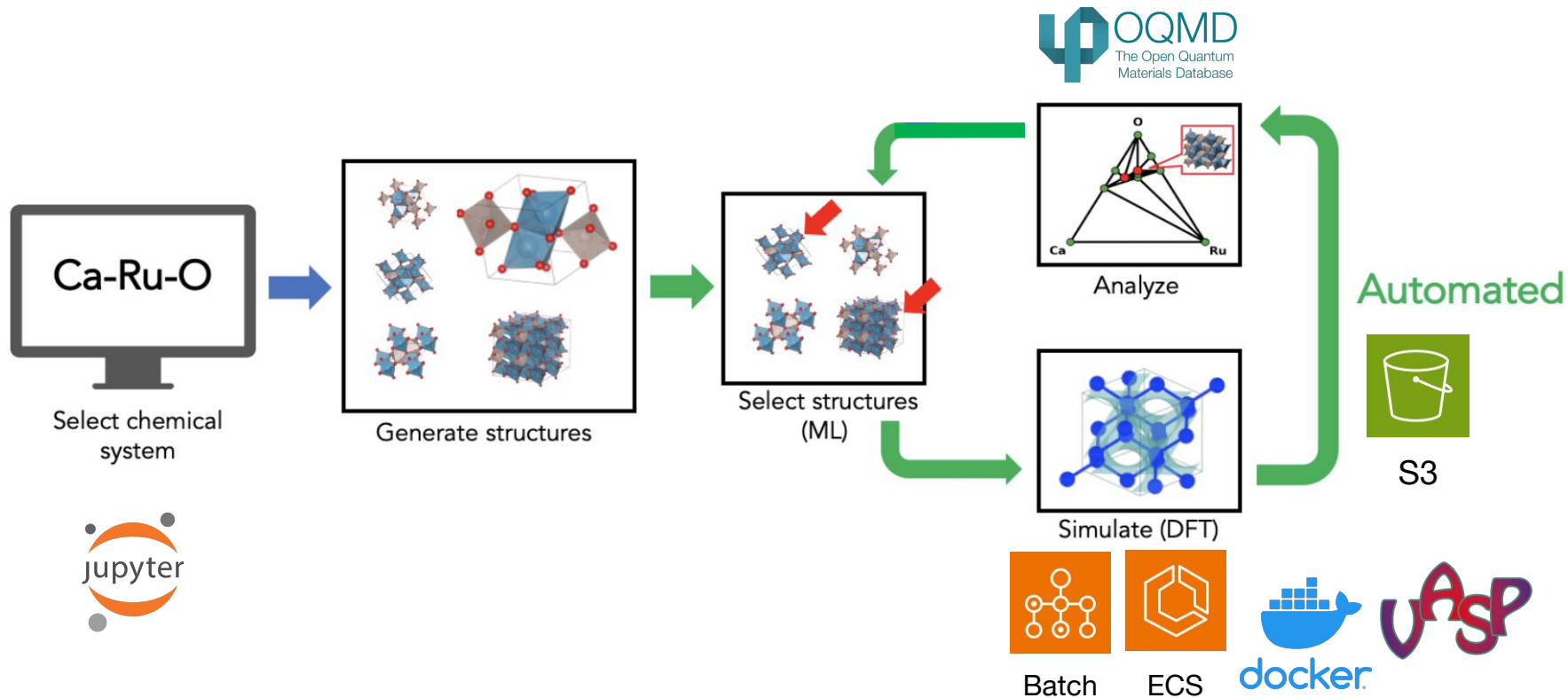
CAMD stores structures and data

# Active learning to discover materials



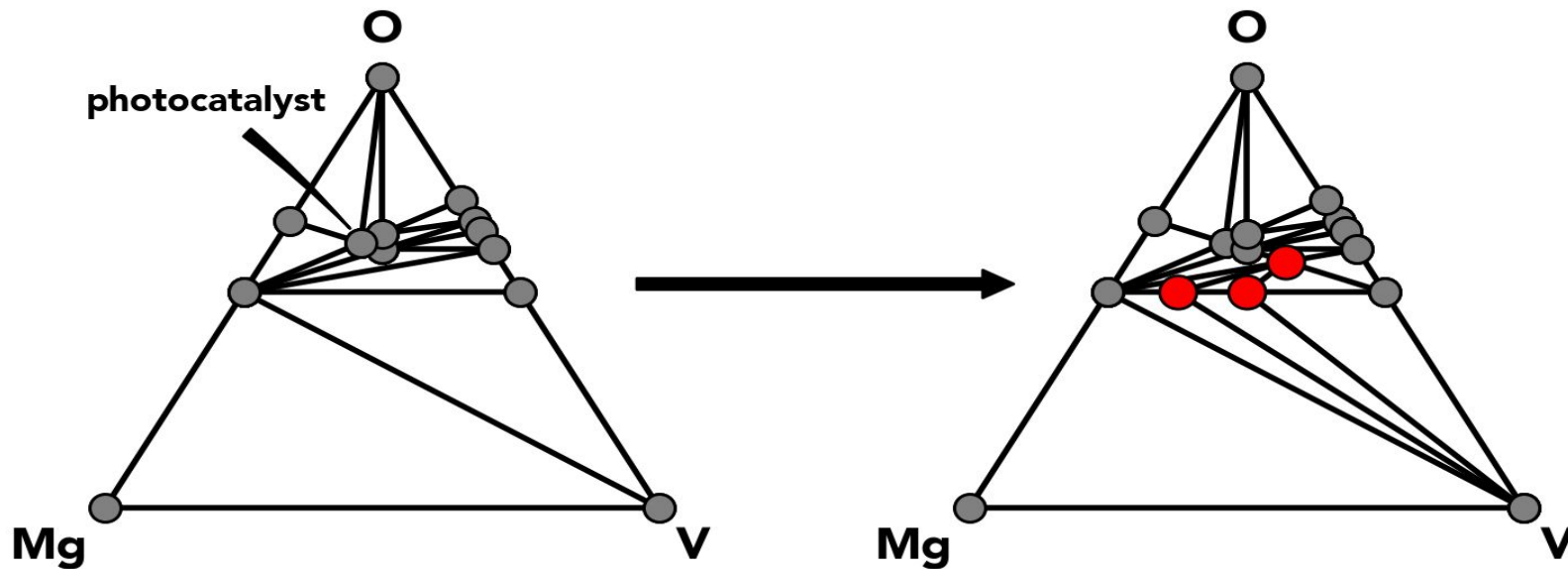
“Autonomous Intelligent Agents for Accelerated Materials Discovery” J.H. Montoya *et al.*, *Chem Sci* **11**, 8517-8532 (2020).

# Active learning to discover materials



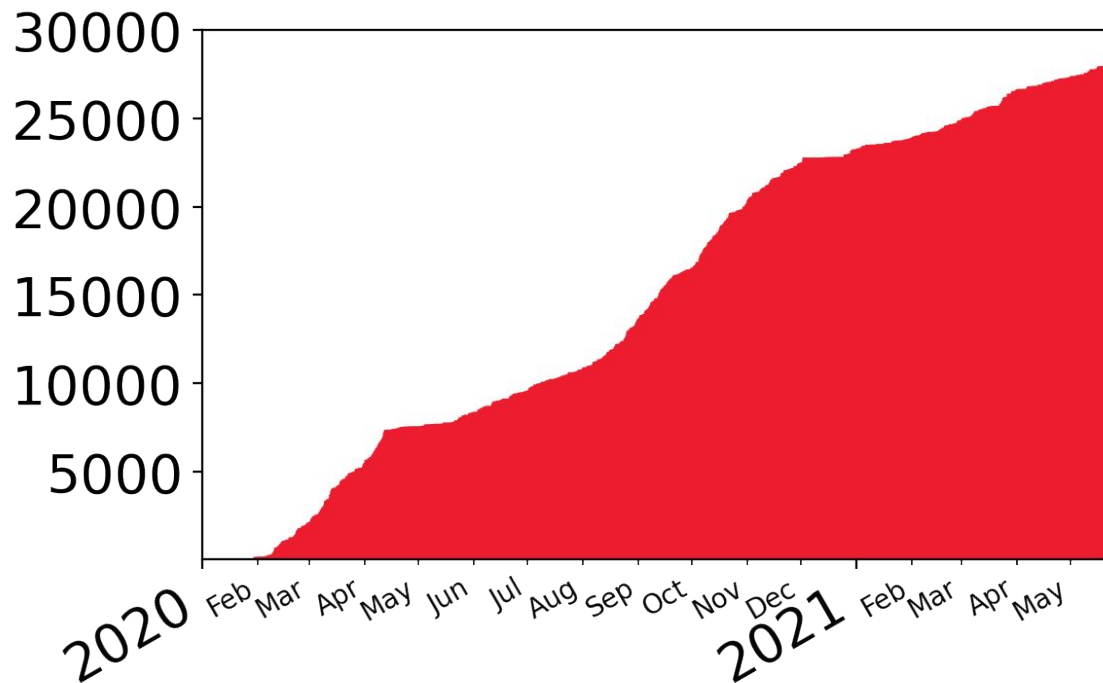
“Autonomous Intelligent Agents for Accelerated Materials Discovery” J.H. Montoya *et al.*, *Chem Sci* **11**, 8517-8532 (2020).

# CAMD finds new (meta)stable structures



“Novel inorganic crystal structures predicted using autonomous simulation agents”, W. Ye, X. Lei, *et al.*, *Sci Data* **9**, 302 (2022).

# CAMD finds a lot of metastable structures



<https://data.matr.io/7/>

CAMD discovers 2-3 new structures within 200 meV of the convex hull per hour

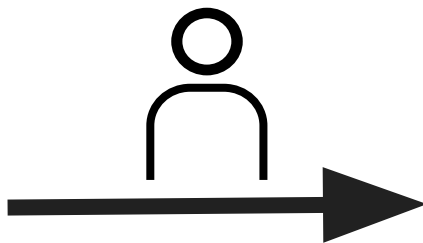
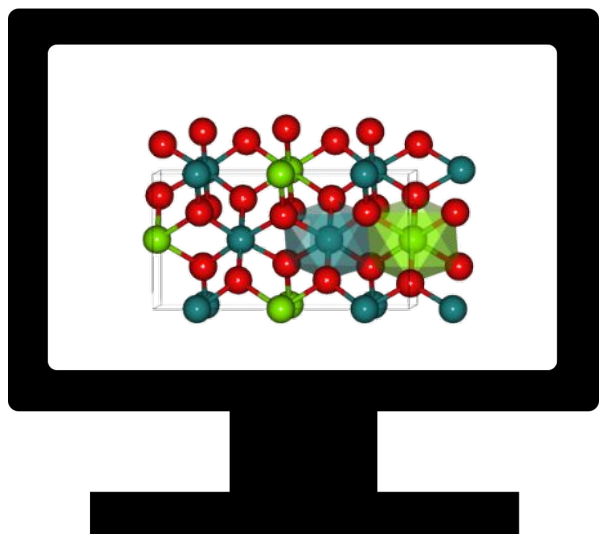
CAMD spends \$3.00 per metastable structure

CAMD's infrastructure is scalable

“Novel inorganic crystal structures predicted using autonomous simulation agents”, W. Ye, X. Lei, *et al.*, *Sci Data* **9**, 302 (2022).

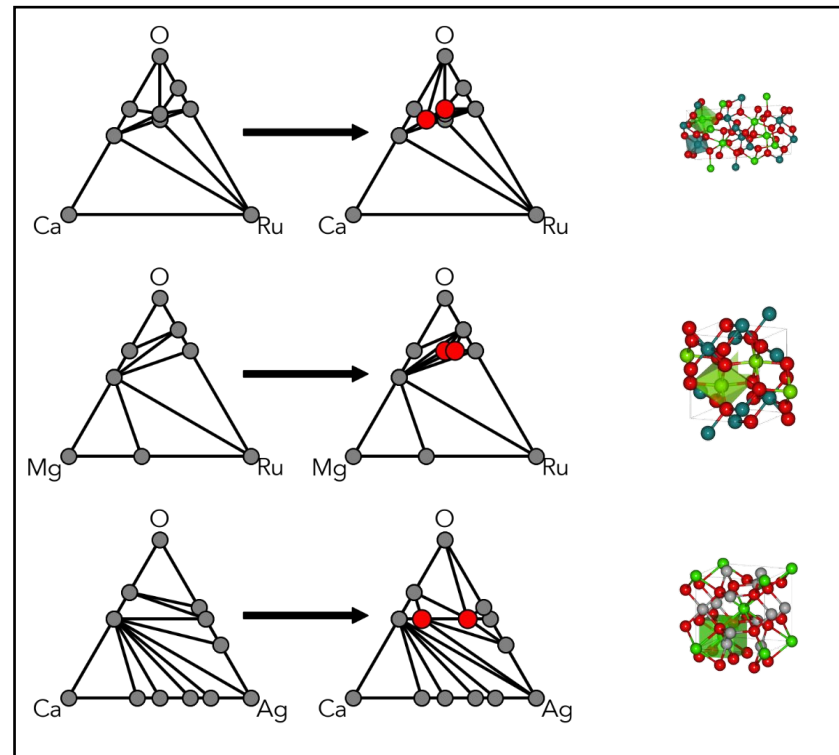
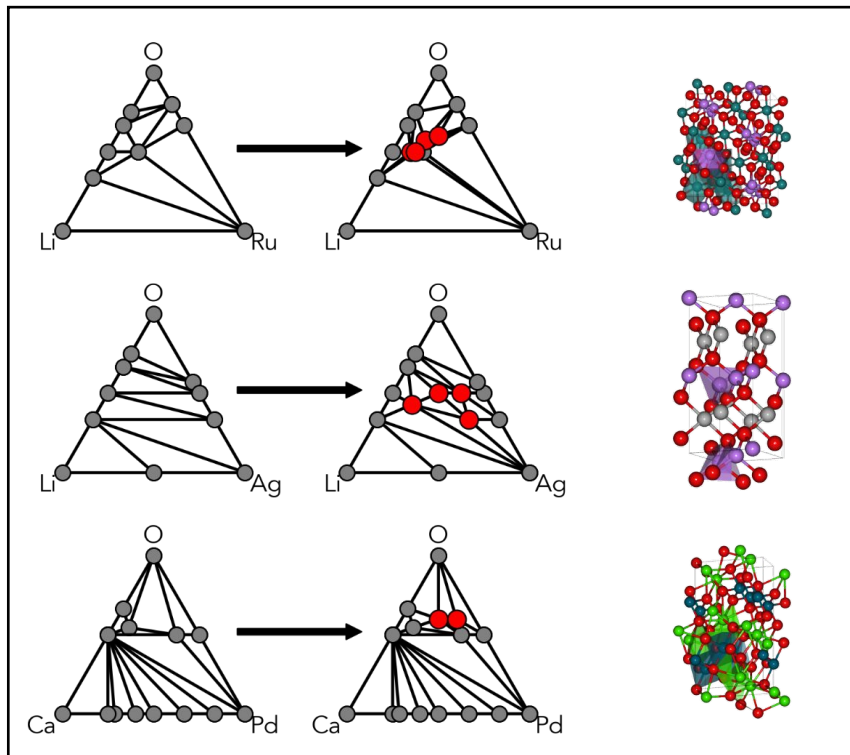


# We tried to make CAMD materials



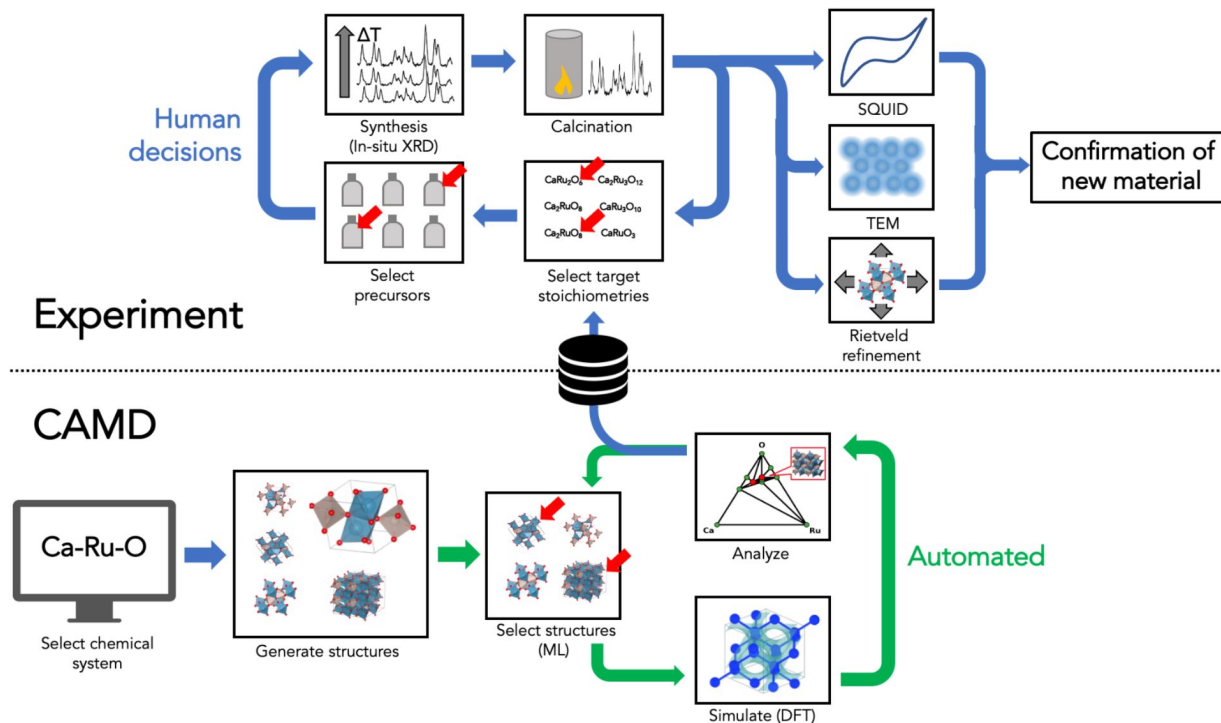
“Computer-assisted discovery and rational synthesis of ternary oxides”, J.H. Montoya *et al.*, DOI: [10.26434/chemrxiv-2023-n4pz9](https://doi.org/10.26434/chemrxiv-2023-n4pz9) (2023).

# Six CAMD inspired systems were selected



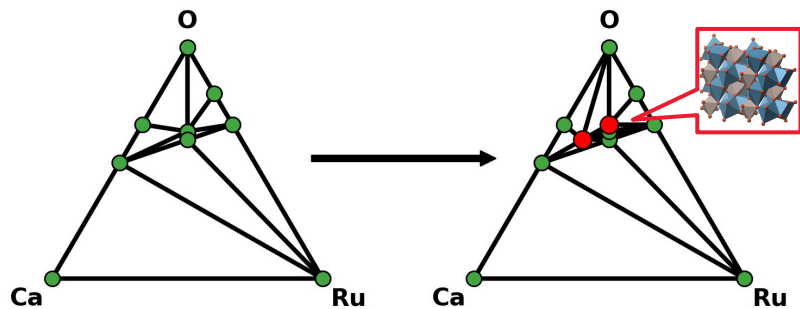
“Computer-assisted discovery and rational synthesis of ternary oxides”, J.H. Montoya *et al.*, DOI: [10.26434/chemrxiv-2023-n4pz9](https://doi.org/10.26434/chemrxiv-2023-n4pz9) (2023).

# Materials discovery workflow



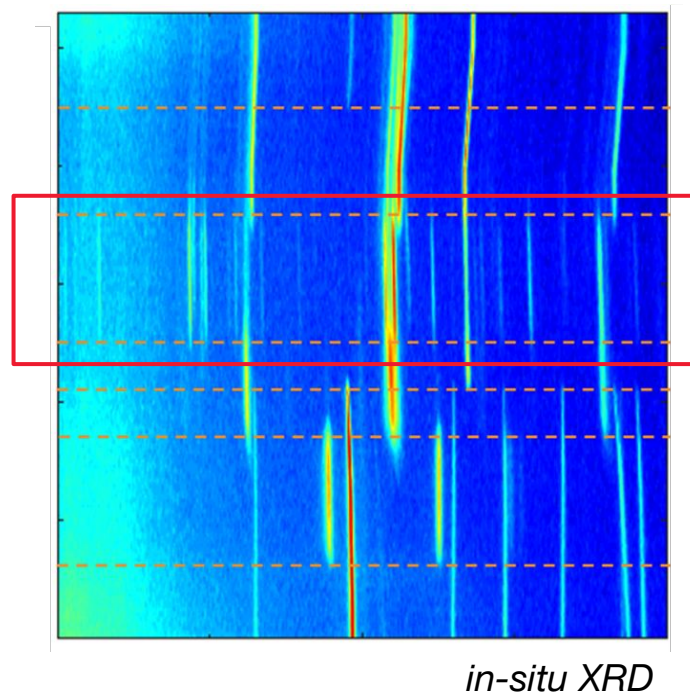
“Computer-assisted discovery and rational synthesis of ternary oxides”, J.H. Montoya *et al.*, DOI: [10.26434/chemrxiv-2023-n4pz9](https://doi.org/10.26434/chemrxiv-2023-n4pz9) (2023).

# A new material predicted by CAMD was synthesized



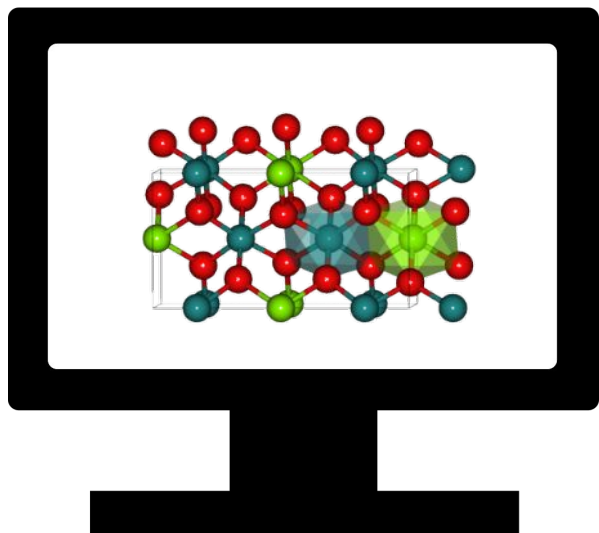
CAMD predicted a novel material containing Ru, a common element in catalyst materials.

In-situ, variable temp XRD experiments confirm a new phase exists, appears 700-1100°C

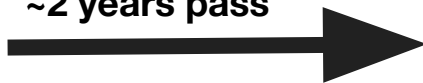


# Other examples start to sound the same

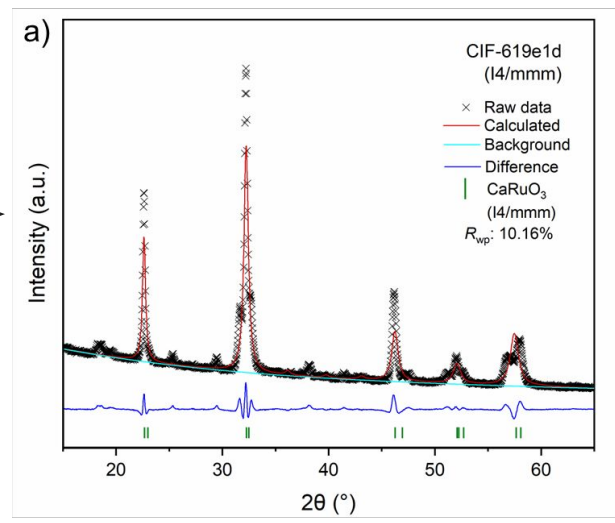
$10^4$  designed on the computer



~2 years pass



$10^0$  confirmed in the lab





# Computation-mediated discovery - challenges

## Efficiency is too low

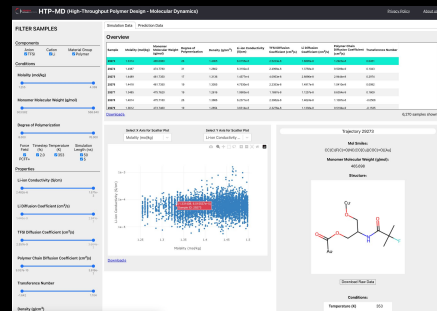
- Automating expt and expt analyses

## Experimental feedback is very difficult

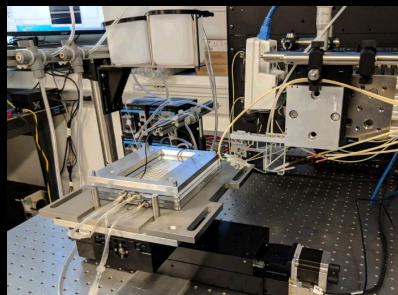
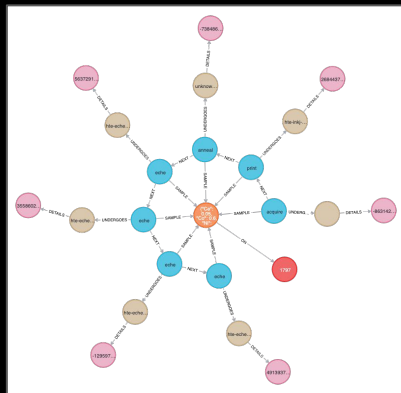
- Datasets: dark data
- Data-driven analyses

## Lack of theory to enable design

- Datasets: defect structures
- Datasets: kinetic properties
- Integrate data-theory
- Multiscale models



...and more research  
Polymers, catalysts, ML





**TOYOTA**  
RESEARCH INSTITUTE

# Outlook

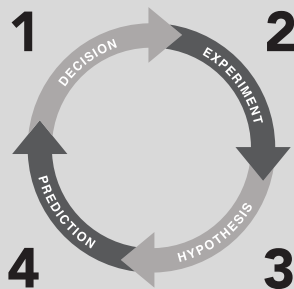
Bridging the gap between experiment & simulation

# Challenges with accelerating materials discovery

## Efficiency is too low

### Our approach

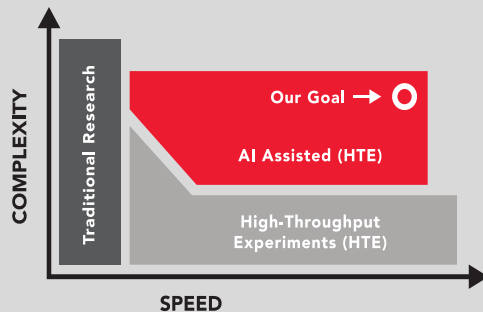
Automation & closed loop discovery.



## Experimental feedback is very difficult

### Our approach

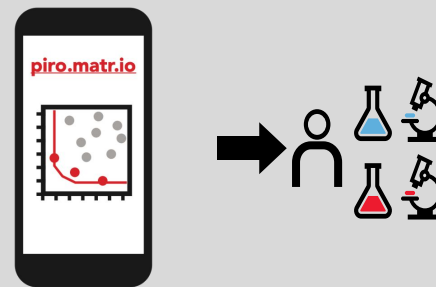
AI powered data analysis.



## Lack of theory to enable design

### Our approach

Actionable input to experiments



# Outlook

- Highlights
  - Battery lifetime
  - Novel materials
- Workflow software allows better integration of compute and experiment
- Industry research uses minimal or mid-scale compute -- but also will benefit from exascale compute to address fundamental questions



**Thank You**