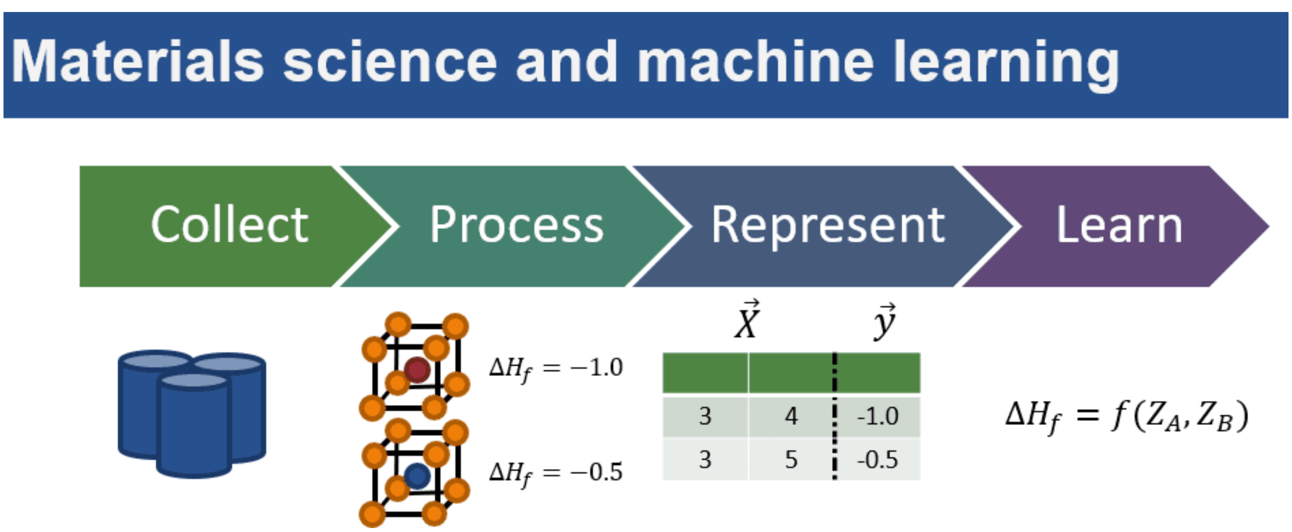
Lecture-2: Computational Material Science

With the increase of experimental and computational data, the field of materials informatics (MI) has grown quickly in recent years. One important task of MI is to use existing materials data to predict properties for new materials by employing mathematics and information science methods.

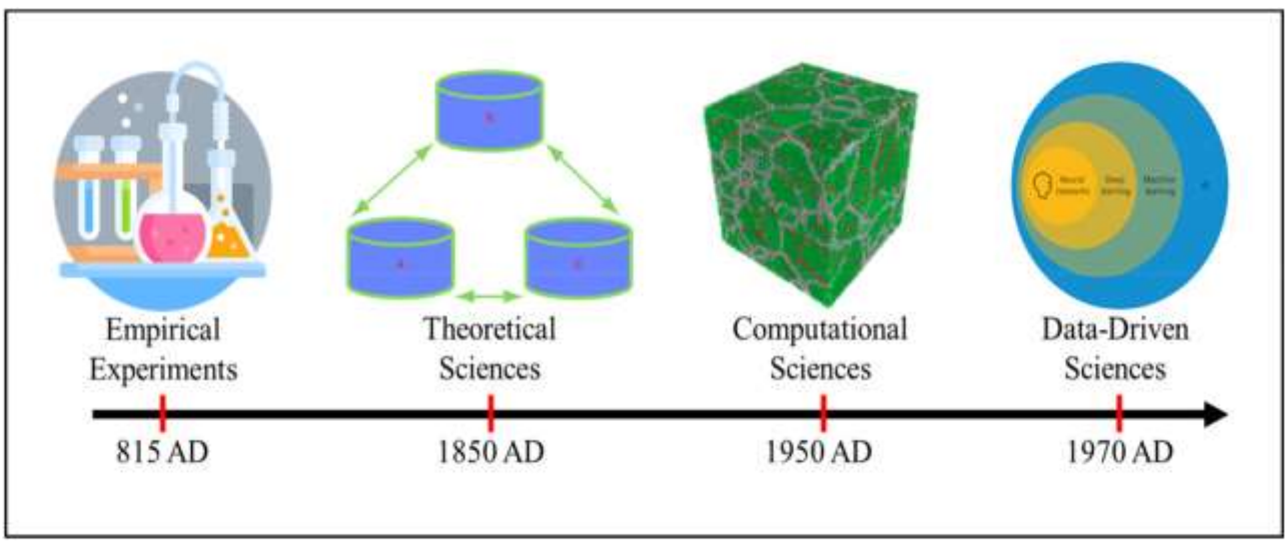
ML is a branch of artificial intelligence (AI) that aims to build models trained from past data and situations. It has started to play a significant role in materials science due to its ability to learn behaviors and trends from available data without knowing the underlying physical mechanisms.
Materials scientists are key users of leadership class computing; their studies are critical to improving economic security and competitiveness, national security and human welfare. Finding and understanding new materials is complex, expensive, and time consuming, often taking over 20 years for research to make its way from experiment to industrial application. To accelerate this process, researchers and facilities are motivated to work together to develop data platforms and tools that are accessible to a broad set of scientific researchers.



In machine learning, a data representation is a combination of key variables believed to best correlate with what is trying to be modeled. If your model is meant to differentiate between cats and dogs.

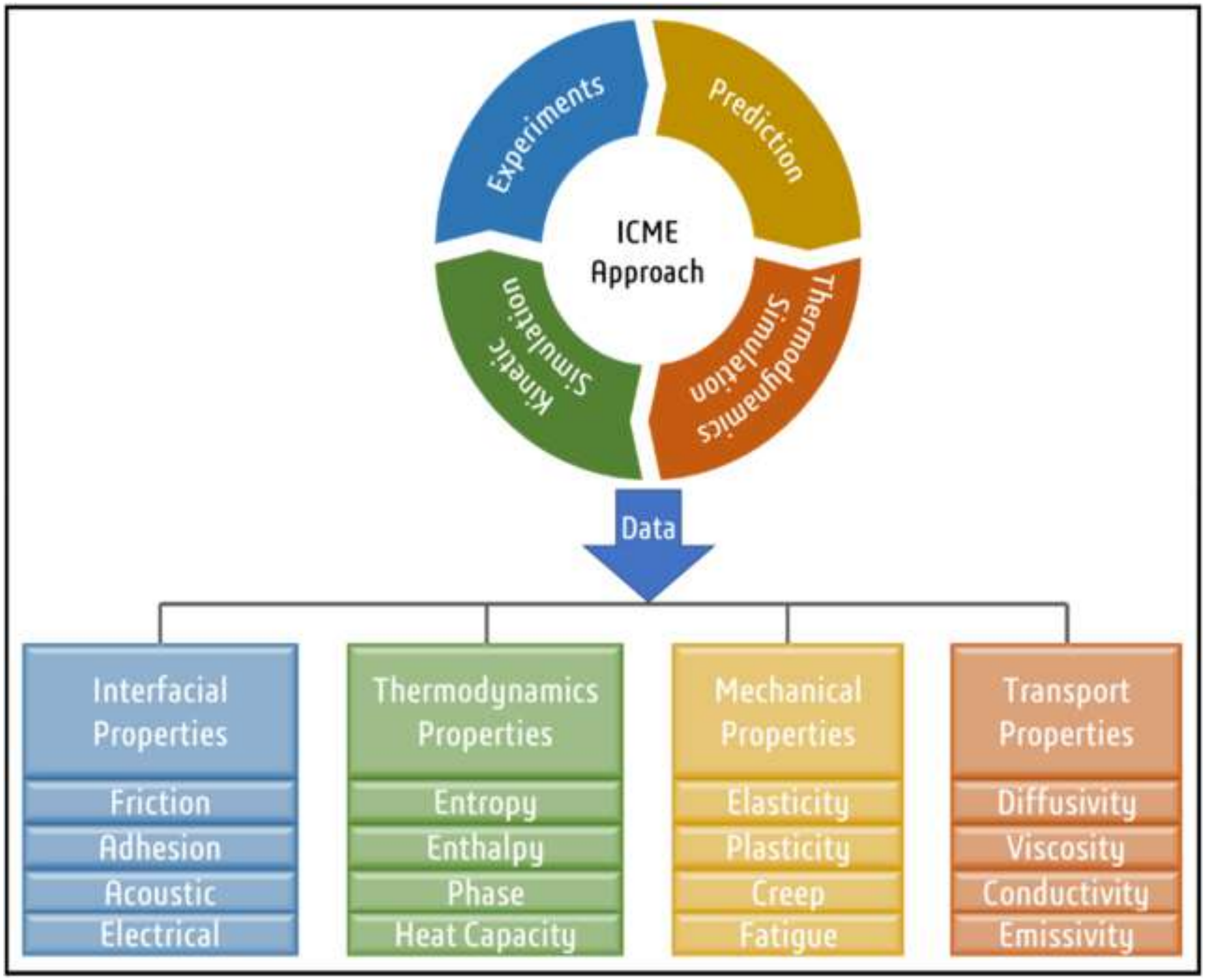
Scientists selected a representation for their data in this project by determining what ‘atomic structure’ actually means in terms of inputs to the model (e.g., what determines force on a projectile, and how does one quantify the effect of this force?). They then trained a machine learning algorithm, selecting an algorithm based on the key criterion of highest prediction accuracy as well as feasibility to train with >104 entries, speed of evaluation, and ability to produce a differentiable model.

Data science in material design: critical overview Material discovery and the characterization of properties have four stages (as below).



Material Discovery Stages

This started with empirical approaches focusing on trial-and-error experiments from the stone age. It must be mentioned that many significant discoveries are the result of this type of approach. The last figure illustrate a model-based theoretical approaches underpinned by the laws of thermodynamics, statistical mechanics, and data science. Generally, computations and simulations are based on quantum (density functional theory) and classical Newtonian mechanics (molecular dynamics, phase field modeling and Monte Carlo simulations). The data-driven decision making for alloy design applies statistical mechanics for predictive modeling and relationship mining. The increasing use of theoretical and computational sciences is attributed to increasing application of computational theories in material design and global infrastructure supporting these initiatives. Some of these open- source infrastructures are Materials Project, Materials Genome Initiative (MGI), Automatic flow framework for materials discovery (AFLOW) libraries and the Integrated Computational and Materials Engineering (ICME).



ICME approach to an integrated data driven advance materials development

Benefits Of Data Science And Material Informatics

1. *Efficient use of resources*

Material design by the traditional trial-and-error approach require using amounts of elements. This is attributed to the iteration involved during the process to establish the right composition-microstructure-property connection.

Computational material design using machine learning (ML) helps narrow down the compositional space with no need for actual experiments. This reduces extremely the "wastage" associated with iterative process involved in optimizing the composition-structure-property relationship for the required engineering application. This has been applied to screening the large compositional space of complex concentrated alloys and for identifying elastic isotropic architecture materials. Without computation approaches in this resource constraint era, conventional trial-and-error approaches could be too costly, difficult, time-consuming, and unsustainable. The application of computational approaches has been used to optimize processes in manufacturing and processing plants.

2. *Cost reduction for computational power*

Numerical simulations are greatly dependent on the quality and accuracy of the assumptions and the underlying mathematical formulations. The computations or simulations can be done from personal computers to very sophisticated high speed systems. A personal computer provides enough power for low-level simulation, whereas high level simulations require high-performance- computing resources. While this could take days or months to compute, slight modification of parameters could reduce the cost and time drastically. By using any previously performed simulations, new insights are gleaned, or data could be generated.

The machine learning can replace the actual simulation without losing much accuracy while reducing computational cost considerably. However, the quality of the data set and the size of the data for this ML based model is critical to have a good enough fit.

3. *Reduction in constraints for material composition and design*

Conventional metallic alloy design concentrates on the use of principal elements with addition of secondary alloying elements in small quantities. The drift towards high performance materials for extreme engineering applications have required the development of new materials with improved properties. The complex concentrated alloys concept opens the opportunities for such properties with core effects deviating from most traditional structural and functional materials. This introduces complexity in composition selection to obtain desirable properties. Machine learning methods employed so far eliminates the tedious ممل simulations and experimental approaches needed to establish suitable alloy compositions and phase analysis. This approach, if fully developed, will lead to the discovery of new metallic alloy compositions, and improve on the properties of existing ones.

4. *Improvement in process metallurgy*

Integrated computational materials engineering (ICME) has been proven to drive the adaptation of technologies for material processing. Common ICME based processing approaches blend physics and chemistry based experimental data with simulations, covering a wide range of length and time scales. It also inculcates physics and thermodynamics-based models into industrial design process. After that, the data are gathered from numerical and experimental studies and part is used as training dataset for machine learning model. Here, the machine learning model should be selected based on the available data.

5. *Accelerated design of materials with artificial intelligence integration*

The advancement in materials informatics which promotes big data – driven science has opened a new paradigm for metallic alloy design. Data mining technologies in combination with artificial intelligence platforms offer accessible means for the synergistic applications presenting together experimental, computer simulations and observed theories for knowledge-guided material design. Machine learning offers an advantage of continues optimization of models to make reasonable predictions. This has provided the opportunity for machine learning algorithms to be implemented successfully in predicting new material compositions, structures, properties, and processes. As new algorithms and databases are being developed, machine learning

Feature Engineering

Material features are selected as variables for predicting the phases. These features could be elemental properties like atomic mass, thermodynamics parameters like entropy of mixing, numerical calculations such as the band gap, and/or metallurgical parameters like cold/hot working.

Feature engineering is the design of attributes or features into a ML algorithm to ease the prediction process. A deep understanding of the domain knowledge is essential to achieving better predictive outcomes and model interpretability. It is achieved by **cleaning** (reducing the noise), **selecting**, **manipulating**, and **transforming** data into useable attributes for descriptors in typical supervised learning model. These attributes are built to improve the efficiency and performance of the ML algorithm. The feature or attribute is the measurable input in a predictive model. The accuracy of the predictive model is dependent on **the well-constructed and well-defined target-related features extracted from the historical data**. When this is poorly, the predictive model could either underestimate or overestimate the desired parameters or properties. Other factors such as <u>sparse datasets</u>, <u>unrealistic assumptions</u> resulting from poor understanding of the problem and <u>insufficient datasets</u> could also lead to <u>**under – or overestimation.**</u>