

Lecture #1: What is materials informatics + Course Navigation

Goals

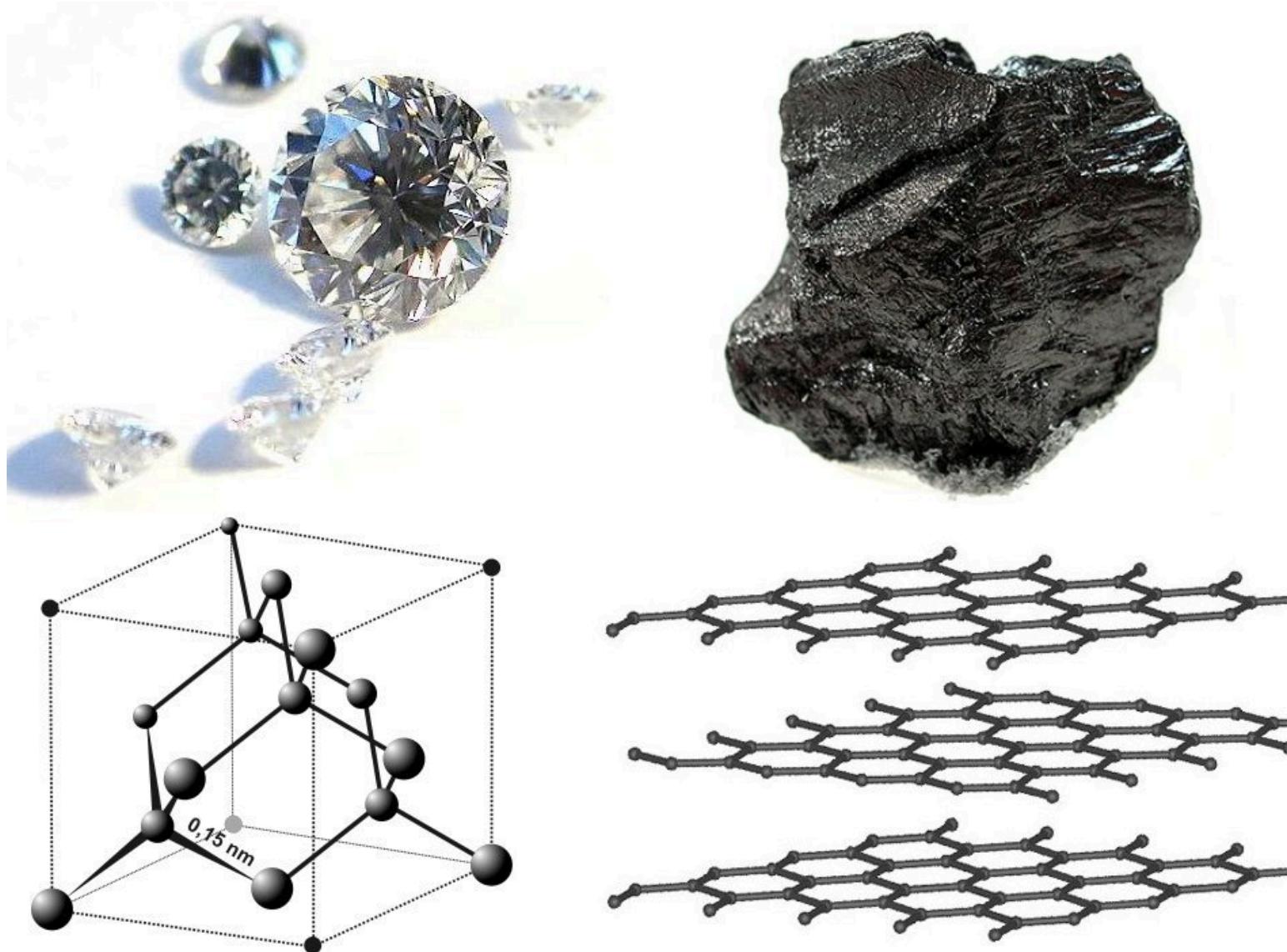
- Provide an understanding of what materials science and materials informatics are
- Describe the course content
- Help students decide if the course is right for them

Agenda

- Goals
- What is materials informatics?
 - Materials science
 - Scales
 - Experiment vs. Theory
 - Materials informatics
 - Examples
- Course navigation
 - ILOs, HW, FP, assessment criteria

Materials science (and engineering)

- studies the relationship between the structure and properties of a material
- and uses this knowledge to improve our lives

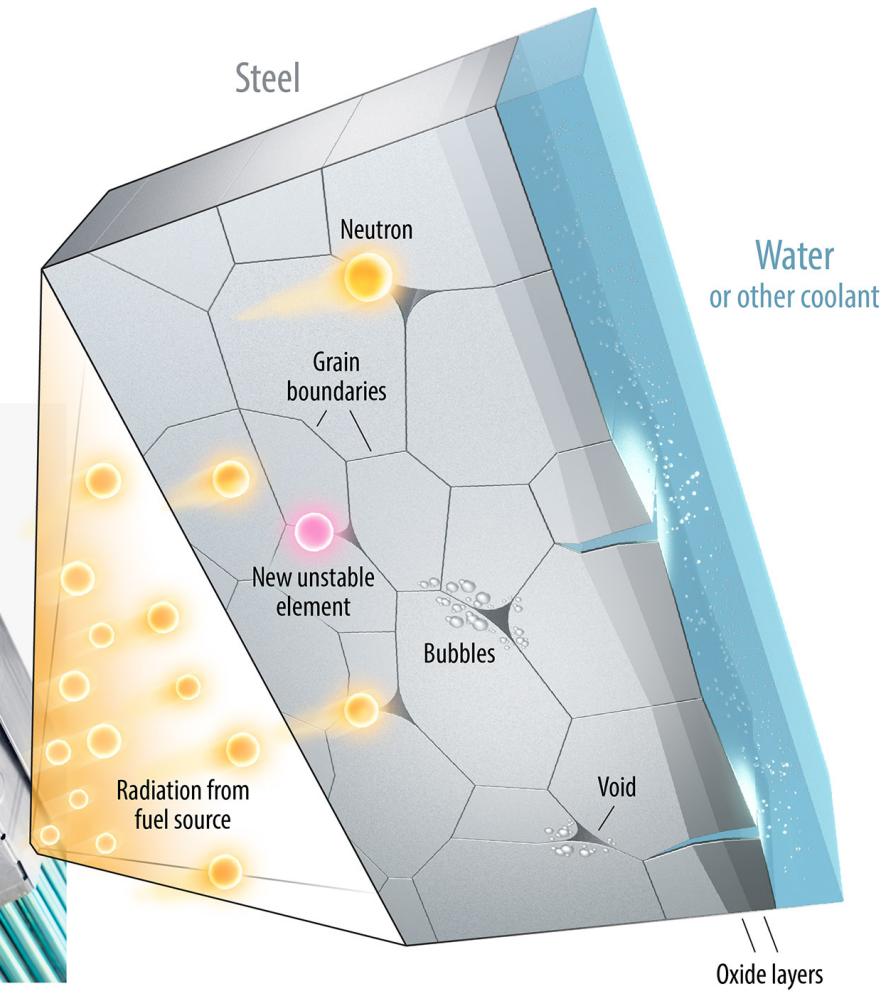


Tasks that materials science solves:

- Design
 - Improve
 - Understand reasons of a failure
- ...

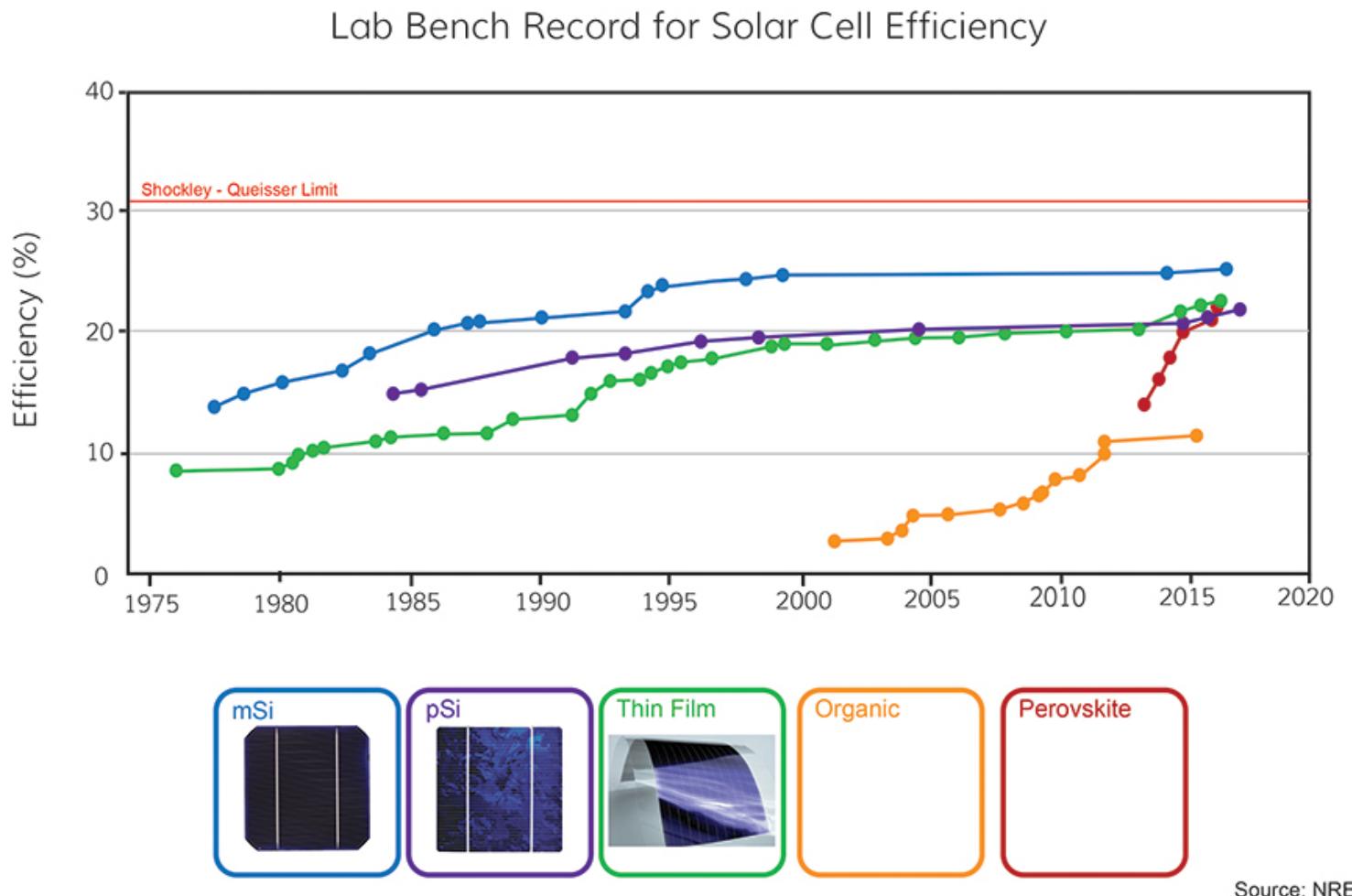
Design

- New material
- With specified properties
- Stable at given conditions



Improve

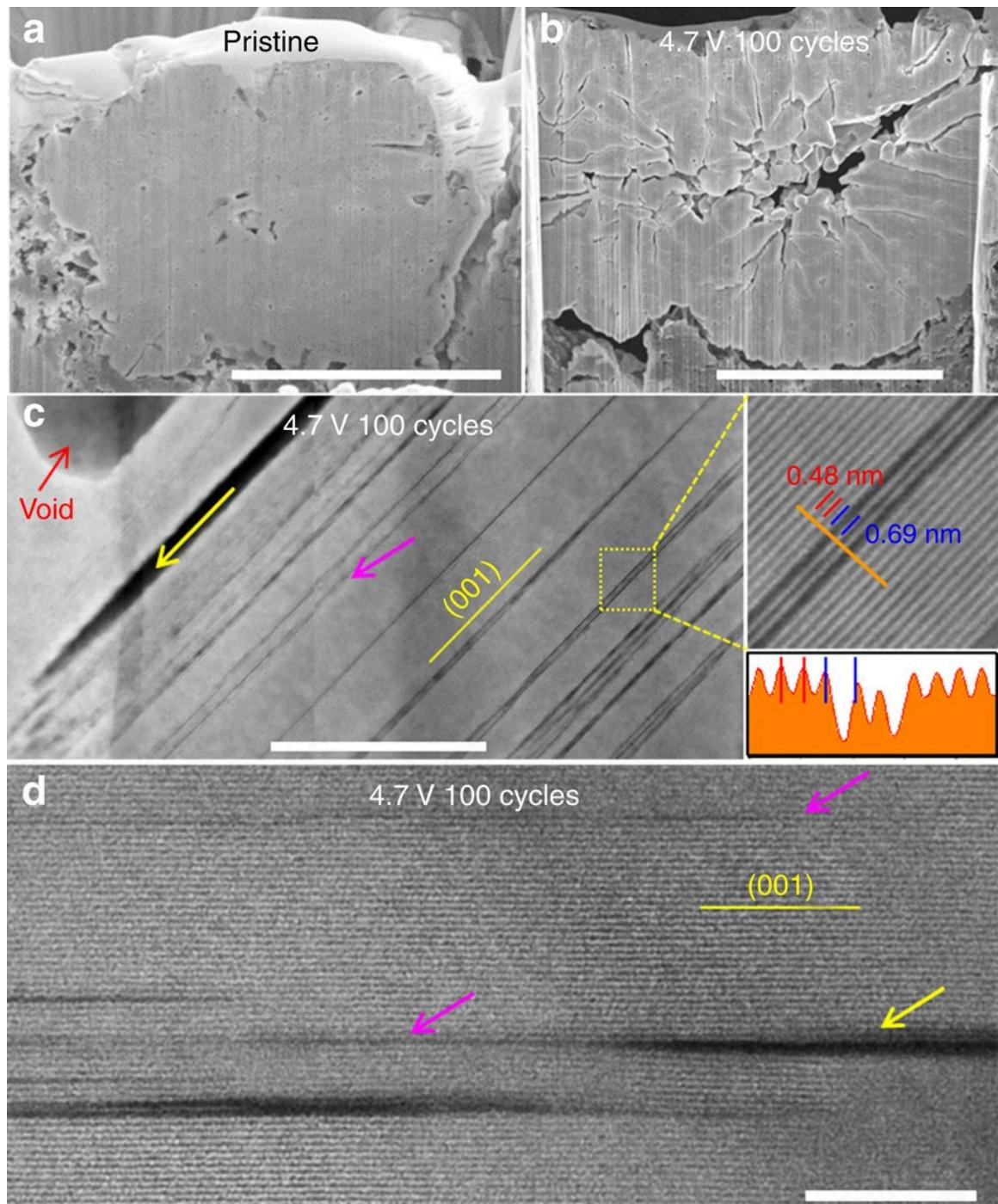
- Better props
- Less expensive
- Easier to process



Understand reasons of a failure

- Why?
- When?
- Can be fixed?

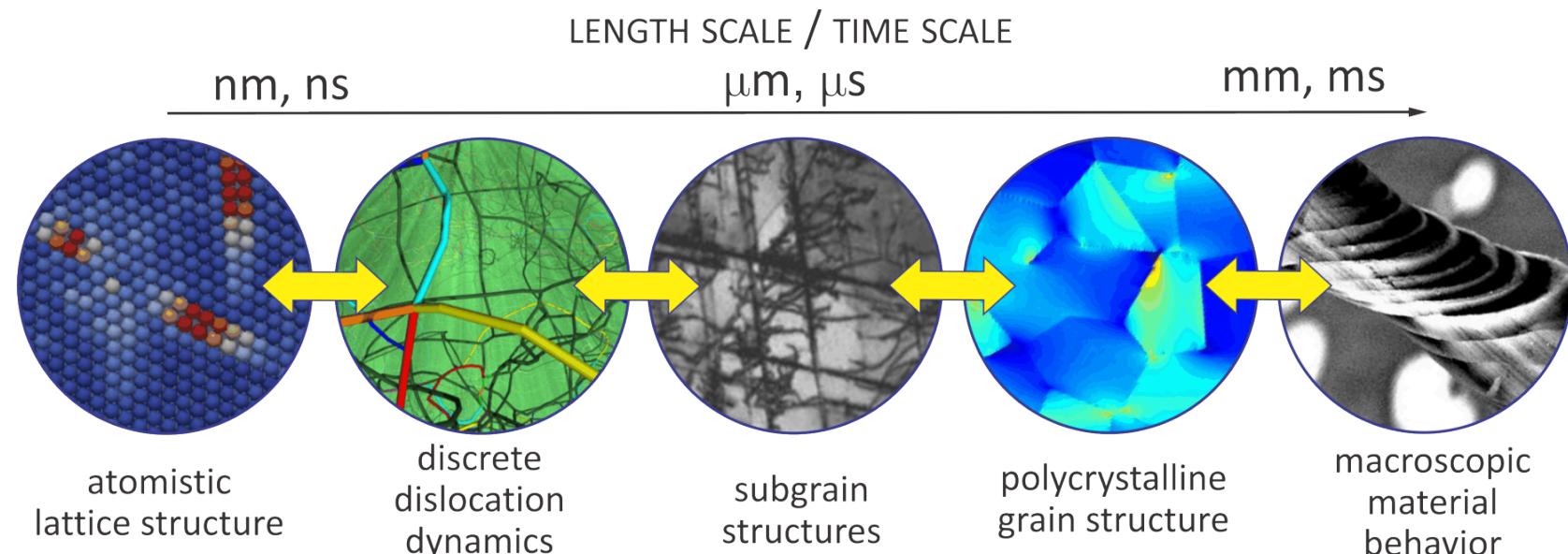
Intragranular cracking as a critical barrier for high-voltage usage of layer-structured cathode for lithium-ion batteries



The materials are characterized

...at various length scales

- atomic
- micro
- meso
- macro



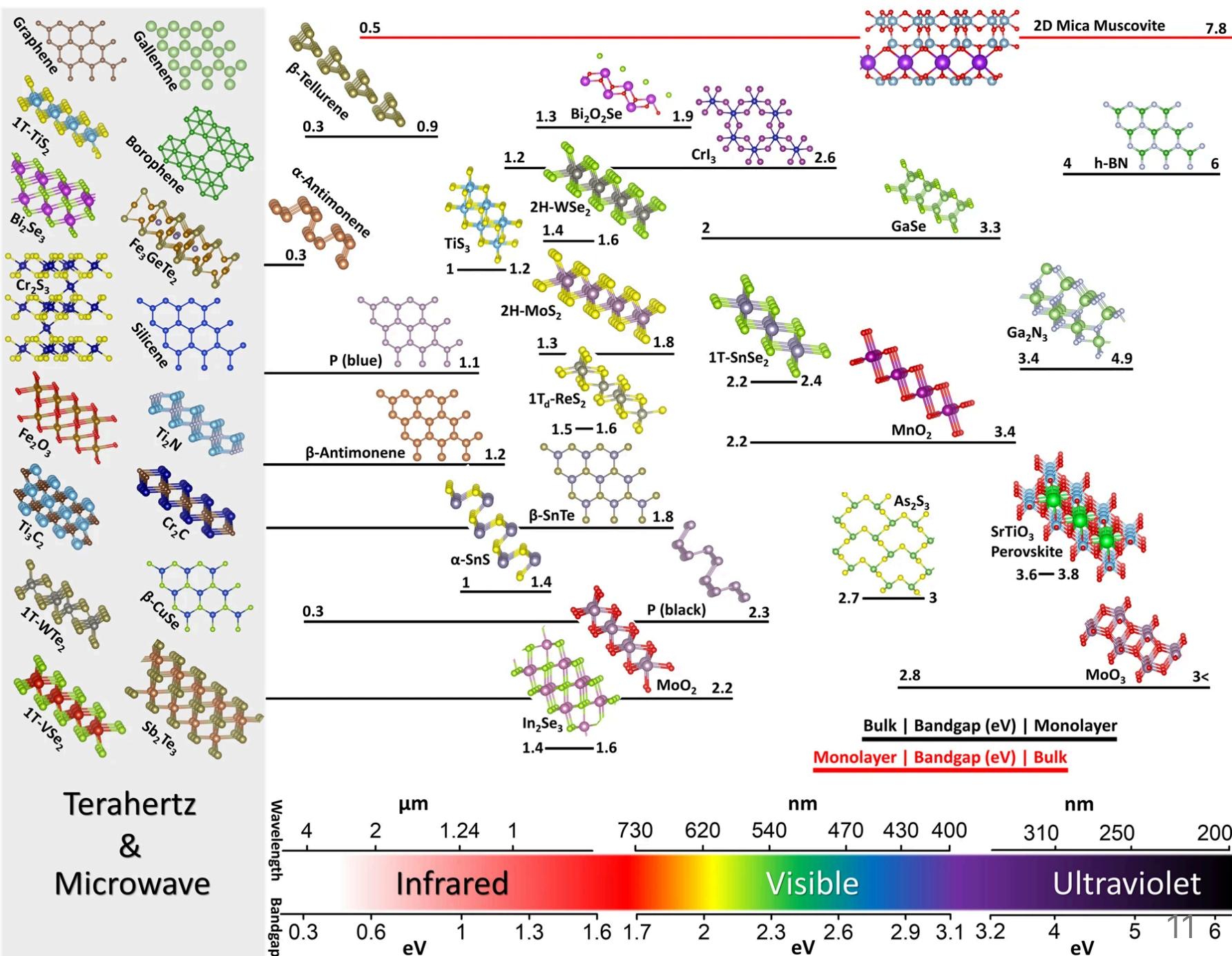
The course

...focuses on the atomic scale

...and inorganic materials

Bandgap engineering of two-dimensional semiconductor materials

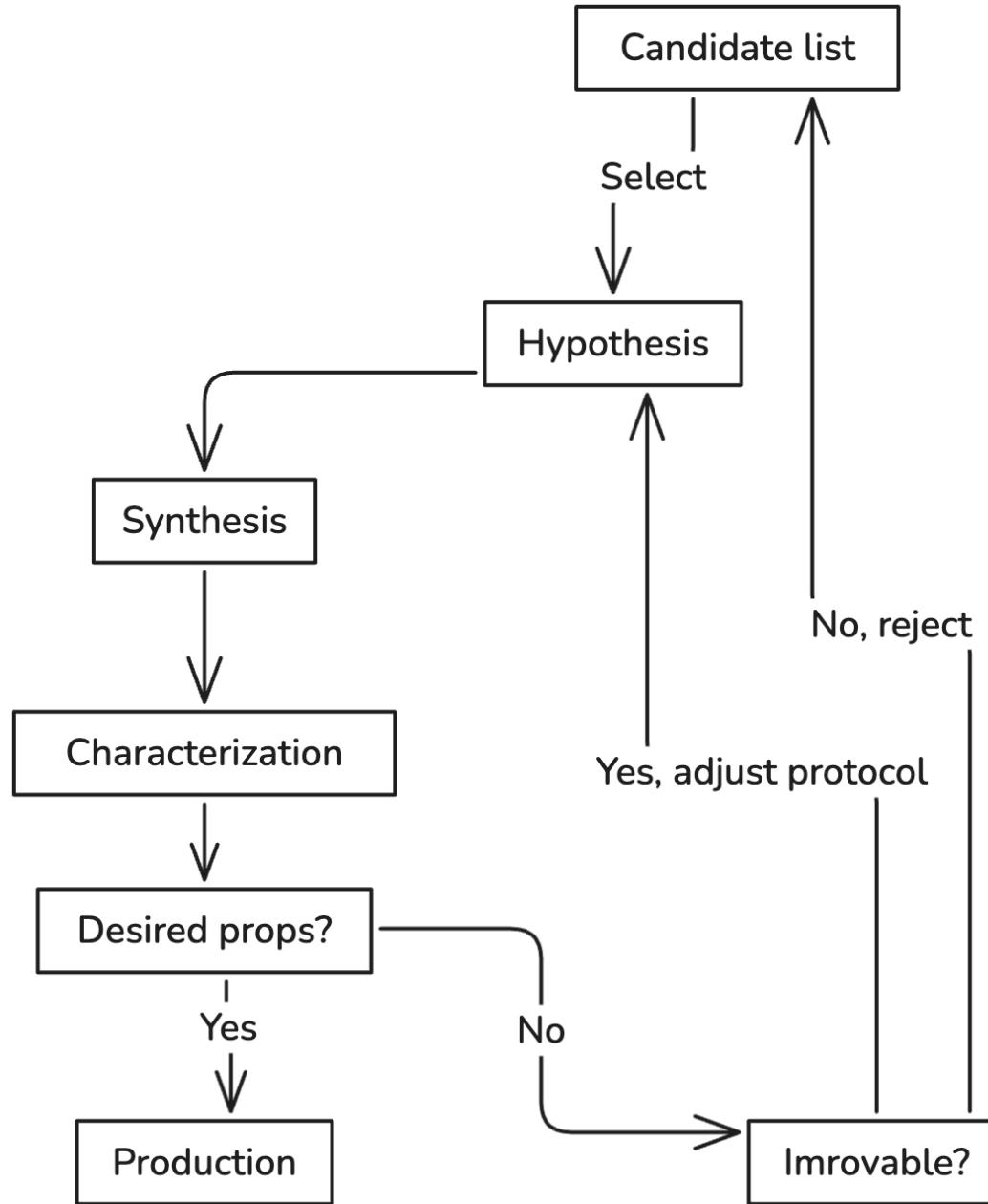
Terahertz
&
Microwave



Trial-and-error in materials science

...is the only way to reach the goal

- Takes ~10 years to design a material



Example: Stainless steels

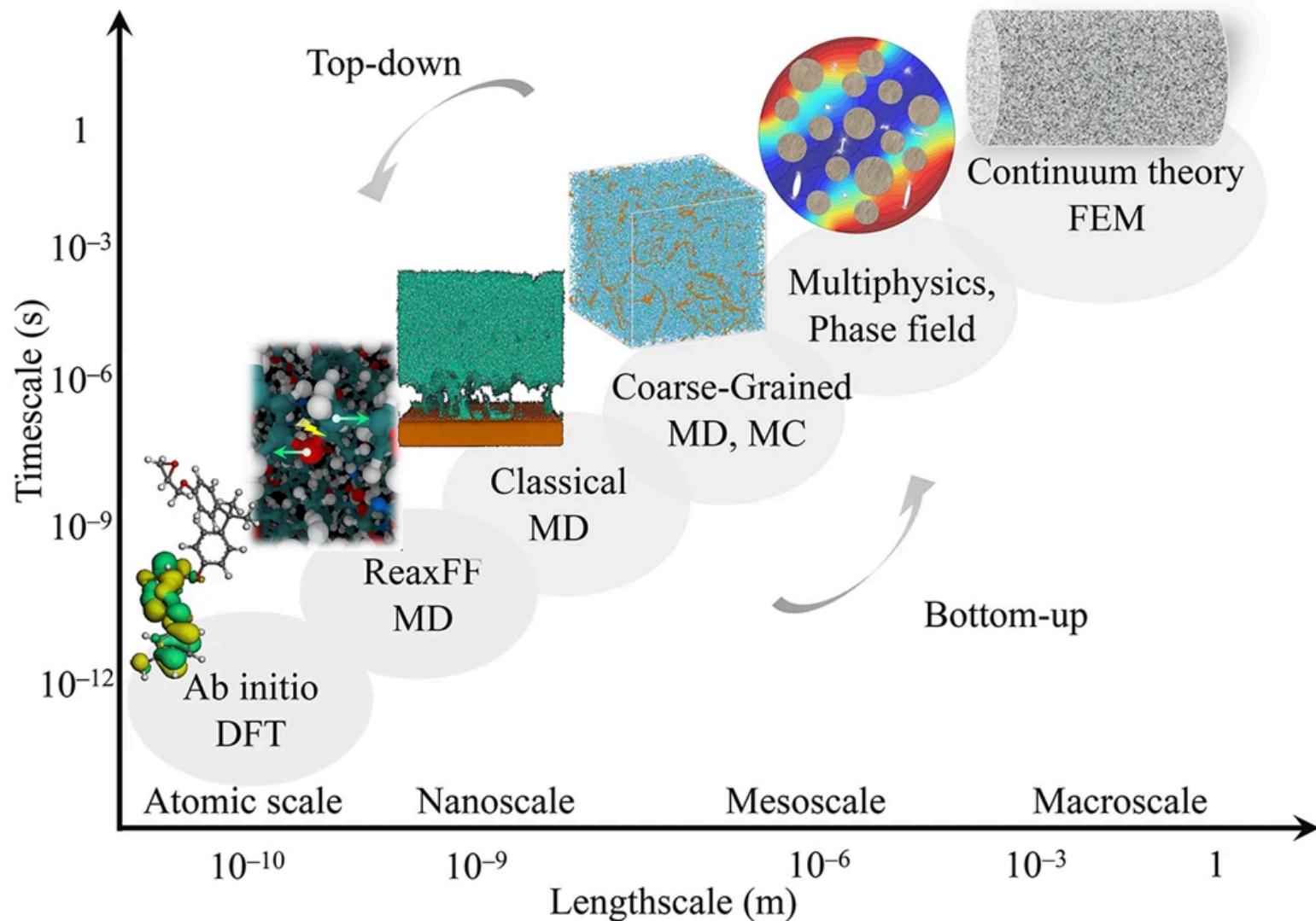
- 1798 - chromium discovery
- 1821 - iron-chromium alloys
- 1861 - the first patent on chromium steels
- 1911 - report on the relationship between the chromium content and corrosion resistance



Trial and error loop can be optimized

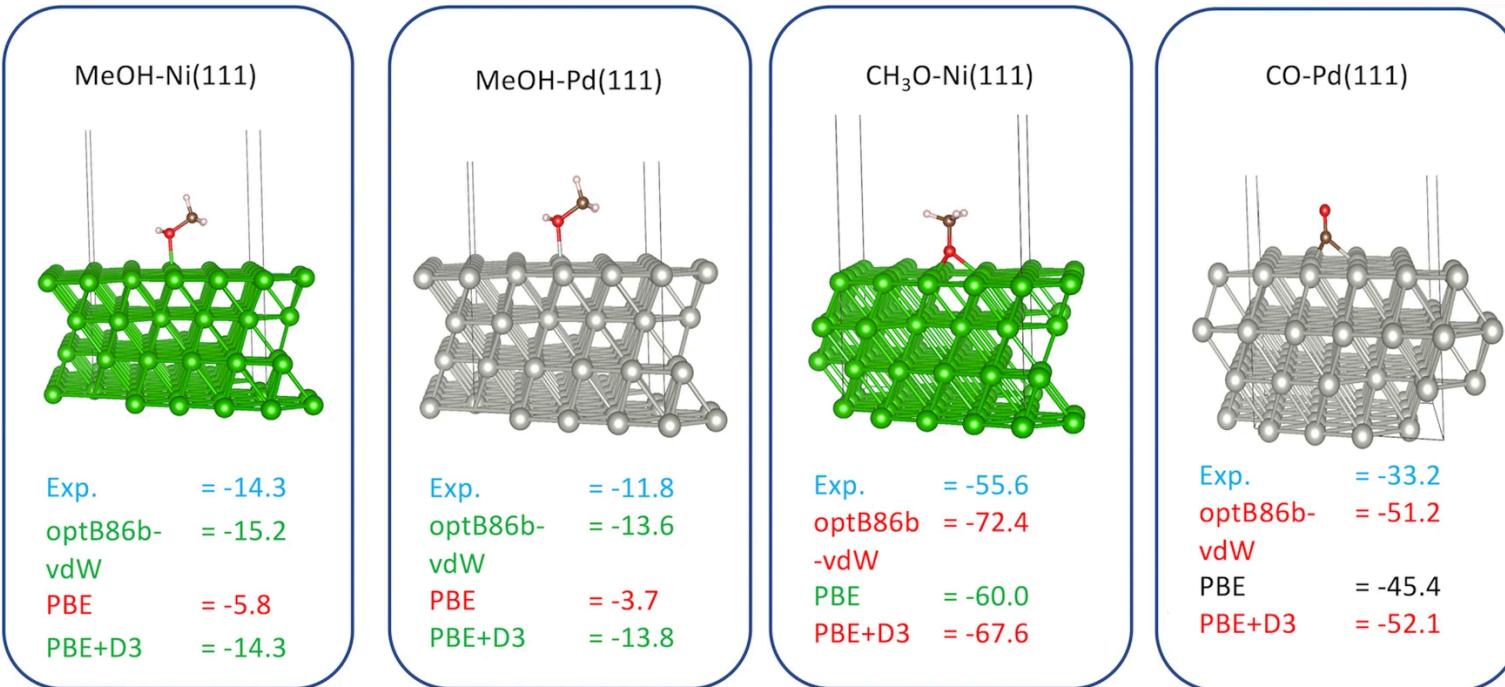
...with the help of theoretical and computational approaches

...by replacing "expensive" experiments with computation



Example: Adsorption energy

- Determines catalytic properties
- Essential for chemical industry



Adsorption energies on transition metal surfaces: towards an accurate and balanced description **accurate and balanced description**

Computational methods

...accelerates the first stage
of the materials design

High-throughput computational-experimental screening
protocol for the discovery of bimetallic catalysts [discovery](#)
[of bimetallic catalysts](#)

Binary systems with 1:1 composition

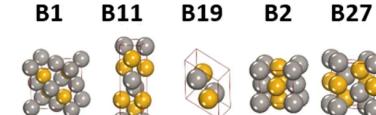
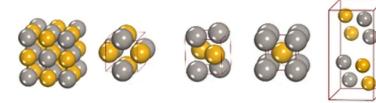
$$^{30}\text{C}_2 = 435 \text{ cases}$$

21	22	23	24	25	26	27	28	29	30
Sc	Ti	V	Cr	Mn	Fe	Co	Ni	Cu	Zn
39	40	41	42	43	44	45	46	47	48
Y	Zr	Nb	Mo	Tc	Ru	Rh	Pd	Ag	Cd

71	72	73	74	75	76	77	78	79	80
Lu	Hf	Ta	W	Re	Os	Ir	Pt	Au	Hg

Data collection

10 crystal structures



4350 candidates

Thermodynamics stability

$\Delta E_f < 0.1 \text{ eV}$

< 250 candidates

DOS similarity

$\Delta \text{DOS}_{2-1} < 2.0$

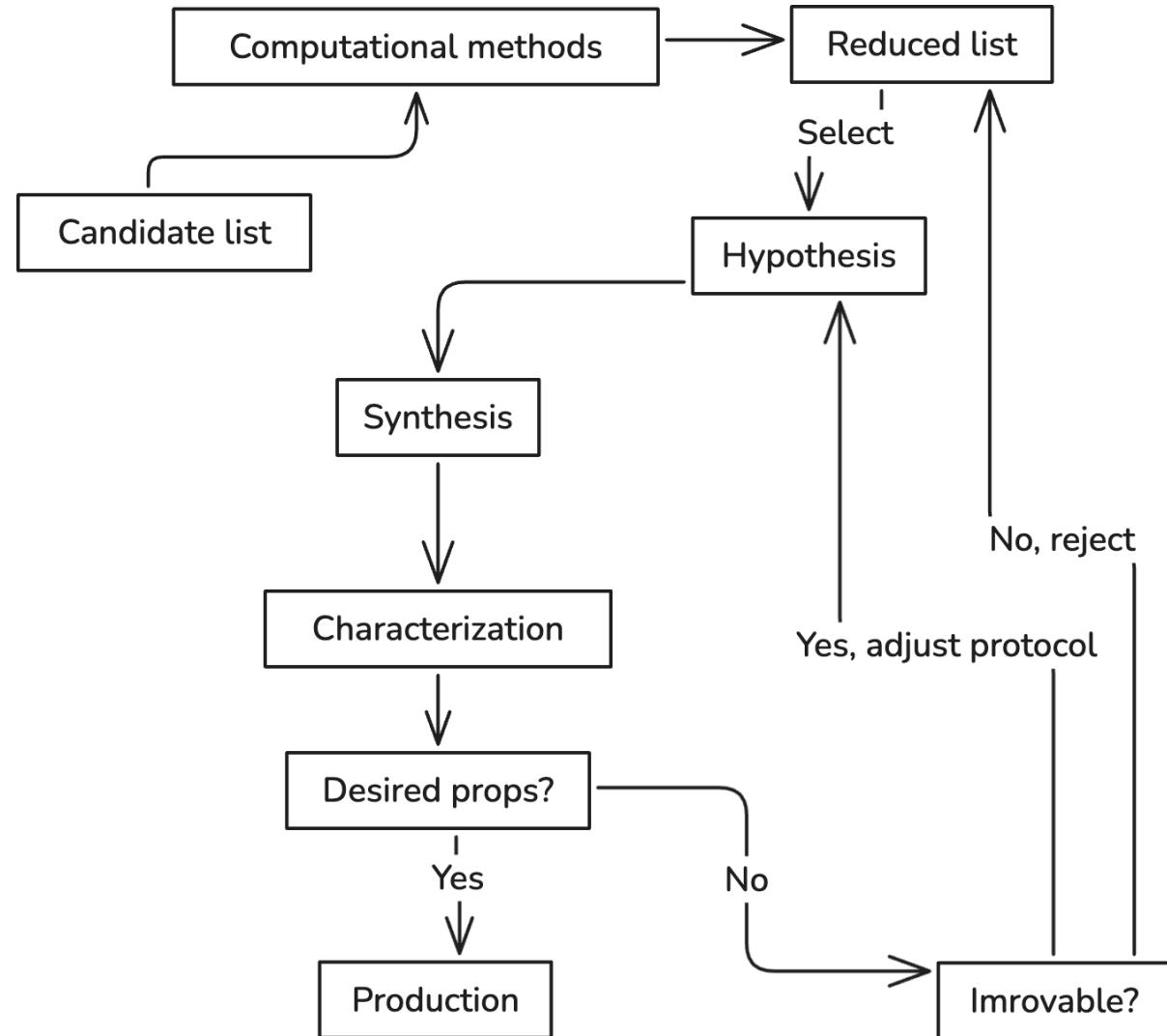
< 20 candidates

Economical?
Feasibility?

Promising catalyst < 10 candidates

Trial-and-error with computational methods

- Takes ~2 years to design a material

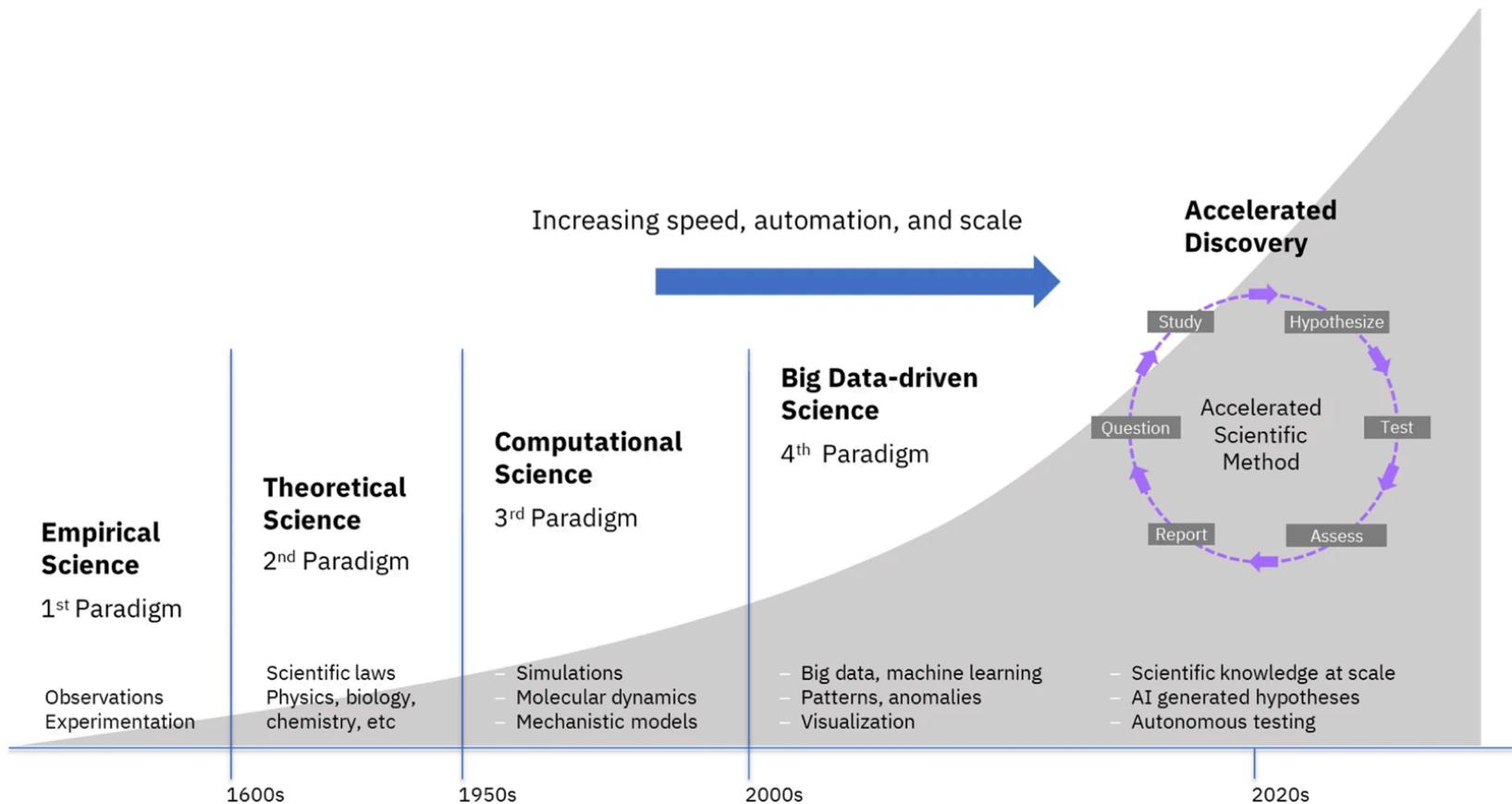


Can it be further optimized?

Materials informatics

...accelerates the solution of materials science problems through the use of data

...mostly by replacing computational methods with surrogate models



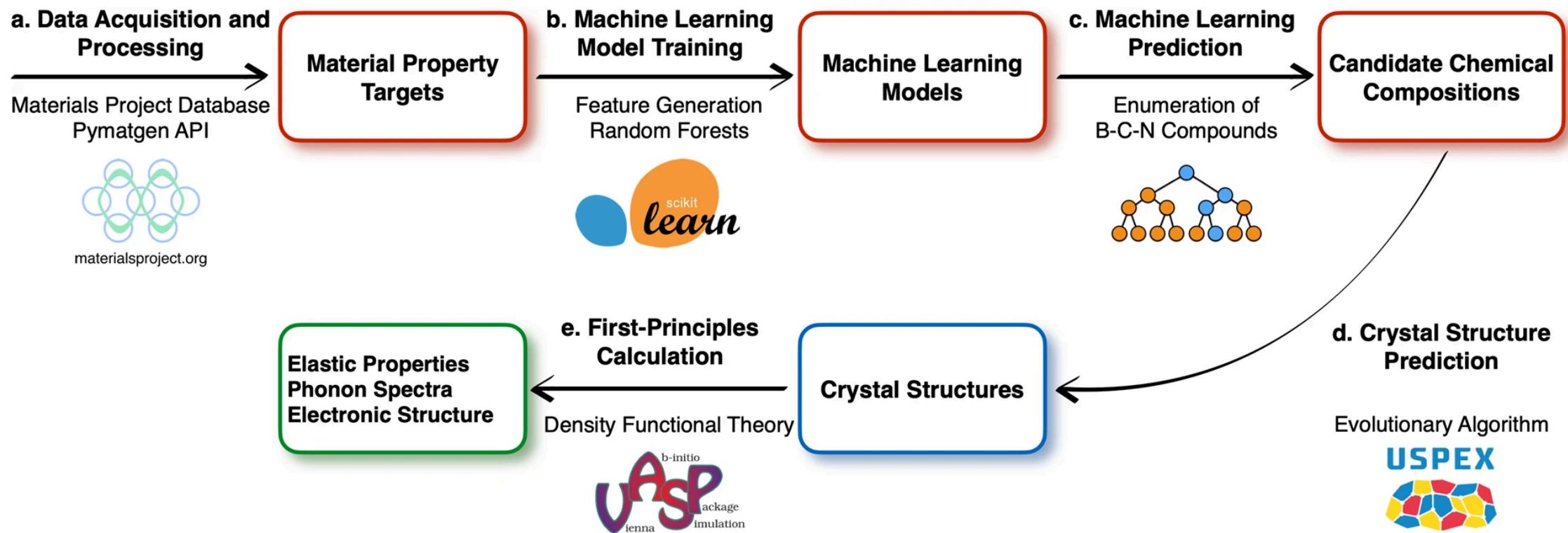
Accelerating materials discovery using artificial intelligence, high performance computing and robotics

"A surrogate model is an engineering method used when an outcome of interest cannot be easily measured or computed, so an approximate mathematical model of the outcome is used instead."

https://en.wikipedia.org/wiki/Surrogate_model

A typical workflow

collect data -> train -> predict -> reduce candidate list -> computational methods -> reduce candidate list -> experiment

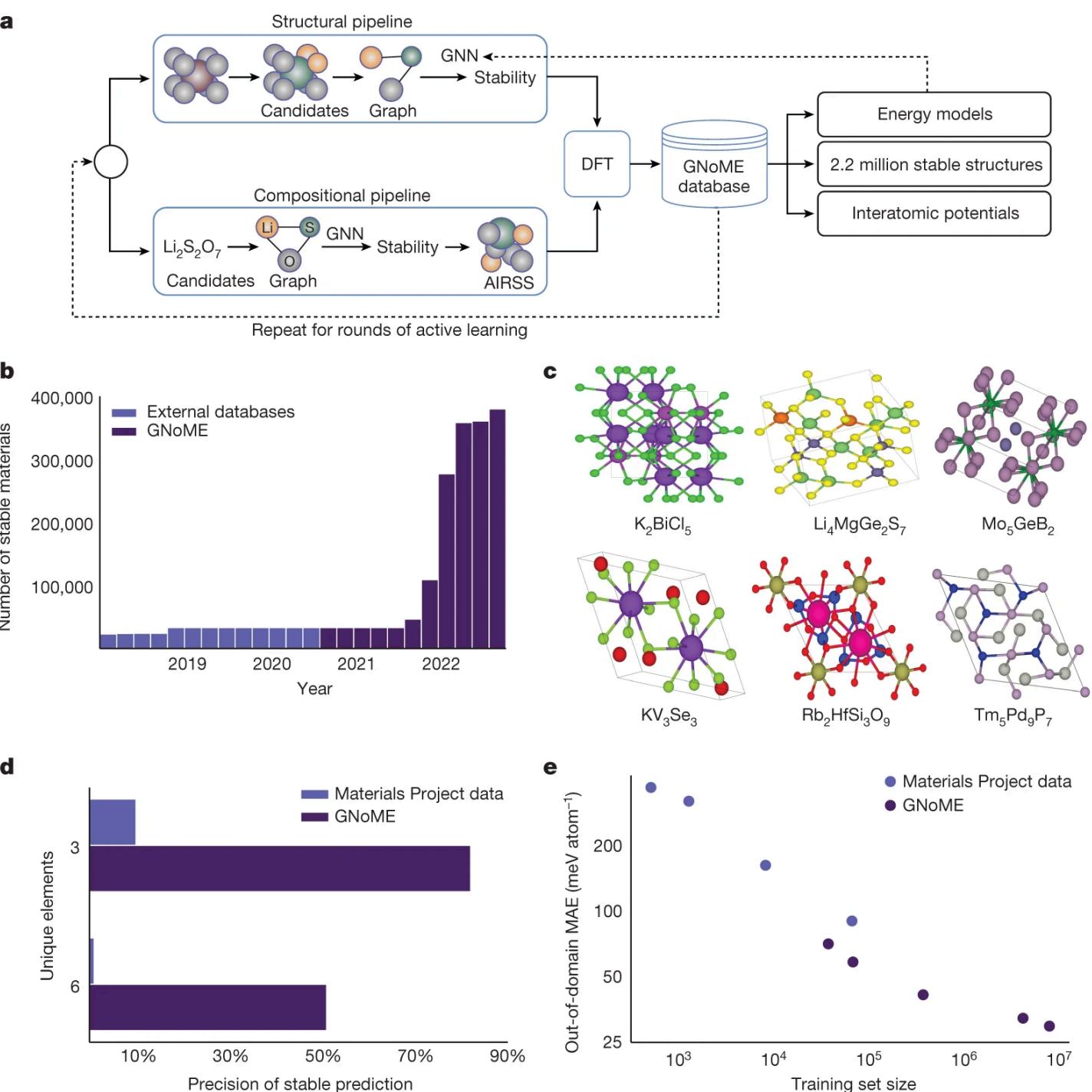


Scaling deep learning for materials discovery

Statement

- 381,000 new stable materials
- 736 structures have been independently experimentally verified

Scaling deep learning for **materials** discovery



Reality

Domain knowledge is essential

- "a very poor fit"
- a different CIF has been used for refinement compared with that in the paper and in the Materials Project
- no evidence for order, and a known disordered version of the compound exists
- the compound is correctly identified but is already reported"

Claimed Phases	1	2	3	4	Claimed Phases	1	2	3	4
$\text{Ba}_2\text{ZrSnO}_6$	X	X	X		$\text{Mg}_3\text{MnNi}_3\text{O}_8$	X		X	
$\text{Ba}_6\text{Na}_2\text{Ta}_2\text{V}_2\text{O}_{17}$	X		X		Mg_3NiO_4		X	X	
$\text{Ba}_6\text{Na}_2\text{V}_2\text{Sb}_2\text{O}_{17}$	X				MgCuP_2O_7		X	X	
$\text{CaCo}(\text{PO}_3)_4$			X		$\text{MgNi}(\text{PO}_3)_4$	X		X	
$\text{CaFe}_2\text{P}_2\text{O}_9$					$\text{MgTi}_2\text{NiO}_6$			X	
$\text{CaMn}(\text{PO}_3)_4$			X		$\text{MgTi}_4(\text{PO}_4)_6$				X
$\text{CaNi}(\text{PO}_3)_4$			X		$\text{MgV}_4\text{Cu}_3\text{O}_{14}$	X	X	X	
$\text{FeSb}_3\text{Pb}_4\text{O}_{13}$			X		Mn_2VPO_7	X		X	
$\text{Hf}_2\text{Sb}_2\text{Pb}_4\text{O}_{13}$			X		$\text{Mn}_4\text{Zn}_3(\text{NiO}_6)_2$			X	
$\text{InSb}_3\text{Pb}_4\text{O}_{13}$			X		MnAgO_2	X			X
$\text{K}_2\text{TiCr}(\text{PO}_4)_3$			X		$\text{Na}_3\text{Ca}_{18}\text{Fe}(\text{PO}_4)_{14}$	X			
$\text{K}_4\text{MgFe}_3(\text{PO}_4)_5$	X				$\text{Na}_7\text{Mg}_7\text{Fe}_5(\text{PO}_4)_{12}$	X			
$\text{K}_4\text{TiSn}_3(\text{PO}_5)_4$	X				$\text{NaCaMgFe}(\text{SiO}_3)_4$		X	X	
KBaPrWO_6	X				$\text{NaMnFe}(\text{PO}_4)_2$	X			
KMn_3O_6	X	X	X		$\text{Sn}_2\text{Sb}_2\text{Pb}_4\text{O}_{13}$			X	
$\text{KNaP}_6(\text{PbO}_3)_8$	X	X	X		$\text{Y}_3\text{In}_2\text{Ga}_3\text{O}_{12}$	X			X
$\text{KNaTi}_2(\text{PO}_5)_2$			X		$\text{Zn}_2\text{Cr}_3\text{FeO}_8$			X	
$\text{KPr}_9(\text{Si}_3\text{O}_{13})_2$	X	X			$\text{Zr}_2\text{Sb}_2\text{Pb}_4\text{O}_{13}$			X	

Lecture #1: Course motivation and navigation

Course structure

- Each class = 1h lecture + 2h seminar
- 11 lectures
- 10 seminars
- 3 HWs
- 1 FP
- 2 FP peer-reviews
- ~10 lecture/seminar recap quizzes

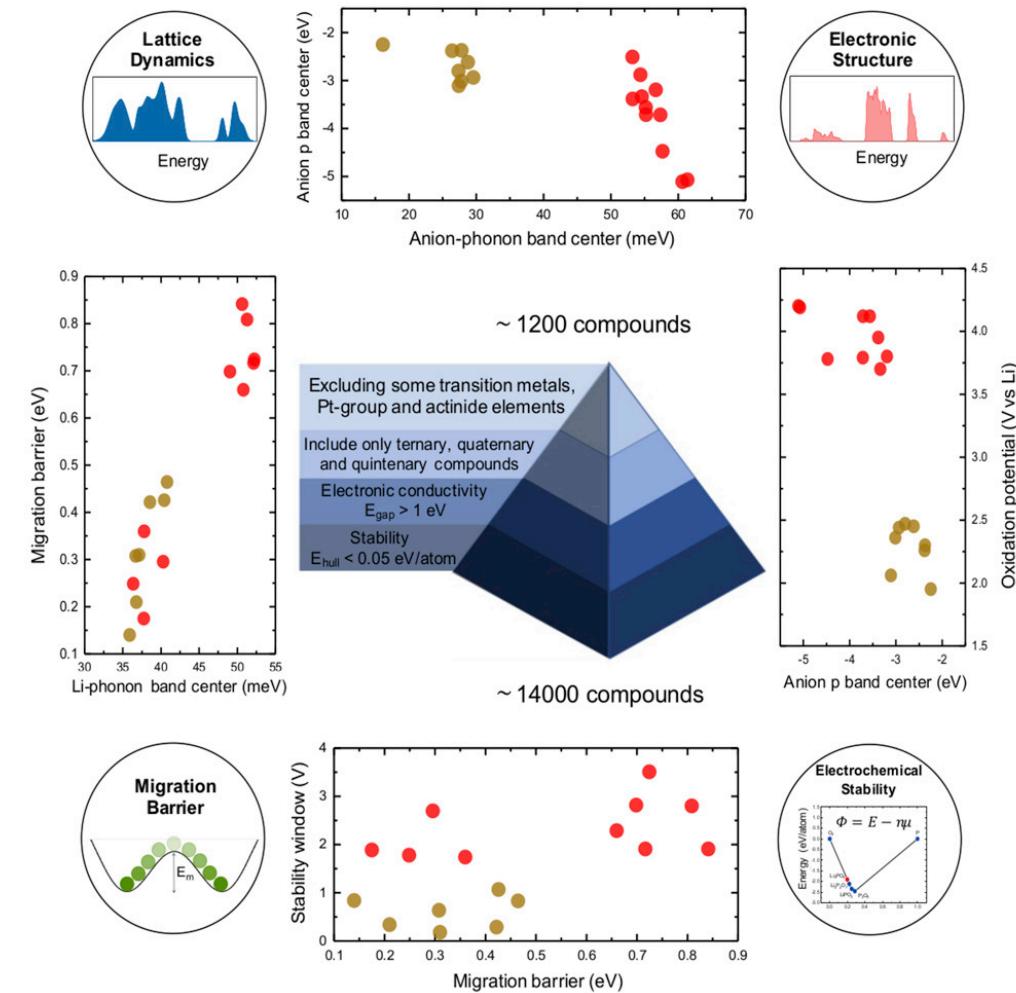
- Class #1: What is materials informatics + Python crash course
- Class #2: Python libraries for atomistic modelling of materials
- Class #3: Data in materials science
- Class #4: Data exploration, visualization, and fitting
- Class #5: Classical ML for materials science pt.1
- Class #6: Classical ML for materials science pt.2
- Class #7: Graph neural networks for materials science pt.1
- Class #8: Graph neural networks for materials science pt.2
- Class #9: Machine learning for molecular simulation
- Class #10: Working on final projects in class
- Class #11: Selected topics in materials informatics
- Class #12: Final project presentations

Course materials

- The course content is stored at the github [repo](#)
- HWs are announced separately in the canvas and a telegram chat

FP example: High throughput screening of solid state electrolytes for Li-ion batteries

- Formulate selection criteria
- Download the data from the database
- Predict the band gap. Predict the ionic conductivity. Perform feature importance study.
- Perform a diffusion simulation study of the best candidate using your favourite universal interatomic potential.
- Write a 3 page article style report
- Prepare a 5 minutes oral presentation



You can propose your own final project if it is in line with the ILOs

On completion of the course you will be able to:

- Apply python libraries and data science tools to solve materials science problems
- Critically evaluate materials informatics literature
- Collect, generate and analyse materials science datasets, including identification of structure-property relationships

Why?

Modern materials science = Materials informatics

Take home message

- Materials informatics accelerates the solution of materials science problems through the use of data science tools
- We focus on atomistic scale and inorganic compounds
- By the end of the course you will be able to apply data science pipelines for solving materials science problems

Announcement

- HW1 is released
- HW1 deadline is Thursday, October 10th at 11.59 pm
- The deadline is strict
- Topics covered in HW1 (the first 4 lectures/seminars):
 - python for science: numpy, pandas, scipy, matplotlib
 - python for materials modeling: ase, pymatgen
 - python for materials informatics: mp_api - the Materials project API
 - exploratory data analysis

Resources

<https://github.com/tilde-lab/awesome-materials-informatics>

<https://github.com/anthony-wang/BestPractices>

<https://github.com/sp8rks/MaterialsInformatics>

<https://github.com/ncfrey/resources>

<https://github.com/eddotman/intro-to-materials-informatics>

<https://enze-chen.github.io/mi-book-2021/intro.html>