THE UNIVERSITY of TENNESSEE KNOXVILLE

MACHINE LEARNING FOR MATERIALS: FROM PCA TO CHATGPT

MSE 494/MSE576

Instructor: Sergei V. Kalinin

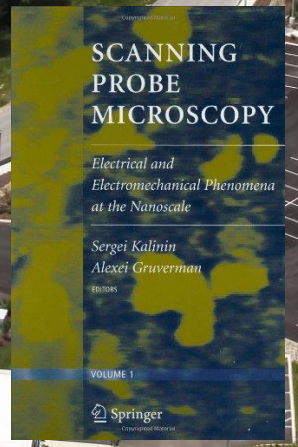
Times and locations: 10:20 am - 11:10 am
MWF, Ferris Hall 502

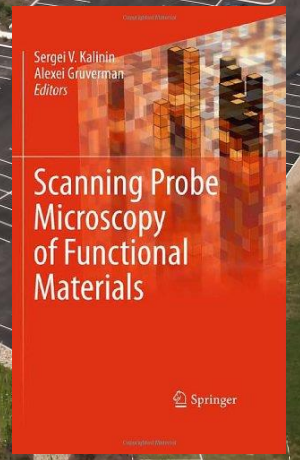
2002 - 2022

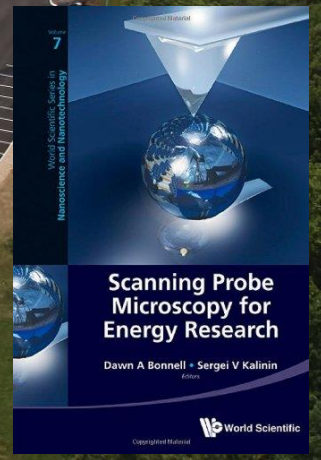
Since 2022

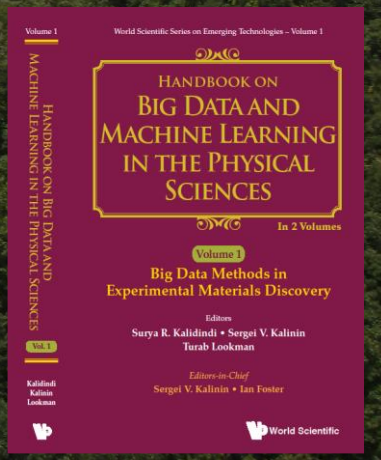
2022 -2023

amazon



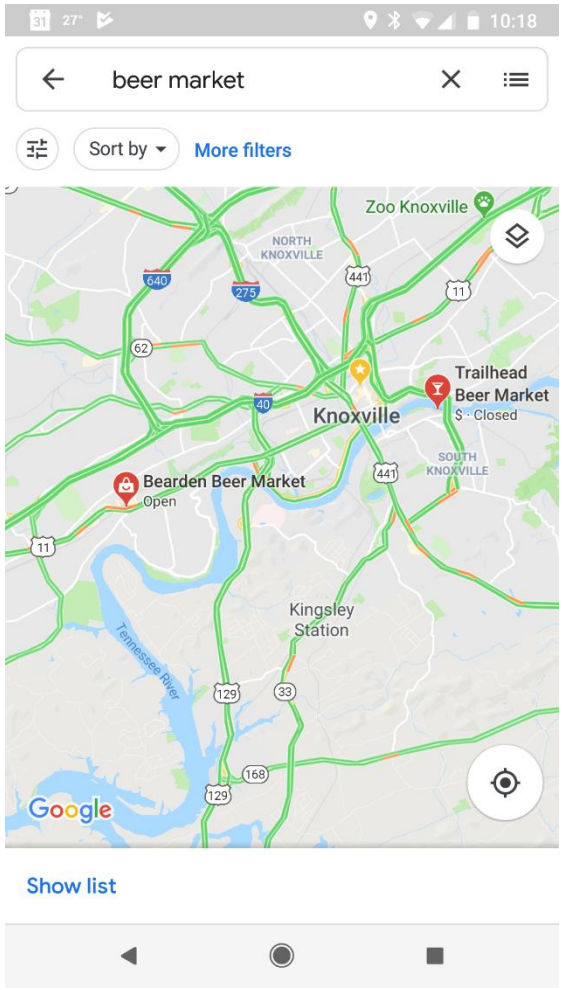


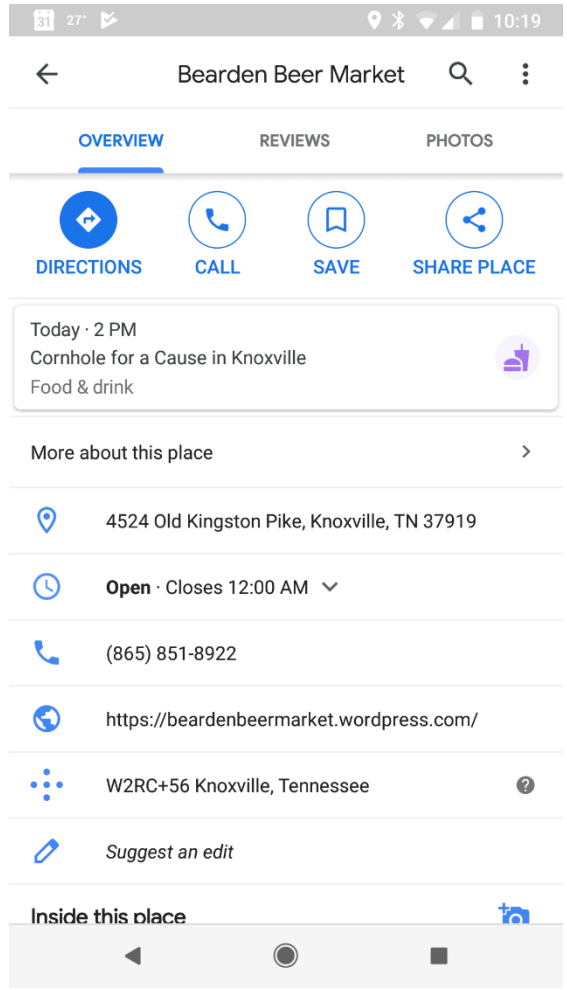




Modern day world

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



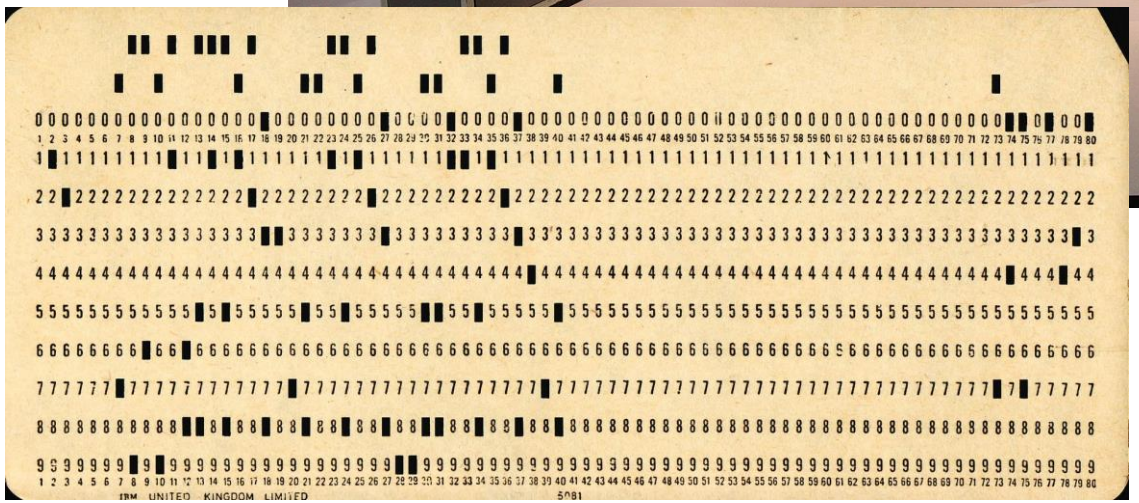


What was science like before 80ies?

- 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

Punch card

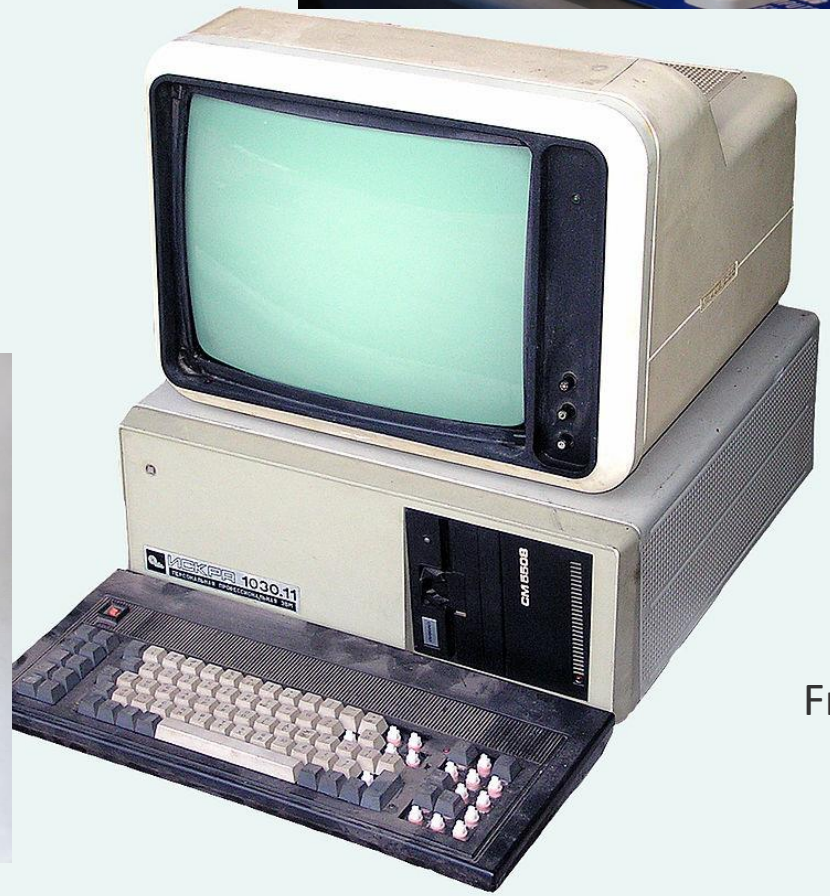


From Wikipedia

Altair(duino)

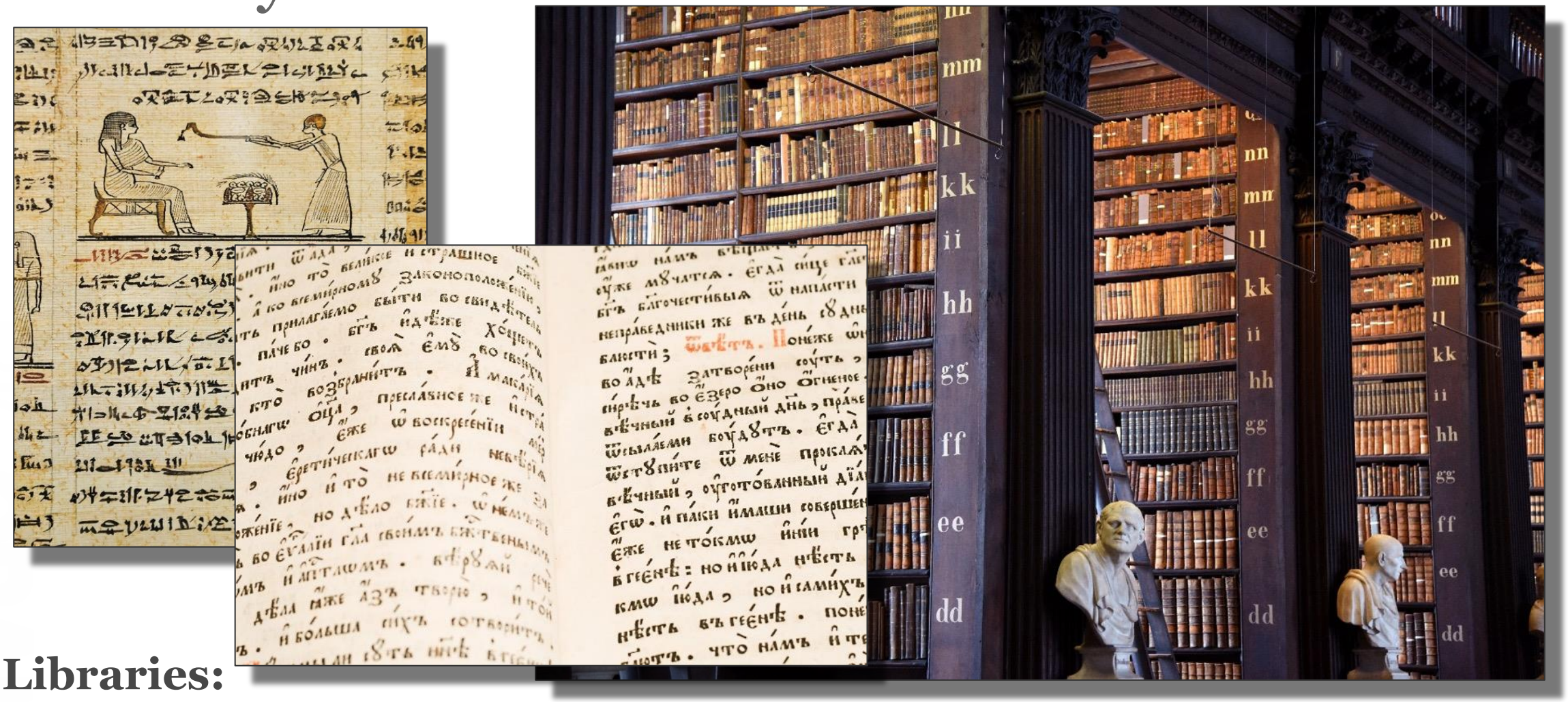
80ies

- 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

What if you want to know more?



Libraries:

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

For the last ~30 years:

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

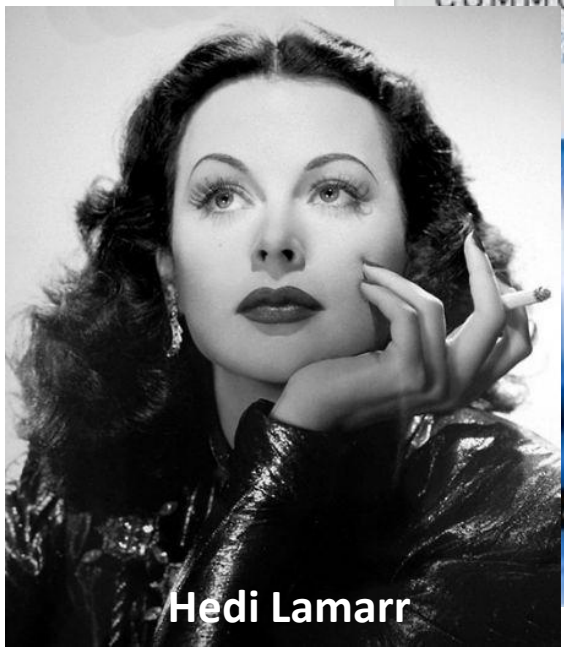
Instrument control (usually limited by manufacturer)

But relatively few changes since then….

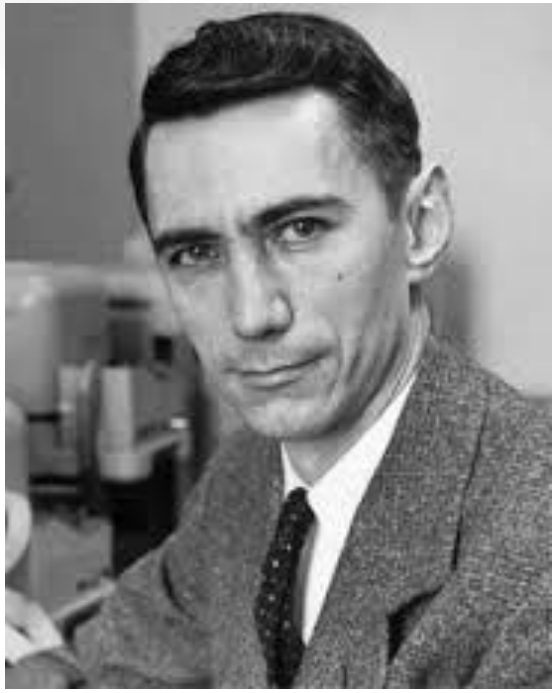
Could anyone have predicted it?

J. C. R. Licklider in 1965. A psycho-acoustician who saw computers as more than calculating machines, he was the first director of ARPA's Information Processing Techniques Office (IPTO). *(Photo courtesy of the MIT Museum)*

Ada Lovelace



Hedi Lamarr



Claude Shannon

Noel Bonnet

What is happening now?

THE UNIVERSITY of TENNESSEE KNOXVILLE

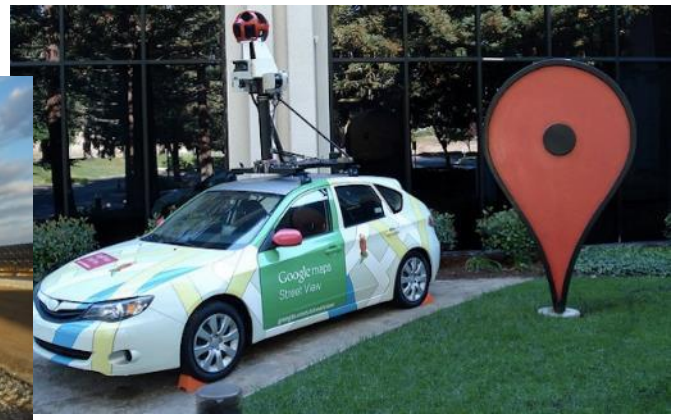
New data technologies

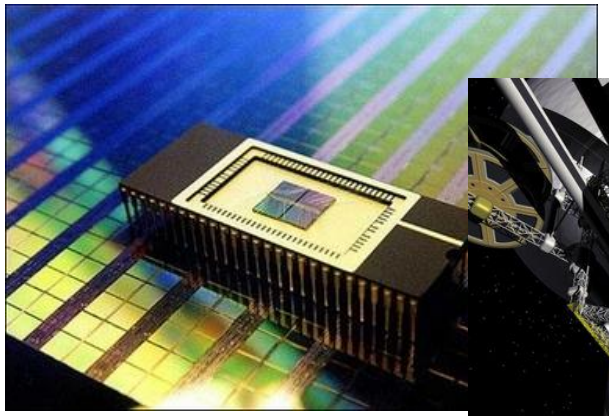
- Searches
- Social networks
- Recommender engines
- Connection to real world
- Large Language models

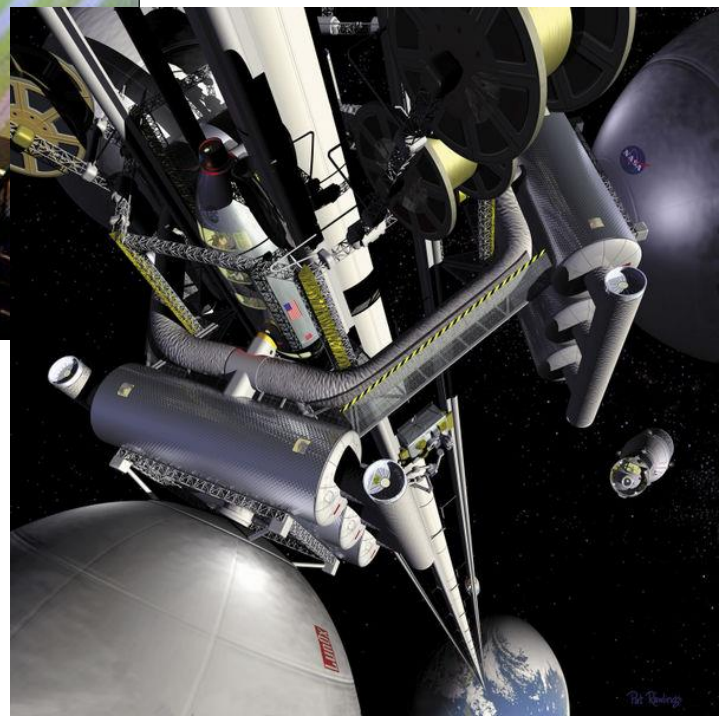
New opportunities:

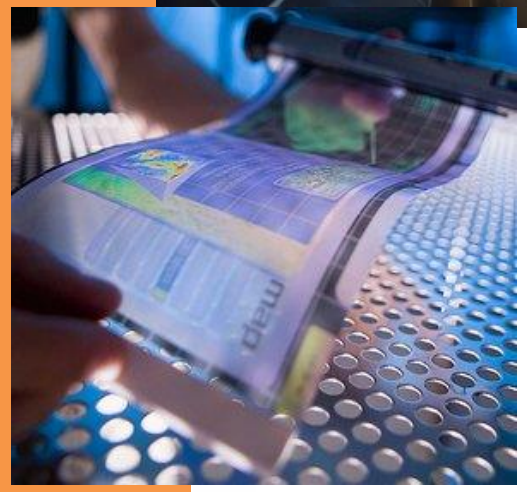
- 3D Printing
- IoT devices
- Laboratory robotics
- Open code
- Text analytics

The 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

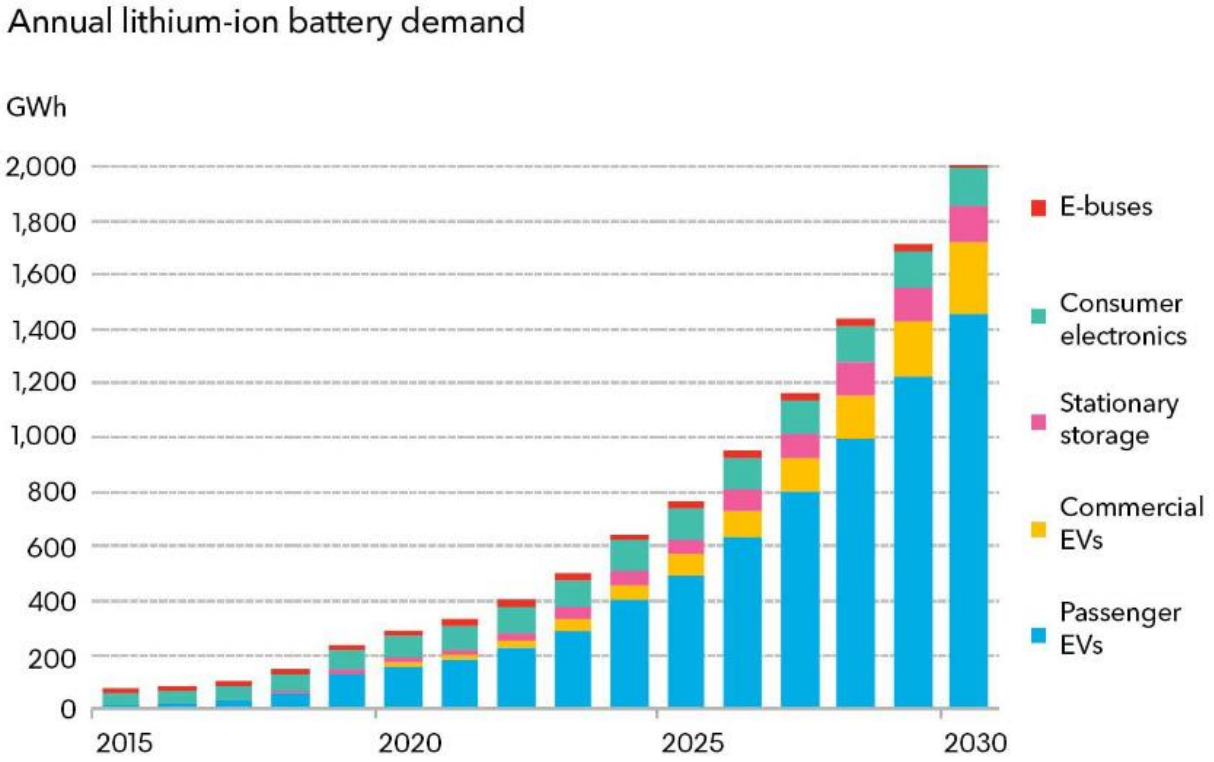
[2]Department of Materials Science and Engineering, University of Wisconsin, Madison, Wisconsin 53706, USA

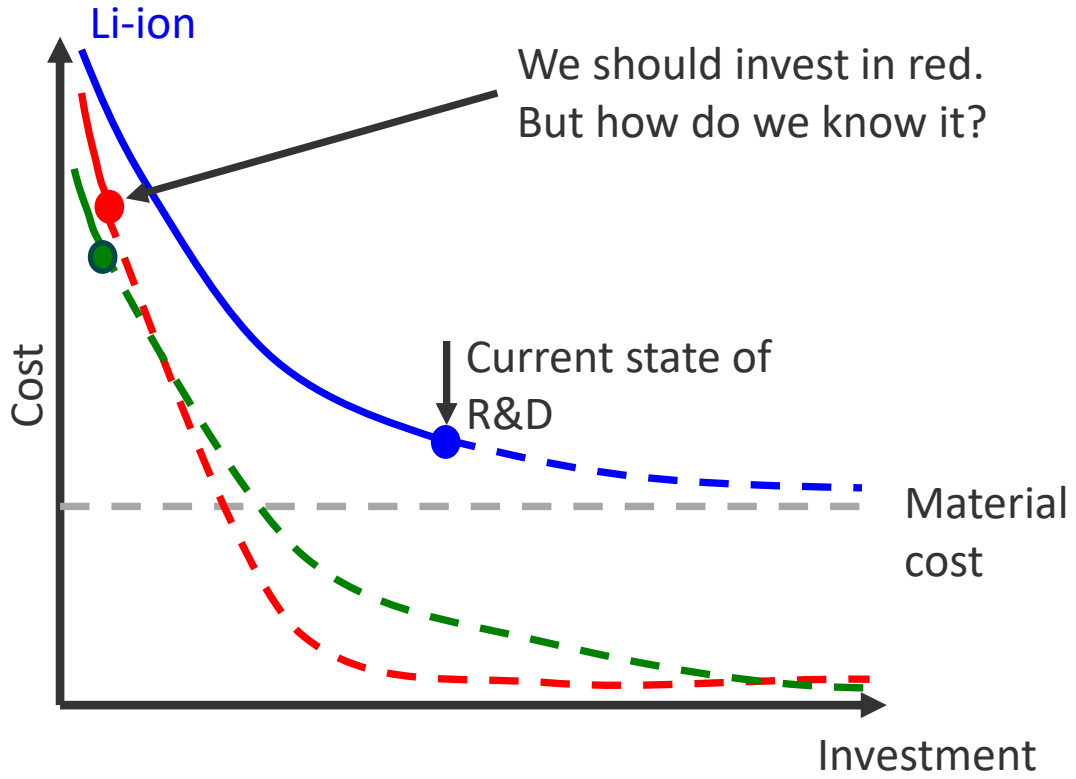
*e-mail: gceder@mit.edu

- **"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

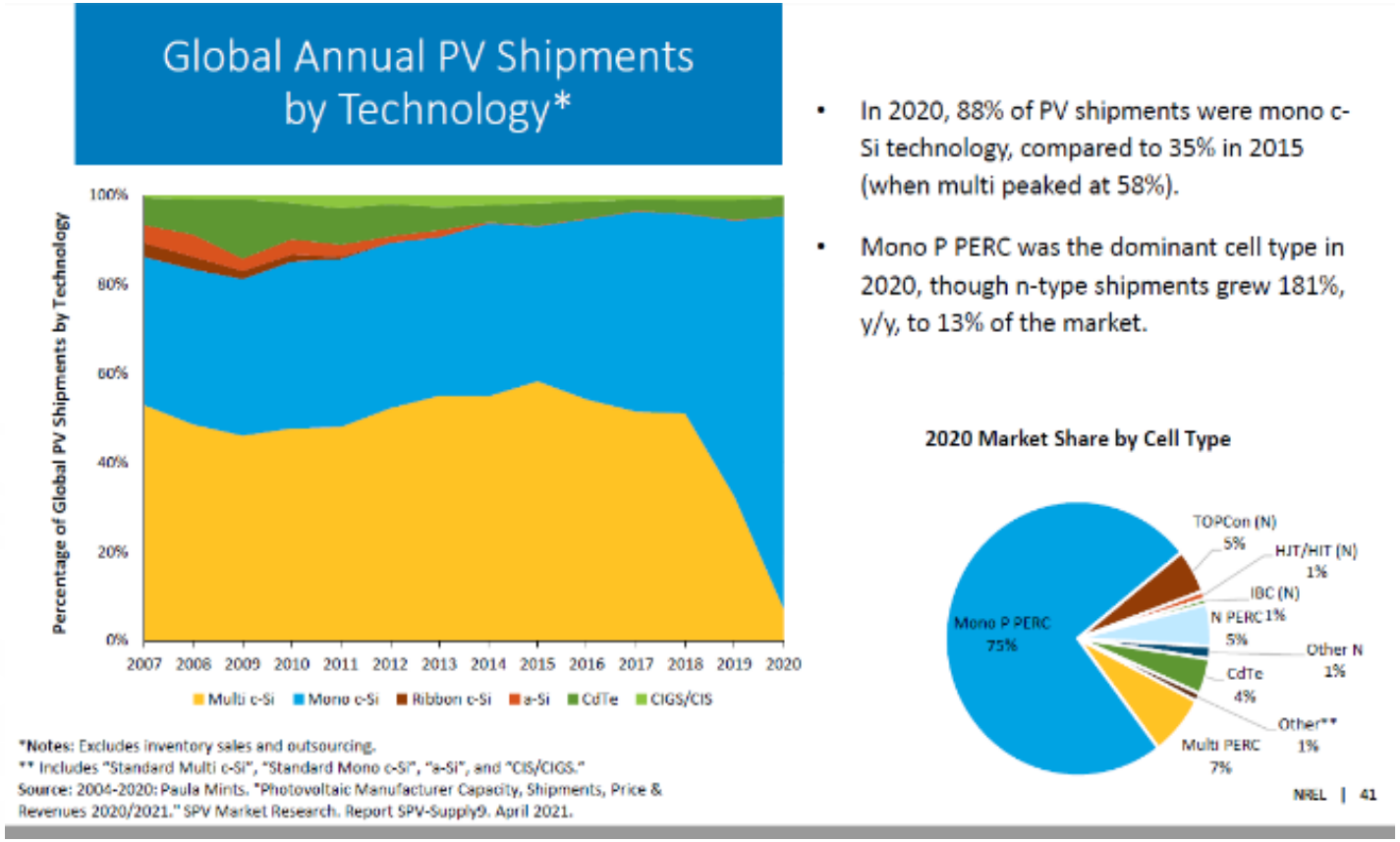




- Batteries are required element of energy transition (EVs, ESS, mobile devices)
- Currently Li-ion is the primary technology
- Optimization of Li-ion batteries takes years (even with same process on new Gigafactory)
- However, it is far from Goldilock zone for ESS or energy transport

- 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?



- In 2020, 88% of PV shipments were mono c-Si technology, compared to 35% in 2015 (when multi peaked at 58%).
- Mono P PERC was the dominant cell type in 2020, though n-type shipments grew 181%, y/y, to 13% of the market.

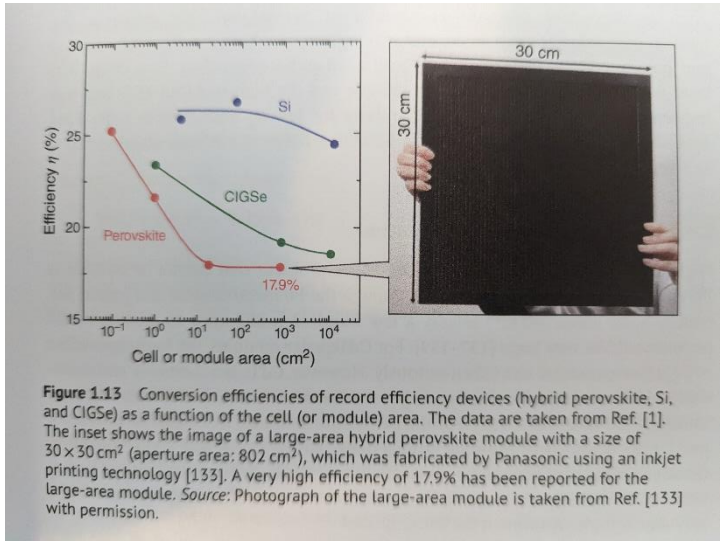
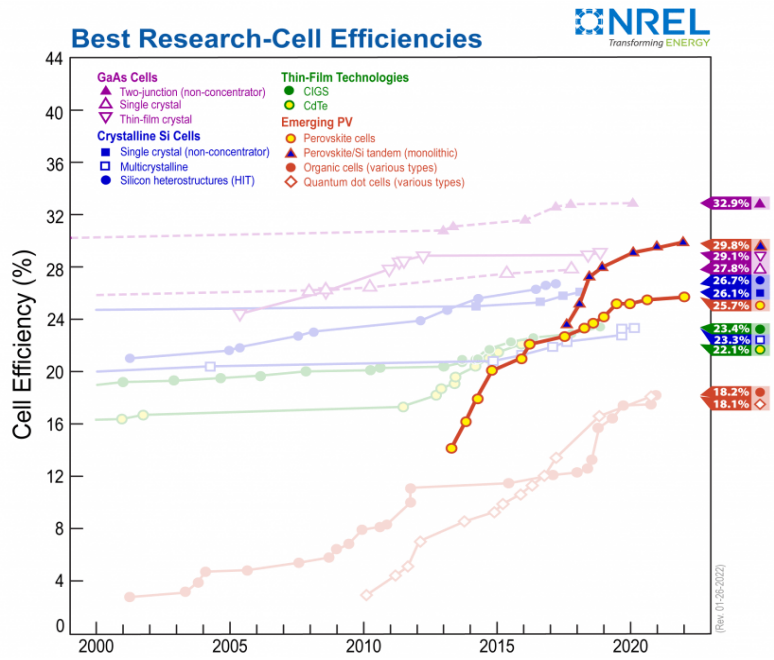
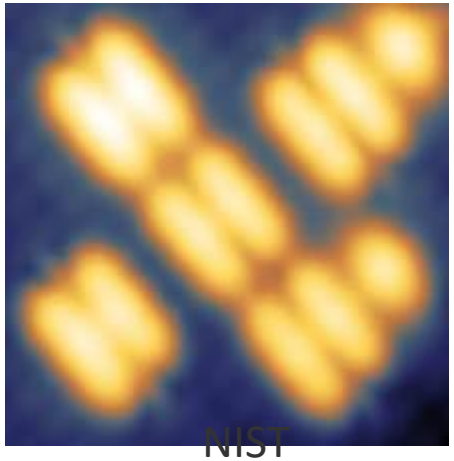


Figure 1.13 Conversion efficiencies of record efficiency devices (hybrid perovskite, Si, and CIGSe) as a function of the cell (or module) area. The data are taken from Ref. [1]. The inset shows the image of a large-area hybrid perovskite module with a size of $30 \times 30\ cm^2$ (aperture area: $802\ cm^2$), 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.



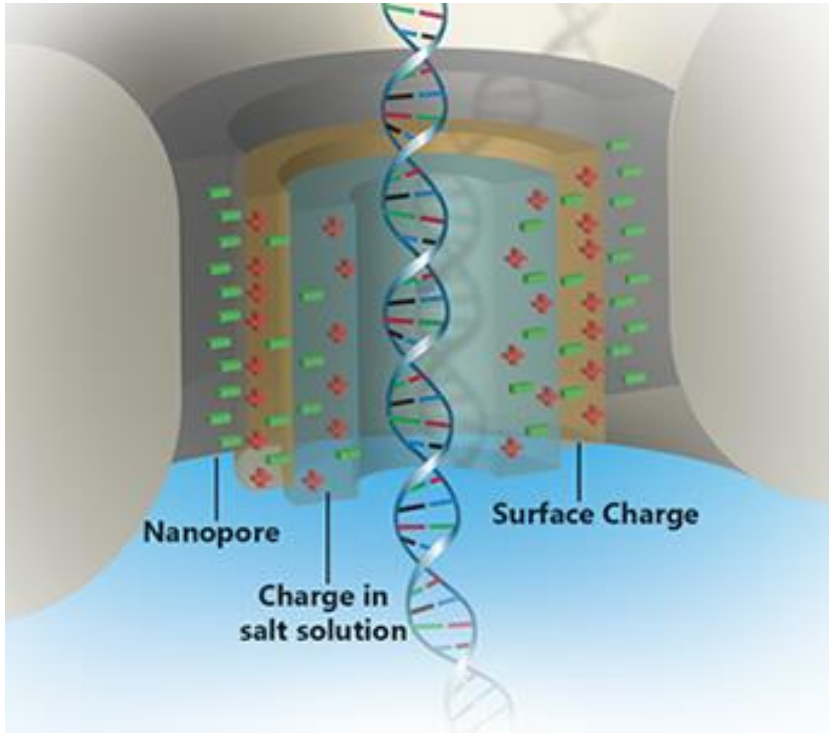
- Solar energy is the fastest growing energy sector
- Si is now reining material – however, it is really not the optimal material for PV (heavy, expensive)!
- Hybrid perovskites can be used as ideal PV materials – if we can make them stable and scale manufacturing!

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

Instrumentation

General synthesis/characterization:

- Multiple data generation tools
- Complex workflows
- 100s at each university in US

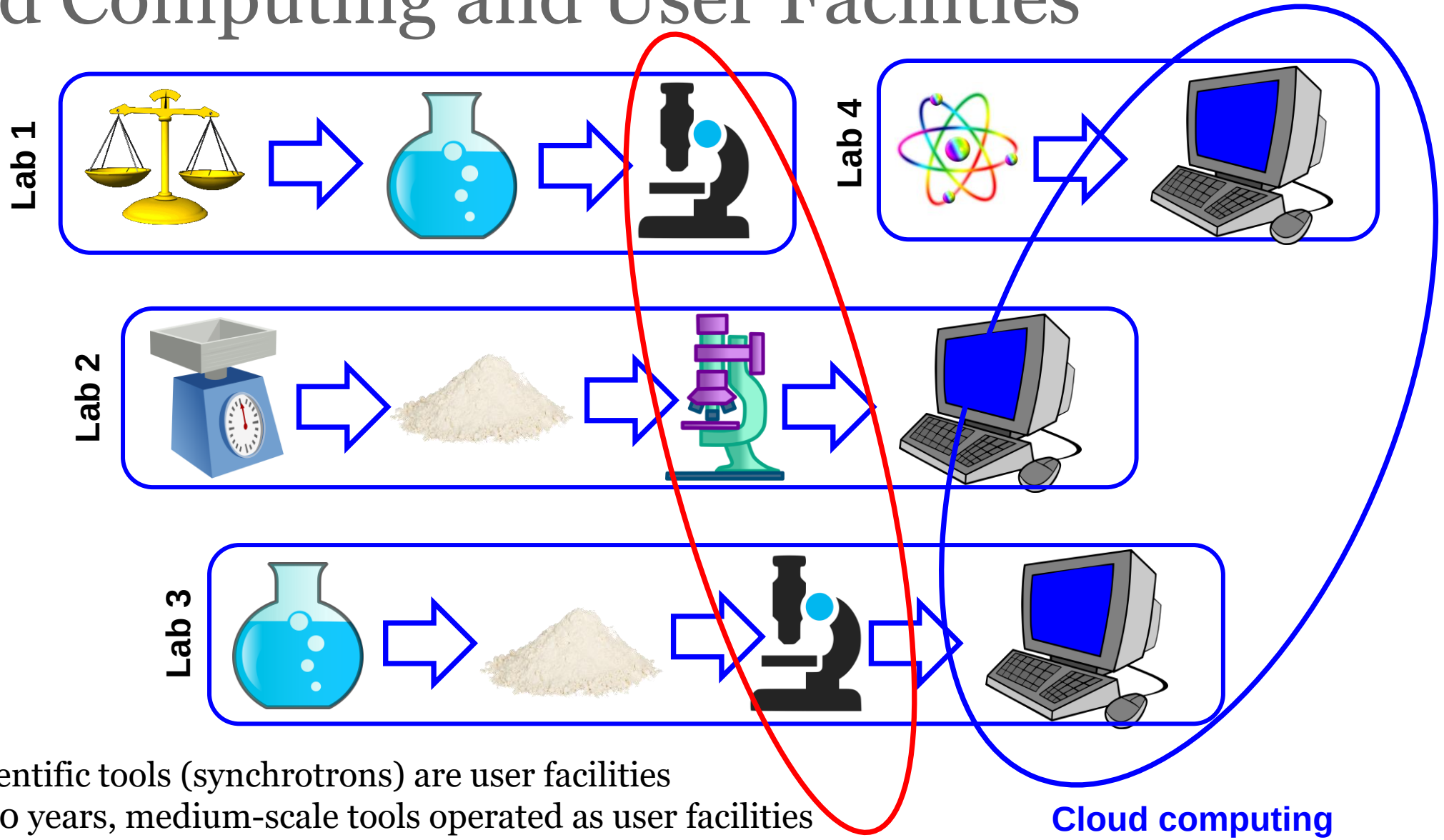
Surface science lab:

- Multiple data generation tools
- Complex controls and workflows
- 10s in each university

Electron microscope:

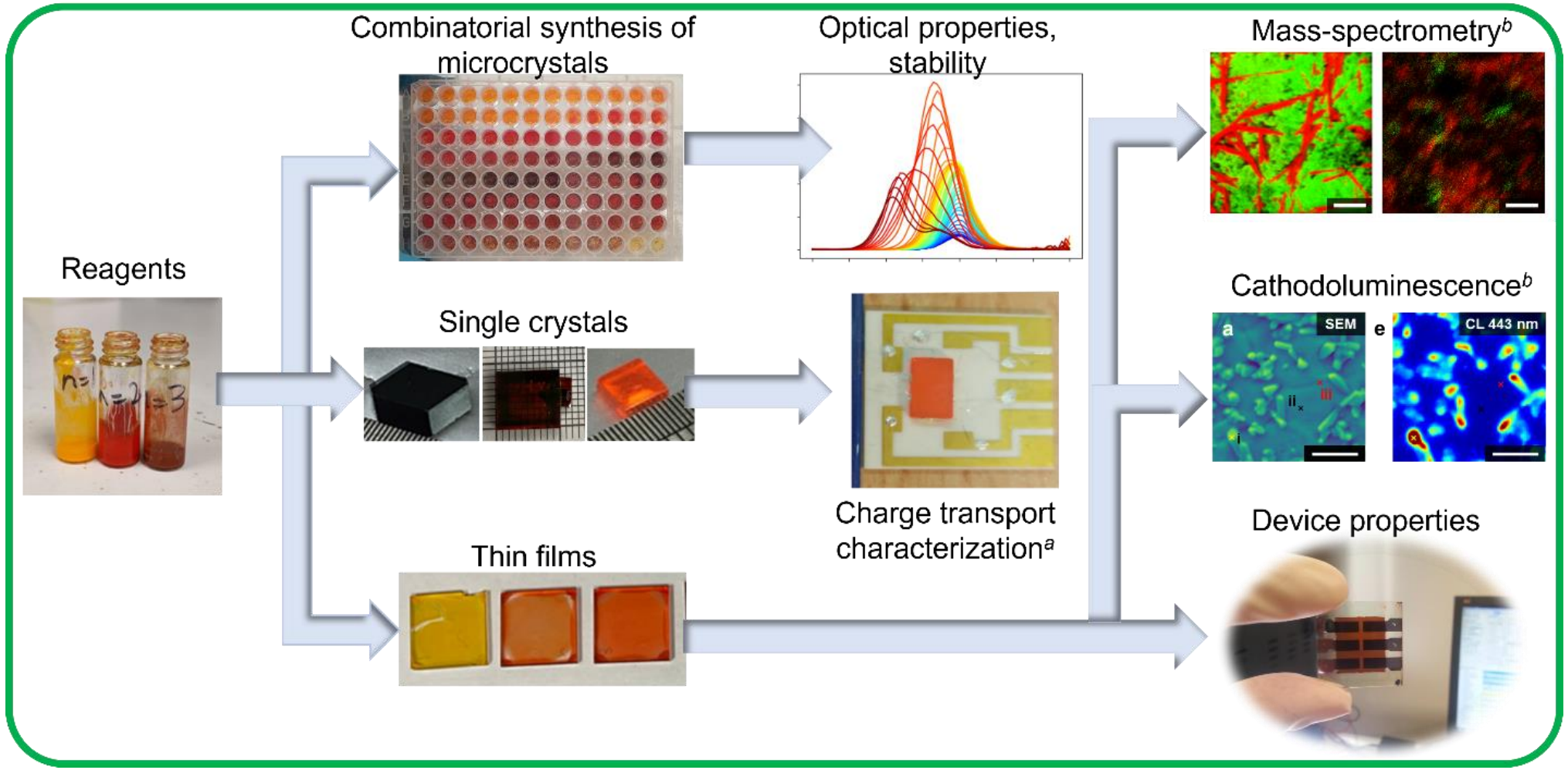
- >100k worldwide
- Can cost up to ~4-5$M
- Can generate data at the ~10GB/s

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?

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?

Course Information

Faculty Contact Information:

Instructor: Prof. Sergei V. Kalinin,
Office: 314 IAMM
E-mail: sergei2@utk.edu
Teaching Assistant: Tommy Wong, cwong13@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_Fall2023

Office Hours:

Friday 1:30 - 3:00 PM are open for 1:1 meetings to discuss any course related item. Can also be made by scheduling via email.

Course Outline

1. Introduction to Materials Science and Machine Learning
2. Understanding the Fundamentals of Machine Learning
3. Exploring Machine Learning in Theory for Models and Property Predictions
4. Applying Machine Learning to Imaging and Characterization
5. Delving into Linear Dimensionality Reduction Methods and Deep Convolutional Neural Networks
6. Unraveling Variational Autoencoders and Generative Models
7. Machine Learning for Process Optimization and Synthesis
8. Machine Learning for Physics Discovery
9. Understanding Large Language Models in Scientific Workflows
10. Learning about Automated Labs

Value Proposition

1. You are interested in ML and AI and would like to try it hands-on on real world problems from materials science and microscopy

2. Learn the basics of the ML methods and build upon this knowledge - from simple principal component analysis to large language models.

3. Explore how ML is being adopted by industry - from IT leaders such as Amazon, Google, and Meta to instrumental, chemical, and materials companies.

4. Learn why next decade of ML will be transition from purely in-silico to real-world materials and device applications, and be a part of this transition.

5. And learn to work backwards from real-world problem to solution.

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.

Outcomes - 1

1. This course aims to provide students with the skills needed to move from data to decisions.

2. Understanding of Materials Science and Machine Learning: Students should expect to gain a solid understanding of materials science, machine learning, and the intersection of the two. They will learn how machine learning can drive the discovery and optimization of materials, which is essential in numerous industries, including technology, manufacturing, energy, and more.

3. Proficiency in Machine Learning Techniques: The course is designed to provide students with basic knowledge of principles of various machine learning techniques, including supervised, unsupervised, and active learning. These skills are essential for careers that involve data analysis, prediction modeling, and artificial intelligence.

Outcomes - 2

1. Practical Application of Machine Learning: Students will learn how to apply machine learning techniques to real-world materials science problems. This includes using these techniques for property prediction, imaging and characterization, and process optimization. These practical applications prepare students for careers where they will need to apply theoretical knowledge to solve practical problems.

2. Experience with Advanced Machine Learning Concepts: The course covers advanced topics such as linear dimensionality reduction methods, deep convolutional neural networks, variational autoencoders, and generative models. This knowledge can help students be at the forefront of the AI field, making them valuable assets to companies investing in these areas.

3. Preparedness for the Future of Science and Industry: With insights into automated labs, large language models in scientific workflows, and federated tools and workflows, students will be prepared for the future of industry. These skills are increasingly important as companies automate processes and incorporate AI into their workflows. This knowledge can help students stand out in the job market and be prepared for the careers of the future.

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:

- Programming Exercises 10%
- Short Quizzes 10%
- Assignments (4+2) 50%
- 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.