# Introduction to Materials Informatics (2025)

Skolkovo Institute of Science and Technology

Dmitry Aksyonov, Assistant professor (Skoltech)

Artem Dembitskiy, Junior researcher (Skoltech), Researcher (AIRI)

Time: TBD

Location: TBD

# 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

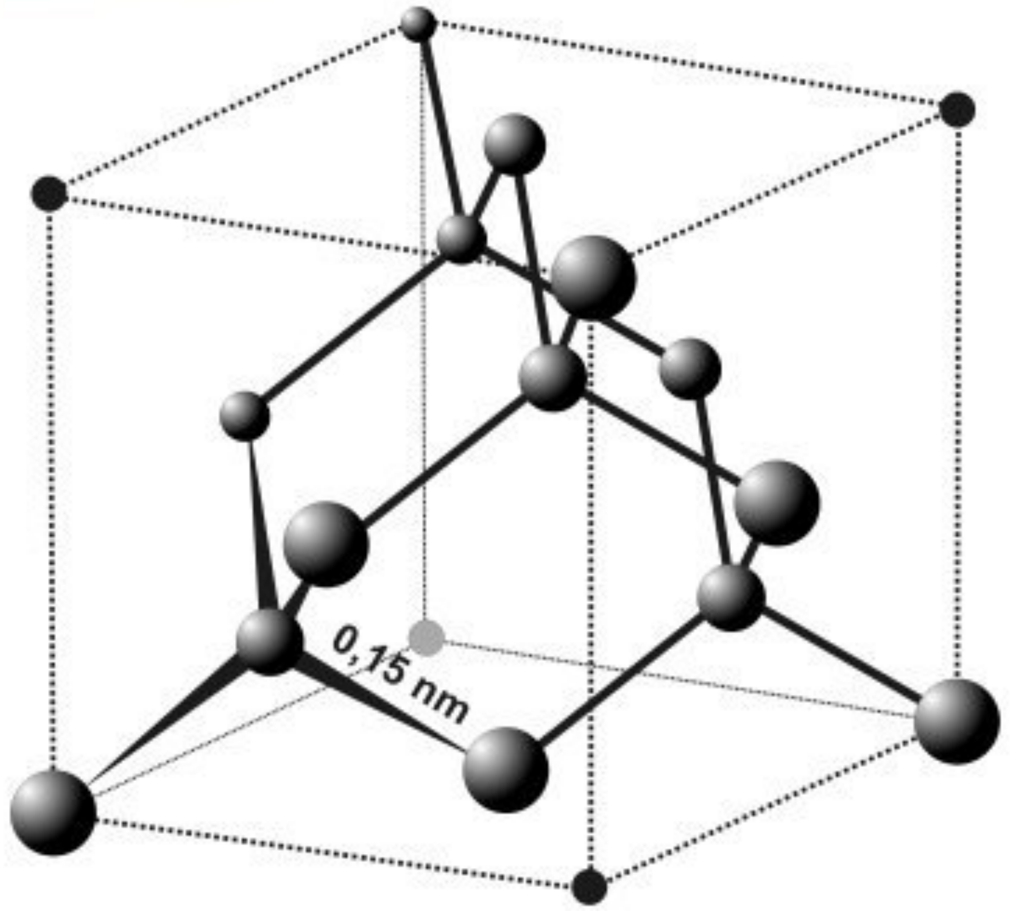
- Goals
- What is materials informatics?
  - Materials science
  - Scales
  - Experiment vs. Theory
  - Materials informatics
  - Examples
- Course navigation
  - ILOs, HW, FP, assessment criteria

# Part #1: Materials science

## Materials science (and engineering)

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

Why are diamond and graphite so different despite the same chemical composition?



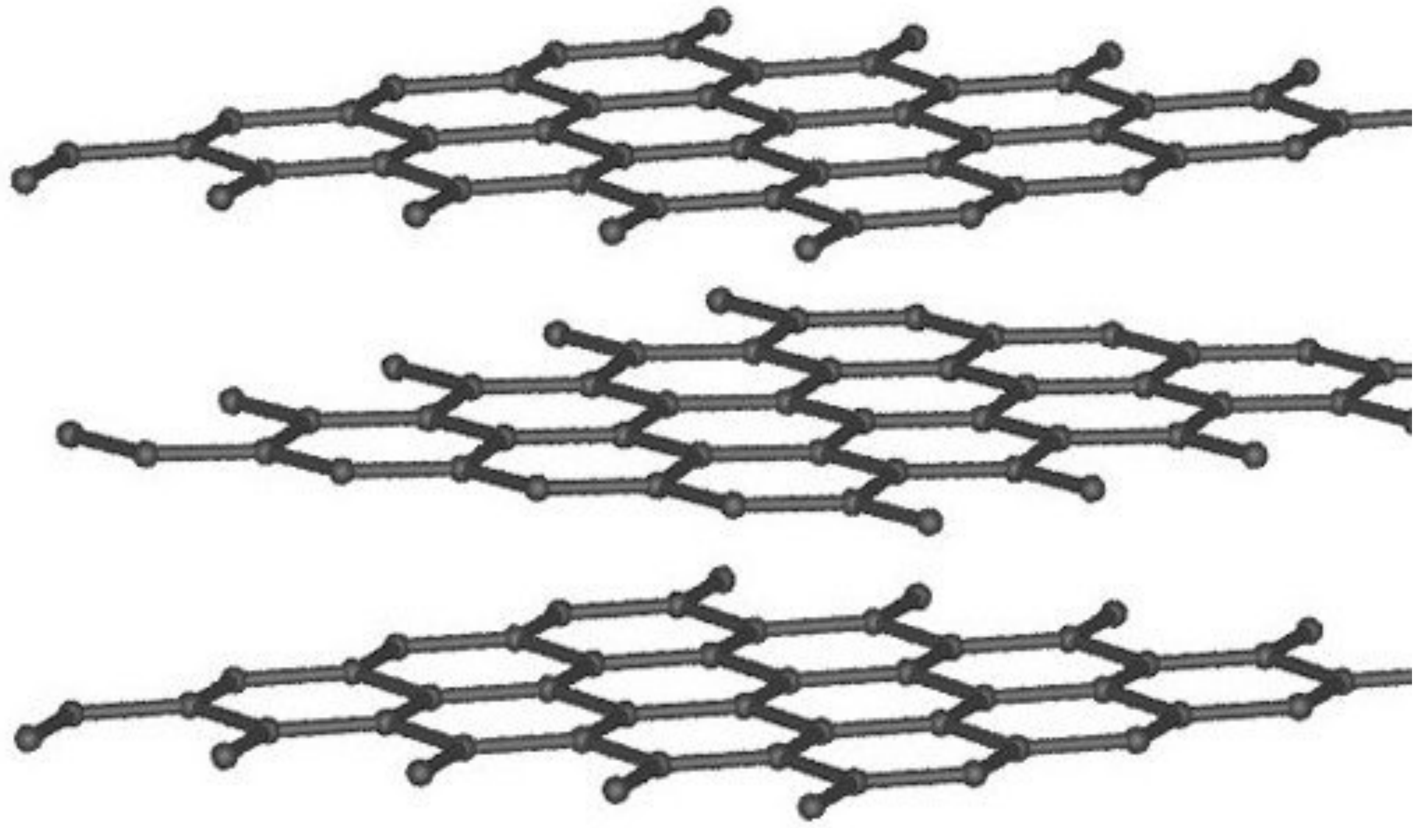



Image credit

## Tasks that materials science solves

- Design
- Improve
- Understand reasons of a failure

## Design

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

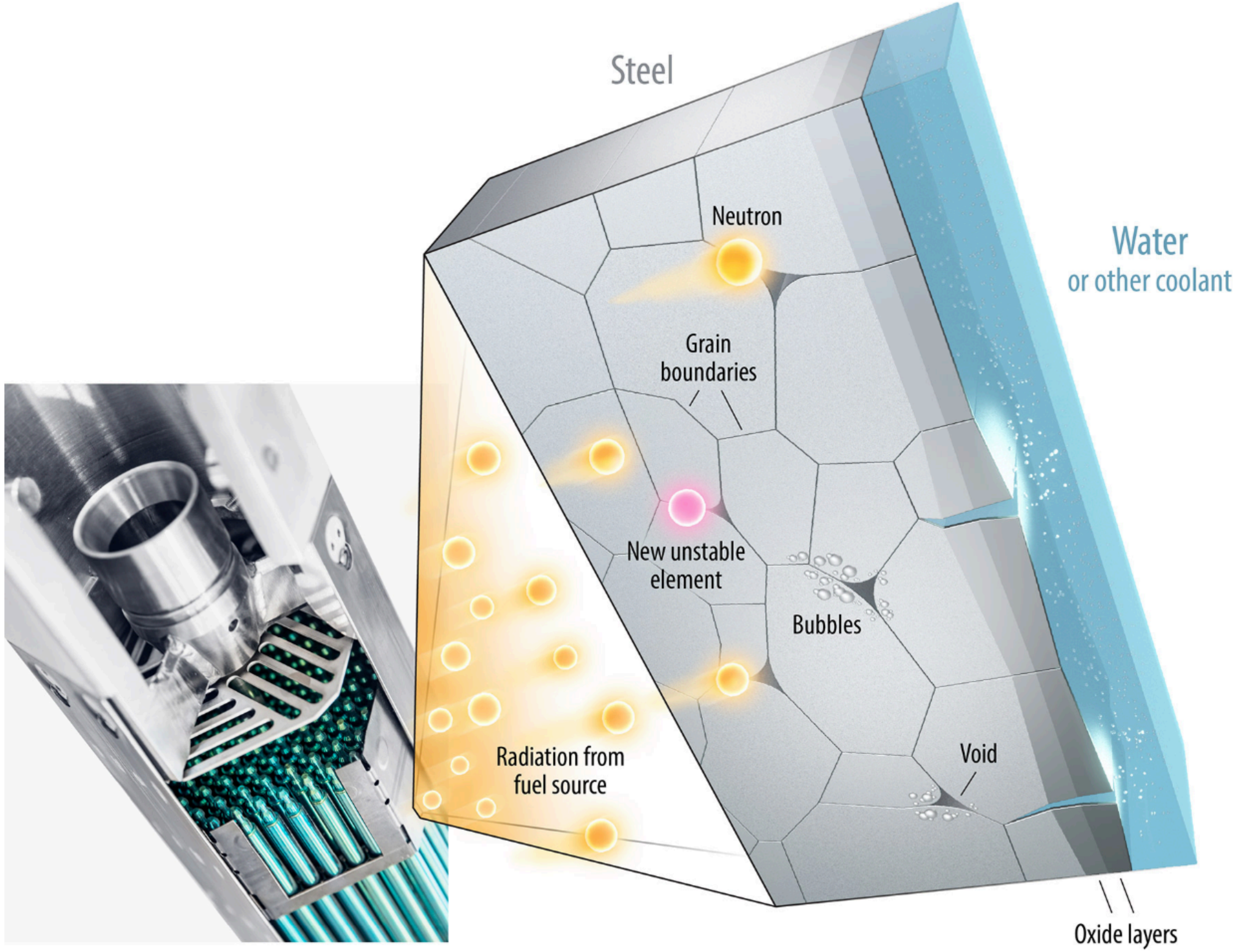
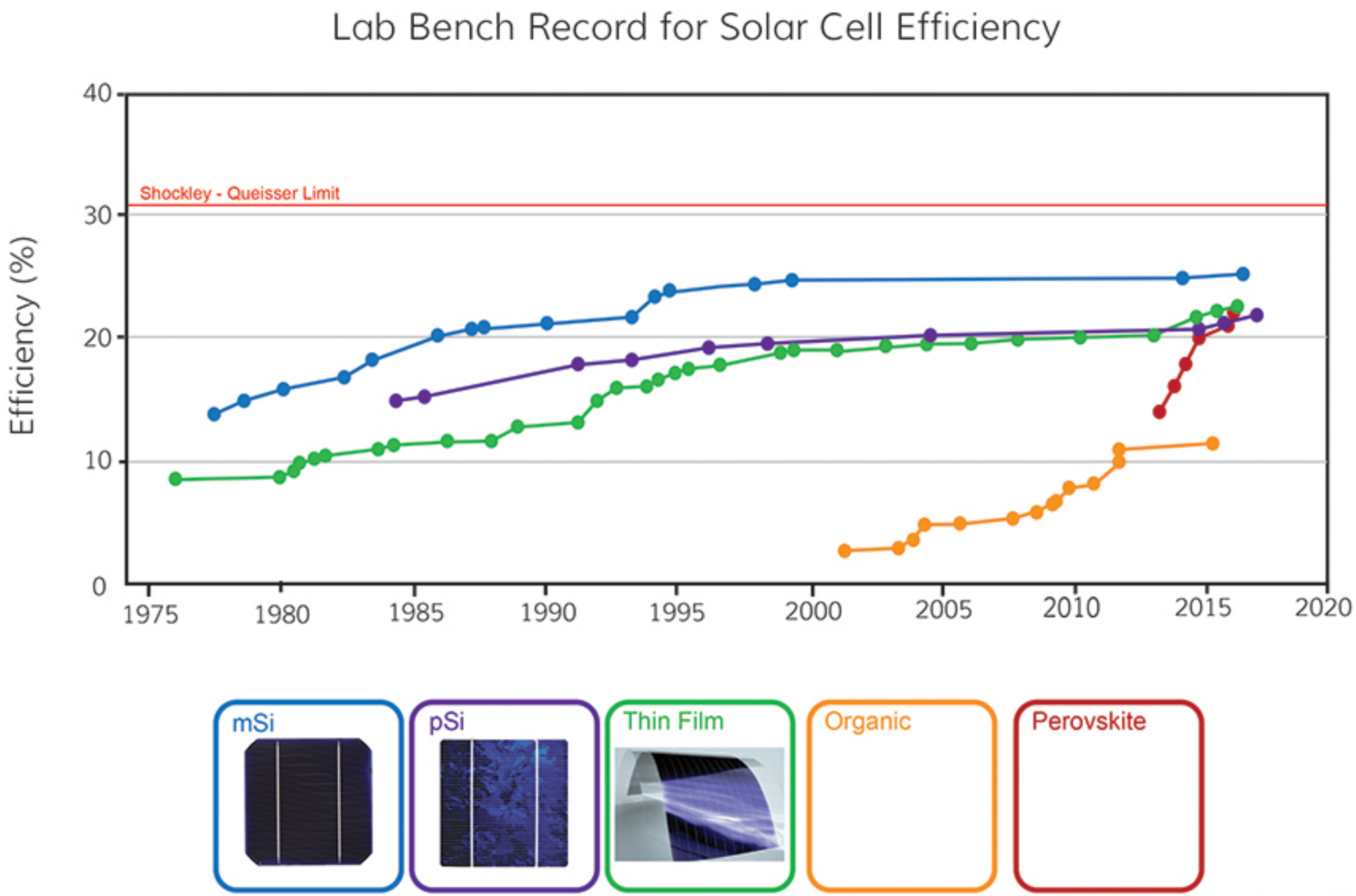



Image credit

# Improve

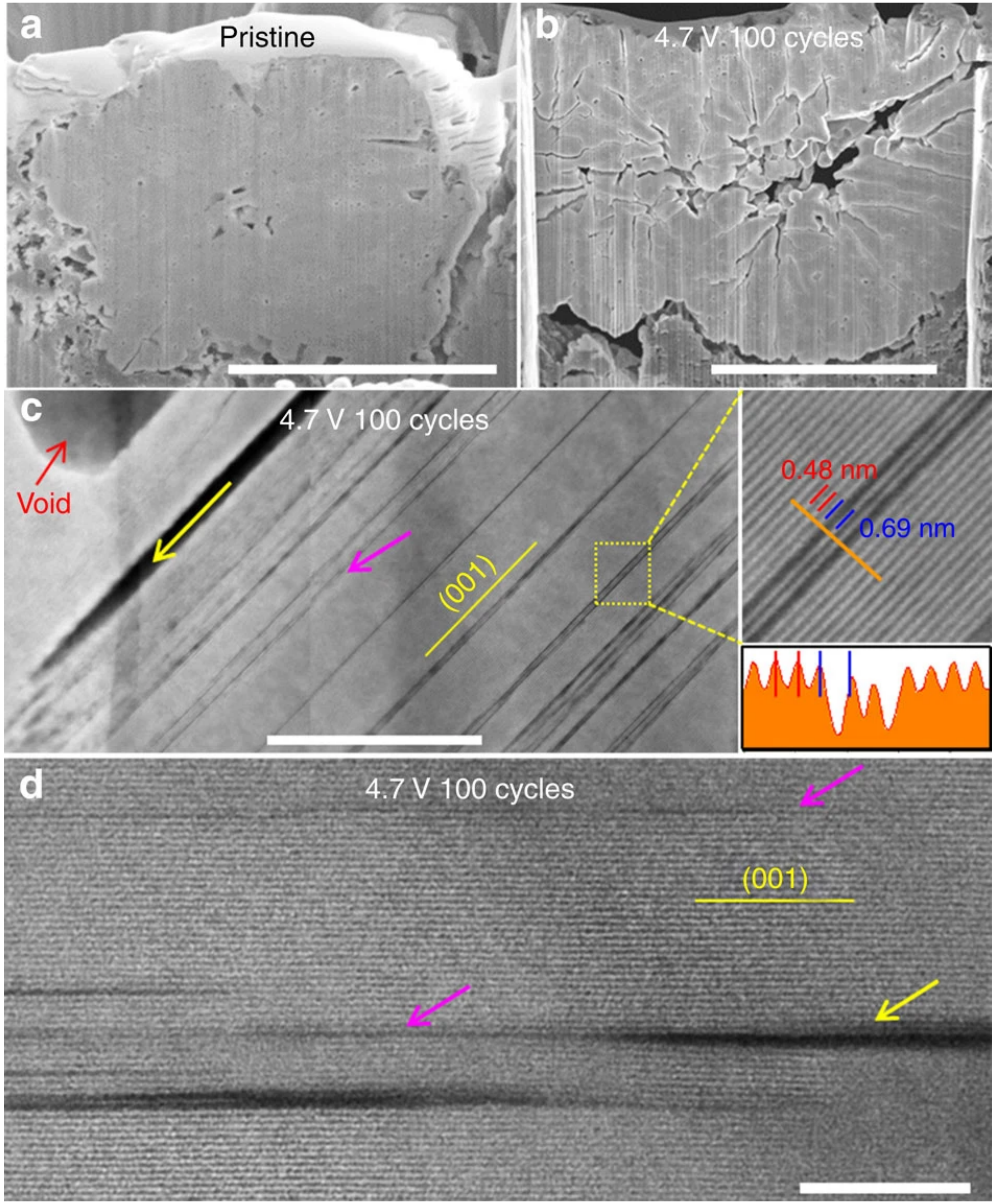
- Better props
- Less expensive
- Easier to process




Source: NREL

## 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

## Materials are characterized

...at various length scales

- atomic
- micro
- meso
- macro

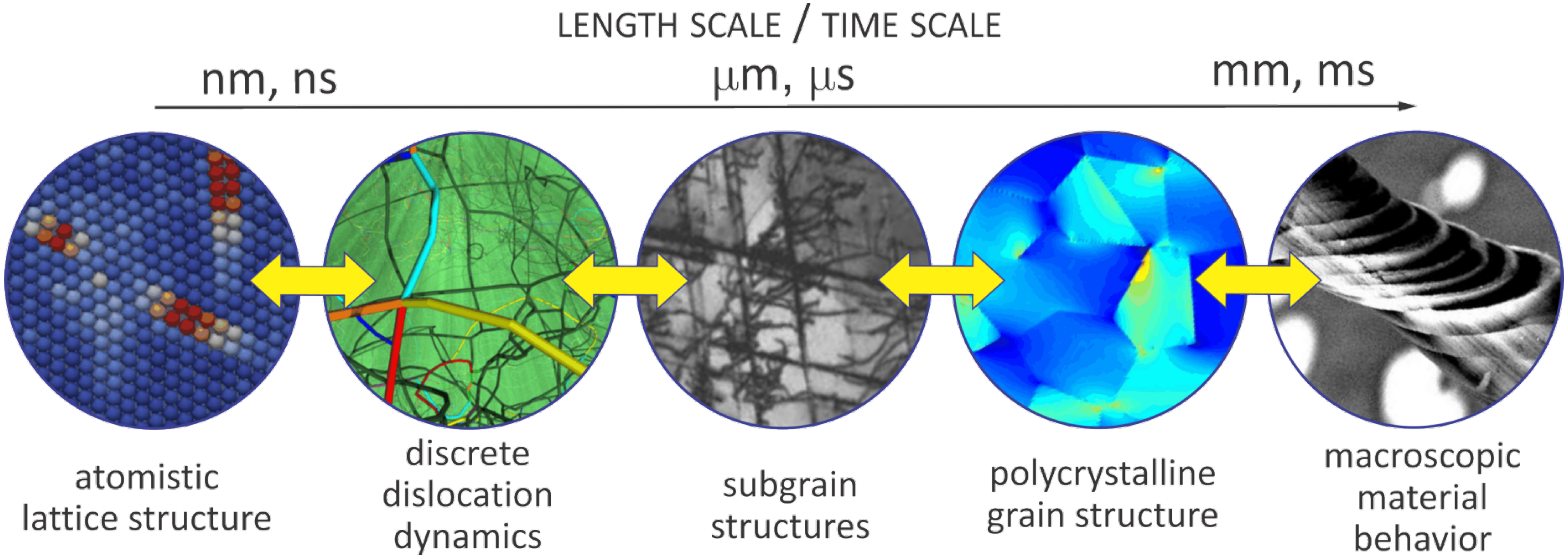
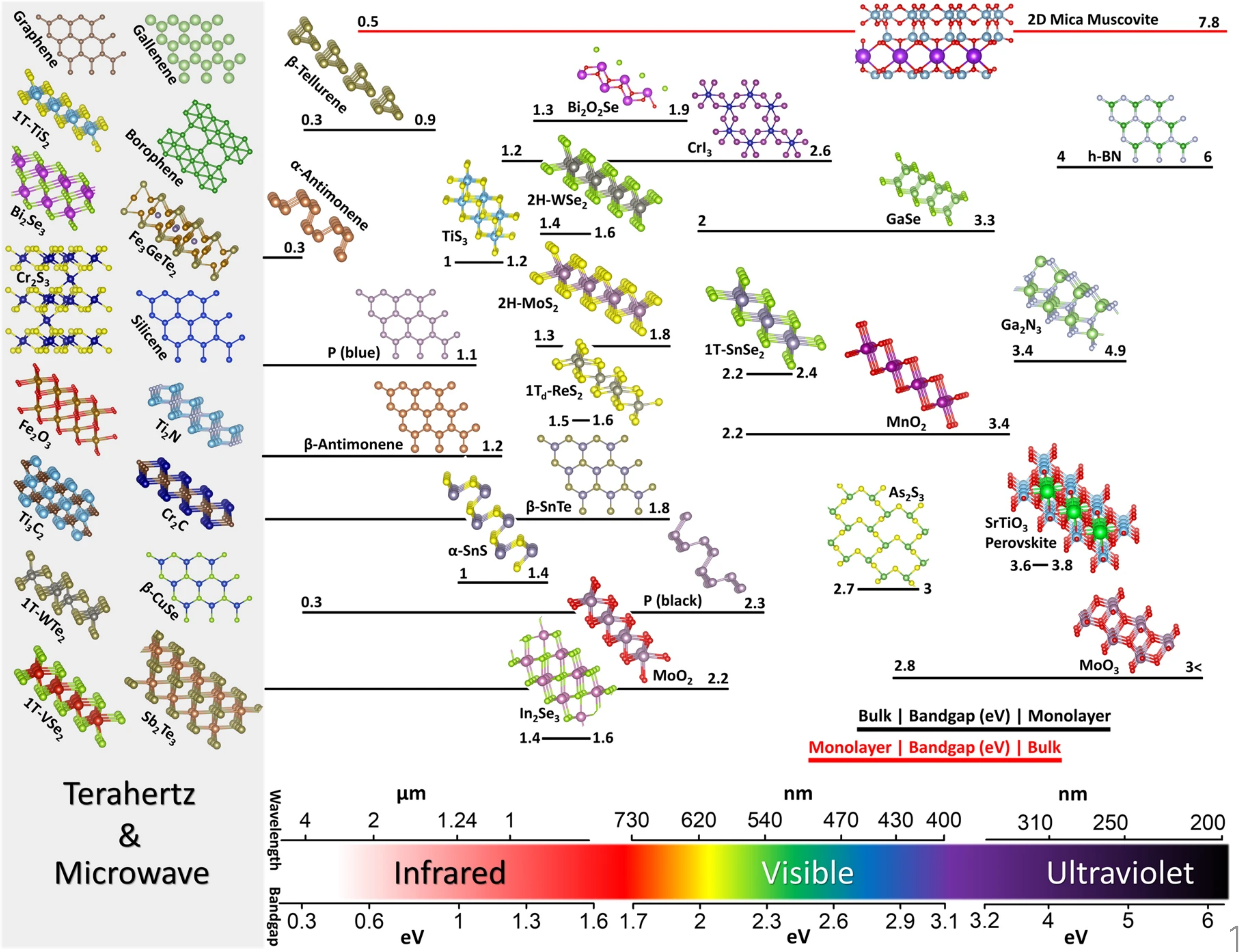



Image credit

# The course focuses

…on the atomic scale

…and inorganic materials



# Part #2: Trial-and-error paradigm

## Stainless steels example

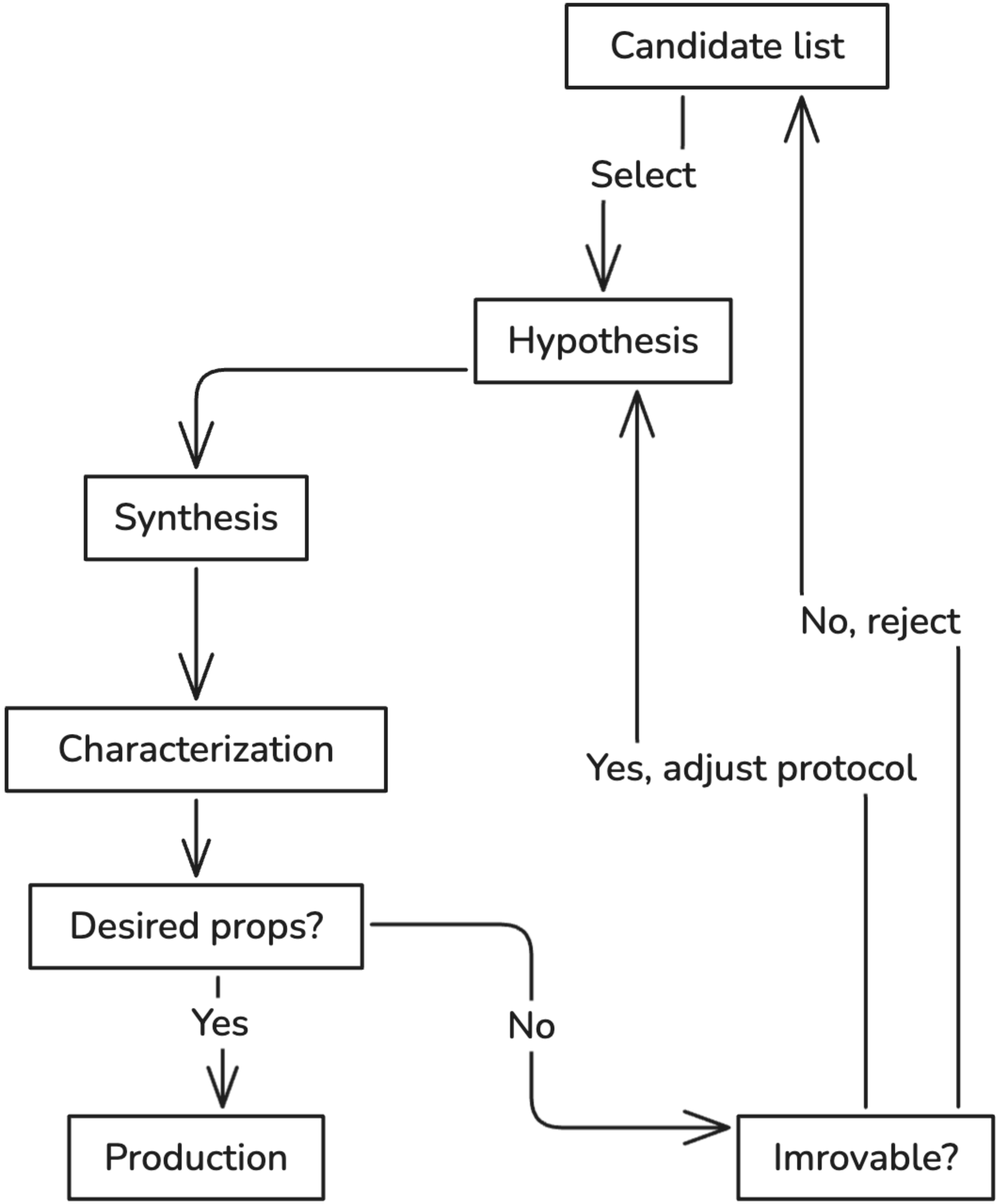
- 1798 - chromium discovery
- 1821 - iron-chromium alloys
- 1861 - the first patent on chromium steels
- 1911 - report on the relationship between the chromium content and corrosion resistance


Image credit

## Trial-and-error in materials science

...is the only way to reach the goal

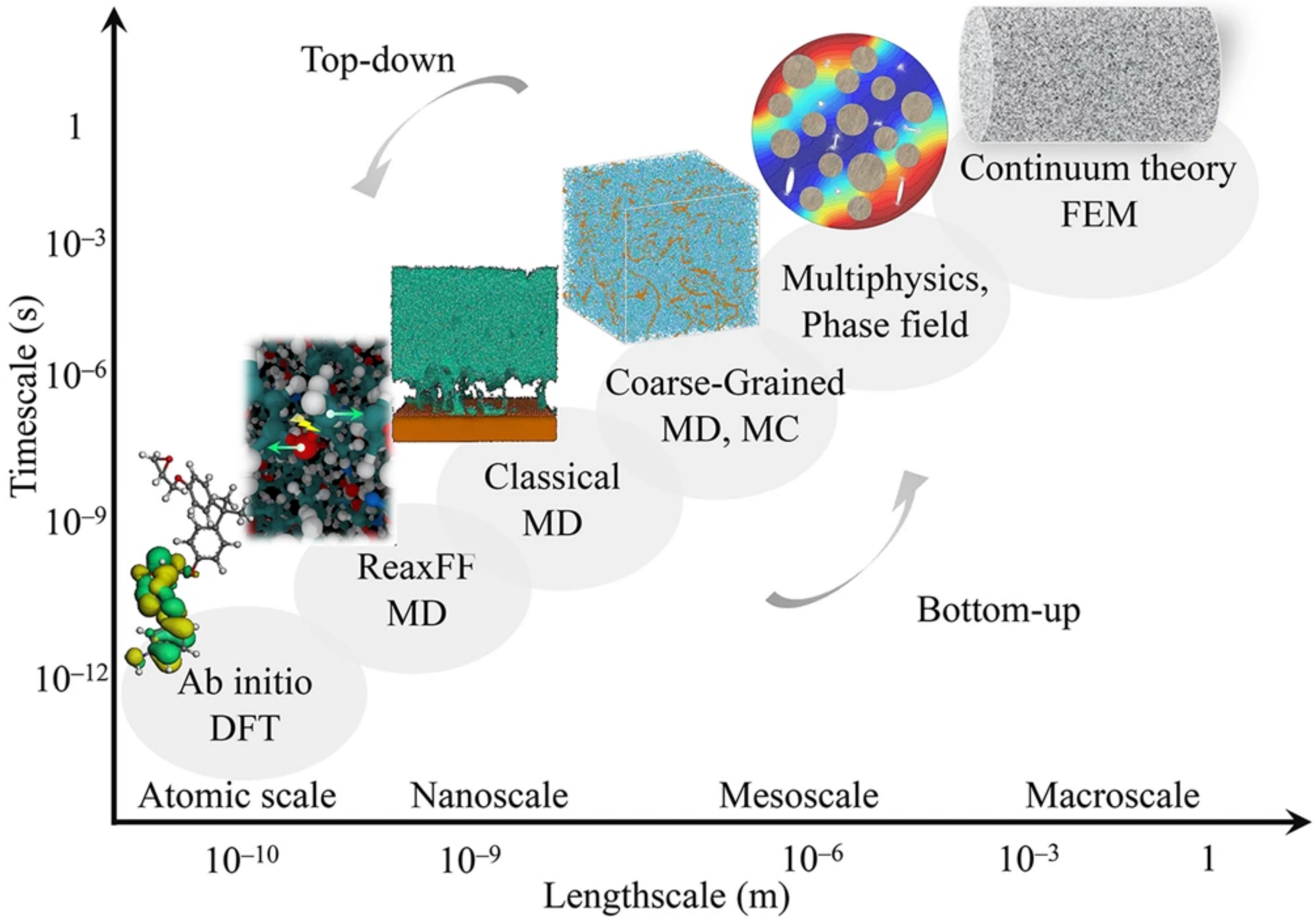
...Takes ~10 years to design a material



## Trial-and-error loop can be optimized

...with the help of theoretical and computational approaches

...by replacing "expensive" experiments with computation

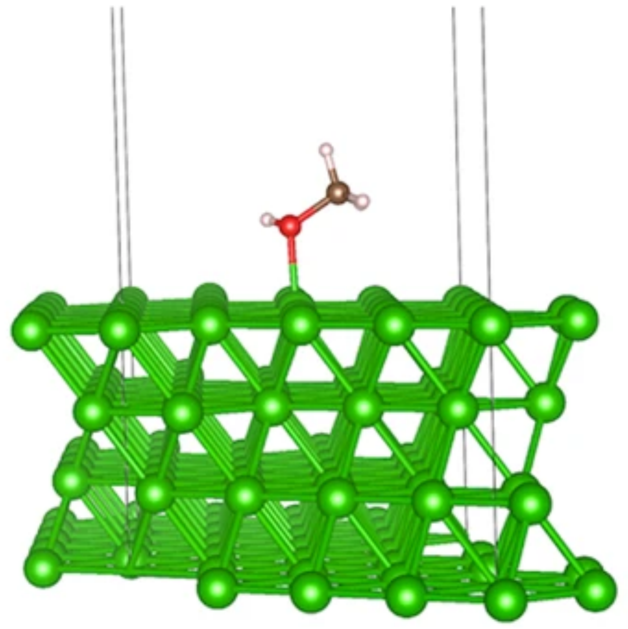



Multiscale mechanics and molecular dynamics simulations of the durability of fiber-reinforced polymer composites

# Example: Adsorption energy

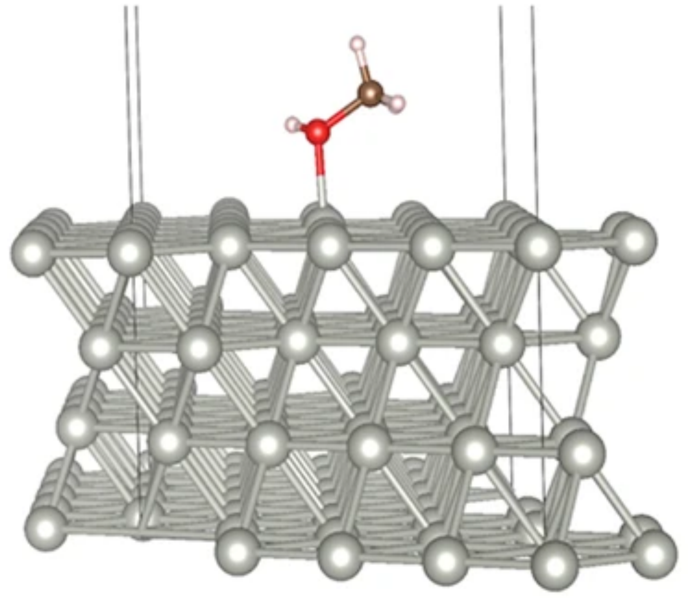
- Determines catalytic properties
- Essential for chemical industry

MeOH-Ni(111)



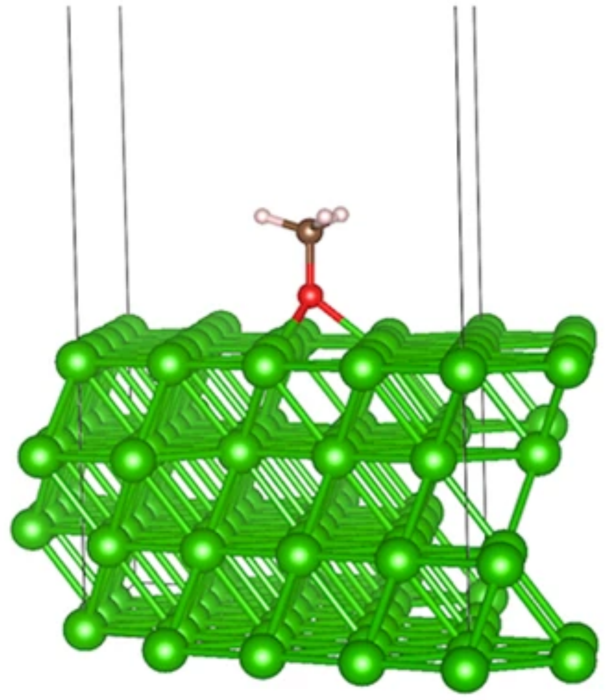
Exp. = -14.3
optB86b-vdW = -15.2
PBE = -5.8
PBE+D3 = -14.3

MeOH-Pd(111)



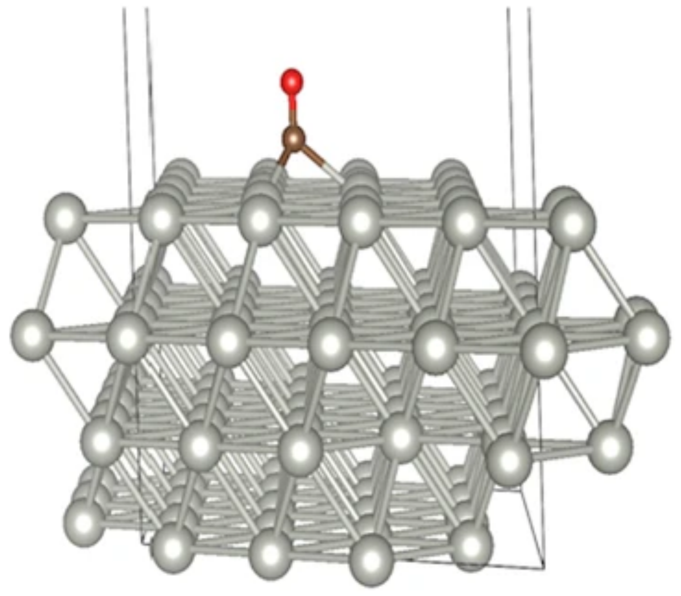
Exp. = -11.8
optB86b-vdW = -13.6
PBE = -3.7
PBE+D3 = -13.8

$CH_3O$-Ni(111)



Exp. = -55.6
optB86b-vdW = -72.4
PBE = -60.0
PBE+D3 = -67.6

CO-Pd(111)

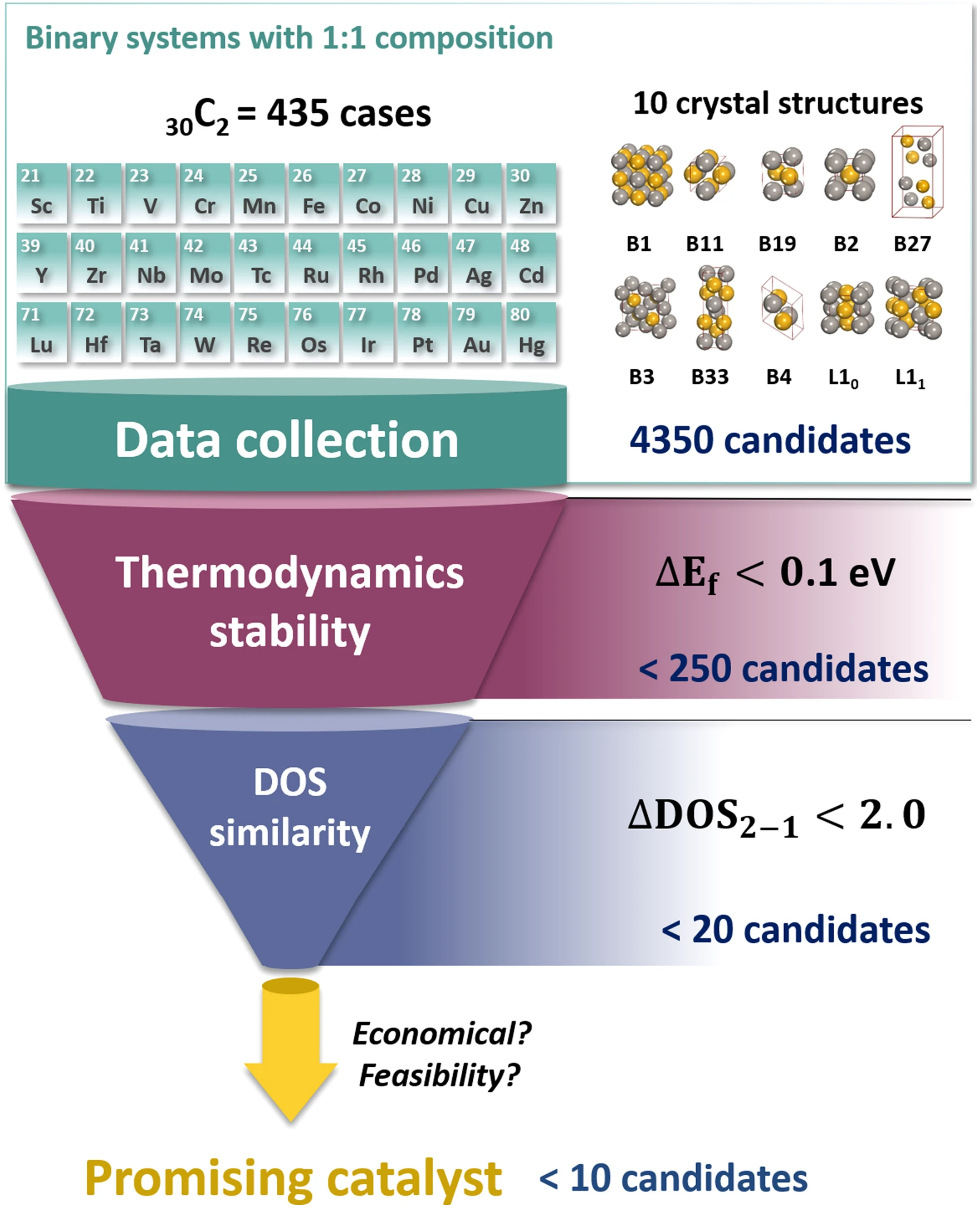


Exp. = -33.2
optB86b-vdW = -51.2
PBE = -45.4
PBE+D3 = -52.1

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

# Computational methods

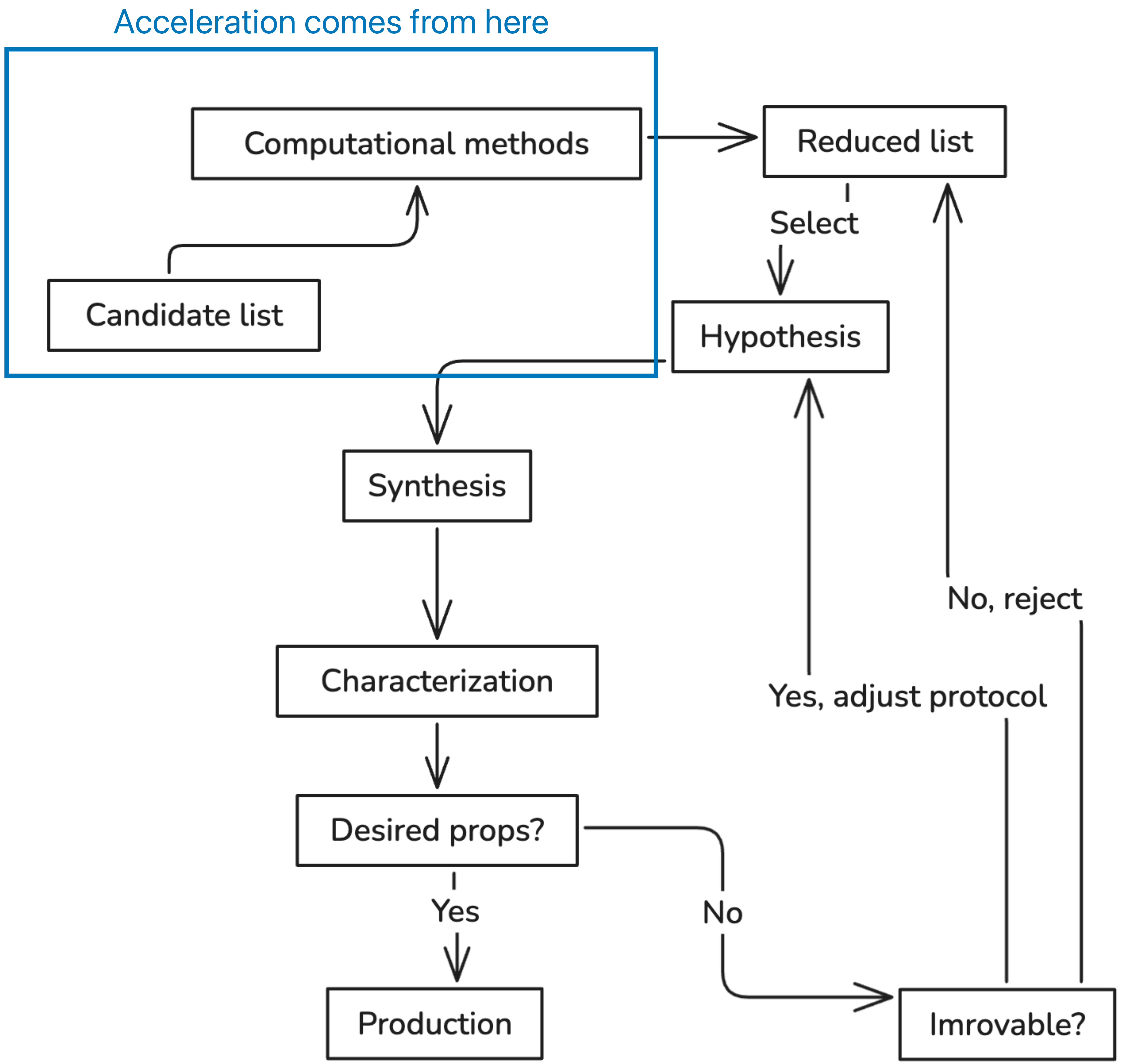
...accelerate the first stage of the materials design




High-throughput computational-experimental screening protocol for the discovery of bimetallic catalysts discovery of bimetallic catalysts

## Trial-and-error with computational methods

…takes ~2 years to design a material

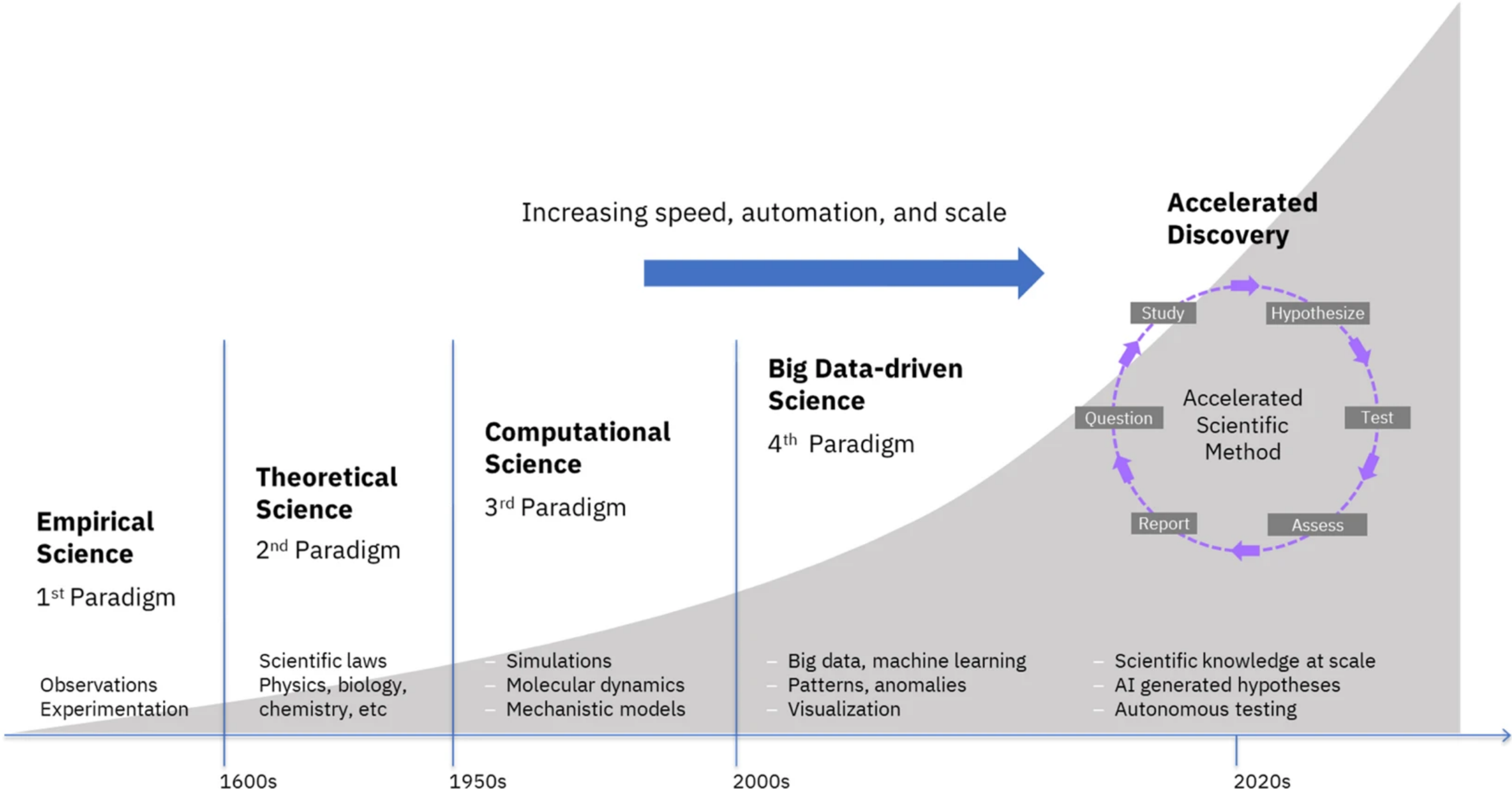


# Can it be further optimized?

# Materials informatics

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

...mostly by replacing computational methods with surrogate models

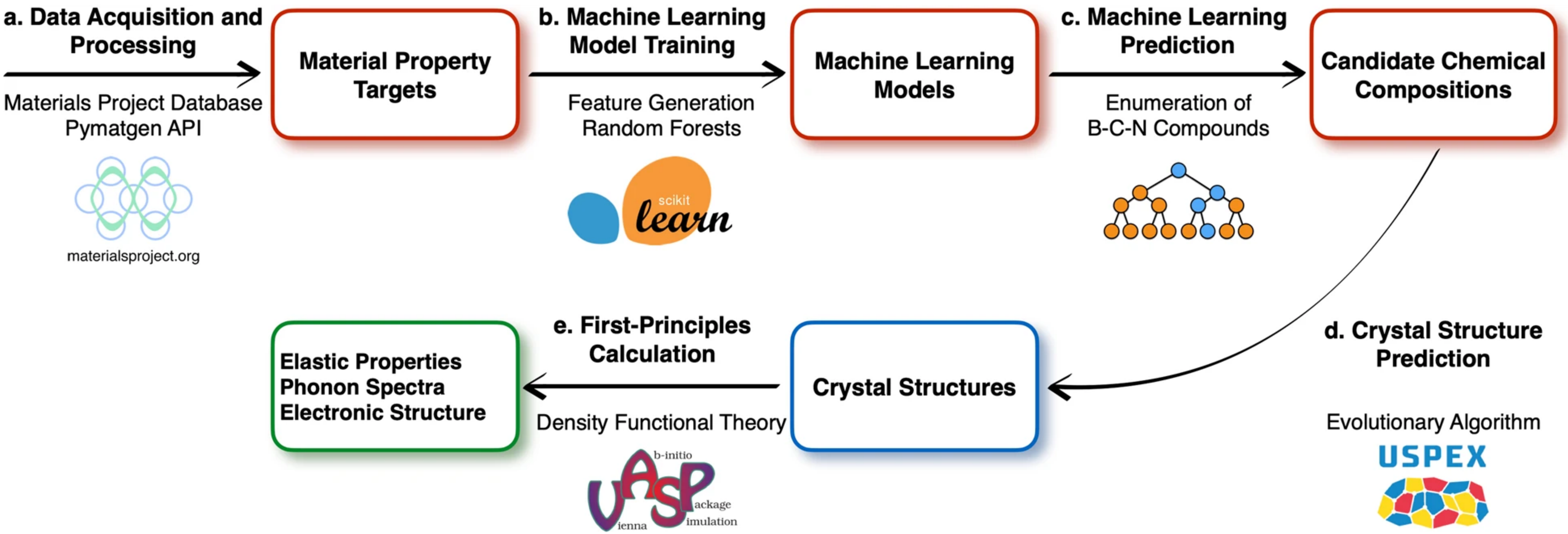


“A **surrogate model** is an engineering method used when an **outcome of interest cannot be easily measured or computed**, so an **approximate** mathematical **model** of the outcome is used instead.”

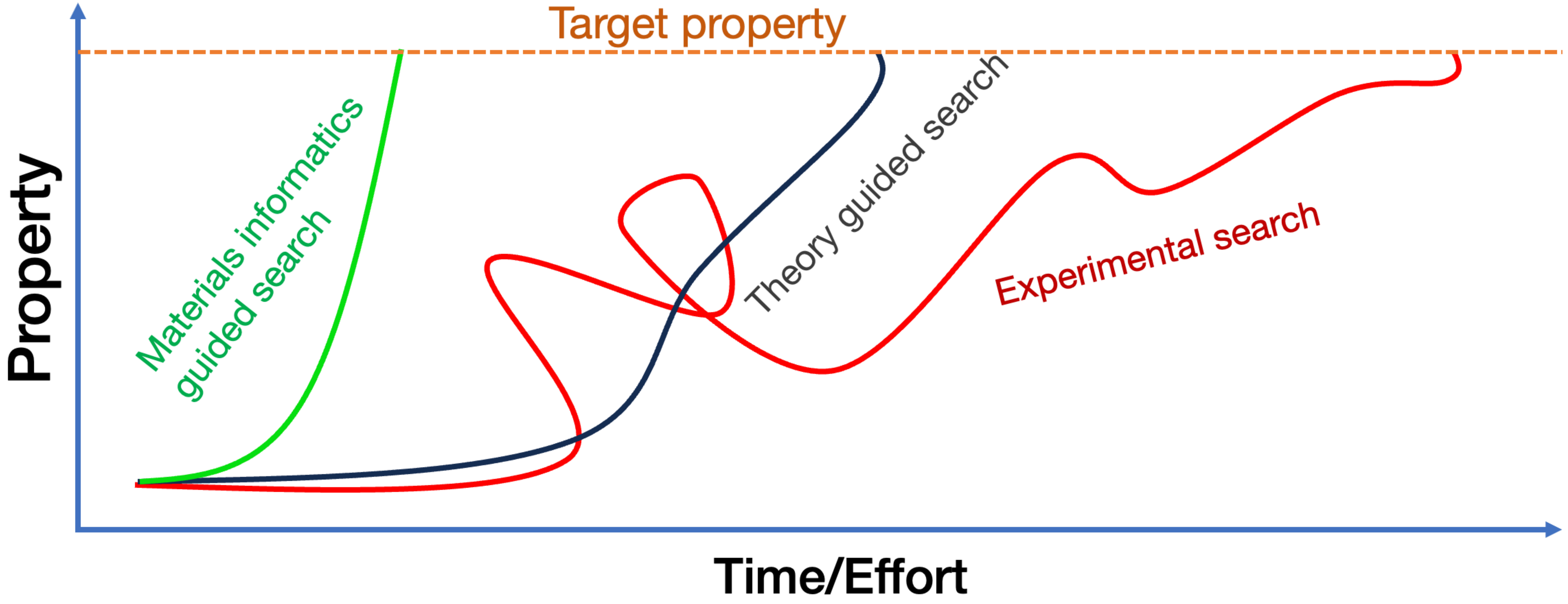
https://en.wikipedia.org/wiki/Surrogate_model

# A typical materials informatics workflow

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



# All in all: Materials informatics paradigm for materials design



# Part #3: Course motivation and navigation

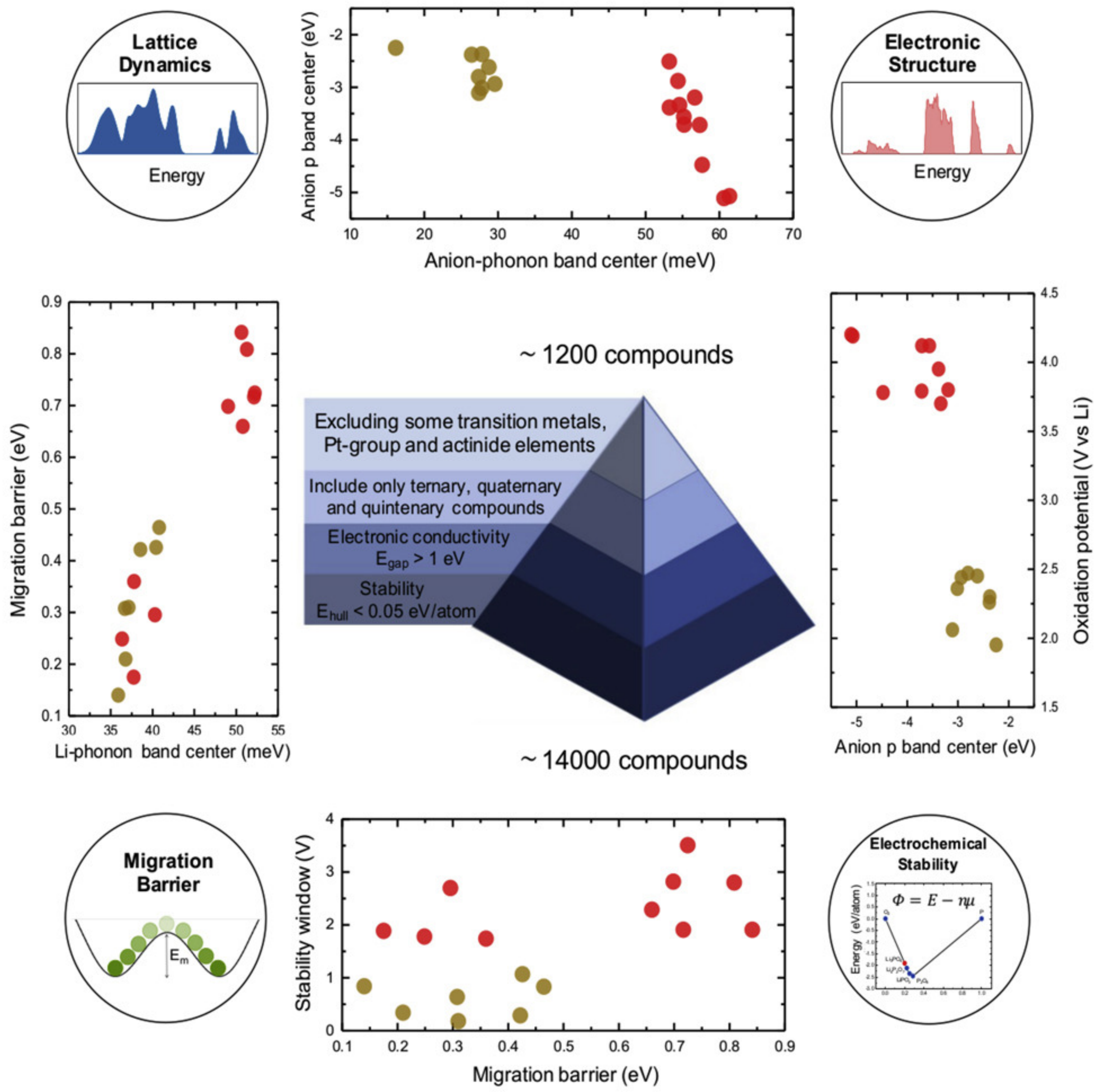
## Course structure

- Each class = 1h lecture + 2h seminar
- 11 lectures
- 10 seminars
- 2 HWs (coding) + 1 paper critical review
- 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 (students present critical reviews)
- Class #12: Final project presentations

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

- Formulate selection criteria
- Download the data from the database
- Prepare a dataset of ionic conductivity parameters
- Training: Select and train model/Interpret results (feature importance, etc.)
- Inference: Predict properties for candidate materials
- Perform a diffusion simulation study for the best candidate using universal interatomic potential
- Write a 3 page article style report
- Prepare a 5 minutes oral presentation



## Course materials

- The course content is stored at the GitHub repo
- HWs are announced separately in the canvas and a telegram chat

# Your final project must be 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
- Collect, generate and analyse materials science datasets, including identification of structure-property relationships
- Critically evaluate materials informatics literature

# 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: 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