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

Miao Liu, Sheng Meng

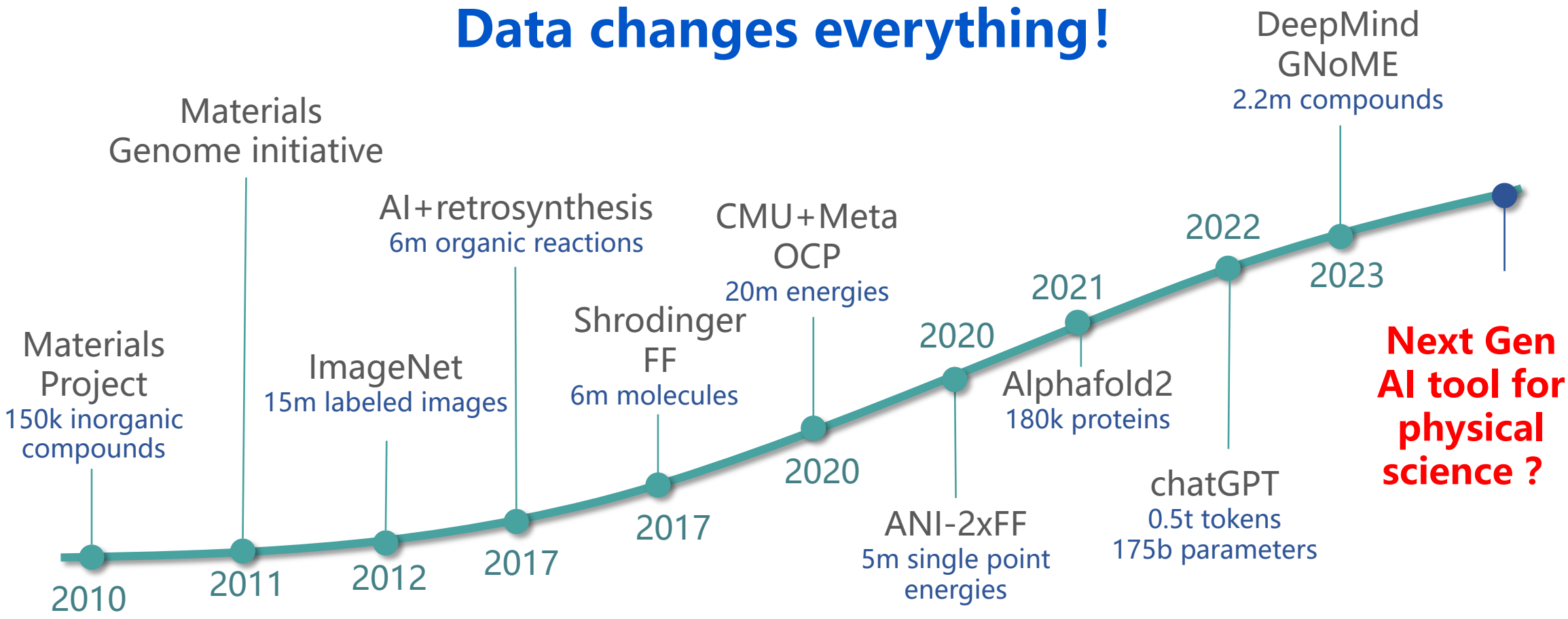
中国科学院物理研究所

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!



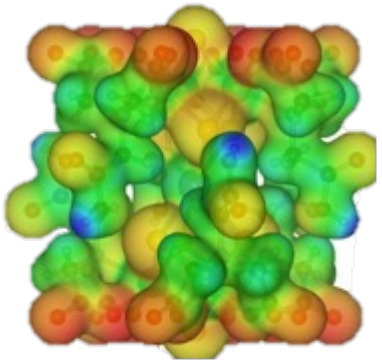
1 计算材料科学的技术路线

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

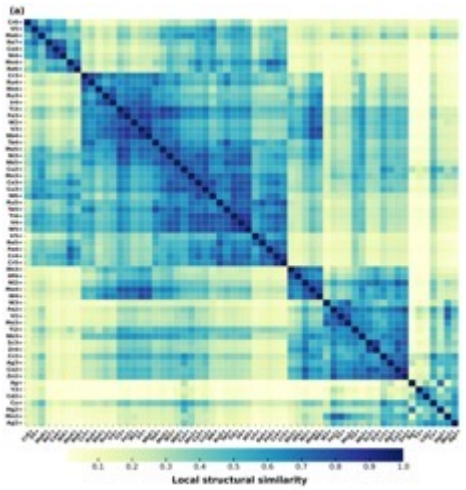
DFT Theory

$$\hat{H}\left|\Psi\right\rangle = E_{gs}\left|\Psi\right\rangle$$

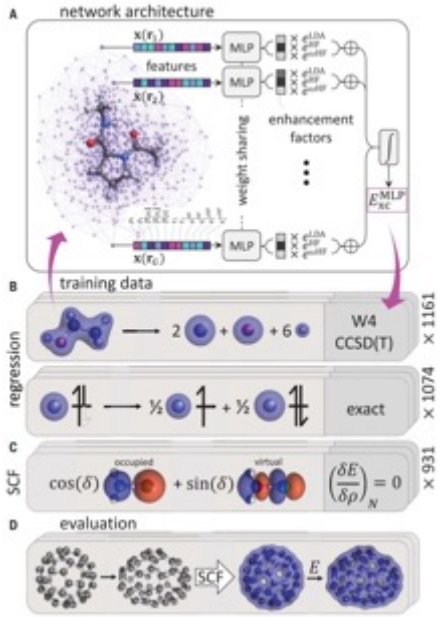
Numerical solution



Data



AI tool



(1965) 2th paradigm 3th paradigm 4th paradigm

Electronic-Structure Theory & Materials Genomics

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

OQMD
美国 西北大学 (2012)

AFLOW
Automatic - FLOW for Materials Discovery
美国 Duke大学 (2012)

NOMAD
德国 Fritz-Haber-Institut (2015)

NIST
美国 国家标准局 (2016)

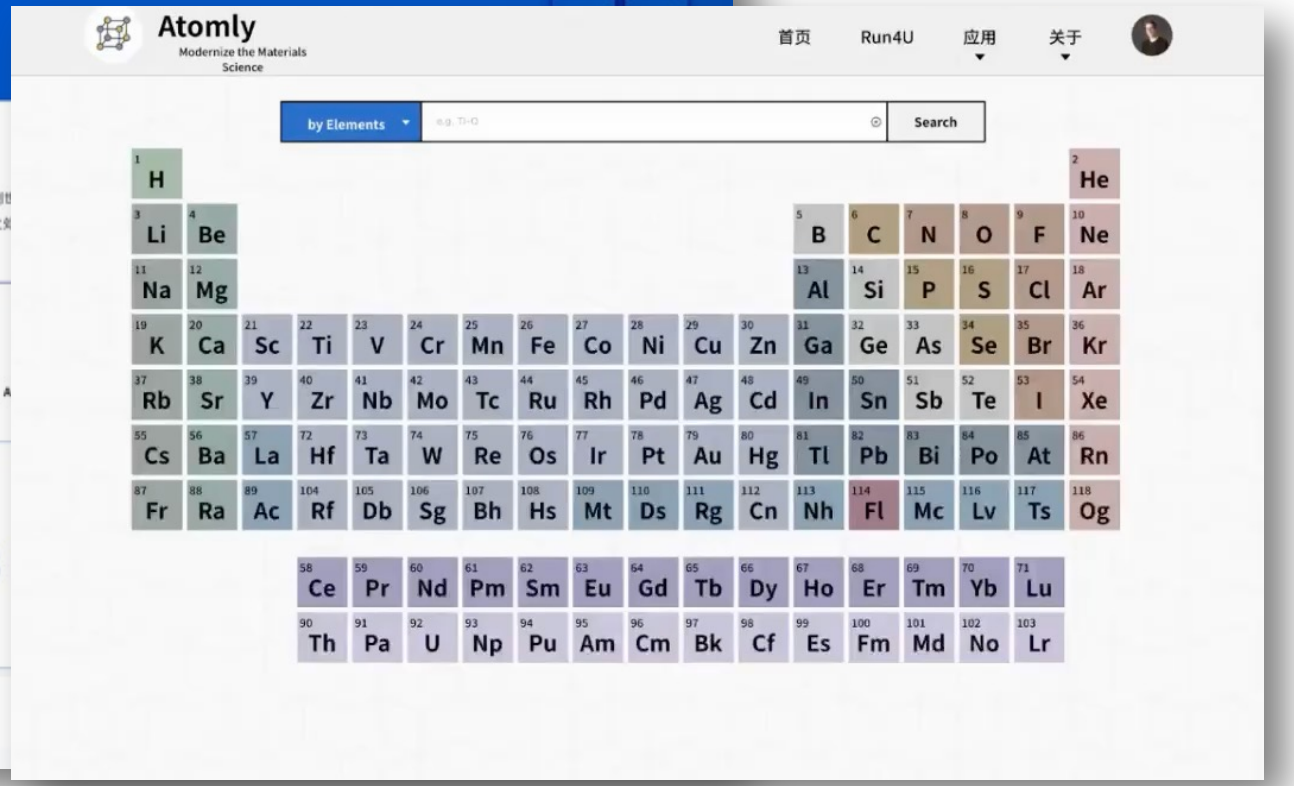
AiiDA MATERIALS CLOUD
瑞士 NCCR (2015)

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材料科学数据库

https://atomly.net

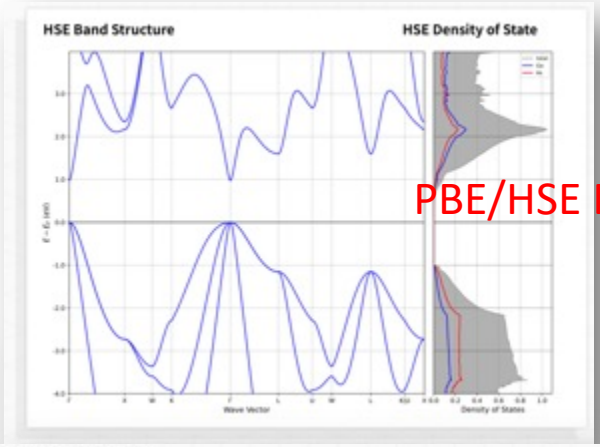


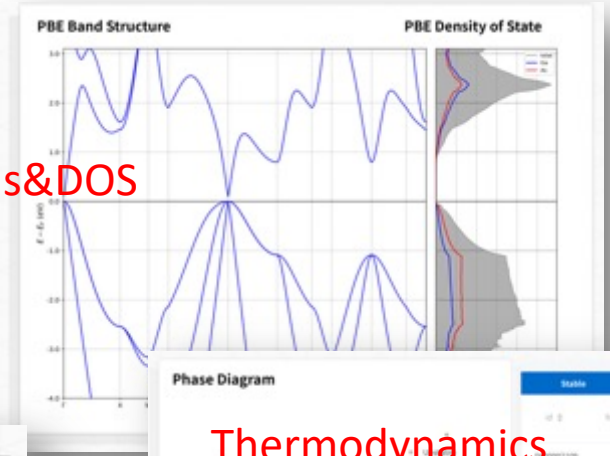
- **自主开发、自主计算**
- **我国首个世界级材料计算数据库**
- **数据质量、数量达标世界级数据库**
- **对我国用户全免费开放**

- 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

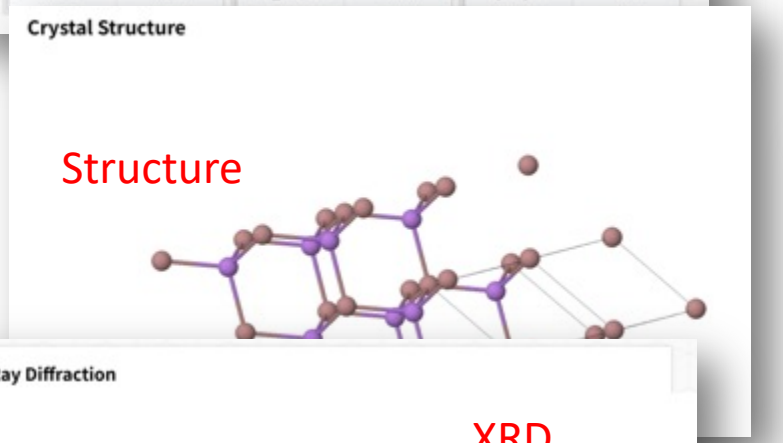
Atomly材料科学数据库 （Demo） https://atomly.net

Basic info

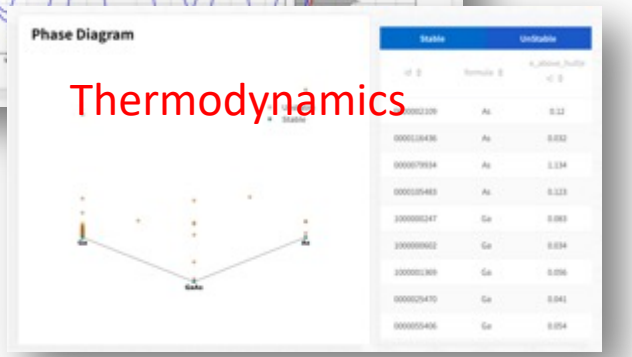




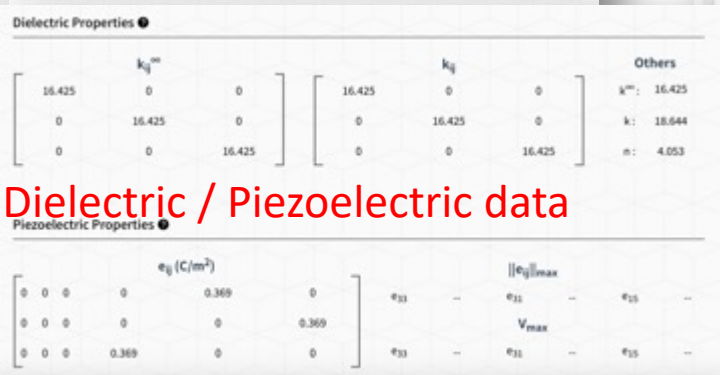
PBE/HSE Bands&DOS



Structure

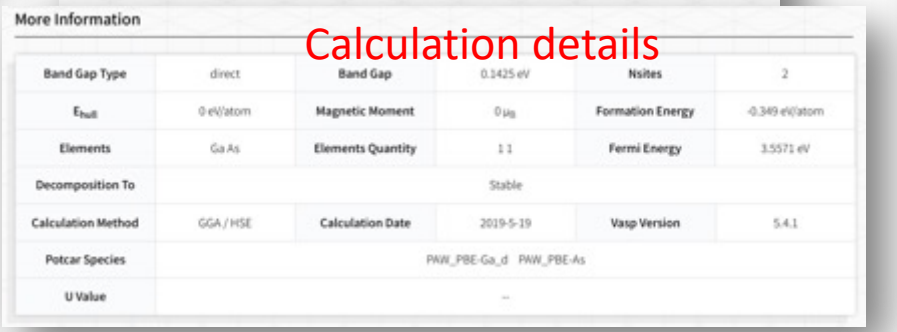


Thermodynamics

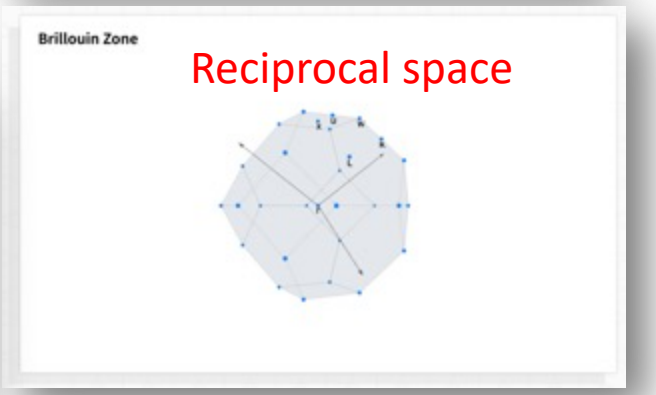


Dielectric / Piezoelectric data

XRD



Calculation details



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

=2 x MP

=4 x MP

=3 x MP

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)

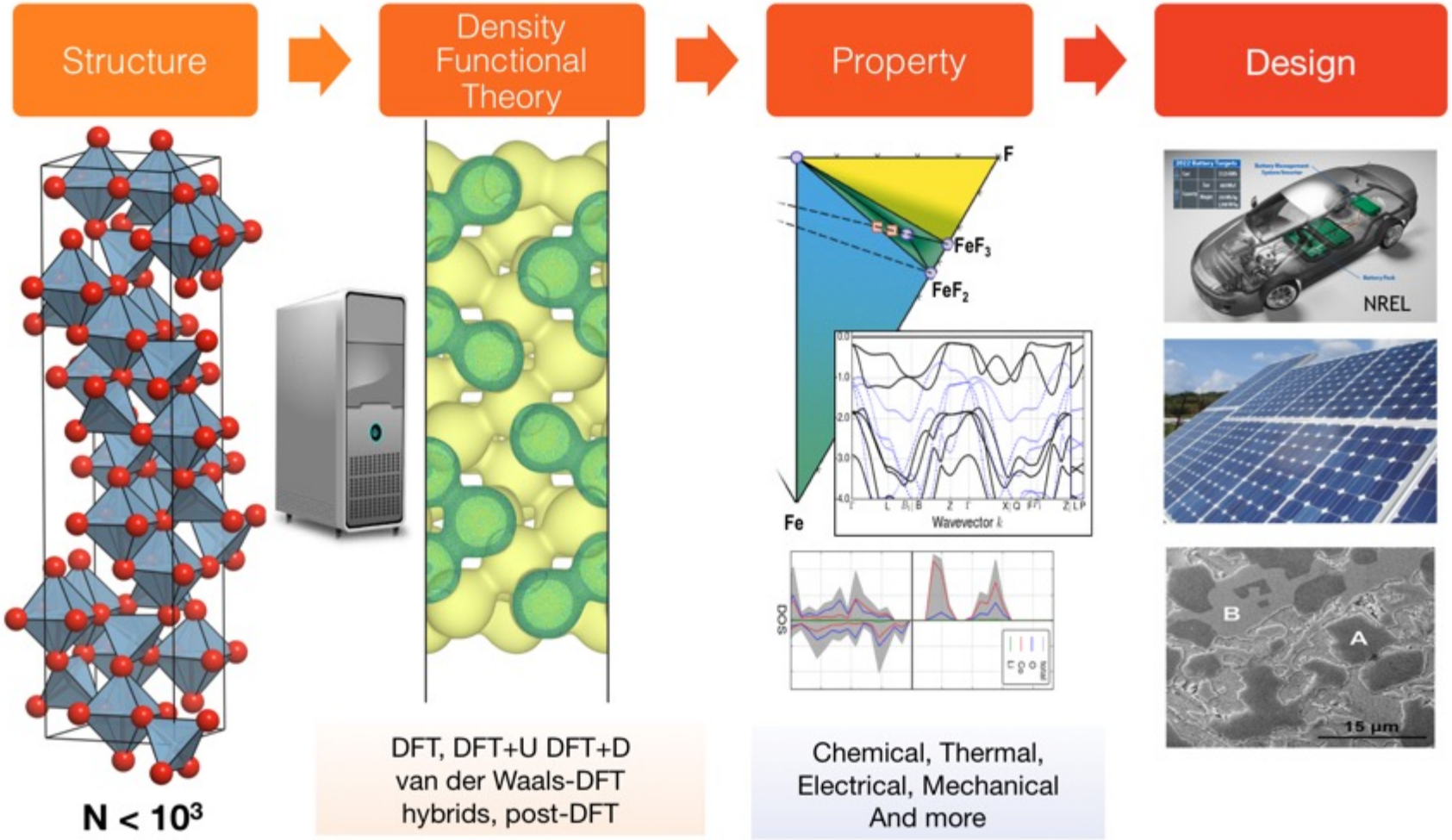
- ✓ 实验料数据集 (2025)

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

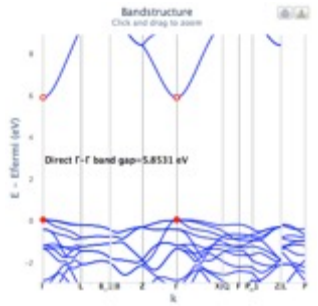
3 DFT计算



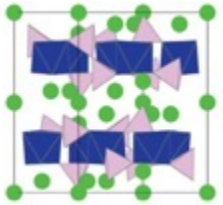
原子结构
电子结构

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



Atomic positions

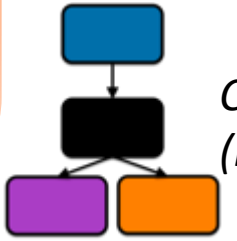


- 自主知识产权
- DFT计算全自动化
- 流程高度标准化
- 结果高度标准化
- 10^3级别高通量计算
- 前瞻性的数据库架构

Materials Data

Submitted Materials

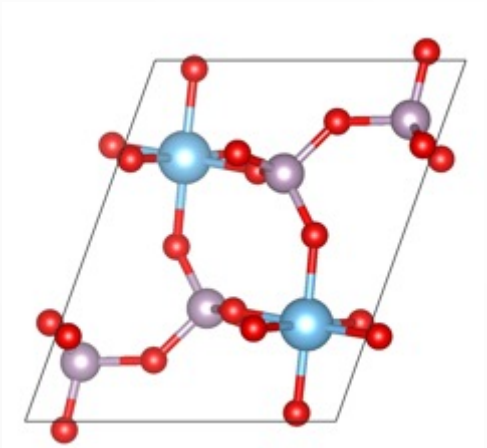
Calculation Workflows



Codes to run (in sequence)

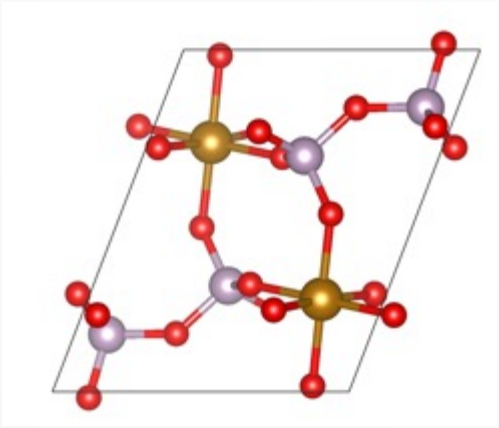
Supercomputers

3 计算未知材料（元素替换产生新结构）

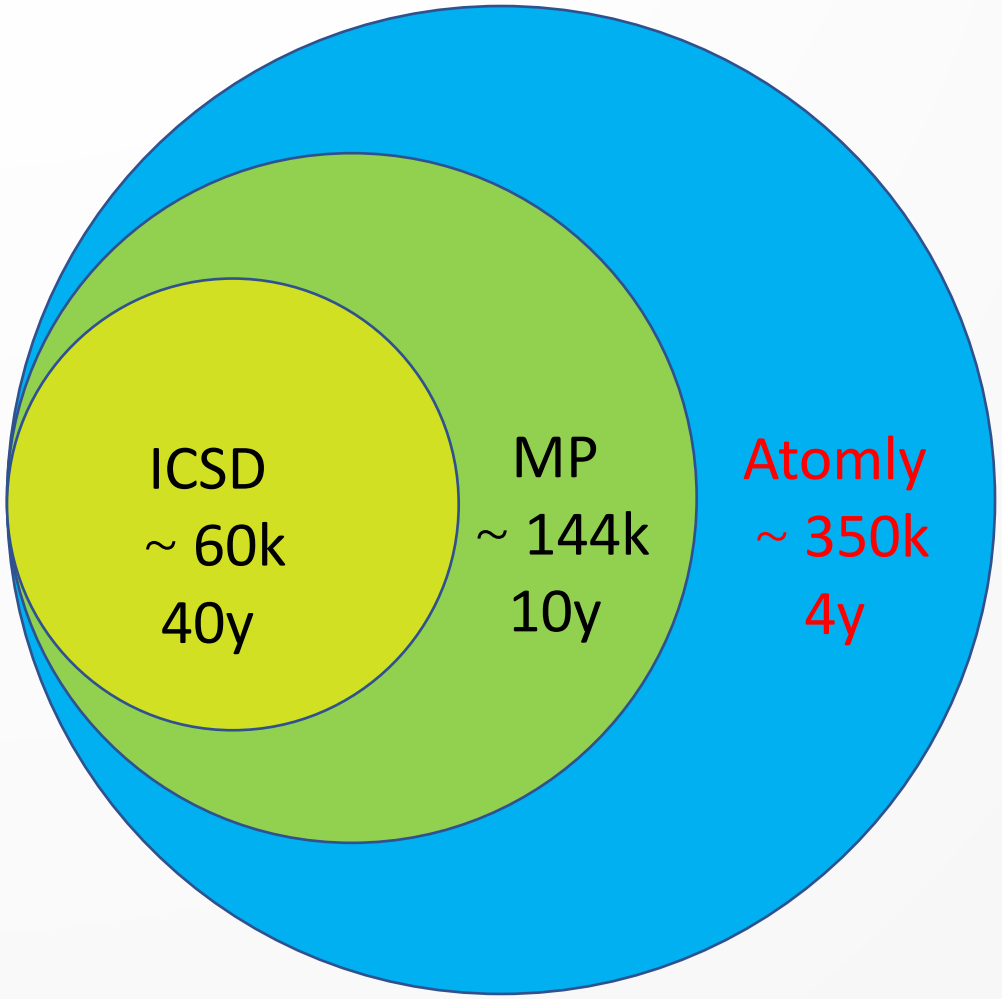


FeP_2O_7

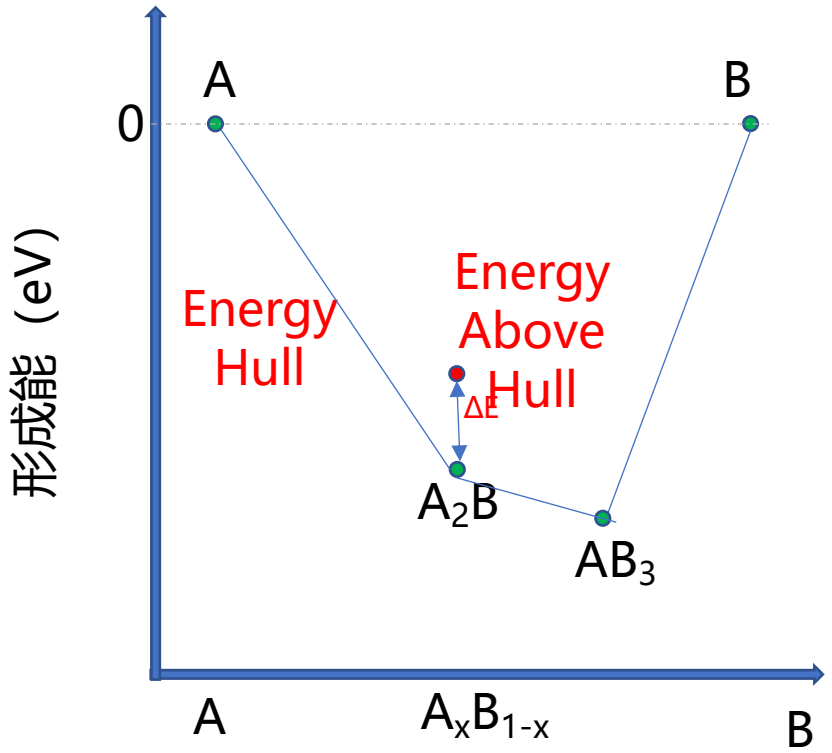
元素替换

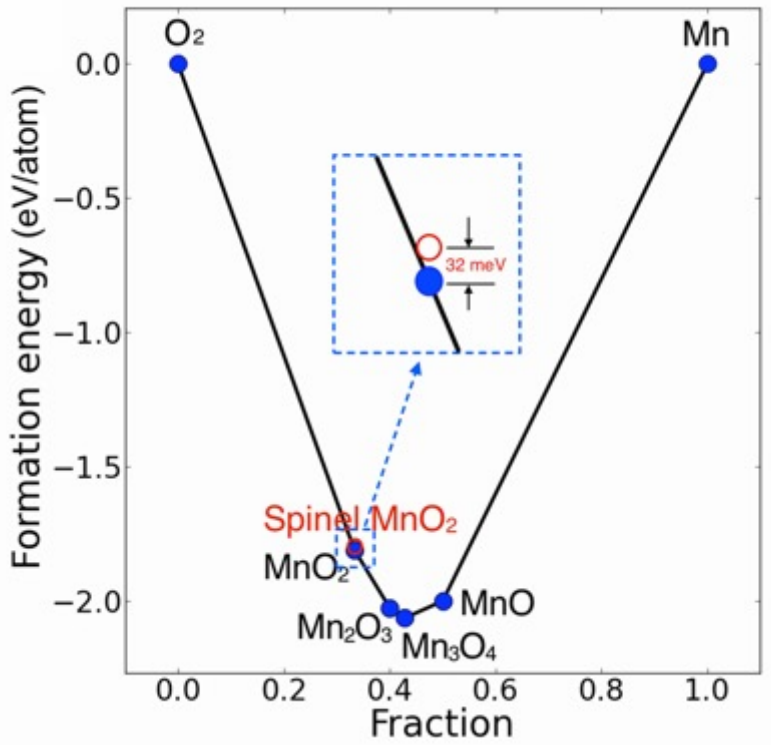


TiP_2O_7



3 探索未知相空间（热力学稳定相）





在50k新材料中

E^hull < 100meV /atom: 30540 compounds

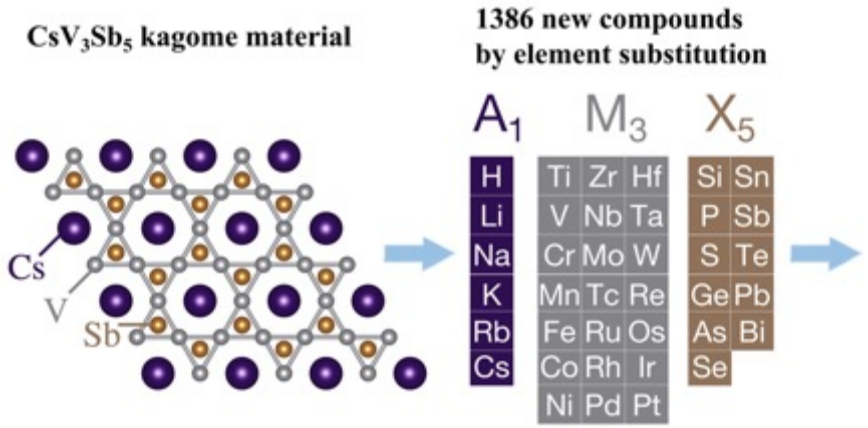
E^hull < 50meV /atom: 16511 compounds

E^hull < 10meV/atom : 6940 compounds

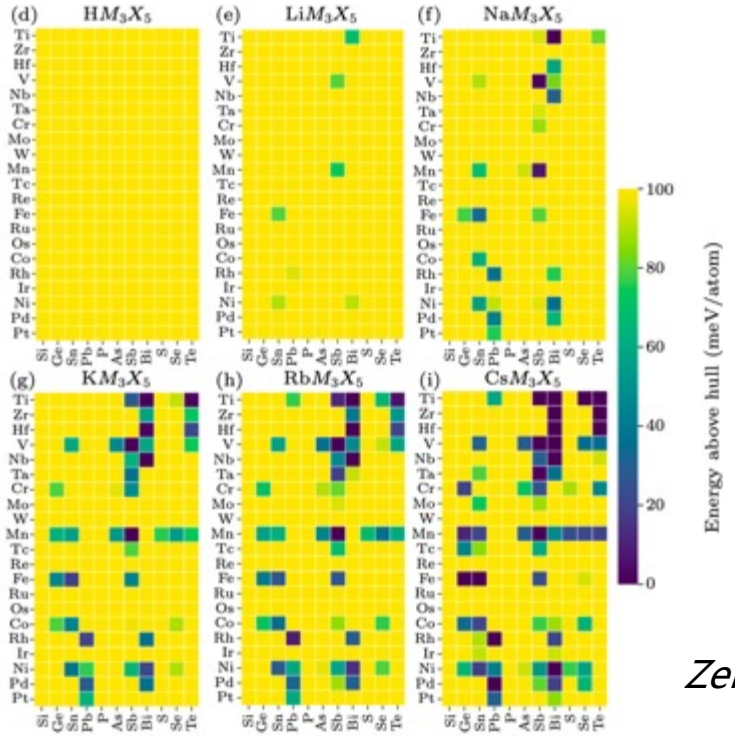
存在有大量未知稳定相

4 拓扑材料发现

扩展化学空间



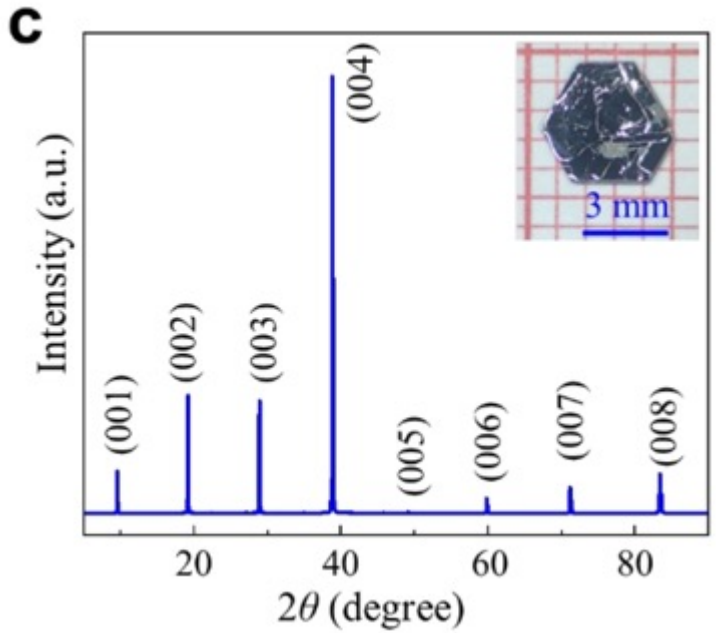
高通量计算



CPL, 39, 047402 (2022) Express letters

成功预言新材料，加速材料发现过程

实验验证



CPB, 31, 097405 (2022)

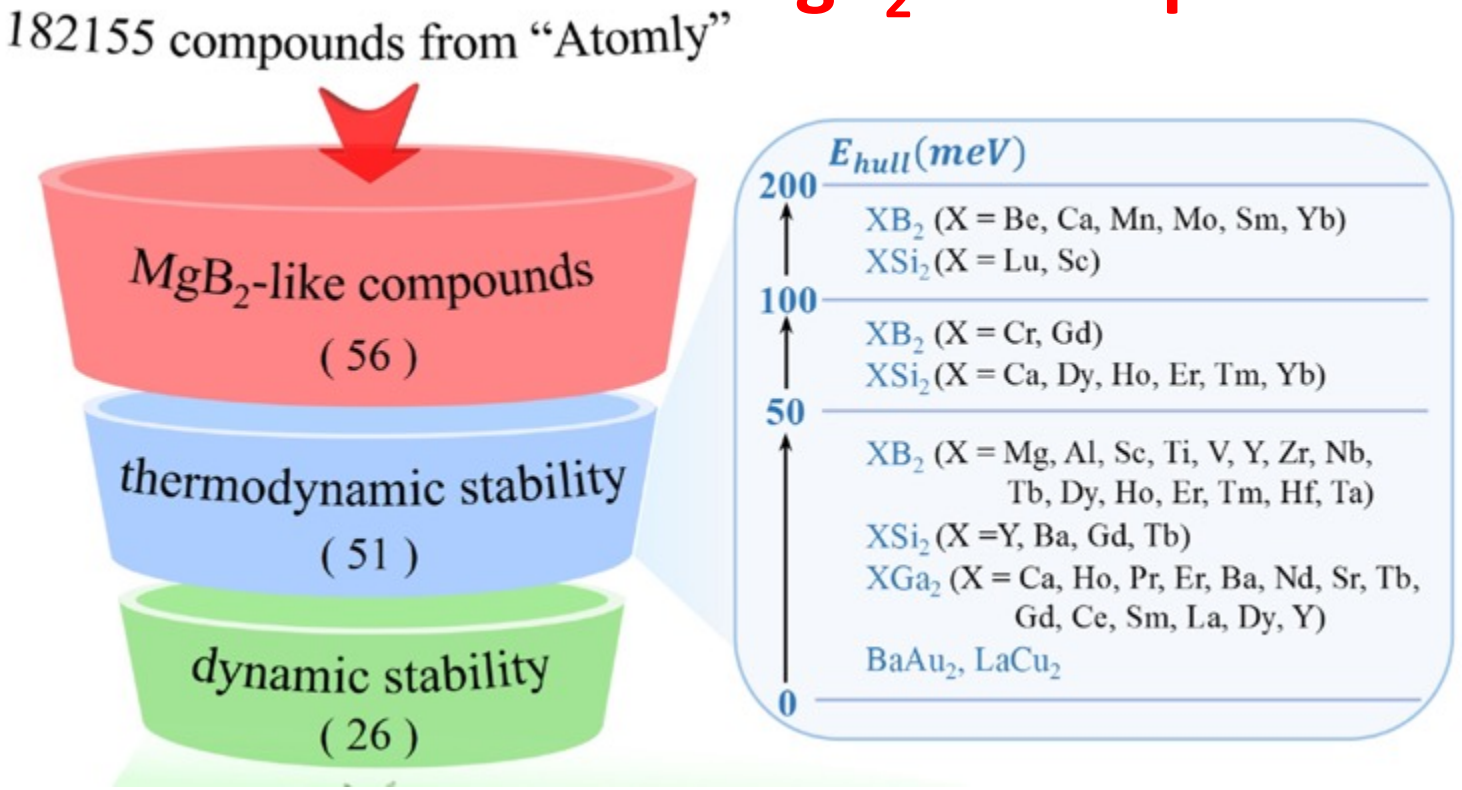
Zeitschrift für Naturforschung B, 77, 11(2022)

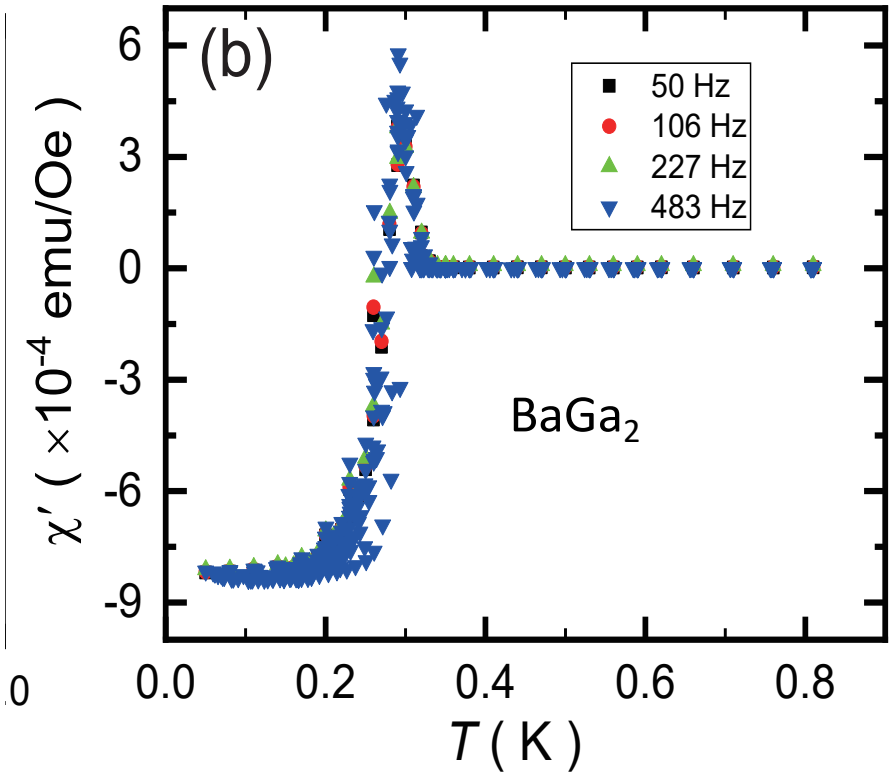
PRL, 131, 026701 (2023)

Nature Comm, 14, 4089 (2023)

Nature Phys (2023)

MgB_2-like superconductors

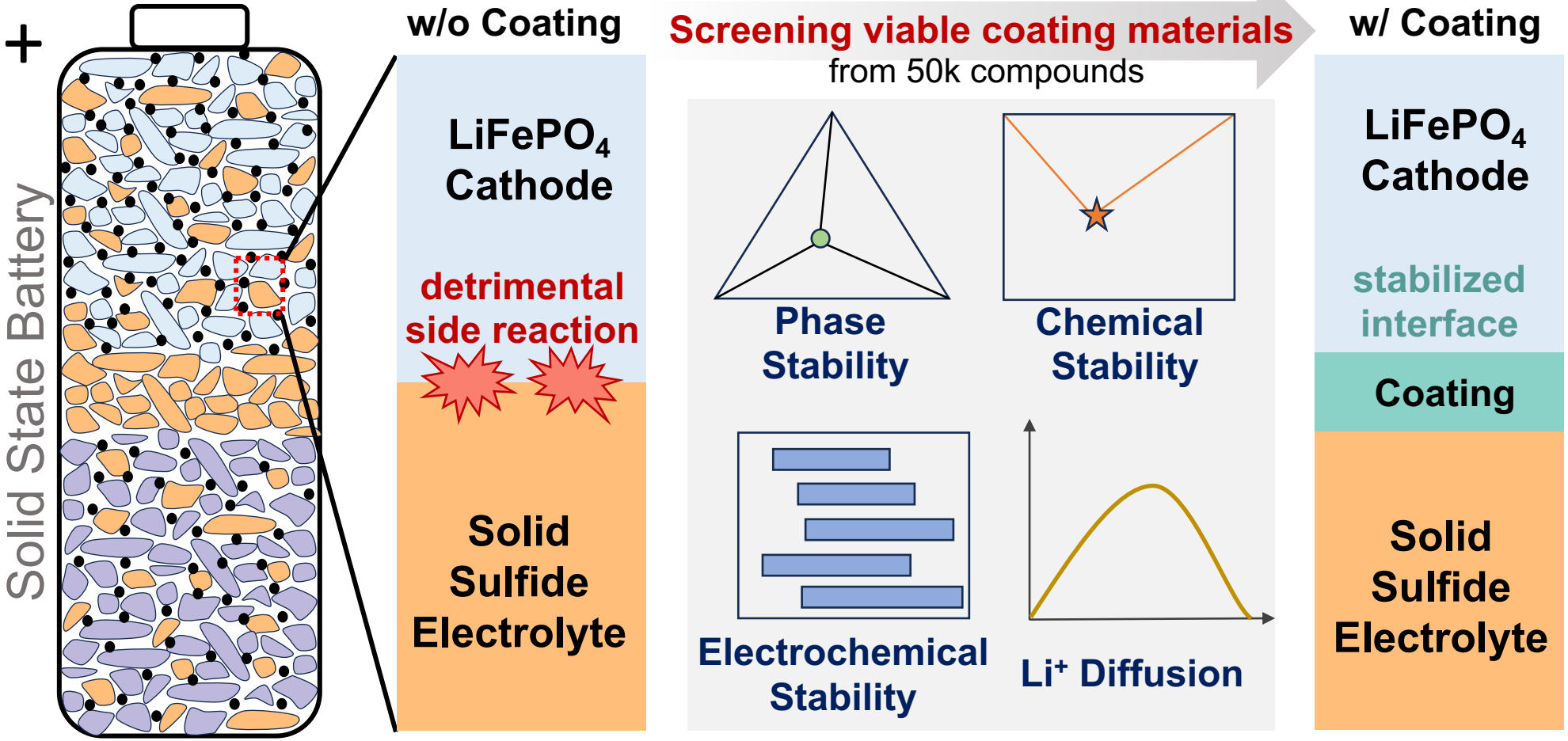




End-to-end screening the targeted compound in a once and for all fashion

Miao Liu, **PRB**, 105, 214517 (2022)

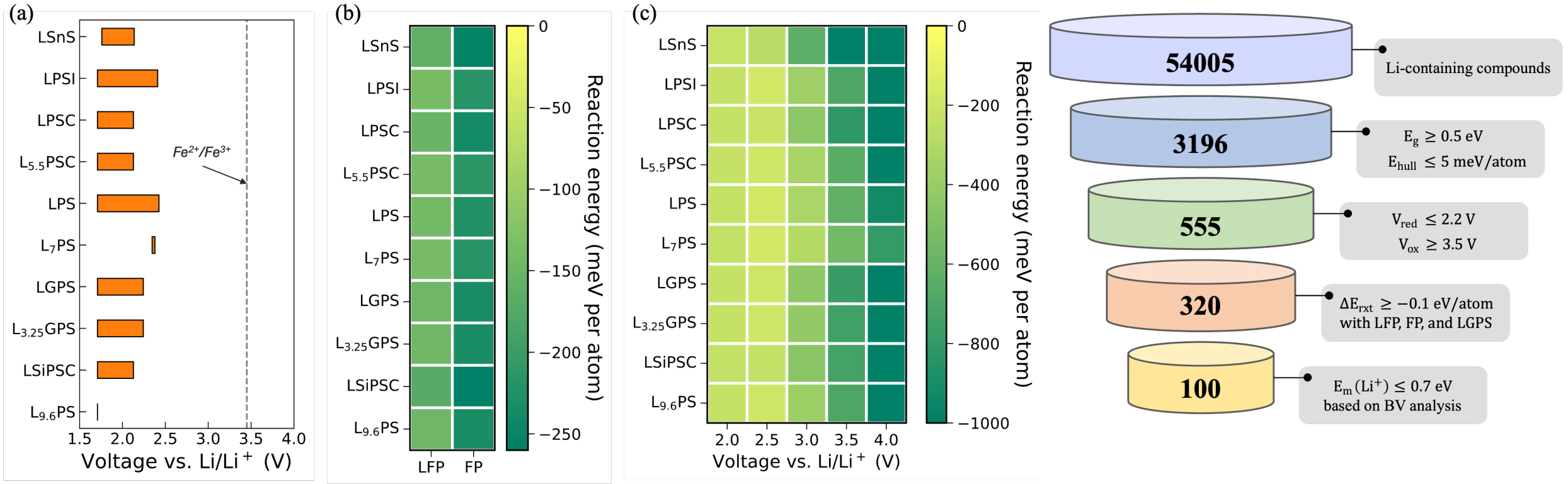
固态电池材料界面



J. Mater. Chem. A, 2024,**12**, 3954

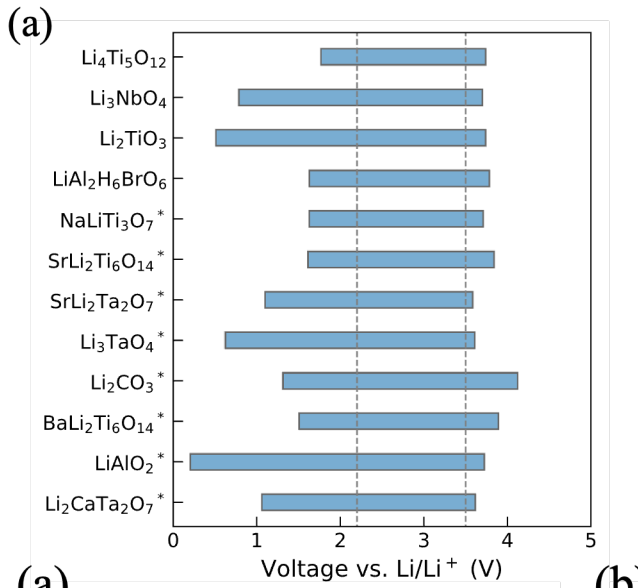
固态电池材料界面

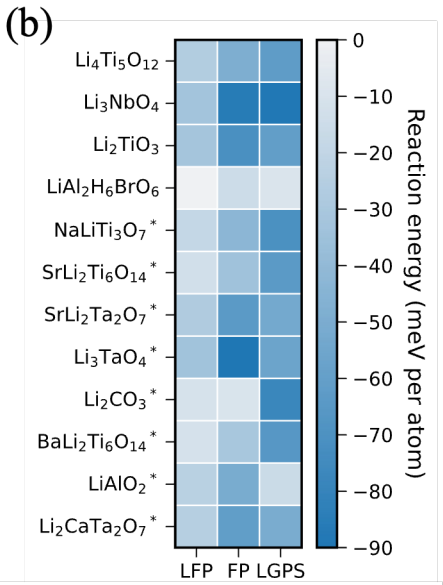
LFPO vs 硫化物固态电解质

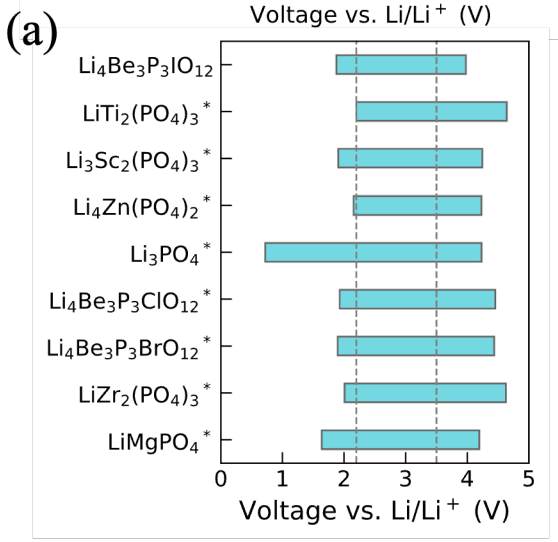


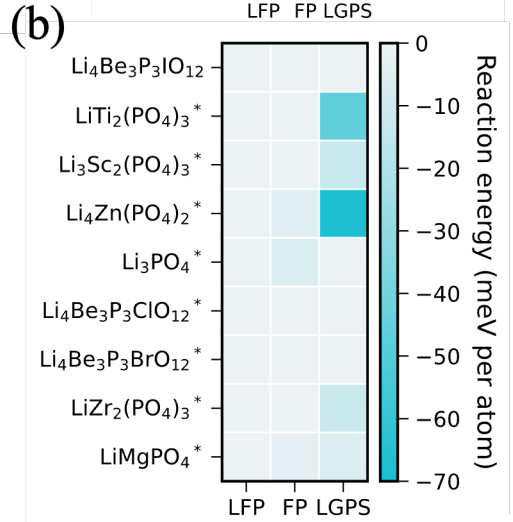
J. Mater. Chem. A, 2024,**12**, 3954

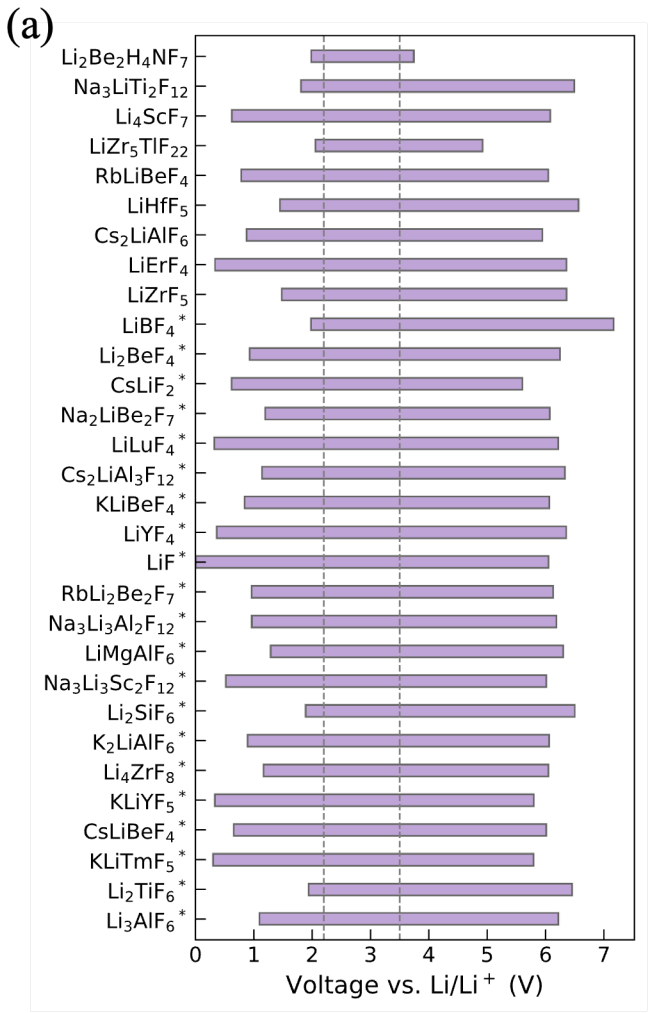
固态电池材料界面

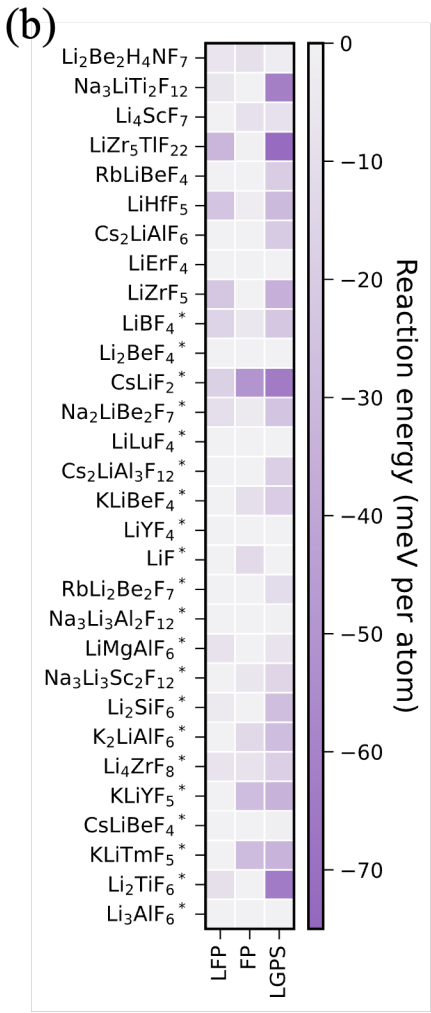






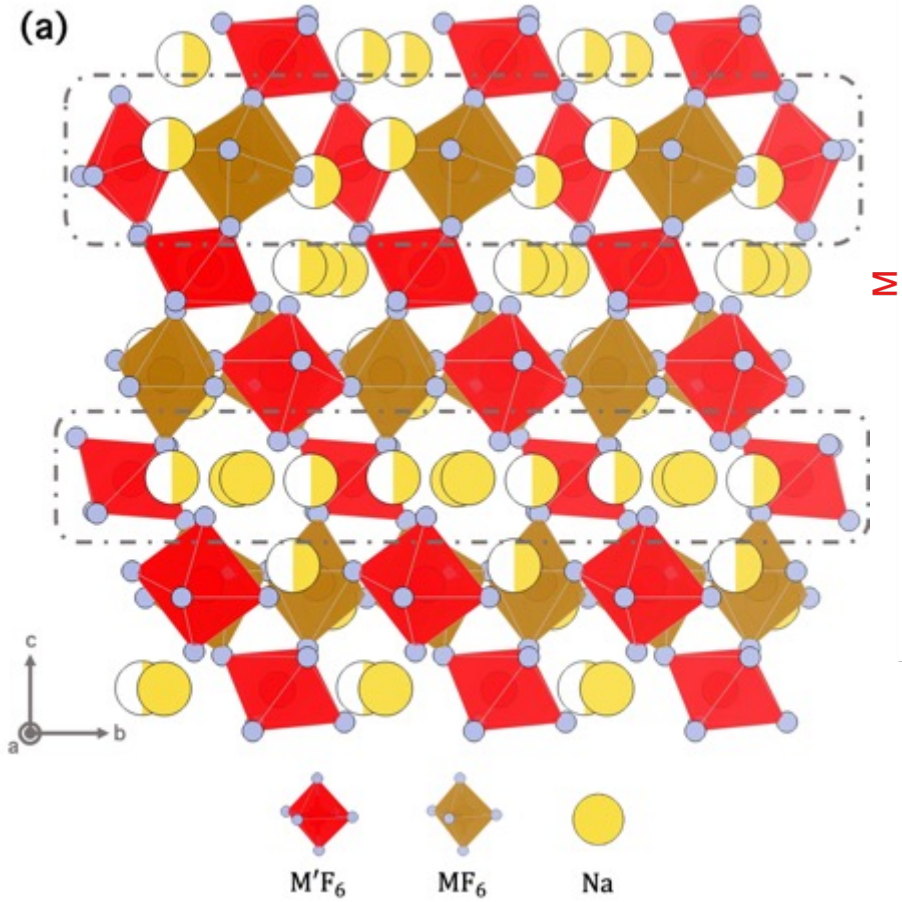


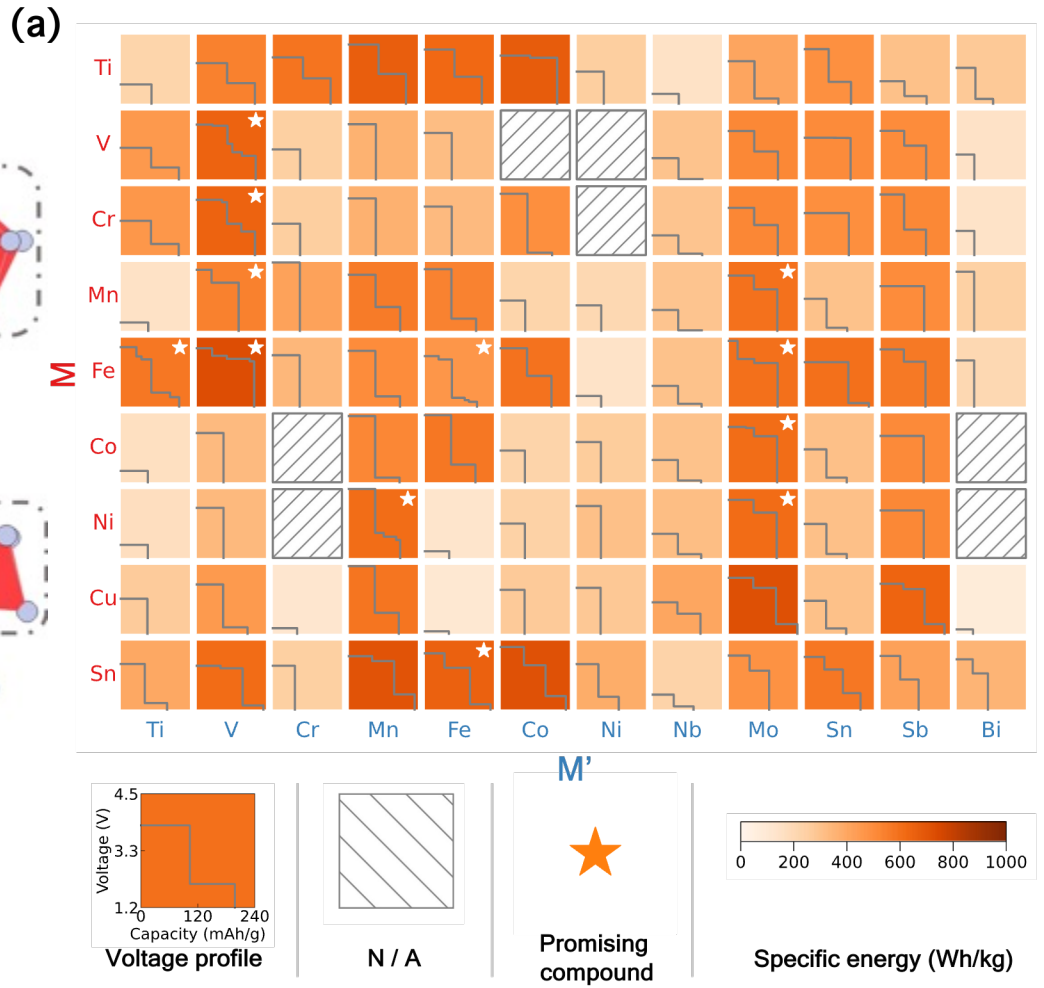


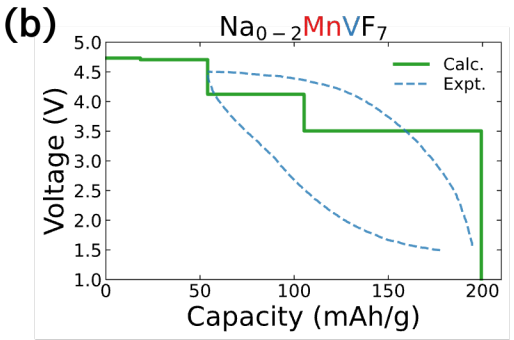


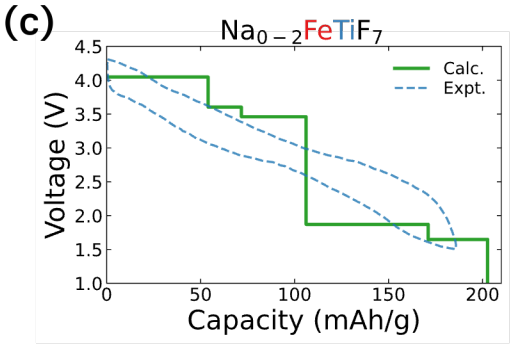
J. Mater. Chem. A, 2024,**12**, 3954

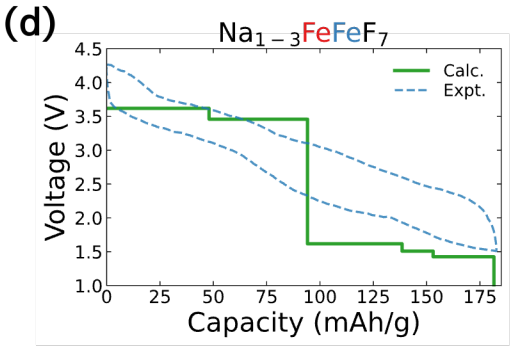
5 Na电池正极





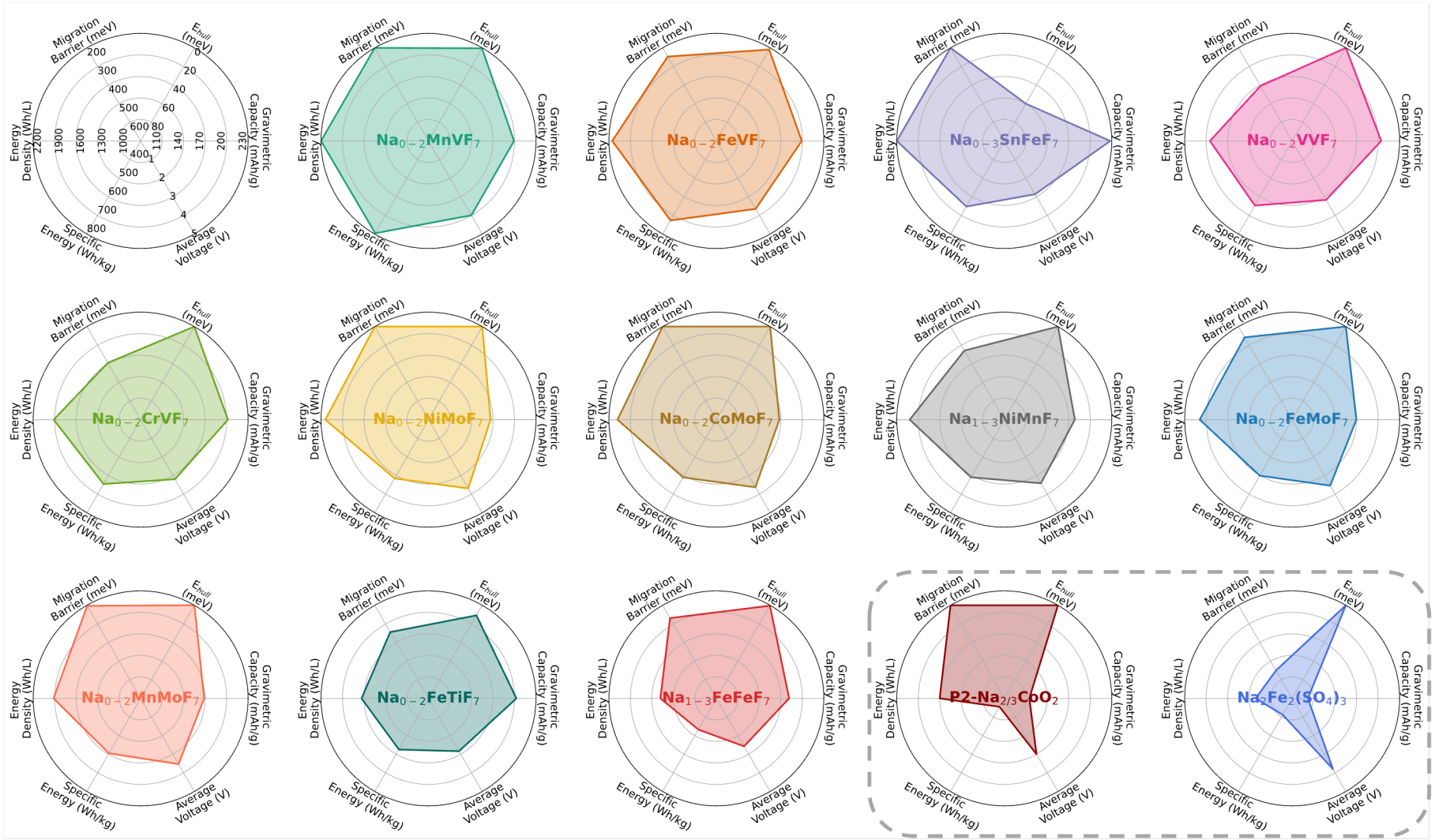






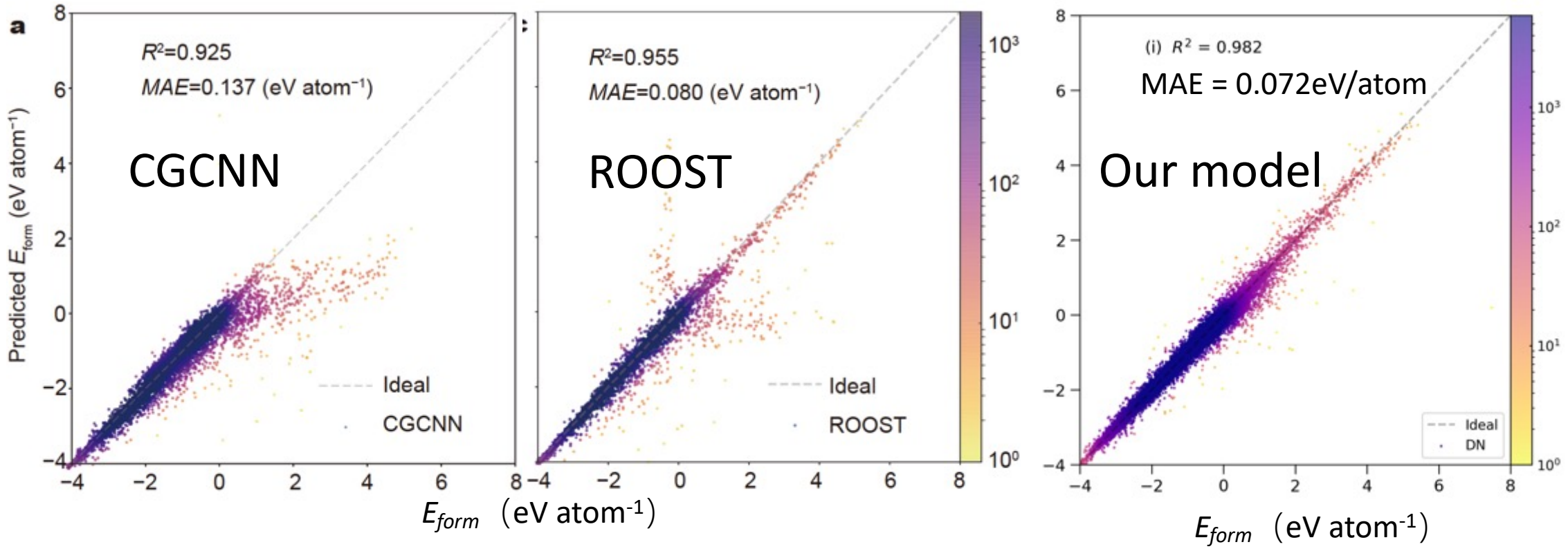
J. Mater. Chem. A, 2024,**12**, 14709

5 Na电池正极



J. Mater. Chem. A, 2024,**12**, 14709

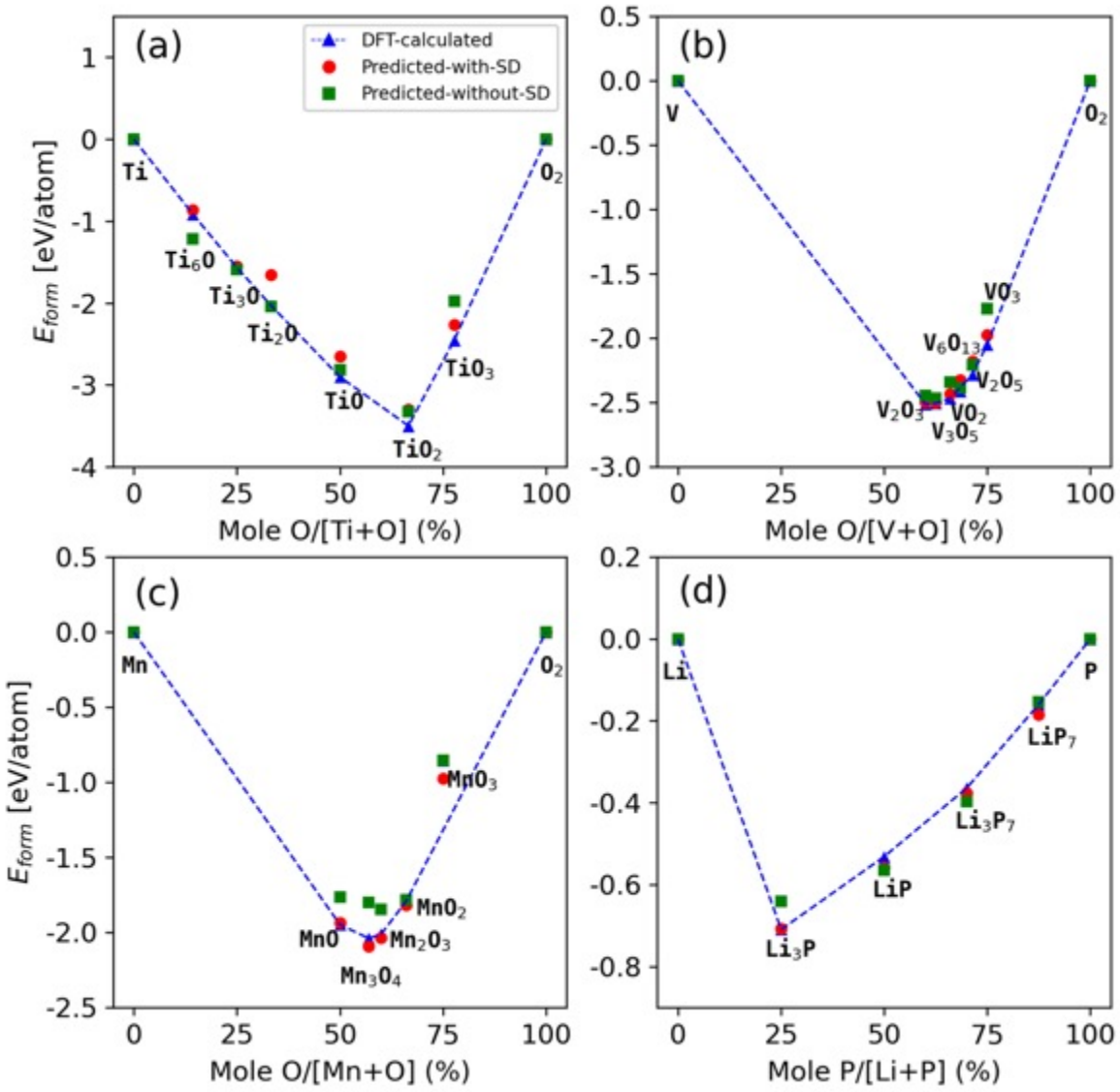
6 人工智能模型，good, bad and ugly



	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）



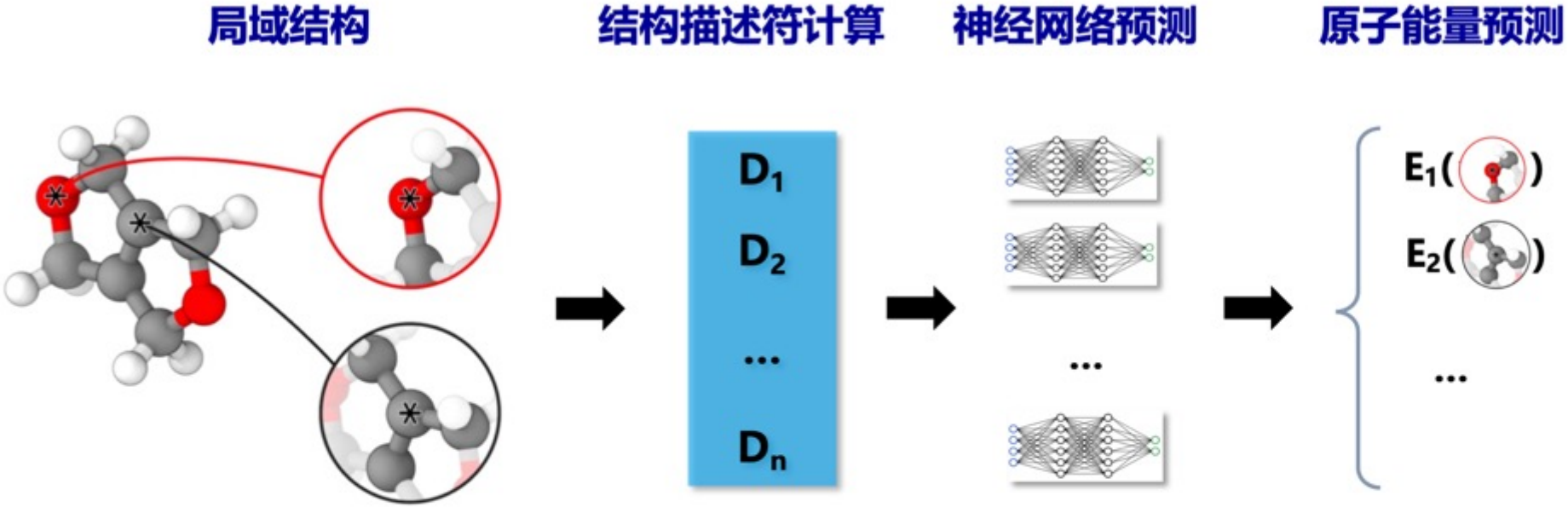
秒级预测材料的热力学稳定性

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 $f_{j\to i} = \frac{\partial E_i}{\partial r_j} = \frac{\partial E_i}{\partial d_i}\frac{\partial d_i}{\partial r_j}$

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

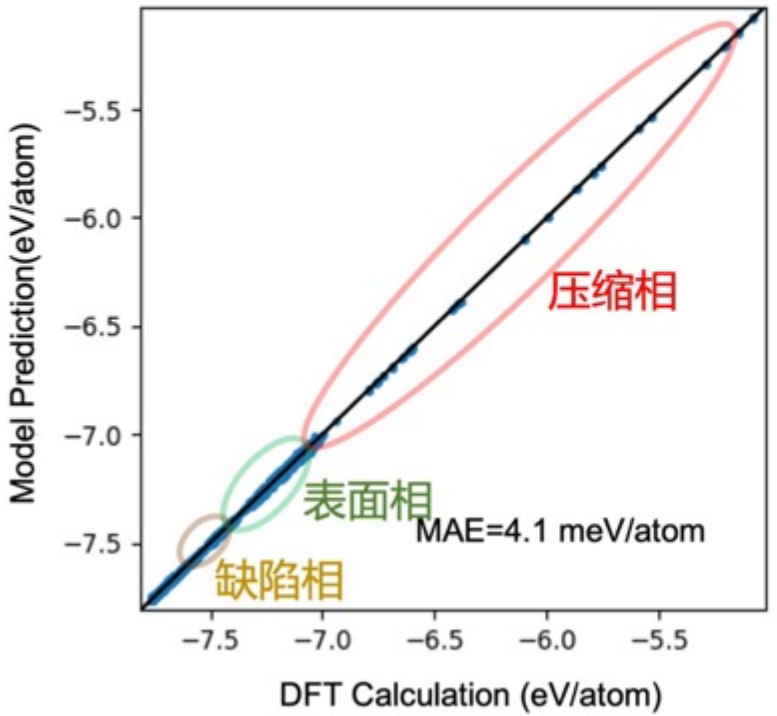
原子应力 $v_{\alpha\beta}^i = \sum_k^{N(i)} r_{k\to i}^{\alpha} f_{k\to i}^{\beta}$

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

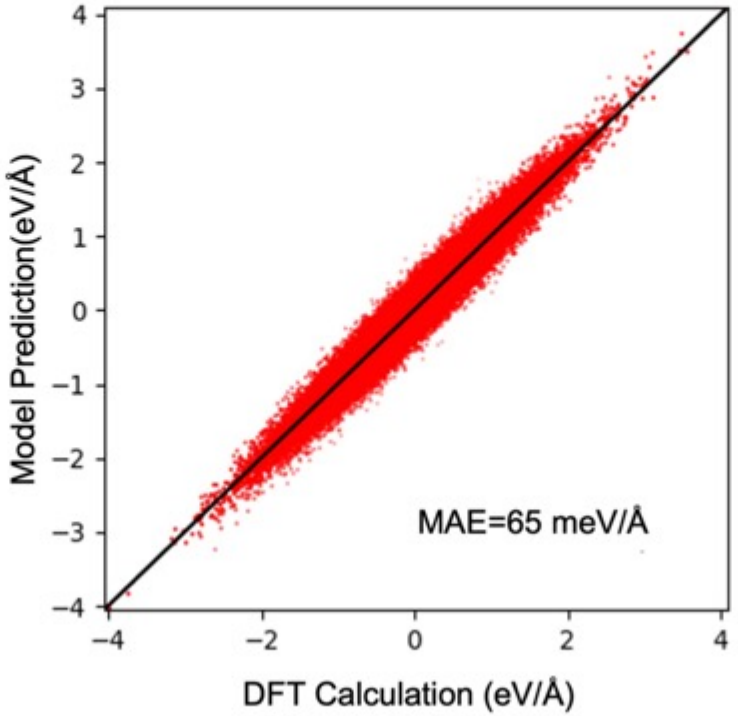
7 人工智能模型预测结合能（V2）

高精度Ti金属机器学习力场

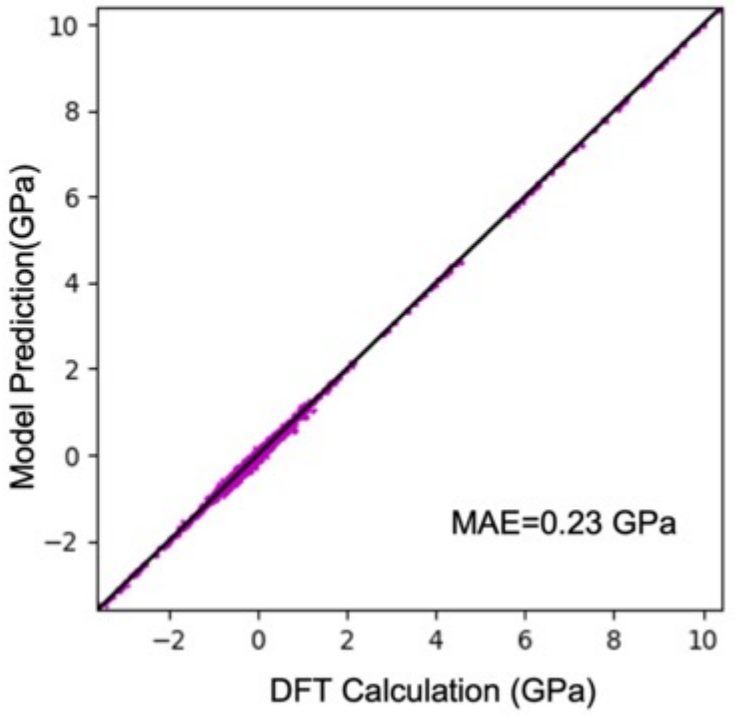
能量误差



受力误差



应力误差



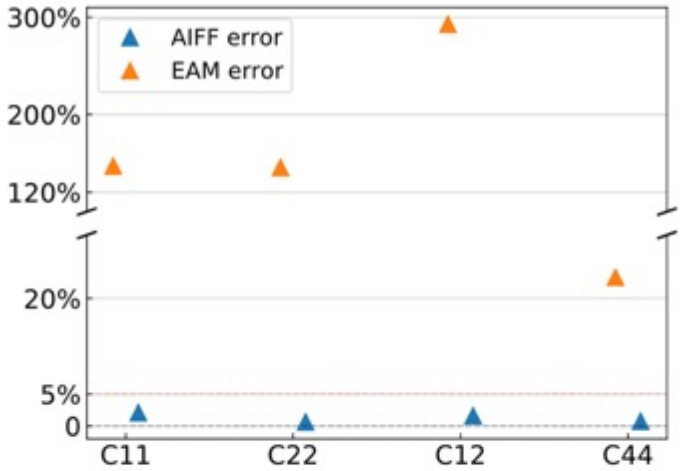
7 人工智能模型预测结合能（V2）

弹性属性

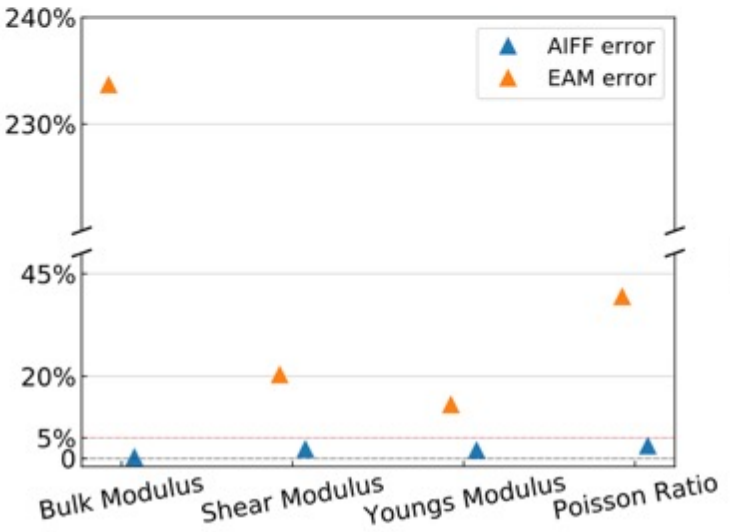
$$\sigma_i = \sum_{j=1}^{6} C_{ij}\varepsilon_j$$

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

	DFT	AIFF	EAM
C_{11} (GPa)	176.77 (176.1[a])	173.06	436.52
C_{22} (GPa)	177.37	178.49	435.41
C_{12} (GPa)	95.55(86.9[a])	97.09	375.41
C_{33} (GPa)	198.33(190.5[a])	193.0	414.26
C_{44} (GPa)	39.18 (50.8[a])	39.47	48.32
Bulk Modulus (GPa)	118.78 (110.0[a])	119.10	396.34
Shear Modulus (GPa)	44.11	43.13	35.13
Youngs Modulus (GPa)	117.76(109.1[a])	115.45	102.37
Poisson Ratio (μV)	0.33	0.34	0.46



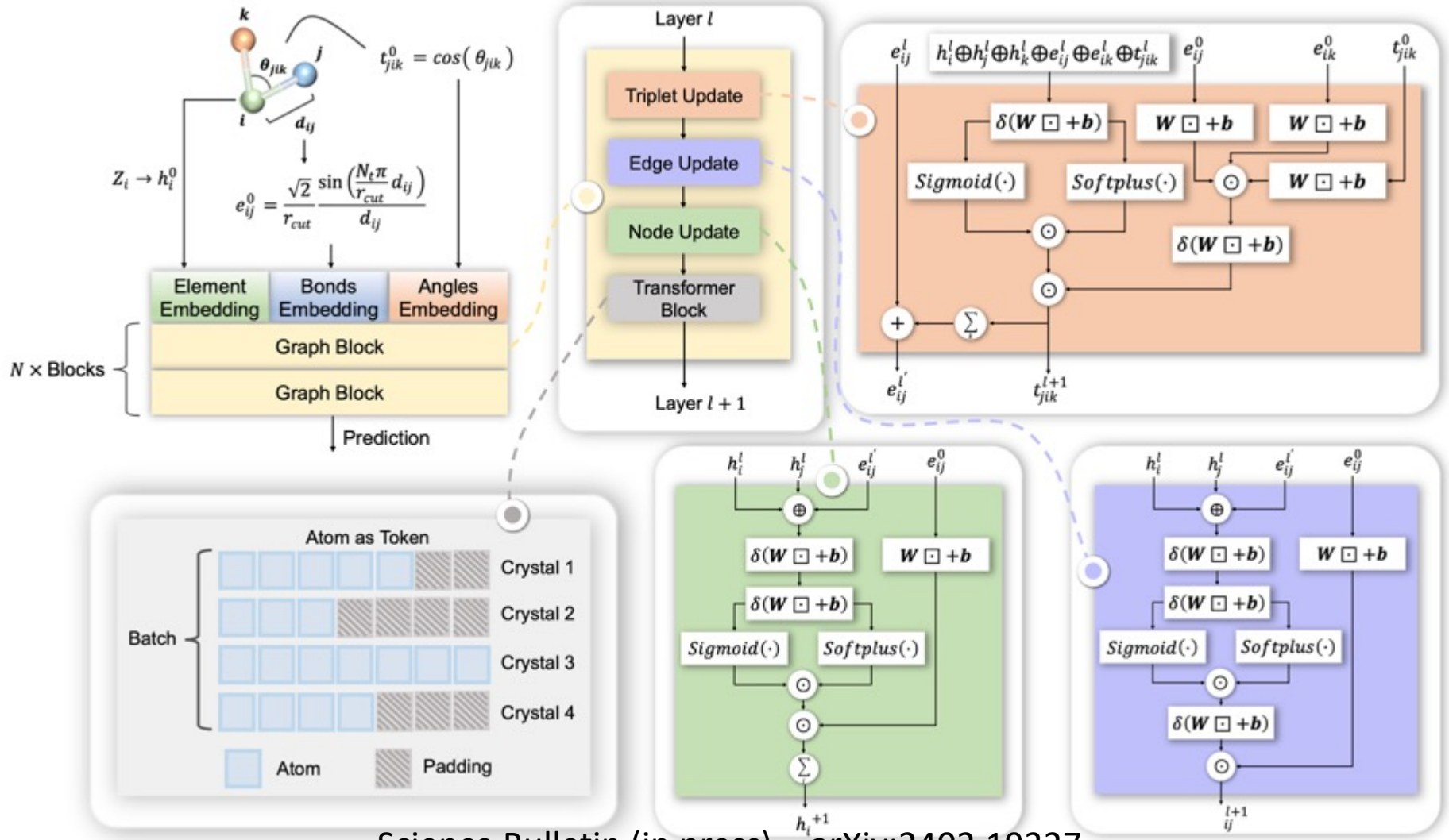
弹性常数



弹性模量

通用力场 （适用于所有元素组合）

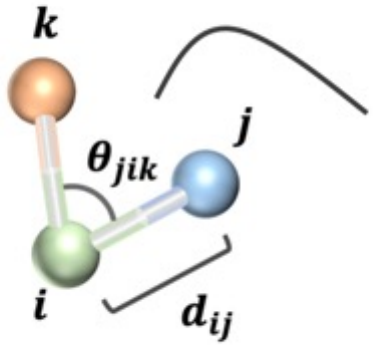
Graph-based Pretrained Transformer Force Field (GPTFF)



Science Bulletin (in press), arXiv:2402.19327

通用力场 （GPTFF）

角向量更新



$$t^0_{jik} = cos(\theta_{jik})$$

嵌入式网络 ⟶

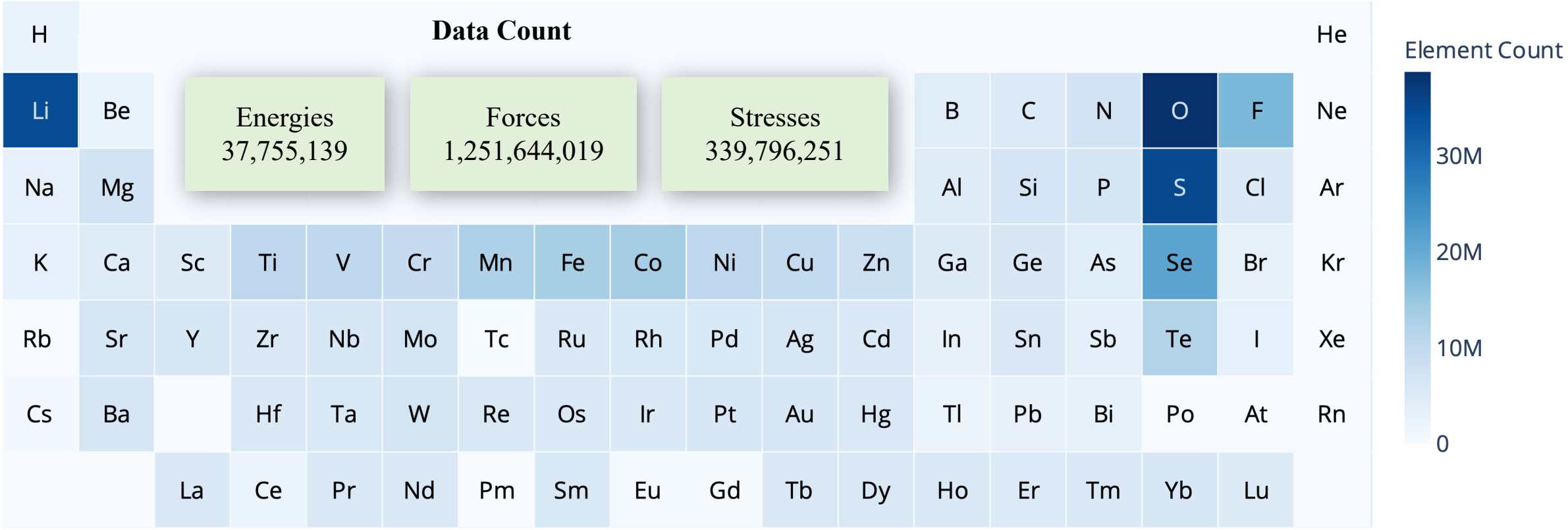
$$t^{l+1}_{jik} = \left(\quad \oplus \quad \oplus \quad \oplus \quad \oplus \quad \oplus \quad \right)$$

原子 j　原子 i　原子 k　键 ij　键 ik　键角 jik

Science Bulletin (in press)， arXiv:2402.19327

通用力场 （GPTFF）

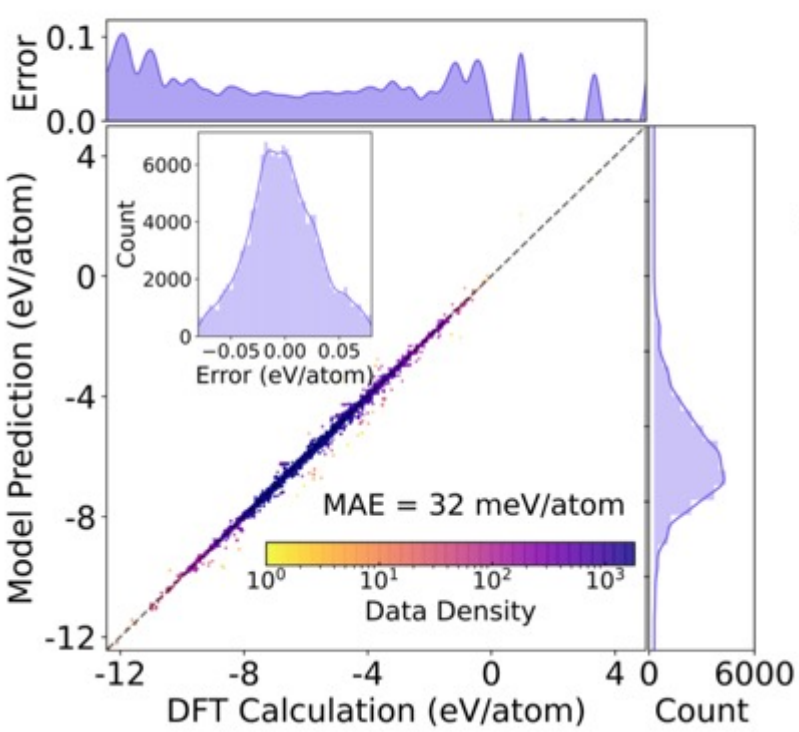
Graph-based Pretrained Transformer Force Field (GPTFF)

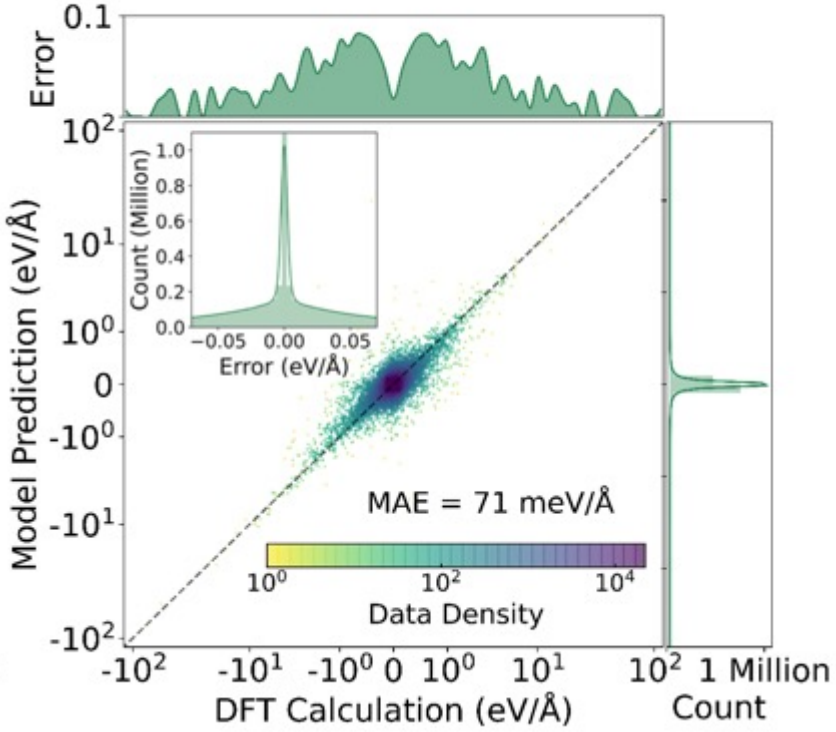


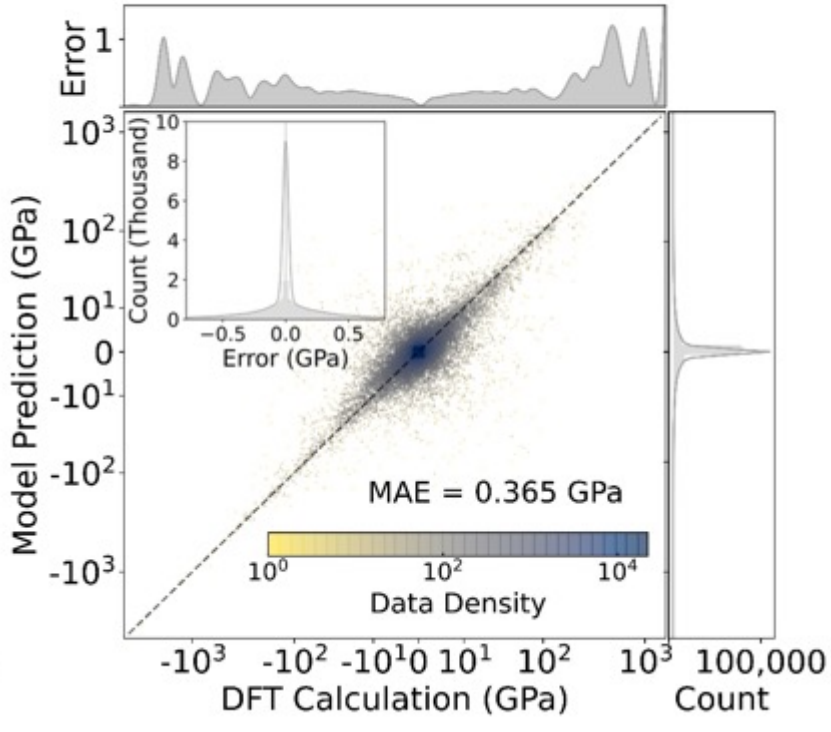
Science Bulletin (in press)， arXiv:2402.19327

通用力场 （GPTFF）

验证指标





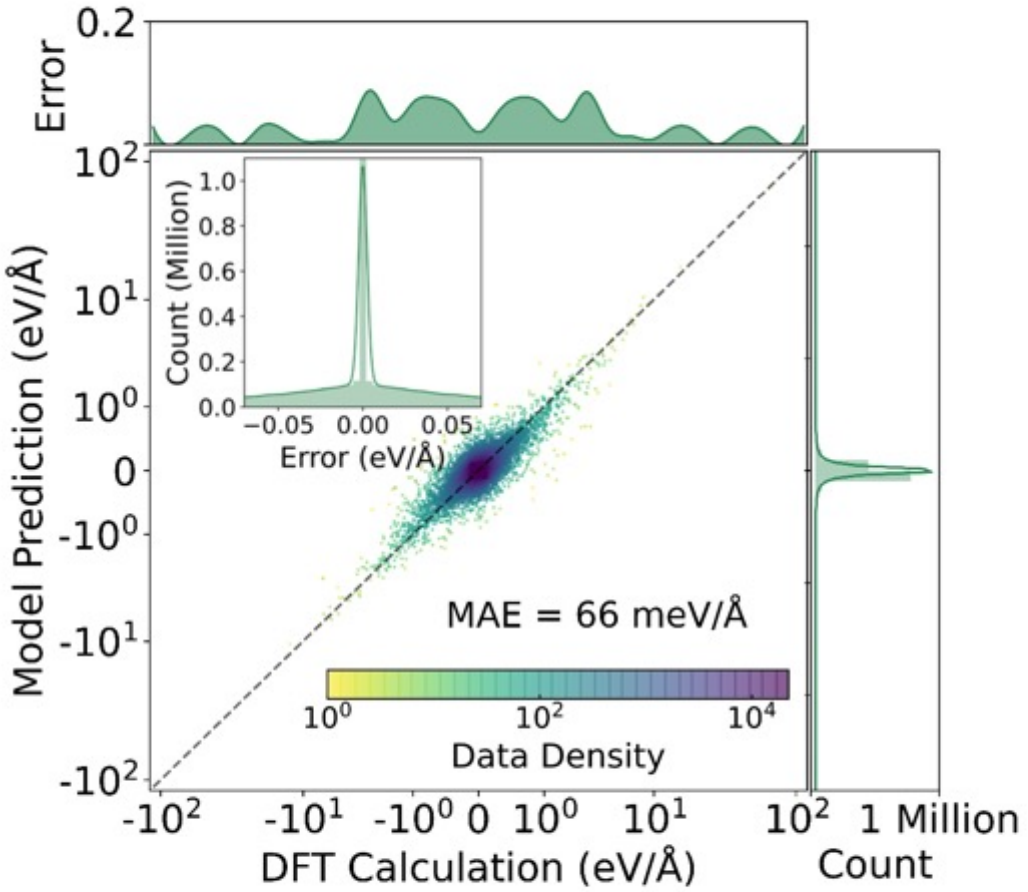


模型	能量误差(meV/atom)	训练数据(Million)
GPTFF	32	37.8
M3GNet	35	0.18
CHGNet	33	1.5

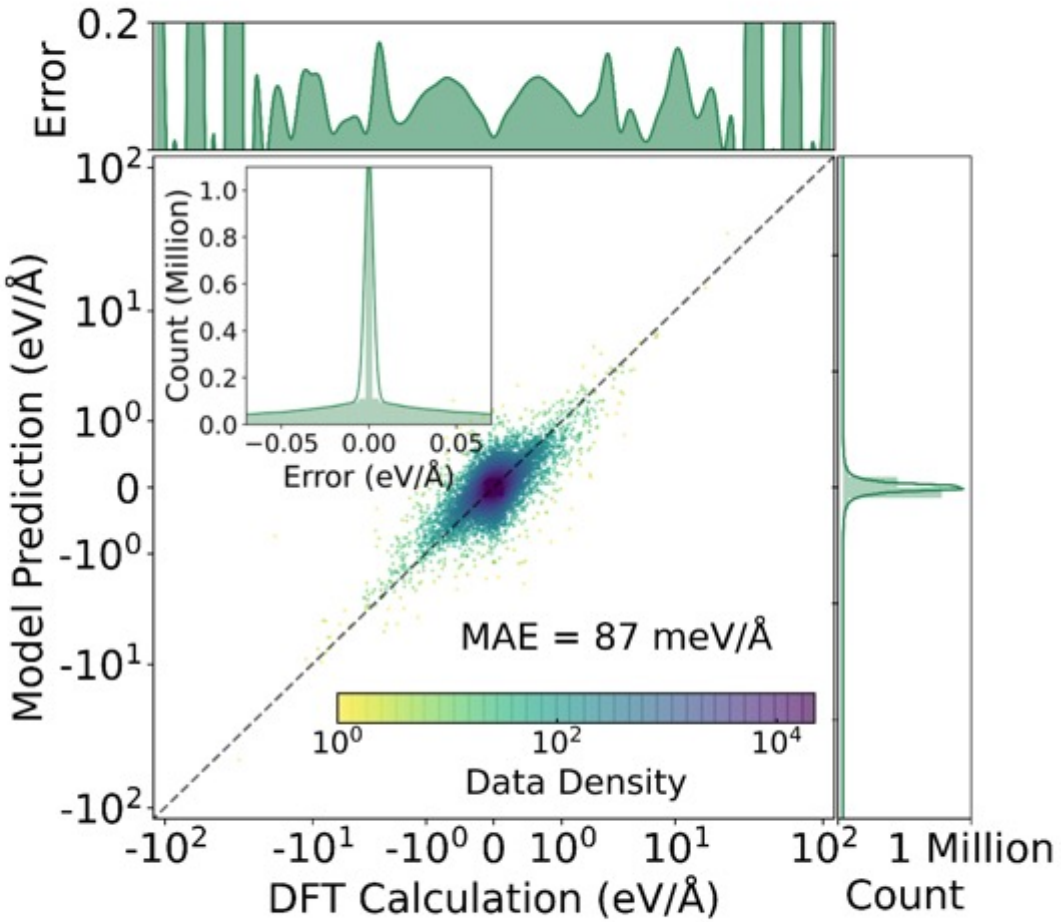
Science Bulletin (in press)， arXiv:2402.19327

通用力场 （GPTFF）

性能比较



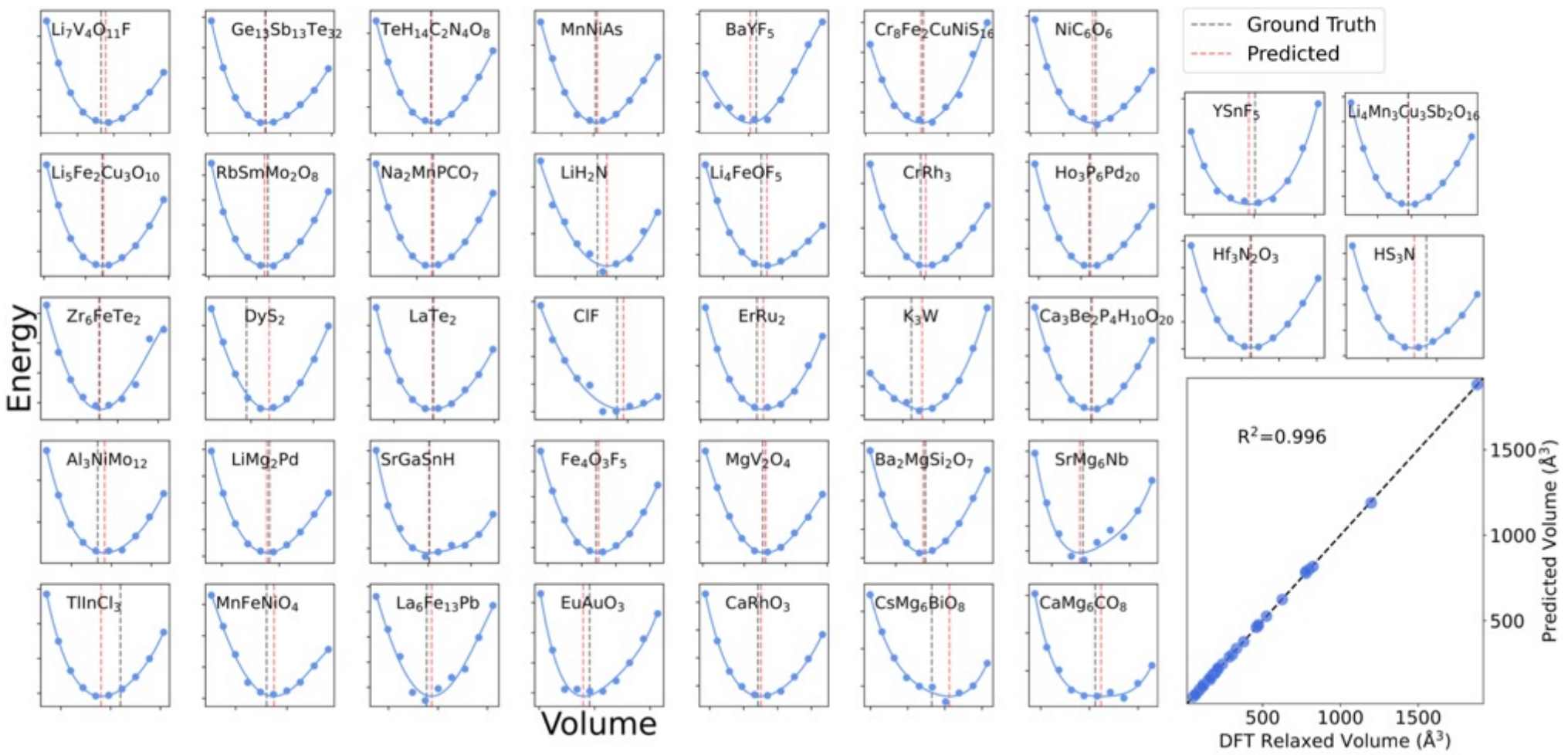
GPTFF 预测结果



CHGNet预测结果

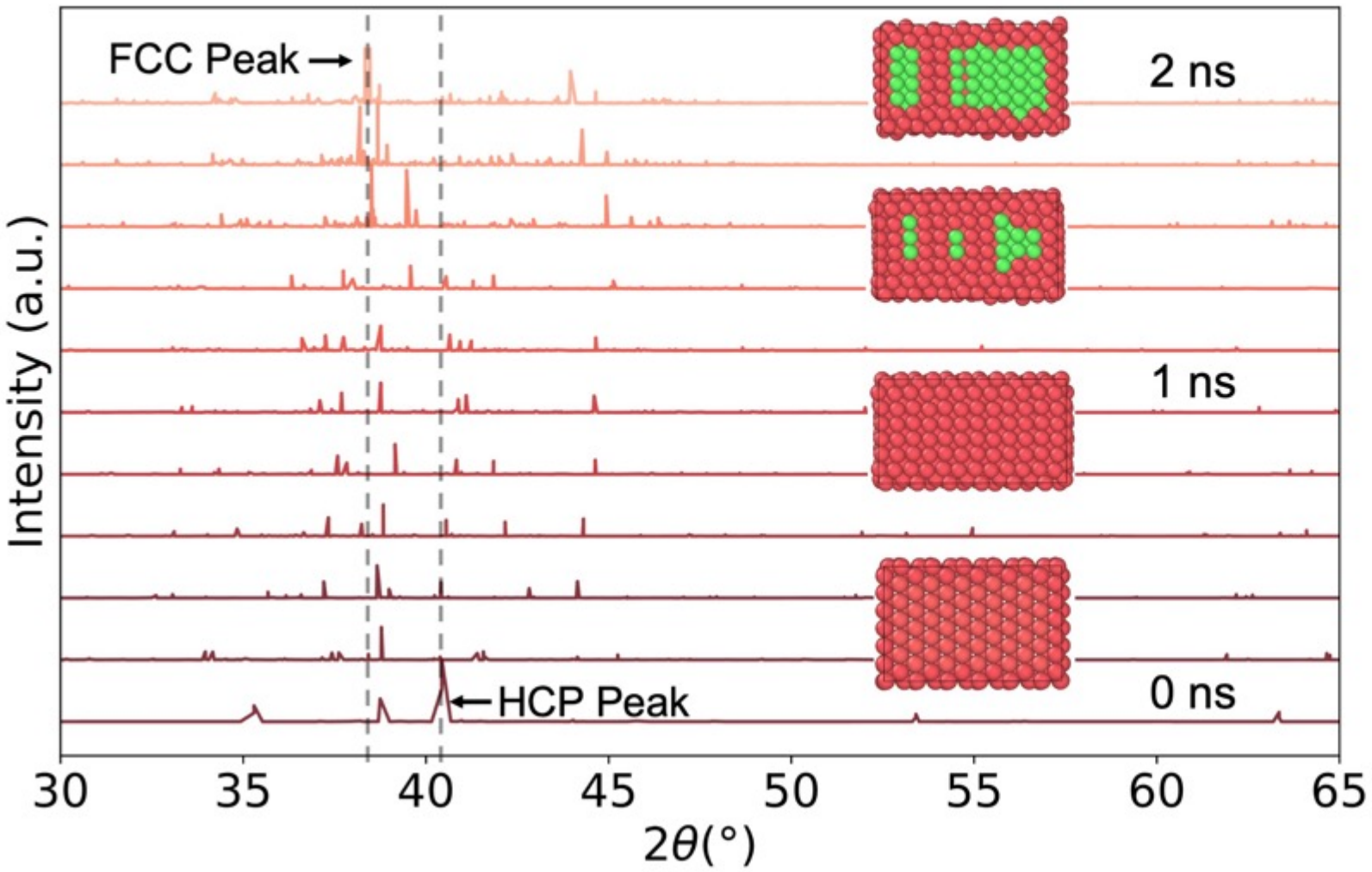
通用力场 （GPTFF)

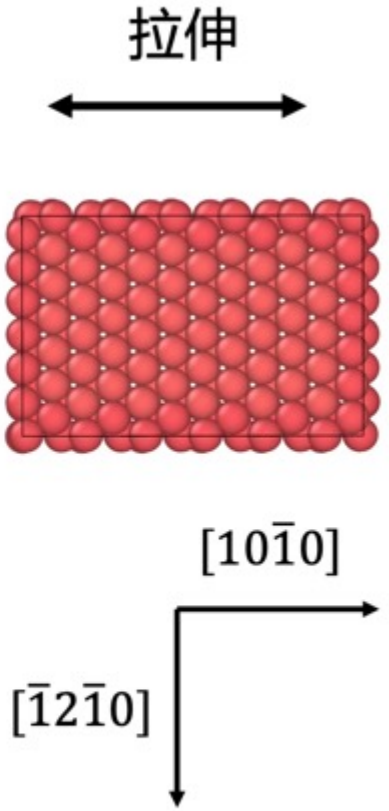
物态方程



通用力场 （GPTFF）

Ti体系的HCP→FCC相变

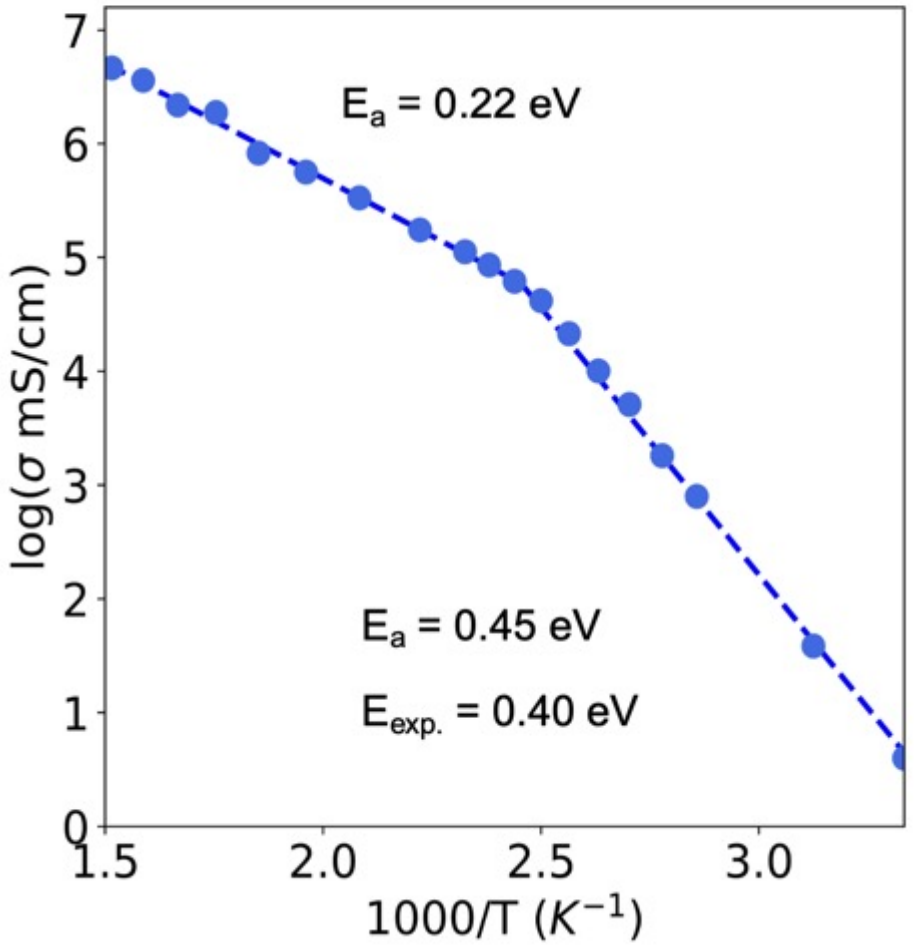




通用力场 （GPTFF）

Li_3YCl_6离子电导率

阿伦尼乌斯关系

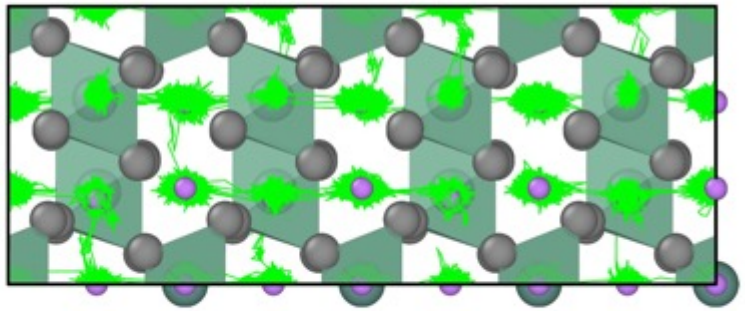


$\sigma_{300K} = 0.6\ \mathrm{mS/cm}$

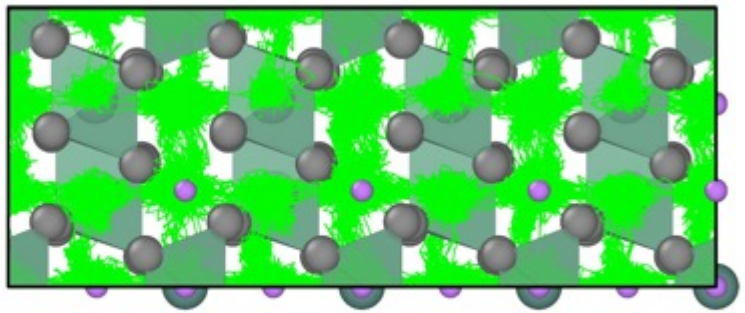
$\sigma_{exp.} = 0.51\ \mathrm{mS/cm}$

Li^+扩散轨迹

300K下的扩散轨迹

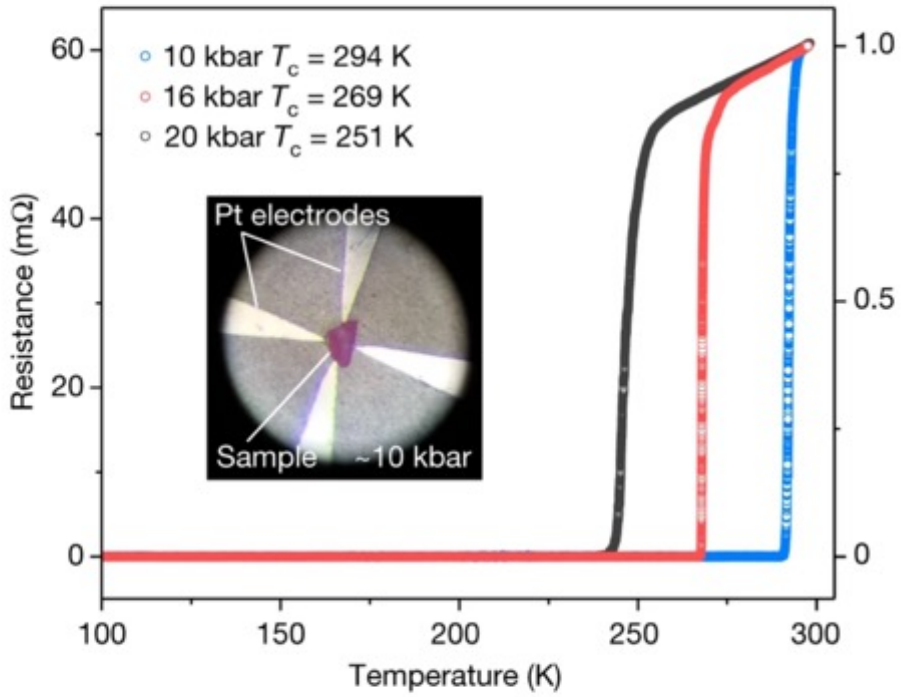


500K下的扩散轨迹

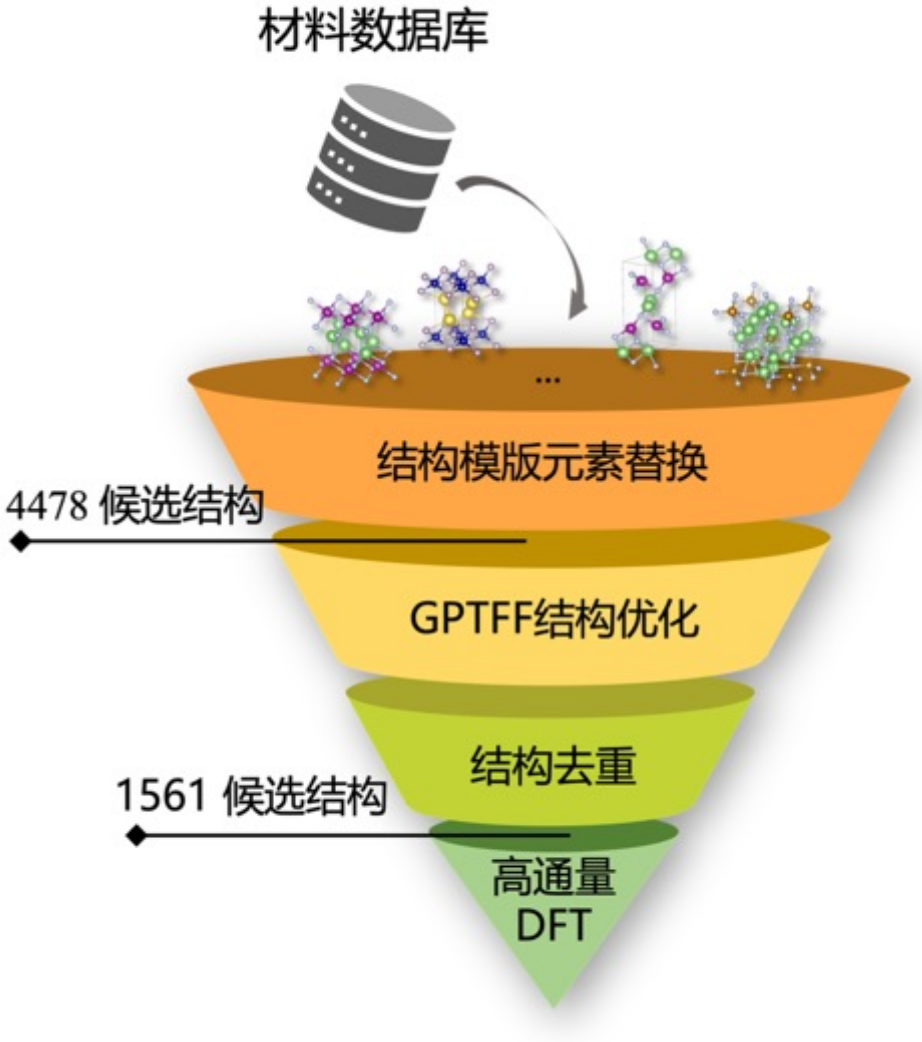


通用力场 （GPTFF）

Lu-H-N相图的快速构建

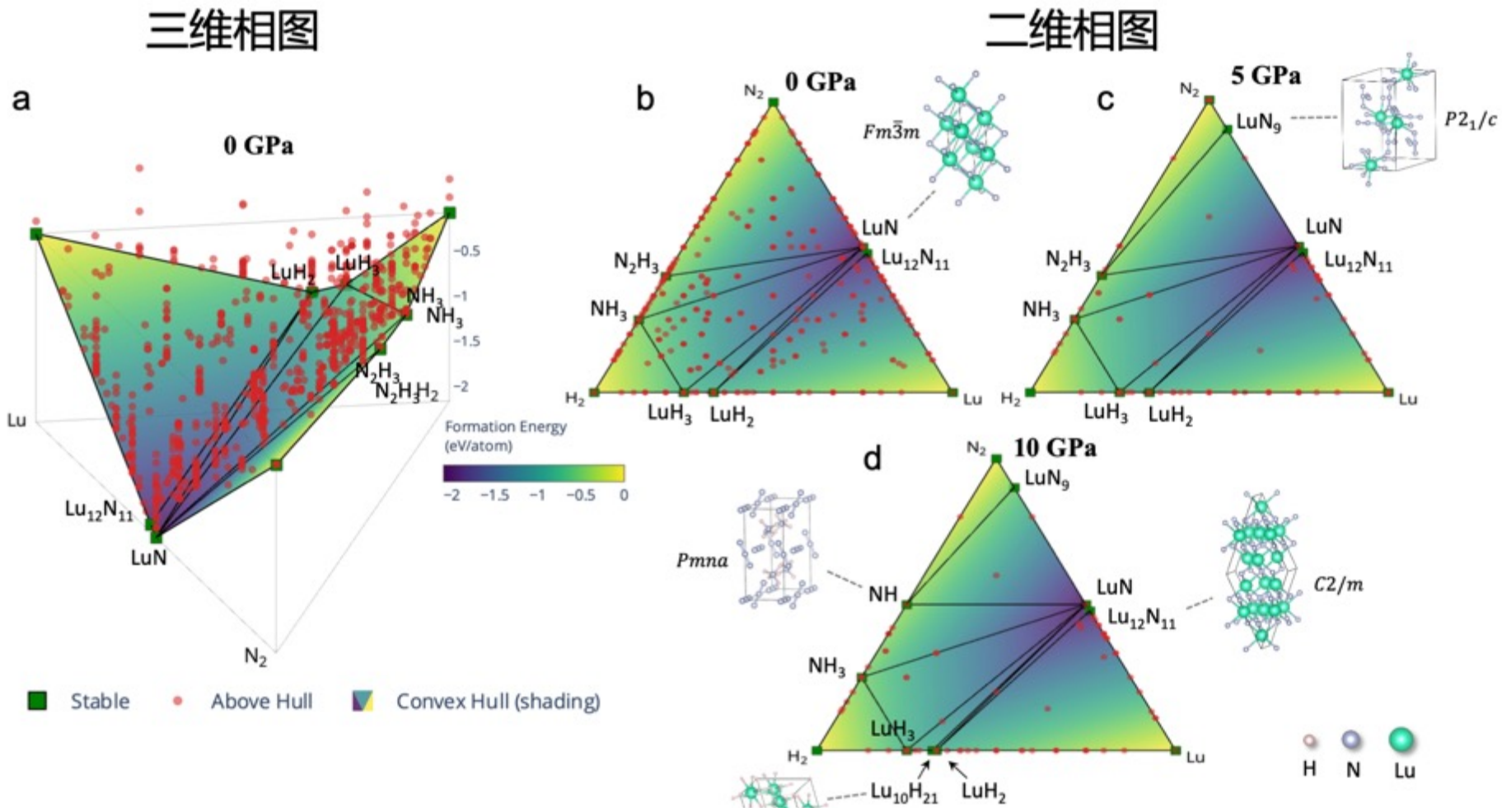


RETRACTED ARTICLE: Nature, 2023, 615(7951): 244-250.



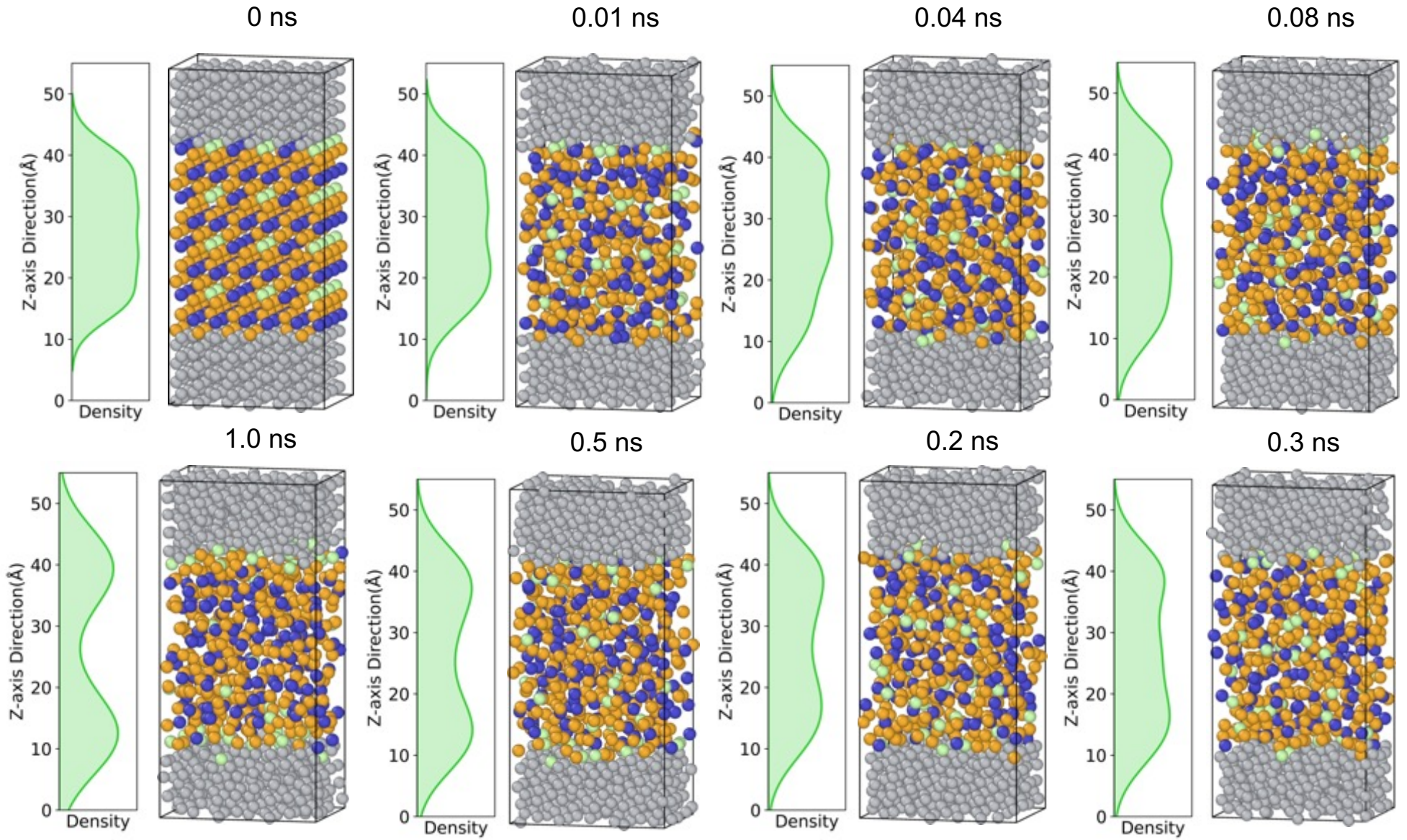
通用力场 （GPTFF）

Lu-H-N相图的快速构建



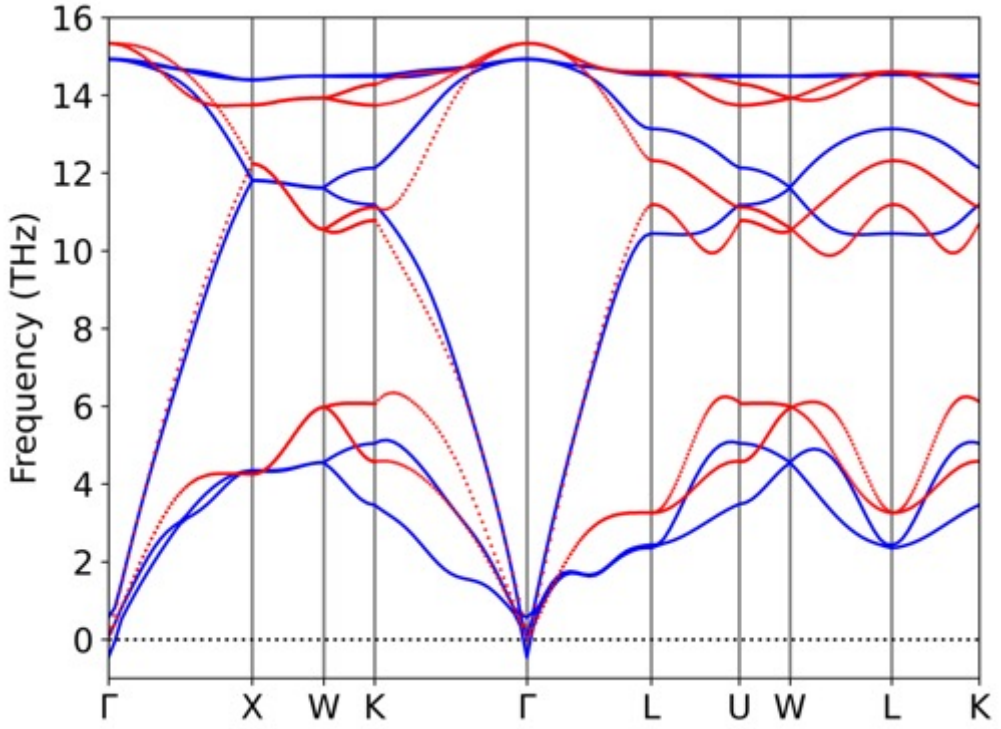
Chin. Phys. Lett. 40 057401 (2023)

通用力场 （GPTFF）

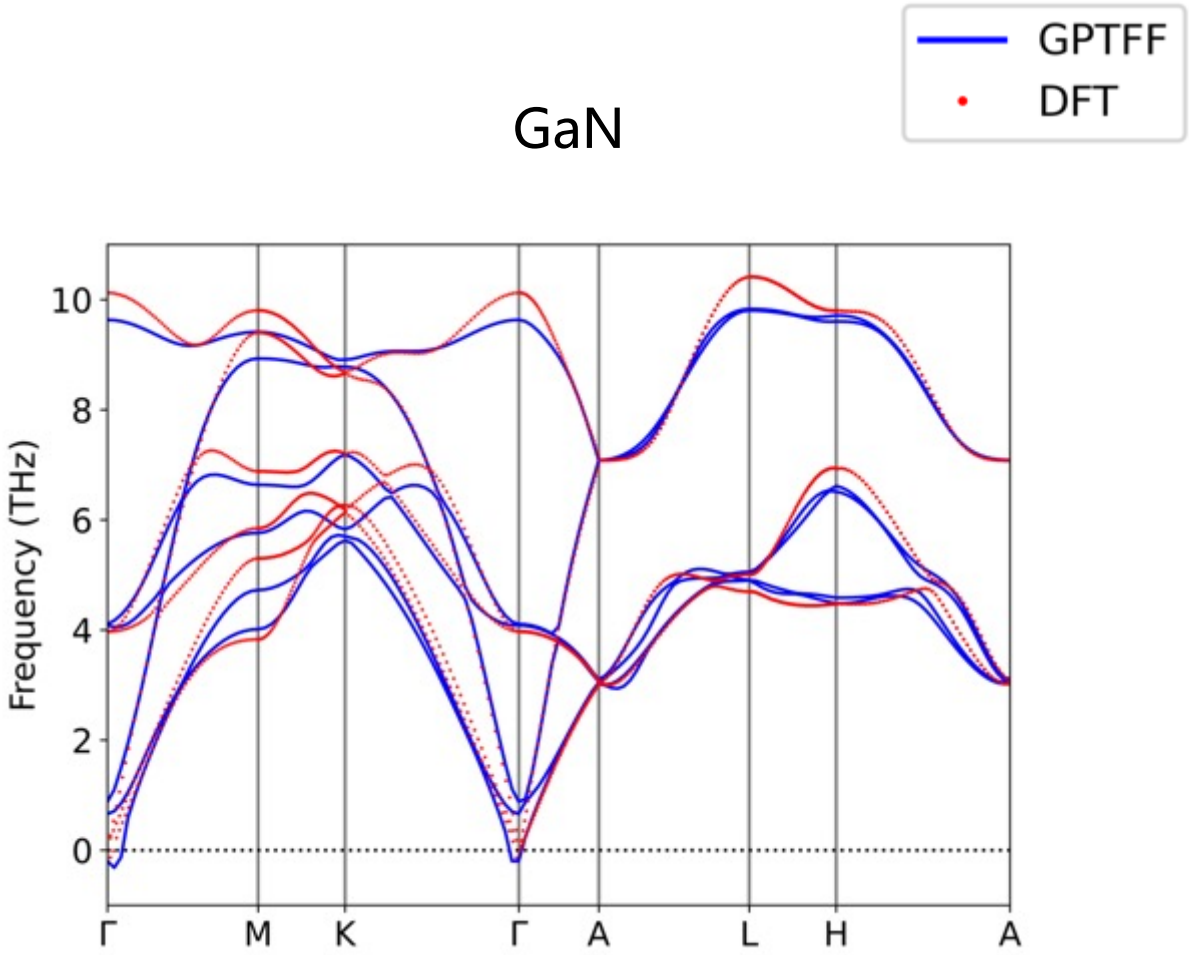

0 ns
0.01 ns
0.04 ns
0.08 ns
1.0 ns
0.5 ns
0.2 ns
0.3 ns
Z-axis Direction(Å)
Density
0
10
20
30
40
50

通用力场 （GPTFF）

Si



GaN



通用力场 （GPTFF）

结构优化

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

```python
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)
```

通用力场 （GPTFF）

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

```python
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'), # Traje
                   append_trajectory=True)
dyn.run(100000)
```

GPTFF

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

The race

Meta

Orbital materials

Microsoft

DeepMind

Model	F1 ↑	DAF ↑	Prec ↑	Acc ↑	TPR ↑	TNR ↑	MAE ↓	RMSE ↓	R^2 ↑	Training
eqV2	0.917	6.047	0.924	0.975	0.910	0.986	0.020	0.072	0.848	3M (102.4
ORB	0.880	6.041	0.924	0.965	0.841	0.987	0.028	0.077	0.824	3M (32.1M
MatterSim	0.859	5.646	0.863	0.957	0.856	0.975	0.026	0.080	0.812	17M (Ma
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.
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

科学数据开放平台

隐私计算技术、数据可用不可见

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


SONGSHAN LAKE
MATERIALS LABORATORY
松山湖材料实验室

谢谢大家，请您批评指正！