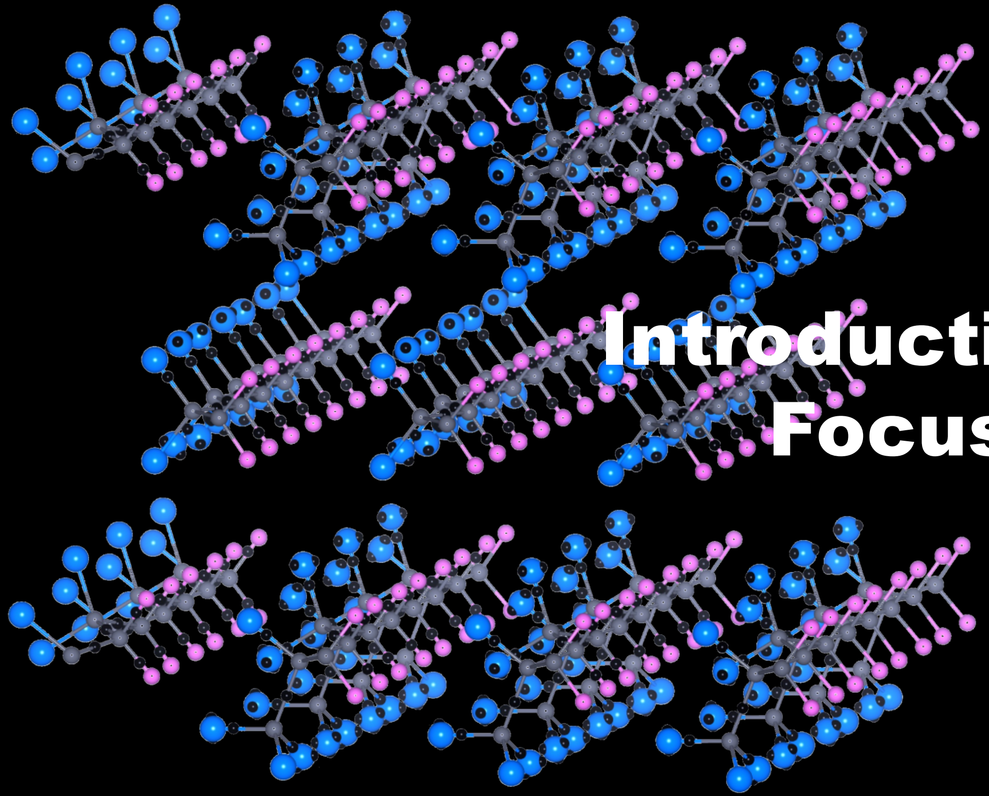
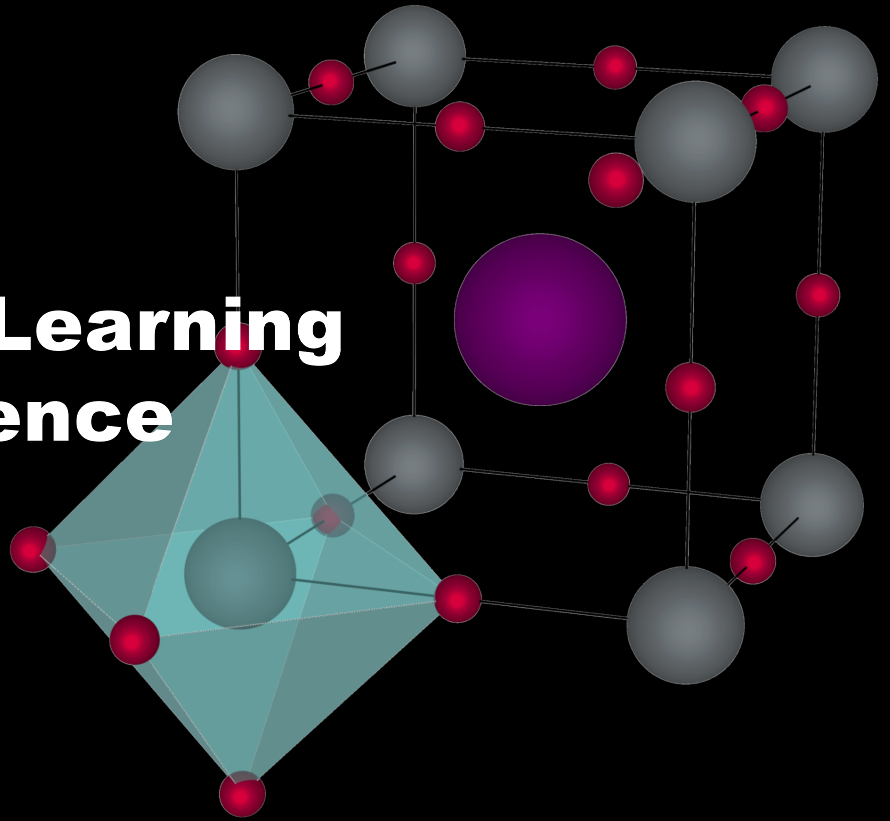




Introduction to Machine Learning
Focus: Materials Science

Workshop Presentation by:

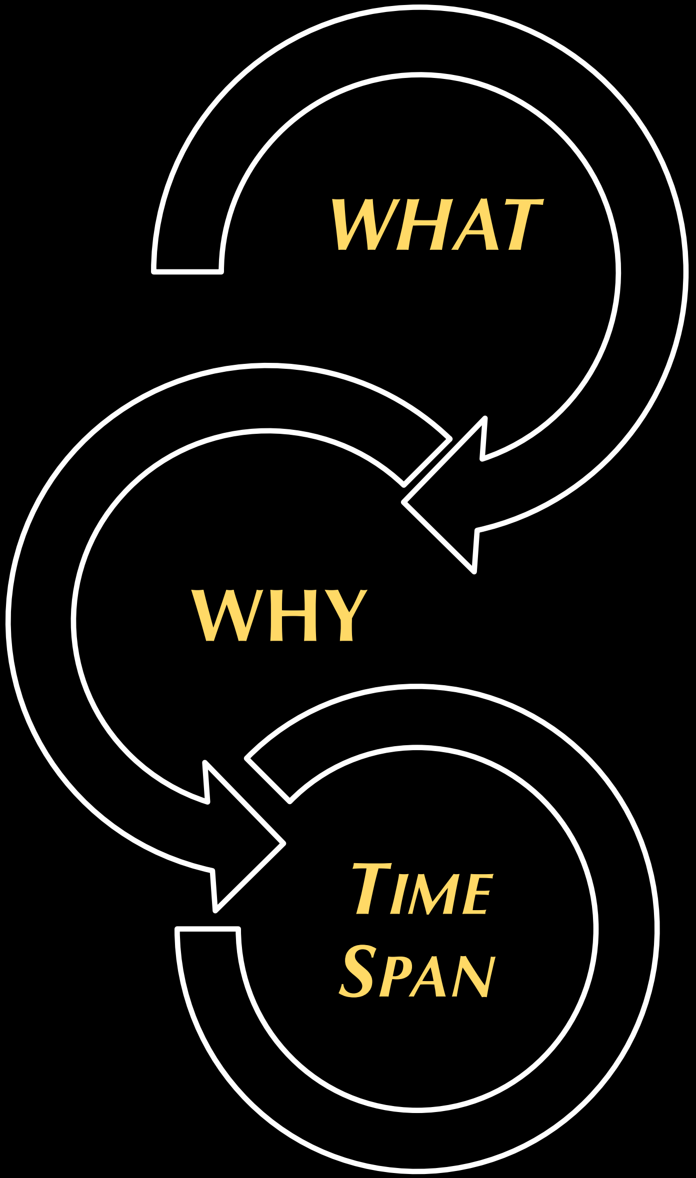
Ayana Ghosh[1,2]

[1]Department of Materials Science & Engineering, University of Connecticut, Storrs, CT 06269, USA
[2]Theoretical Division, Los Alamos National Laboratory, Los Alamos, NM 87545, USA

Supercomputing Challenge Kickoff 2018-19

Taken at Old Town Square, Prague

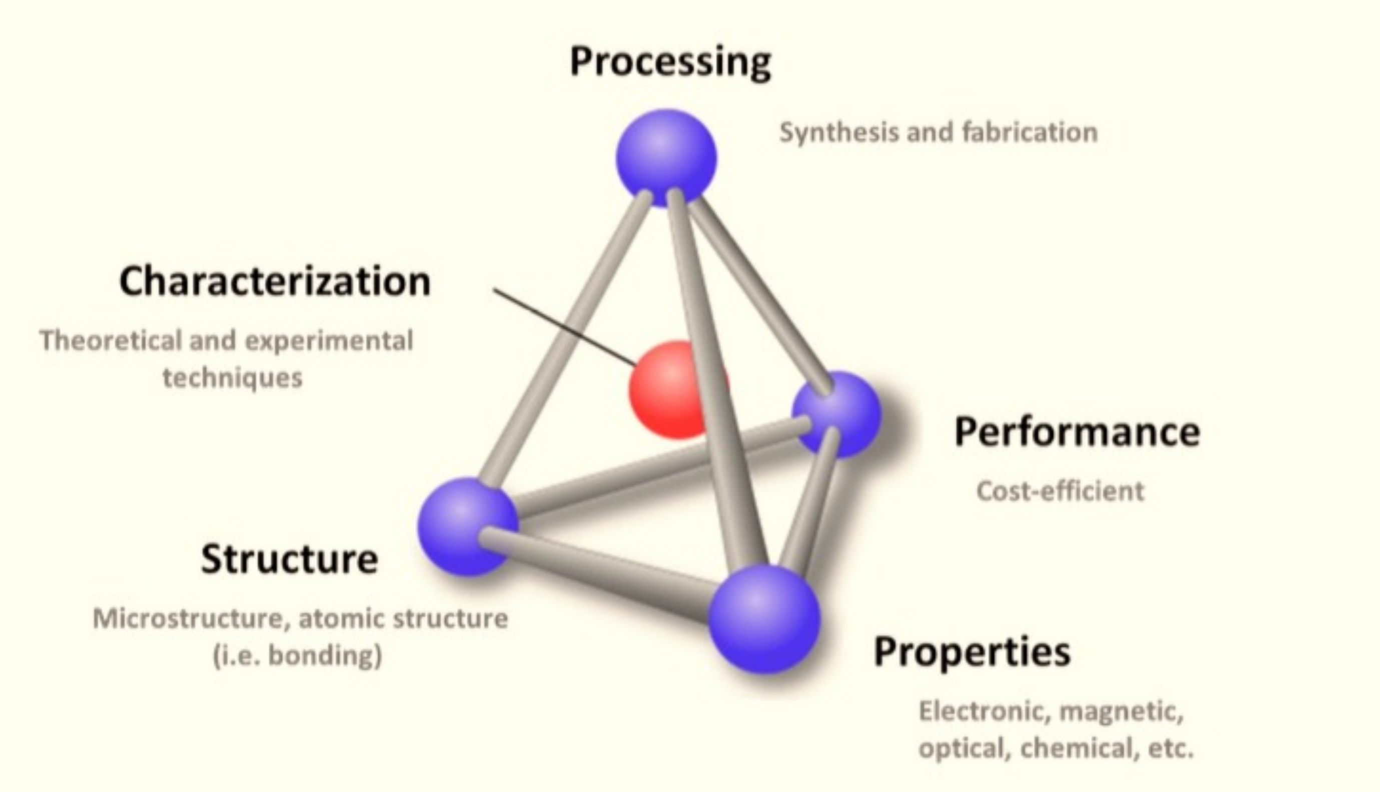
OUTLINE



- *What is Materials Science and Engineering ?*
- *Why should you be interested?*
  - *Latest Trends*
    - *Use of Machine Learning (Big Data-driven science)*
- *Fundamentals of Materials ~25 minutes*
- *Use of Machine Learning ~ 35 minutes*

WHAT ?

Materials Science & Engineering (MSE) Tetrahedron



- ***Materials Science develops the fundamental understanding of the relationships and structure of materials.***

- ***Materials Engineering uses this understanding to engineer (design) materials for real-life applications.***

MATERIALS CLASSIFICATION: LET'S IDENTIFY…

https://cdn-a.william-reed.com/var/wrbm_gb_food_pharma/storage/images/5/9/9/1/2501995-5-eng-GB/Coke-passes-green-for-Go!-with-UK-traffic-light-nutrition-labeling.jpg

https://img.etsystatic.com/il/d1eb21/1416040507/il_570xN.1416040507_bepi.jpg?version=0

http://cleanleap.com/sites/default/files/images/additional/2918/coke_bottles.jpg

Metals

Ceramics

Polymers

Materials Classification

Ready for Quiz!!!

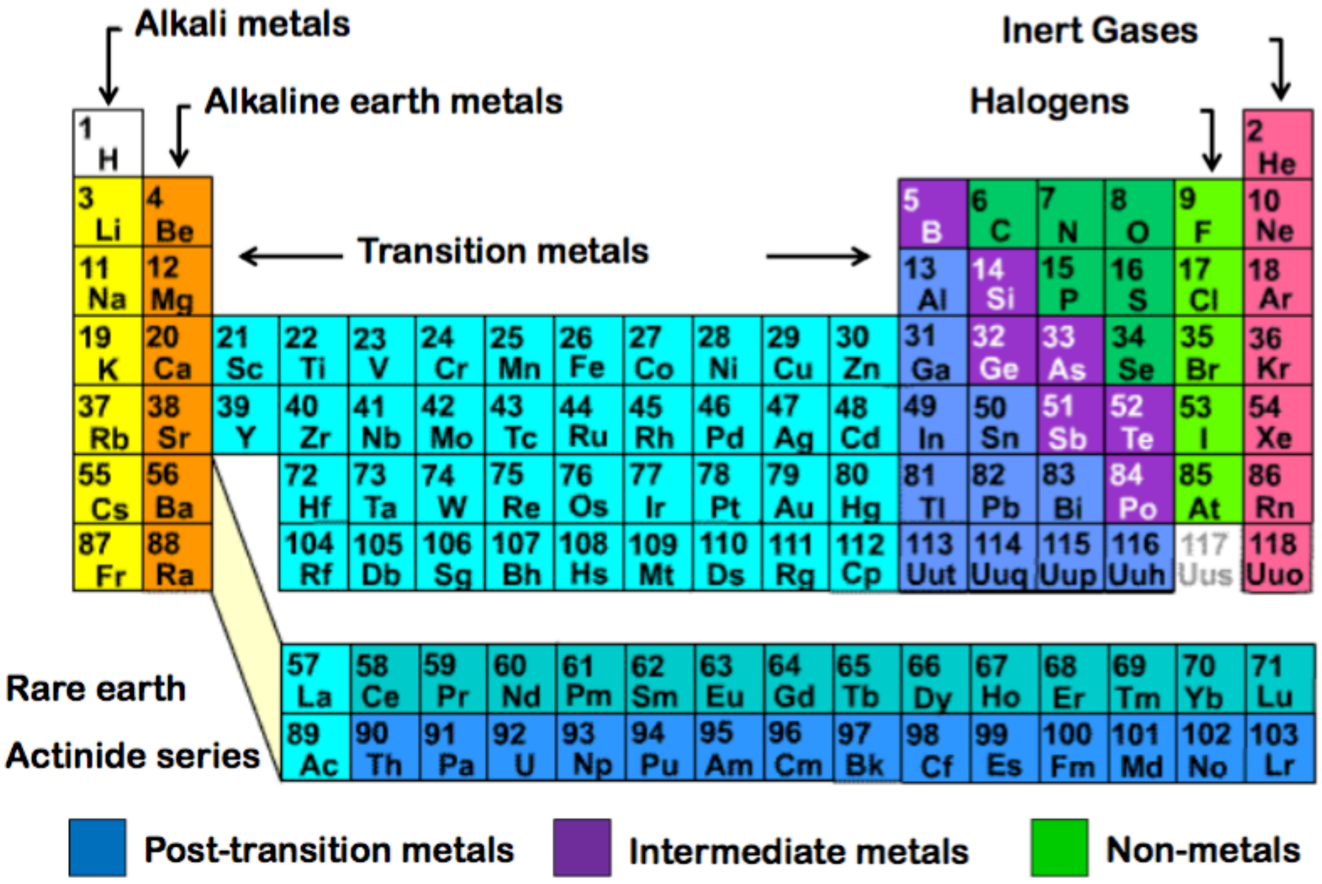
<u>Subclasses</u>: Composites, Semiconductors, Smart Materials , Biomaterials

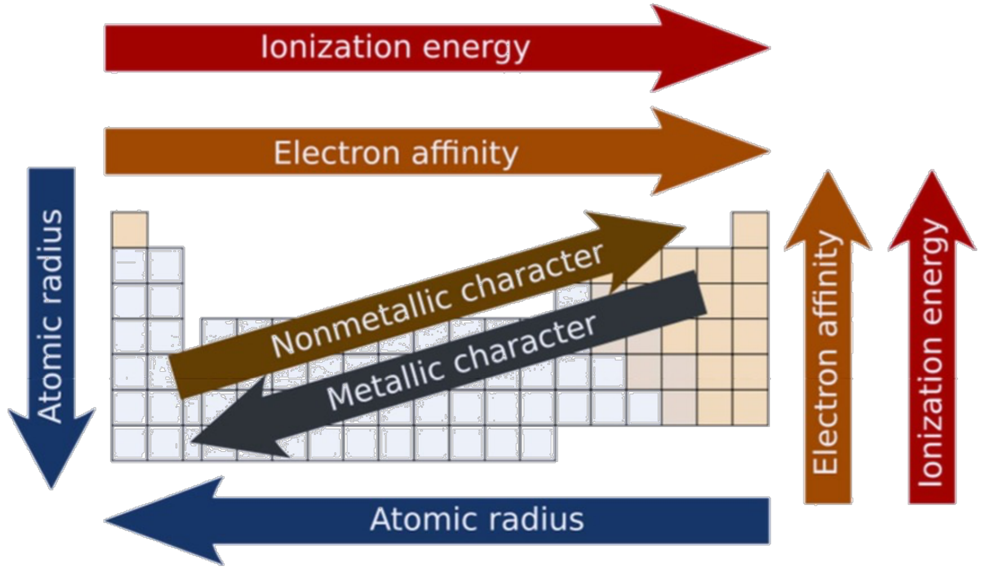
Can you Classify ?

Captured at Santa Fe, NM

Captured at Kouty, Czech Republic

PERIODIC TABLE





Characteristics of Metals, Ceramics and Polymers

- *Pure elements or combinations of metallic elements (alloys)*
- *Metallic bonding*
- *Good electrical conductors*
- *Good heat conductors*
- *Shiny appearance - not transparent*
- *Strong (High Strength, Stiffness)*
- *Deformable*
- *At times Magnetic (Fe)*

- *Compounds between metallic and non-metallic elements*
- *Ionic or covalently bonded*
- *Hard*
- *Brittle*
- *Electrical insulators*
- *Poor thermal conduction*
- *Heat & corrosion resistant*
- *Can be transparent or opaque*

- *Organic compounds based on C, H and other non-metallic elements*
- *Covalent and secondary bonding*
- *Huge variety of properties*
- *Low densities*
- *Non conductors*
- *Low melting points*
- *Can be very flexible*

Selection of Materials for Real-Life Applications

Determine the Application

Required Properties for appropriate uses

(Mechanical, Electrical, Thermal, Magnetic, Optical)

Pick Candidates

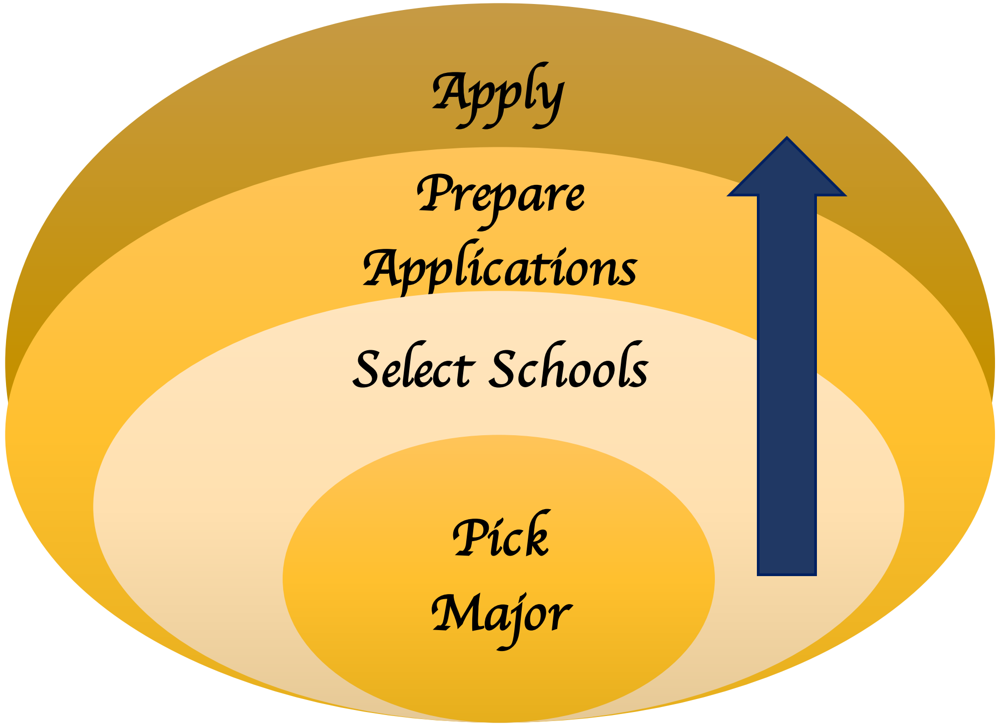
(Develop Understanding about Structure, Composition) – Does it relate to the Application of interest ?

Processing

(Make the materials and also improve properties according to the requirements of applications)

Casting, Sintering, Vapor Deposition, Doping, Forming, Joining, Annealing

Analogy



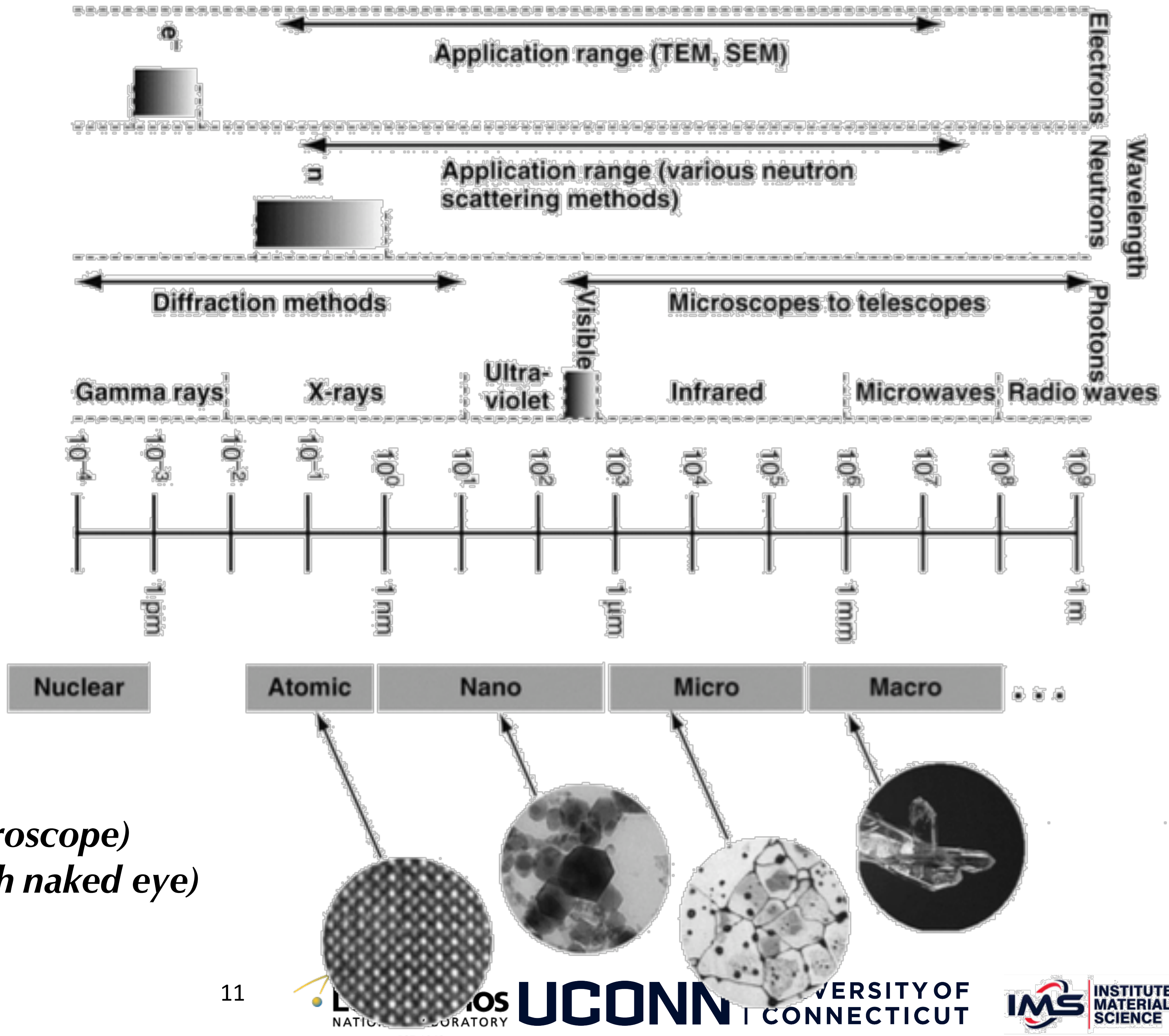
QUICK RECAP

- ***What is Materials Science and Engineering ?***

Materials Classes → Classification → Characteristics Applications → Relate Structures to Properties

STRUCTURE: LENGTH SCALES

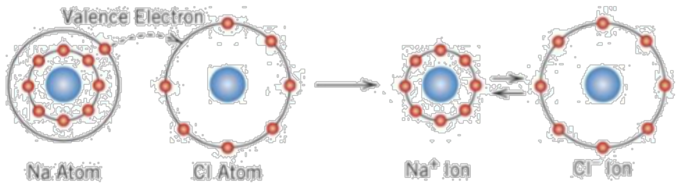
- ***Structures at many length levels***
  - *Electronic (sub-atomic)*
  - *Atomic (molecular, chemical)*
  - *Crystal (group of atoms)*



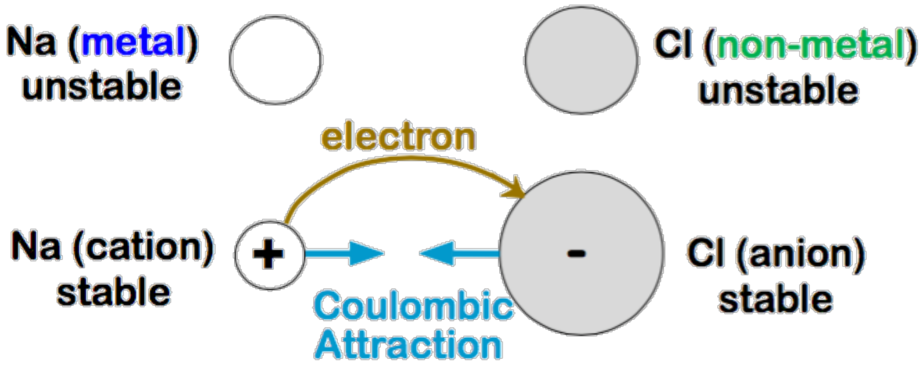
  - *Microstructure (visible with microscope)*
  - *Macrostructure (you can see with naked eye)*

BONDING: PRIMARY

Ionic Bonding



Example: NaCl

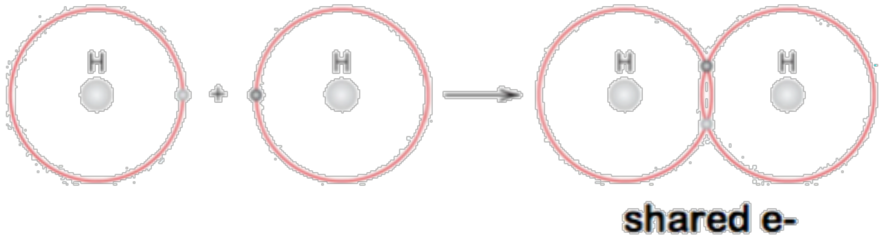


electrostatic attraction between oppositely charged ions

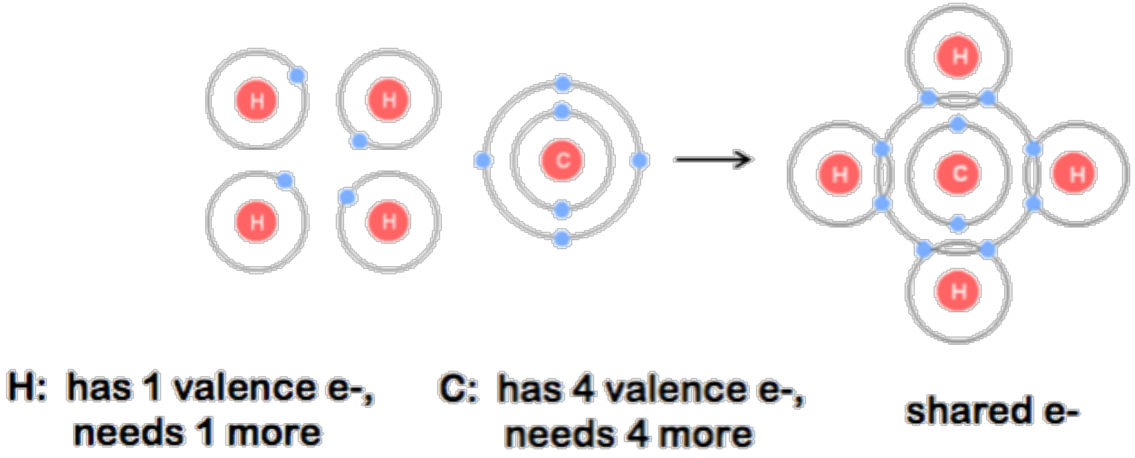
- Elements with large electronegativity difference
- Non-directional, strong
- Generally Ceramics
  - Non-conducting
  - Hard
  - Brittle
  - Thermal Insulator

Covalent Bonding

H_2



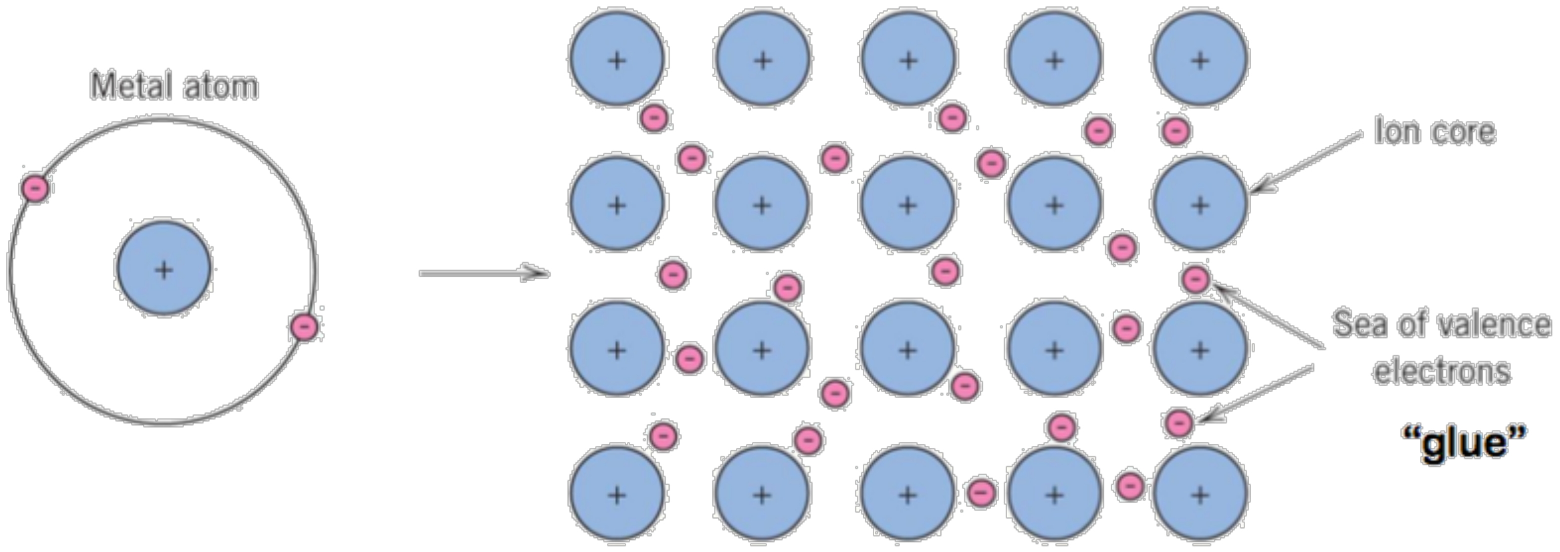
CH_4



- Elements with small electronegativity difference
- Generally polymers, semiconductors
  - Non-conducting
  - Hard
  - Brittle
  - Insulating

BONDING: PRIMARY

Metallic Bonding



- Availability of free electrons
- Mixed Ionic-covalent character
- Generally metals, alloys
  - Good electrical and thermal conductors

Quick Quiz!!!

- *Identify types of bonding*
  - *Diamond*
  - *Glass (silicates)*
  - *Plastics*

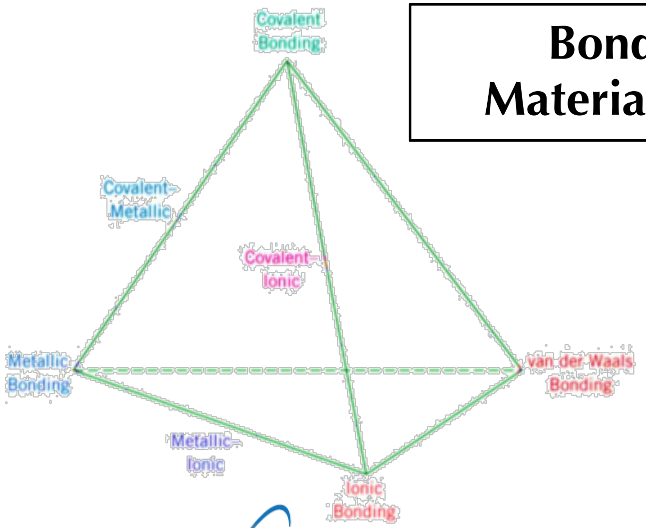
Secondary Bonding

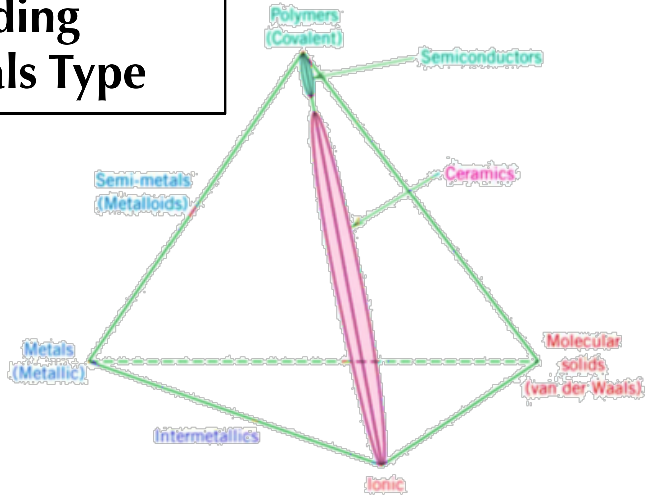
Bonding and Properties

Bond Type	Bond Energy (KJ/mol)
Ionic Non-directional (ceramics)	625-1550
Covalent Directional (polymers, semiconductors, even few ceramics)	520-1250
Metallic Non-directional (metals)	100-800
Secondary Directional (inter-chain polymers, inter-molecular in molecular crystals)	<40

- Higher the bond energy, more difficult it is to break the bond
- <u>Consequences…</u>
  - Strongly bonded compounds – Higher melting temperature
  - Weakly bonded compounds – Higher coefficients of thermal expansion

Bonding
Materials Type





➢ ***What is Materials Science and Engineering ?***

Materials Classes

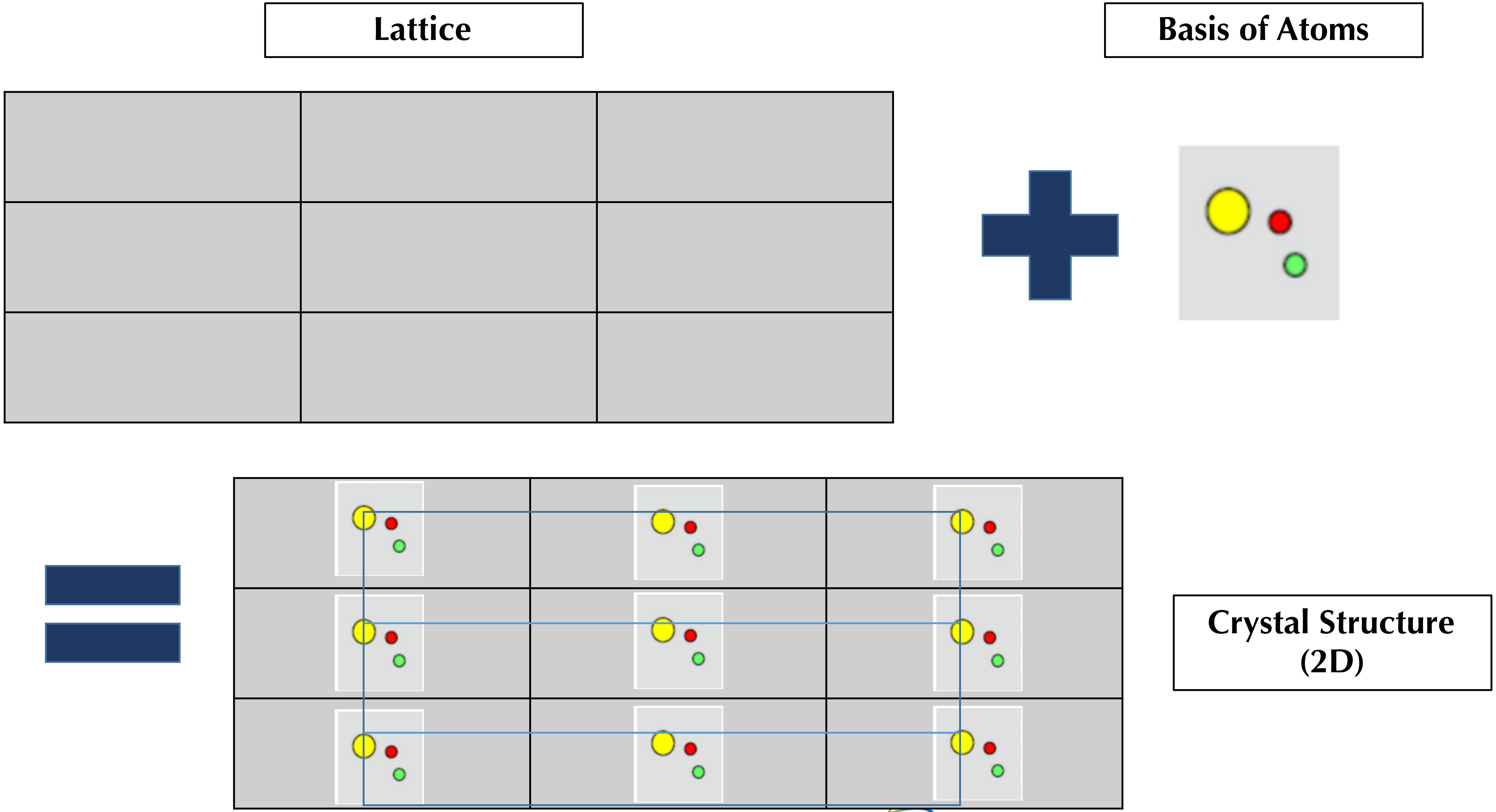
Classification

Characteristics

Applications

Relate Structures to Properties

CRYSTAL STRUCTURES IN 2D



How can you make them !!!

Let's play with Vesta & More!
(Follow Handouts)

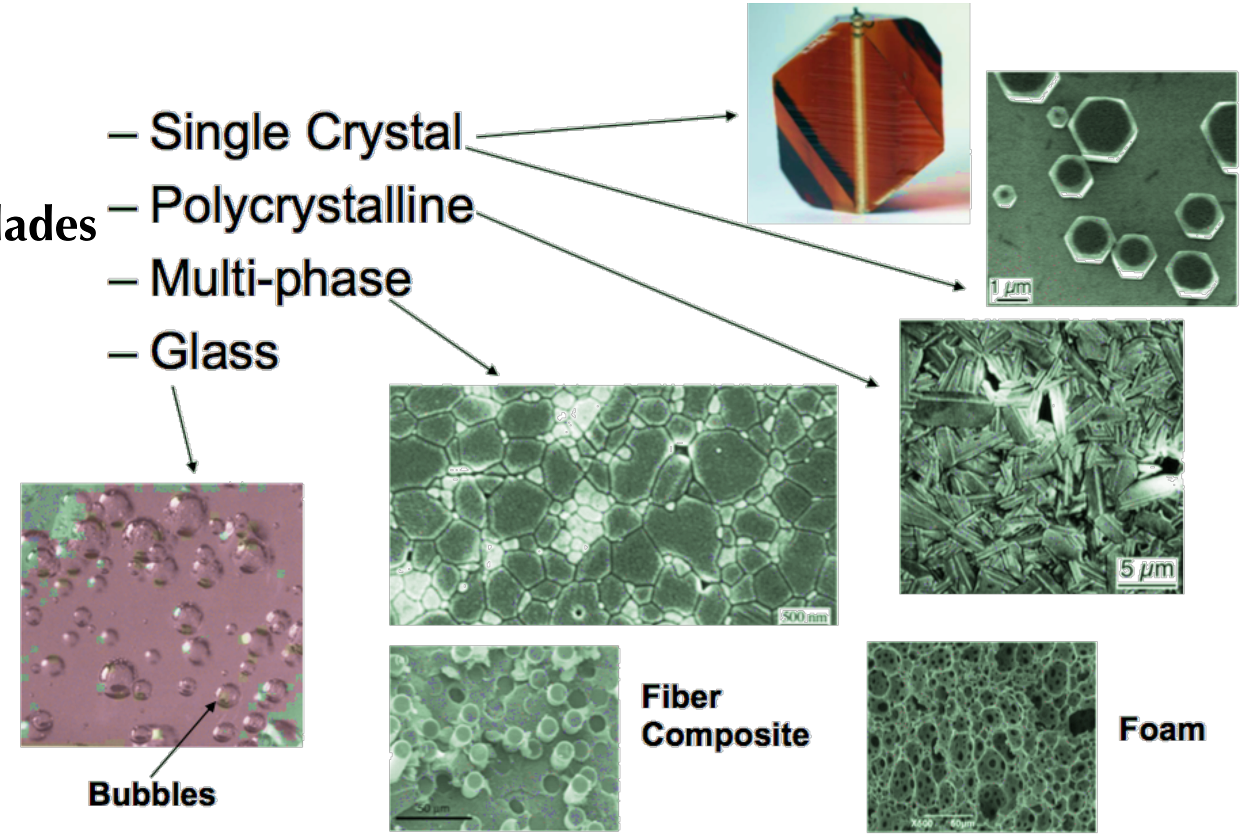
- Key Parameters to look for …
  - Lattice parameters
  - Unit cell volume
  - Angles
  - Bond Length
  - Lattice Planes

- Can you classify the type of bonding in these ?
- Atomic Packing Factor, Coordination numbers, types of crystals systems

Crystals as Building Blocks

- Engineering Applications

  - Single Crystals
    - Anisotropic
    - Diamond single crystals for abrasives
    - Single crystal nickel alloys for turbine blades



  - Polycrystalline
    - Most engineering materials are polycrystalline
    - Isotropic

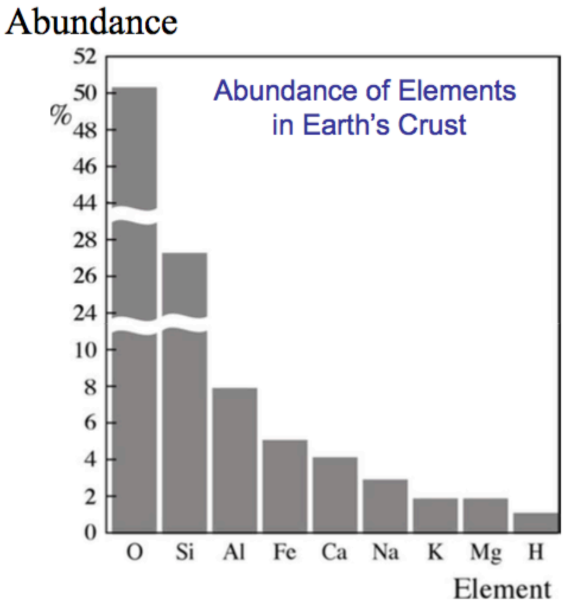
MATERIALS PROCESSING (ENGINEERING)

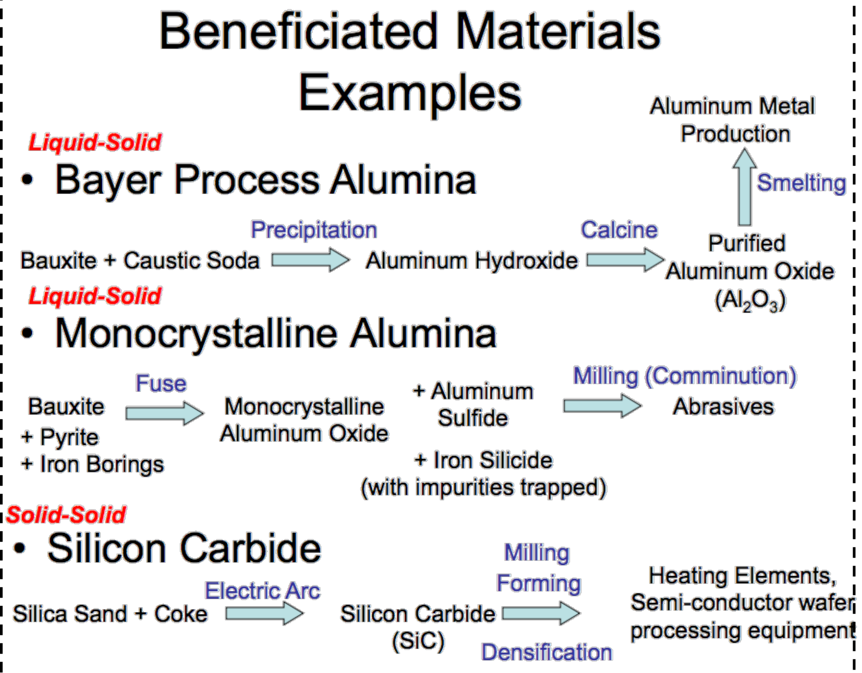
Raw Materials

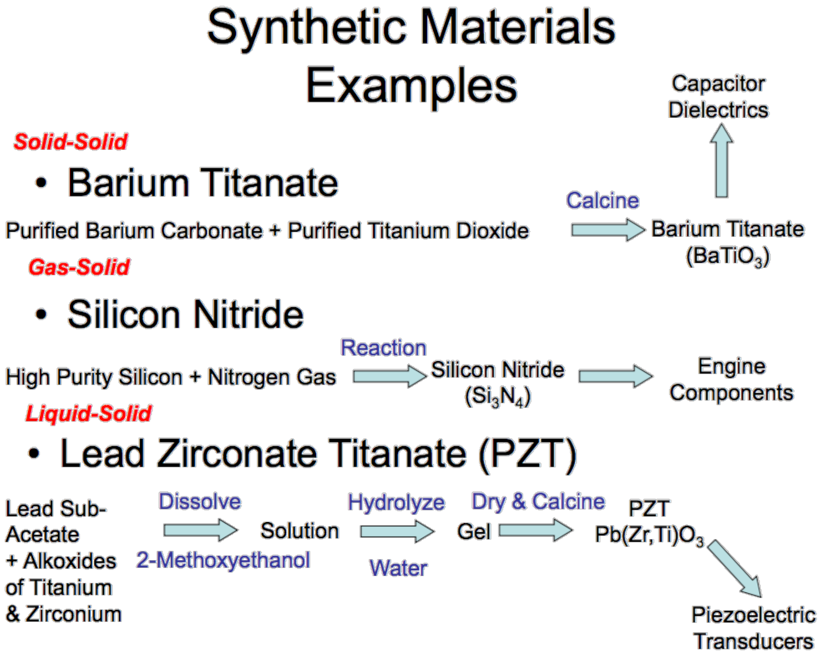
Naturally Occurring Minerals

Beneficiated Materials

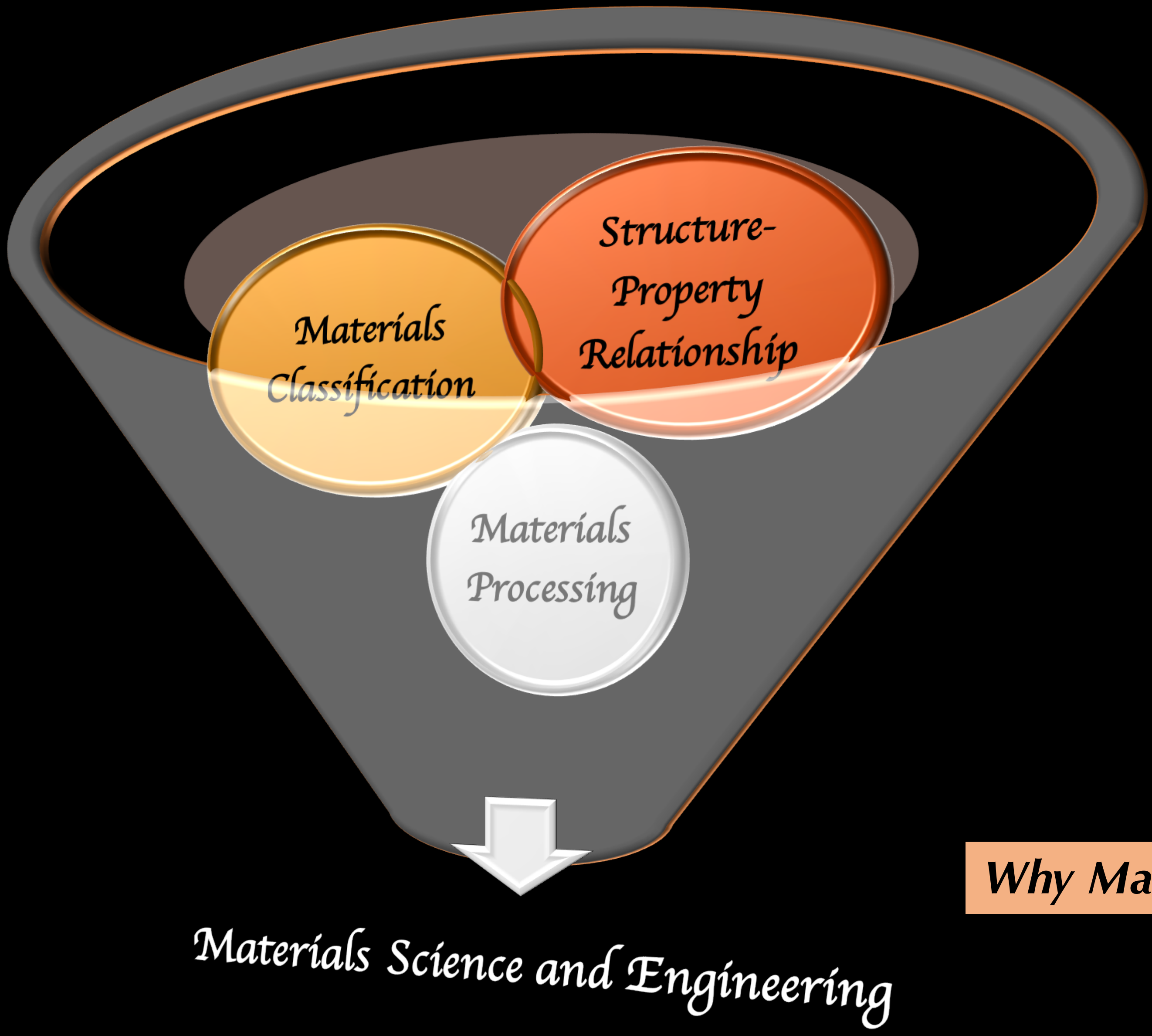
Synthetic Materials





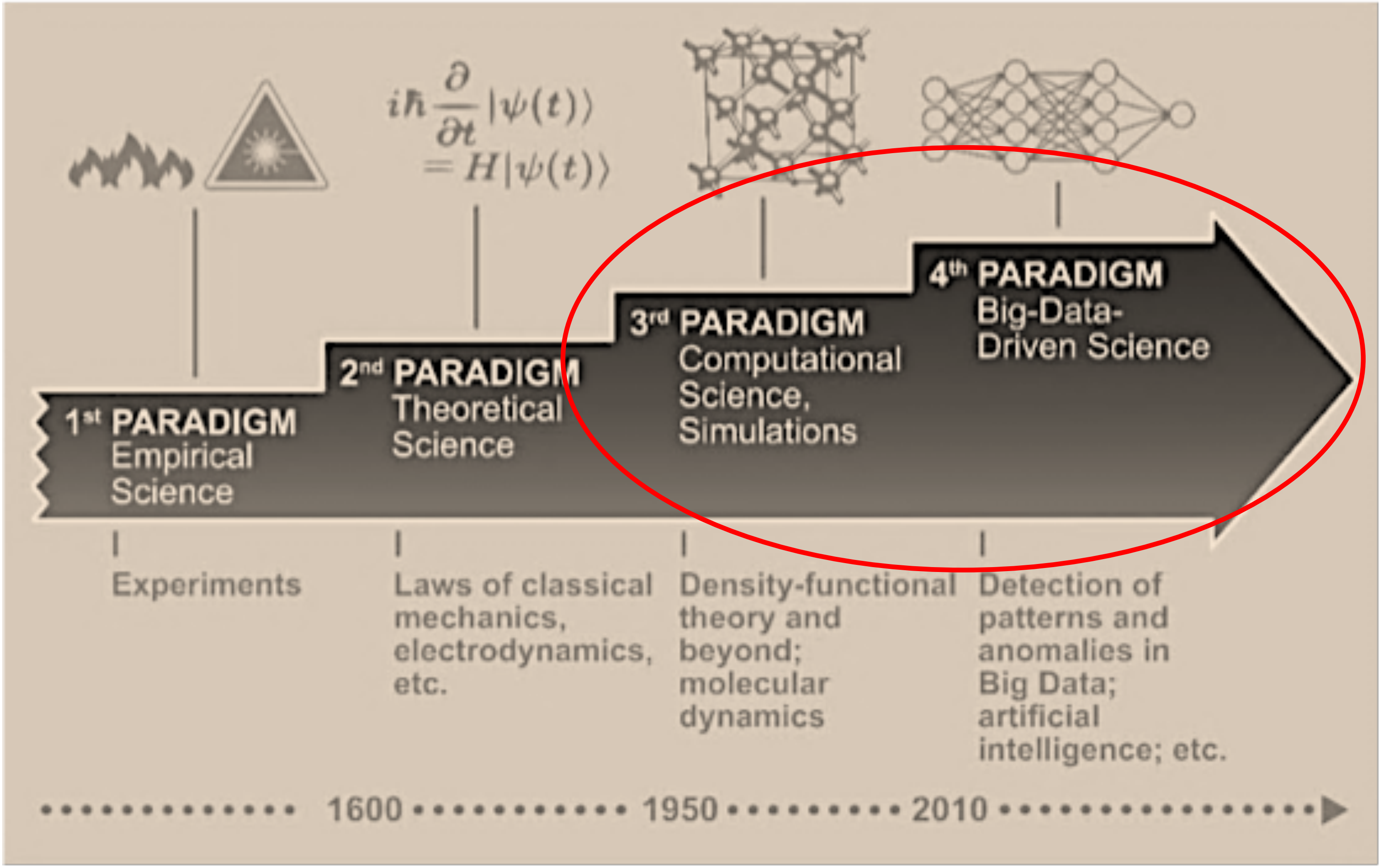


Materials Science and Engineering



Why Machine Learning?

WHY MACHINE LEARNING IN MATERIALS SCIENCE?



https://www.nomad-coe.eu/uploads/images/News/NOMAD_new_paradigms_material_science.png

Useful Weblinks:

NOMAD:

https://www.nomad-coe.eu/index.php?page=centre-of-excellence

Materials Project:

https://materialsproject.org

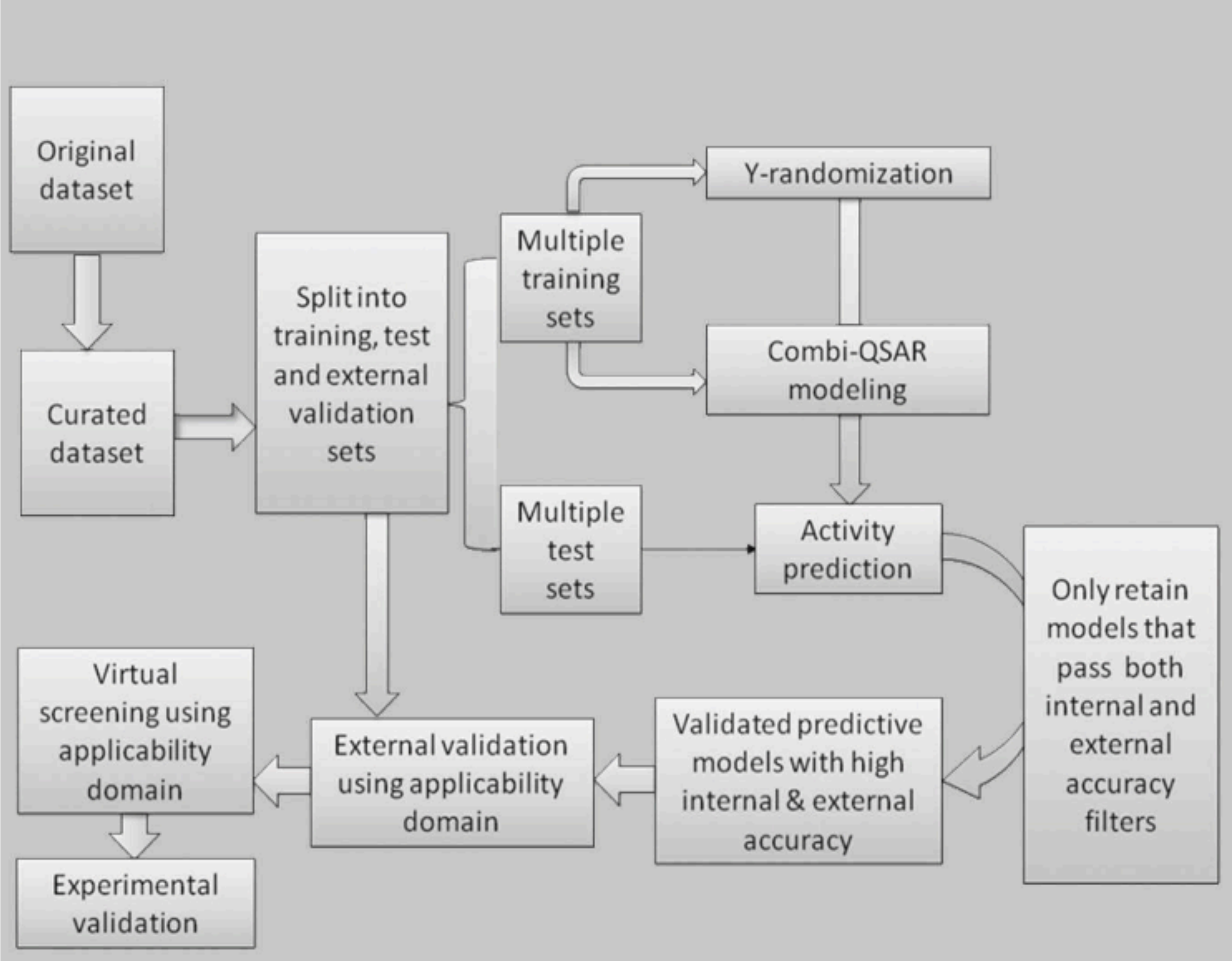
ICSD:

https://icsd.fiz-karlsruhe.de/search/basic.xhtml

Los Alamos NATIONAL LABORATORY EST.1943 UCONN UNIVERSITY OF CONNECTICUT IMS INSTITUTE OF MATERIALS SCIENCE

Introduction to Machine Learning

- QSPR (Quantitative Structure Property Relationship) -> Establish relations between structure of a molecule and its chemical property



Example Problem ?

Tropsha, Alexander. "Best practices for QSAR model development, validation, and exploitation." Molecular informatics 29.6-7 (2010): 476-488.

Introduction to Machine Learning: Drug Discovery

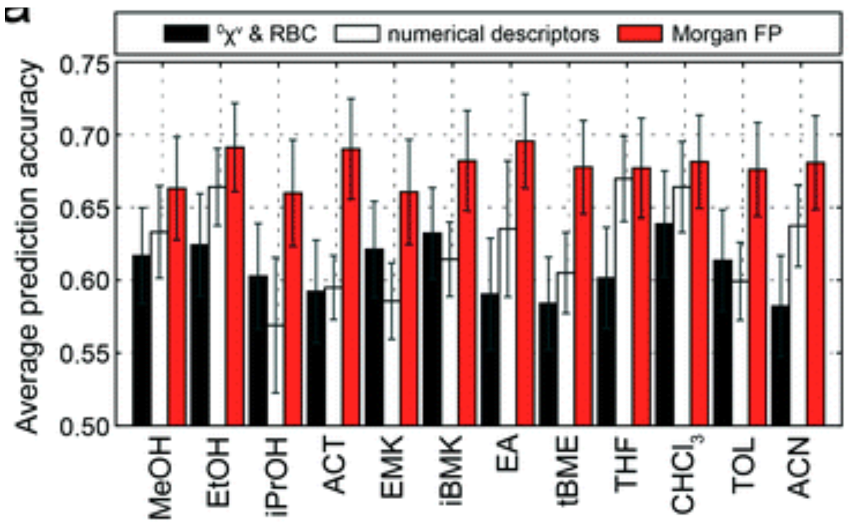
- Crystalline or non-crystalline

	Non-crystalline	crystalline	Total
Training	13440	13453	22733
Test	4480	4485	8965
Total	17920	17938	35858

- Molecular Descriptors
- Build models
- Accuracy: 79%

Wicker, J. G., & Cooper, R. I. (2015). Will it crystallise? Predicting crystallinity of molecular materials. CrystEngComm, 17(9), 1927-1934.

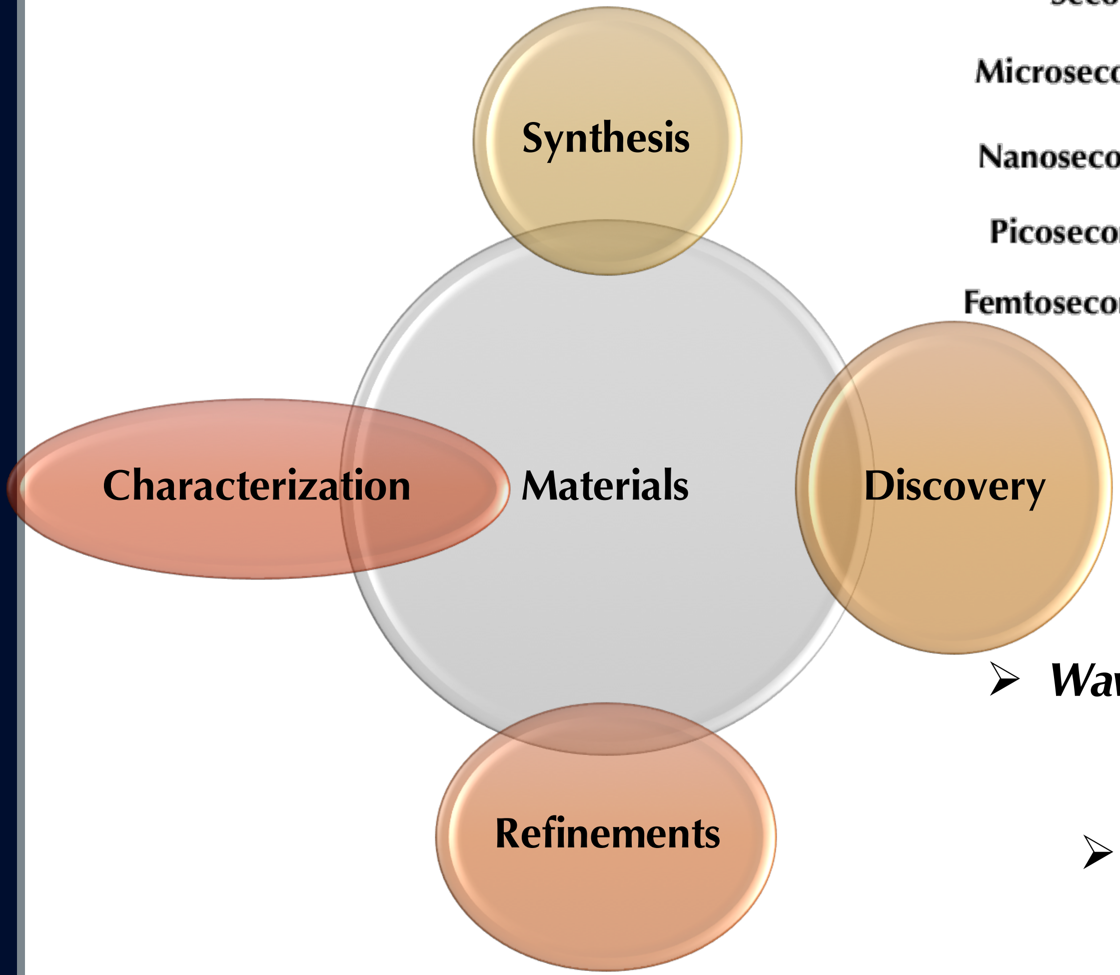
- Crystalline or non-crystalline, solvents dependency
- 319 small molecules in 18 different solvents
- Total of 5710 compounds
  - Training (50%)
- Molecular Descriptors, build models, accuracy

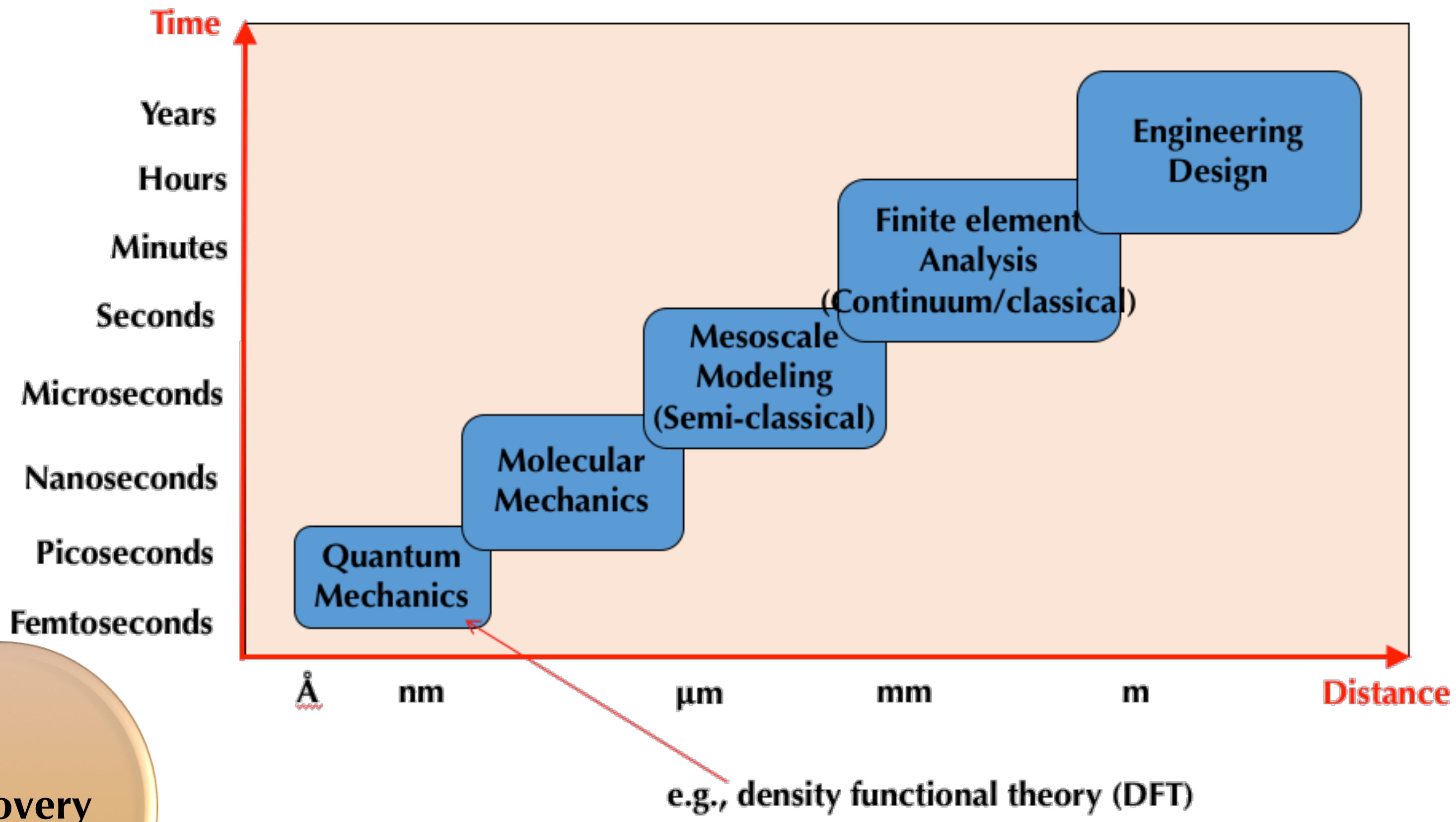


Pillong, M., Marx, C., Piechon, P., Wicker, J. G., Cooper, R. I., & Wagner, T. (2017). A publicly available crystallisation data set and its application in machine learning. CrystEngComm, 19(27), 3737-3745.

COMPUTATIONAL MATERIALS SCIENCE? (MACHINE LEARNING IN MATERIALS SCIENCE)

- *Need for computational approach applicable to all length scales*





- *Wavefunction -> DFT -> tight binding, force fields -> continuum mechanics, rate equation -> finite element modelling*
- *Electronic -> atomistic -> microstructure -> macroscale*

How Does it Work?

Simulated (Synthetic) Data

- *Chemical formula*
- *DFT Simulations:*
  - *Optimized Geometry*
  - *Force fields*

Descriptors (Features)

- *Structural or chemical*
- *Sometimes correlated*
- *Identifying most important descriptors*

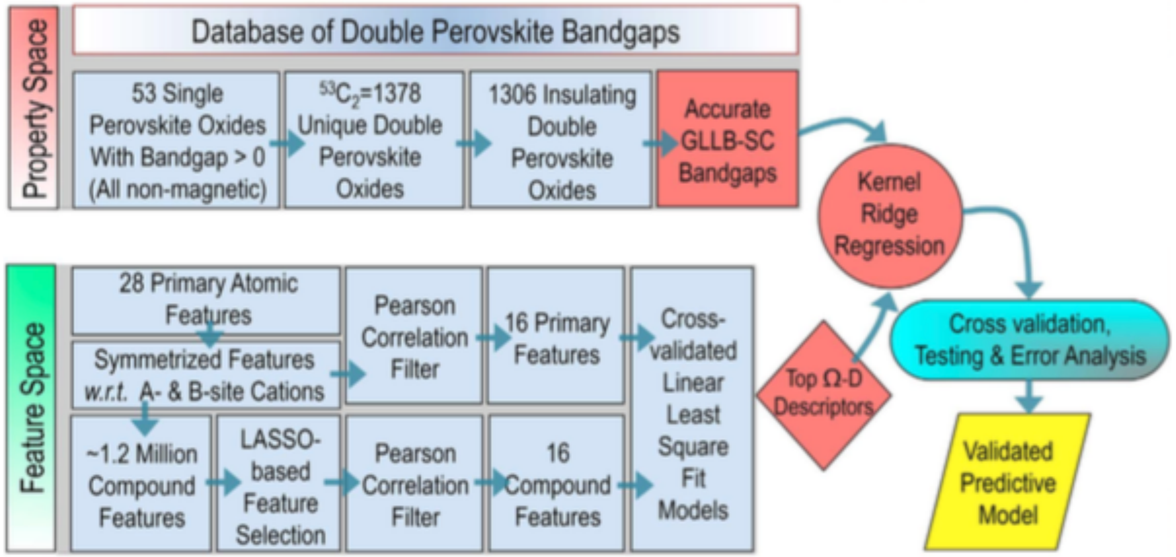
Endpoint

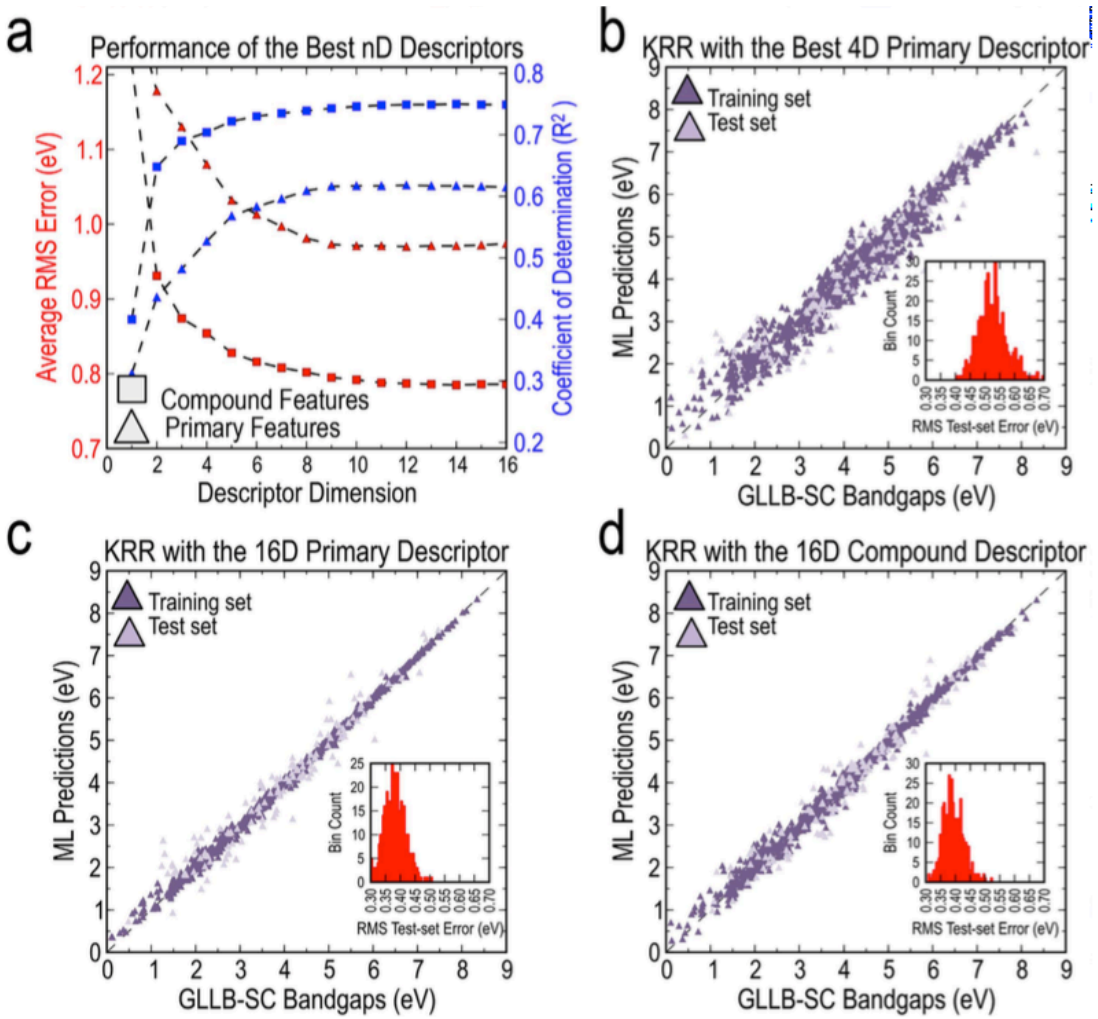
- *Energy*
- *Conductivity*
- *Band Gap*
- *Polarization*
- *Magnetic Property*

- ***Establish a quantitative trend between descriptors and endpoint***
  - ***Algorithms dependent***
- ***If based on simulated data only, experimental validations are needed***
- ***Accelerate the process of new materials discovery***

Example Problem ?

Predict Band Gaps for Double Perovskites

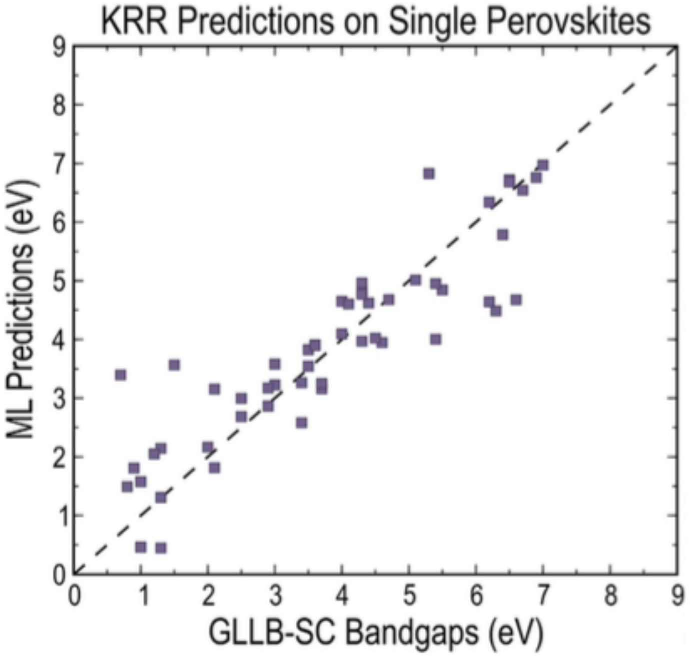




Pilania, G., Mannodi-Kanakkithodi, A., Uberuaga, B. P., Ramprasad, R., Gubernatis, J. E., & Lookman, T. (2016). Machine learning bandgaps of double perovskites. Scientific reports, 6, 19375.

Predict Band Gaps

Performance of KRR ML models with primary descriptors					Performance of KRR ML models with compound descriptors				
Descriptor	Training set (90%)		Test set (10%)		Descriptor	Training set (90%)		Test set (10%)	
Dimension	*rms* error (eV)	R^2	*rms* error (eV)	R^2	Dimension	*rms* error (eV)	R^2	*rms* error (eV)	R^2
1-D	0.959	0.634	1.056	0.546	1-D	0.954	0.638	1.056	0.545
2-D	1.059	0.554	1.114	0.493	2-D	0.888	0.686	0.879	0.685
3-D	0.689	0.811	0.867	0.692	3-D	0.774	0.762	0.777	0.753
4-D	0.306	0.963	0.501	0.897	4-D	0.716	0.796	0.742	0.775
5-D	0.270	0.971	0.529	0.885	5-D	0.566	0.872	0.637	0.834
6-D	0.302	0.968	0.521	0.889	6-D	0.502	0.900	0.594	0.855
7-D	0.356	0.949	0.510	0.892	7-D	0.426	0.928	0.554	0.875
8-D	0.281	0.969	0.429	0.925	8-D	0.437	0.924	0.563	0.870
9-D	0.219	0.981	0.406	0.932	9-D	0.399	0.937	0.550	0.876
10-D	0.139	0.992	0.397	0.935	10-D	0.300	0.964	0.484	0.904
11-D	0.178	0.987	0.413	0.930	11-D	0.249	0.975	0.455	0.915
12-D	0.074	0.998	0.393	0.936	12-D	0.222	0.980	0.451	0.917
13-D	0.137	0.993	0.365	0.944	13-D	0.187	0.986	0.418	0.928
14-D	0.110	0.995	0.397	0.935	14-D	0.169	0.989	0.413	0.930
15-D	0.087	0.997	0.377	0.941	15-D	0.144	0.992	0.376	0.942
16-D	0.080	0.997	0.371	0.939	16-D	0.132	0.993	0.360	0.947



Pilania, G., Mannodi-Kanakkithodi, A., Uberuaga, B. P., Ramprasad, R., Gubernatis, J. E., & Lookman, T. (2016). Machine learning bandgaps of double perovskites. Scientific reports, 6, 19375.

Key words

- Machine Learning:
- Alternative method where all possible data don't need to be computed.
  - Conventional (Shallow) Learning:
    - Based on existent experimental data /theoretical simulations
      - Diverse dataset
    - Define a Target Property – Endpoint (This is the property you want to predict)
    - Data mining: Extracting information already available, as applicable to predict Endpoint
    - Descriptors: Properties that are related to target property, which, when varied, may affect the target property

Key words

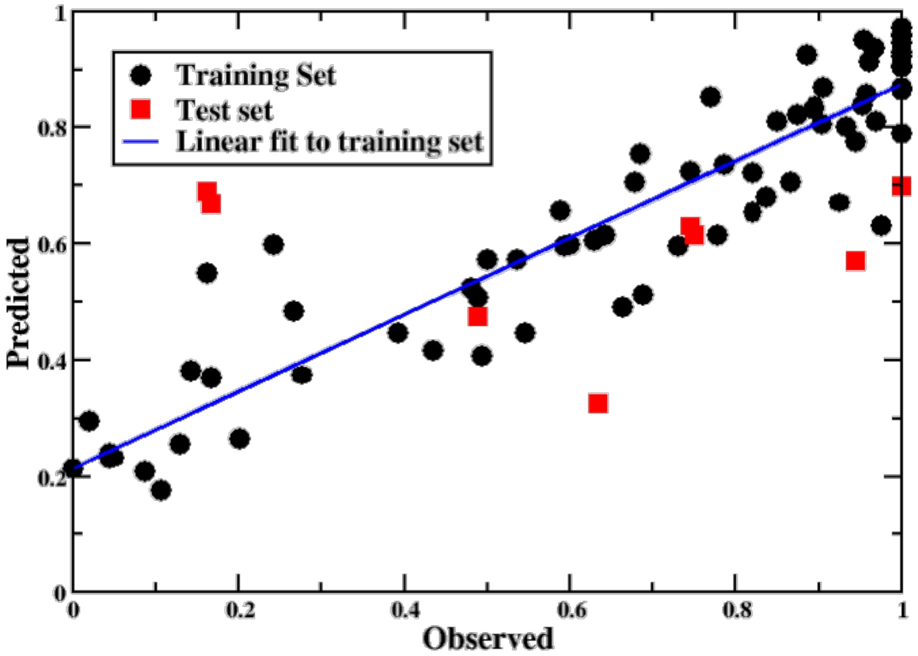
- Identify what type of problem it is: classification or regression

- Choosing Algorithms (Classification or Regression):
  - Linear
  - non-linear (Random Forest, Support Vector Machine etc.)

- Model Development:
  - Divide dataset (training (on which your model will be built on), test (on which you will test your model) and validation set (test the true capability of model)
  - Both training and test sets are subjected to vary and are used during model building
  - Validation set is kept untouched and used at the very end.

- How do you decide what's a good model ???

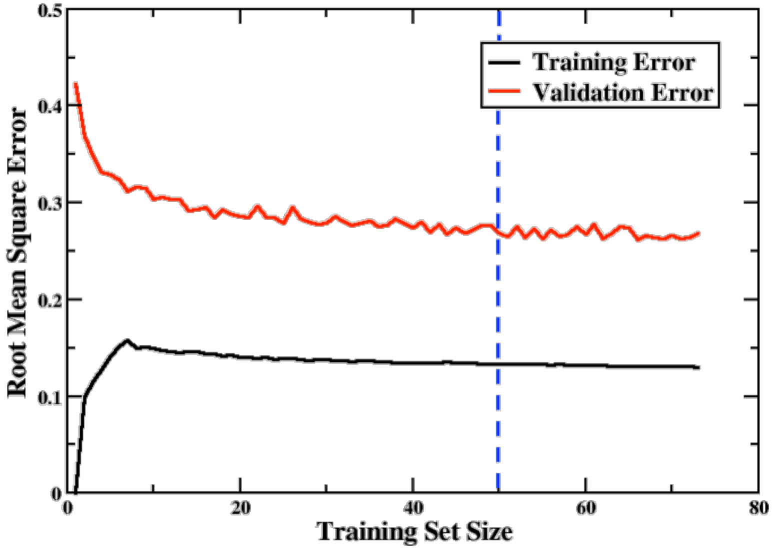
Key Points with quick Example plots

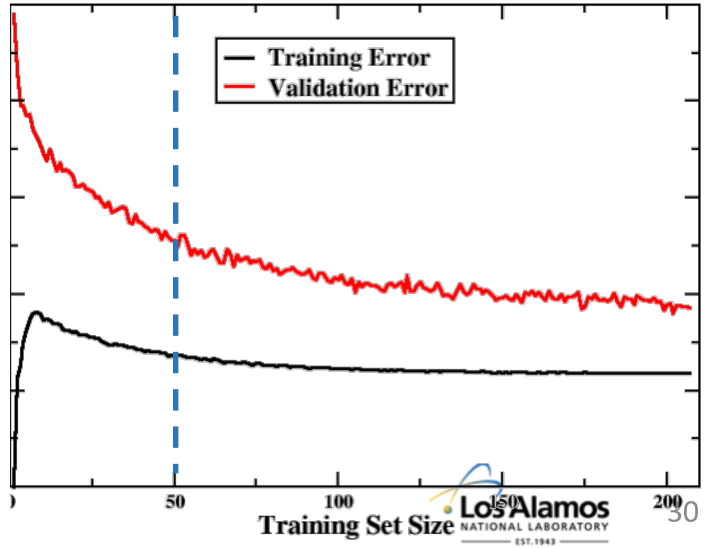


Dataset	RMSE	MAE	R^2
Train	0.14	0.11	0.86
Test	0.19	0.15	0.00

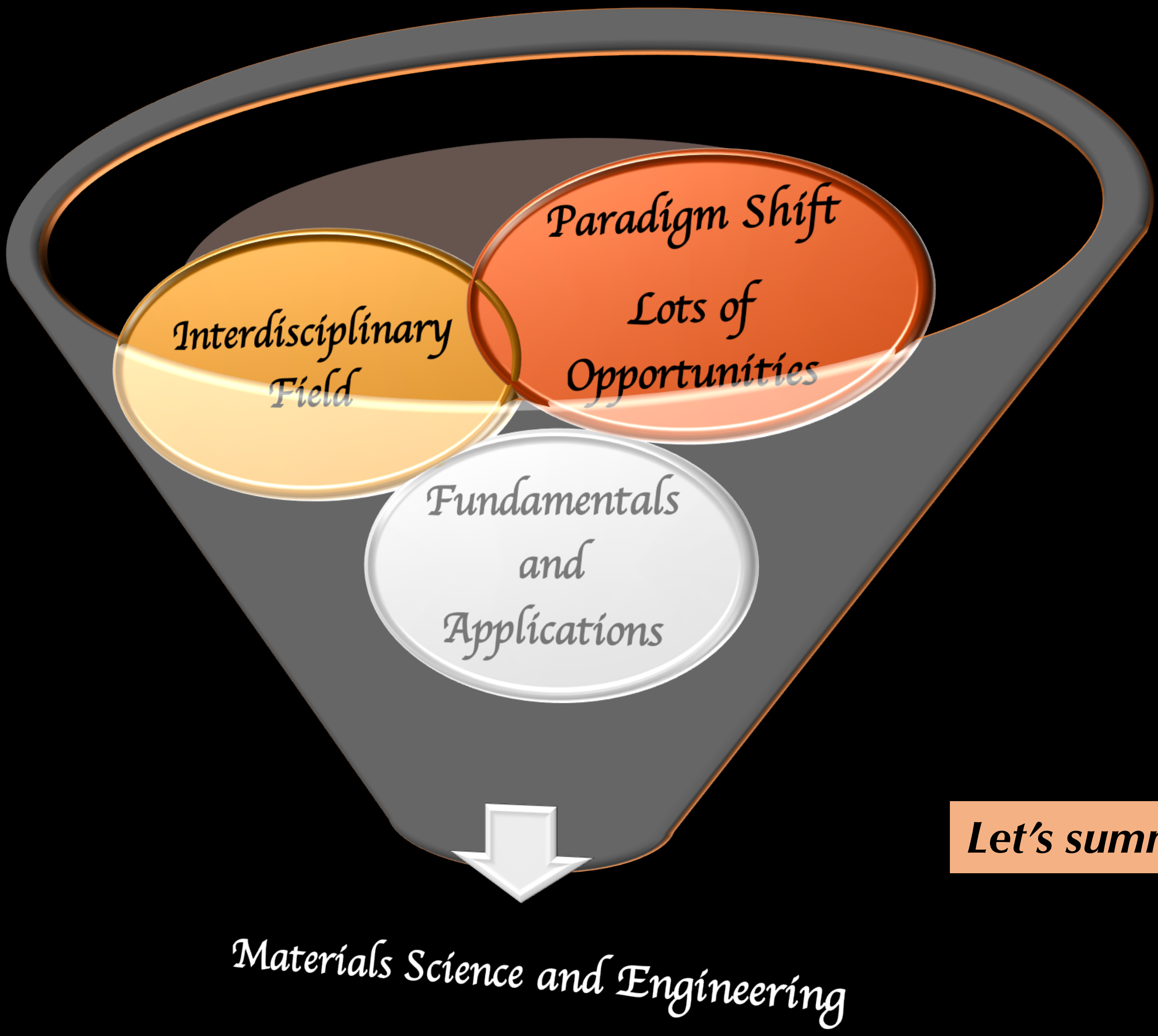
Training Set Size = 70, Test Set Size = 8

- For a comparatively small dataset sizes, these error values for training and test sets are going to vary: statistical fluctuation of some sorts
- Cannot be used to determine the optimal performance of a model
- Learning Curve:
  - Model predictability over a varying number of training sets





MATERIALS SCIENCE AND ENGINEERING



Let's summarize!

SUMMARY I (WHAT I BRIEFLY TALKED ABOUT)

Medicinal Applications

Computational / Experimental Data

Machine Learning

Crystallizability of organic compounds

Synthetic Data

Learn the physics

Novel Organic Ferroelectrics

Novel Inorganic functional materials

- Multiferroicity (magnetic and polar)
- Thermal conductivity
- Large Optical Band-gap Materials

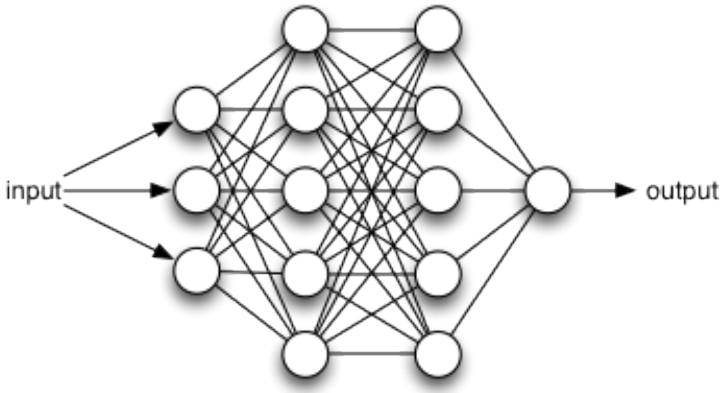
Technological Applications

DAILY USES (WHAT I DIDN'T)

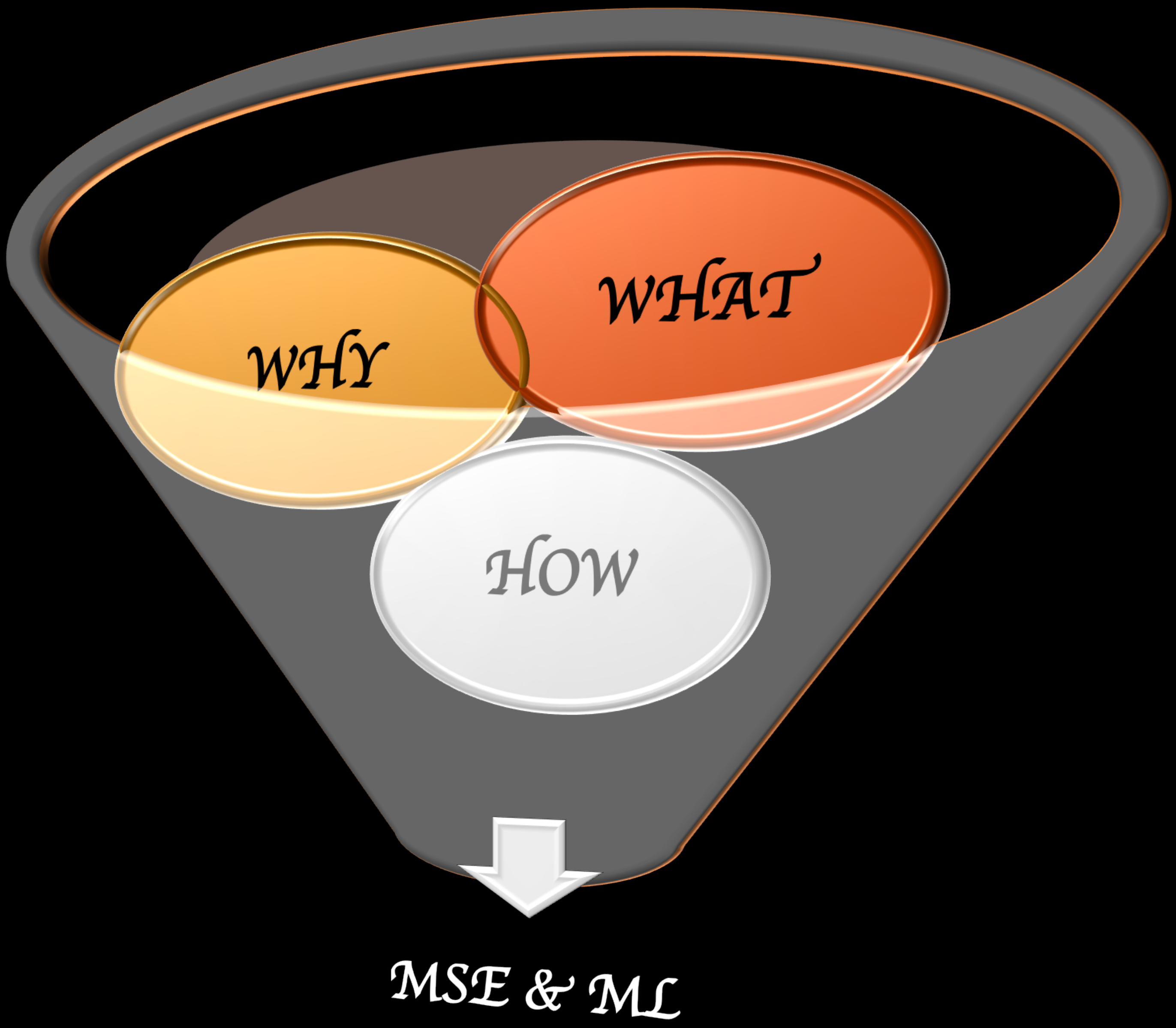
Search Engines
Maps
E-mails
Shopping
Self-driven cars
Face-recognition
Smartphones

Machine (Deep) Learning

Big Data-set
Train Neural-net with layers
(No descriptors)
Prediction



Machine Learning in Materials Science



ACKNOWLEDGEMENTS

Questions!!!

Supercomputing Challenge Kickoff Organizers

LANL Advisor: Dr. Jian-Xin Zhu

UConn Advisor: Prof. Serge Nakhmanson

All Attendees of this workshop

Contact: ayana.ghosh@uconn.edu