Machine Learning Accelerated CALYPSO Structure Prediction Method

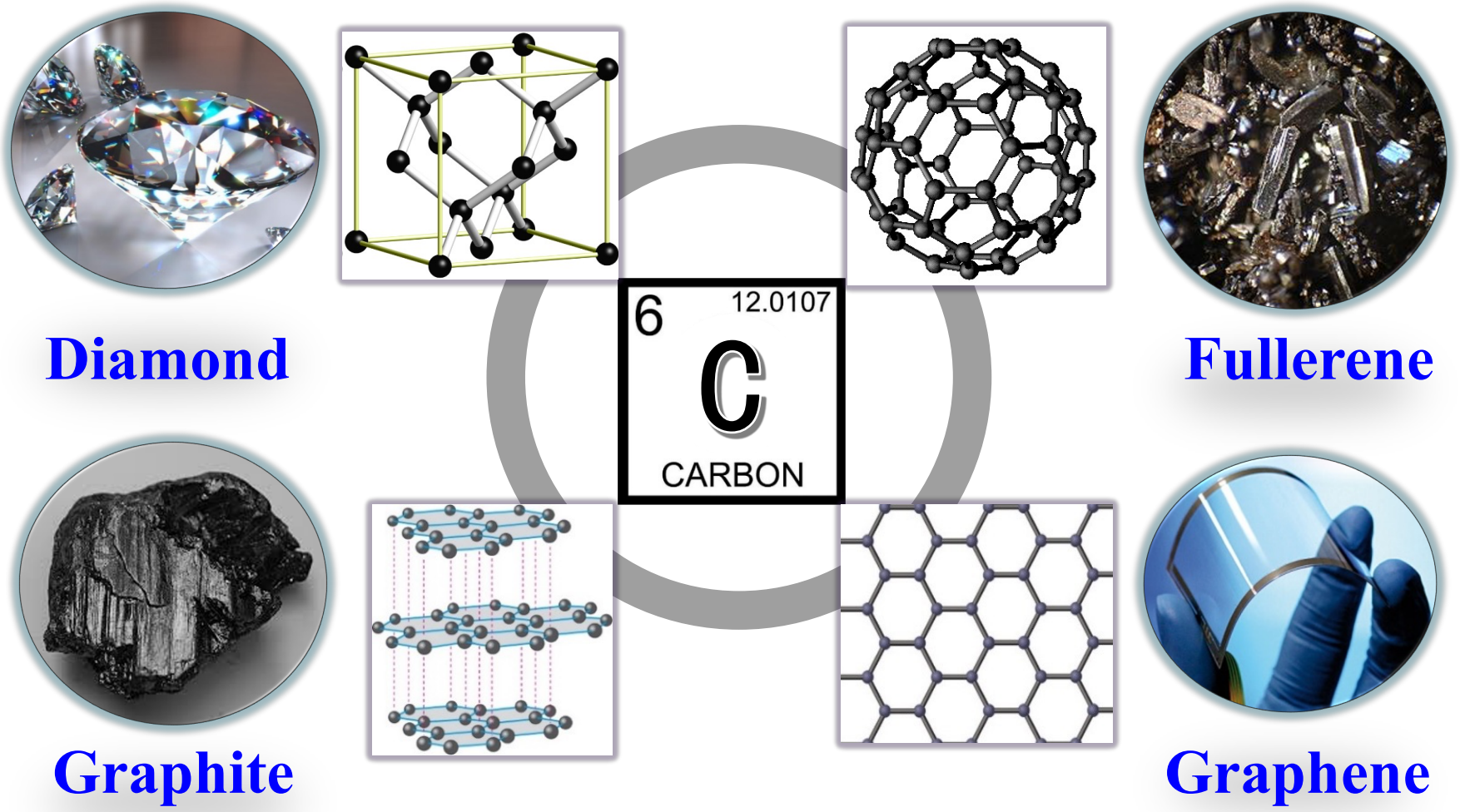
Pengyue Gao, Jian Lv, Yanchao Wang, and Yanming Ma

Jilin University, China

Quantum Computing & Machine Learning Workshop, Jilin, August 6 - 8, 2024

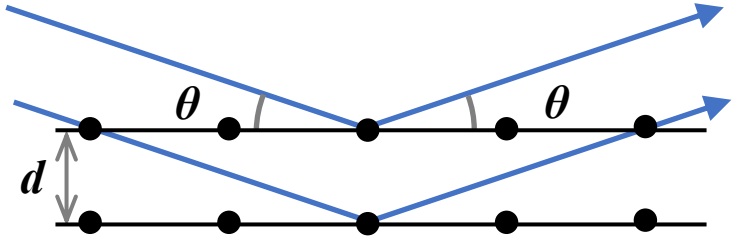
Crystal Structure

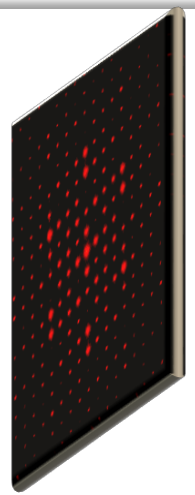
The crystal structure determines the macroscopic properties of the matter and is the basis for the study of material science.

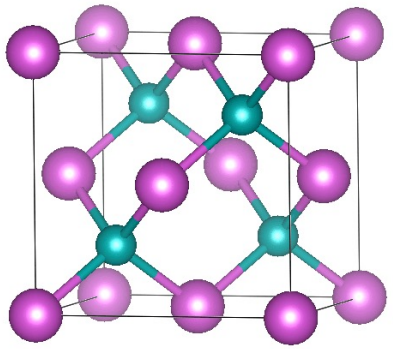


Experimental method for determining crystal structures

$$n\lambda = 2d\sin\theta$$





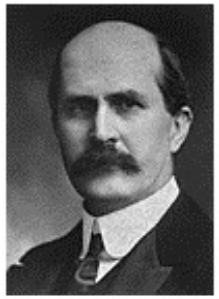


Crystal Structure Solution Based on XRD Technique

The Nobel Prize in Physics 1914

Max von Laue

The Nobel Prize in Physics 1915



Sir William Henry Bragg

William Lawrence Bragg

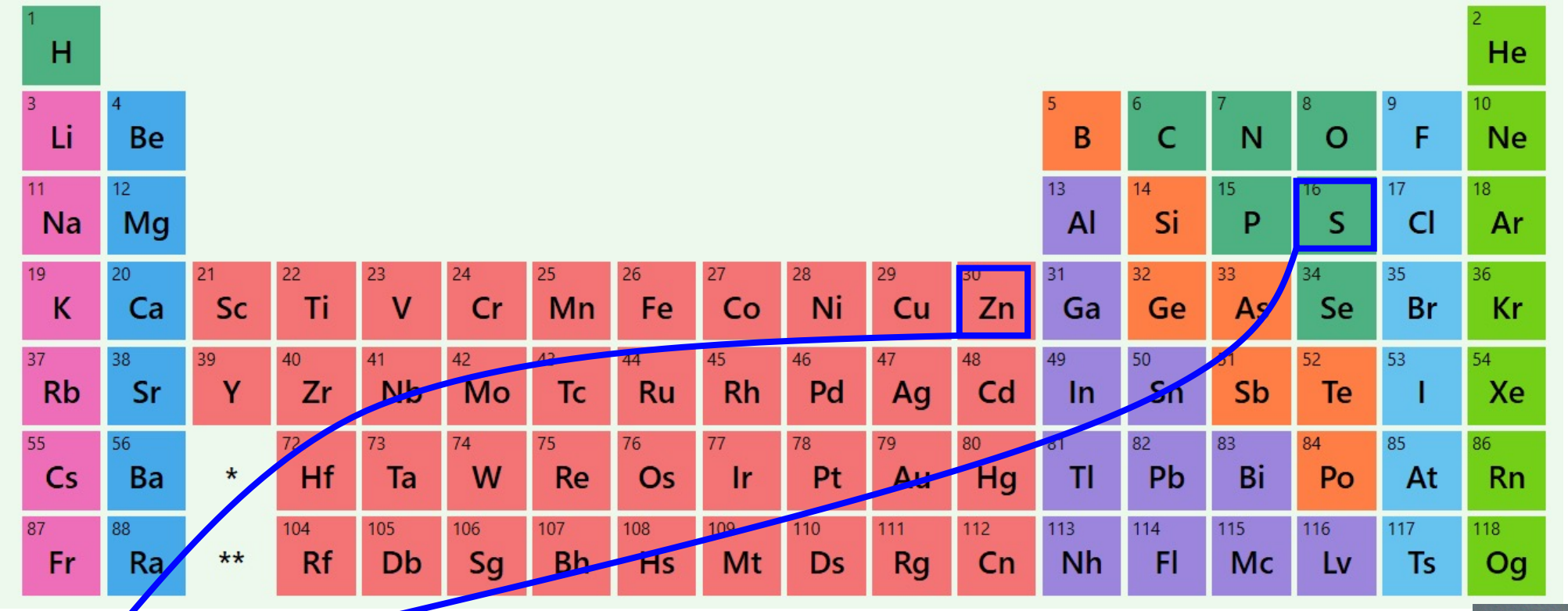
The Nobel Prize in Chemistry 1985

Herbert A. Hauptman

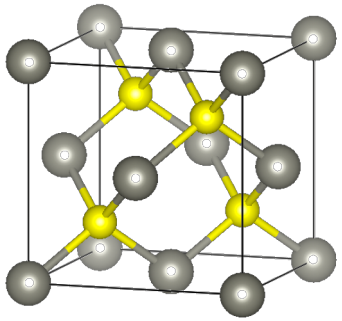
Jerome Karle

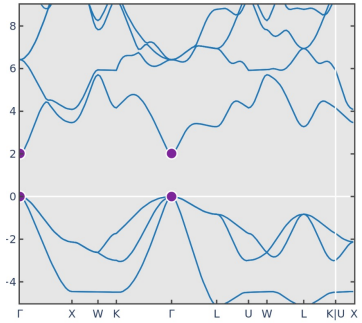
Five scientists win Nobel Prizes for contributions to resolving crystal structures

Computational Material Discovery



ZnS

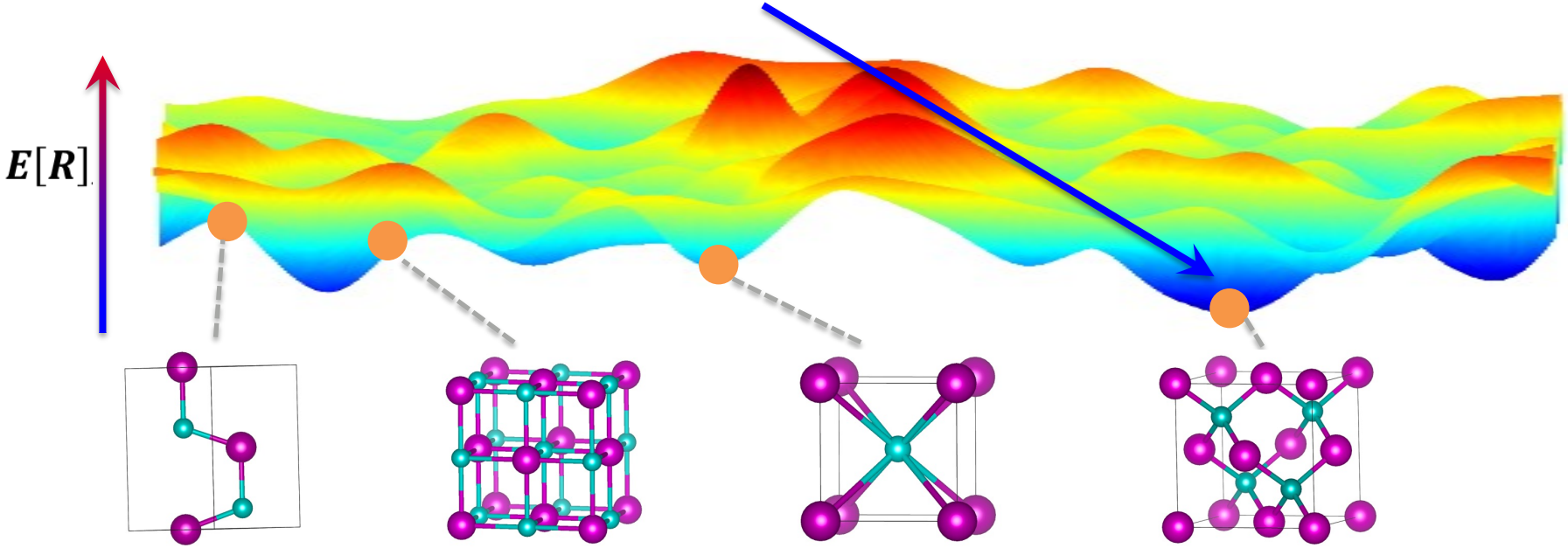




Composition **Structure** **Property** **Synthesis**

Born−Oppenheimer potential energy surface (PES)

Structure prediction target: finding the lowest energy point of the PES



$$E[R_0] = min\, E[R], R \in \mathbb{R}^{3N+3}$$ (N is atom number)

Scientific challenge: because of the **high dimension** and **multi-valley** properties of the PES, the determination of its lowest energy state is a typical NP-hard problem, which can not be solved by analytic methods.

CALYPSO crystal structure prediction

www.calypso.cn

CALYPSO Package, a swarm-intelligence based computational method that is able to predict crystal structures of materials at given information of chemical compositions.

Numerical Solution of Potential Energy Surface Based on Swarm Intelligence

I. Simplify

Physical constraints are used to simplify the potential energy surface.

II. Decompose

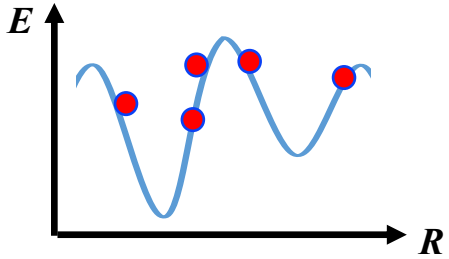
Develop quantitative characterization methods for structures, decompose potential energy surfaces.

III. Solve

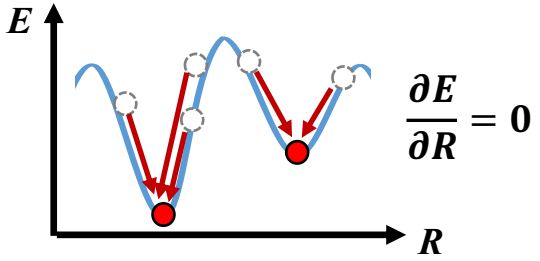
Heuristic swarm intelligence algorithm is introduced to solve the potential energy surface.

Key Techniques of CALYPSO Method

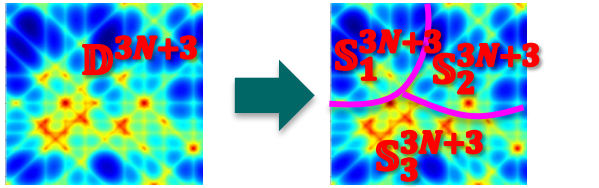
Generation → Symmetry Restriction; Atomic distance Restriction



Local Optimization → DFT/Empirical

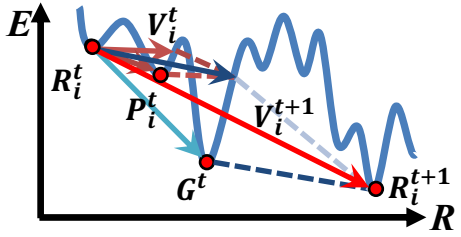


Characterization → BCM/CCF



Initial PES Decomposed PES

Evolution → PSO; SABC



CALYPSO is widely used by international peers

CALYPSO is used by more than 4,700 scholars in 77 countries under copyright agreements, allowing users to solve a range of scientific problems in multidisciplinary fields such as physics, chemistry, and materials.

- **>3,340 domestic users in >400 institutes**
- **>1,360 foreign users in >600 institutes**

6 Ivy League universities

- Cornell University
- Yale University
- ……

45 TOP50 universities

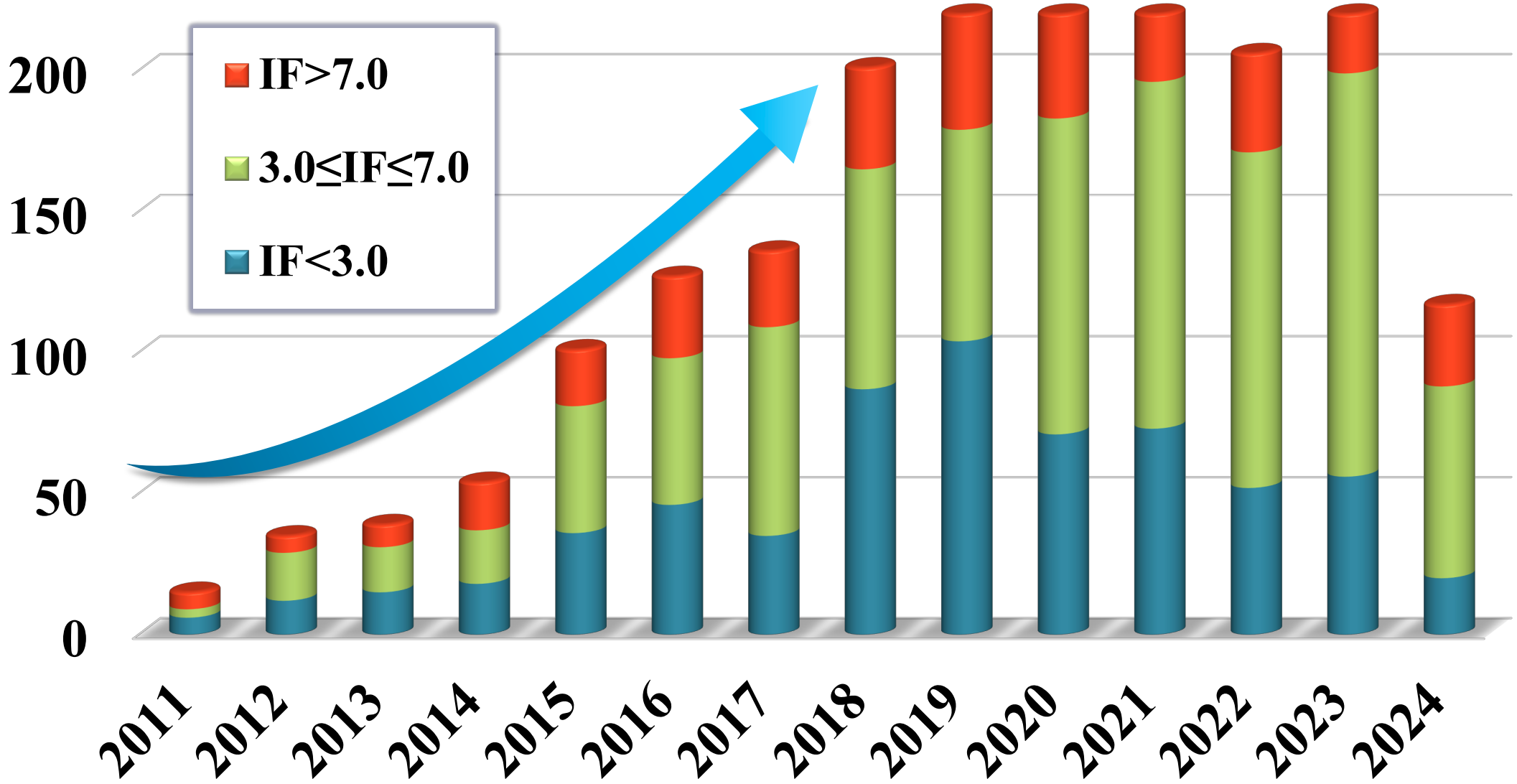
- MIT
- Stanford University
- ……

17 internationally research institutes

- National Laboratory for Renewable Energy
- Max Planck Institute
- ……

Users published 1,963 papers using CALYPSO in Nature sub-journals, PRL, etc.

IF<3.0: 631 papers; 3.0≤IF≤7.0: 944 papers; IF>7.0: 388 papers



CALYPSO won the second prize of National Natural Science Award in 2019

Project name： CALYPSO Crystal Structure Prediction Method and Its Application

Persons：Yanming Ma, Yanchao Wang, Jian Lv, Hanyu Liu, and Hui Wang

国家自然科学奖

证 书

为表彰国家自然科学奖获得者，特颁发此证书。

项目名称：CALYPSO 晶体结构预测方法与应用

奖励等级：二等

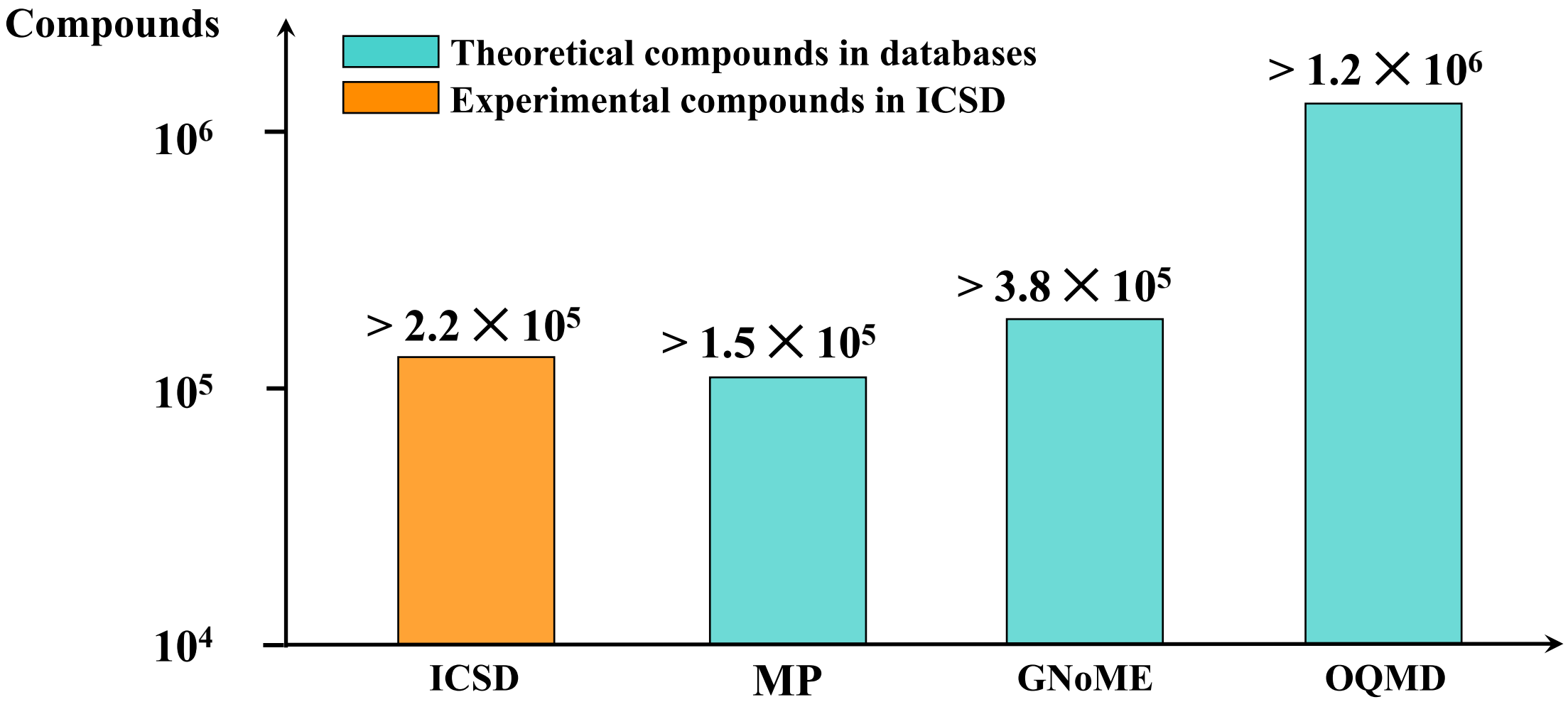
获 奖 者：马琰铭（吉林大学）

中华人民共和国国务院

2019 年 12 月 18 日

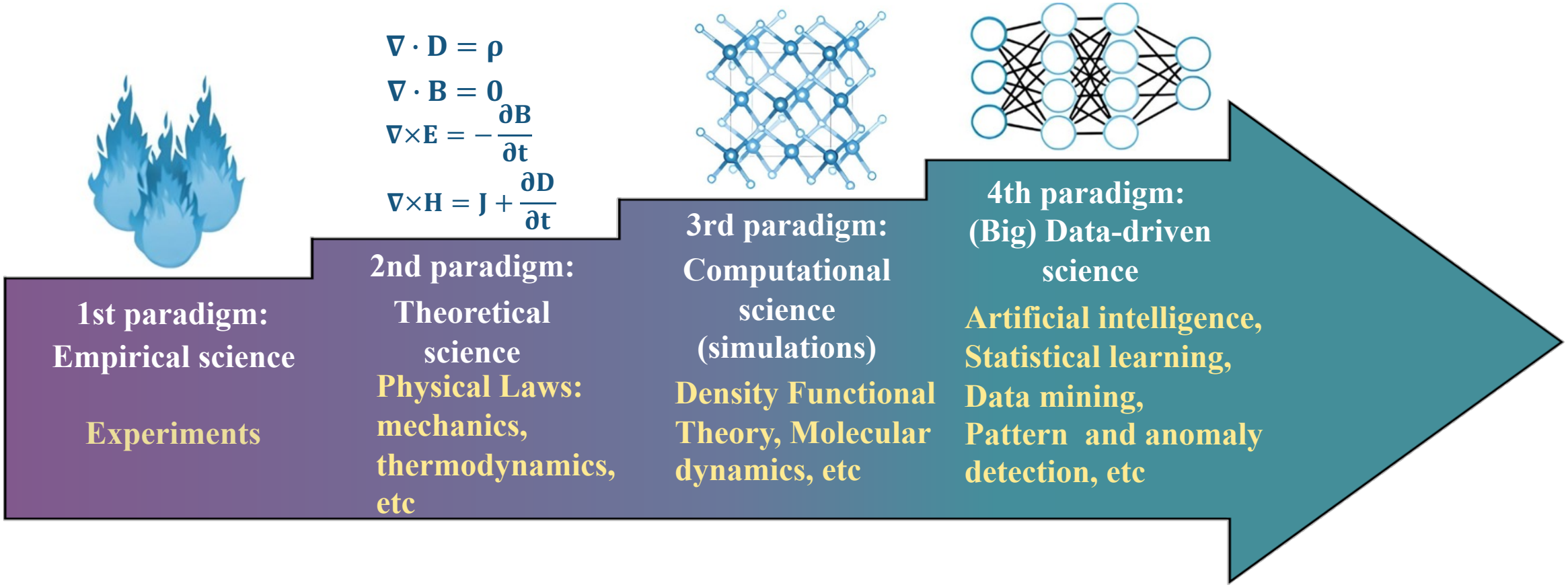
证书号：2019-Z-102-2-04-R01

Crystal Structure Database



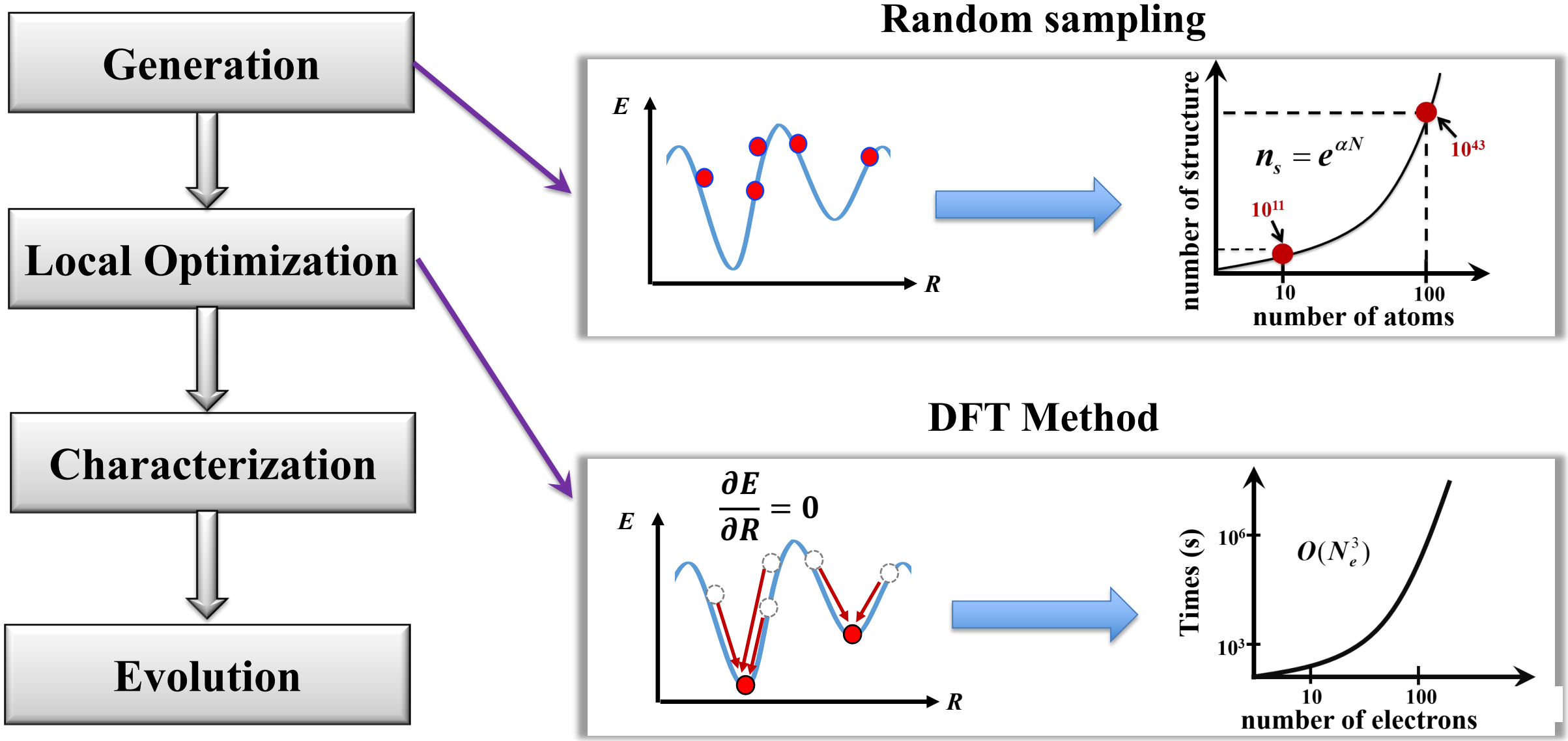
Data-driven science

Research methods centered on big data, which involves collecting, processing, analyzing, and mining data to reveal underlying patterns, drive scientific discoveries, and promote theoretical innovation.



Gabriel *et al.*, J. Phys. Mat. 2, 032001 (2019)

Machine Learning Accelerated Key Techniques



Data-driven method for material discovery

Database

- Structures
- Energies
- Forces
- Electronic properties

……

Machine Learning Potential

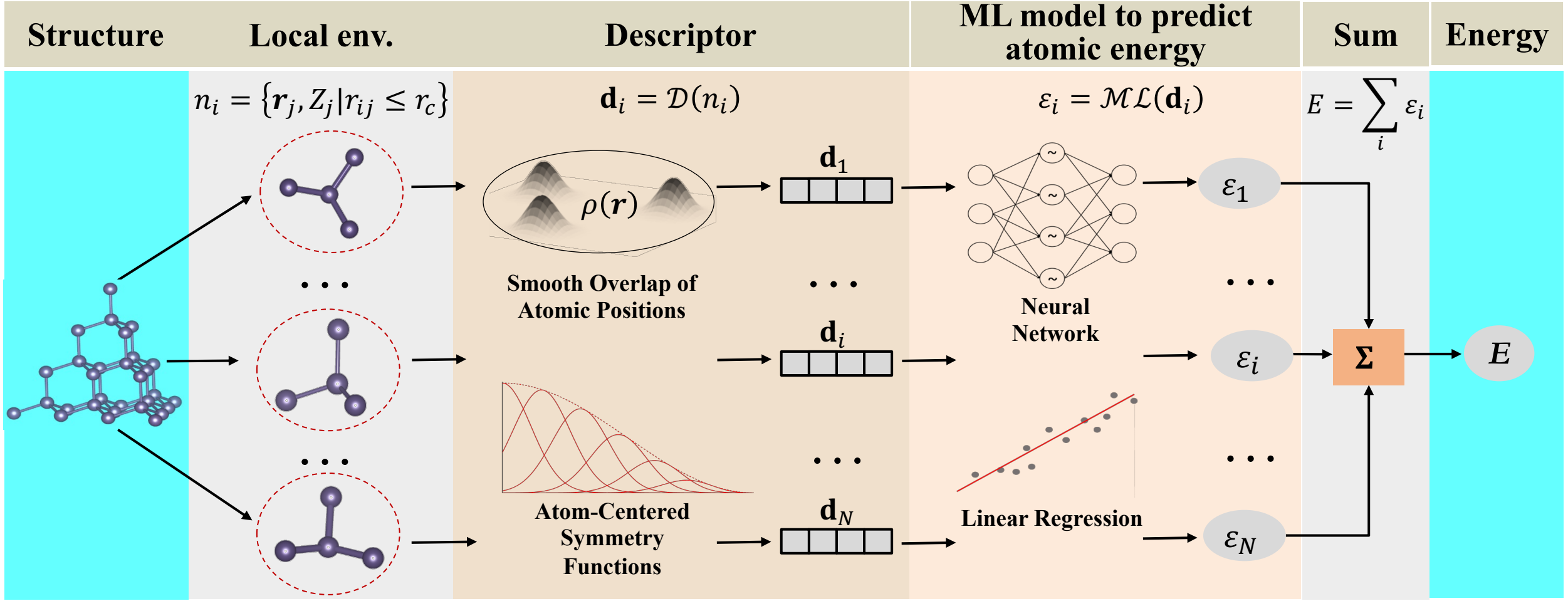
Reconstructing the PES to accelerate quantum mechanical calculations.

Generative model

Efficiently sampling the configuration space to propose stable and novel compounds.

Machine learning potentials (MLPs)

MLPs learn from ab initio data to reconstruct PES, enabling efficient and accurate prediction of material energies and atomic forces for large-scale, high-throughput simulations.



Machine learning potentials (MLPs)

A variety of MLPs have been developed, evolving from descriptor-based approaches to graph neural network architectures that eliminate the need for handcrafted descriptors.

Descriptor-based approaches

- **High-Dimensional Neural Network**
  Behler, Parrinello, PRL 98, 146401(2007)
- **Gaussion Approximation Potential**
  Bartók et al., PRL104, 136403 (2010)
- **Spectral Neighbor Analysis Potential**
  Thompson et al., JCP. 285, 316 (2015)
- **Moment Tensor Potential**
  Shapeev, Multiscale Model. Simul. 14, 1153 (2016)
- **Deep Potential**
  Zhang et al., PRL 120, 143001 (2018)

……

Graph neural network-based approaches

- **SchNet**
  Schütt et al., JCP 148, 241722 (2018)
- **PhysNet**
  Unke and Meuwly, J. Chem. Theory Comput. (2019)
- **DimeNet**
  Gasteiger et al., NeurIPS (2021)
- **NequIP**
  Simon et al., Nat. Commun. 13, 2453 (2022)
- **NewtonNet**
  Haghighatlari et al., Digit. Discov. 1, 333 (2022)

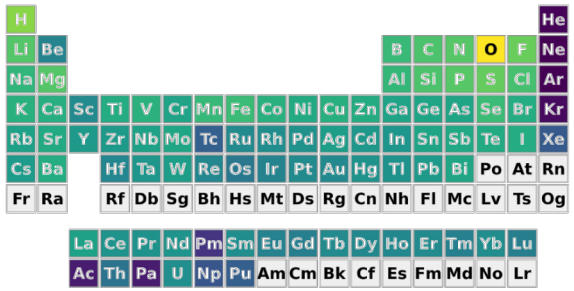
……

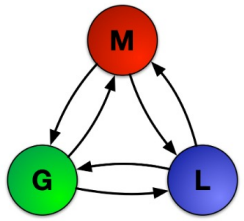
Universal interatomic potentials (UIPs)

UIPs aim to be broadly applicable, providing accurate predictions of energies, atomic forces across the periodic table.

Ab initio Database

Material Project JARVIS-DFT ...

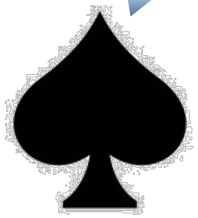




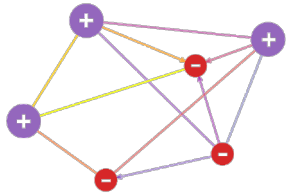
M3GNet (2022)
Materials Virtual Lab

MatterSim (2024)
Microsoft AI4S

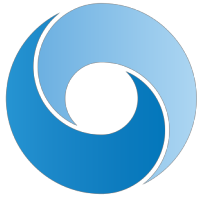
DPA2 (2023)
DP Technology



MACE (2022)
MACE Group

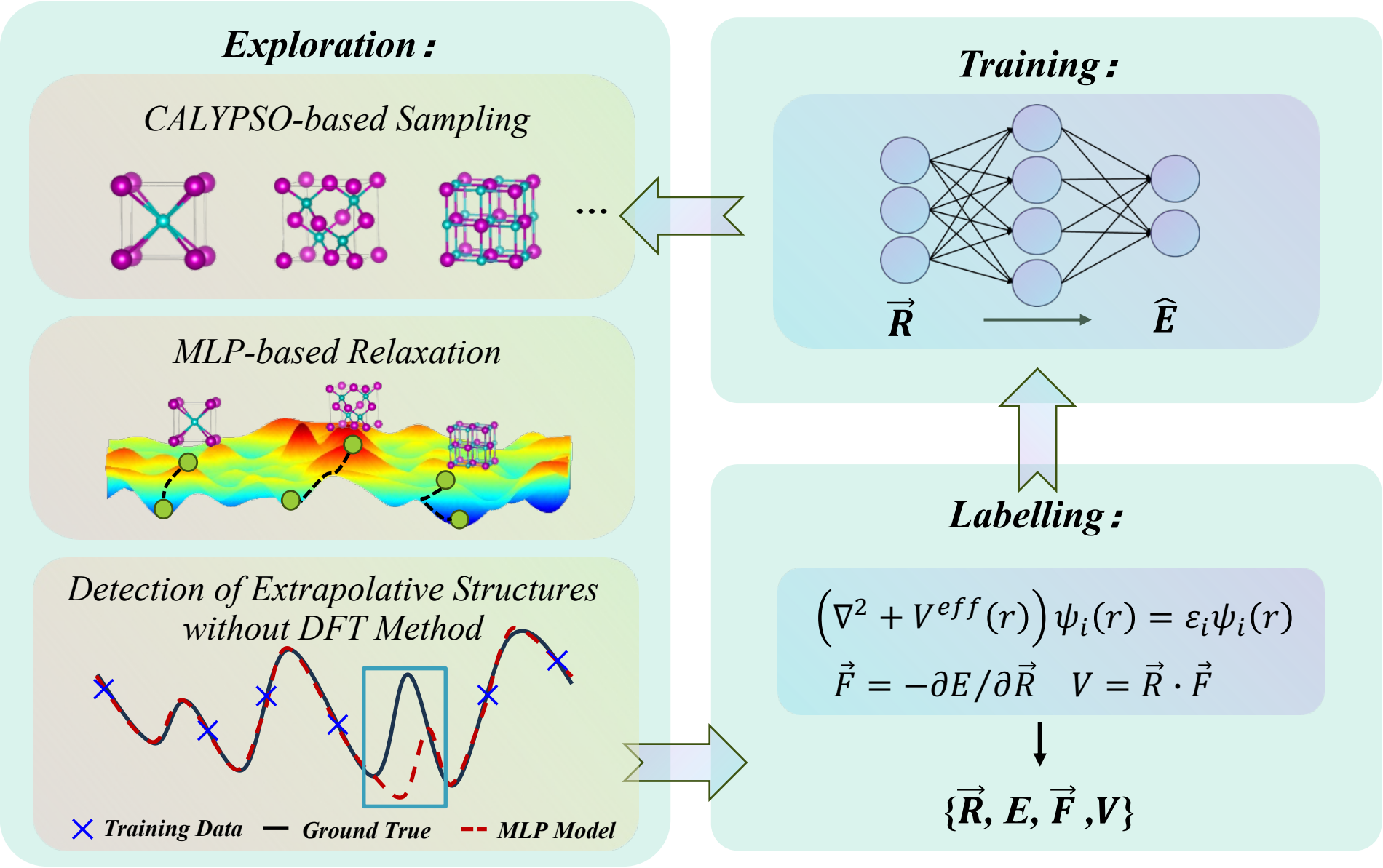


CHGNet (2023)
Ceder Group



GNoME (2023)
NequIP architecture
Google DeepMind

MLP Accelerated CALYPSO Structure Prediction Scheme



Accelerated CALYPSO method by Gaussian Approximation Potential (GAP)

Atomic environment descriptor

- **Radial Symmetry Function Responds to Atomic Distance Information**

$$W_i^{\mathrm{rad}} = \sum_{j\neq i}^{N} g(Z_j) e^{-\eta(r_{ij}-\mu)^2} f_{ij}$$

- **Angular Symmetry Function Responds to Bonding Direction Information**

$$W_i^{\mathrm{ang}} = 2^{1-\zeta} \sum_{j\neq i}^{N} \sum_{k\neq i,j}^{N} h(Z_j, Z_k) \left(1 + \lambda \cos\theta_{ijk}\right)^{\zeta} \times e^{-\eta(r_{ij}-\mu)^2} e^{-\eta(r_{ik}-\mu)^2} e^{-\eta(r_{jk}-\mu)^2} f_{ij} f_{ik} f_{jk}$$

J. Behler et al., JCP 134, 074106 (2011)

M. Gastegger et al., JCP 148, 241709 (2018)

Regression model

- **Bayes Rule:**

$$P(t_{N+1}|\mathbf{t}) = \frac{P(\mathbf{t}|t_{N+1})P(t_{N+1})}{P(\mathbf{t})}$$

- **Atomic Energy :**

$$\varepsilon_* = \boldsymbol{k}_*^T \boldsymbol{Q}_{MM}^{-1} \boldsymbol{C}_{MN} \boldsymbol{L}^T (\Lambda + \boldsymbol{\sigma}^2 \boldsymbol{I})^{-1} \boldsymbol{E}$$

- **Variance of the predicted atomic energy:**

$$\boldsymbol{\sigma}_{\hat{t}}^2 = \boldsymbol{C}(\mathrm{x}_*, \mathrm{x}_*) - \boldsymbol{k}_*^T (\boldsymbol{C}_{MM}^{-1} - \boldsymbol{Q}_{MM}^{-1}) \boldsymbol{k}_* + \boldsymbol{\sigma}^2$$

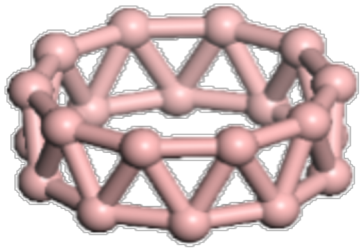
- **Variance of the predicted total energy:**

$$\Sigma_u = \frac{1}{N} \sum_{i\in u} \sigma_i(\boldsymbol{d}_i)$$

Bartok et al., PRL 104, 136403 (2010)

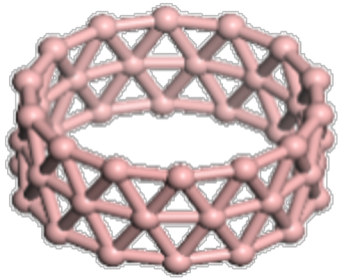
Application 1 : the prediction of Boron clusters with GAP

Boron clusters are rich in bonding environments and are ideal test systems for building machine learning potentials and structure predictions.



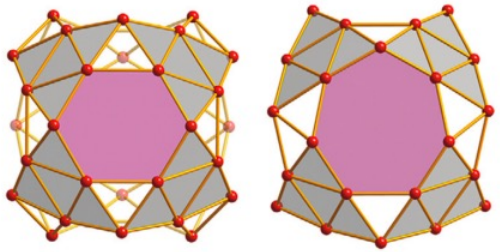
Double-Ring Tube

PNAS 102, 961 (2005)



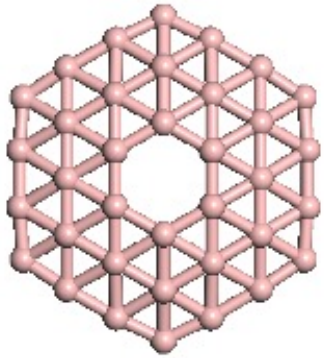
Three-Ring Tube

JCP 129, 024903 (2008)



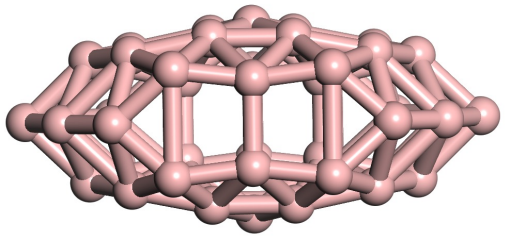
Fullerene-like structure

Nat. Chem. 6, 727 (2014)



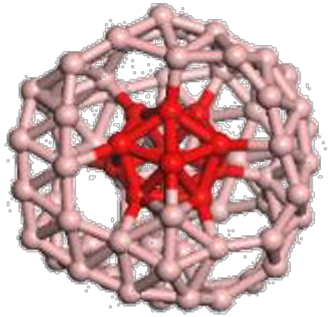
Planar structure

Nat. Commun. 5, 3113 (2014)



Bilayer structure

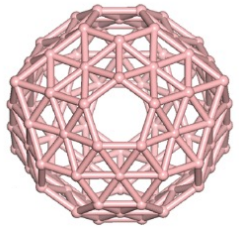
Nanoscale 9, 13905 (2017)



Core-shell structure

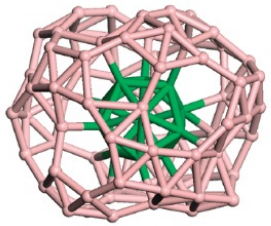
JPCA 114, 9969 (2010)

What's the structure of B_{84}?



Cage structure

PRL 100, 165504 (2008)



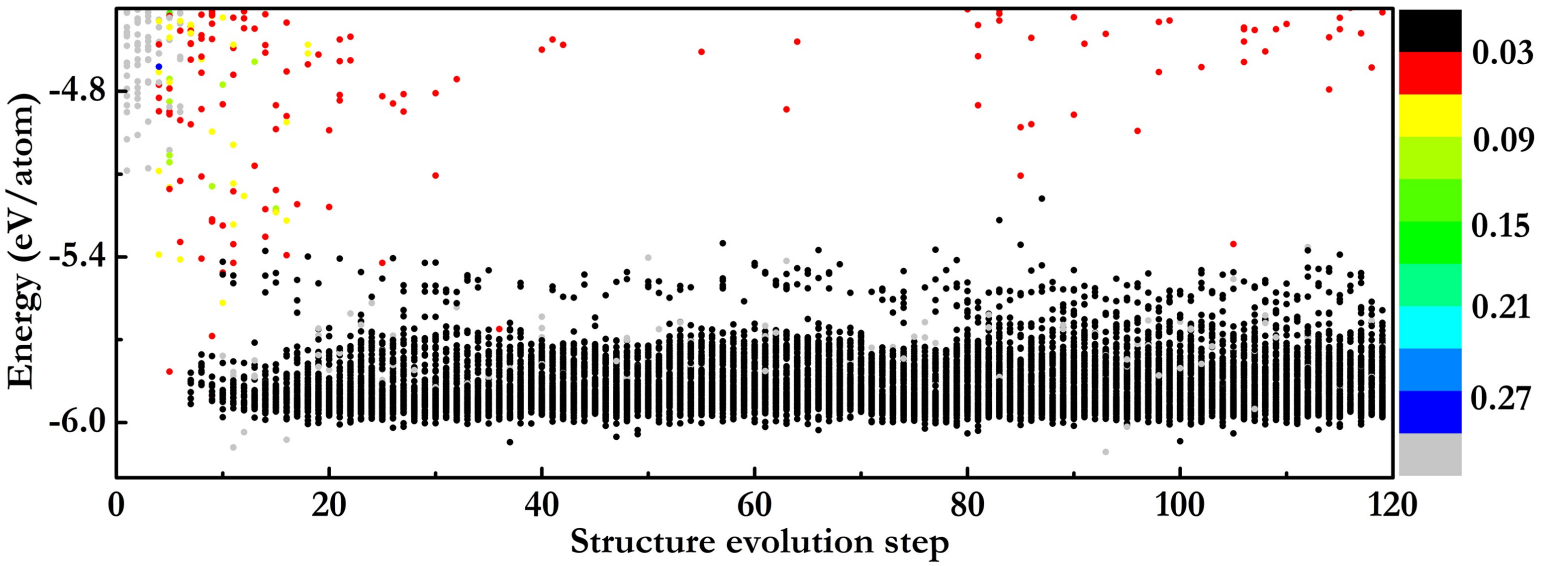
Core-shell structure

JPCA 2245 (2010)

Quasi-planar structure

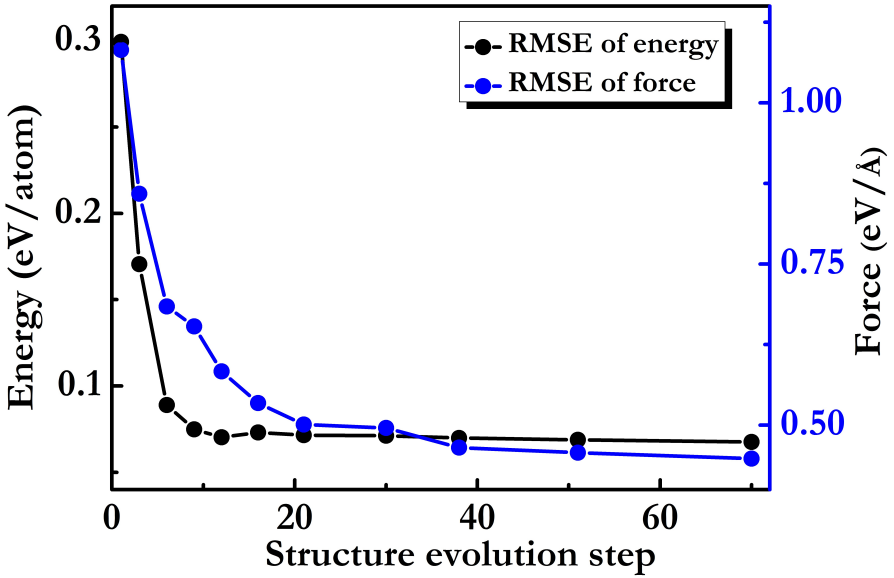
Nanoscale 7, 5055 (2015)

Accelerated structure prediction of B_{84} with GAP



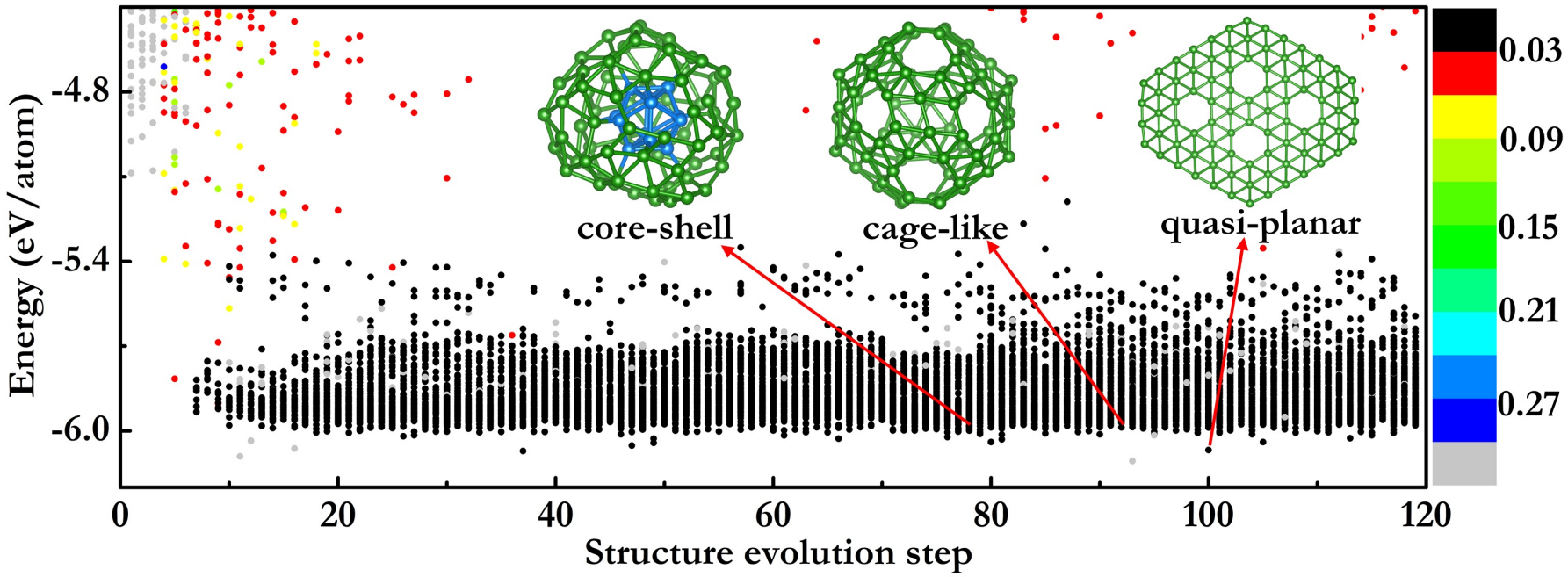
Energy evolution in structure prediction

The color of data points represents the variance of predicted total energy.



Evolution of RMS error of energy and force

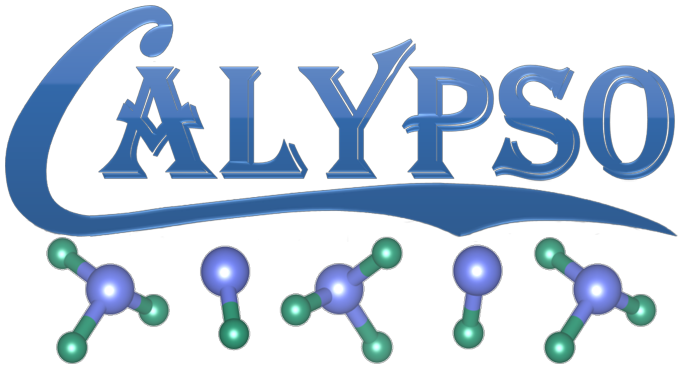
The prediction results of B_{84} cluster



- Proposed a core-shell structure of B_{84} cluster with the lowest energy so far
- The computational cost is substantially reduced by 1–2 orders of magnitude if compared with full DFT-based structure searches

Tong, Lv*, Wang and Ma* et al., Fara. Discuss. 211, 31 (2018)

Interfacing CALYPSO and other machine learning potentials



DP/DPA/DPA2 — DP Technology

CHGNET — Ceder Group

M3GNET — Materials Virtual Lab

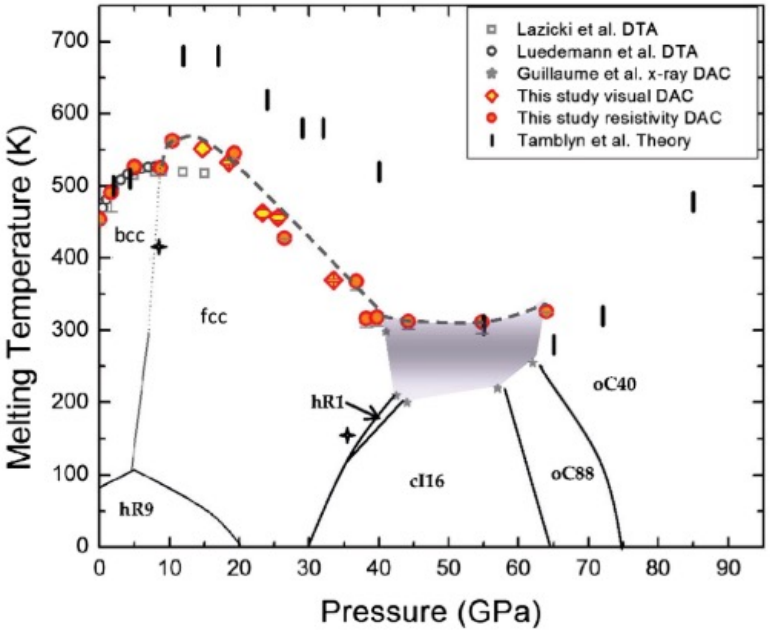
MACE — MACE Group

……

Application 2 : the prediction of Li high-pressure phases

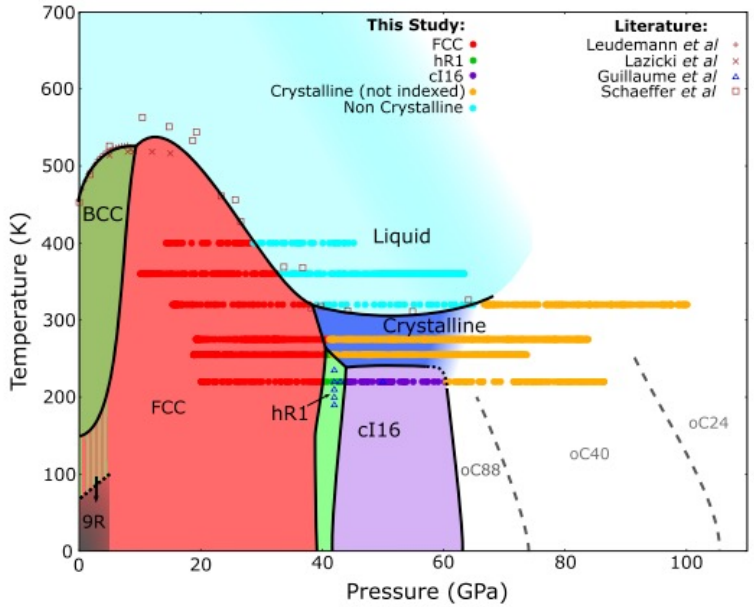
It is found that a new phase may exist in the range of 40-60 GPa and 200-300 K, and the structure is difficult to be determined for a long time.

- **Phase diagram of lithium (resistivity measurement)**



Anne et al., PRL 109, 185702 (2012)

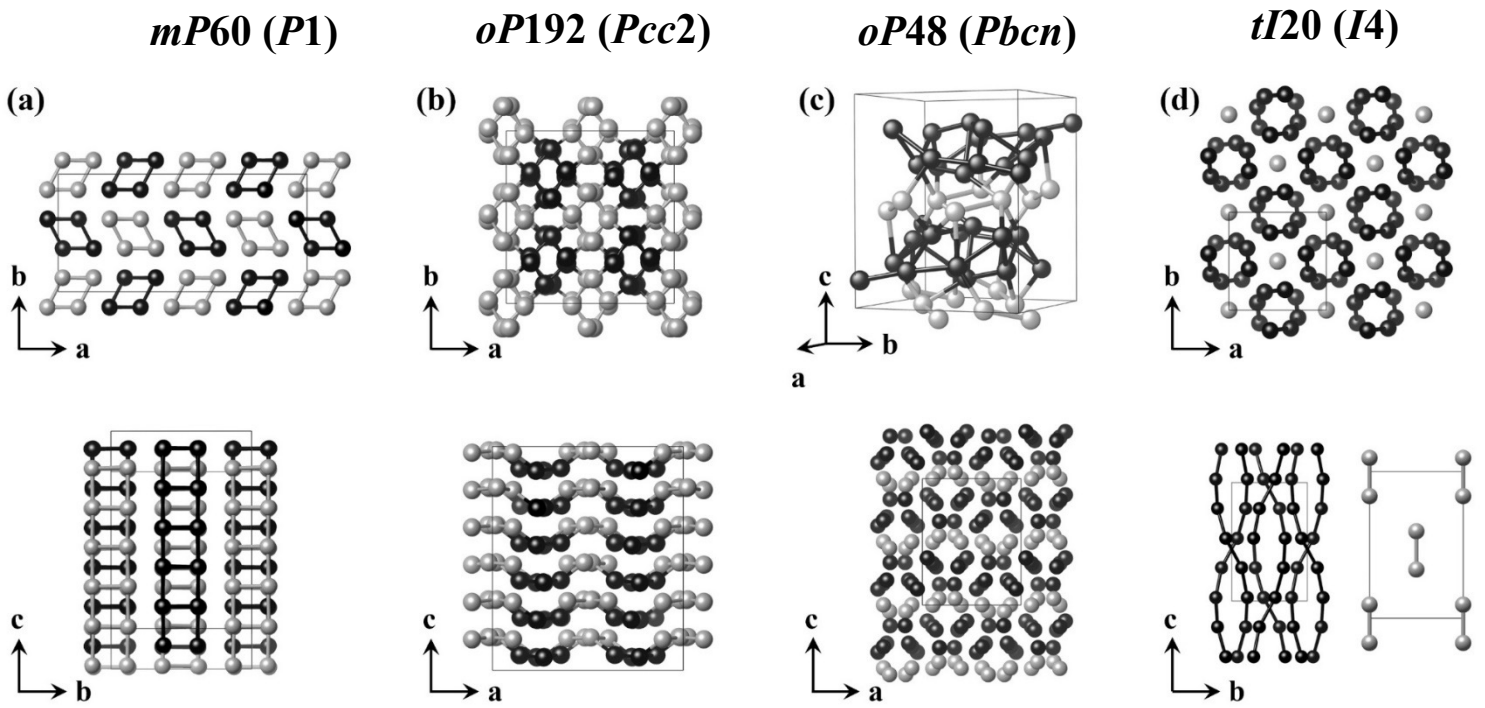
- **Phase diagram of lithium (XRD measurement)**

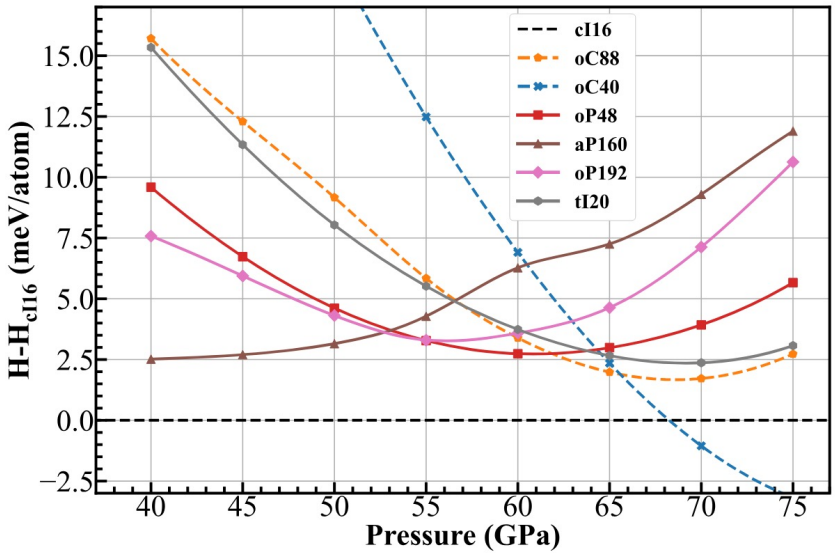


Mungo et al., PRL 123, 065701 (2019)

Several Li high-pressure phases are discovered

- **Traning set：** experimental structures of Li and perturbed configurations
- **Structure search：** CALYPSO+Deep Potential，1-200 atoms/cell，600,000 structures
- **DFT optimization：** VASP，5000 structures



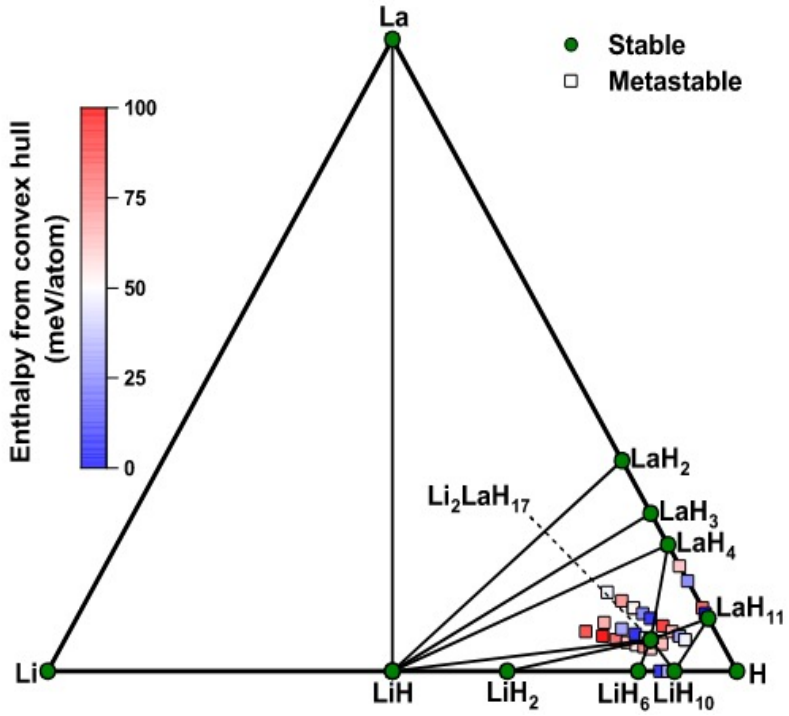


Enthalpy curve of elementary lithium

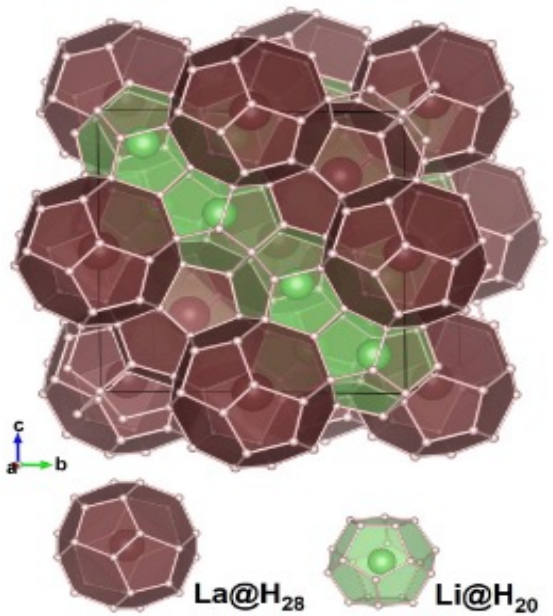
Wang *et al.*, Nat. Commun. 14, 2924 (2023)

Application 3 : the prediction of hydrogen-rich Li-La-H compounds

The Li_2LaH_{17} compound has been predicted at high pressure using the DFT method in our previous work.



Phase diagram of Li-La-H system at 300 GPa

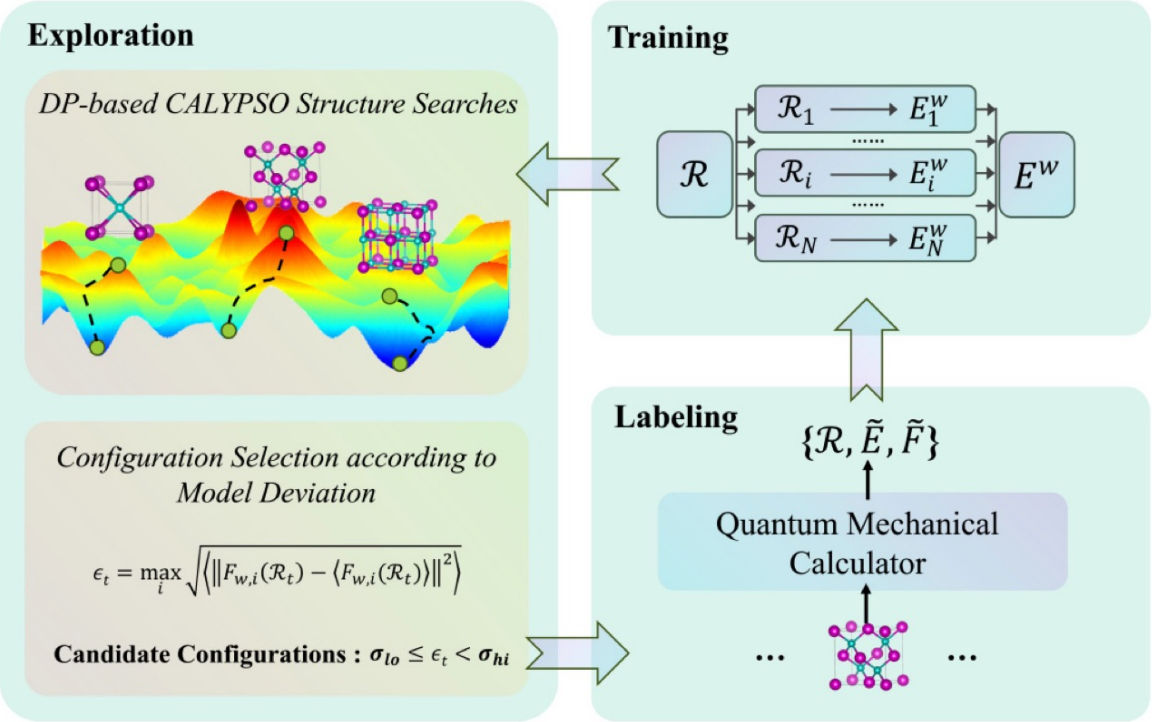


The crystal structure of Li_2LaH_{17} (T_c = ~150K @ 160GPa)

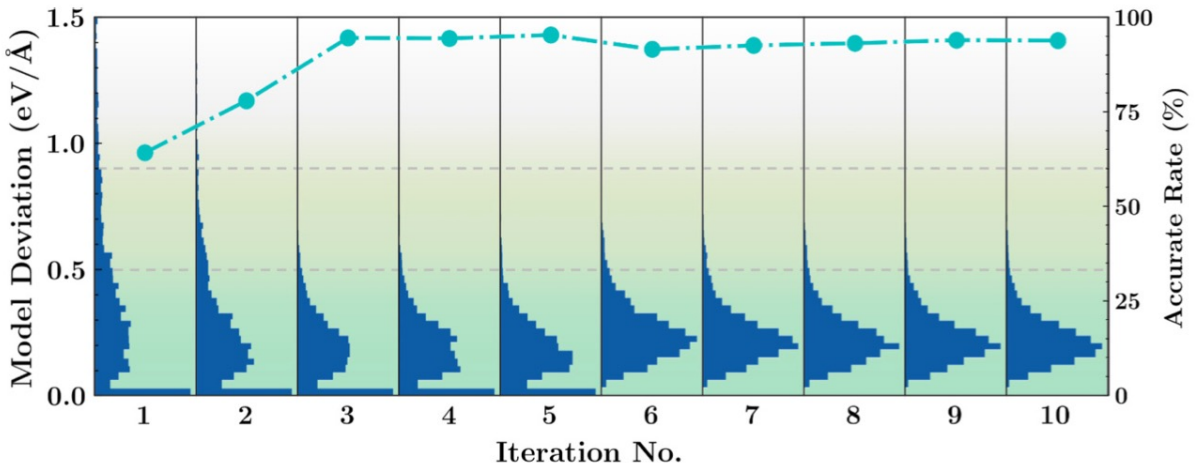
Sun *et al.*, Phys. Rev. B 106, 024519 (2022)

A concurrent learning scheme

For Li-La-H system, the initial ensemble of DP models was constructed based on 2036 random structures, and 10 iterations of CALYPSO structure searches (2100 structures for Li-La-H compounds) were carried out to construct Deep Potential.



Workflow of a concurrent learning scheme

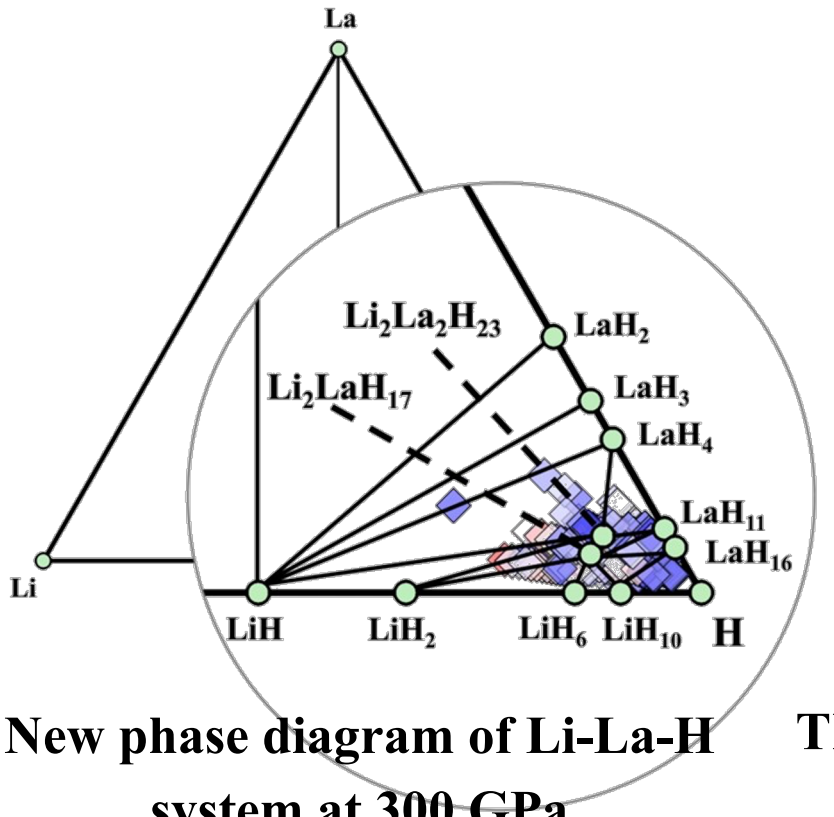


The model deviation distribution during the iterations

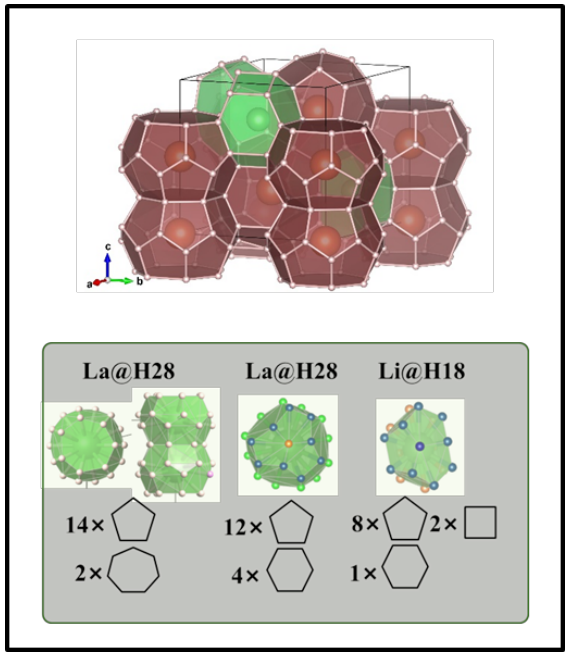
Wang *et al.*, Phys. Rev. B 109, 094117 (2024)

A new compound $Li_4La_4H_{46}$ is discovered

A new stable structure $Li_2La_2H_{23}$ was discovered after searching 300,000 configurations with CALYPSO and Deep Potential.



New phase diagram of Li-La-H system at 300 GPa



The crystal structure of $Li_2La_2H_{23}$

(Tc = ~130K @ 300GPa)

Structure search with DFT
DFT number: 140,000

VS

Structure search with MLP
DFT number: ~1549

Wang *et al.*, Phys. Rev. B 109, 094117 (2024)

Data-driven method for material discovery

Database

- Structures
- Energies
- Forces
- Electronic properties

……

Machine Learning Potential

Reconstructing the PES to accelerate quantum mechanical calculations.

Generative model

Efficiently sampling the configuration space to propose stable and novel compounds.

Generative model

Generative models are optimized by minimizing the difference between the generated distribution and true data distribution through techniques like maximum likelihood estimation or adversarial training.

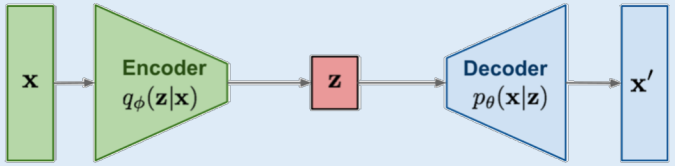
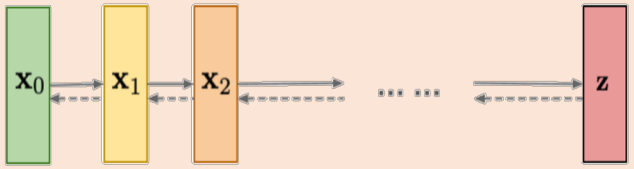
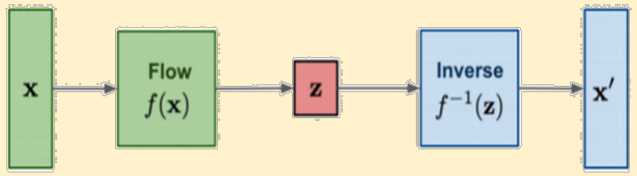
- **Real data distribution:** $p(x)$, **which is unknown and needs to be learned from the dataset.**
- **Estimated distribution:** $q_\theta(x)$, **which is described by a neural network with a set of parameters** θ.

Generative modeling frames learning as an optimization problem where the loss corresponds to the KL divergence between $q_\theta(x)$ **and** $p(x)$ **:**

$$\mathcal{L}(\theta) = D_{KL}[p(x) \parallel q_\theta(x)] = \mathbb{E}_{x\sim p}\left[\log \frac{p(x)}{q_\theta(x)}\right]$$

Generative model for crystal structures

Crystal structure prediction performance of various generative model using different architectures on Materials Project dataset

Architectures	Methods	CSP performances on MP20	
		Match Rate (%)	RMSE
VAE 	iMatGen Noh, *et al.*, Matter, 1, 1370 (2019)	/	/
	CDVAE Xie, *et al.*, arXiv 2110.06197 (2022)	33.90	0.1045
	Cond-CDVAE Luo, *et al.*, arXiv 2403.10846 (2024)	/	/
Diffusion model 	DiffCSP Jiao, *et al.*, arXiv 2309.04475 (2023)	51.49	0.0631
	MatterGen Zeni, *et al.*, arXiv 2312.03687 (2023)	/	/
	UniMat Yang, *et al.*, arXiv 2311.0923 (2023)	/	/
Flow model 	GM4CSP Luo, *et al.*, in progress	58.43	0.1425
	FlowMM Miller, *et al.*, arXiv 2406.04713 (2024)	61.39	0.0566

The CALYPSO high-pressure structure database

A database for high-pressure structures collected from the CALYPSO community.

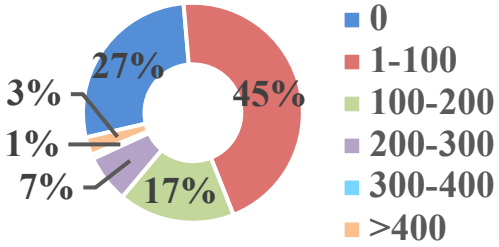
670,979 structures

86 elements

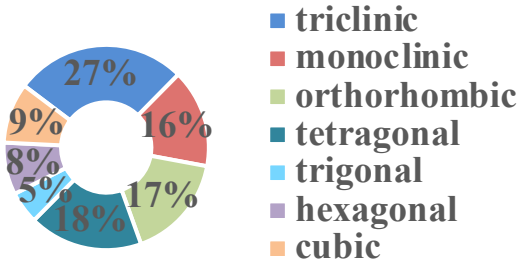
85,824 compositions

114,733 prototypes

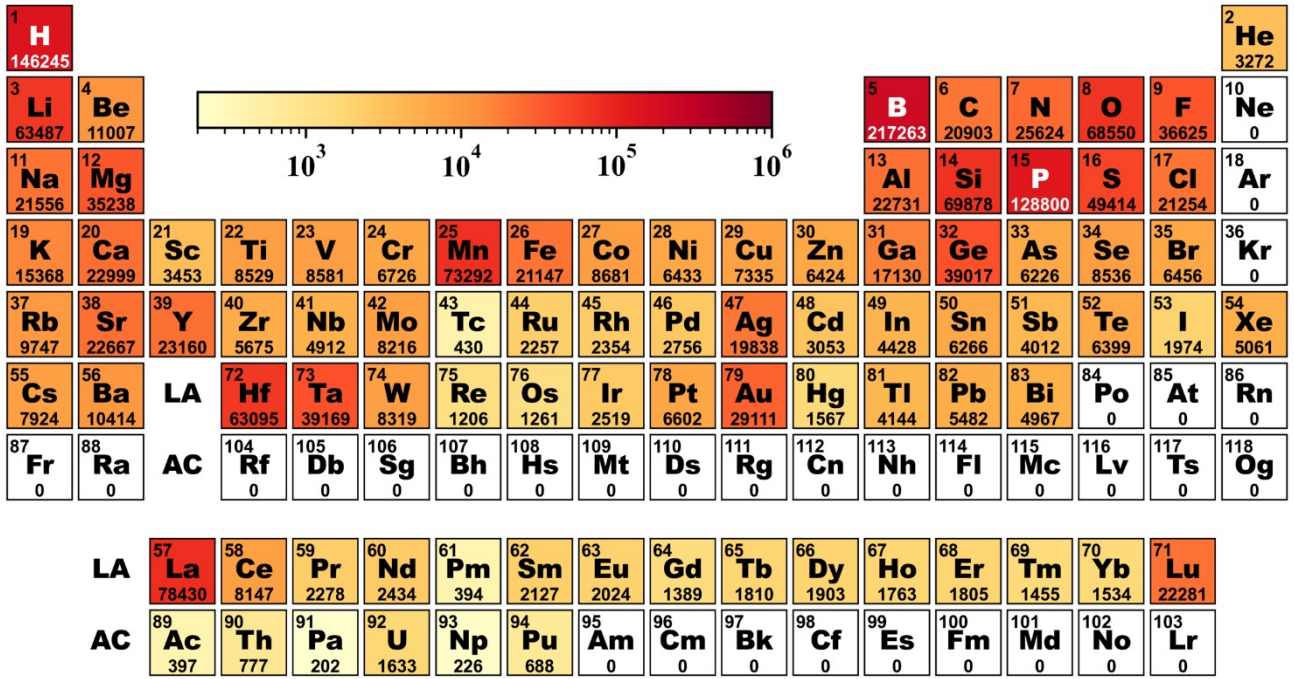
Pressure (GPa)



Crystal system



Structure counts of each element

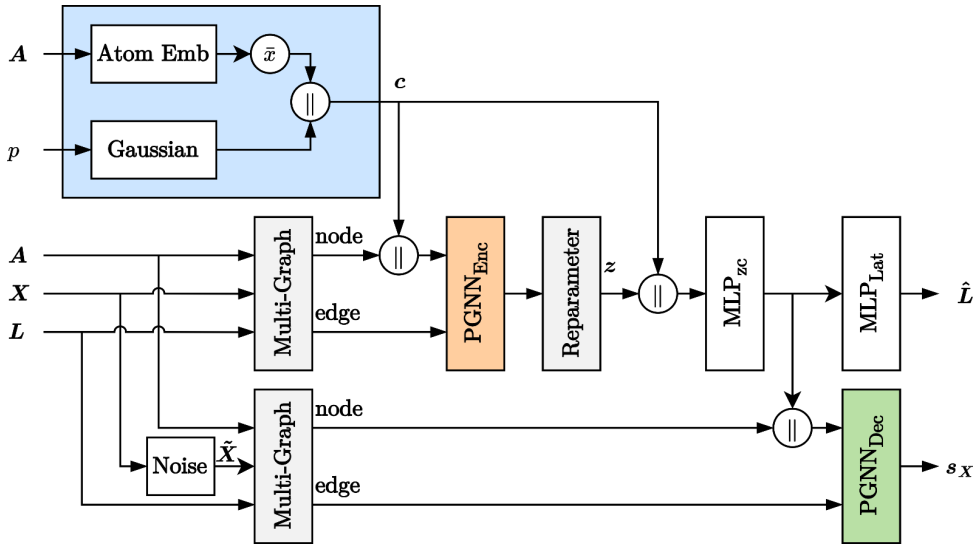


Luo *et al.*, arXiv 2403.10846 (2024)

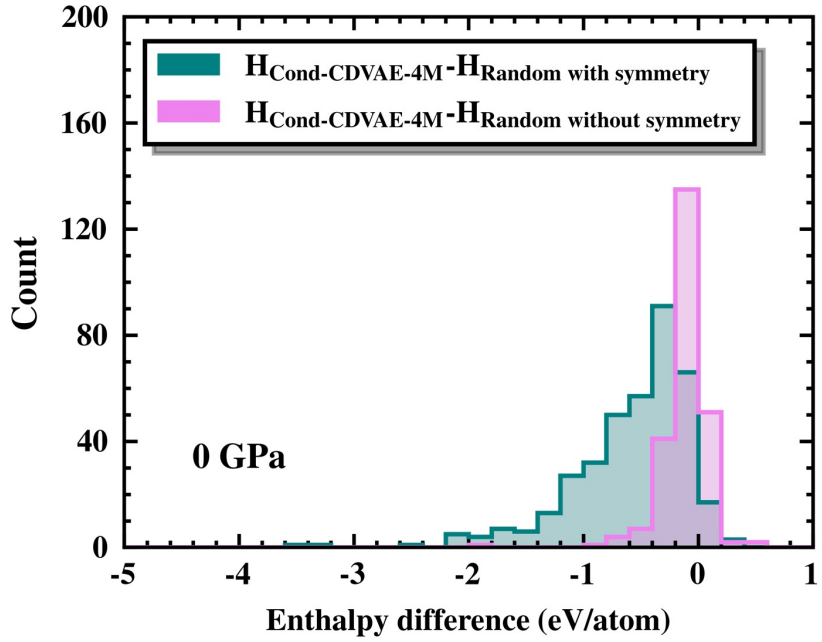
Generative model for dense matters

We developed a universal generative model for crystal structure prediction through a conditional crystal diffusion variational autoencoder approach, tailored to accommodate user-defined material and physical parameters such as composition and pressure.

Cond-CDVAE architecture



final energy difference of 500 structures at 0GPa

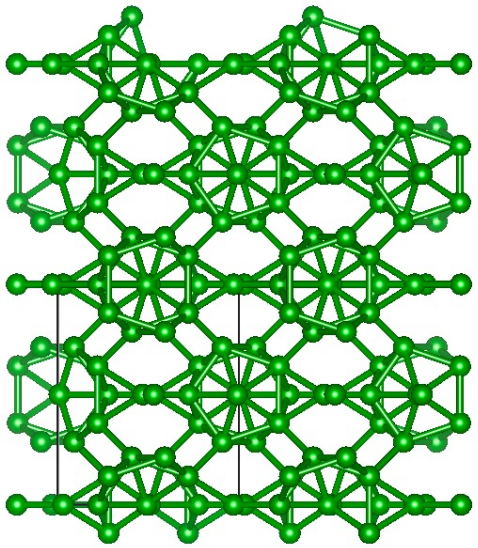


Luo *et al.*, arXiv 2403.10846 (2024)

The performance of generative model for structure search

Search high-pressure phases of Li, B and SiO_2 by generative model and CSP with generating 1000 samples for each run.

Structures	N_{atoms}	P (GPa)	N_{model}	Runs	N_{CSP}	Runs
Li						
$cI16$	16	50	566.0	2/5	**50.0**	**3/3**
B						
α-B_{12}	36	0	**74.0**	1/5	392.3	**3/3**
γ-B_{28}	28	50	**341.0**	**5/5**	-	0/3
α-Ga-type	8	100	**58.0**	**5/5**	78.0	3/3
SiO_2						
α-quartz	9	0	**62.8**	**5/5**	189.0	3/3
coesite	24	5	**328.0**	**3/5**	-	0/3
rutile-type	6	50	**5.0**	**5/5**	90.0	3/3
$CaCl_2$-type	6	80	**10.2**	**5/5**	61.0	3/3
α-PbO_2-type	12	100	**21.6**	**5/5**	74.7	3/3
pyrite-type	12	300	**28.8**	**5/5**	66.0	3/3

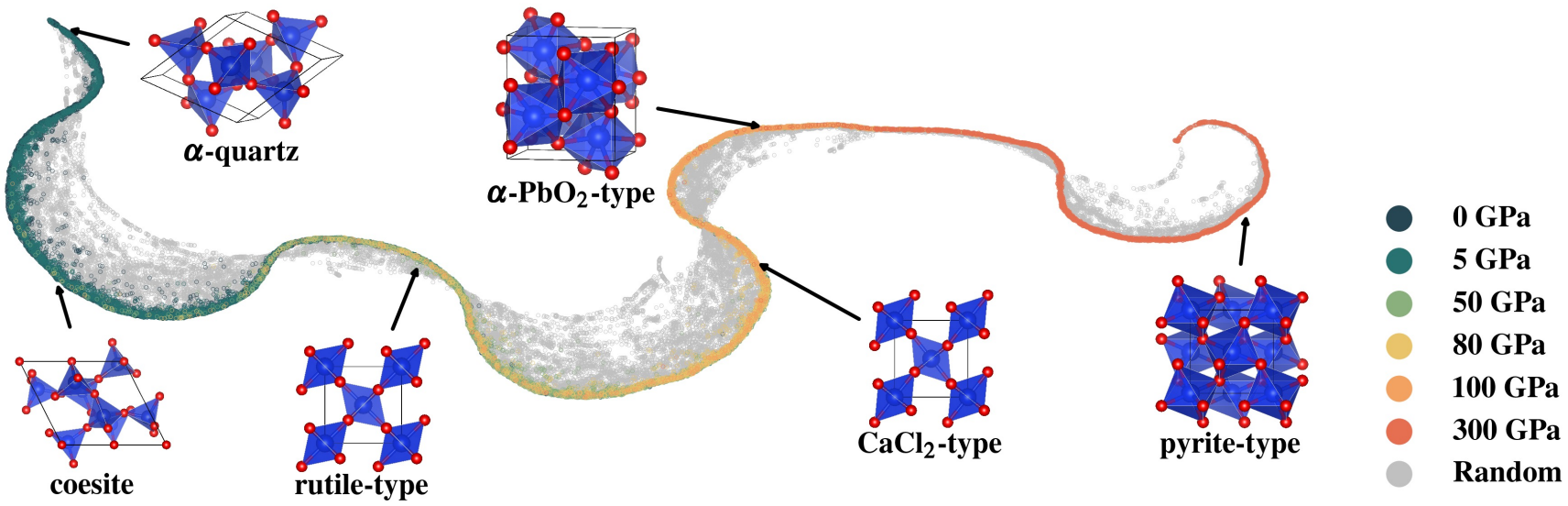


the crystal structure of γ-B_{28}

N_{model} and N_{CSP} denote the average number of structural samplings required to locate the global minimum

The manifold learning of silica

Two-dimensional projection of silica structures generated by generated model and random sampling.



Si (4CN), O (2CN) **Si (6CN), O (3CN)** **Si (6CN), O (4CN)**

CN：coordination number

Luo *et al.*, arXiv 2403.10846 (2024)

Summary

- **CALYPSO is able to predict crystal structures of materials at given information of chemical compositions.**

- **We have developed machine learning potential accelerated CALYPSO structure prediction method and predicted the structures of B_{84}, Li and $Li_2La_2H_{23}$.**

- **We built a database for high-pressure structures and developed a universal generative model for generating crystal structures.**

Acknowledgements

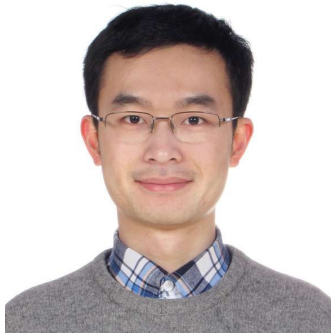
Prof. Yanming Ma
Jilin University

Prof. Yanchao Wang
Jilin University

Prof. Jian Lv
Jilin University

- **Prof. Changfeng Chen** Nevada University
- **Prof. Quan Li** Jilin University
- **Prof. Hanyu Liu** Jilin University
- **Prof. Yu Xie** Jilin University
- **Dr. Qunchao Tong**

National University of Defense Technology



Han Wang
Researcher
IAPCM

Xiaoyang Wang
Assistant Researcher
IAPCM

Xiaoshan Luo
Doctoral candidate
Jilin University

Zhenyu Wang
Doctoral candidate
Jilin University

Thanks very much for you attention!

CALYPSO

www.calypso.cn

2006 — **Idea**

2010 — **First paper and Copyright**

2013 — **First workshop**

2015 — **1000 users**

2016 — **Walter Kohn Prize**

2018 — **The First Prize of the Natural Science Award from the Ministry of Education**

2019 — **The second class prize of the state natural science award**

2024 — **1900 papers**