MACHINE LEARNING FOR MATERIALS

MSE 404/MSE504

Instructor: Sergei V. Kalinin

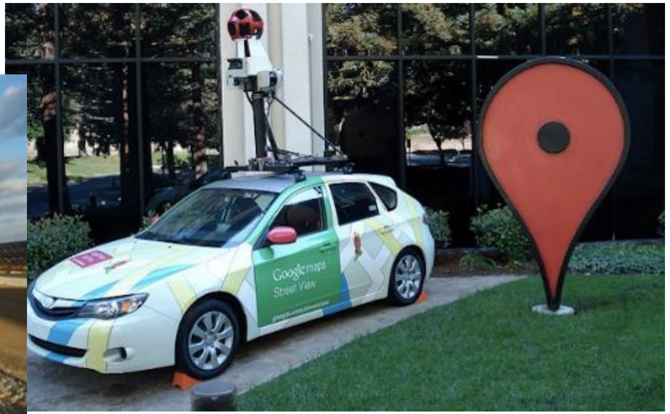
Times and locations: 10:20 am - 11:10 am MWF, Ferris Hall 502

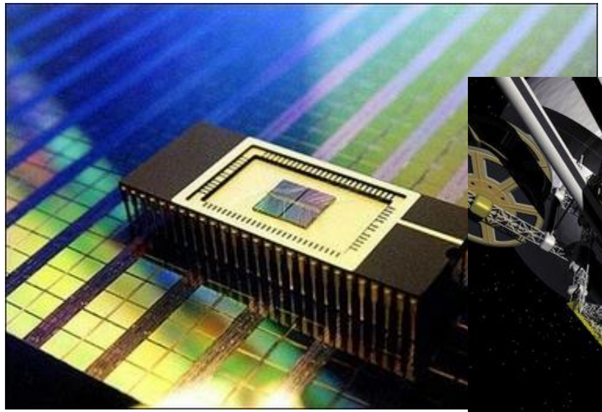

2002 - 2022
Since 2022
2022 -2023
SCANNING PROBE MICROSCOPY
Electrical and Electromechanical Phenomena at the Nanoscale
Sergei Kalinin
Alexei Gruverman
EDITORS
VOLUME 1
Springer
Sergei V. Kalinin
Alexei Gruverman
Editors
Scanning Probe Microscopy of Functional Materials
Springer
Scanning Probe Microscopy for Energy Research
Dawn A Bonnell • Sergei V Kalinin
World Scientific
Volume 1
World Scientific Series on Emerging Technologies – Volume 1
HANDBOOK ON BIG DATA AND MACHINE LEARNING IN THE PHYSICAL SCIENCES
In 2 Volumes
Volume 1
Big Data Methods in Experimental Materials Discovery
Editors
Surya R. Kalidindi • Sergei V. Kalinin
Turab Lookman
Editors-in-Chief
Sergei V. Kalinin • Ian Foster
World Scientific

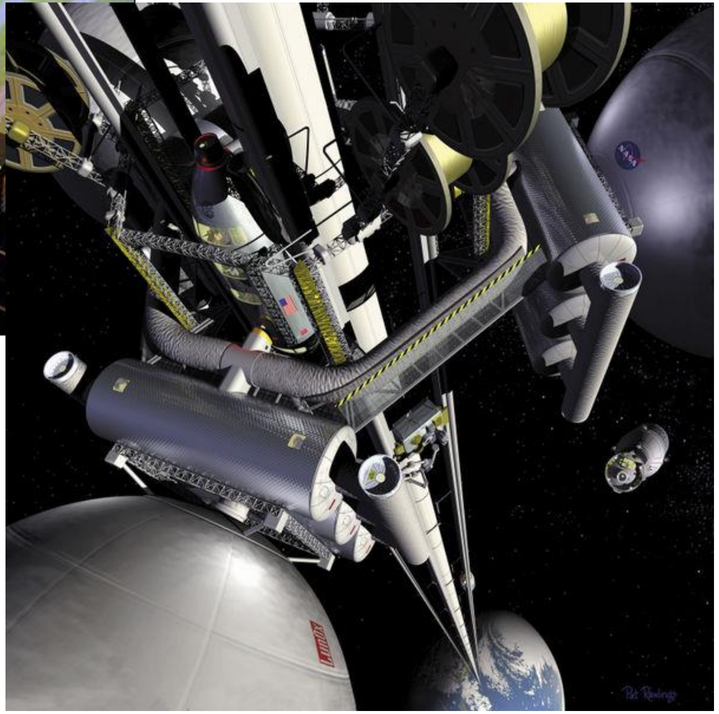
Materials Science cannot change overnight….

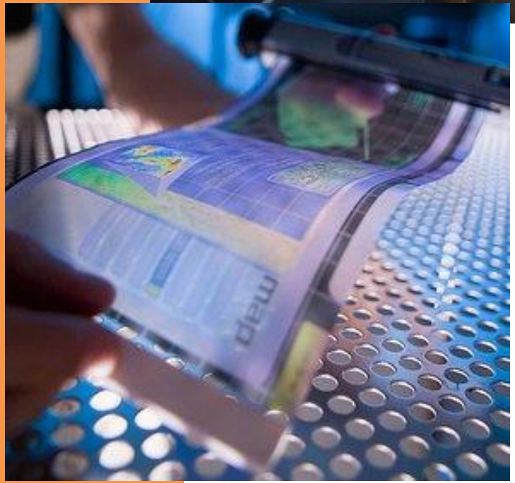
… but what is happening now comes very close

The World is Material Opportunity









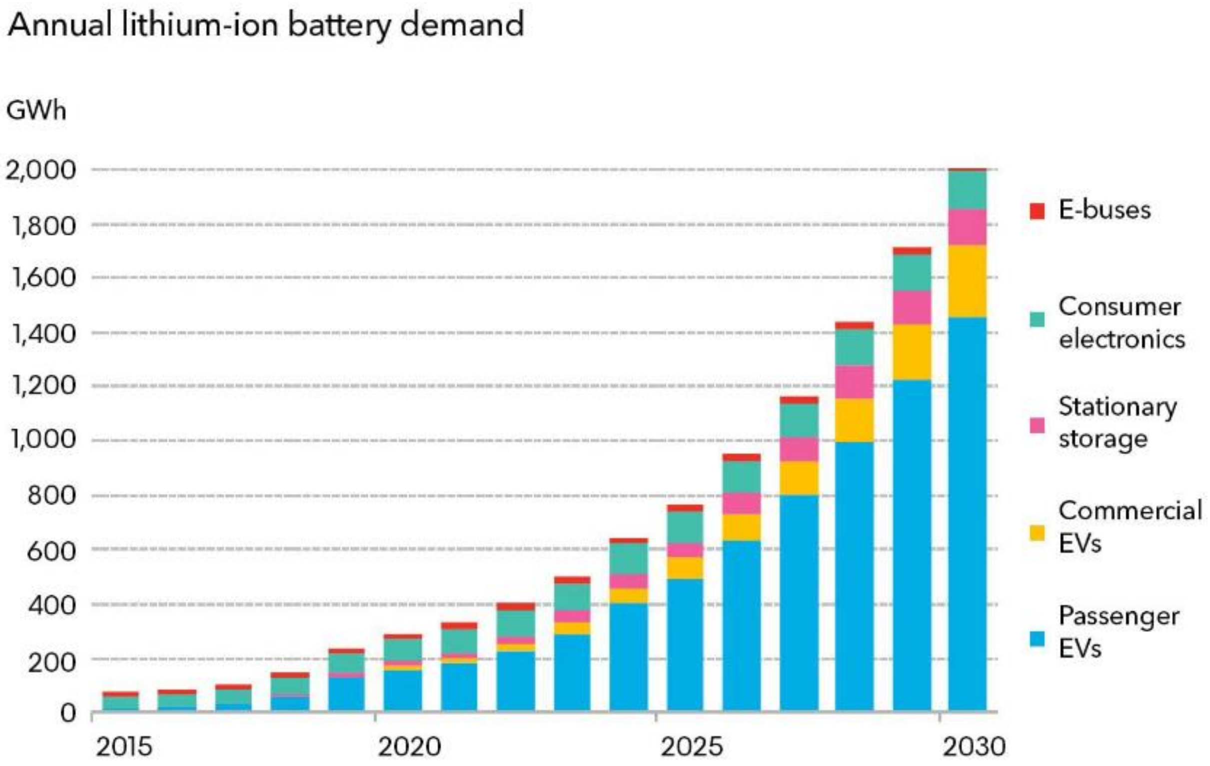
Predicting crystal structure by merging data mining with quantum mechanics

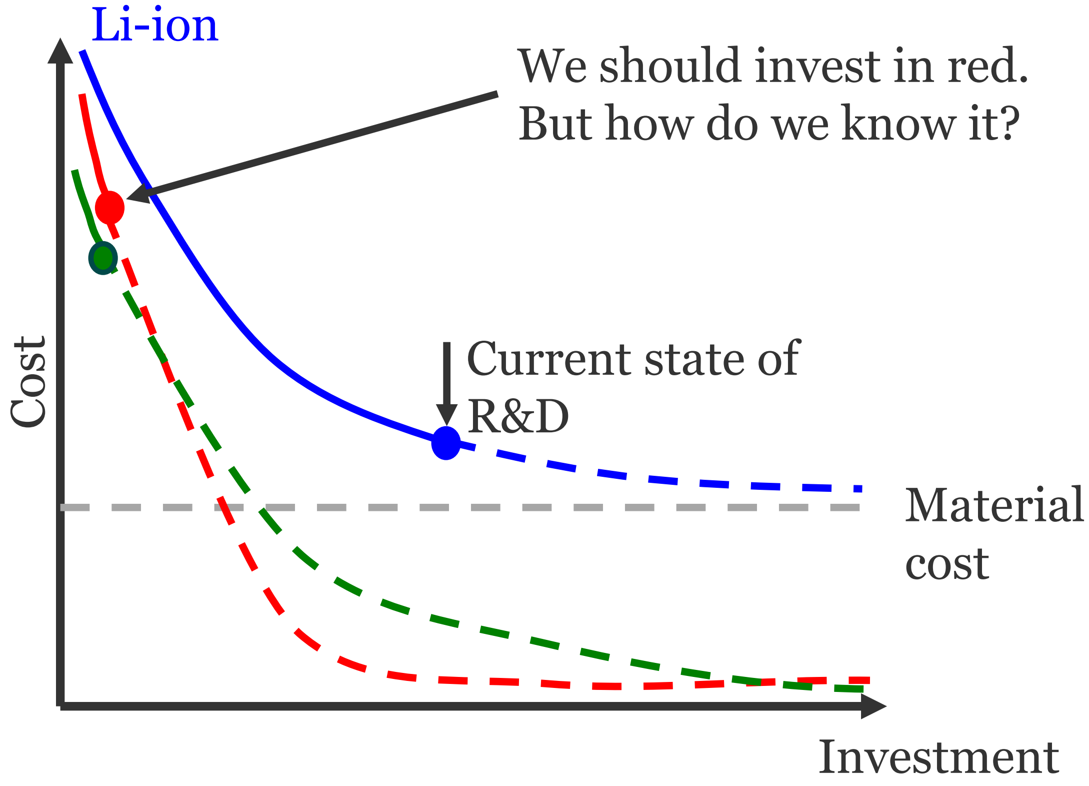
CHRISTOPHER C. FISCHER[1], KEVIN J. TIBBETTS[1], DANE MORGAN[2] AND GERBRAND CEDER[1]*
[1]Department of Materials Science and Engineering, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139, USA
[2]Department of Materials Science and Engineering, University of Wisconsin, Madison, Wisconsin 53706, USA
*e-mail: gceder@mit.edu

- **"Improve":** Renewable energy, self-driving cars, transparent displays, memory technologies
- **"Discover":** Room temperature superconductivity, high mechanical stress materials
- **"Engineer":** Quantum computing, single-atom catalysts, biomolecules

Functionality, manufacturability, cost

Batteries: Li-ion and Beyond

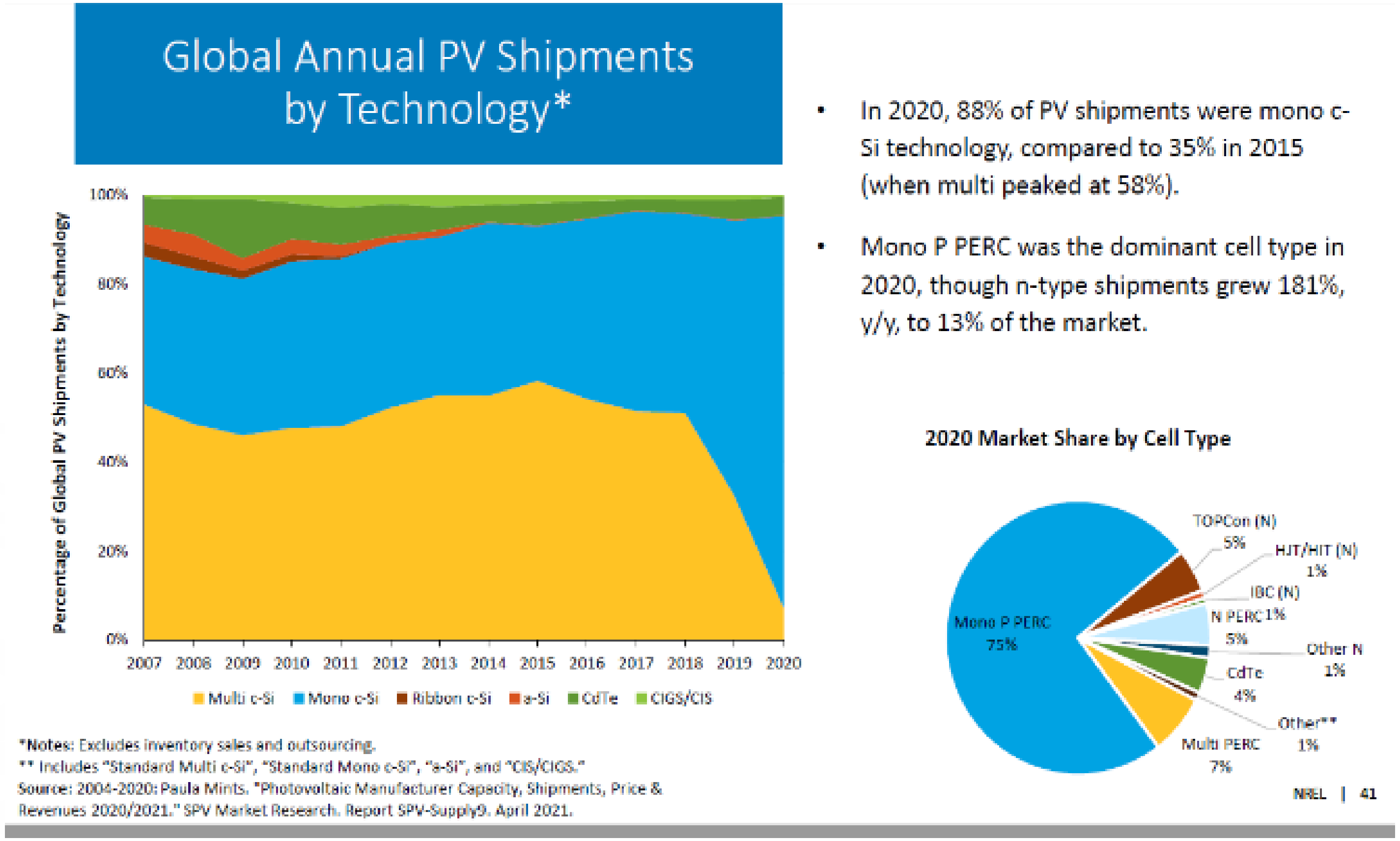




- Batteries are required element of energy transition (EVs, ESS, mobile devices)
- Currently Li-ion is the primary technology
- Optimization of Li-ion batteries takes years (even with same process on new Gigafactory)
- However, it is far from Goldilock zone for ESS or energy transport

- How can we optimize usage and safety for Li-ion batteries in EVs?
- How do we select beyond Li technologies for ESS?

Solar Energy: Will Silicon Ever Reign?



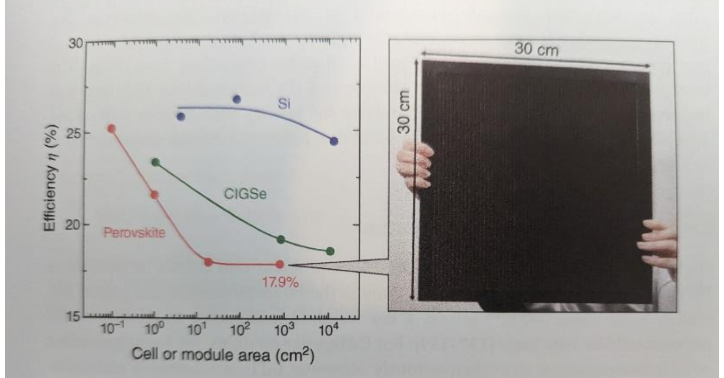
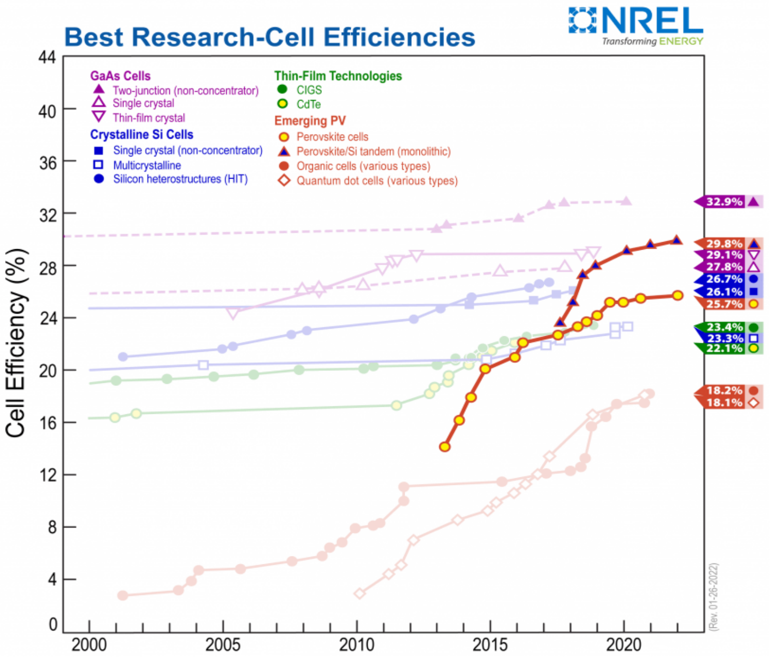
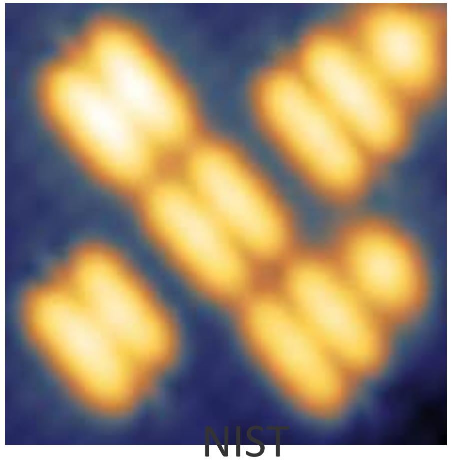


Figure 1.13 Conversion efficiencies of record efficiency devices (hybrid perovskite, Si, and CIGSe) as a function of the cell (or module) area. The data are taken from Ref. [1]. The inset shows the image of a large-area hybrid perovskite module with a size of 30 × 30 cm² (aperture area: 802 cm²), which was fabricated by Panasonic using an inkjet printing technology [133]. A very high efficiency of 17.9% has been reported for the large-area module. *Source:* Photograph of the large-area module is taken from Ref. [133] with permission.

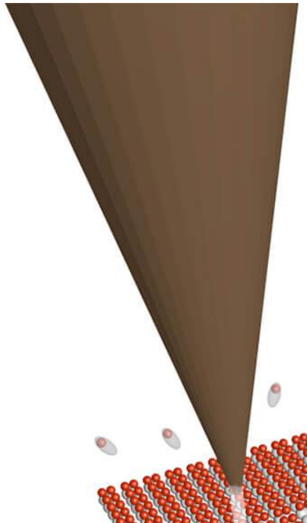
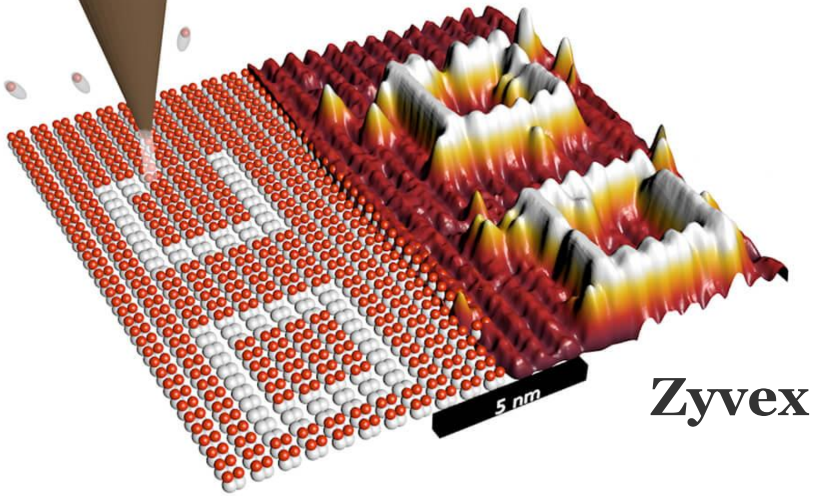


- Solar energy is the fastest growing energy sector
- Si is now reining material – however, it is really not the optimal material for PV (heavy, expensive)!
- Hybrid perovskites can be used as ideal PV materials – if we can make them stable and scale manufacturing!

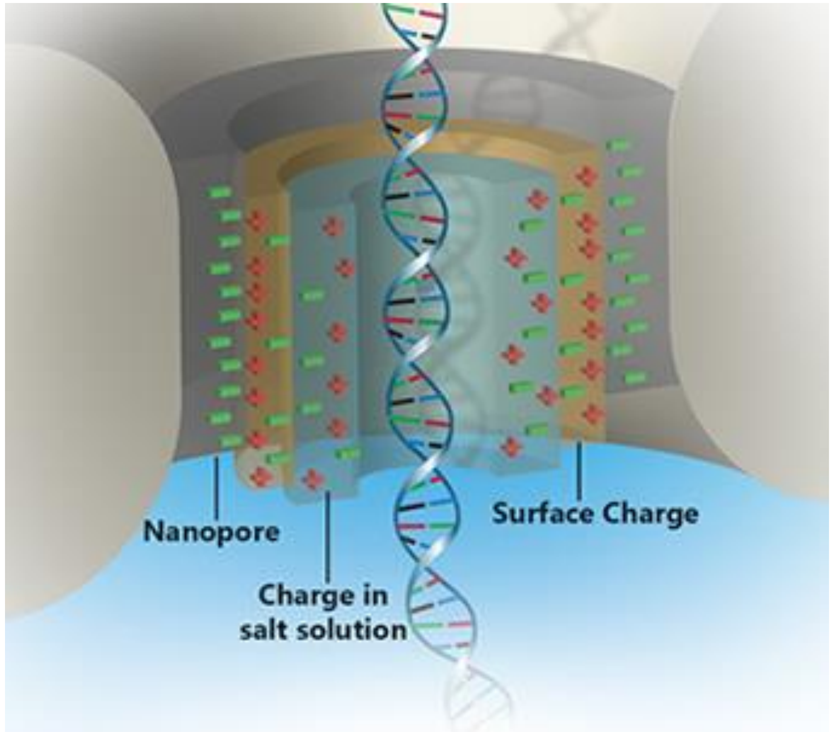
Quantum Computing and Single Molecule Bio



NIST

Zyvex Labs



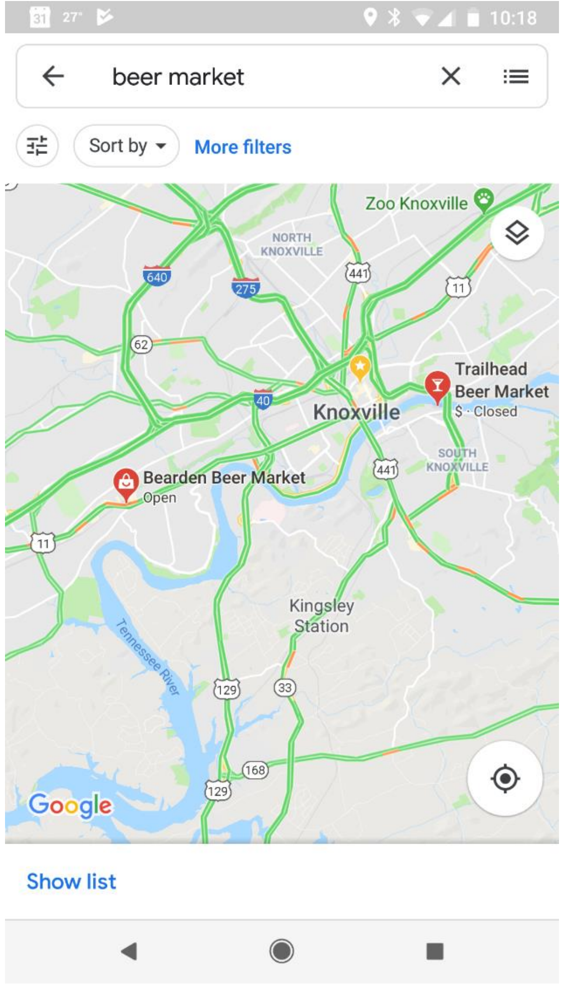
Oxford Nanopore

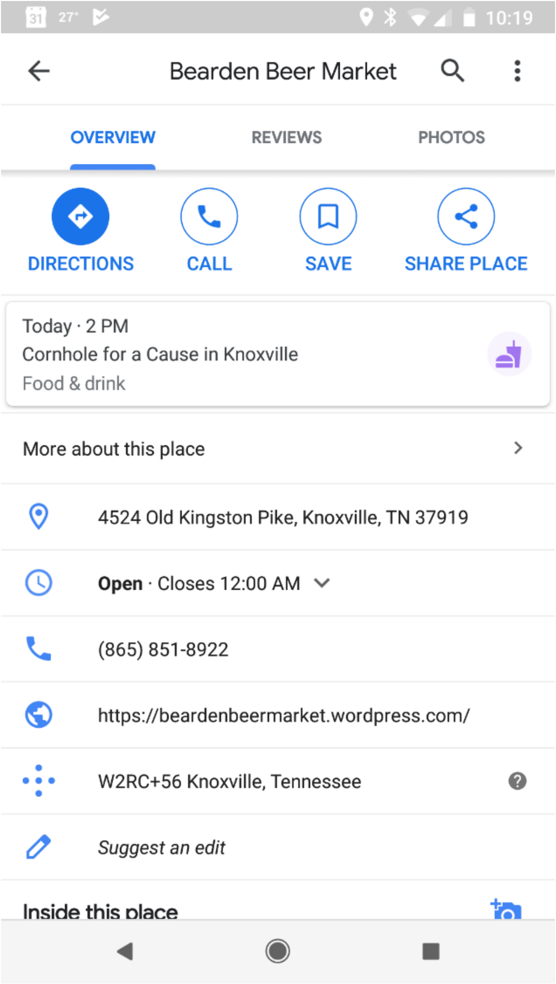
- Direct atomic fabrication: quantum communications and quantum computing, environmental sensing
- Single-molecule biological devices

- Success story 1: cryo-electron microscopy
- Success story 2: nanoelectron diffraction

Modern day world

- Google
- Facebook
- Yelp
- Netflix
- Uber
- Lyft
- …





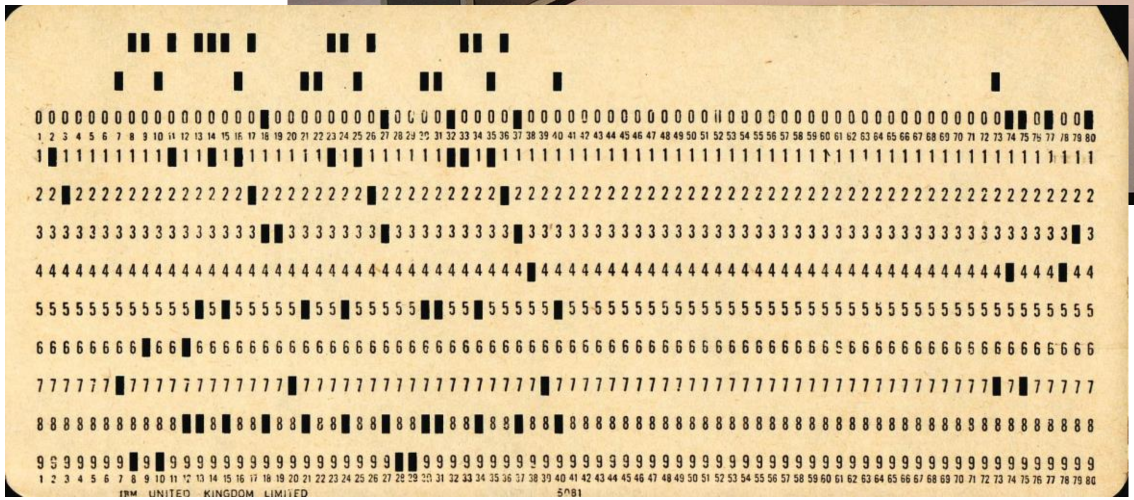
What was science like before 80ies?

- Computers existed only for specialized applications
- Many years of training before you can use one
- … and even if you can, required much patience

BESM-6 Computer

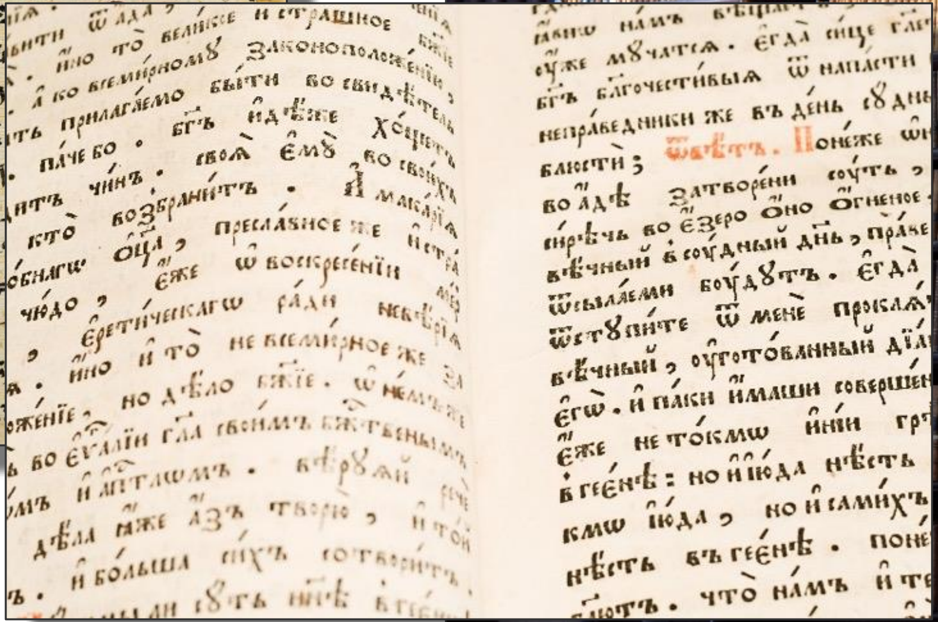
From Wikipedia

Punch card



Altair(duino)

What if you want to know more?

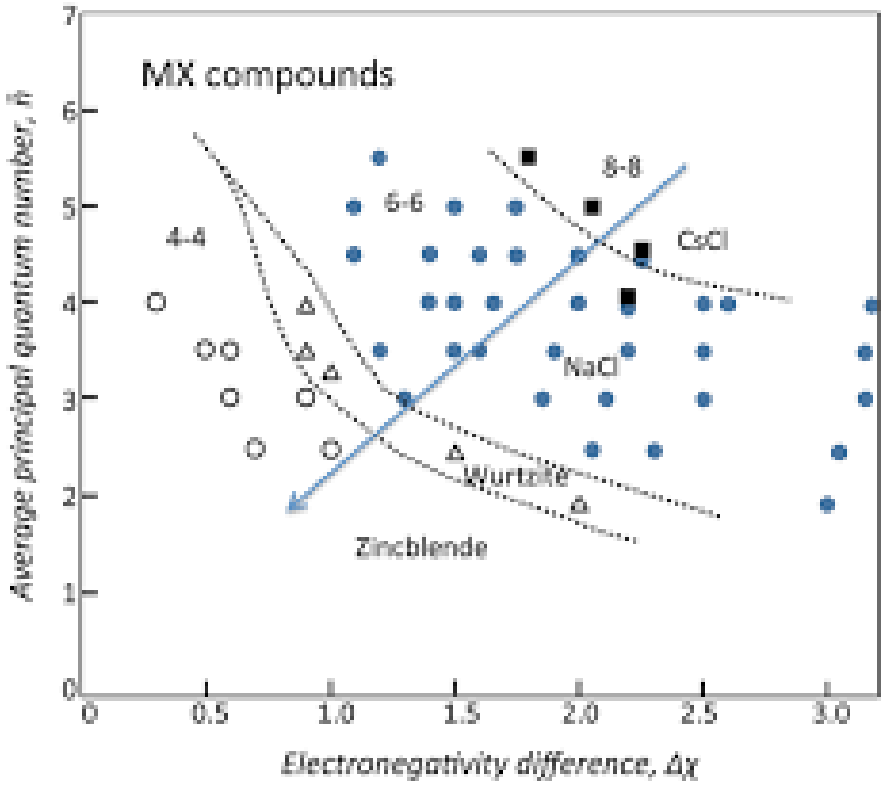


Libraries:

- Annual abstract books to find papers
- Dewey system to find books
- Collections: Landolt-Bornstein, etc

Was there machine learning for materials then?

Mooser-Pearson diagrams



- **Electronegativity:**

1	H 2.20																		He
2	Li 0.98	Be 1.57												B 2.04	C 2.55	N 3.04	O 3.44	F 3.98	Ne
3	Na 0.93	Mg 1.31												Al 1.61	Si 1.90	P 2.19	S 2.58	Cl 3.16	Ar
4	K 0.82	Ca 1.00		Sc 1.36	Ti 1.54	V 1.63	Cr 1.66	Mn 1.55	Fe 1.83	Co 1.88	Ni 1.91	Cu 1.90	Zn 1.65	Ga 1.81	Ge 2.01	As 2.18	Se 2.55	Br 2.96	Kr 3.00
5	Rb 0.82	Sr 0.95		Y 1.22	Zr 1.33	Nb 1.6	Mo 2.16	Tc 1.9	Ru 2.2	Rh 2.28	Pd 2.20	Ag 1.93	Cd 1.69	In 1.78	Sn 1.96	Sb 2.05	Te 2.1	I 2.66	Xe 2.60
6	Cs 0.79	Ba 0.89	*	Lu 1.27	Hf 1.3	Ta 1.5	W 2.36	Re 1.9	Os 2.2	Ir 2.20	Pt 2.28	Au 2.54	Hg 2.00	Tl 1.62	Pb 1.87	Bi 2.02	Po 2.0	At 2.2	Rn 2.2
7	Fr >0.79[en 1]	Ra 0.9	**	Lr 1.3[en 2]	Rf	Db	Sg	Bh	Hs	Mt	Ds	Rg	Cn	Nh	Fl	Mc	Lv	Ts	Og
			*	La 1.1	Ce 1.12	Pr 1.13	Nd 1.14	Pm –	Sm 1.17	Eu –	Gd 1.2	Tb 1.1	Dy 1.22	Ho 1.23	Er 1.24	Tm 1.25	Yb –		
			**	Ac 1.1	Th 1.3	Pa 1.5	U 1.38	Np 1.36	Pu 1.28	Am 1.13	Cm 1.28	Bk 1.3	Cf 1.3	Es 1.3	Fm 1.3	Md 1.3	No 1.3		

- **Nephelauxetic effect:**

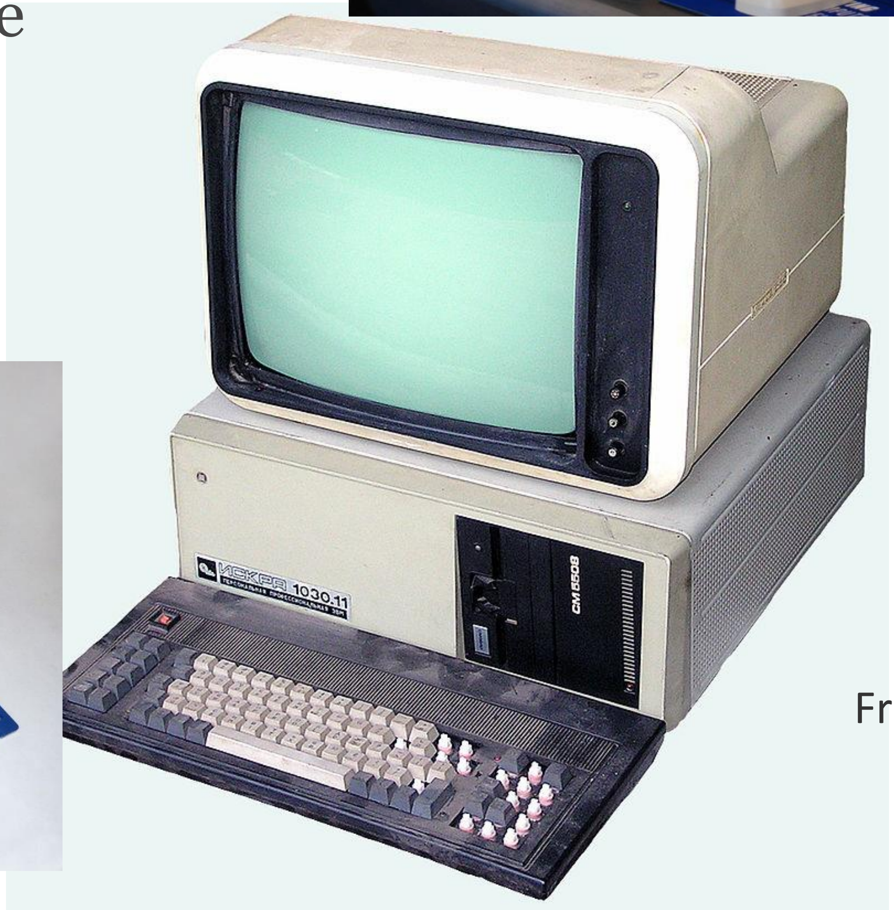
$F^- < H_2O < NH_3 <$ en $< [NCS - N]^- < Cl^- < [CN]^- < Br^- < N_3^- < I^-$

- **Solvent mixtures laws in organic chemistry**

Why does it matter? Physics and human heuristics is difficult to beat!

80ies

- First broadly available personal computers
- Can start programming (and get results) overnight
- First specialized scientific software
- What you have is what you bring



From Wikipedia

The 1980-2010 period:

Scientific information access:
- ISI and other data bases
- Electronic journals

General information
- Google and other search engines

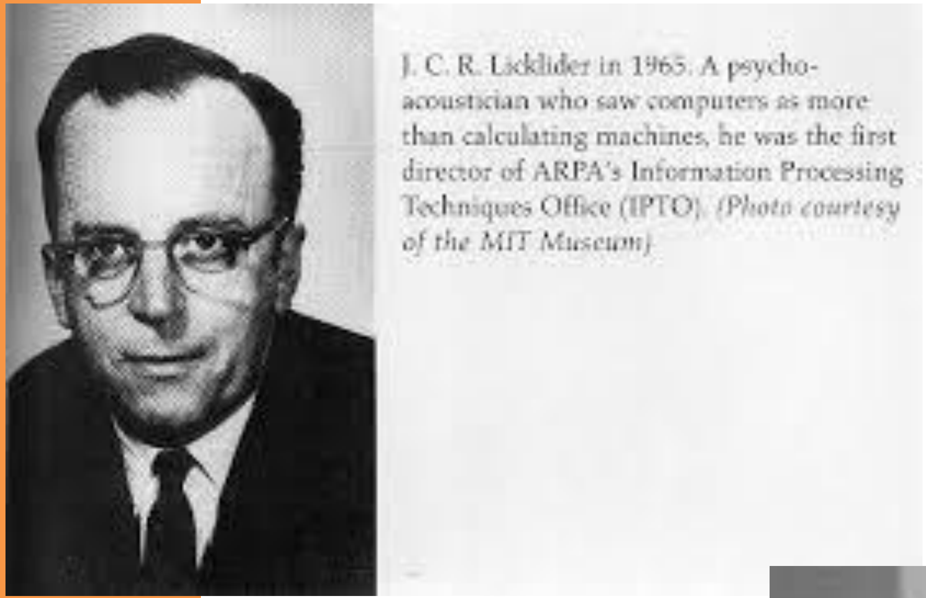
Scientific software:
- Word
- Origin
- Digital micrograph

Programming languages
- MatLab
- Igor Pro

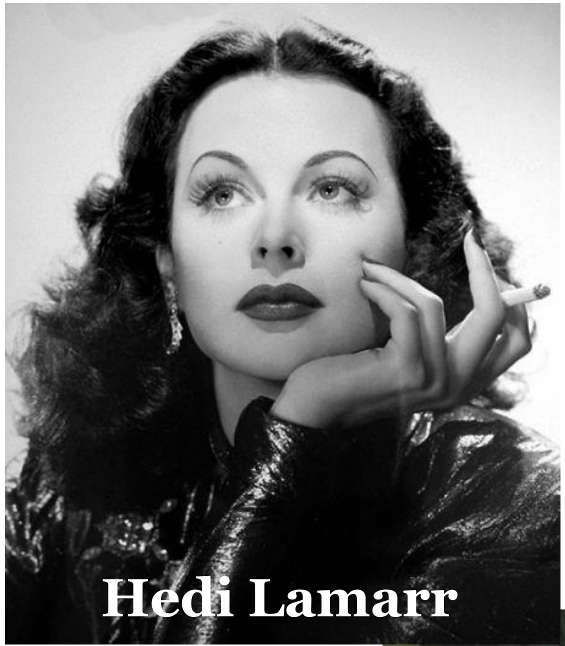
Instrument control (usually limited by manufacturer)

But relatively static….

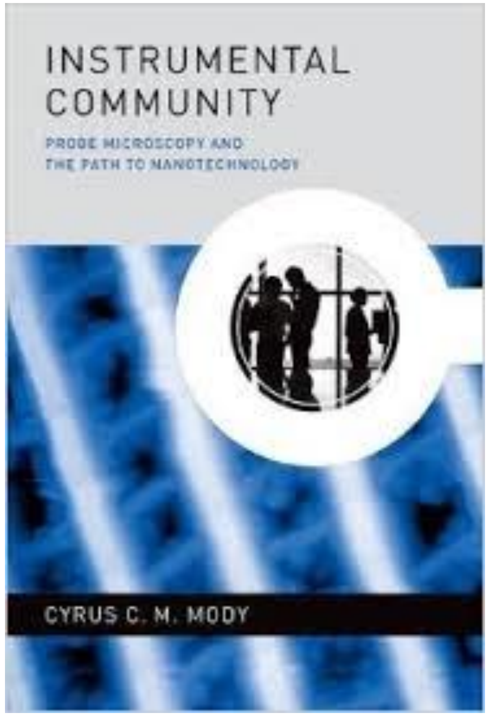
Could anyone have predicted it?

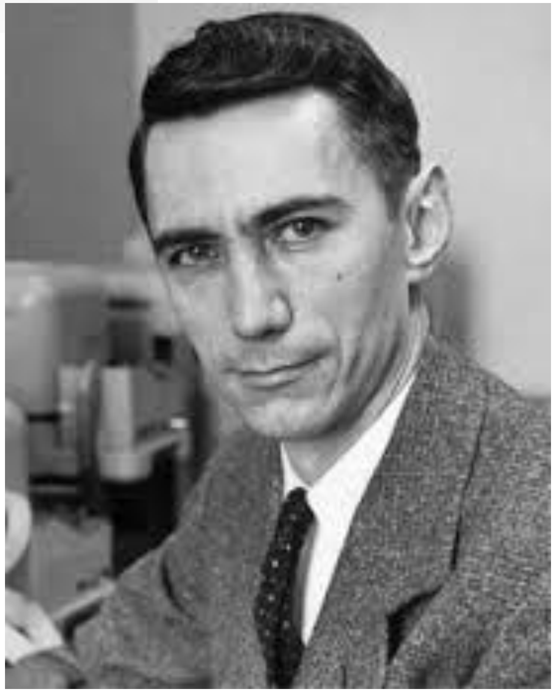


Ada Lovelace

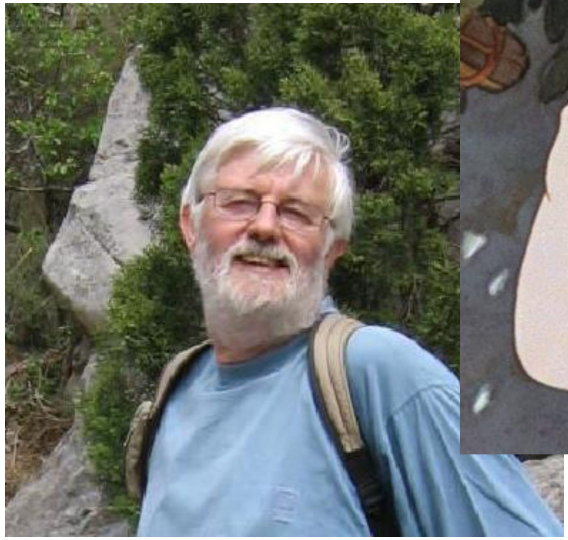


Hedi Lamarr





Claude Shannon



Noel Bonnet

What is happening now?

New data technologies

- Searches
- Social networks
- Recommender engines
- Connection to real world
- Large Language models

New opportunities:

- 3D Printing
- IoT devices
- Laboratory robotics
- Open code
- Text analytics

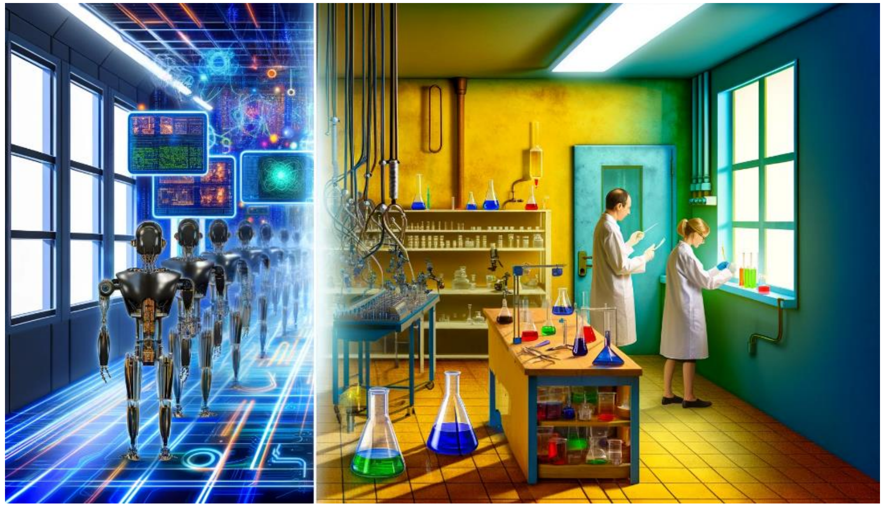
The inflection point (for theory): 2006

Predicting crystal structure by merging data mining with quantum mechanics

CHRISTOPHER C. FISCHER[1], KEVIN J. TIBBETTS[1], DANE MORGAN[2] AND GERBRAND CEDER[1]*
[1]Department of Materials Science and Engineering, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139, USA
[2]Department of Materials Science and Engineering, University of Wisconsin, Madison, Wisconsin 53706, USA
*e-mail: gceder@mit.edu

Publication by Gerd Ceder paper that is broadly seen as the inflection point launching Materials Genome Initiative in US and equivalent programs worldwide

Launch of AWS (Amazon Web Services) made cloud computing a reality – allowing businesses and scientists alike have access to computational resources without the need to build and maintain clusters



Machine learning in theory:

- Homogeneous workflows
- Known causal structure/lack of exogeneous factors
- Requires know-how, but relatively low entry barrier
- Easy to scale (given the funding)

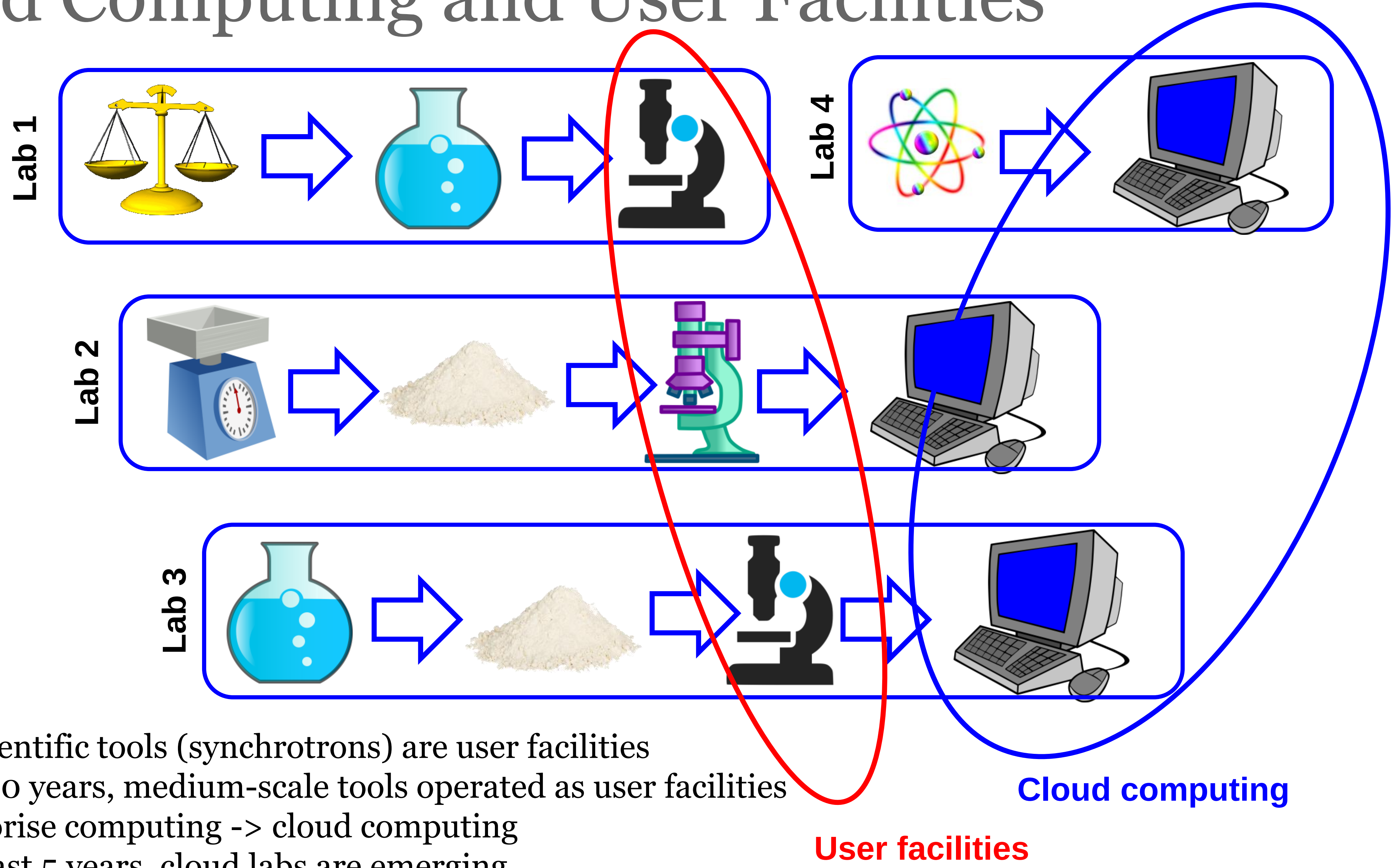
Instrumentation

General synthesis/characterization:
- Multiple data generation tools
- Complex workflows
- 100s at each university in US

Surface science lab:
- Multiple data generation tools
- Complex controls and workflows
- 10s in each university

Electron microscope:
- >100k worldwide
- Can cost up to ~4-5$M
- Can generate data at the ~10GB/s

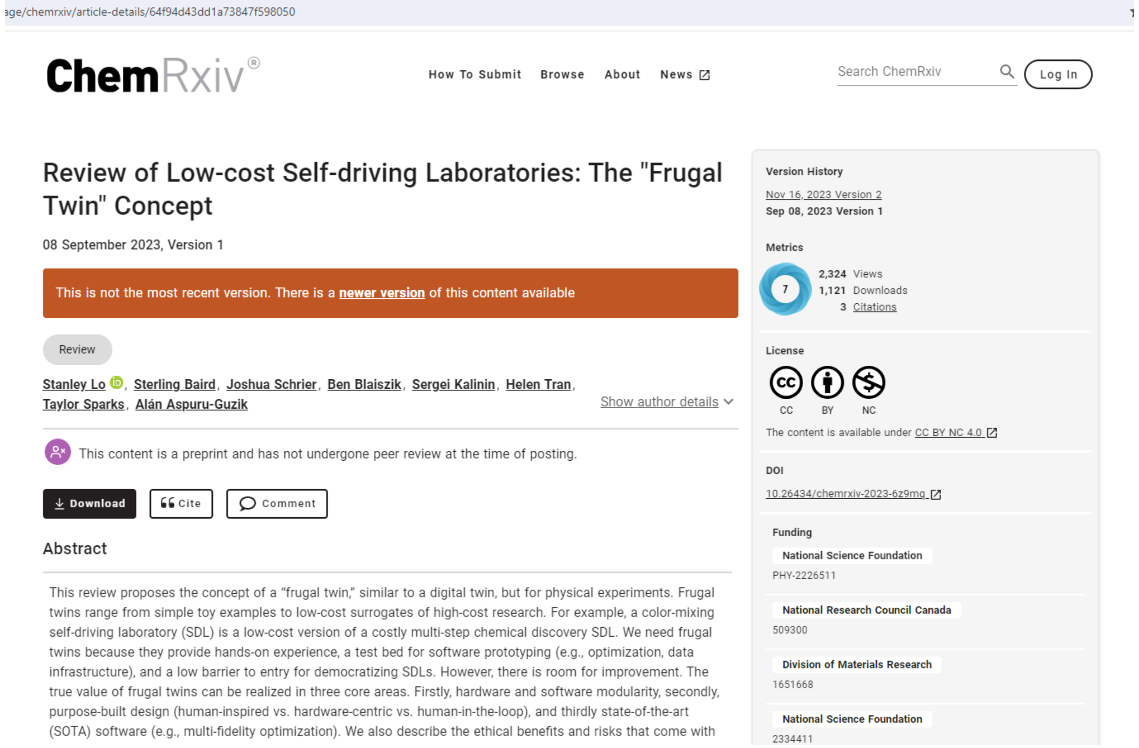
Cloud Computing and User Facilities



- Big scientific tools (synchrotrons) are user facilities
- For ~20 years, medium-scale tools operated as user facilities
- Enterprise computing -> cloud computing
- Over last 5 years, cloud labs are emerging
- What about the workflows?

The inflection point (for experiment): ~2020

- **Before 2010:** A number of (usually) confidential efforts in industry
- **2010 - 2015:** Early adopters and visionaries (Cronin, Maryama, Kusne, etc)
- **2015 – 2020:** The time of engineers
- **2020 – now:** Automated experimentation becomes broadly available with very low cost entry barriers


age/chemrxiv/article-details/64f94d43dd1a73847f598050

ChemRxiv®   How To Submit   Browse   About   News   Search ChemRxiv   Log In

Review of Low-cost Self-driving Laboratories: The "Frugal Twin" Concept

08 September 2023, Version 1

This is not the most recent version. There is a newer version of this content available

Review

Stanley Lo, Sterling Baird, Joshua Schrier, Ben Blaiszik, Sergei Kalinin, Helen Tran, Taylor Sparks, Alán Aspuru-Guzik   Show author details

This content is a preprint and has not undergone peer review at the time of posting.

Download   Cite   Comment

Abstract

This review proposes the concept of a "frugal twin," similar to a digital twin, but for physical experiments. Frugal twins range from simple toy examples to low-cost surrogates of high-cost research. For example, a color-mixing self-driving laboratory (SDL) is a low-cost version of a costly multi-step chemical discovery SDL. We need frugal twins because they provide hands-on experience, a test bed for software prototyping (e.g., optimization, data infrastructure), and a low barrier to entry for democratizing SDLs. However, there is room for improvement. The true value of frugal twins can be realized in three core areas. Firstly, hardware and software modularity, secondly, purpose-built design (human-inspired vs. hardware-centric vs. human-in-the-loop), and thirdly state-of-the-art (SOTA) software (e.g., multi-fidelity optimization). We also describe the ethical benefits and risks that come with

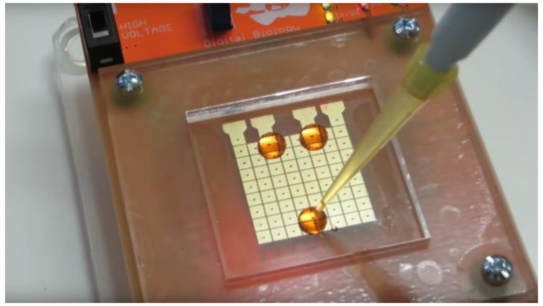
Version History
Nov 16, 2023 Version 2
Sep 08, 2023 Version 1

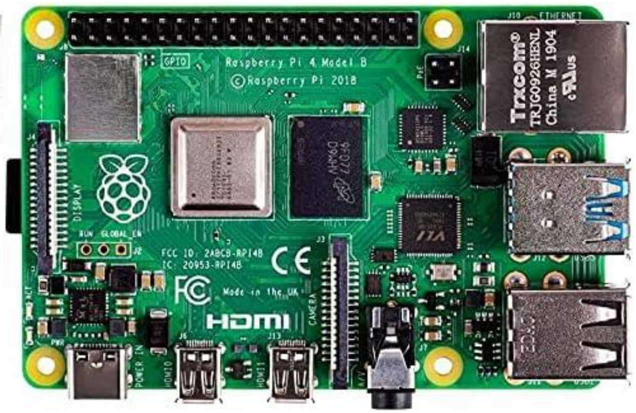
Metrics
7
2,324 Views
1,121 Downloads
3 Citations

License
CC BY NC
The content is available under CC BY NC 4.0

DOI
10.26434/chemrxiv-2023-6z9mq

Funding
National Science Foundation
PHY-2226511
National Research Council Canada
509300
Division of Materials Research
1651668
National Science Foundation
2334411

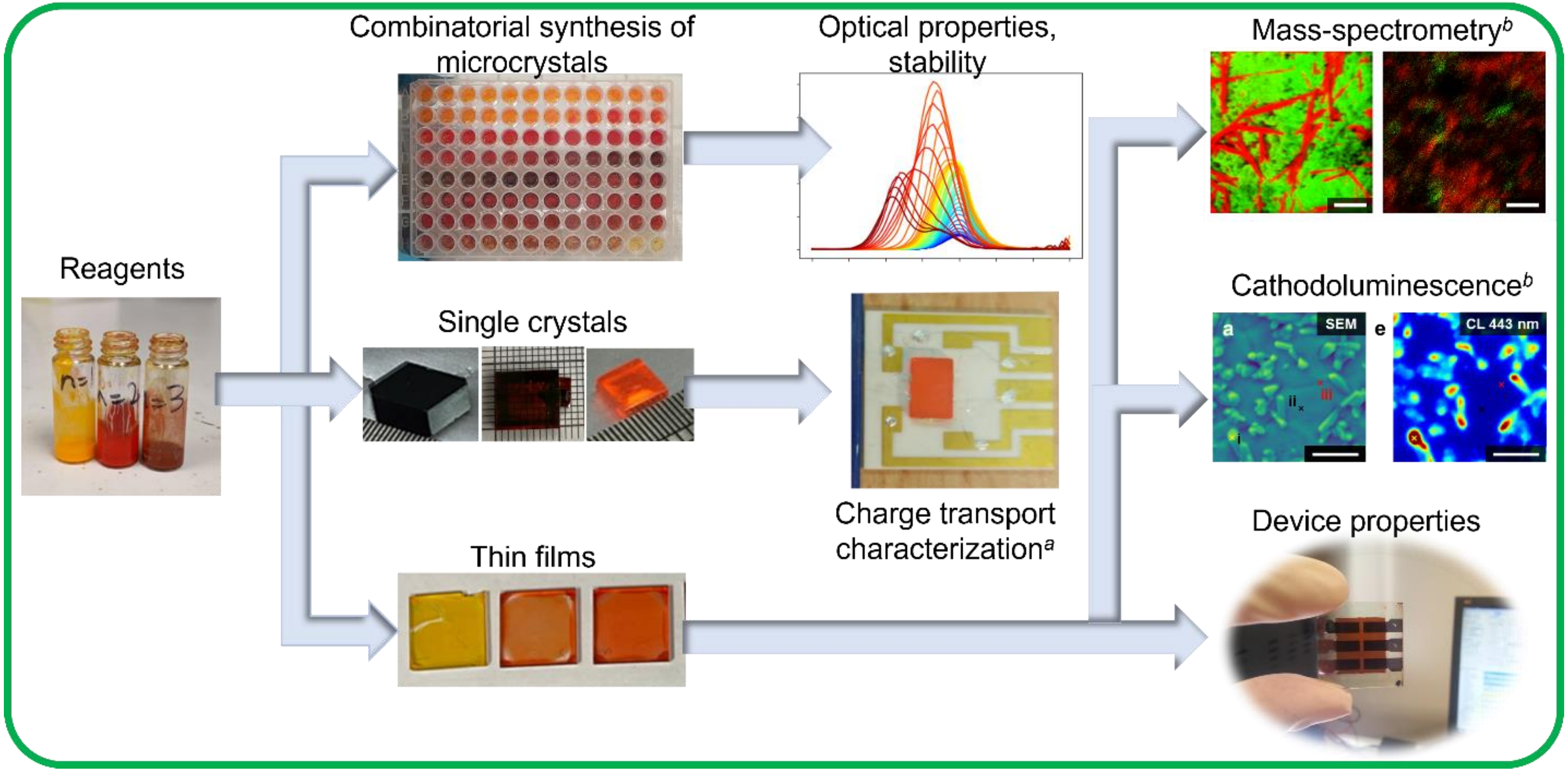




Why do we need machine learning?

To help us make decisions!

What is A Workflow?



- **Workflow:**
- Ideation, orchestration, implementation
- Domain specific language
- Dynamic planning: latencies and costs
- Reward and value functions

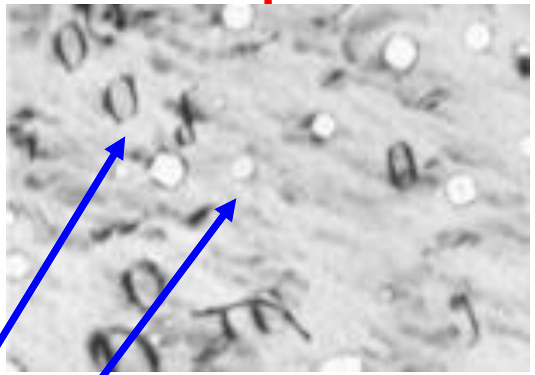
Designed in academia and adopted by industry

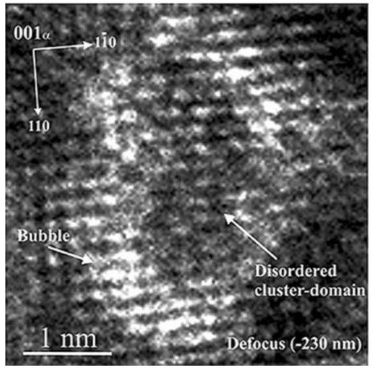
- Are they optimal?
- Can we design them better?
- Can they be changed dynamically?

Workflows for Nuclear Materials Design

Traditional experiment:
1. Always based on workflows
2. Ideated, orchestrated, and implemented by humans
3. The "gain of value" during the workflow implementation is uncertain

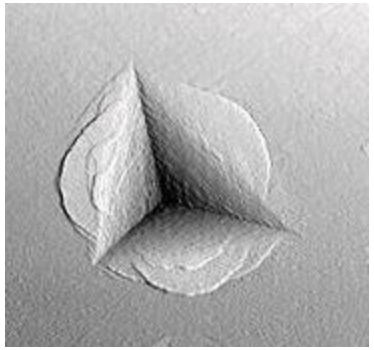
Materials selection





Prior knowledge

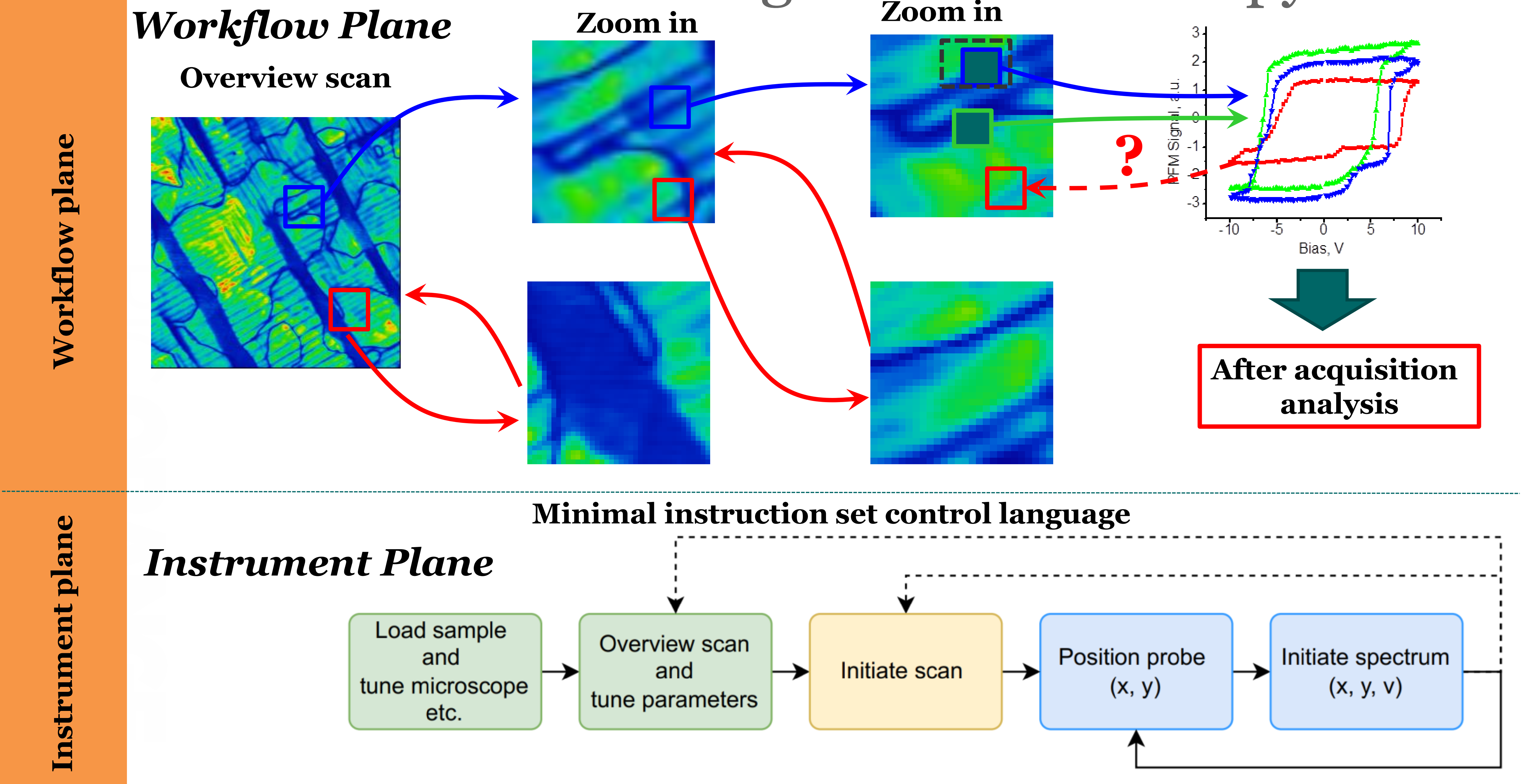
Operator curiosity



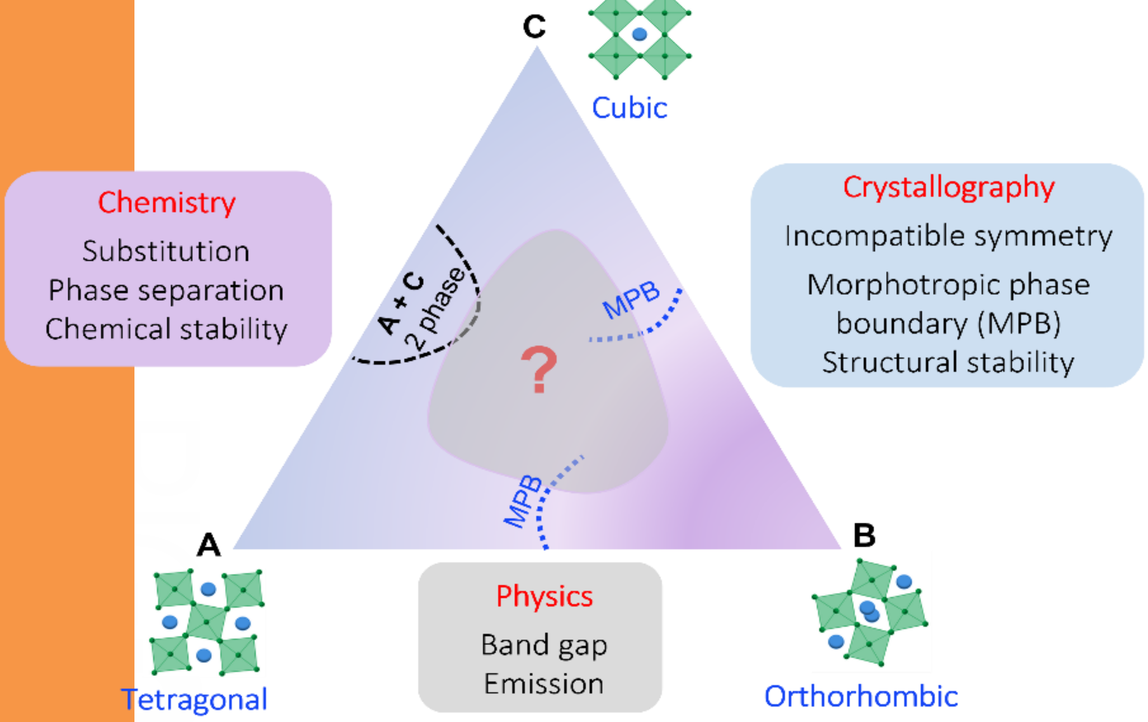
Value

Value of the step is key element:
- Either based on prior knowledge
- Or defined in a sense of the reinforcement learning Q-function

Workflows in Scanning Probe Microscopy



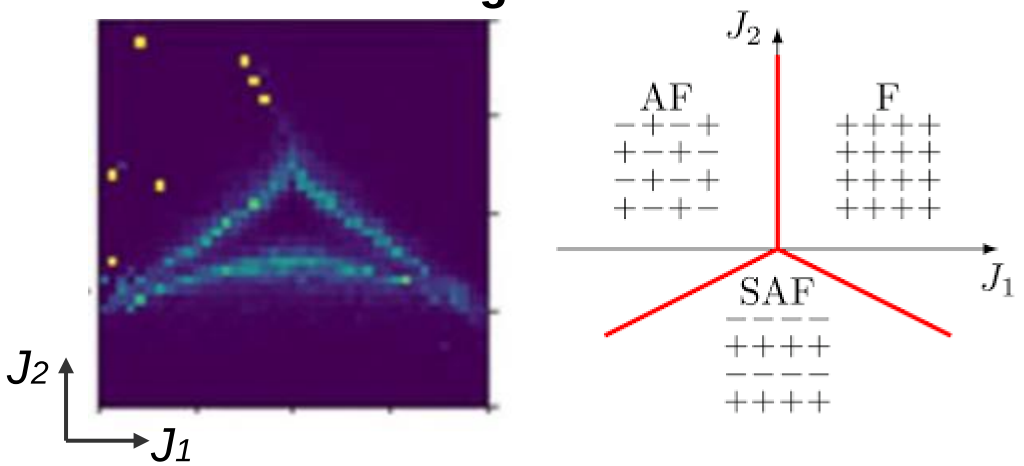
Why synthesis (or theory)?



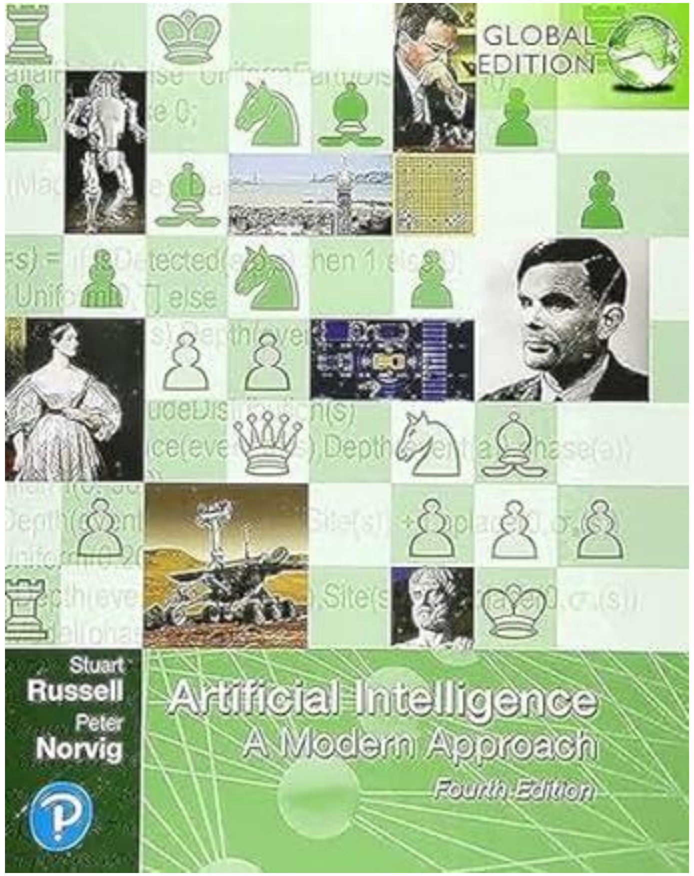
- Automated synthesis in its simplest form requires some way to navigate phase diagrams
- In more complex form, processing space.
- Ideally, incorporate physical knowledge

- Similar problem - theory

Ising model



Somewhat remarkably, almost all AI research until very recently has assumed that the performance measure can be exactly and correctly specified in the form of utility or reward function



In 2012-2014, large number of theory-trained workforce (DFT, MD, FEA) has moved into AI/ML fields

Many are returning to their original domain areas, building start-ups and corporate research centers. Many of these gain valuation very rapidly (In Silico, Schroedinger, etc)

The value created by AI/ML is orders of magnitude below investment (autonomous driving, radiology, cashier less stores)

Real-world industries and manufacturing are looking for work force with the combination of **domain** and **ML/AI** expertise

Course Information

Faculty Contact Information:

Instructor:	Prof. Sergei V. Kalinin,
Office:	314 IAMM
E-mail:	sergei2@utk.edu
Teaching Assistant:	Sheryl Sanchez, ssanch18@vols.utk.edu

Instructor Availability:
Please don't hesitate to email me with updates, questions, or concerns. I will typically respond within 24 hours during the week and 48 hours on the weekend. I will notify you if I will be out of town and if connection issues may delay a response.

Meeting Time: 10:20 am - 11:10 am MWF, Ferris Hall 502

The lectures and materials will be posted on Canvas and at GitHub:
https://github.com/SergeiVKalinin/MSE_Fall2024

Office Hours:
Friday 1:30 - 3:00 PM are open for 1:1 meetings to discuss any course related item. Please set up time via email.

Prerequisites

To be successful in this course you will need a general background in materials science. Python or similar programming experience, while not essential, will be extremely useful. Students without any prior programming experience should expect to spend extra time outside of class learning basic skills.

This and that

Learning Environment:
The class will be delivered as in-person lectures. The Jupyter notebooks, code libraries, and videos provided. Weekly programming exercises will be assigned via Google Colabs and those students wishing to interact with the instructor in person should attend office hours.

Use of ChatGPT:
Strongly encouraged both for programming and written assignments. However, the students have to be aware of the limitations of the generative models.

Grading & Policies:

• Homework assignments	40%
• Mid-terms (2)	30%
• Final Project & Presentation	30%

Reference Materials

I will provide copies of lecture notes, presentations, and Colabs on GitHub and Canvas. There is no specific textbook for the course and we will take material from a variety of sources including:

- Andrew Bird et al, *Python Workshop – Second Edition*, https://subscription.packtpub.com/book/programming/9781804610619/1

- Sebastian Raschka, *Machine Learning with PyTorch and Scikit-Learn*, https://subscription.packtpub.com/book/data/9781801819312/1

- Rowel Atienza, *Advanced Deep Learning with TensorFlow 2 and Keras - Second Edition*, https://www.packtpub.com/product/advanced-deep-learning-with-tensorflow-2-and-keras-second-edition/9781838821654

Homework 1:

- Create new Colab, https://colab.google/
- Chapter 1, Python Workshop.