

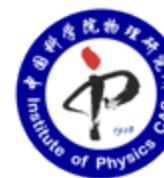
# GPTFF: A high-accuracy out-of-the-box universal AI force field for arbitrary inorganic materials



Miao Liu, Sheng Meng

中国科学院物理研究所

松山湖材料实验室

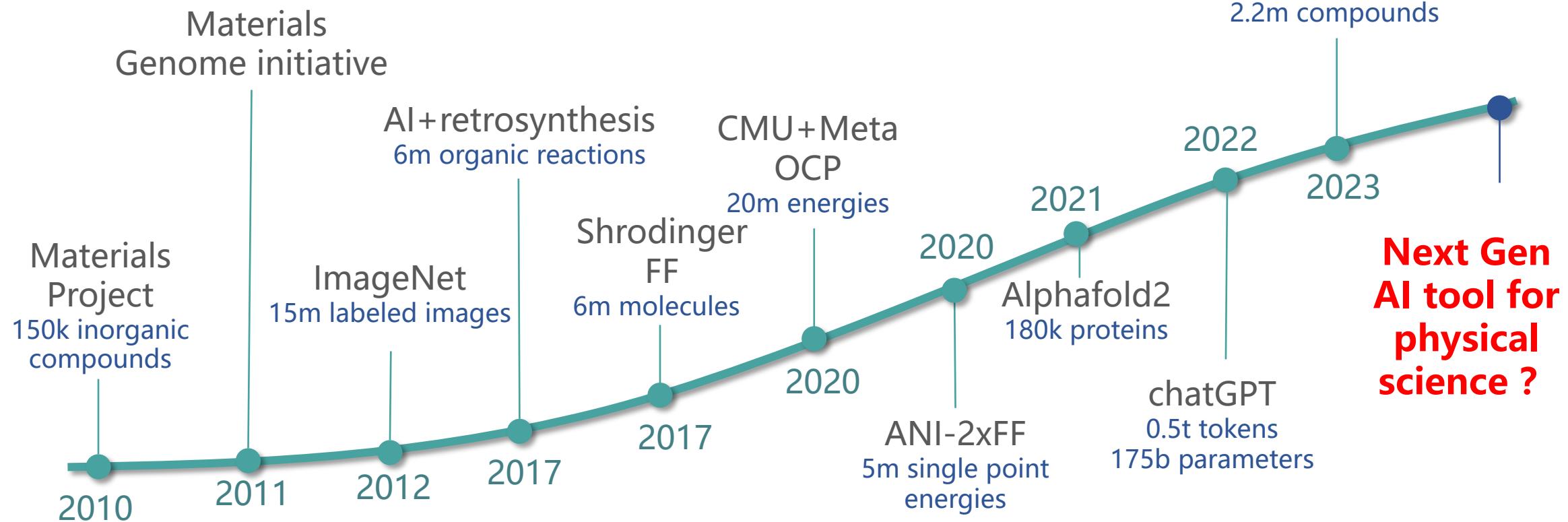


北京 2024/11/12

# 1 数据科学带来的启示?



## Data changes everything!



# 1 计算材料科学的技术路线

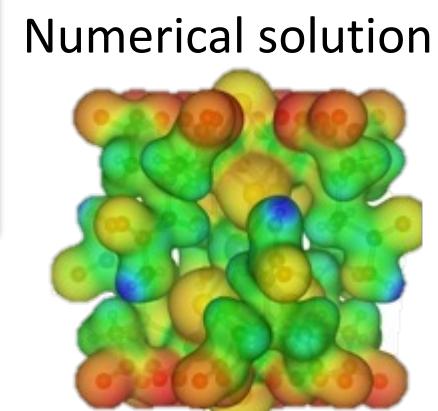


$$i\hbar \frac{\partial}{\partial t} \Psi = \hat{H} \Psi \longrightarrow \text{Applications}$$

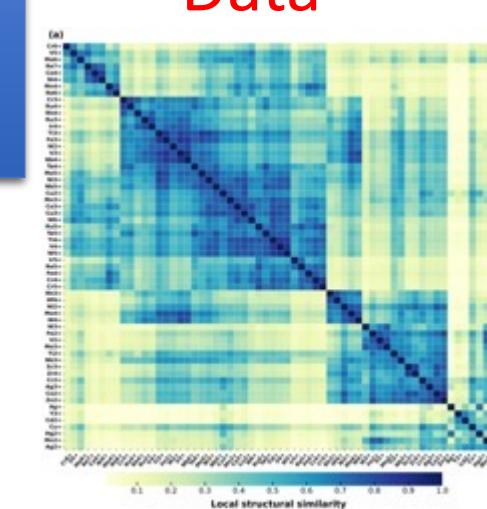


DFT Theory  
 $\hat{H} |\Psi\rangle = E_{gs} |\Psi\rangle$

(1965) 2th paradigm

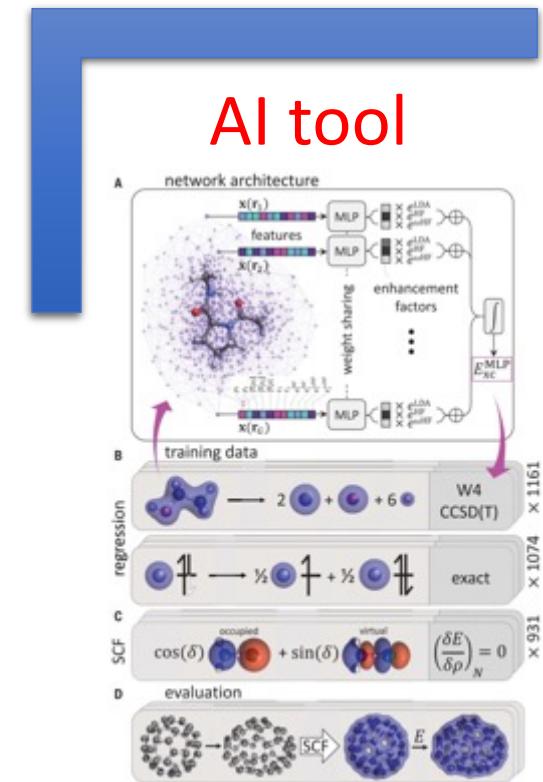


Numerical solution

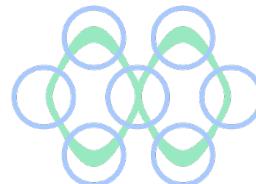


Data

Electronic-Structure Theory & Materials Genomics



# 2 世界级材料数据库（计算）



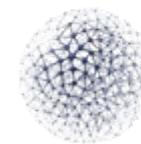
The Materials Project

美国 劳伦斯国家实验室 (2010)

12万条高质量材料数据



美国 西北大学 (2012)



AFLOW  
Automatic-FLOW for Materials Discovery

美国 Duke大学 (2012)



美国 国家标准局 (2016)



德国 Fritz-Haber-Institut (2015)



MATERIALS  
CLOUD

瑞士 NCCR (2015)

## GNoME

美国 google(2023)

220万条材料数据  
(公开38万个)



## Atomly材料科学数据库

<https://atomly.net>



Atomly Modernize the Materials Science

Atomly 材料科学数据库

用高通量第一性原理，推动材料科学演进  
自主知识产权 · 世界级材料数据库 · 人工智能

Atomly Modernize the Materials Science

by Elements e.g. Ti-O Search

1 H 2 He  
3 Li 4 Be  
11 Na 12 Mg  
19 K 20 Ca 21 Sc 22 Ti 23 V 24 Cr 25 Mn 26 Fe 27 Co 28 Ni 29 Cu 30 Zn 31 Ga 32 Ge 33 As 34 Se 35 Br 36 Kr  
37 Rb 38 Sr 39 Y 40 Zr 41 Nb 42 Mo 43 Tc 44 Ru 45 Rh 46 Pd 47 Ag 48 Cd 49 In 50 Sn 51 Sb 52 Te 53 I 54 Xe  
55 Cs 56 Ba 57 La 72 Hf 73 Ta 74 W 75 Re 76 Os 77 Ir 78 Pt 79 Au 80 Hg 81 Tl 82 Pb 83 Bi 84 Po 85 At 86 Rn  
87 Fr 88 Ra 89 Ac 104 Rf 105 Db 106 Sg 107 Bh 108 Hs 109 Mt 110 Ds 111 Rg 112 Cn 113 Nh 114 Fl 115 Mc 116 Lv 117 Ts 118 Og  
58 Ce 59 Pr 60 Nd 61 Pm 62 Sm 63 Eu 64 Gd 65 Tb 66 Dy 67 Ho 68 Er 69 Tm 70 Yb 71 Lu  
90 Th 91 Pa 92 U 93 Np 94 Pu 95 Am 96 Cm 97 Bk 98 Cf 99 Es 100 Fm 101 Md 102 No 103 Lr

Atomly.net 致力于开创世界领先的材料科学数据平台，数据带你触手可得！

346,396 Compounds

- **自主开发、自主计算**
- **我国首个世界级材料计算数据库**
- **数据质量、数量达标世界级数据库**
- **对我国用户全免费开放**

- 34万+无机晶体材料
- 34万+能带和DOS
- 快捷搜索&可视化晶体结构
- 高计算精度及物性推演
  - 键能、形成能
  - 热力学稳定性
  - XRD
  - 化学反应能
  - 磁性
  - 电子结构信息
  - 介电常数
  - 力学张量
- 用户提交结构计算
- 新材料不断扩充中

# Atomly材料科学数据库 (Demo) <https://atomly.net>



**Basic info**

**GaAs** Compound Identification 0000021046

**Material Details**

Unit Cell Parameter		Energies		Symmetry Property	
a: 4.066 Å	α: 60°	Energy: -8.275 eV	Hull: F-42-3		
b: 4.066 Å	β: 60°	Energy Atom: -4.138 eV	Space Group: F-43m		
c: 4.066 Å	γ: 60°	Energy Vdw: -8.883 eV	Point Group: -43m		
Volume: 47.528 Å <sup>3</sup>		Energy Vdw/Mom: -4.442 eV	Crystal System: cubic		

**Crystal Structure**

**Structure**

**X-Ray Diffraction**

**XRD**

Radiation Wavelength (Å): 1.54184 CuKa Particle Size (nm): 200 Generate

**HSE Band Structure**      **HSE Density of State**

**PBE/HSE Bands&DOS**

**Dielectric Properties**

$$\begin{bmatrix} k_{ij}^{xx} & & \\ & 16.425 & 0 & 0 \\ & 0 & 16.425 & 0 \\ & 0 & 0 & 16.425 \end{bmatrix} \quad \begin{bmatrix} k_{ij} & & \\ & 0 & 16.425 & 0 \\ & 0 & 0 & 16.425 \end{bmatrix} \quad \begin{array}{l} k^x: 16.425 \\ k^y: 18.644 \\ n: 4.053 \end{array}$$

**Others**

**Dielectric / Piezoelectric data**

**Piezoelectric Properties**

$$\begin{bmatrix} e_{ij} (C/m^2) & & \\ & 0 & 0 & 0 & 0.369 & 0 \\ & 0 & 0 & 0 & 0 & 0.369 \\ & 0 & 0 & 0.369 & 0 & 0 \end{bmatrix} \quad \begin{array}{l} ||e_{ij}||_{max} \\ e_{33} \dots e_{11} \dots e_{22} \dots e_{33} \dots e_{11} \dots e_{22} \dots \\ V_{max} \end{array}$$

**More Information**

Band Gap Type	direct	Band Gap	0.1425 eV	Nsites	2
E <sub>hull</sub>	0 eV/atom	Magnetic Moment	0 μ <sub>B</sub>	Formation Energy	-0.349 eV/atom
Elements	Ga As	Elements Quantity	1 1	Fermi Energy	3.5571 eV
Decomposition To					Stable
Calculation Method	GGA / HSE	Calculation Date	2019-5-19	Vasp Version	5.4.1
Potcar Species	PAW_PBE-Ga_d PAW_PBE-As				
U Value	---				

**Calculation details**

**PBE Band Structure**      **PBE Density of State**

**Phase Diagram**

**Thermodynamics**

**Stable**      **Unstable**

$\Delta E$ : 0 Formula: 0  $\mu_{\text{Mol}}/\text{Hartree}$ : < 0

000002109	As	0.12
0000010430	As	0.032
0000079514	As	0.124
0000015483	As	0.123
1000000247	Ge	0.003
1000000662	Ge	0.034
1000001369	Ge	0.036
0000029470	Ge	0.041
0000005406	Ge	0.054

**Brillouin Zone**

**Reciprocal space**



# 数据库世界领先

	Materials Project	Atomly
计算精度	$10^{-4}$ eV/atom	$10^{-5}$ eV/atom
材料结构	145k	349k
电子结构	76k	343k
杂化泛函	N/A	2k
介电常数	4k	12k
力学性质	14k	16k
小分子	63k	2M
电化学材料	5k	9k

已大幅超 世界第一的Materials Project

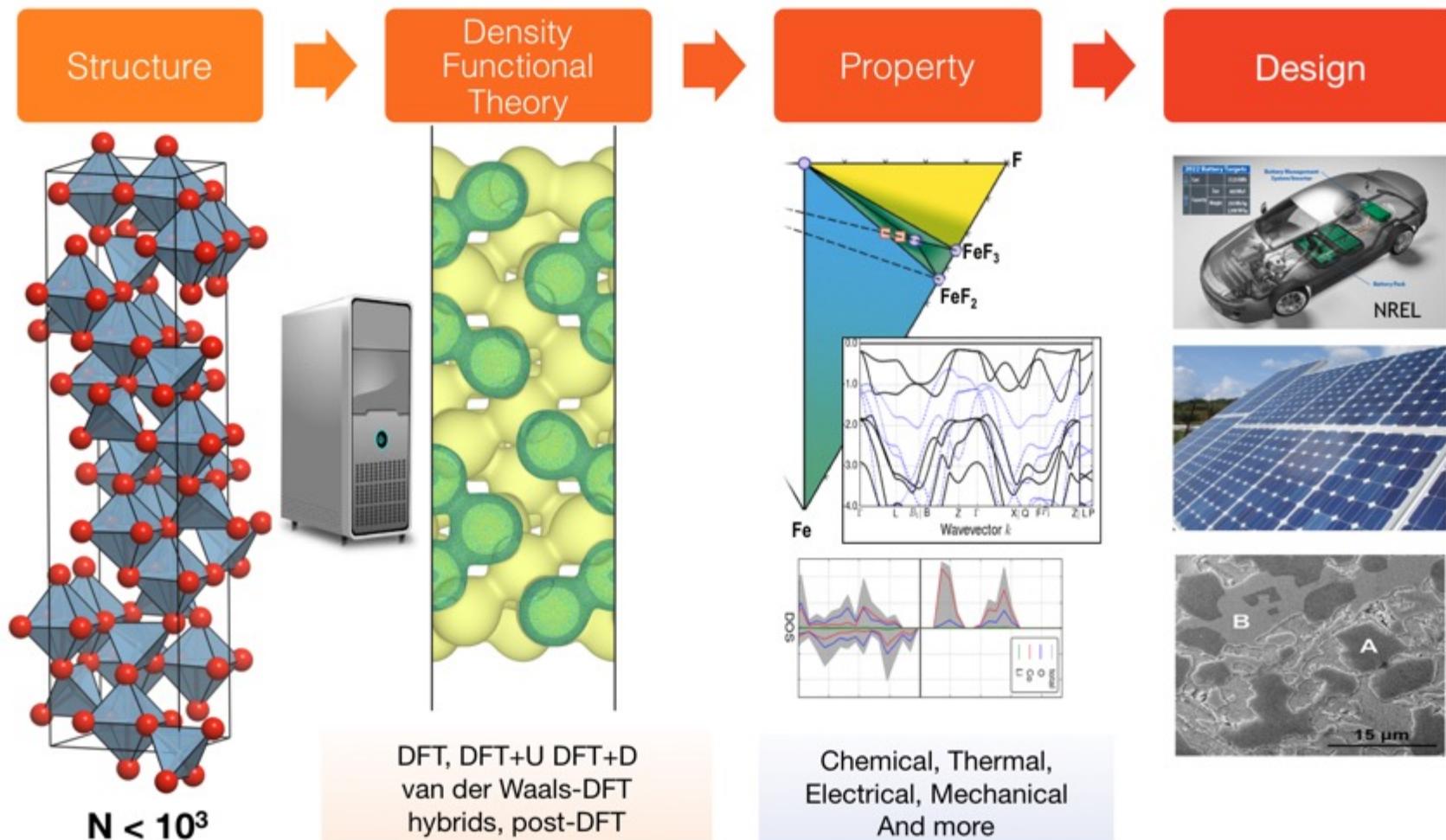
为领域提供海量优质数据

- 用户覆盖全国
- 70万次点击量
- 5300+用户

## Upcoming datasets

- ✓ 低维材料数据集 (2023)
- ✓ 单原子催化 (2023)
- ✓ MOF数据集 (2023)
- ✓ 小分子数据集 (2023)
- ✓ TB数据集 (2023)
- ✓ 极端条件数据集 (2024)
- ✓ 实验料数据集 (2025)

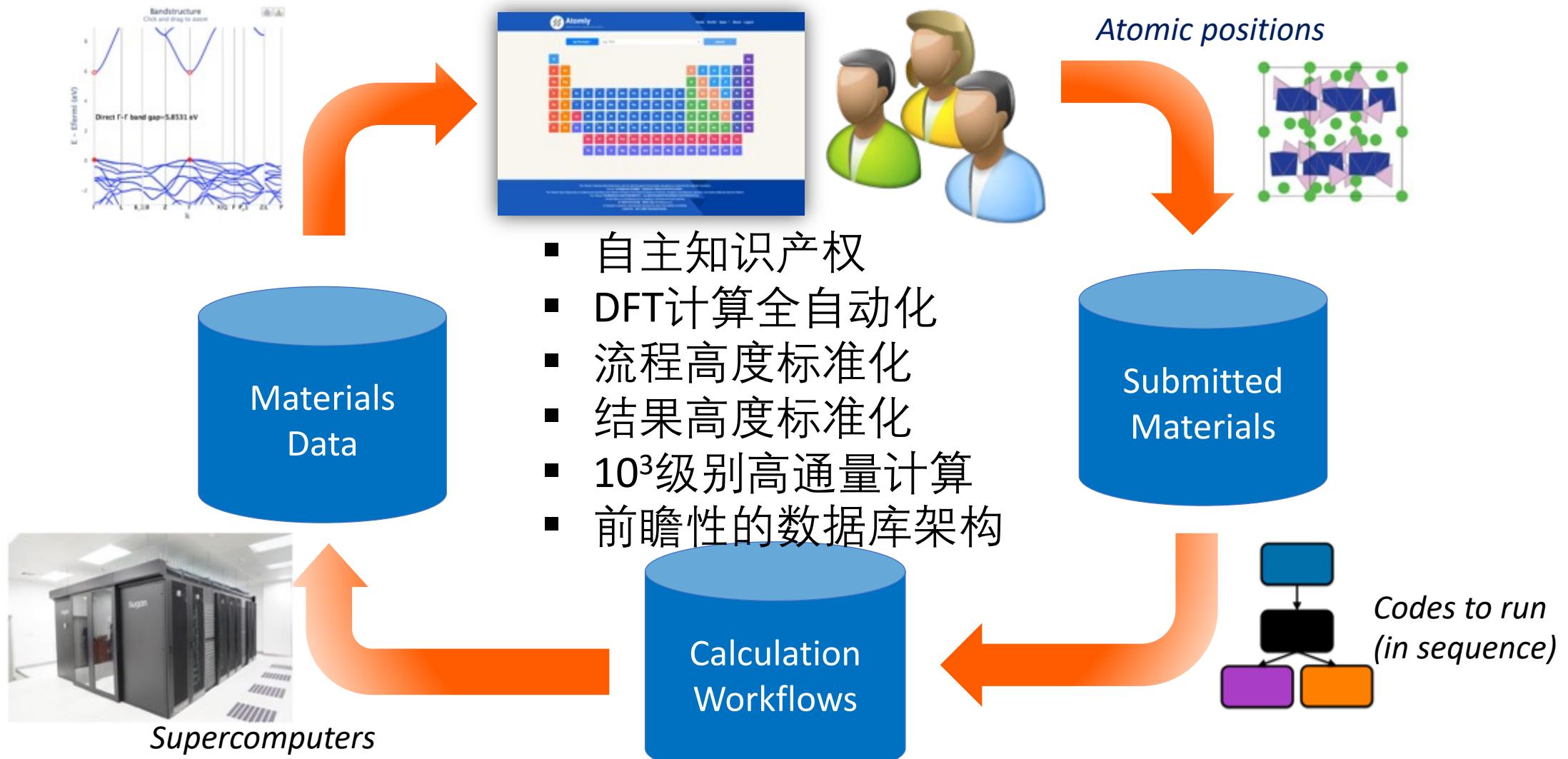
# 3 DFT计算



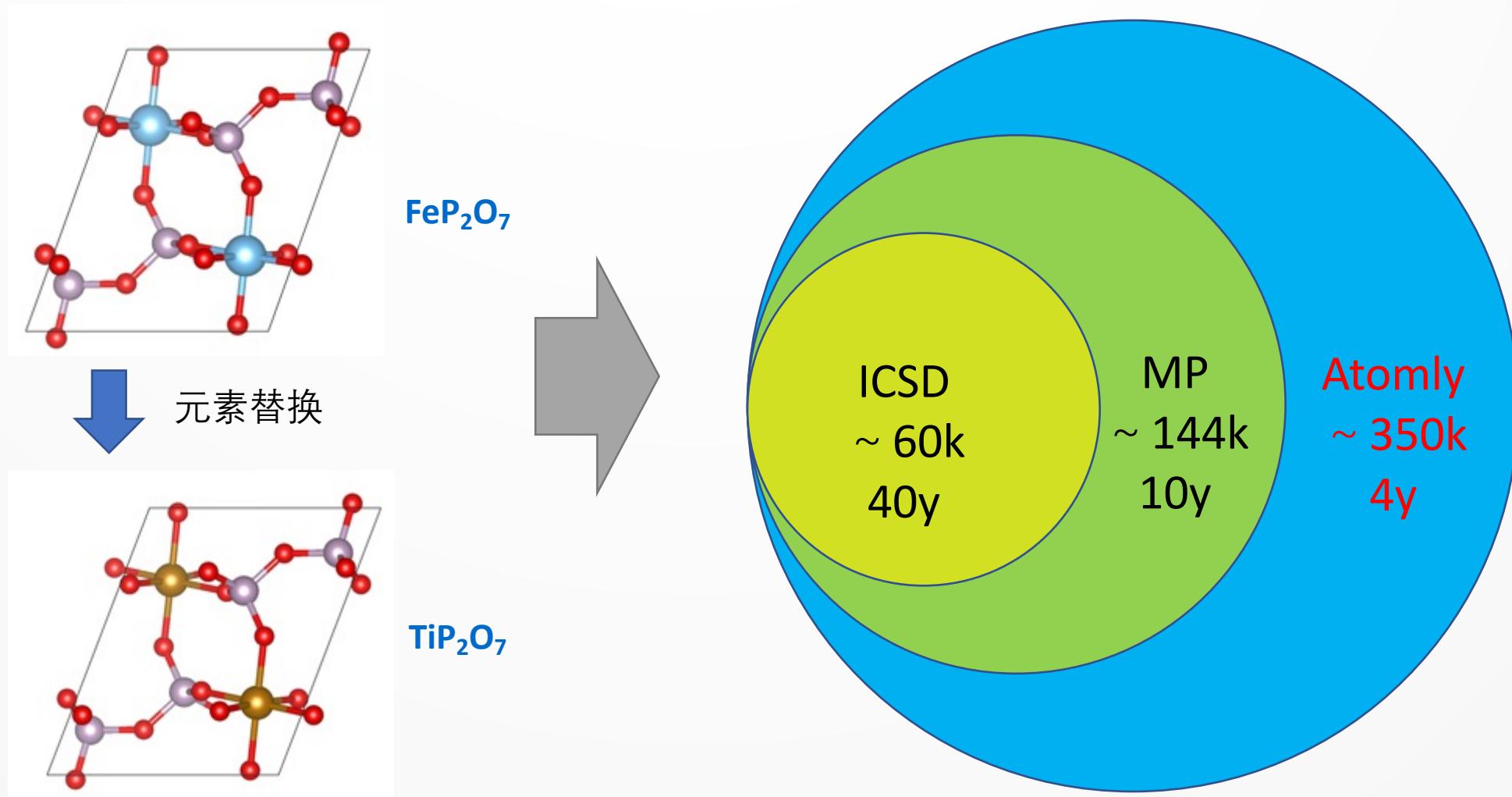
原子结构  
电子结构

XRD  
热力学稳定性  
电荷疏运  
声子/震动  
光吸收  
力学性质  
表面结构  
磁结构  
.....

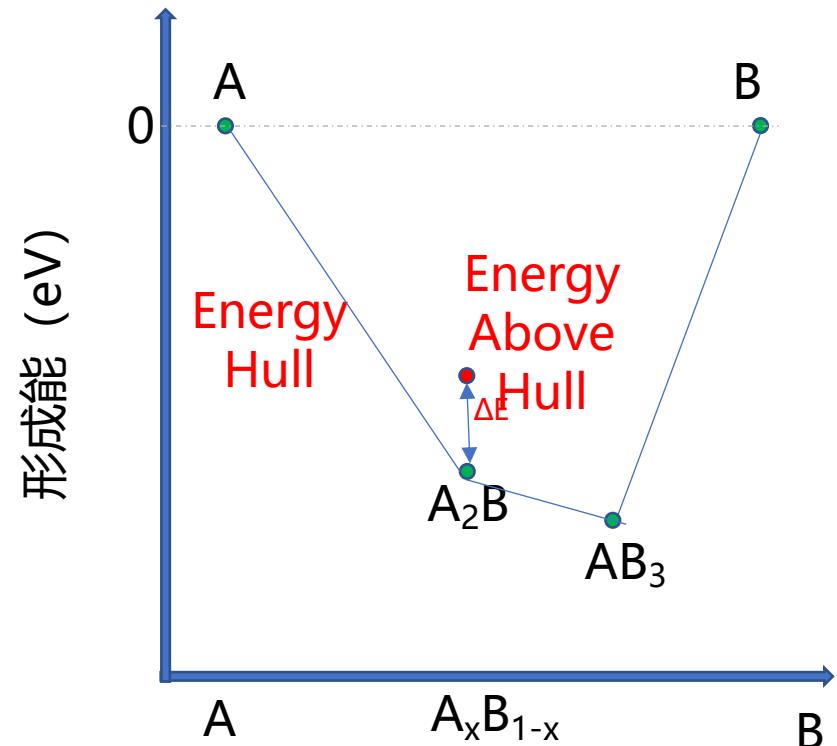
# 3 DFT高通量计算



### 3 计算未知材料（元素替换产生新结构）



### 3 探索未知相空间（热力学稳定相）

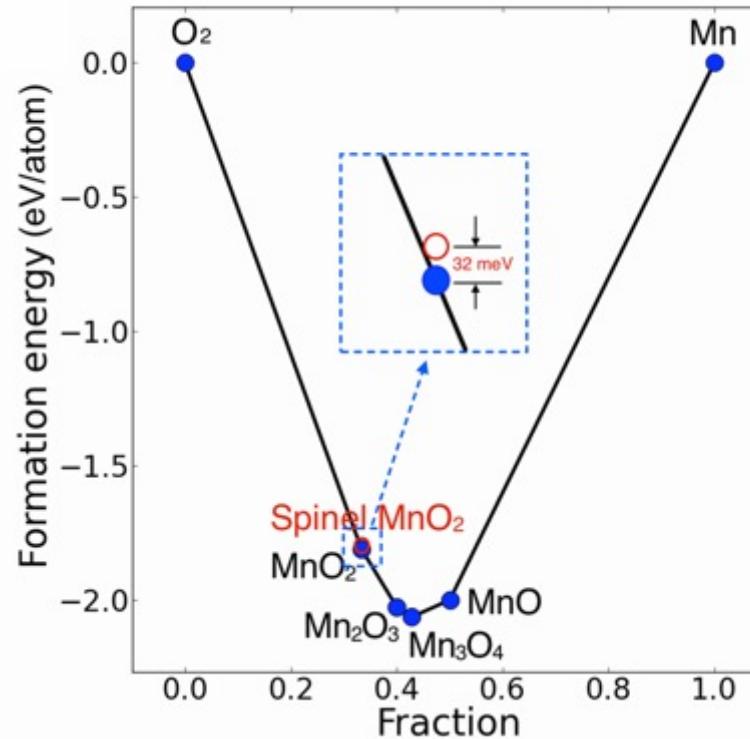


在50k新材料中

$E^{hull} < 100\text{meV / atom}$ : 30540 compounds

$E^{hull} < 50\text{meV / atom}$ : 16511 compounds

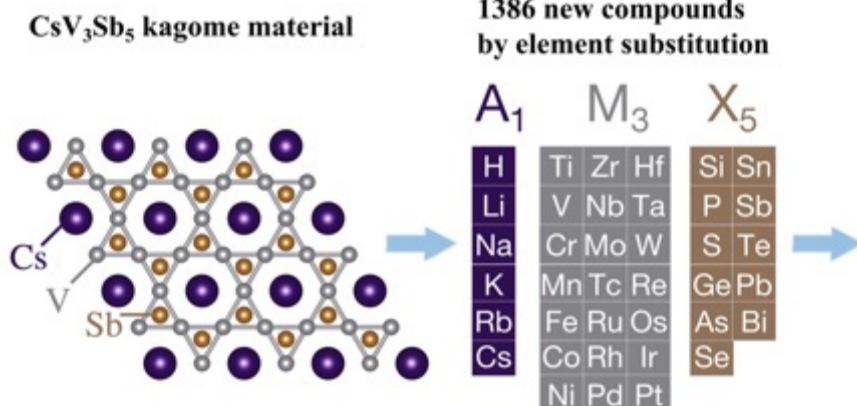
$E^{hull} < 10\text{meV/atom}$  : 6940 compounds



存在有大量未知稳定相

# 4 拓扑材料发现

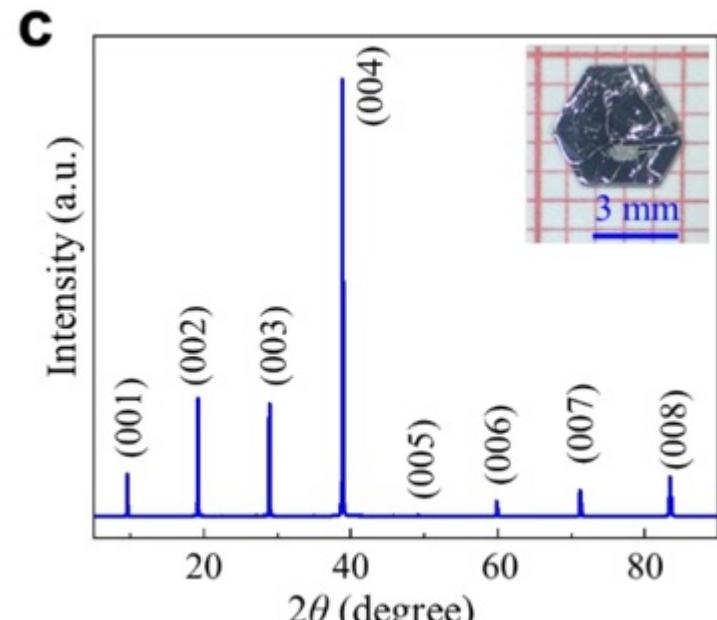
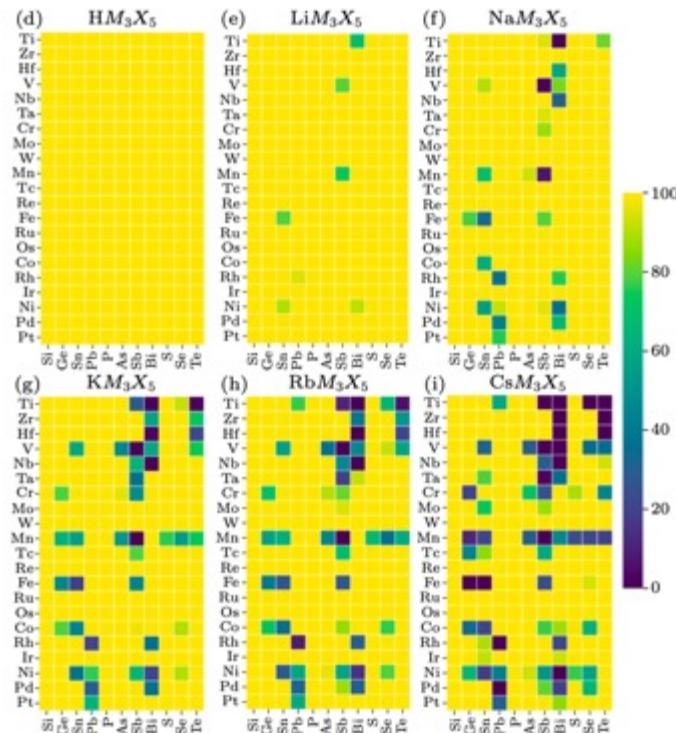
## 扩展化学空间



CPL, 39, 047402 (2022) Express letters

成功预言新材料，加速材料发现过程

## 高通量计算



Zeitschrift für Naturforschung B, 77, 11(2022)

PRL, 131, 026701 (2023)

Nature Comm, 14, 4089 (2023)

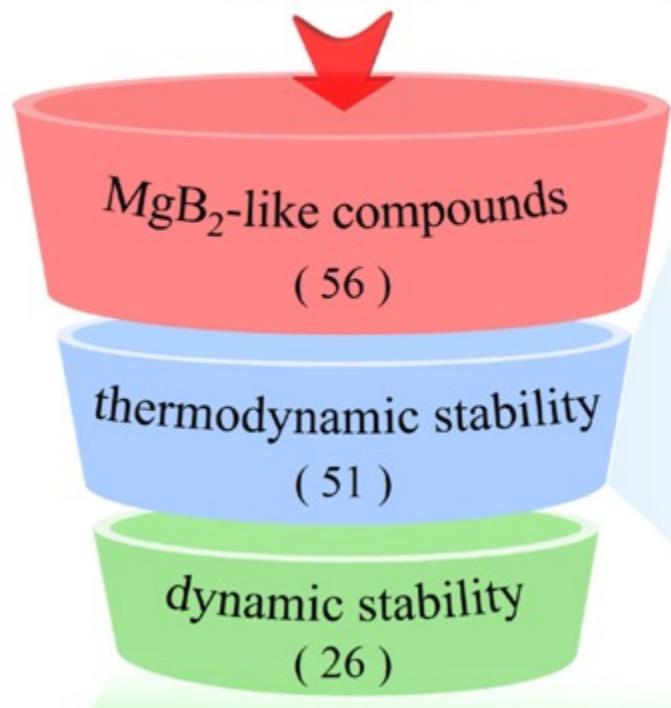
Nature Phys (2023)



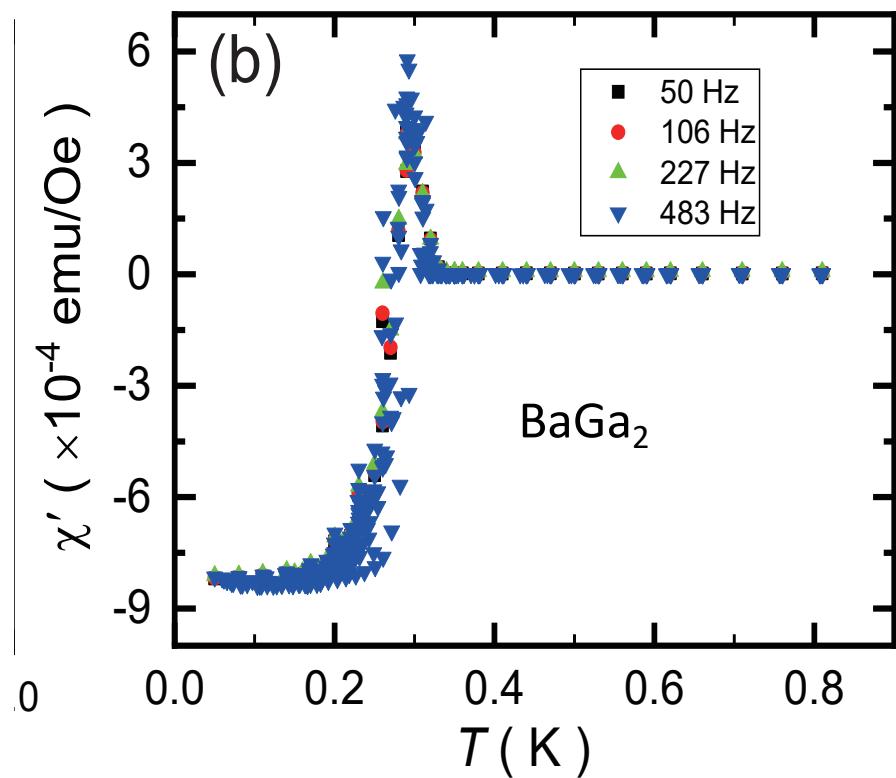
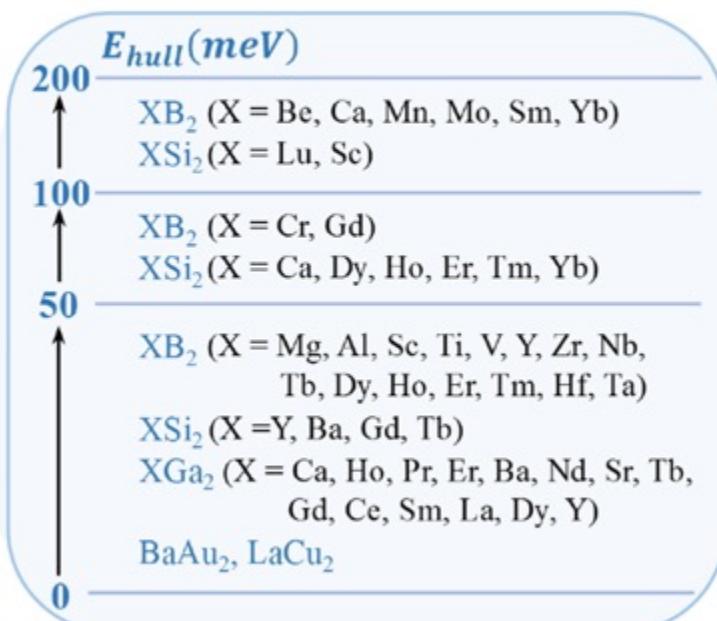
# 4 超导材料搜索



182155 compounds from “Atomly”

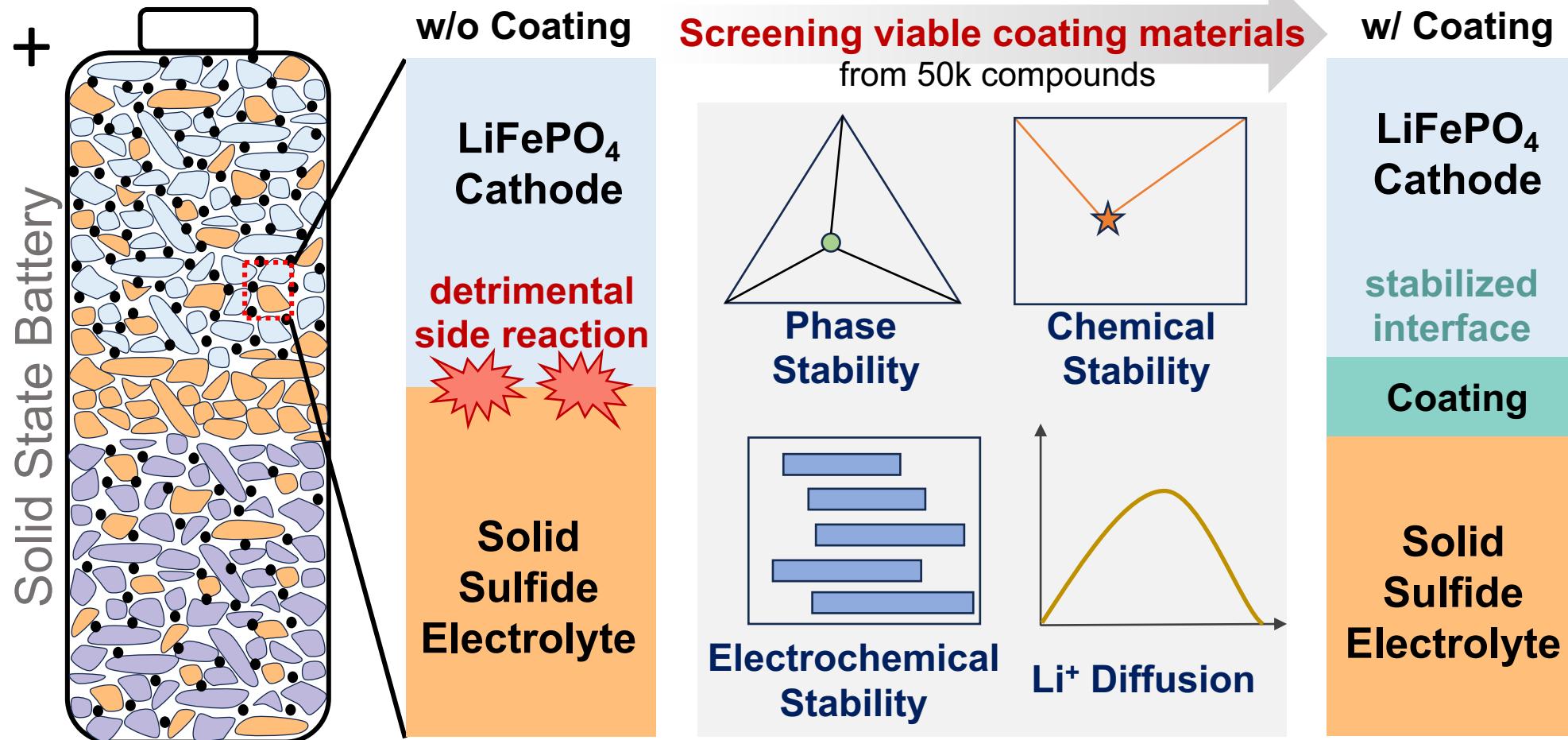


## $MgB_2$ -like superconductors

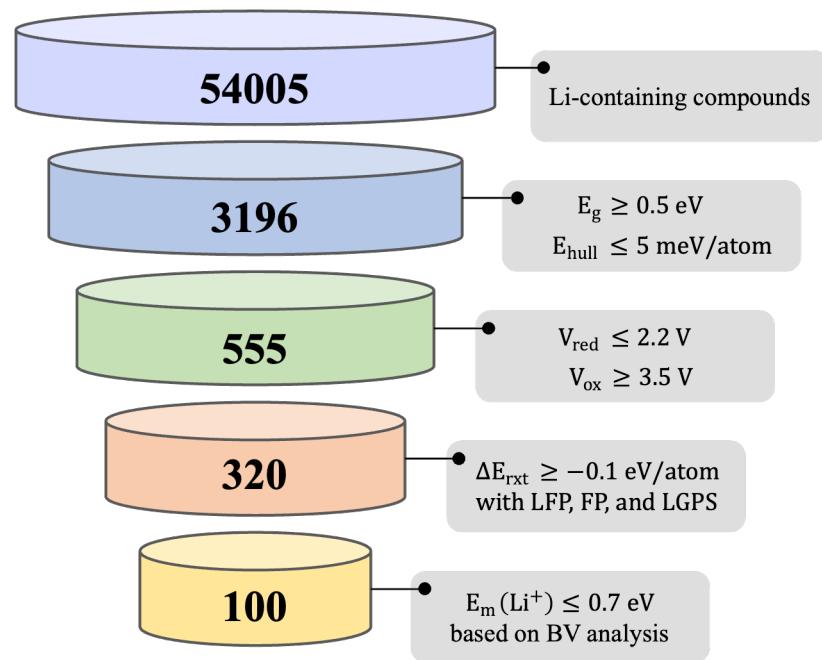
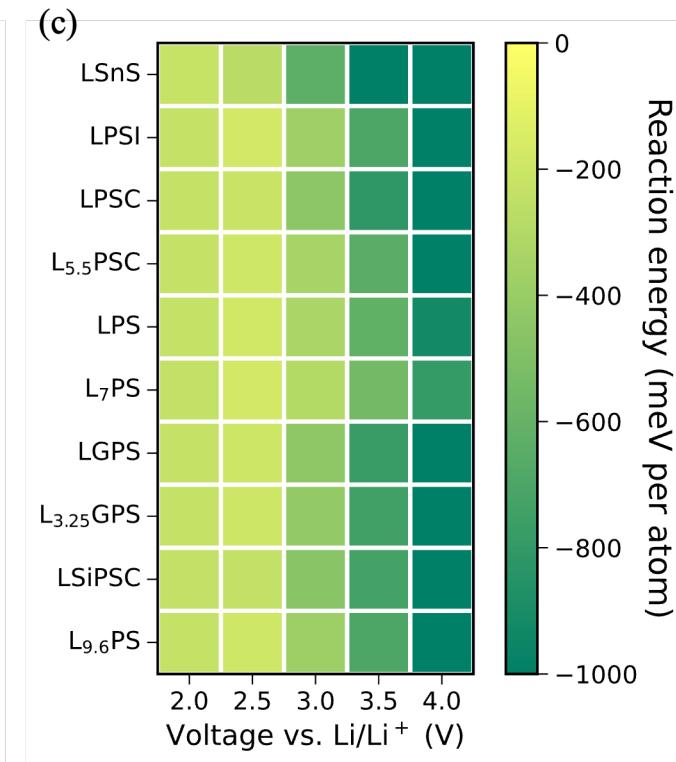
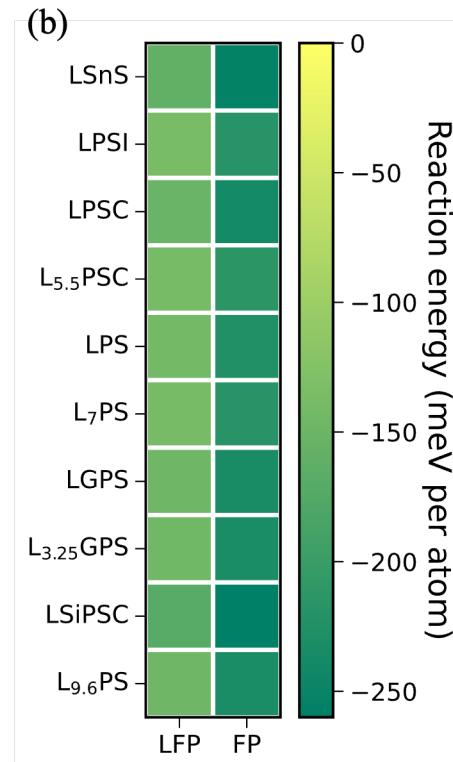
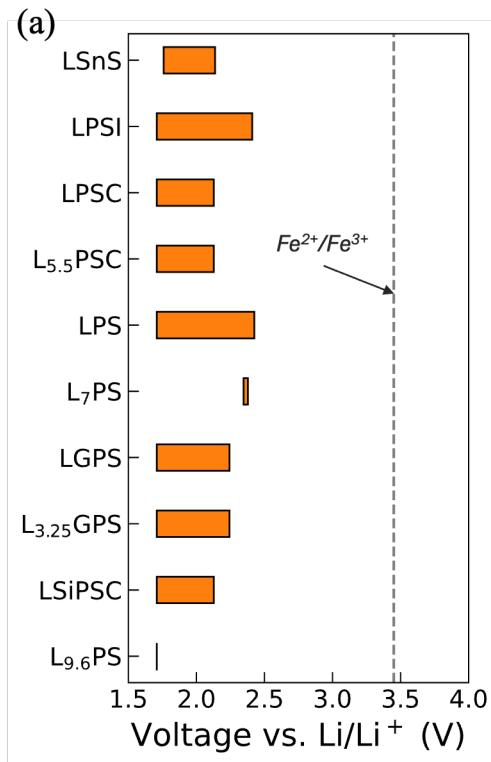


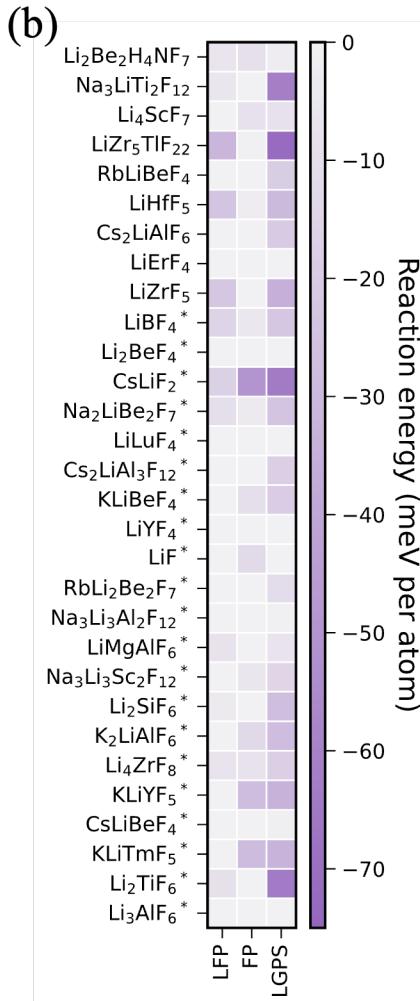
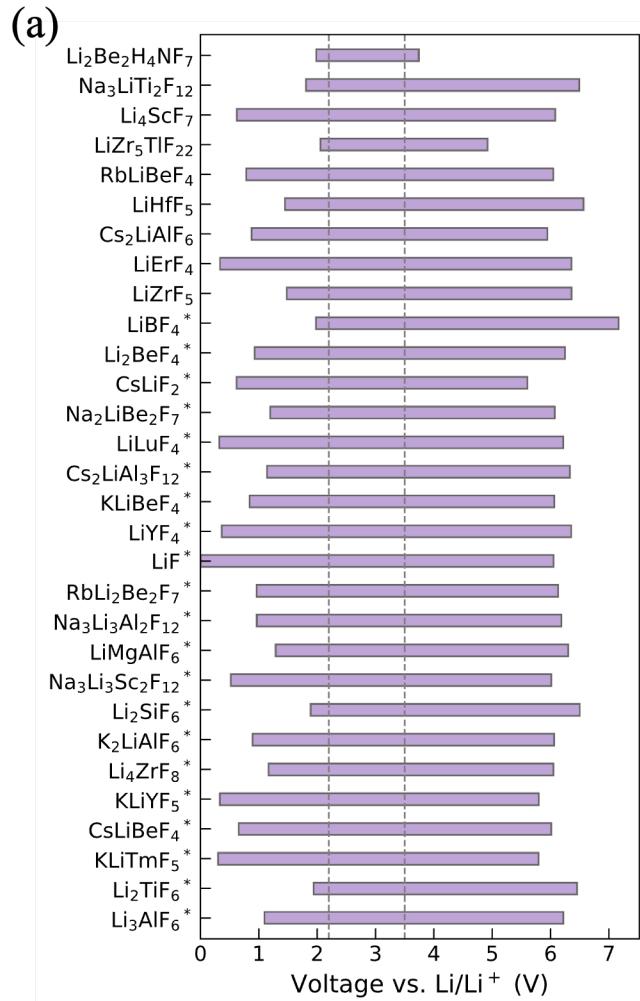
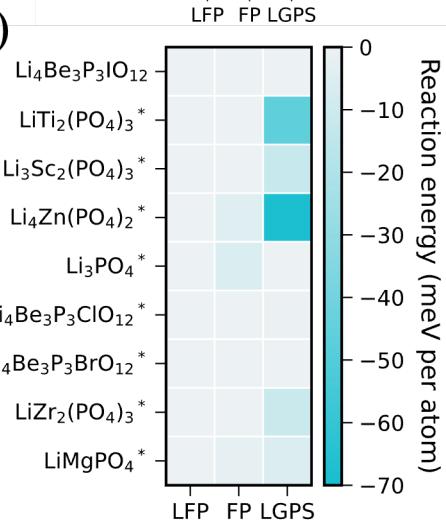
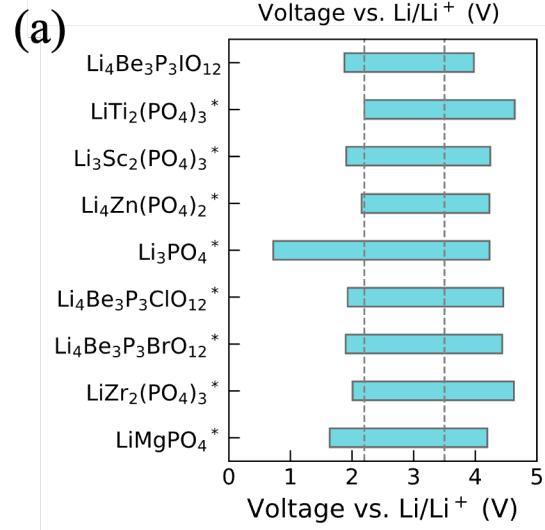
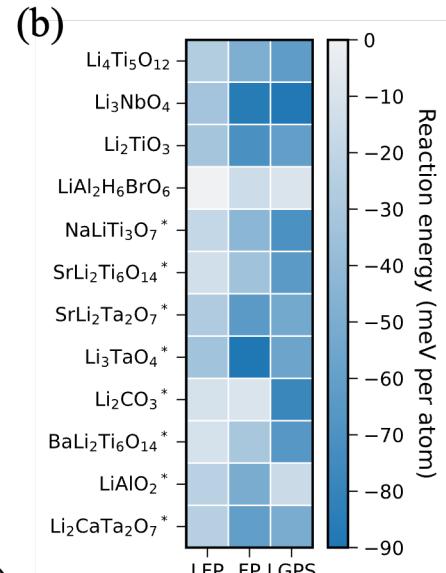
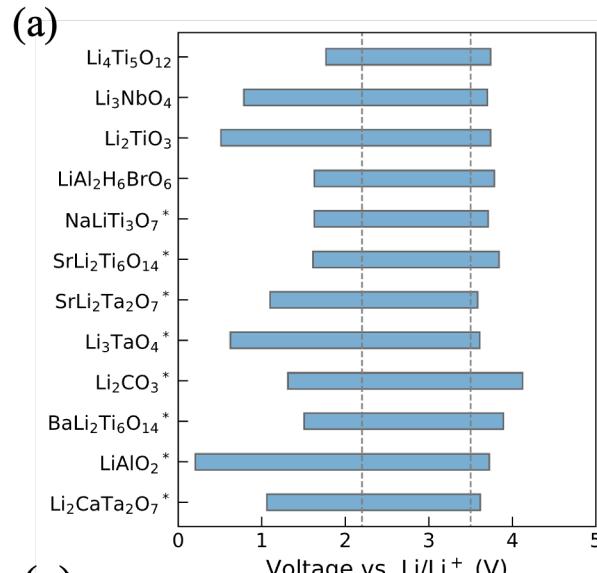
End-to-end screening the targeted compound in a once and for all fashion

Miao Liu, PRB, 105, 214517 (2022)



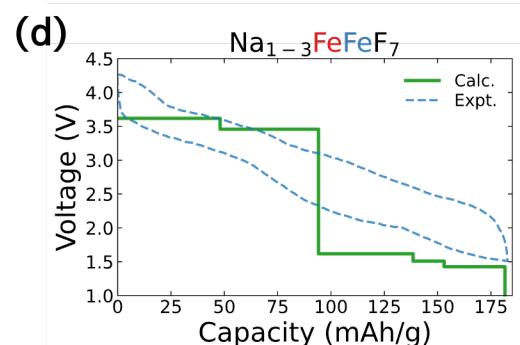
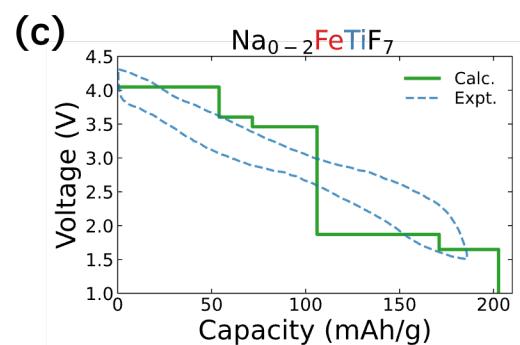
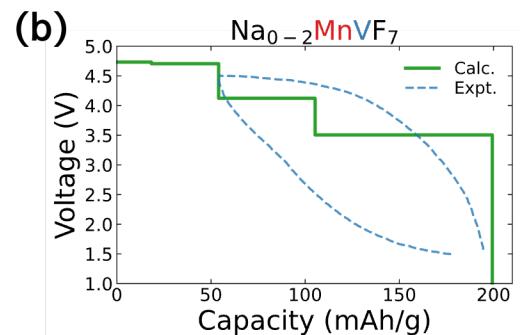
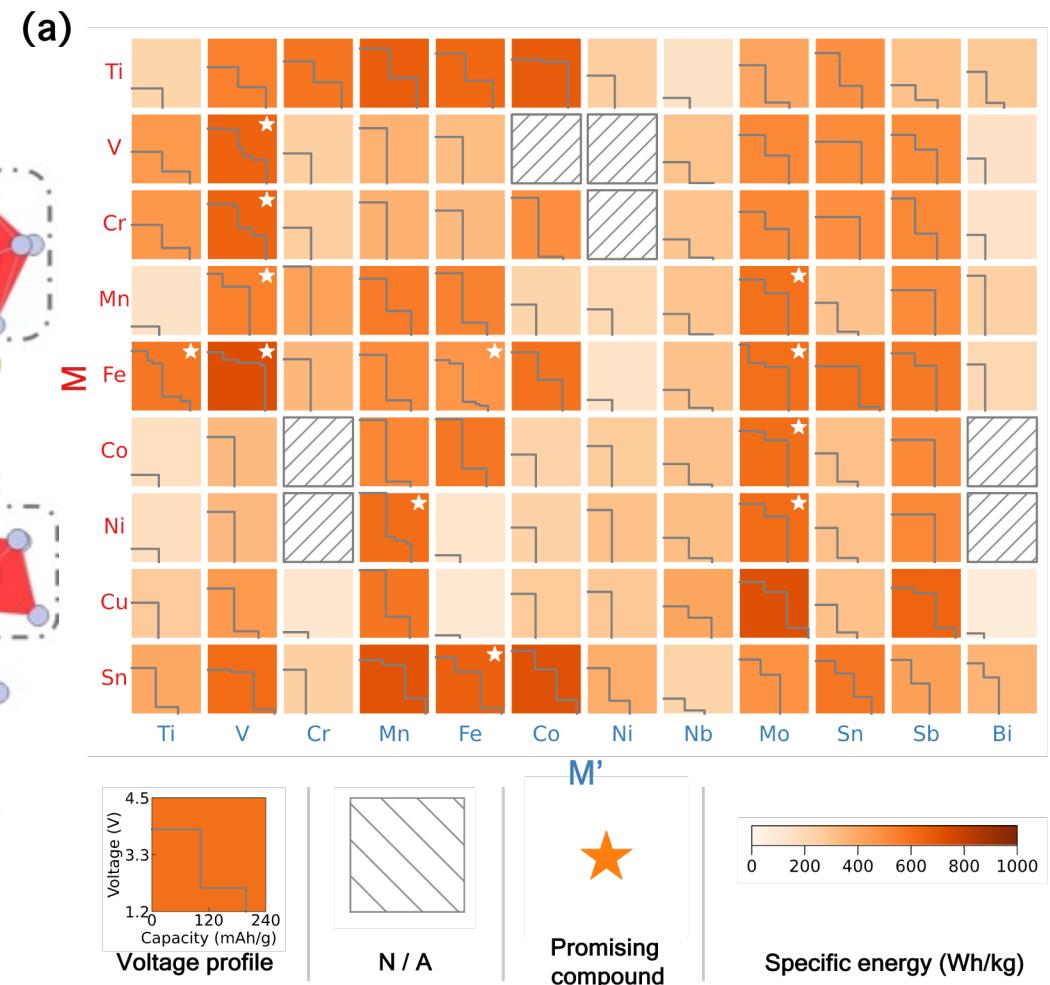
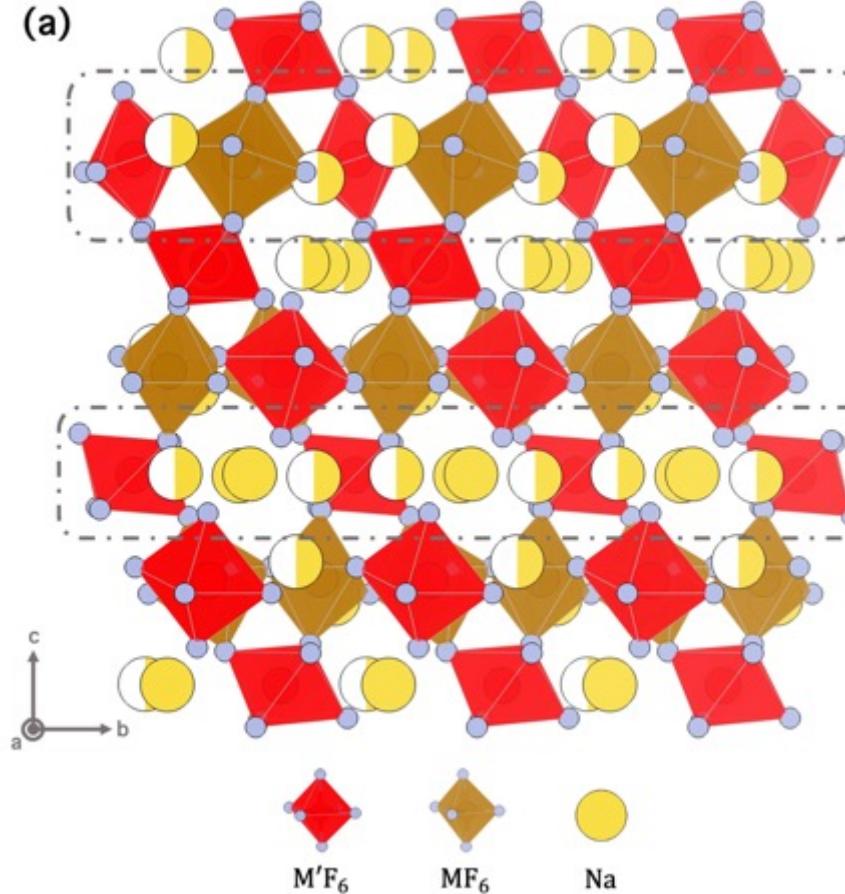
## LFPO vs 硫化物固态电解质

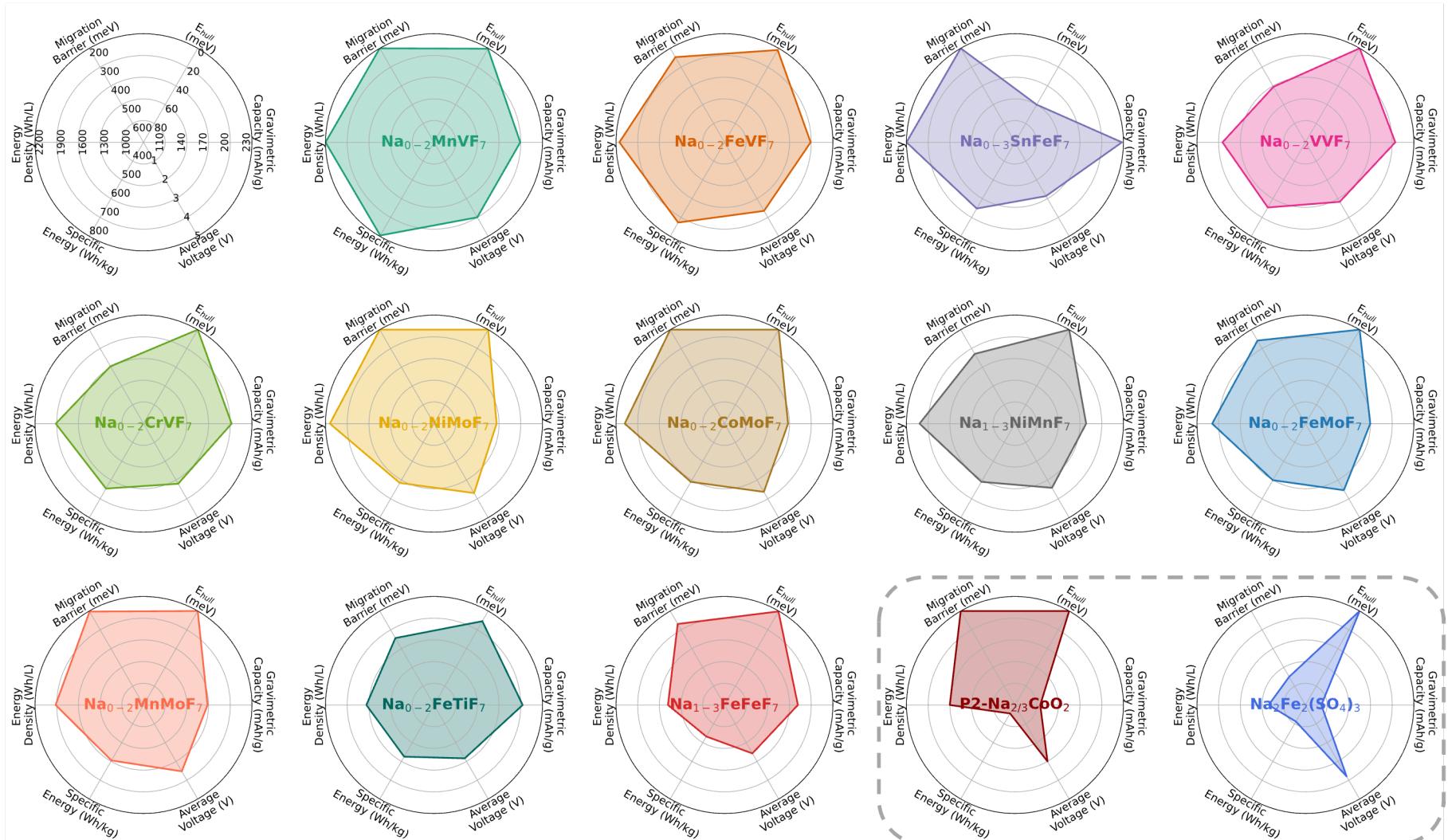




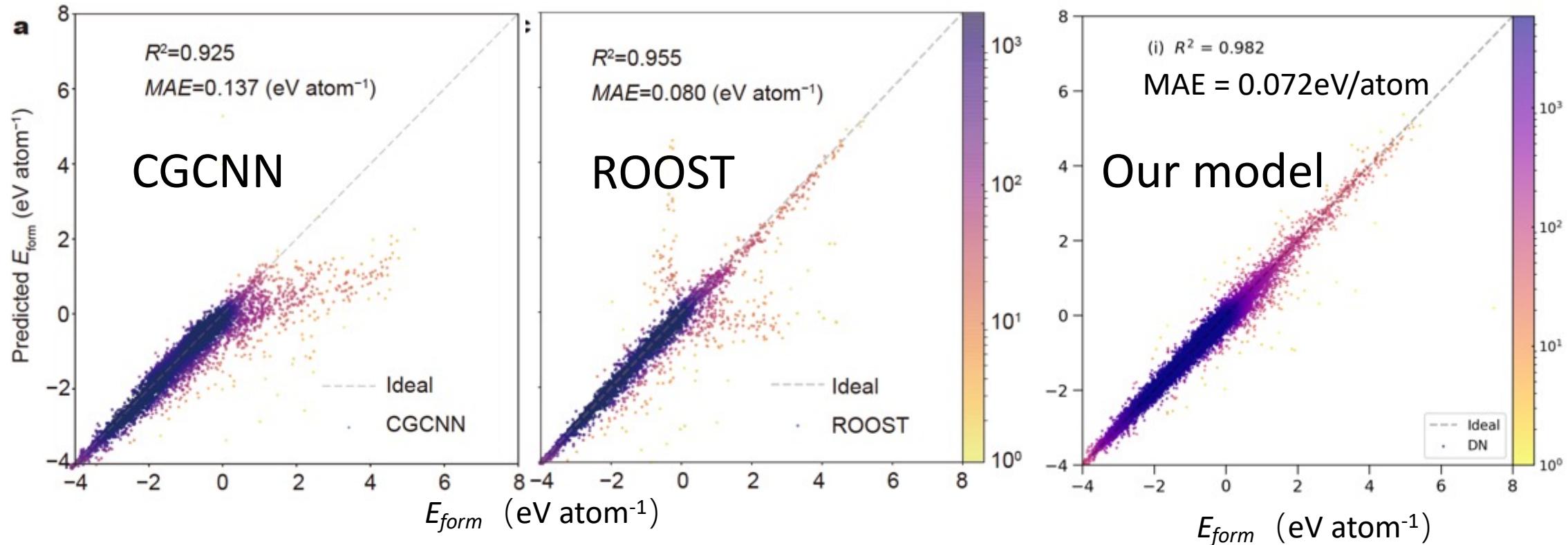
## 5

## Na电池正极





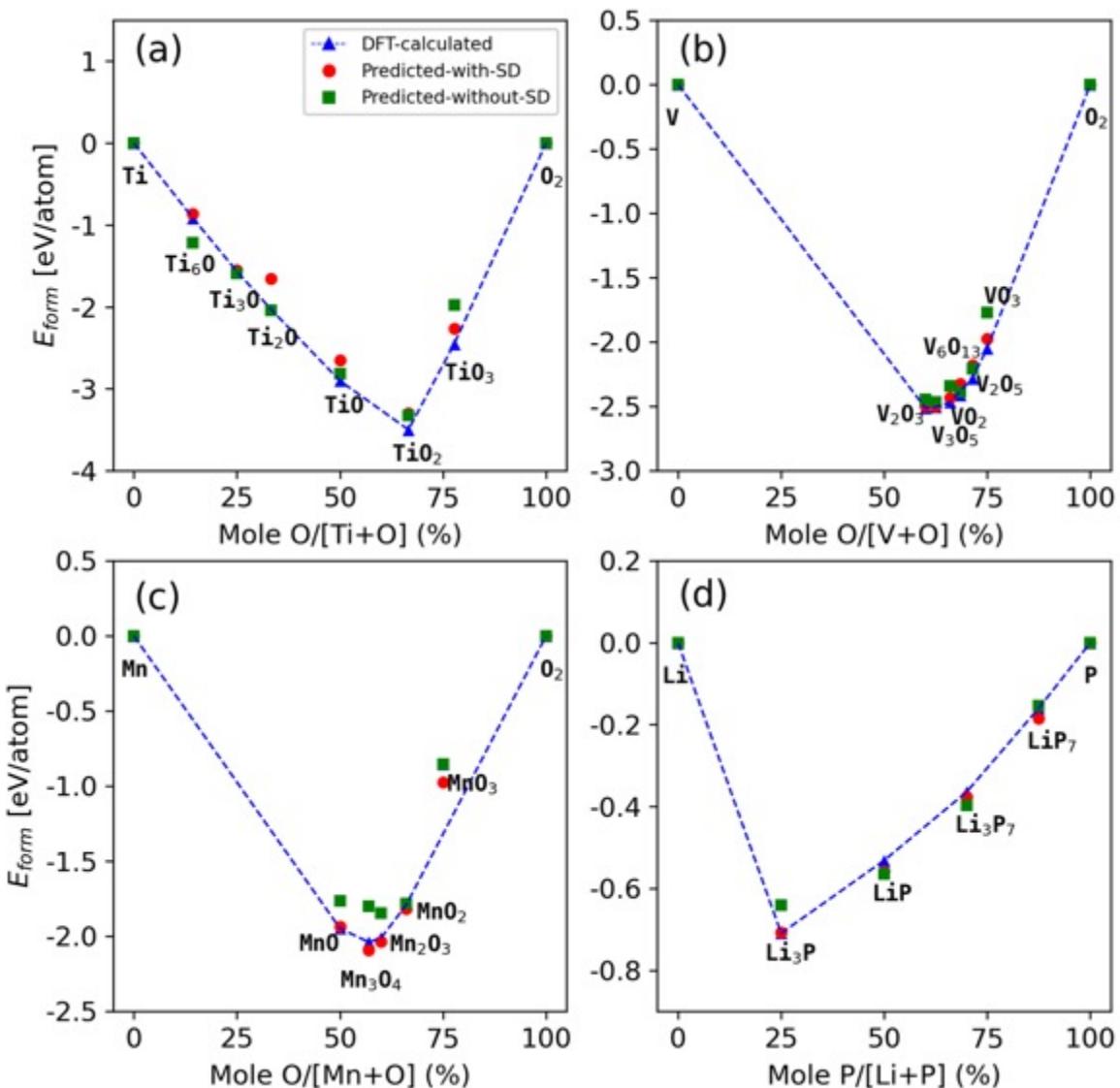
# 6 人工智能模型, good, bad and ugly



	CGCNN	ROOST
文献	0.039eV/atom	0.024eV/atom
实际	0.137eV/atom	0.080eV/atom

Science China Materials, 66, 343 (2023)

# 6 人工智能模型预测结合能 (V1)



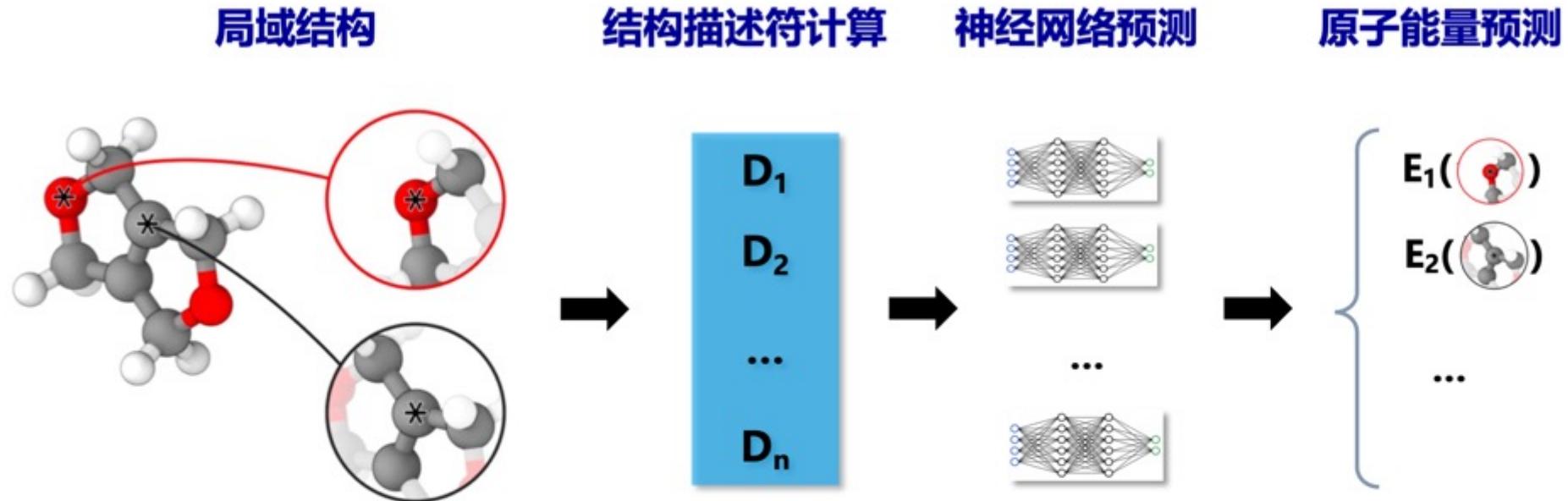
秒级预测材料的热力学稳定性

Ti-O/V-O/Mn-O/Li-P 体系  
精度高

可用于快速探索未知材料体系

Science China Materials, 66, 343 (2023)

# 7 人工智能模型预测结合能 (V2)



**原子受力**  $f_{j \rightarrow i} = \frac{\partial E_i}{\partial r_j} = \frac{\partial E_i}{\partial d_i} \frac{\partial d_i}{\partial r_j}$

$$f_i = \sum_k^{N(i)} f_{k \rightarrow i}$$

**原子应力**  $v_{\alpha\beta}^i = \sum_k^{N(i)} r_{k \rightarrow i}^\alpha f_{k \rightarrow i}^\beta$

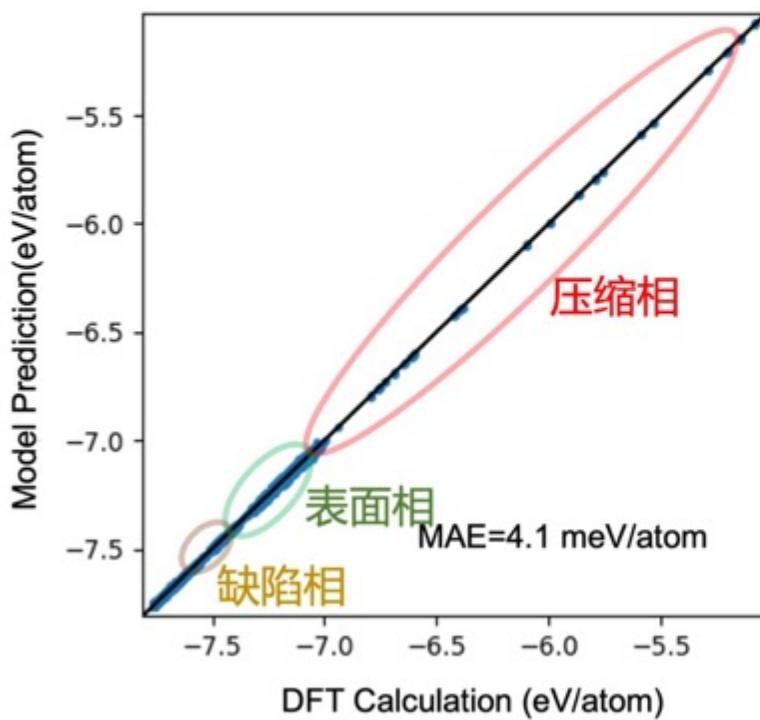
**整体应力**  $v_{\alpha\beta} = \sum_i^n v_{\alpha\beta}^i$

# 7 人工智能模型预测结合能 (V2)

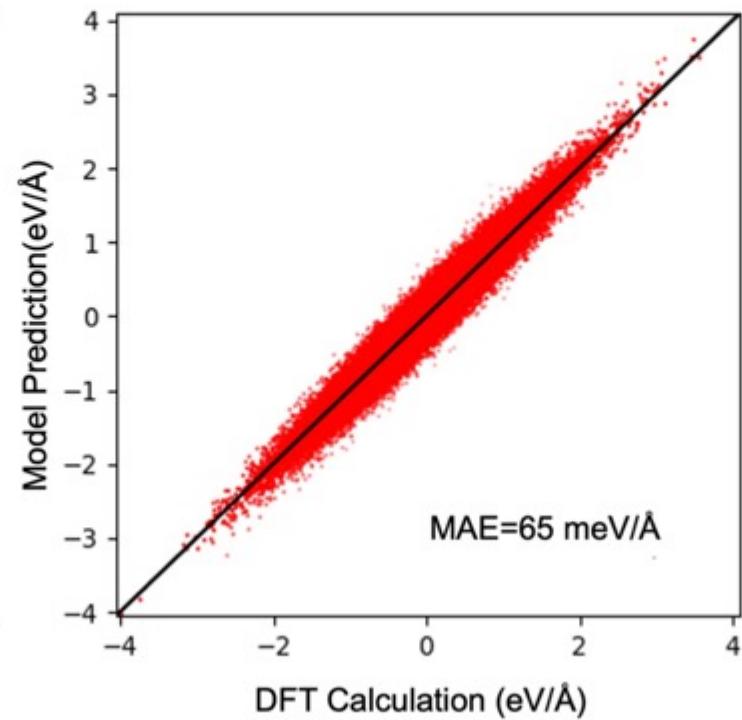


## 高精度Ti金属机器学习力场

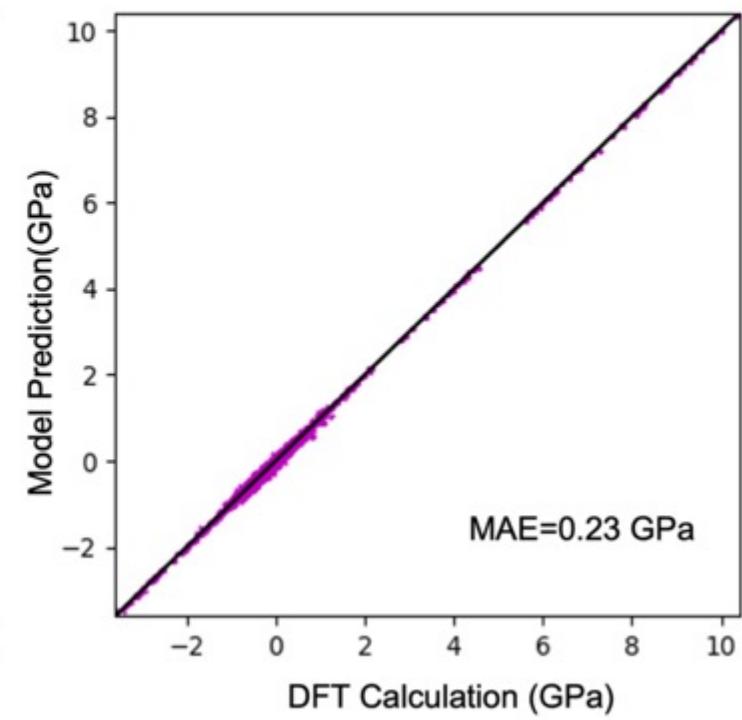
能量误差



受力误差



应力误差



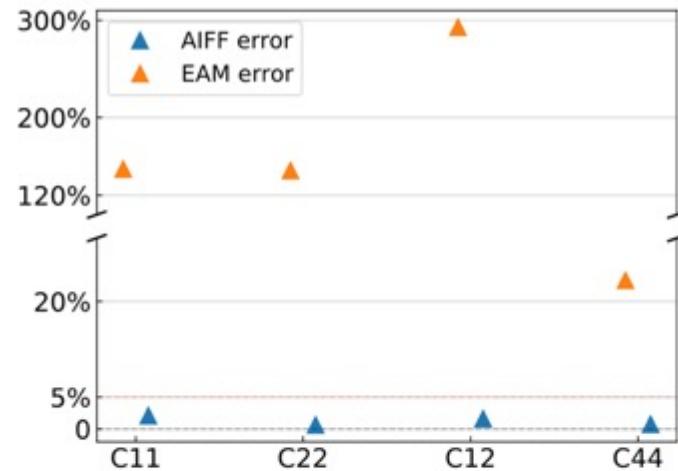
# 7 人工智能模型预测结合能 (V2)

## 弹性属性

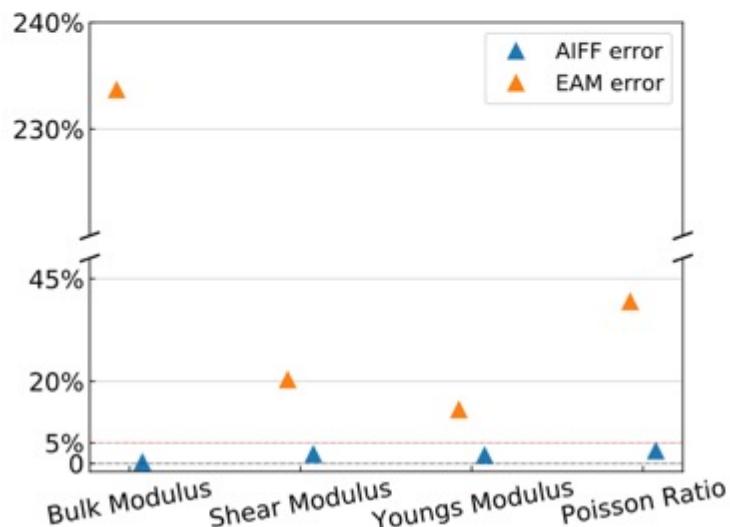
$$\sigma_i = \sum_{j=1}^6 C_{ij} \epsilon_j$$

$$\begin{pmatrix} \sigma_{11} \\ \sigma_{22} \\ \sigma_{33} \\ \sigma_{12} \\ \sigma_{23} \\ \sigma_{31} \end{pmatrix} = \begin{pmatrix} C_{11} & C_{12} & C_{13} & 0 & 0 & 0 \\ C_{12} & C_{22} & C_{23} & 0 & 0 & 0 \\ C_{13} & C_{23} & C_{33} & 0 & 0 & 0 \\ 0 & 0 & 0 & C_{44} & 0 & 0 \\ 0 & 0 & 0 & 0 & C_{55} & 0 \\ 0 & 0 & 0 & 0 & 0 & C_{66} \end{pmatrix} \begin{pmatrix} \epsilon_{11} \\ \epsilon_{22} \\ \epsilon_{33} \\ \epsilon_{12} \\ \epsilon_{23} \\ \epsilon_{31} \end{pmatrix}$$

	DFT	AIFF	EAM
C <sub>11</sub> (GPa)	176.77 (176.1 <sup>a</sup> )	173.06	436.52
C <sub>22</sub> (GPa)	177.37	178.49	435.41
C <sub>12</sub> (GPa)	95.55(86.9 <sup>a</sup> )	97.09	375.41
C <sub>33</sub> (GPa)	198.33(190.5 <sup>a</sup> )	193.0	414.26
C <sub>44</sub> (GPa)	39.18 (50.8 <sup>a</sup> )	39.47	48.32
Bulk Modulus (GPa)	118.78 (110.0 <sup>a</sup> )	119.10	396.34
Shear Modulus (GPa)	44.11	43.13	35.13
Youngs Modulus (GPa)	117.76(109.1 <sup>a</sup> )	115.45	102.37
Poisson Ratio ( $\mu V$ )	0.33	0.34	0.46



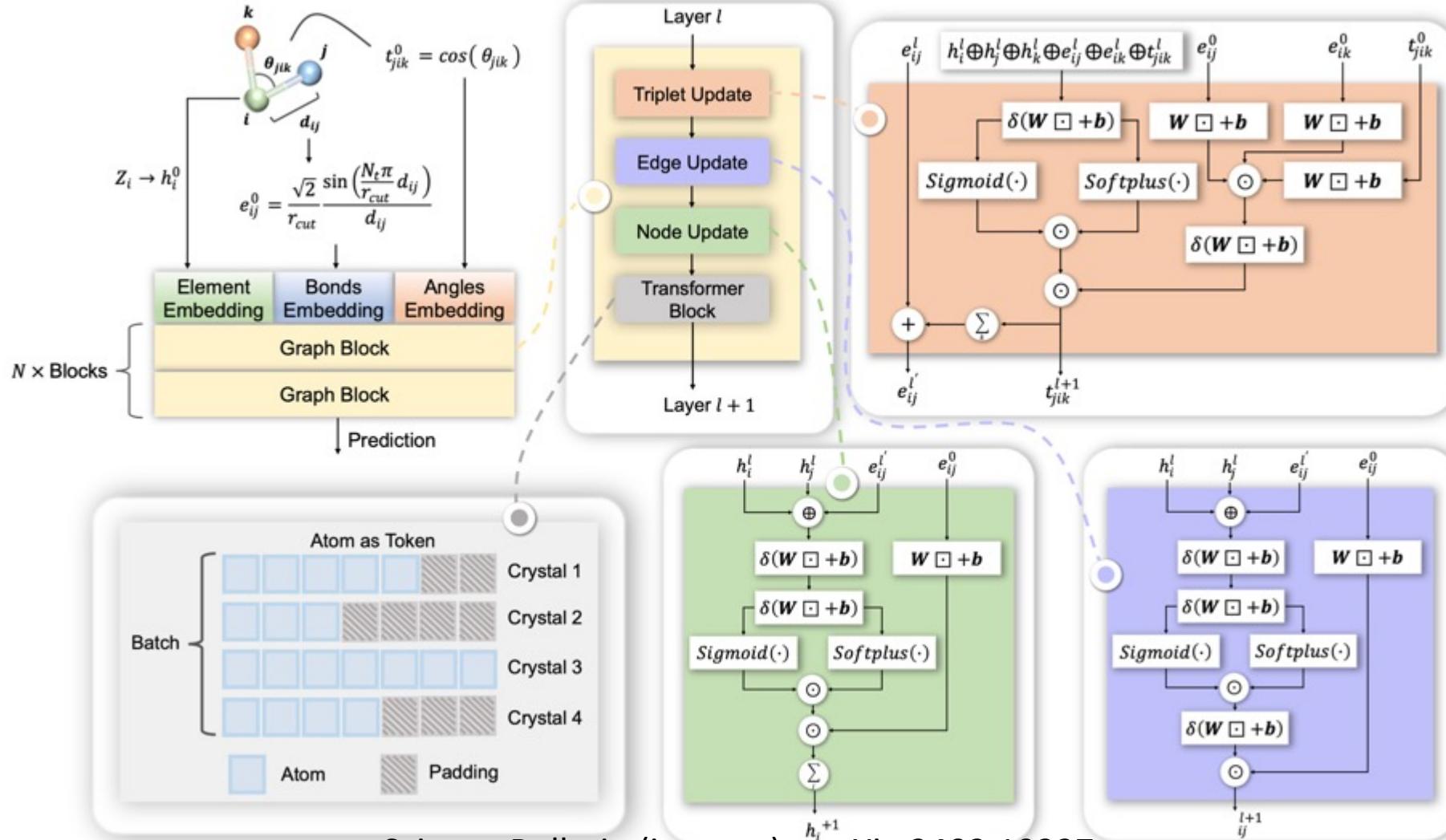
## 弹性常数



## 弹性模量

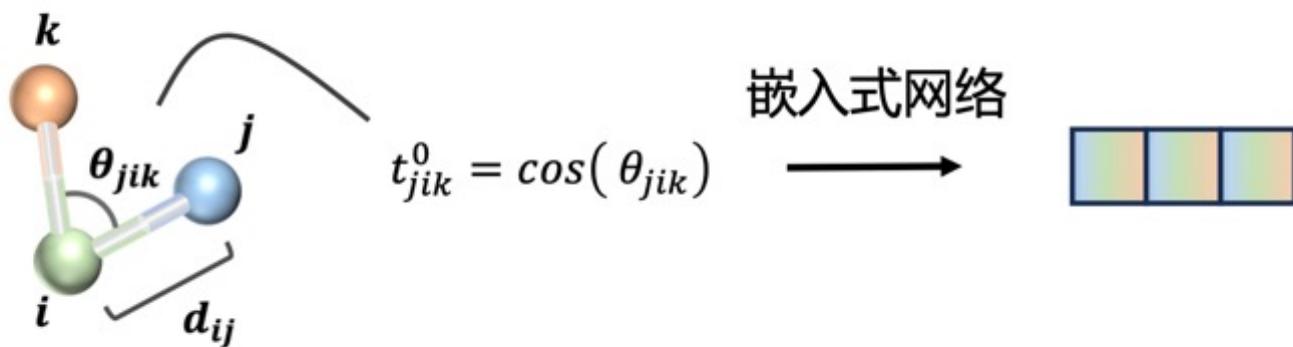
# 8 通用力场 (适用于所有元素组合)

## Graph-based Pretrained Transformer Force Field (GPTFF)



# 8 通用力场 (GPTFF)

## 角向量更新

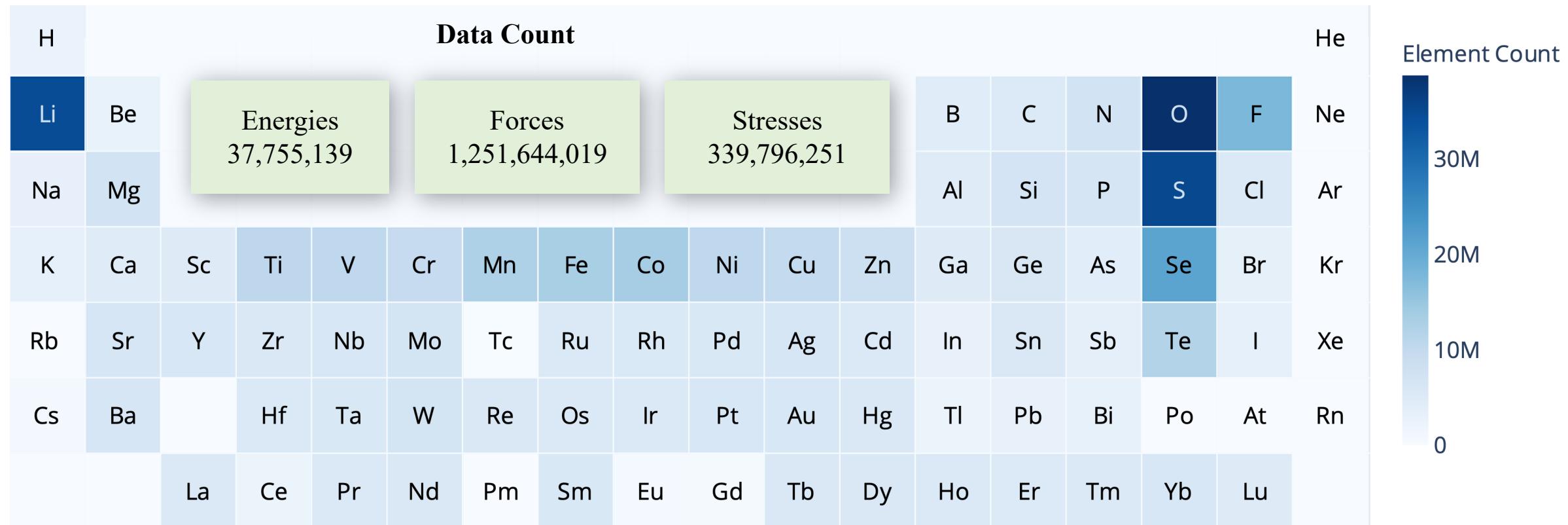


$$t_{jik}^{l+1} = \left( \begin{array}{c|c|c|c|c|c} \text{原子 } j & \text{原子 } i & \text{原子 } k & \text{键 } ij & \text{键 } ik & \text{键角 } jik \end{array} \oplus \begin{array}{c|c|c|c|c|c} \text{原子 } j & \text{原子 } i & \text{原子 } k & \text{键 } ij & \text{键 } ik & \text{键角 } jik \end{array} \oplus \begin{array}{c|c|c|c|c|c} \text{原子 } j & \text{原子 } i & \text{原子 } k & \text{键 } ij & \text{键 } ik & \text{键角 } jik \end{array} \oplus \begin{array}{c|c|c|c|c|c} \text{原子 } j & \text{原子 } i & \text{原子 } k & \text{键 } ij & \text{键 } ik & \text{键角 } jik \end{array} \oplus \begin{array}{c|c|c|c|c|c} \text{原子 } j & \text{原子 } i & \text{原子 } k & \text{键 } ij & \text{键 } ik & \text{键角 } jik \end{array} \oplus \begin{array}{c|c|c|c|c|c} \text{原子 } j & \text{原子 } i & \text{原子 } k & \text{键 } ij & \text{键 } ik & \text{键角 } jik \end{array} \right)$$

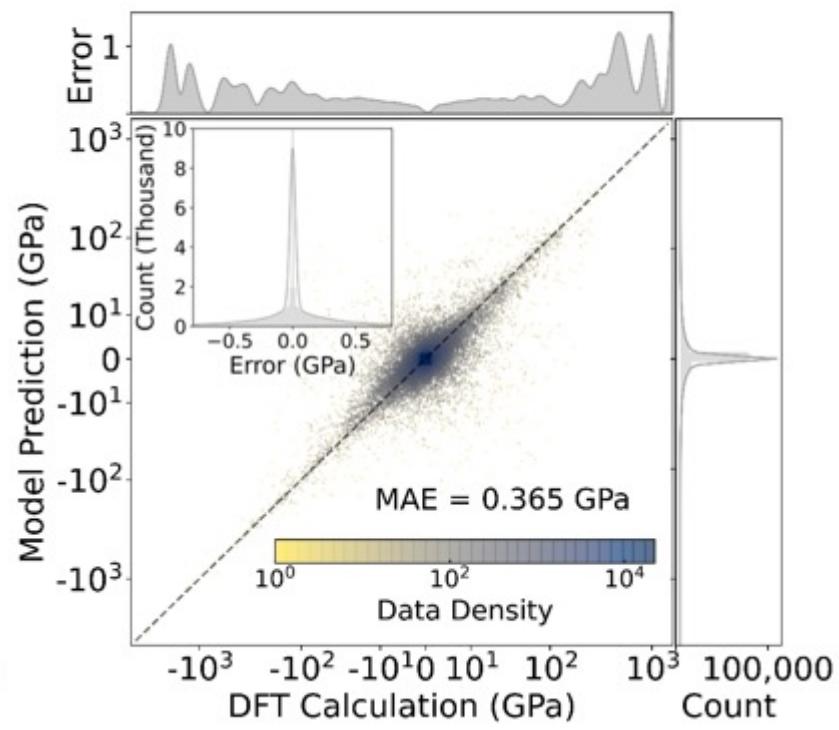
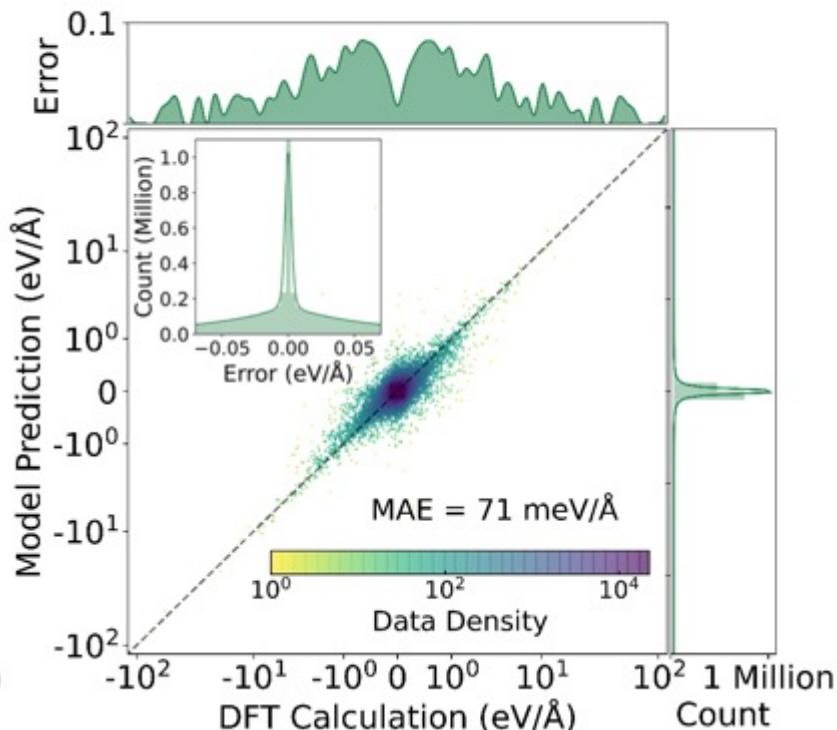
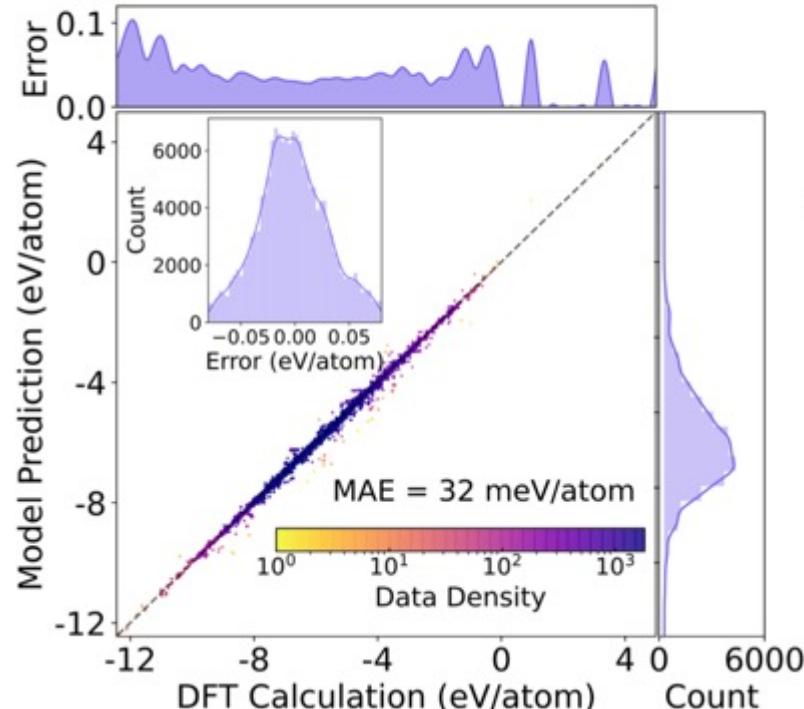
# 8 通用力场 (GPTFF)



## Graph-based Pretrained Transformer Force Field (GPTFF)



# 8 通用力场 (GPTFF)

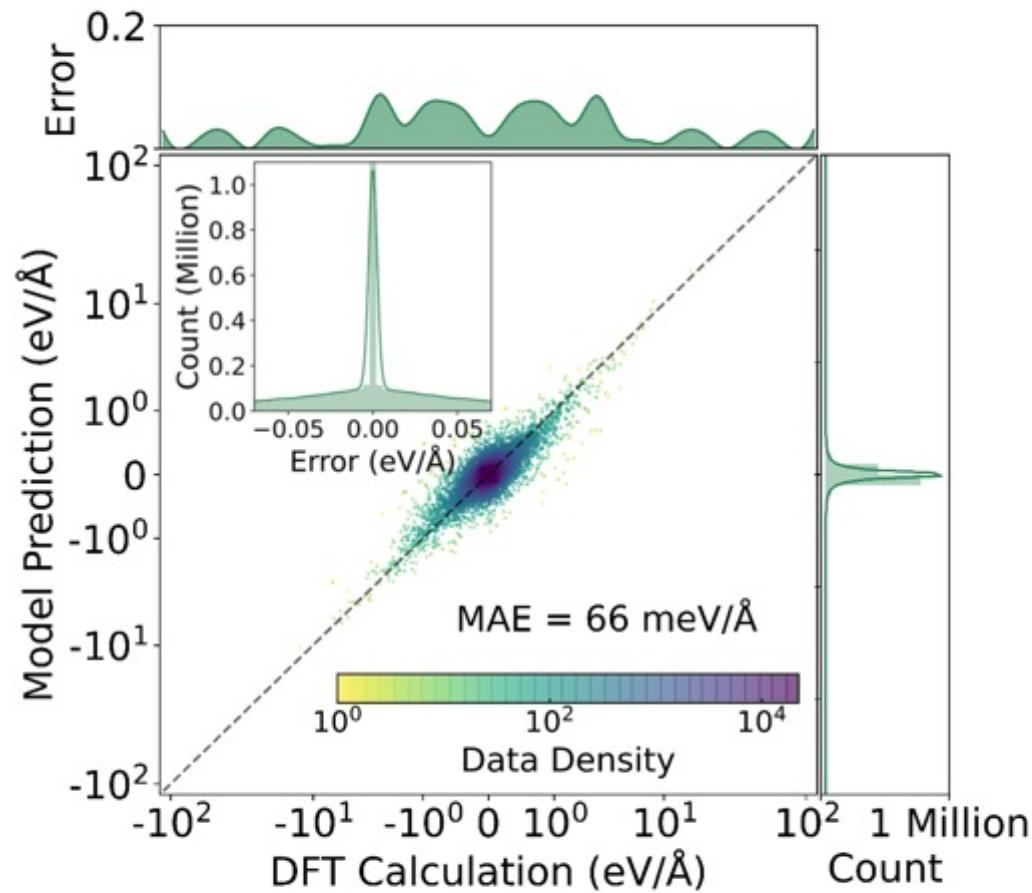


模型	能量误差(meV/atom)	训练数据(Million)
GPTFF	32	37.8
M3GNet	35	0.18
CHGNet	33	1.5

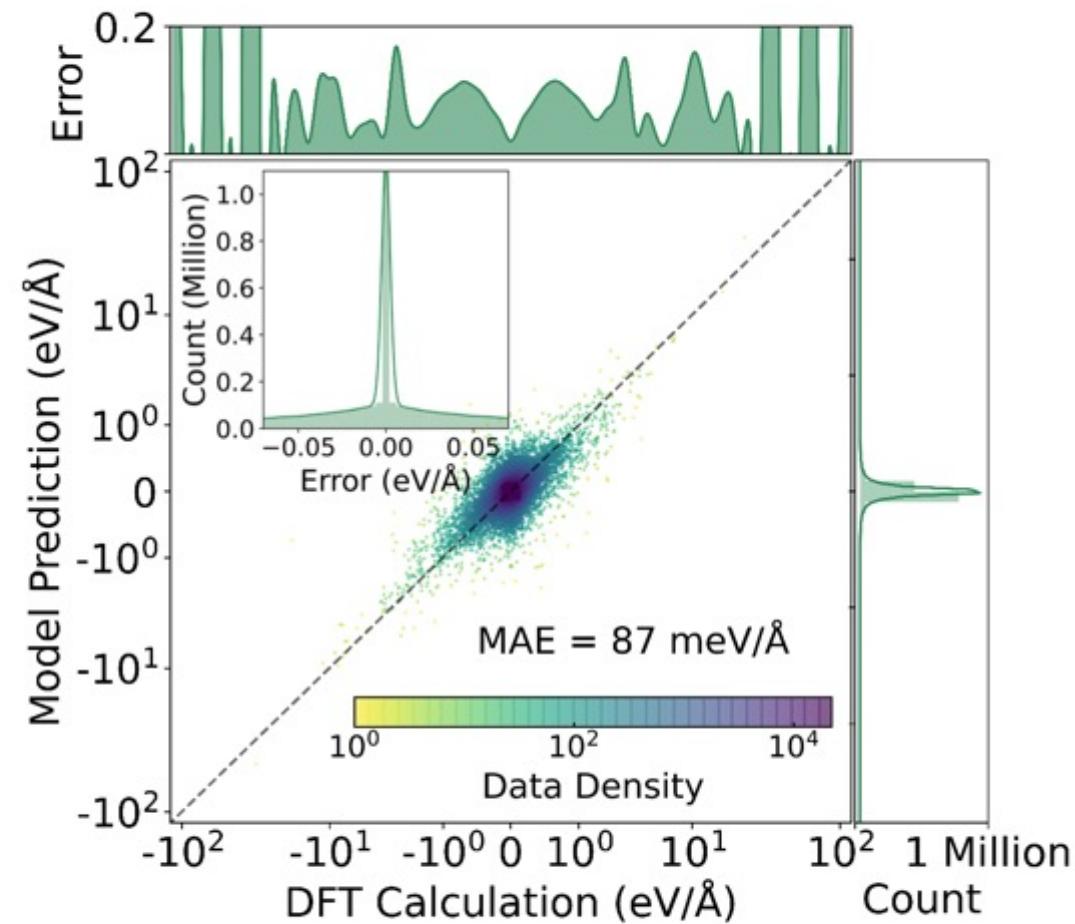
# 8 通用力场 (GPTFF)



## 性能比较



GPTFF 预测结果

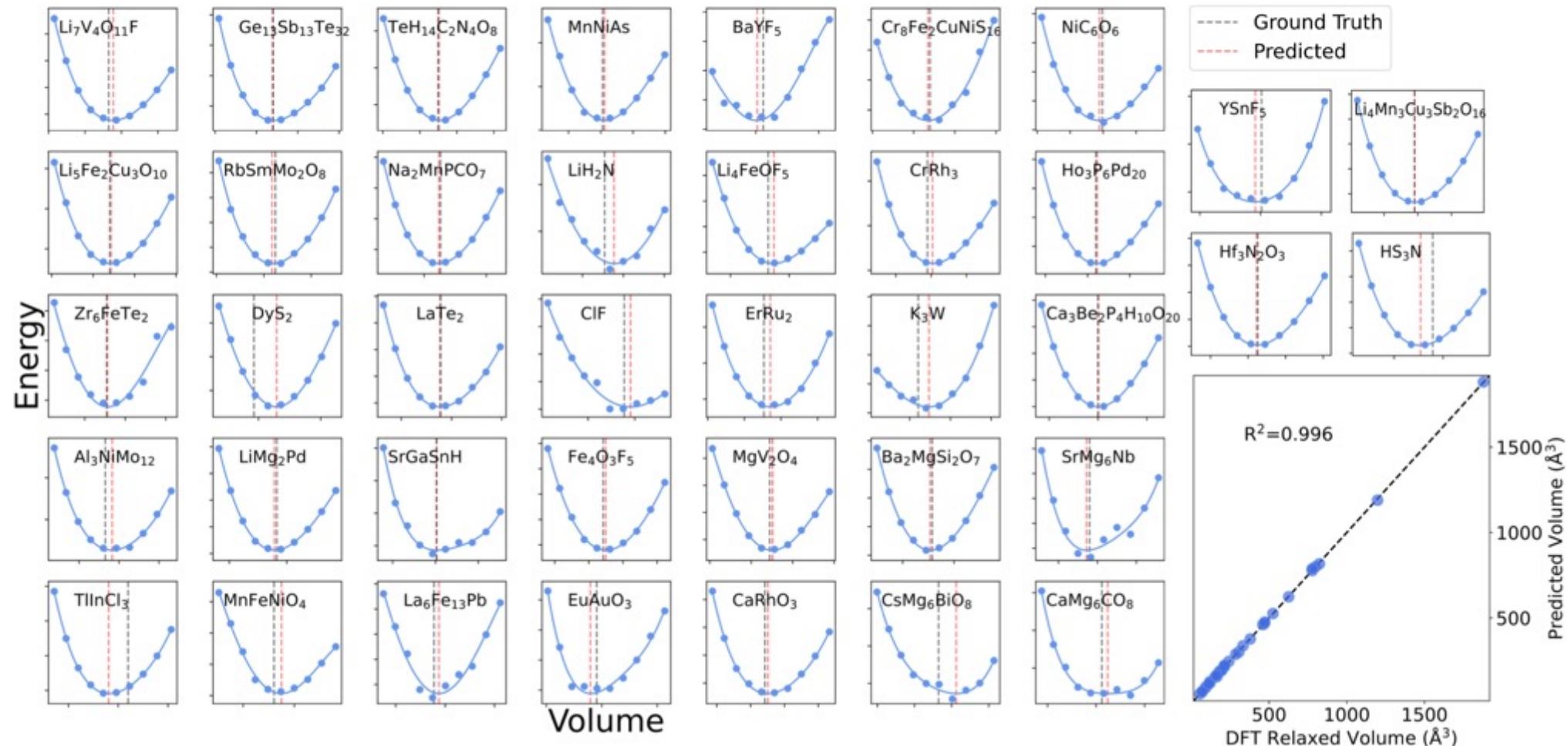


CHGNet 预测结果

# 8 通用力场 (GPTFF)



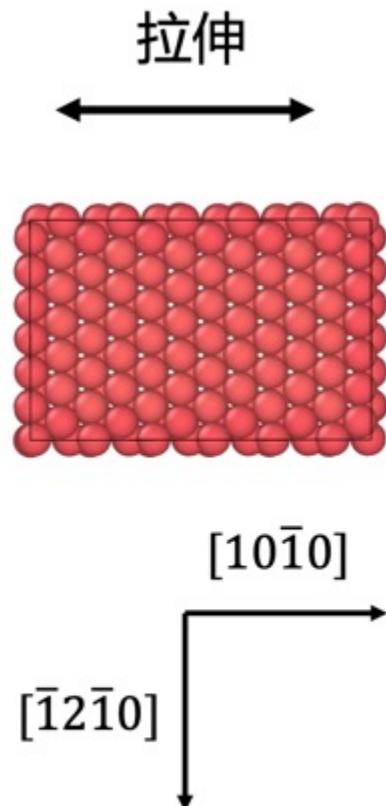
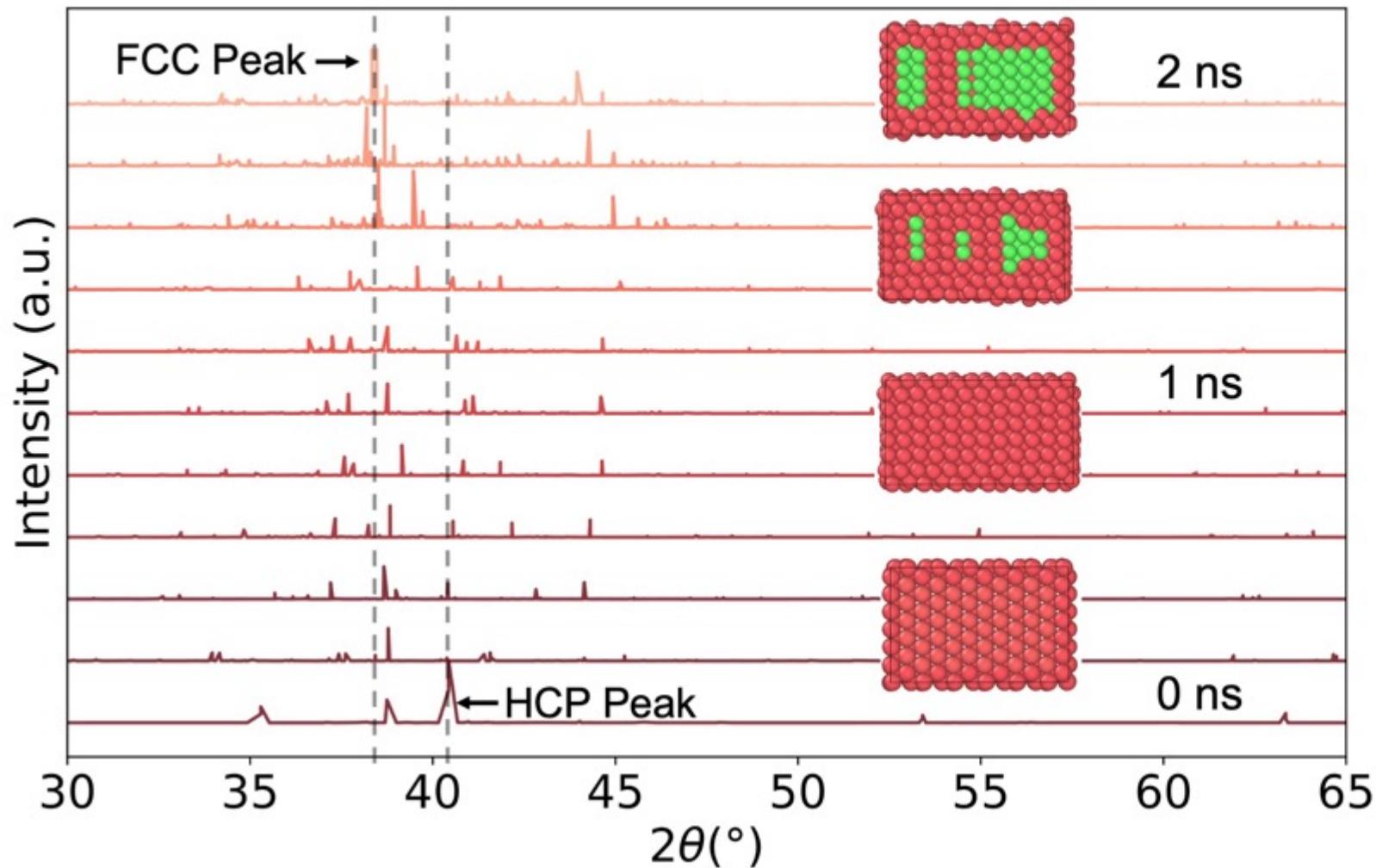
## 物态方程



# 8 通用力场 (GPTFF)



## Ti体系的HCP→FCC相变

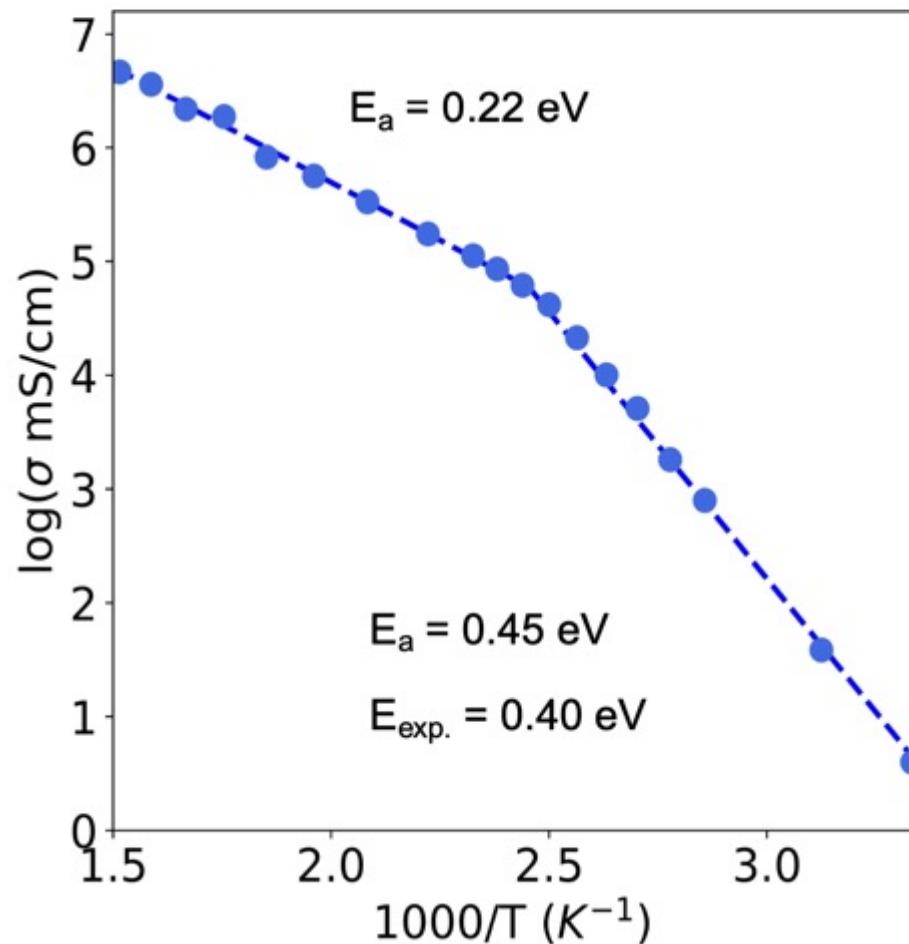


# 8 通用力场 (GPTFF)



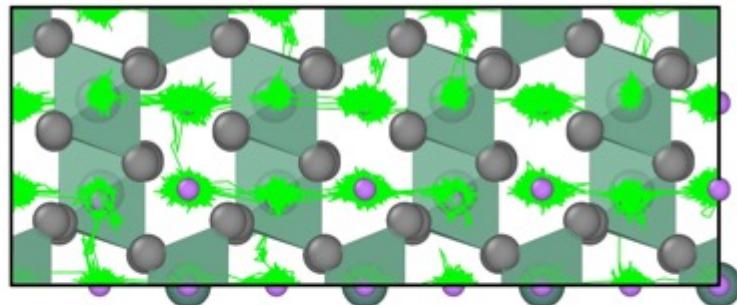
## $\text{Li}_3\text{YCl}_6$ 离子电导率

阿伦尼乌斯关系

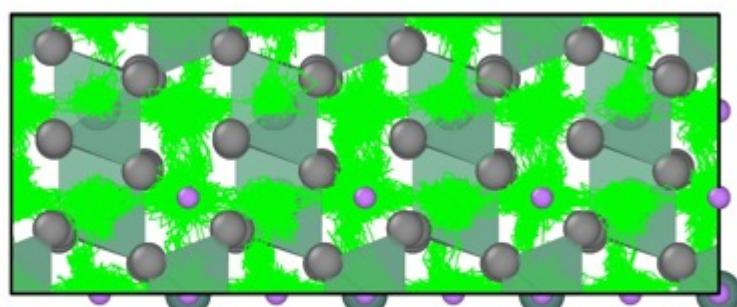


$\text{Li}^+$ 扩散轨迹

300K下的扩散轨迹



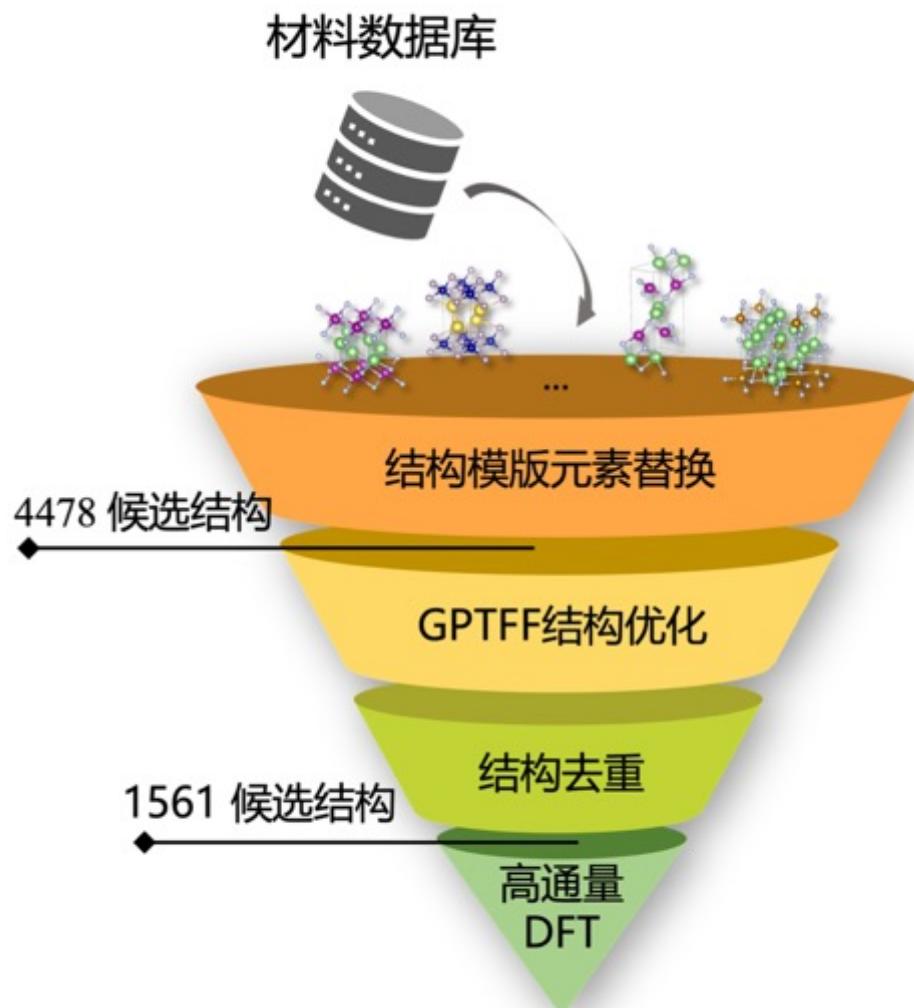
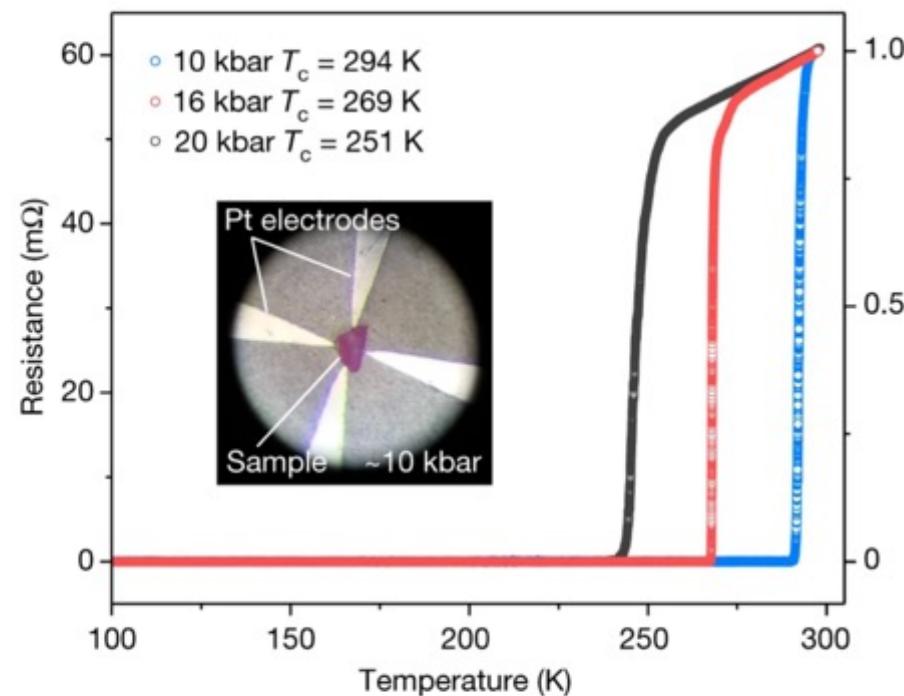
500K下的扩散轨迹



$$\begin{aligned}\sigma_{300K} &= 0.6 \text{ mS/cm} \\ \sigma_{\text{exp.}} &= 0.51 \text{ mS/cm}\end{aligned}$$

# 8 通用力场 (GPTFF)

## Lu-H-N相图的快速构建



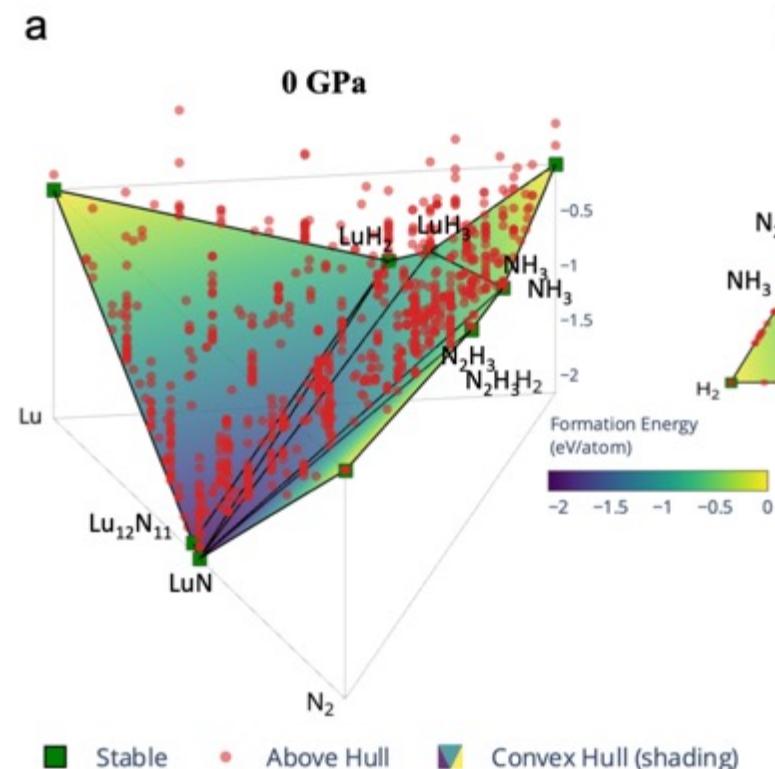
RETRACTED ARTICLE: Nature, 2023, 615(7951): 244-250.

# 8 通用力场 (GPTFF)

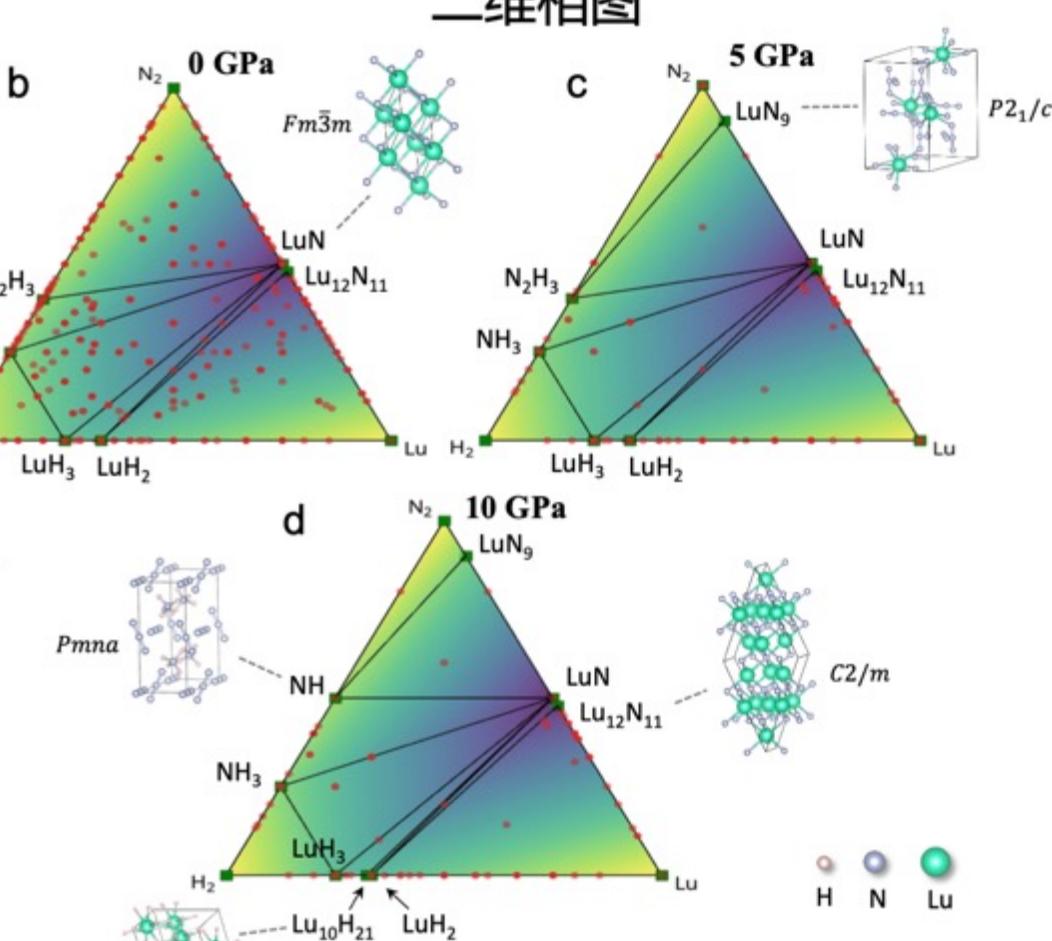


## Lu-H-N相图的快速构建

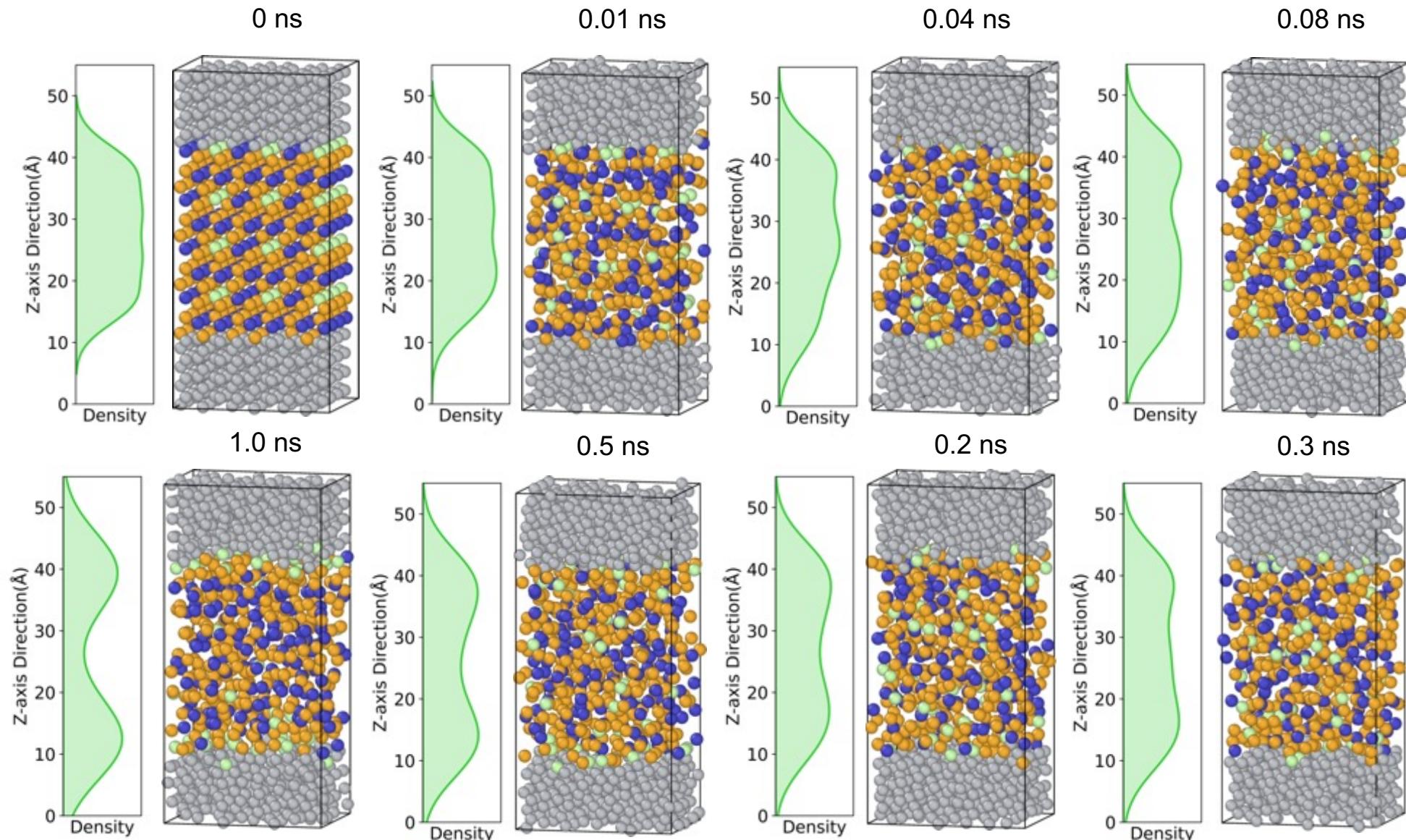
三维相图



二维相图



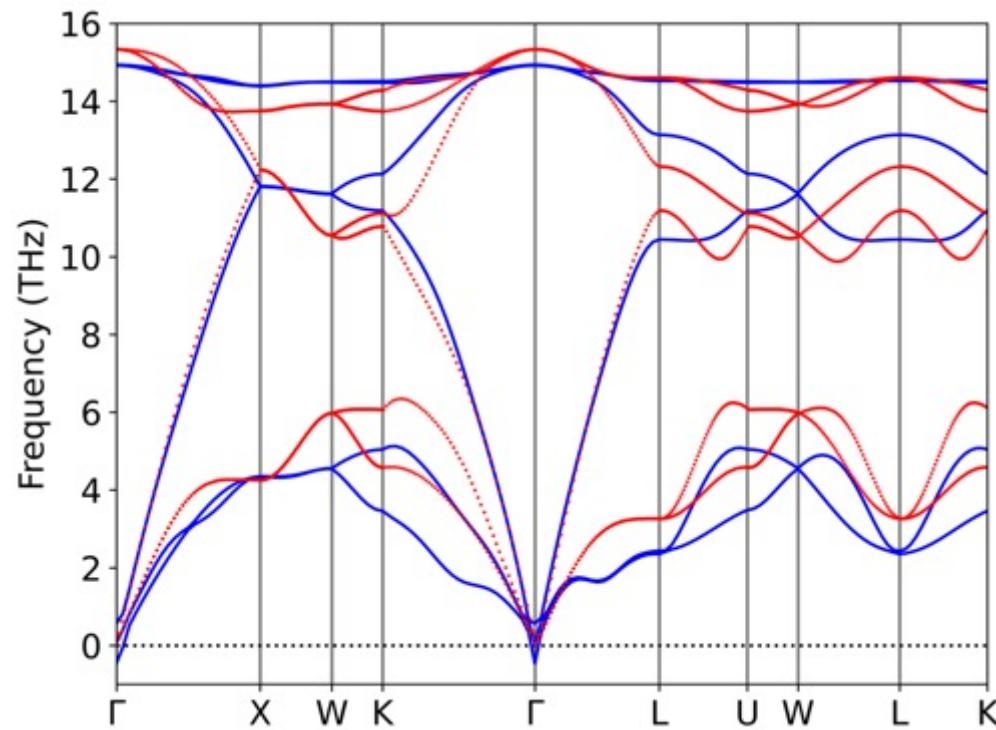
# 8 通用力场 (GPTFF)



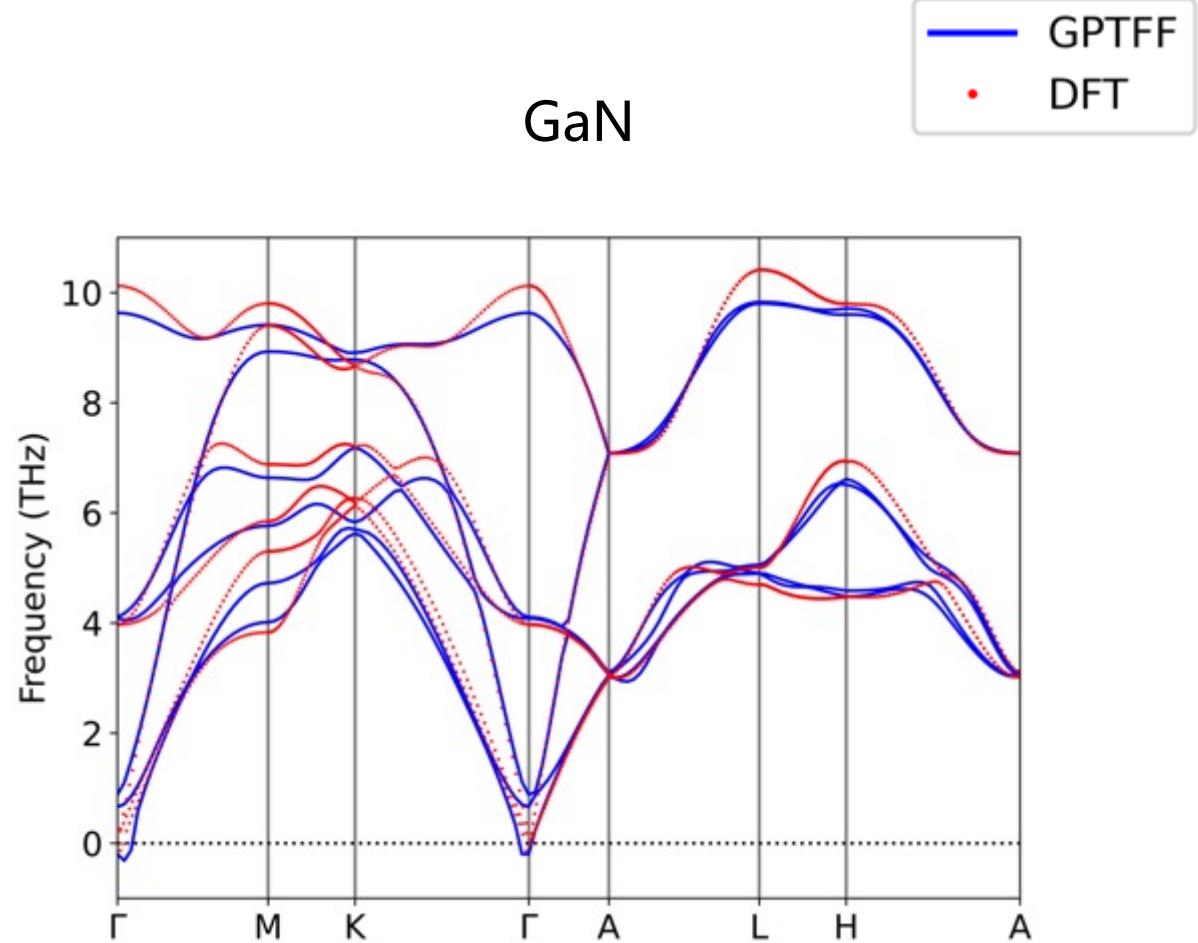
# 8 通用力场 (GPTFF)



Si



GaN



# 8 通用力场 (GPTFF)



## 结构优化

<https://github.com/atomly-materials-research-lab/GPTFF>

```
from gptff.model.mpredict import ASECalculator
from pymatgen.core import Structure
from pymatgen.io.ase import AseAtomsAdaptor
from ase.optimize.fire import FIRE
from ase.constraints import ExpCellFilter, StrainFilter

model_weight = "pretrained/gptff_v1.pth"
device = 'cuda' # or cpu
p = ASECalculator(model_weight, device) # Initialize the model and load weights

struc = Structure.from_file('POSCAR_structure') # Read structure

adp = AseAtomsAdaptor()
atoms = adp.get_atoms(struc)
atoms.set_calculator(p)

optimizer = ExpCellFilter(atoms)

FIRE(optimizer).run(fmax=0.01, steps=100)
```

# 8 通用力场 (GPTFF)



MD

<https://github.com/atomly-materials-research-lab/GPTFF>

```
from gptff.model.mpredict import ASECalculator
from pymatgen.core import Structure
from pymatgen.io.ase import AseAtomsAdaptor
from ase import Atoms, units
from ase.md.nvtberendsen import NVTBerendsen
import os

model_weight = "pretrained/gptff_v1.pth"
device = 'cuda' # or cpu
p = ASECalculator(model_weight, device) # Initialize the model and load weights

struc = Structure.from_file('POSCAR_structure') # Read structure

adp = AseAtomsAdaptor()
atoms = adp.get_atoms(struc)
atoms.set_calculator(p)

save_dir = './results_path'
os.makedirs(save_dir, exist_ok=True)

temp = 430 # unit (K)
dyn = NVTBerendsen(atoms=atoms,
                    timestep=2 * units.fs,
                    temperature=temp, # unit (K)
                    taut=200*units.fs,
                    loginterval=20, # Save md information and trajectory every 20 steps
                    logfile=os.path.join(save_dir, f'output.txt'), # Information printer
                    trajectory=os.path.join(save_dir, f'Li3PO4_nvt_out_{temp}K.trj'), # Trajectory file
                    append_trajectory=True)

dyn.run(100000)
```

**GPTFF**

- 开源
- 开箱即用
- CPU or GPU
- pip install
- finetune
- ASE

# 9 The race



	Model	F1 ↑	DAF ↑	Prec ↑	Acc ↑	TPR ↑	TNR ↑	MAE ↓	RMSE ↓	R² ↑	Training
Meta	eqV2	0.917	6.047	0.924	0.975	0.910	0.986	0.020	0.072	0.848	3M (102.4)
Orbital materials	ORB	0.880	6.041	0.924	0.965	0.841	0.987	0.028	0.077	0.824	3M (32.1M)
Microsoft	MatterSim	0.859	5.646	0.863	0.957	0.856	0.975	0.026	0.080	0.812	17M (Ma)
DeepMind	GNoME	0.829	5.523	0.844	0.955	0.814	0.972	0.035	0.085	0.785	6M (89.0M)
	eqV2 DeNS	0.815	5.042	0.771	0.941	0.864	0.953	0.036	0.085	0.788	146K (1.6)
	ORB MPtrj	0.765	4.702	0.719	0.922	0.817	0.941	0.045	0.091	0.756	146K (1.6)
	SevenNet	0.724	4.252	0.650	0.904	0.818	0.919	0.048	0.092	0.750	146K (1.6)
	MACE	0.669	3.777	0.577	0.878	0.796	0.893	0.057	0.101	0.697	146K (1.6)
	CHGNet	0.613	3.361	0.514	0.851	0.758	0.868	0.063	0.103	0.689	146K (1.6)
	M3GNet	0.569	2.882	0.441	0.813	0.803	0.813	0.075	0.118	0.585	63K (188.1)
	ALIGNN	0.567	3.206	0.490	0.841	0.672	0.872	0.093	0.154	0.297	155K (MI)
	MEGNet	0.510	2.959	0.452	0.826	0.585	0.870	0.130	0.206	-0.248	133K (MI)
	CGCNN	0.507	2.855	0.436	0.818	0.605	0.857	0.138	0.233	-0.603	155K (MI)
	CGCNN+P	0.500	2.563	0.392	0.786	0.693	0.803	0.113	0.182	0.019	155K (MI)
	Wrenformer	0.466	2.256	0.345	0.745	0.719	0.750	0.110	0.186	-0.018	155K (MI)

# 9 科学数据开放平台



 **OSDI** 科学数据开放平台  
Open Scientific Data Infrastructure (OSDI)

<https://openscidata.net> 已上线

- ❖ Atomly数据共享利器
- ❖ 保障数据生产者权益，保护数据资产
- ❖ 数据共享，AI模型共享
- ❖ 面向全世界开放

隐私计算技术、数据可用不可见





SONGSHAN LAKE  
MATERIALS LABORATORY

松山湖材料实验室

谢谢大家， 请您批评指正！