



中国科学技术大学
University of Science and Technology of China

热烈祝贺安徽理工大学“基础物理研究中心”成立！

基于图深度学习加速的晶体结构预测方法发展

林子敬

2024-05-23

目录

1. 研究背景

1.1. 大数据与人工智能时代

1.2. 晶体结构预测方法介绍

2. SCCOP方法发展

2.1. 晶胞生成与网格剖分

2.3. SCCOP核心方法

2.2. 图神经网络预测模型

2.4. SCCOP整体流程

3. SCCOP测试与应用

3.1. SCCOP算例测试

3.3. 有序空位组分 CuIn_5Se_8 结构搜索

3.2. 硼同素异形体搜索

3.4. SCCOP软件获取和使用

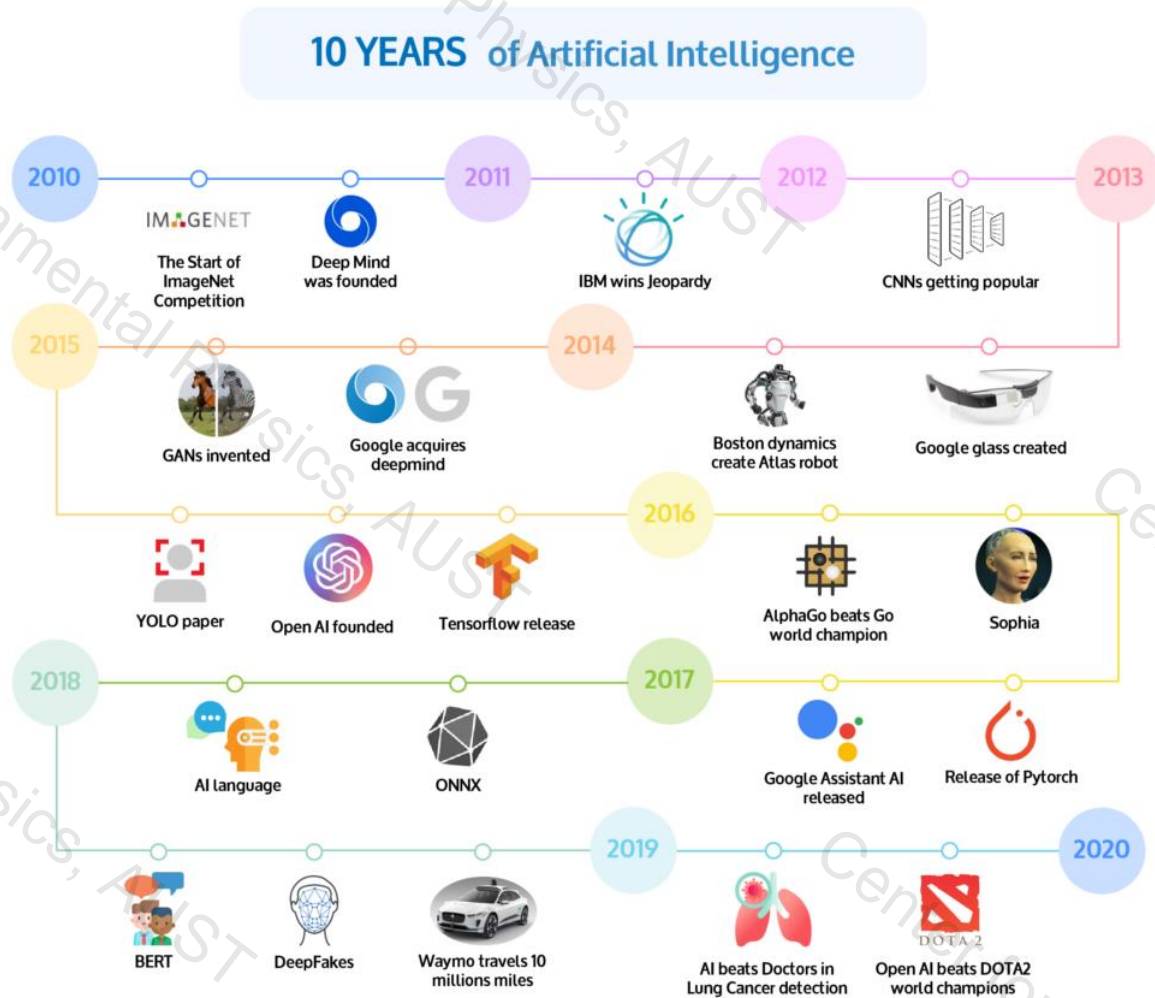
4. 总结与展望

4.1. SCCOP方法与应用总结

4.2. SCCOP软件发展总结

1.1a 大数据和人工智能时代

2010-2020年AI发展史:



代表性事件:

2012年 卷积神经网络CNN在图像识别大赛 ImageNet中获得冠军, 击败所有传统算法



2016年 人工智能AlphaGo在围棋比赛中, 以4比1 击败世界冠军李世石



2019年 人工智能AlphaStar在策略游戏星际争霸中, 击败世界冠军Serral



2020年 人工智能AlphaFold2在蛋白质结构预测竞赛 (CASP14) 中, 首次达到实验精度

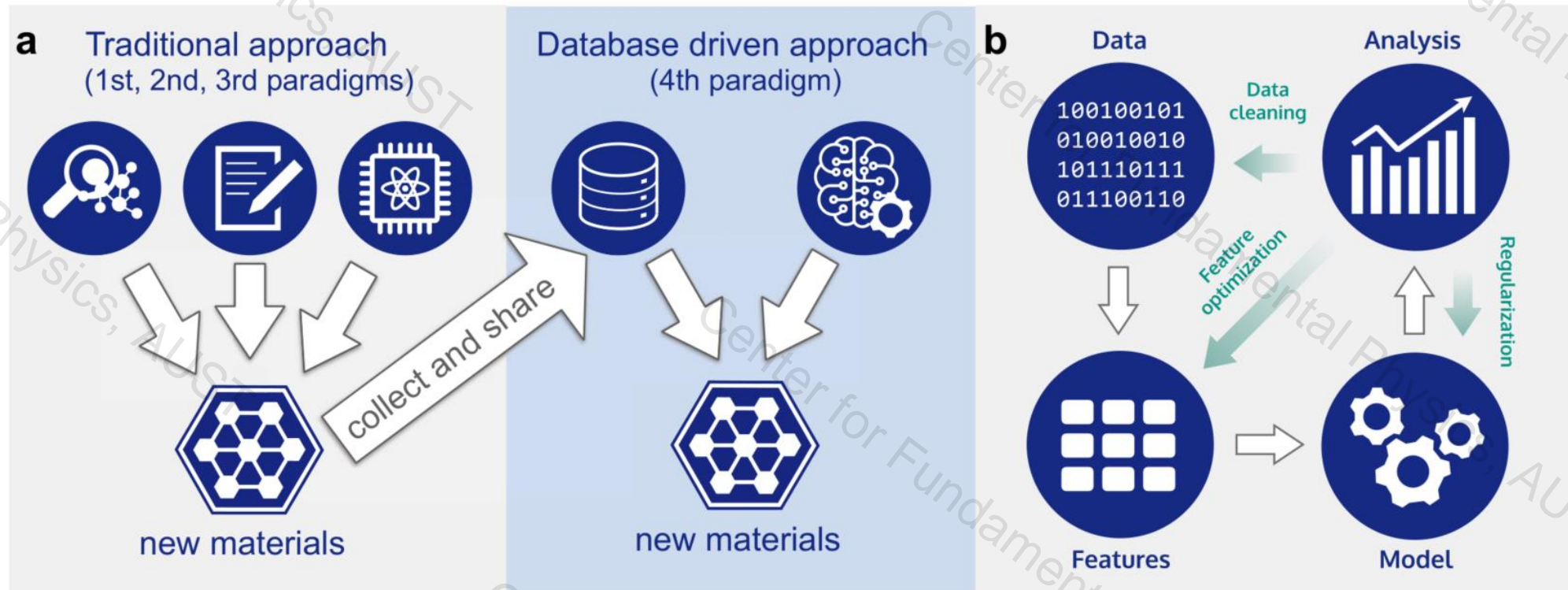
1.1b 大数据和AI驱动的科学研究的

科学研究的四大范式:

实验、理论、计算

数据驱动的第四研究范式

数据、特征、模型、分析



科学研究中，数据通常来源于实验和预测，通过机器学习方法加速结构预测将极大推动数据收集

材料数据库的核心 → 晶体结构

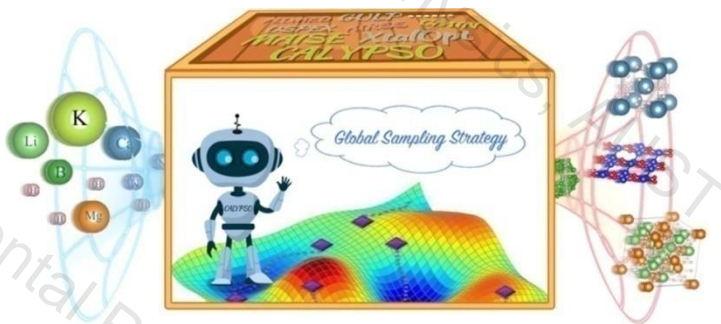
1.2 晶体结构预测 (CSP) 问题 --- 背景介绍

CSP问题定义:

化合物组分



晶体结构



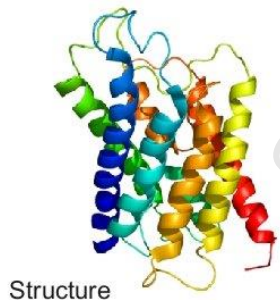
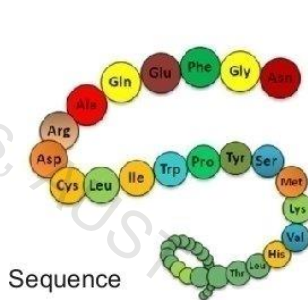
预测给定组分最低自由能的结构

难点: 可能结构数目随原子数超指数型增长

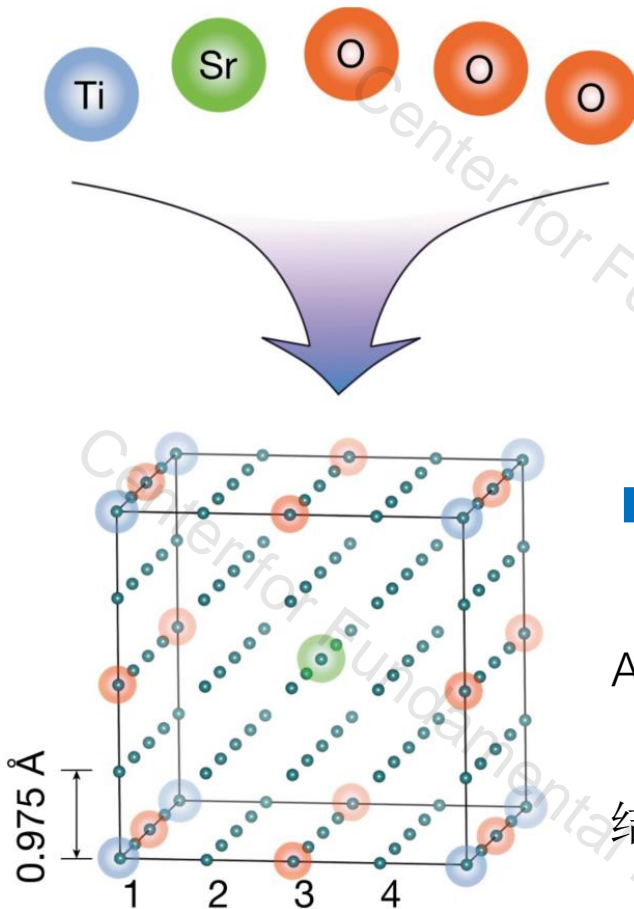
氨基酸序列



蛋白质结构



空间网格化:



解决CSP问题的一般步骤

1. 实空间均匀划分网格;
2. 原子在离散位点上排序;
3. 找最低能的一组排序。



CSP实际上是原子排序问题

AB (20): 10^{30}

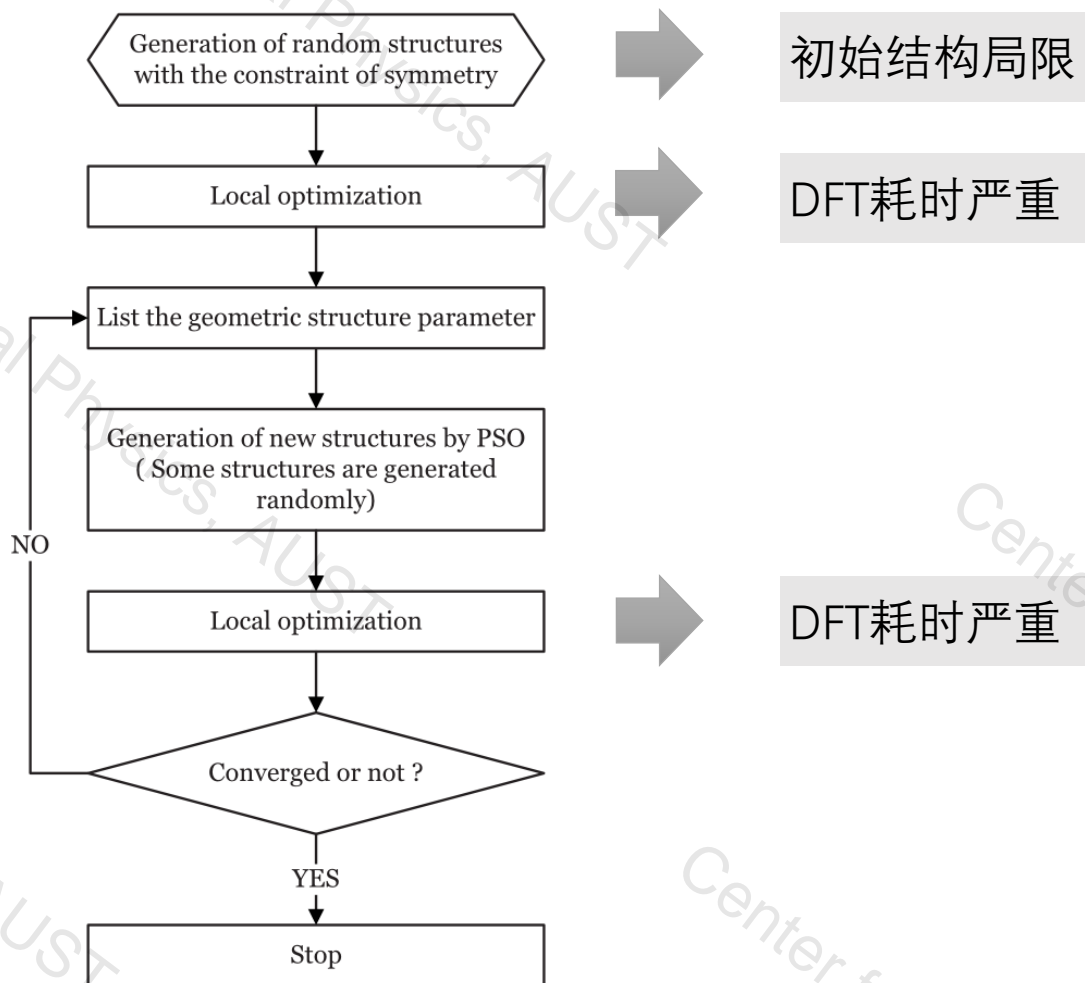
位点组合 原子组合

结构数估计:

$$C = \binom{V/\delta^3}{N} \prod_i \binom{N^*}{n_i}$$

代表性的基于DFT的晶体结构预测方法

经典的CALYPSO/USPEX流程:



流程总结:

1. 按照空间群随机采样生成结构;
2. 结构优化(DFT优化);
3. 粒子群/遗传算法生成新结构;
4. 回到步骤2。

方法优势:

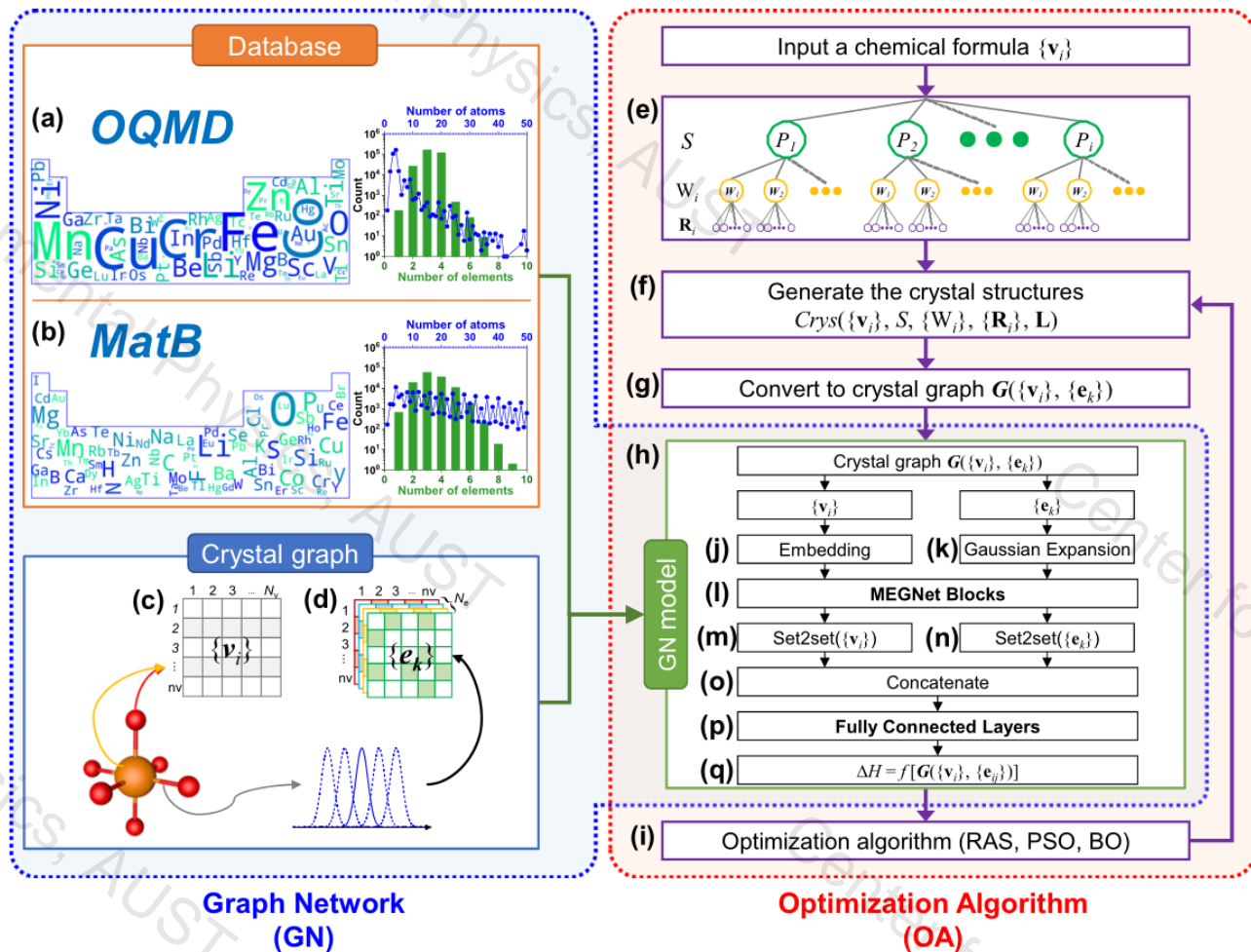
- 无需任何先验信息且结果精度高。

潜在问题:

- 结构优化耗时增长过快;
- 构型搜索空间增长超指数增长。

基于机器学习的晶体结构预测方法1 --- GN-OA

图神经网络-优化算法流程:



流程总结:

0. 数据库上预训练模型;
1. 按照空间群随机采样生成结构;
2. ML优化结构;
3. 返回步骤1。

ML方法优势:

- 结构优化耗时随原子数增长缓慢;
- 低能结构的特征被嵌入预测模型。

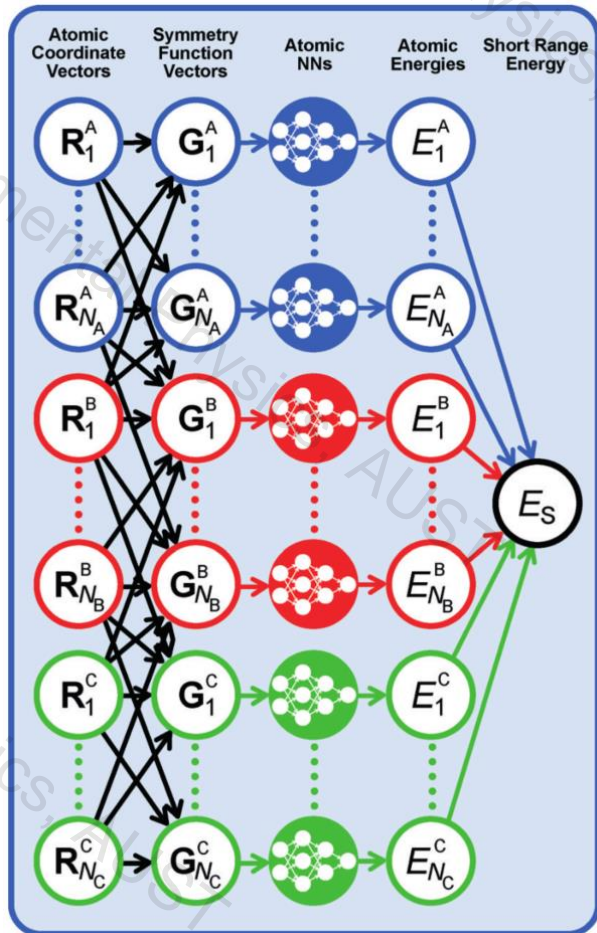
目前存在问题:

- 搜索到的**结构精度有限**;
- 对未知体系的**有效性无法保证**。

基于机器学习的晶体结构预测方法2 --- MAGUS

机器学习力场结构优化:

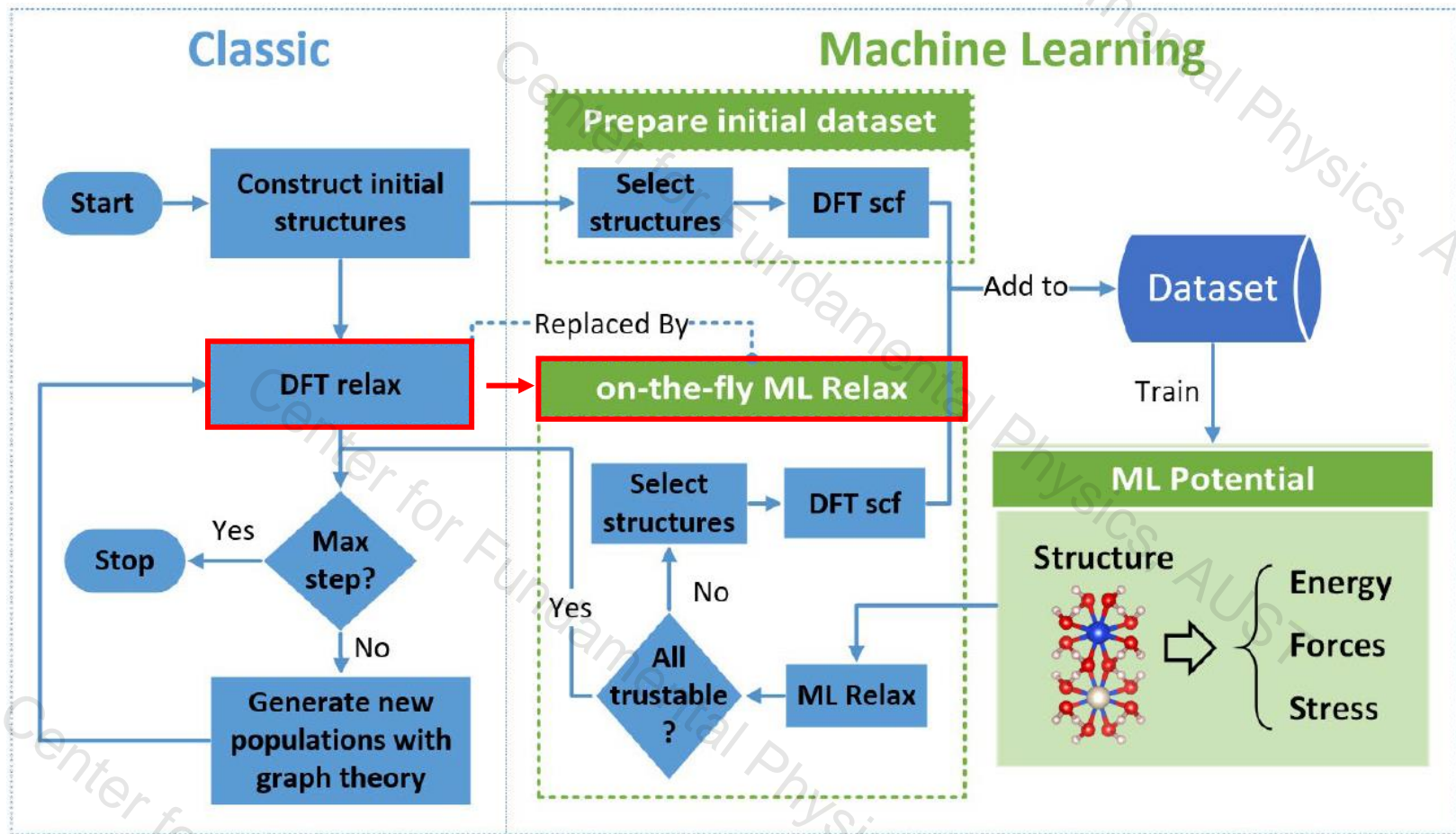
输入: 原子坐标, 输出: 能量



MAGUS:

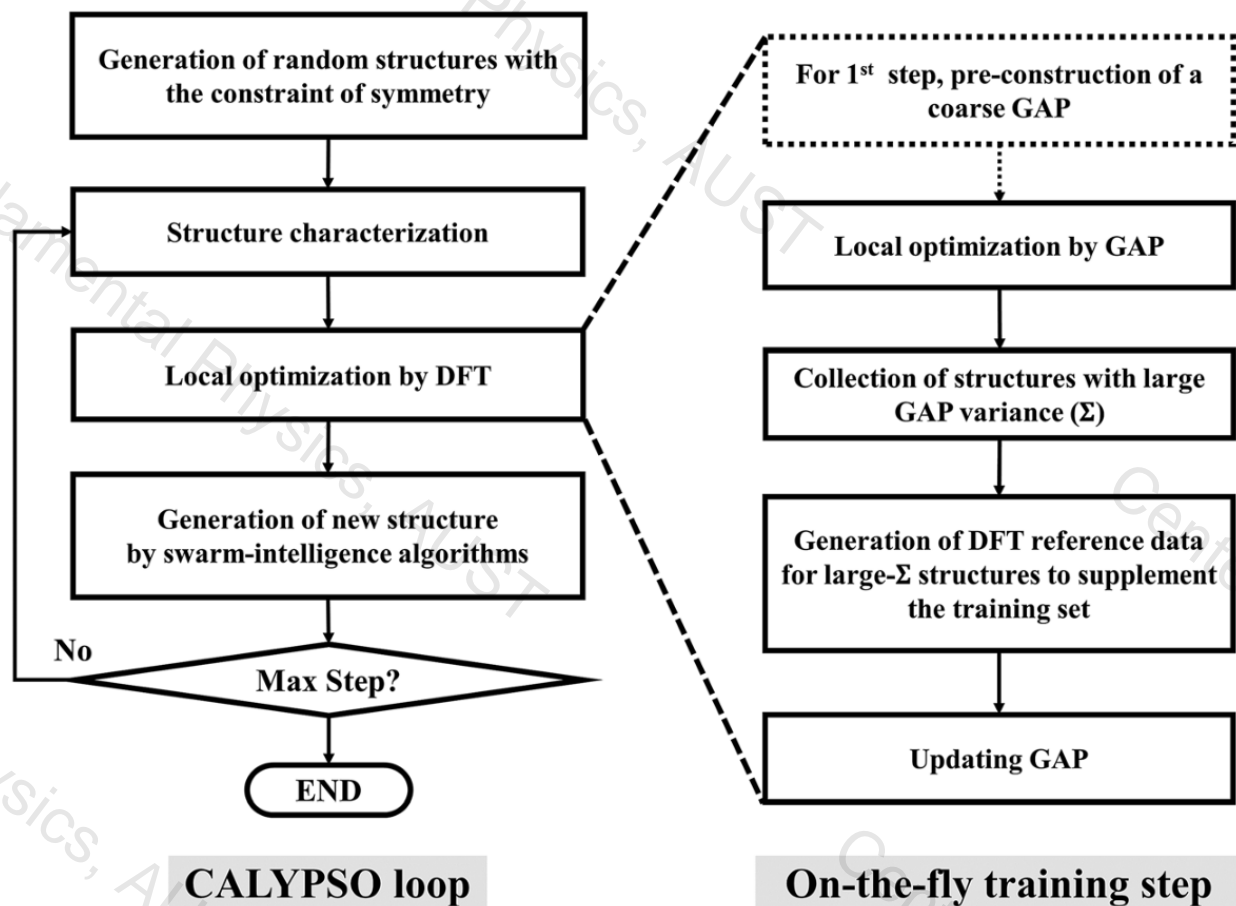
1. 随机采样结构生成

2. 机器学习力场优化结构 (替换DFT优化)



典型的DFT-ML结合的晶体结构预测方法

CALYPSO/USPEX中的ML方法:



流程总结:

1. 按照空间群随机采样生成结构;
2. 训练ML模型;
3. 结构优化(ML+DFT优化);
4. 粒子群/遗传算法生成新结构;
5. 回到步骤2。

方法优势:

- 利用ML (力) 缓解DFT优化的耗时。

存在问题:

- ML的优势未得到充分利用;
- 有效性与可靠性待进一步提高。

1.3 Symmetry-based crystal combinatorial optimization program (SCCOP) 的基本思路和目标

方法比较:

- ◆ 基于DFT的方法: 高精度但耗时长;
- ◆ 基于ML的方法: 耗时短但精度有限;
- ◆ 目前的DFT-ML方法: 耗时短具备一定精度, 但依赖现有数据库且模型存在不确定性。

基本思路:

1. 结构采样: 发展高效的初始结构生成方法, 便于ML模型进行高通量筛选;
2. 结构搜索: 发展高效的势能面划分方法, 使得有潜力的结构能被快速定位;
3. 结构优化: 结合ML的结构优化方法, 尽可能减少DFT结构优化耗时。

目标:

基于机器学习发展一套不依赖数据库的快速且高精度晶体结构预测方法

2. SCCOP方法发展

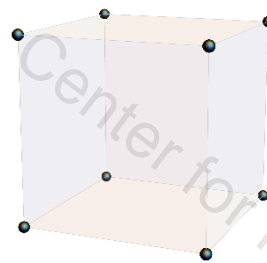
2.1 晶胞生成方法 --- 按晶系生成

七大晶系的晶胞参数:

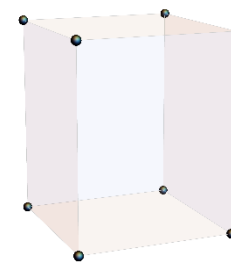
晶系	长度	夹角
立方晶系	$a = b = c$	$\alpha = \beta = \gamma = 90^\circ$
四方晶系	$a = b \neq c$	$\alpha = \beta = \gamma = 90^\circ$
正交晶系	$a \neq b \neq c$	$\alpha = \beta = \gamma = 90^\circ$
菱方晶系	$a = b = c$	$\alpha = \beta = \gamma \neq 90^\circ$
单斜晶系	$a \neq b \neq c$	$\gamma \neq \alpha = \beta = 90^\circ$
三斜晶系	$a \neq b \neq c$	$\alpha \neq \beta \neq \gamma \neq 90^\circ$
六方晶系	$a = b \neq c$	$\alpha = \beta = 90^\circ, \gamma = 120^\circ$

晶胞示例:

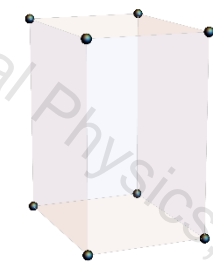
立方晶系



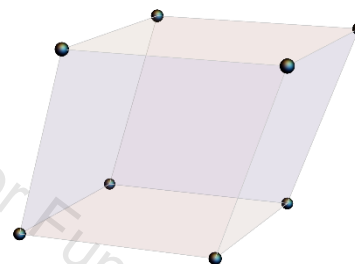
四方晶系



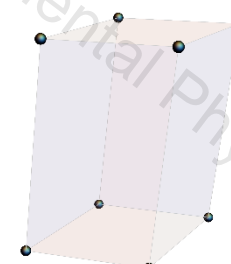
正交晶系



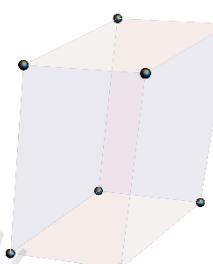
菱方晶系



单斜晶系



三斜晶系



六方晶系



常使用高斯分布对晶胞参数进行估计

$$\begin{cases} \text{Edge} \sim \text{Gaussian}(\mu_1, \sigma_1) \\ \text{Angle} \sim \text{Gaussian}(\mu_2, \sigma_2) \end{cases}$$

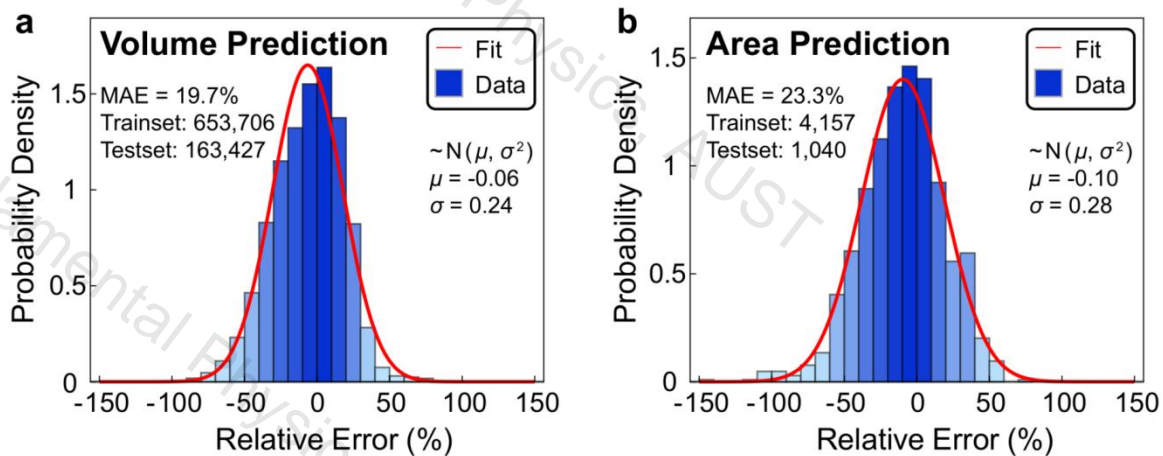
$$\text{Len}_\mu = 5, \text{Len}_\sigma = 1$$

$$\text{Ang}_\mu = 90, \text{Ang}_\sigma = 15$$

根据空间群所属晶系生成晶胞

晶胞生成方法 --- 体积或面积约束

体积或面积预测误差分布:



$$\hat{V}_{opt} = A \sum r_i + B \sum X_i + C \sum E a_i + D$$

$$A = 17.9974 \quad B = -3.1113 \quad C = 4.5057 \quad D = -8.3959$$

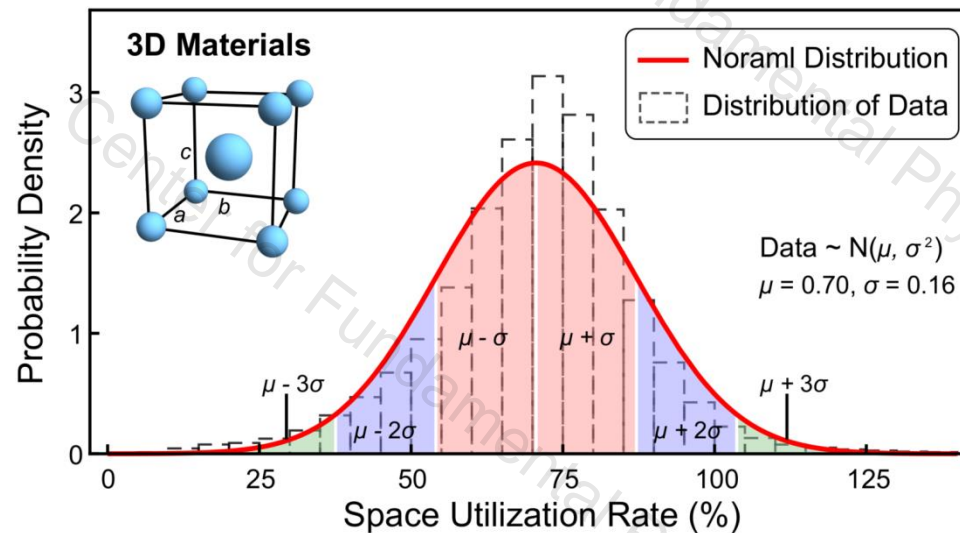
2倍置信区间

$$\frac{1}{1 + \mu + 2\sigma} \cdot \hat{V}_{opt} \leq V_{opt} \leq \frac{1}{1 + \mu - 2\sigma} \cdot \hat{V}_{opt}$$

联立获得

$$\max(\hat{V}_{min}, \tilde{V}_{min}) \leq V_{opt} \leq \min(\hat{V}_{max}, \tilde{V}_{max})$$

空间占有率分布:



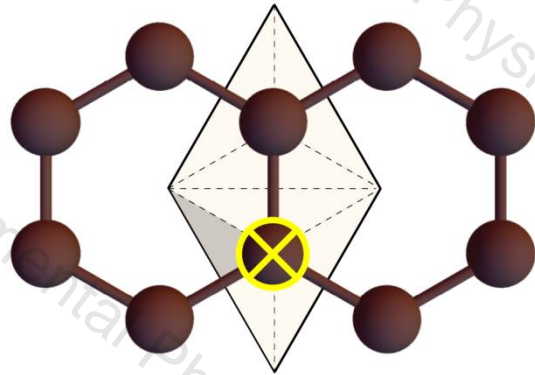
$$UR = \sum V_i / V_{lattice} \text{ or } \sum S_i / S_{lattice}$$

2倍置信区间

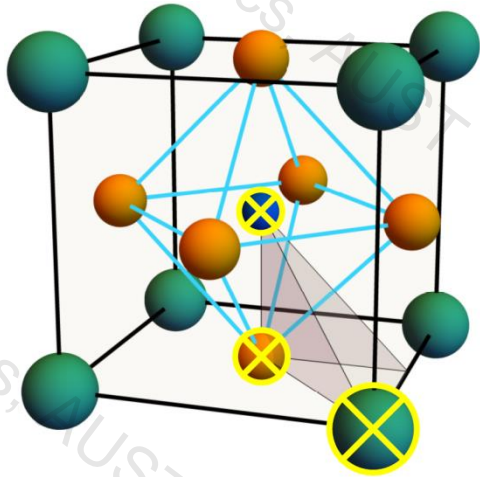
$$\frac{1}{\mu + 2\sigma} \cdot \sum V_i \leq V_{opt} \leq \frac{1}{\mu - 2\sigma} \cdot \sum V_i$$

2.1 简化晶体结构预测问题 --- 对称性：基于DAU的网格划分

Wycoff位点:



2倍位点

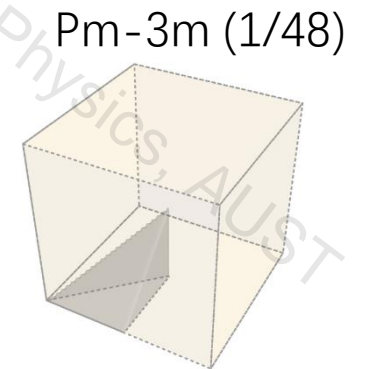
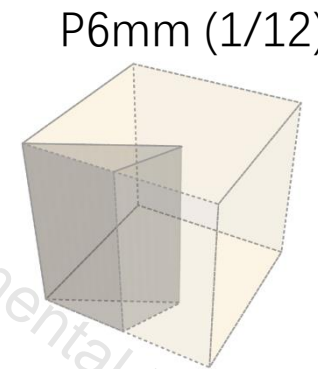
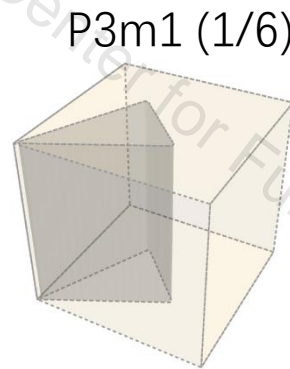
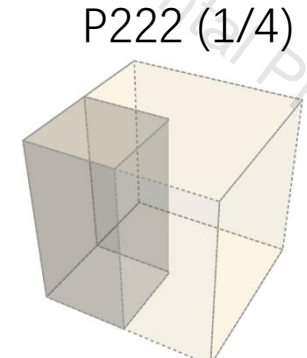
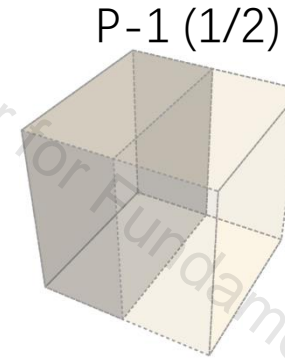
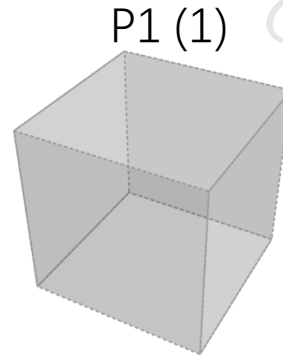


ABX3
1倍位点
3倍位点

利用对称性极大减少自由度

直接不等价单元(direct-space asymetrical unit)

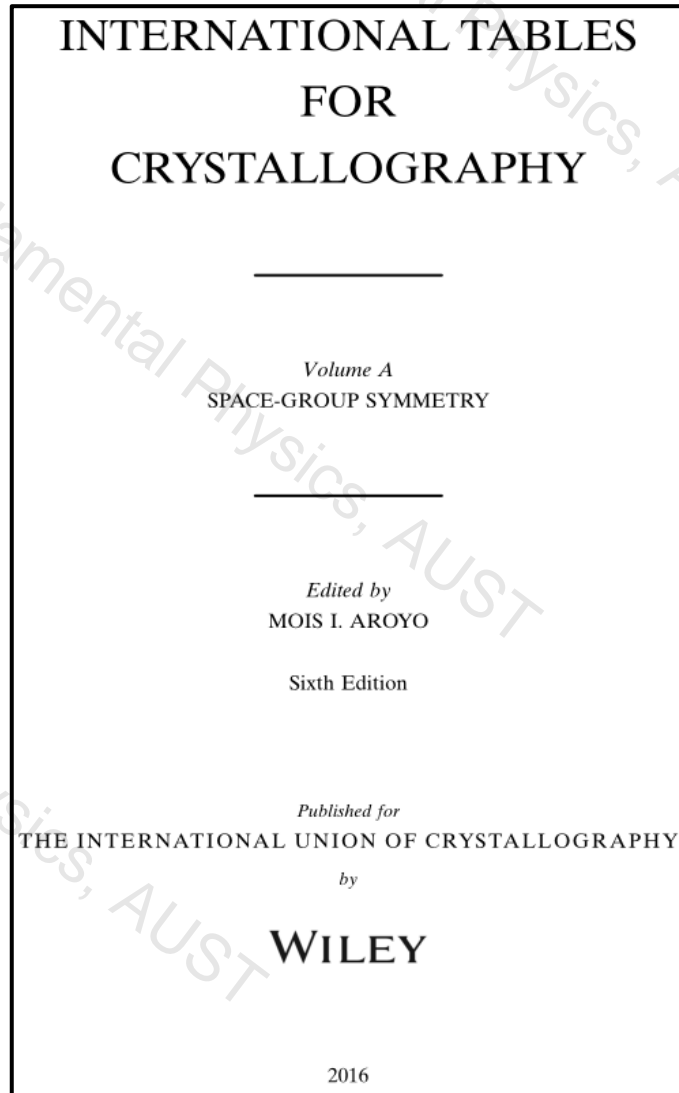
DAU更有利于数据存储和减少近邻计算量



DAU中所有位点独立不重复，通过对称性操作获取原胞中的所有位置

三维空间群的直接不等价单元

国际晶体学表:



不等价单元和Wyckoff位点:

P2

Origin on 2

Asymmetric unit

$$0 \leq x \leq 1; \quad 0 \leq y \leq 1; \quad 0 \leq z \leq \frac{1}{2}$$

→ AU中对结构进行操作更加高效

Symmetry operations

(1) 1 (2) 2 0,y,0

Generators selected (1); $t(1,0,0)$; $t(0,1,0)$; $t(0,0,1)$; (2)

Positions

Multiplicity,
Wyckoff letter,
Site symmetry

Coordinates

Reflection conditions

General:

no conditions

Special: no extra conditions

2	e	1	(1) x, y, z	(2) \bar{x}, y, \bar{z}
1	d	2	$\frac{1}{2}, y, \frac{1}{2}$	
1	c	2	$\frac{1}{2}, y, 0$	
1	b	2	$0, y, \frac{1}{2}$	
1	a	2	$0, y, 0$	

→ 通过Wyckoff位点判定边界, 从而获得DAU

Symmetry of special projections

Along [001] $p1m1$

$\mathbf{a}' = \mathbf{a}_p$ $\mathbf{b}' = \mathbf{b}$

Origin at 0,0,z

Along [100] $p11m$

$\mathbf{a}' = \mathbf{b}$ $\mathbf{b}' = \mathbf{c}_p$

Origin at x,0,0

Along [010] $p2$

$\mathbf{a}' = \mathbf{c}$ $\mathbf{b}' = \mathbf{a}$

Origin at 0,y,0

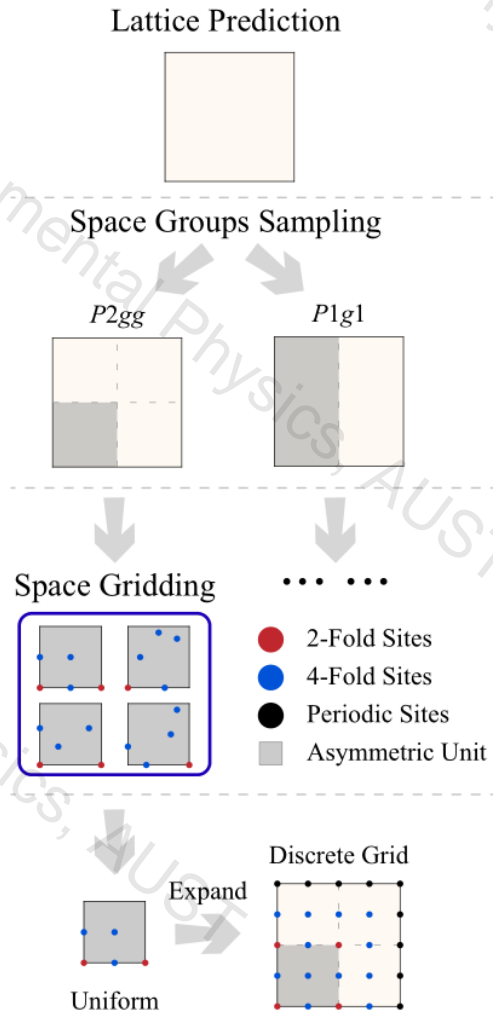
Copyright © 2016 International Union of Crystallography

196

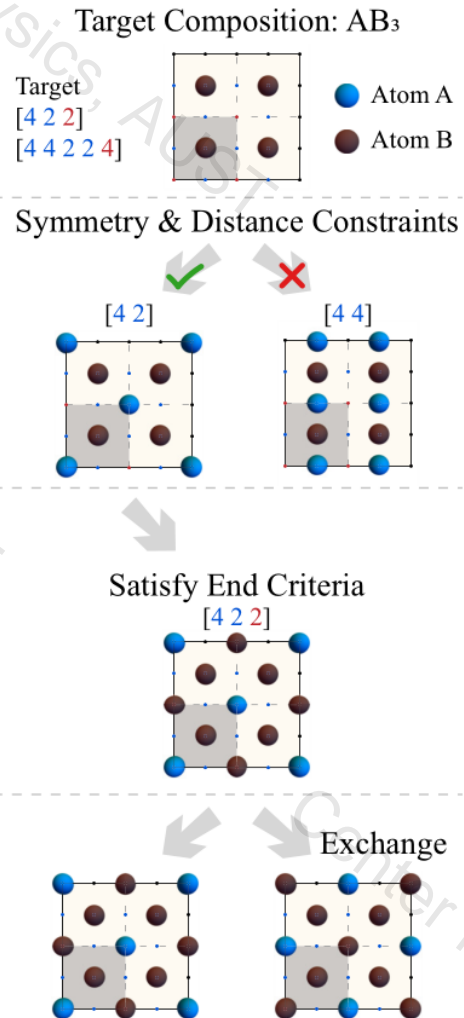
对称性和距离约束的蒙特卡洛树采样

结构采样流程图:

(a) Grid Generation



(b) Structure Sampling

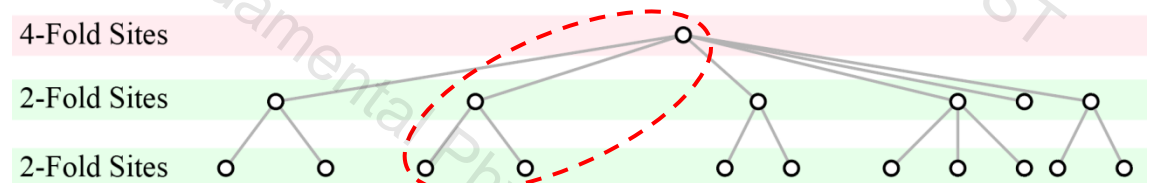


采样流程总结:

1. 均匀划分网格, 减小生成合理结构的难度
2. 根据对称性生成原子分配方案, 例如, $[4,2,2]$
3. 根据分配方案生成树结构, 按照距离约束采样

树结构采样示意图:

树结构随机采样, 采样同时满足对称性和最小距离约束

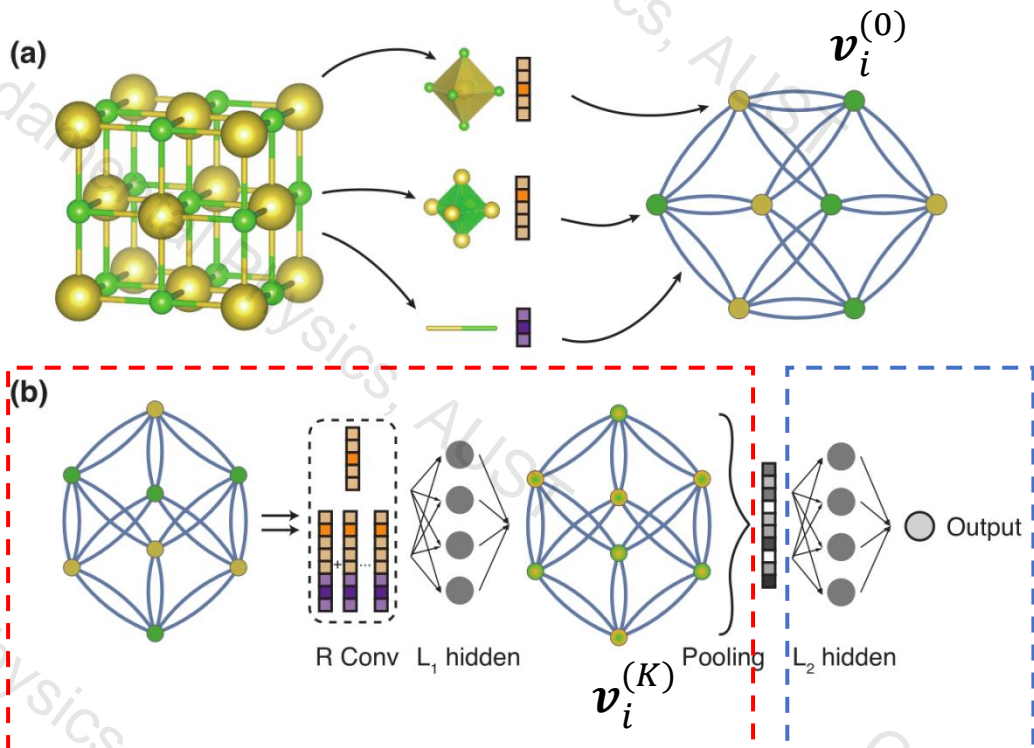


按对称性分配方案 $[4,2,2]$ 和距离约束生成的树结构

2.2 SCCOP使用的机器学习模型

CGCNN示意图:

通过原子间的键连关系构建晶体图



特征提取

能量预测

CGCNN的核心思路:

$$\mathbf{v}_i^{(t+1)} = \mathbf{v}_i^{(t)} + \Delta \mathbf{v}_i^{(t)} \quad \longrightarrow \quad \text{迭代修正原子向量}$$



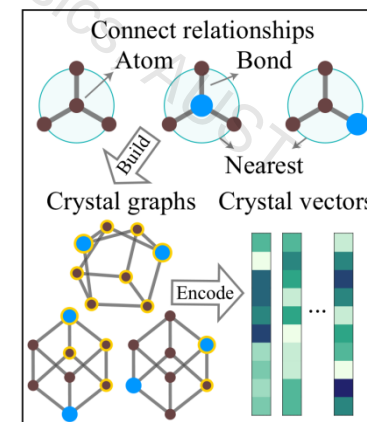
为相互作用赋予权重

$$\mathbf{v}_i^{(t+1)} = \mathbf{v}_i^{(t)} + \sum_{j,k} \underbrace{\sigma(\mathbf{z}_{(i,j)}^{(t)} \mathbf{W}_f^{(t)} + \mathbf{b}_f^{(t)})}_{\text{权重向量}} \odot \underbrace{g(\mathbf{z}_{(i,j)}^{(t)} \mathbf{W}_s^{(t)} + \mathbf{b}_s^{(t)})}_{\text{相互作用向量}}$$

SCCOP-DAUGNN:

直接不等价单元中原子的键连关系与对称区域中原子的键连关系相同。

模型在直接不等价单元上构建, **最大效率提升为空间群的最大对称性。**

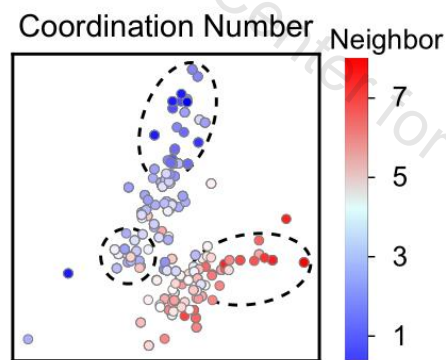
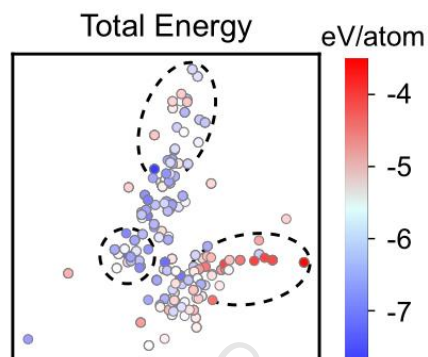
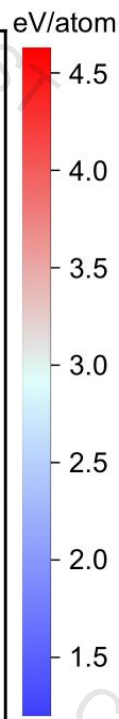
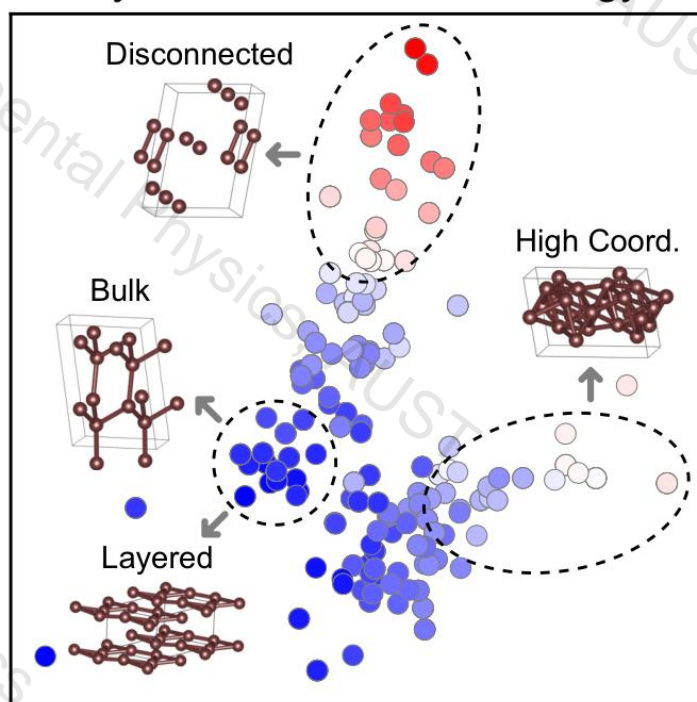


2.3 基于晶体向量的势能面划分

势能面划分:

随机初始化的机器学习模型

Crystal Vector and ML Energy



关键点:

1. 结构根据**键连关系聚类**, 每个簇内部的结构**能量相近**

➤ 选择代表性结构构成训练集

2. **随机初始化**的机器学习模型**并不影响势能面划分**

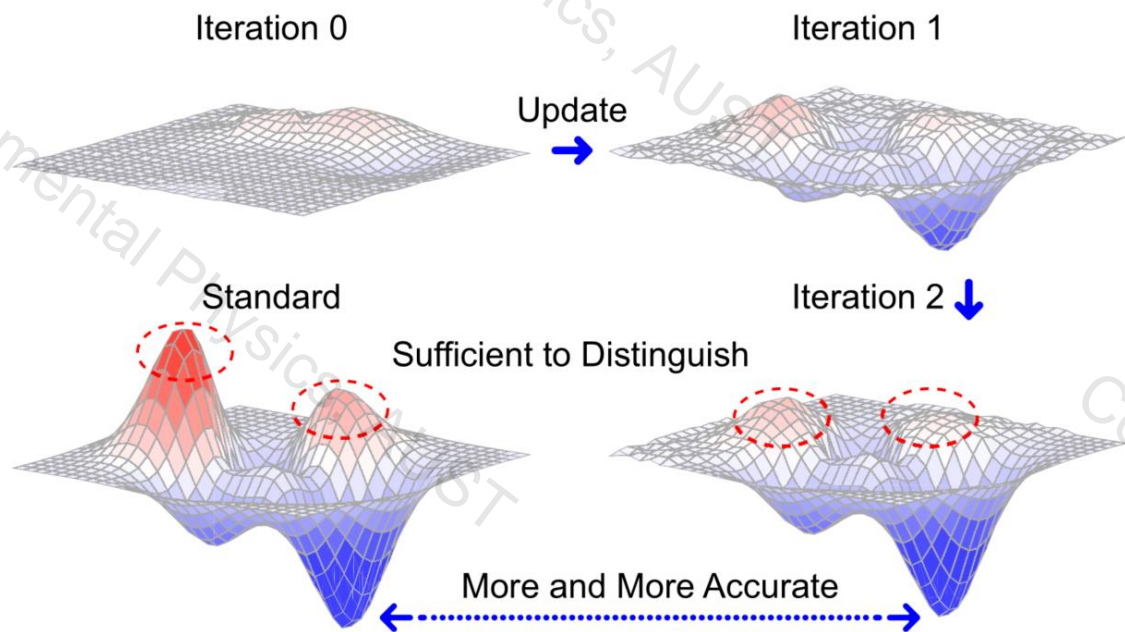
➤ 可以通过多次**迭代势能面划分**, 获得**越来越准确的势能面**

关键点1、2使得SCCOP并不依赖模型预训练

SCCOP核心方法总结

基于DAUGNN的势能面划分:

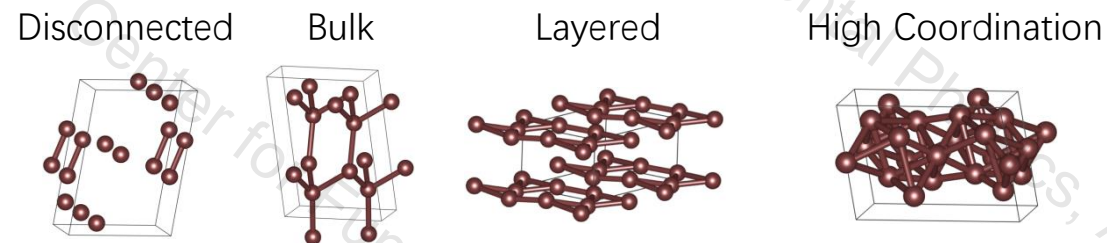
Identify Low Energy Area on PES



1. 通过DAUGNN将采样结构转化为晶体向量;
2. 使用KernelPCA对晶体向量降维;
3. 利用KMeans对降维后的晶体向量聚类。

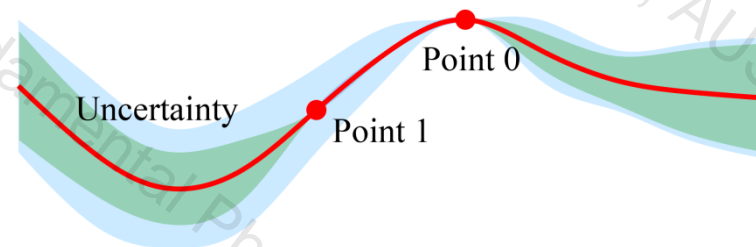
核心方法:

1. 利用图神经网络DAUGNN划分势能面



选择**代表性结构**形成数据集

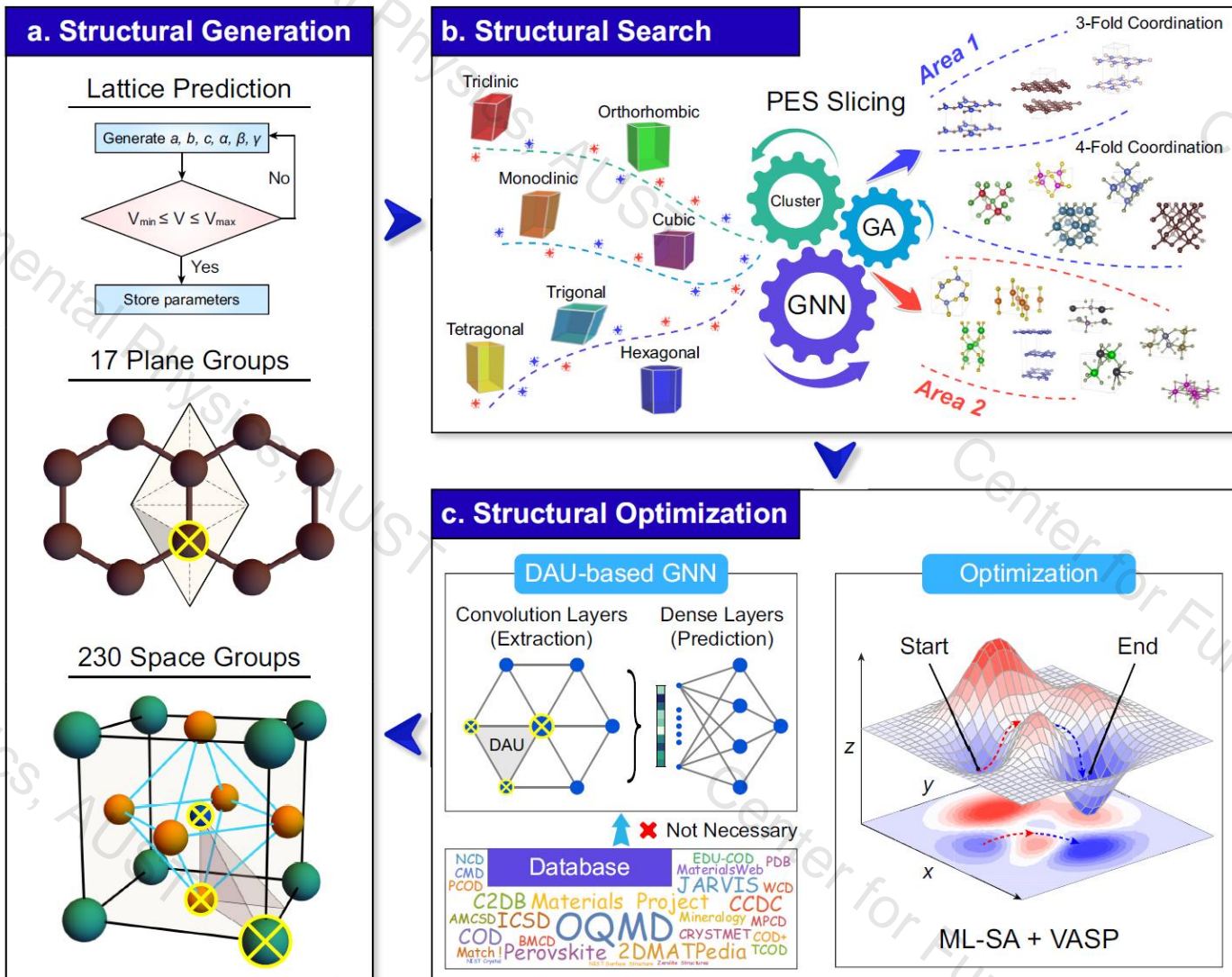
2. 利用模拟退火分区域探索势能面



减少模型搜索过程中的**模型预测不确定性**

2.4 SCCOP搜索流程

SCCOP流程图:



方法创新:

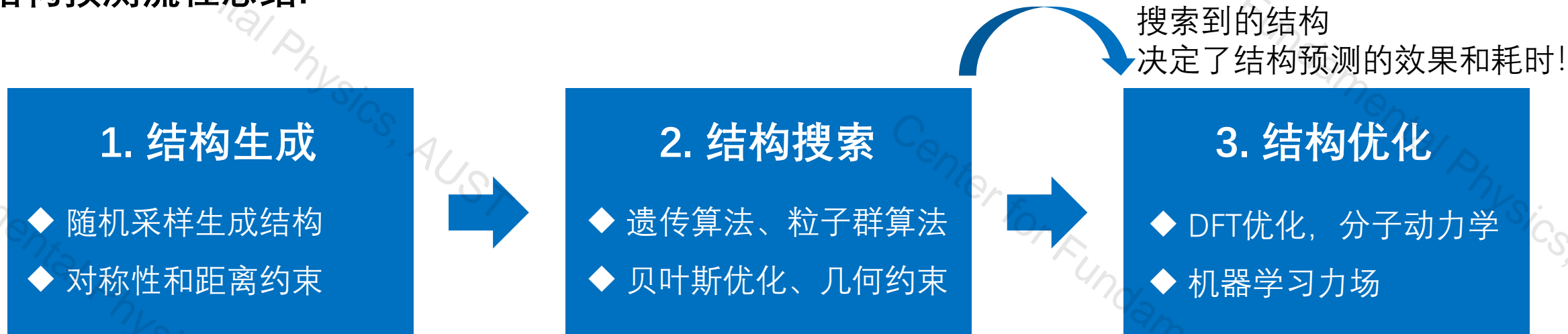
1. 基于**直接不等价单元**的树采样结构生成
2. 基于图神经网络的**势能面划分**
3. 基于图神经网络的**分区域模拟退火**
4. **不依赖数据库训练模型**的晶体结构预测

SCCOP特点:

1. 支持**全空间群搜索**
2. 支持自定义**模板/结构种子搜索**
3. 支持**团簇结构搜索**
4. 支持**大规模/超大规模并行**

晶体结构预测方法总结

结构预测流程总结:



SCCOP

- ◆ 机器学习快速筛选结构
- ◆ 根据晶体向量对结构分类

实现势能面划分
和低能区域定位

晶体结构预测算法总结

- USPEX = 随机采样+遗传算法+DFT
- CALYPSO = 随机采样 + 粒子群算法+DFT
- MAGUS = 随机采样+遗传算法+ML力场
- GN-OA = 随机采样+图神经网络

★ 充分利用ML的两大优势: 1. 快速且保持一定精度, 2. 特征提取

3. 结果与讨论

3.1 SCCOP框架有效性验证 1 --- 固定位点

固定位点搜寻低能原子排列:

基本设置:

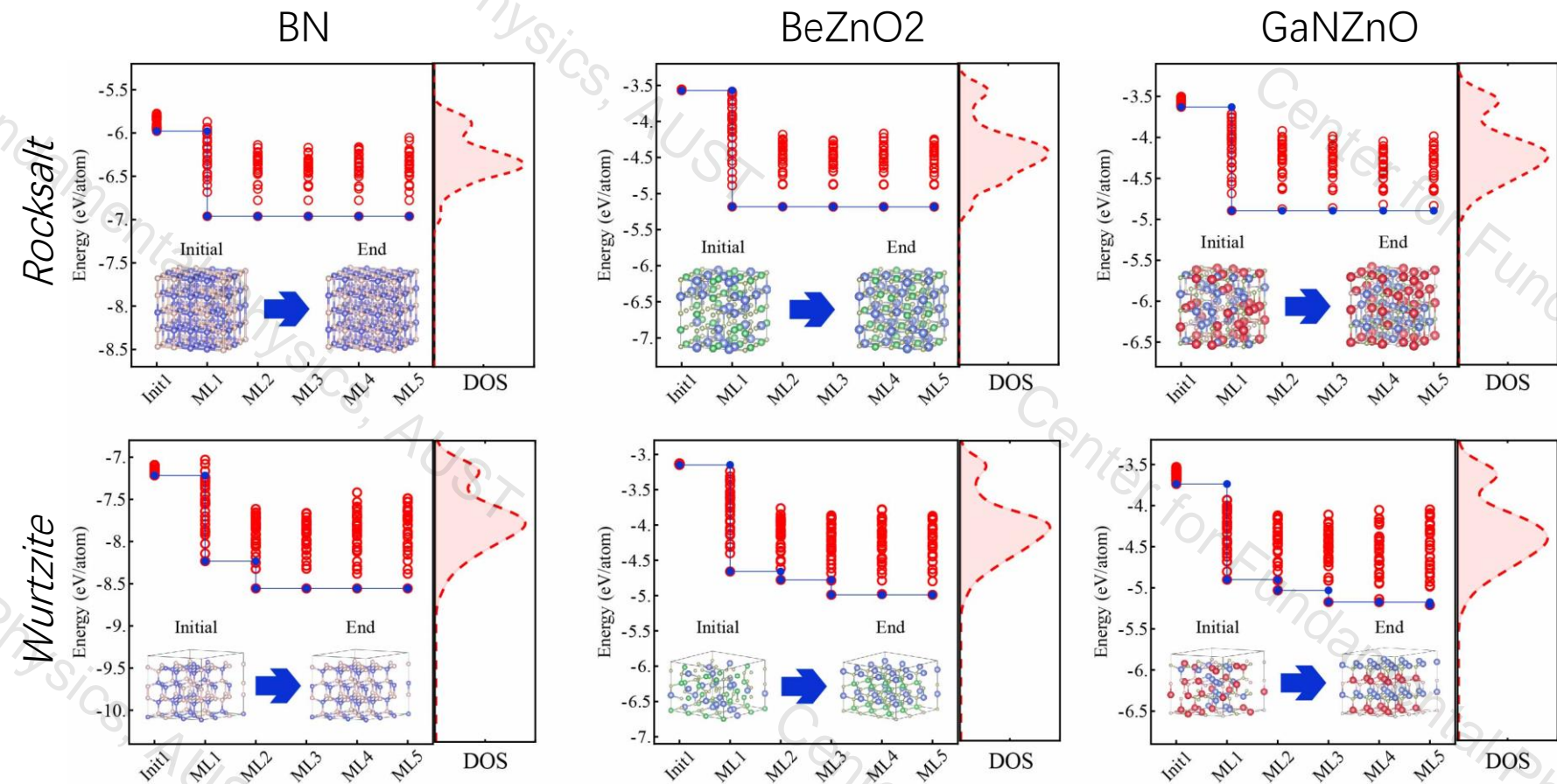
- 每轮30个结构计算单点能
- ML模型优化原子排序5轮

基本流程:

- 结构采样
- 结构搜索 (只使用ML搜索)
- 结构优化 (只使用ML优化)

验证结论:

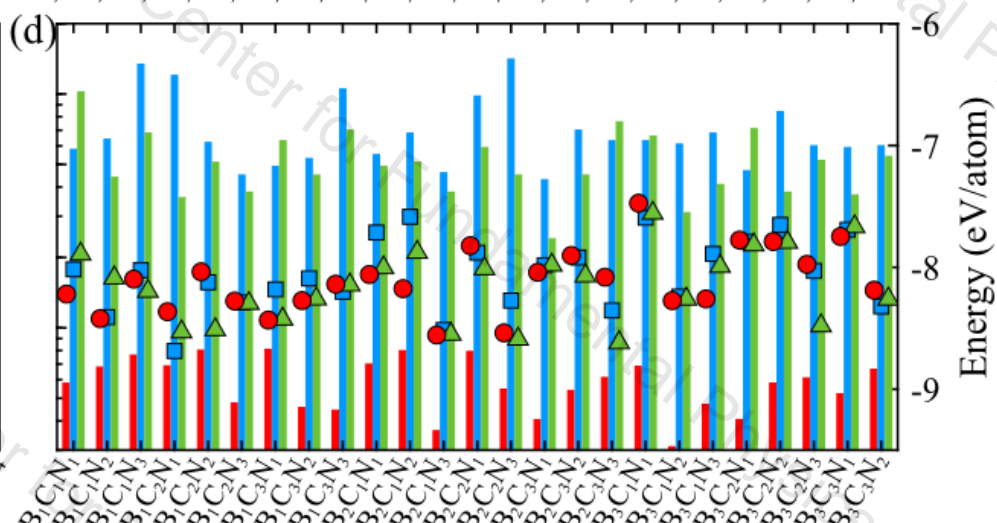
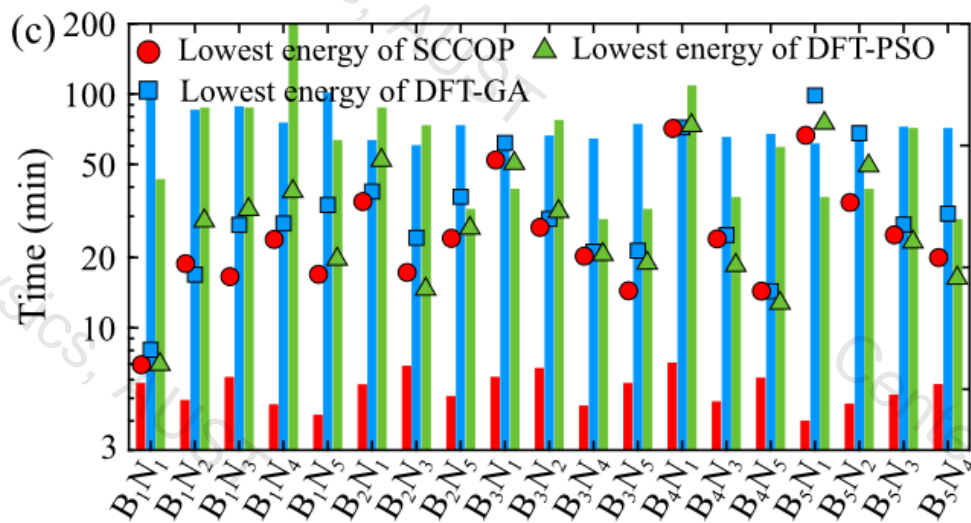
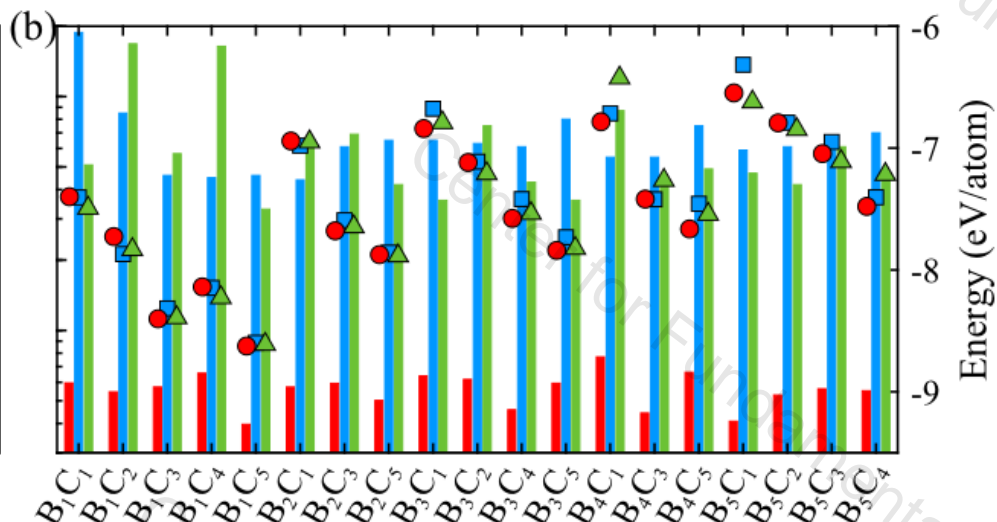
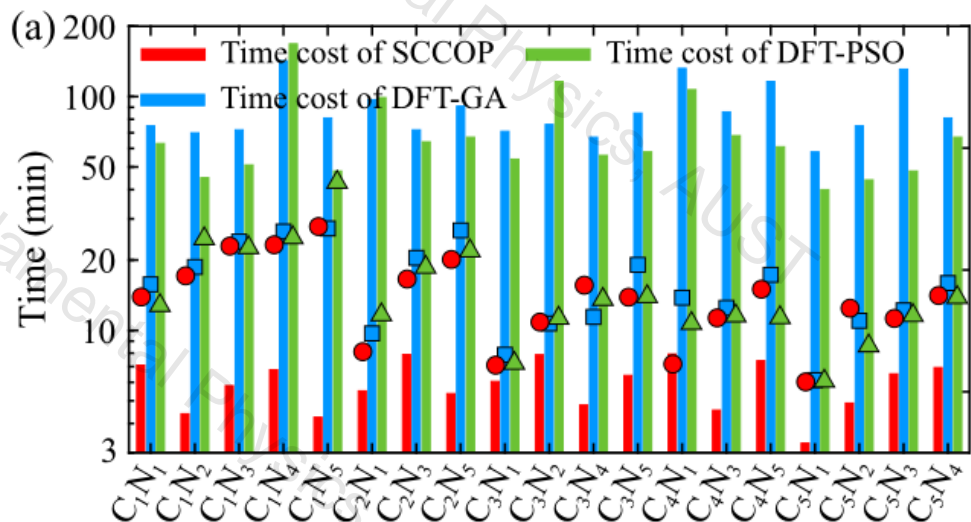
- 选取的ML模型与晶体适配
- ML加速结构预测是可行的



SCCOP框架有效性验证 2 --- 二维材料

BCN二维体系测试:

新发现了28个动力学稳定的二维BCN材料, 包含五种力学性能优异宽带隙材料



最低能量次数
(总共82个组分)

SCCOP: 46

CALYPSO: 37

USPEX: 11

大约十倍加速!

平均耗时(分钟)

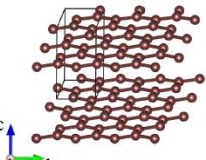
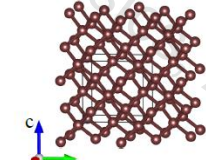
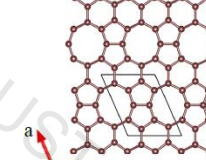
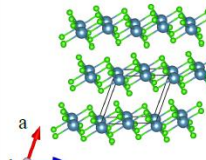
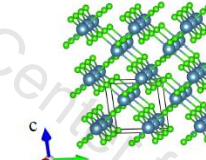
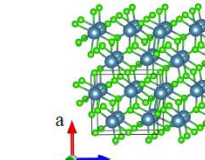
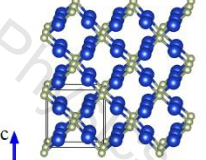
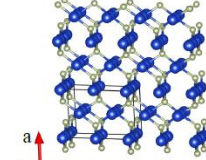
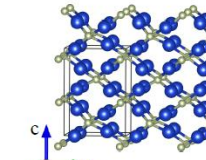
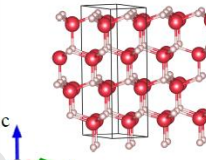
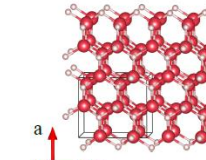
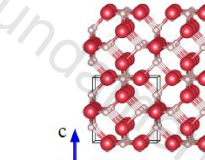
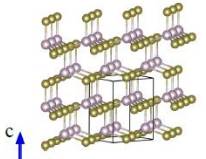
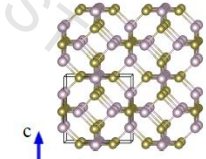
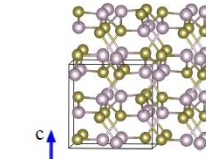
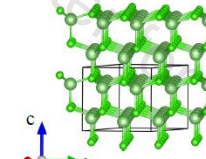
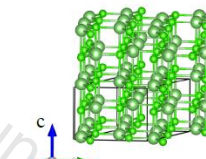
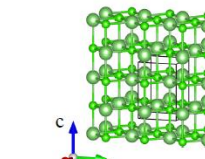
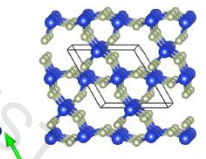
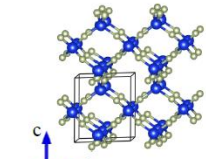
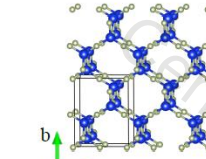
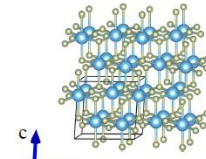
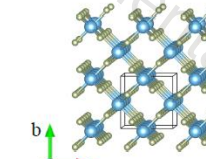
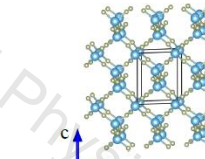
SCCOP: 5.7

CALYPSO: 62

USPEX: 75

SCCOP框架有效性验证 3 --- 三维材料

典型三维材料测试:

C	$R\bar{3}m$ (-9.230) 	$Fd\bar{3}m$ (-9.097) 	$Cmmm$ (-9.024) 	CaCl ₂	$R\bar{3}m$ (-4.275) 	$P2_1$ (-4.271) 	$P2_1/c$ (-4.268) 
Cu ₂ O	$Pn\bar{3}m$ (-4.546) 	Pc (-4.533) 	$P2_1$ (-4.484) 	GaN	$P6_3mc$ (-6.083) 	$Pbca$ (-6.026) 	$Pm\bar{3}n$ (-5.946) 
HgTe	$P6_3mc$ (-1.815) 	$Pm\bar{3}$ (-1.784) 	$P2_13$ (-1.744) 	LiCl	$P6_3mc$ (-3.716) 	$P6_3/m$ (-3.695) 	$Fm\bar{3}m$ (-3.694) 
SiO ₂	$P3_121$ (-7.908) 	$P6_222$ (-7.897) 	$Ima2$ (-7.887) 	TiO ₂	$I4_1/amd$ (-8.978) 	$P\bar{4}2_1m$ (-8.949) 	$P6_422$ (-8.932) 

维度: 3

空间群: 1-230

原胞中原子数: 5-10

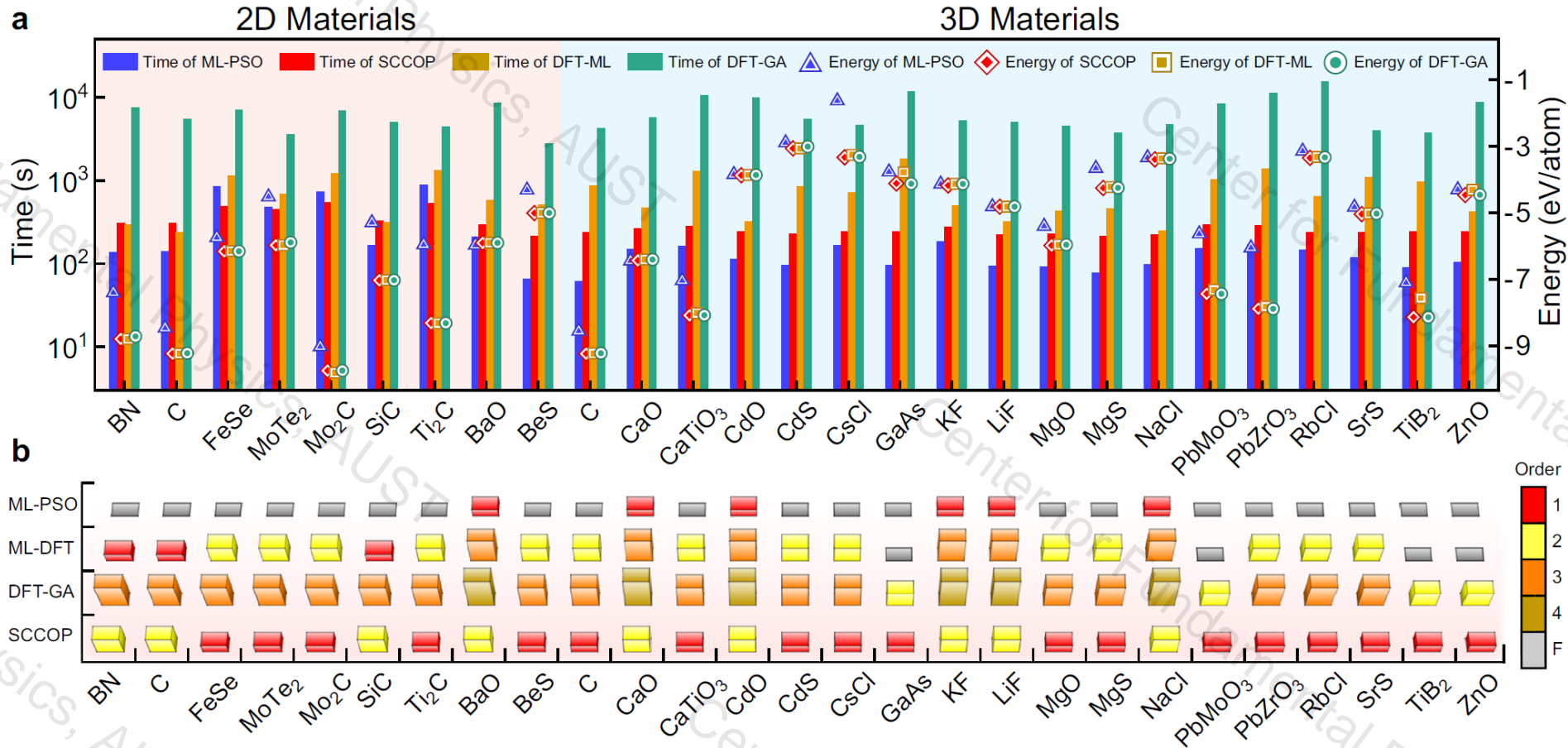
CPU核: 6*48 cores

耗时: ~5 minutes

SCCOP找到了Material
Project数据库中所有
最低能结构

SCCOP框架有效性验证 3 --- 三维材料

不同方法的时间和精度比较:



ML最快，但精度有限
 DFT最慢，但精度最高
 SCCOP取两者优点
 保持DFT精度同时大幅
 缩短时间

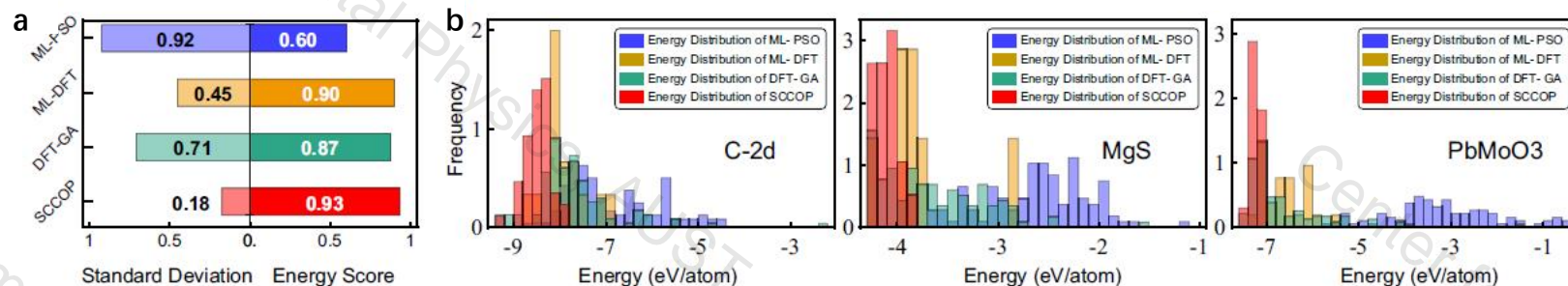
SCCOP显著优于ML-DFT

1. 有效的ML模型
2. 高效的势能面分割

SCCOP在**所有组分**中均找到了**最低能量**，并在**17个组分**中取得了**最短时间**

SCCOP框架有效性验证 3 --- 三维材料

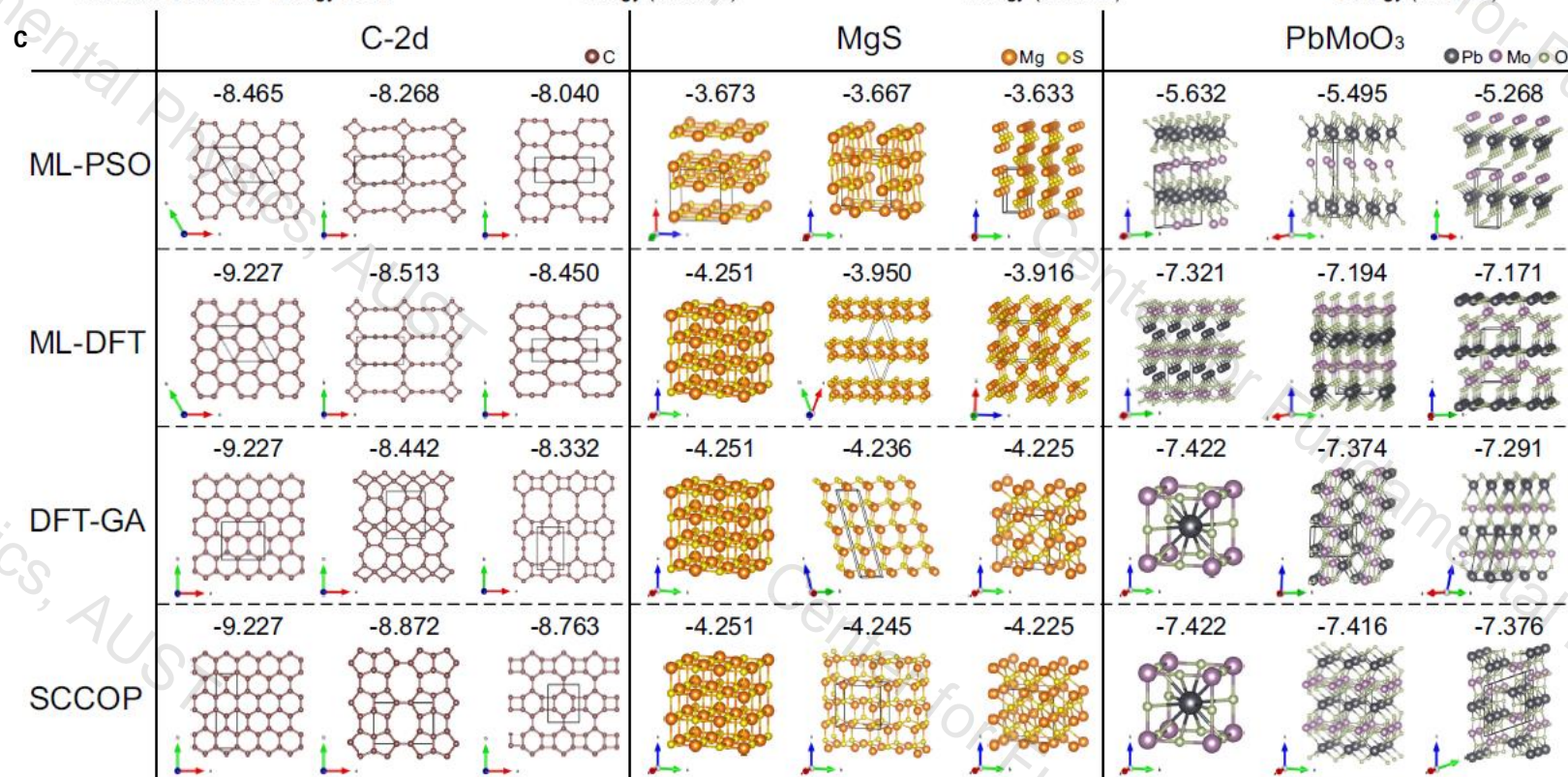
不同方法的低能结构多样性比较:



SCCOP搜索到更丰富的低能构型

具体表现为: 1. **平均能量低**

2. **能量分布集中**

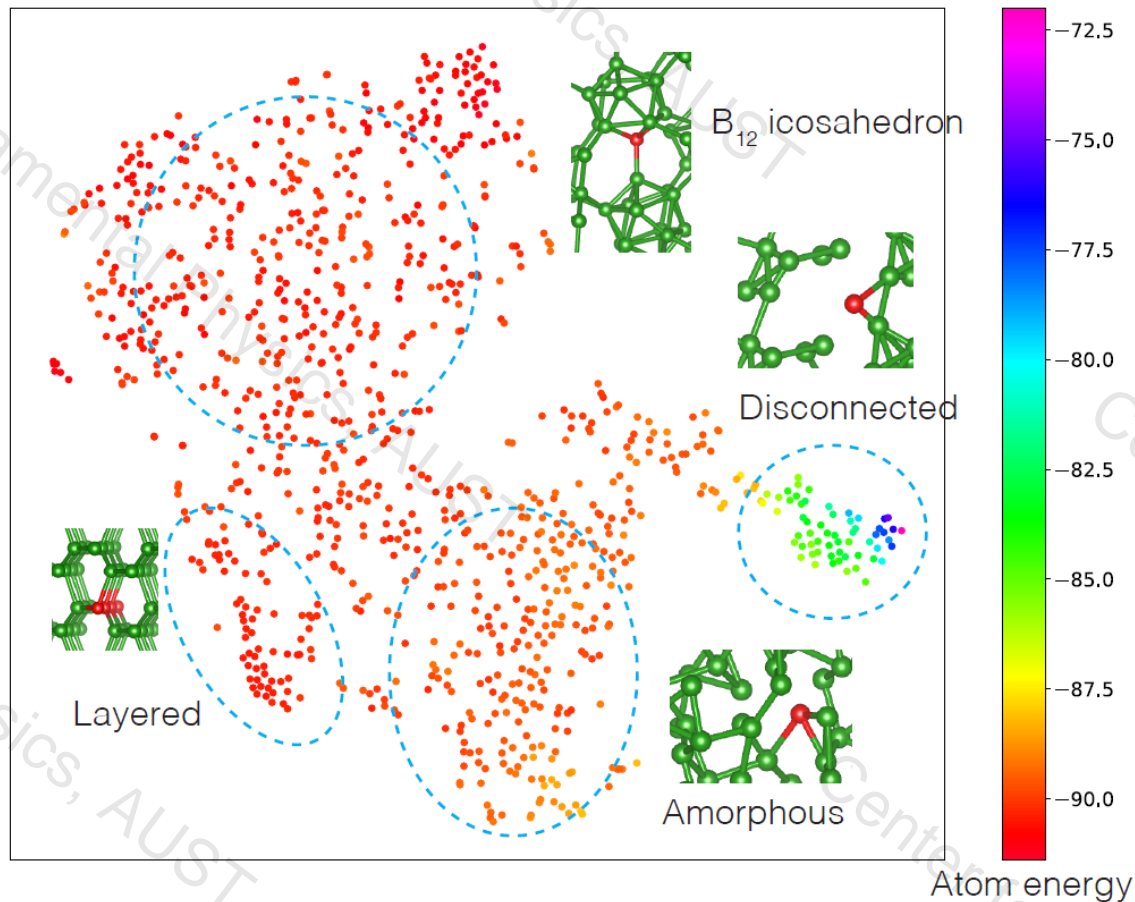


ML-PSO、ML-DFT、DFT-GA、SCCOP在C-2d, MgS和PbMoO₃搜索到的低能结构

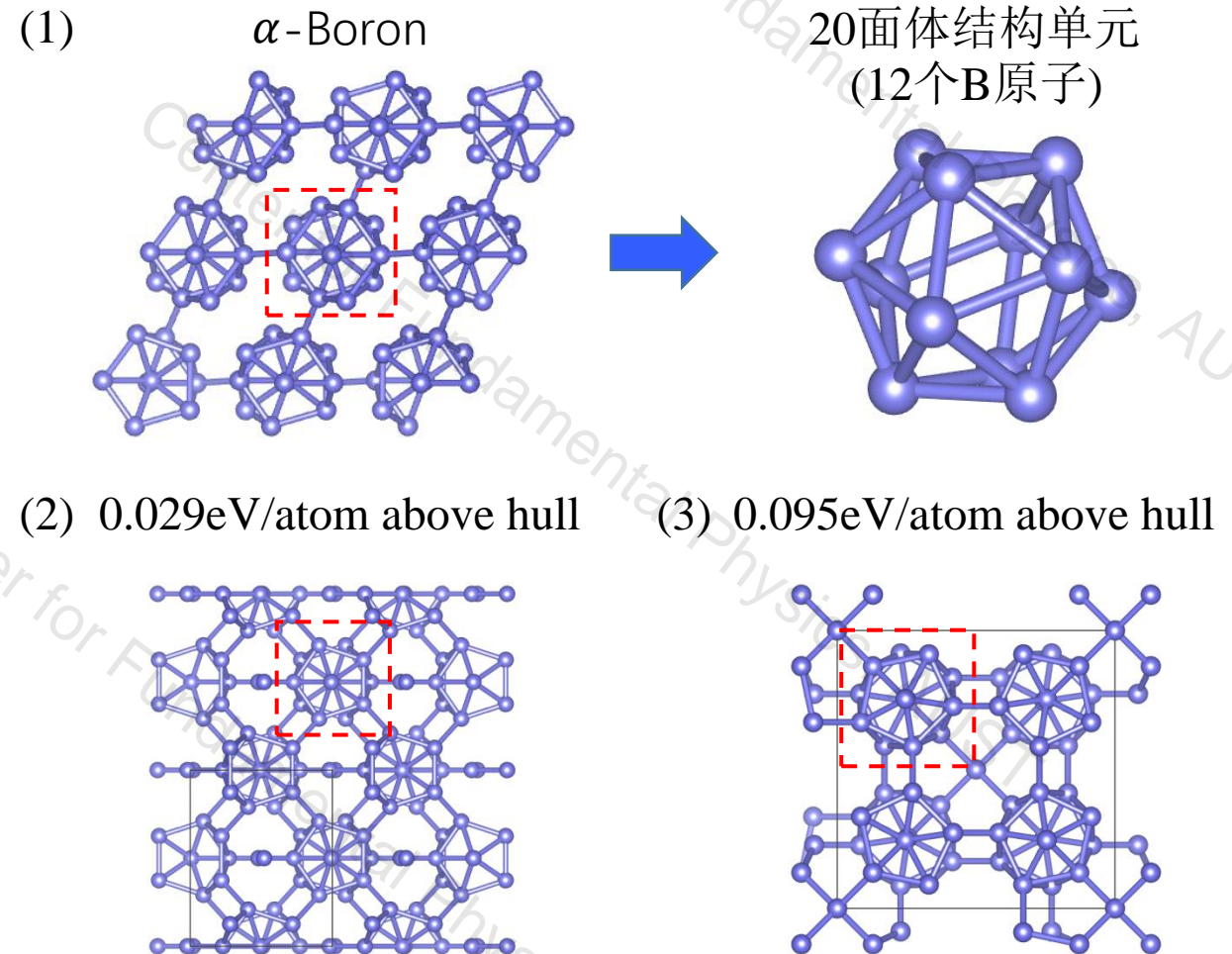
3.2 Boron团簇搜索 --- 背景介绍

Boron复杂的化学环境:

Boron的缺电子特性使其具有复杂的成键



Materials Project中的低能硼结构:



结论: 二十面体作为结构单元广泛存在于低能结构中

Boron团簇搜索 --- 方法介绍

团簇搜索方法:

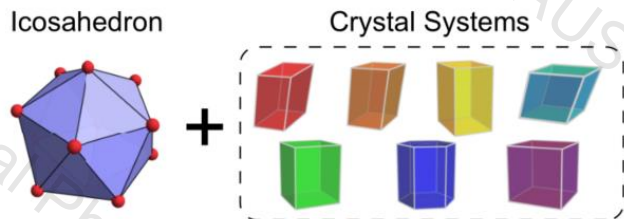
可广泛应用于包含团簇或分子的结构预测

例如：太阳能电池材料MAPbI3中有机分子的朝向

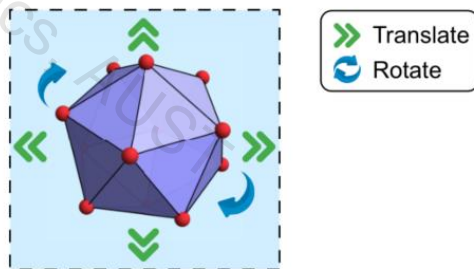


a Generate Grid for Cluster Search

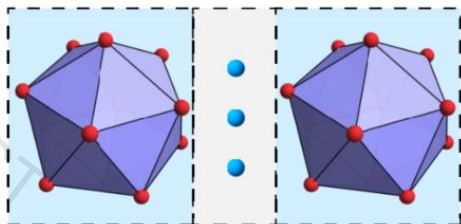
1. Put Boron Cluster into Lattice



2. Rotate and Translate Cluster

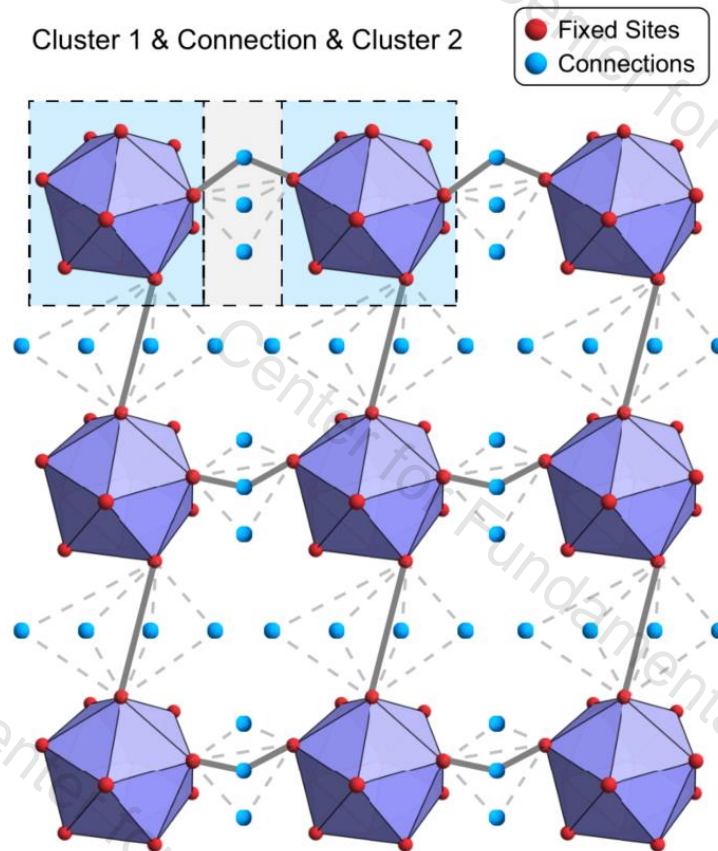


3. Add Connection Points



b Change Connection between Clusters

Cluster 1 & Connection & Cluster 2



流程总结:

整体优化

1. 构建团簇;
2. 优化团簇相对位置。

局部优化

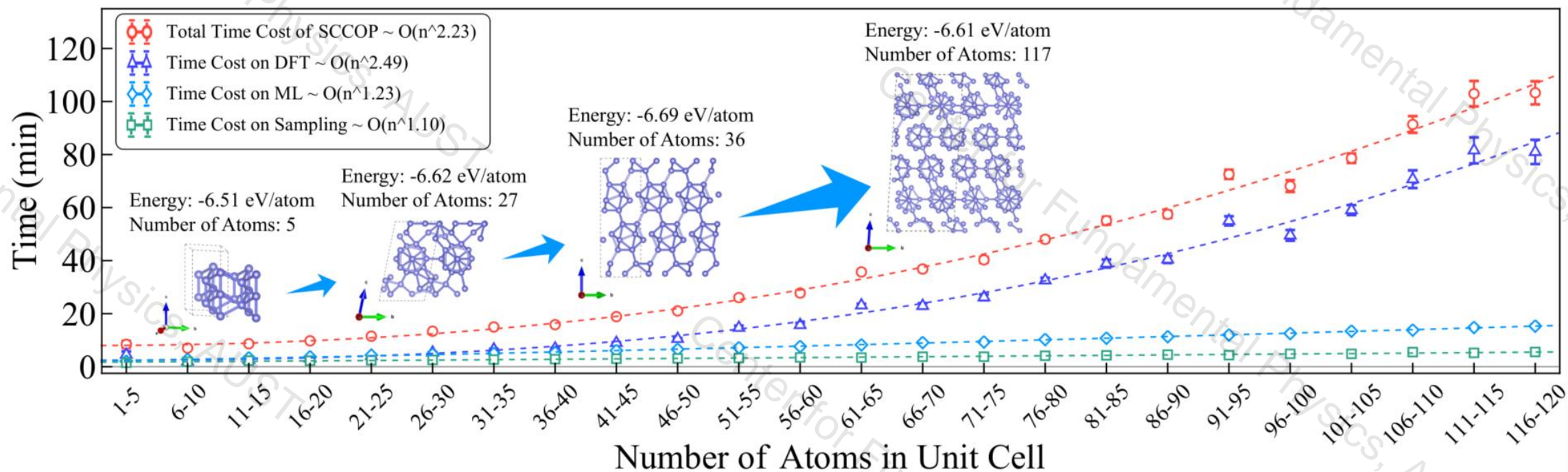
1. 团簇间添加连接点;
2. 优化团簇间连接。

独立变量数目:

$$\begin{array}{ccc} \boxed{12 \times 3} & \longrightarrow & \boxed{3 + 3} \\ \downarrow & & \downarrow \\ \text{平动} & & \text{平动+转动} \end{array}$$

Boron团簇搜索 --- SCCOP耗时和搜索能力测试

SCCOP应用于硼1-120原子的搜索耗时:



USPEX-ML耗时:

12原子: 3.4小时
能量: -6.69 eV/atom

106原子: 24.3小时
能量: -6.69 eV/atom

SCCOP耗时:

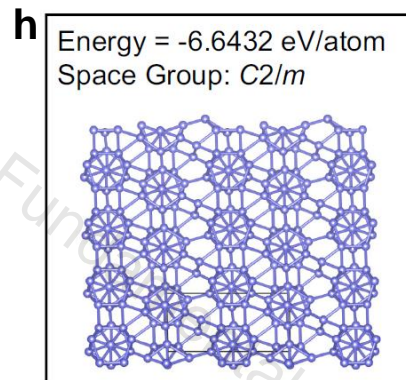
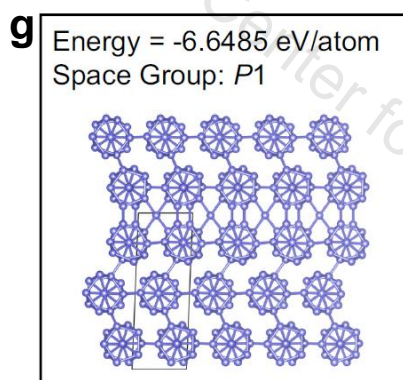
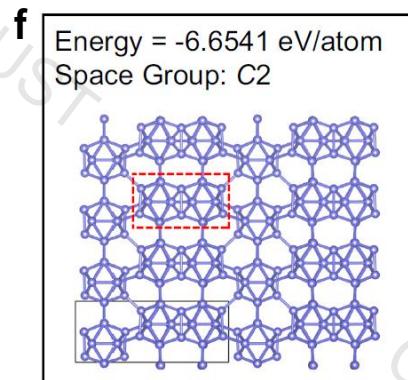
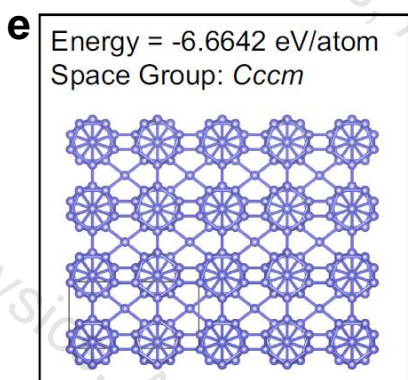
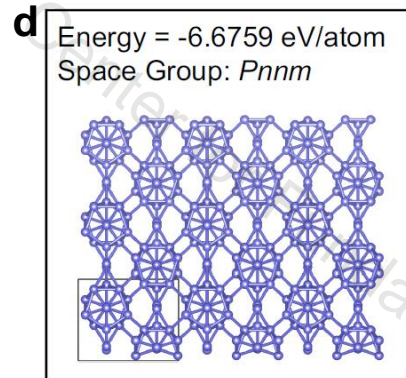
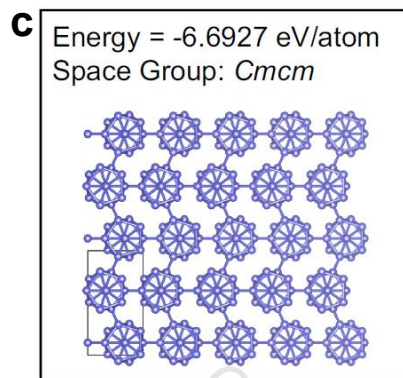
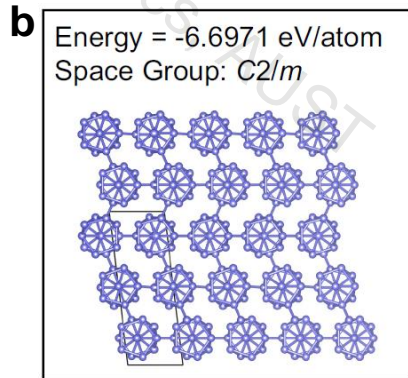
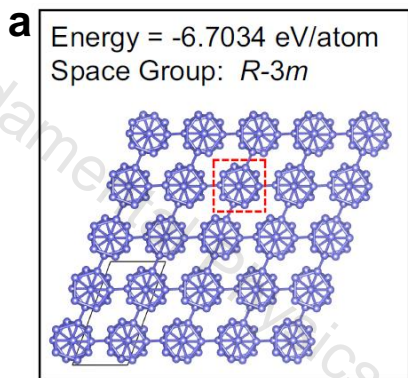
11-15原子: 0.14小时
能量: -6.69 eV/atom

106-110原子: 1.29小时
能量: -6.69 eV/atom

Boron团簇搜索 --- 低能结构

SCCOP搜索到的硼团簇低能结构:

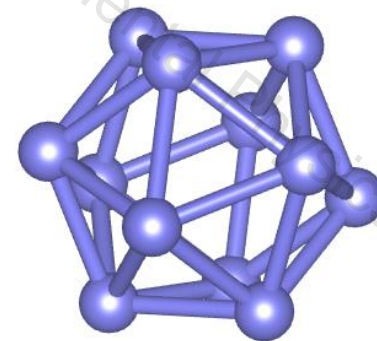
α -Boron



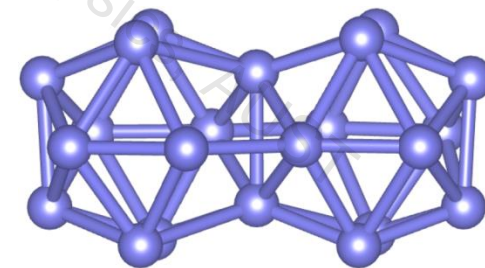
新发现的团簇单元 B_{22}

新发现的 B_{22} 单元:

正20面体 (B_{12})



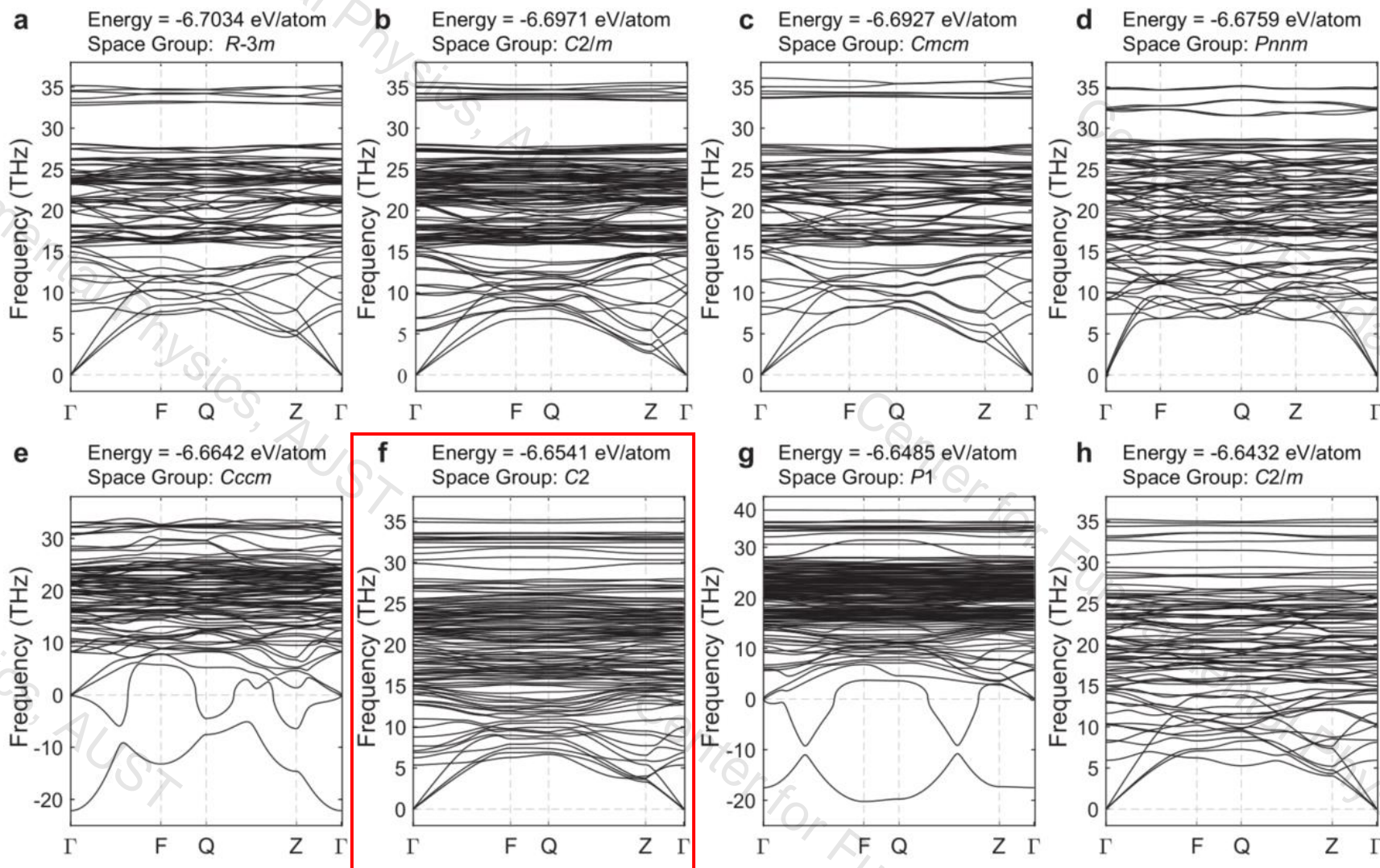
新发现的团簇 B_{22}



通过两个 B_{22} 共享一对原子形成

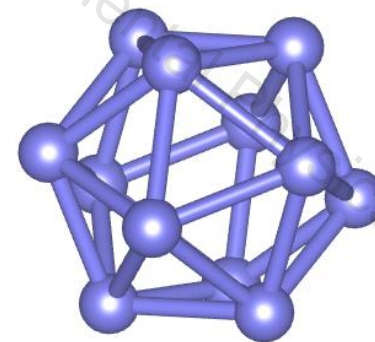
Boron团簇搜索 --- 声子谱计算

低能硼团簇的声子谱:

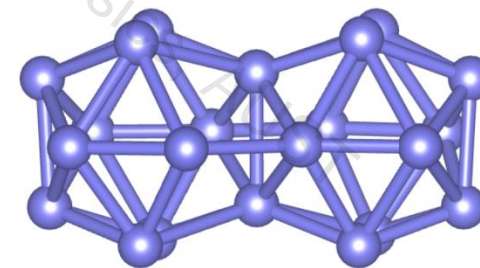


新发现的 B_{22} 单元:

正20面体 (B_{12})



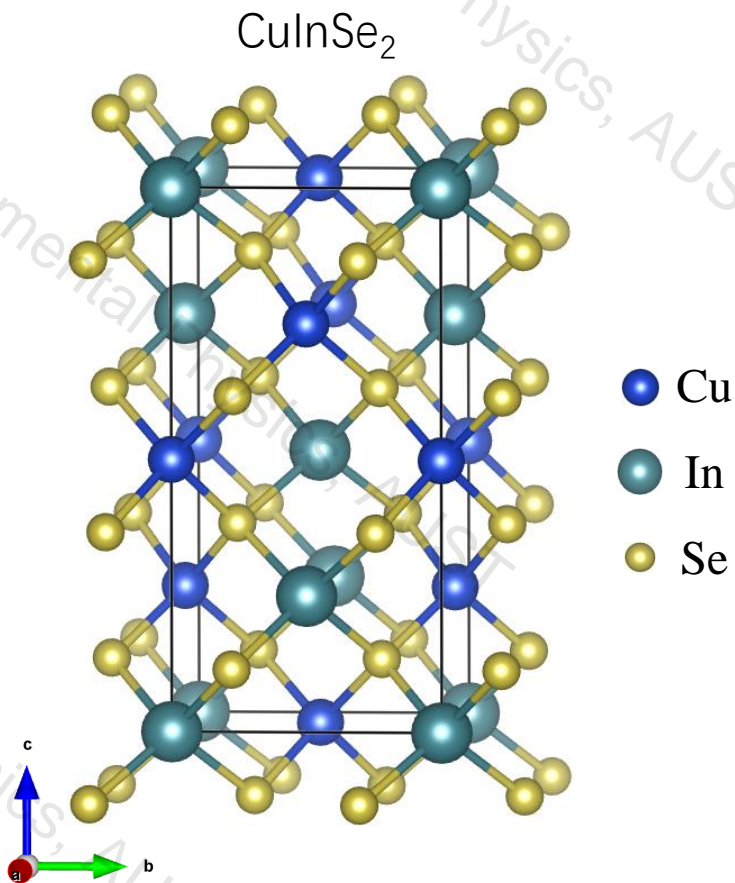
新发现的团簇 B_{22}



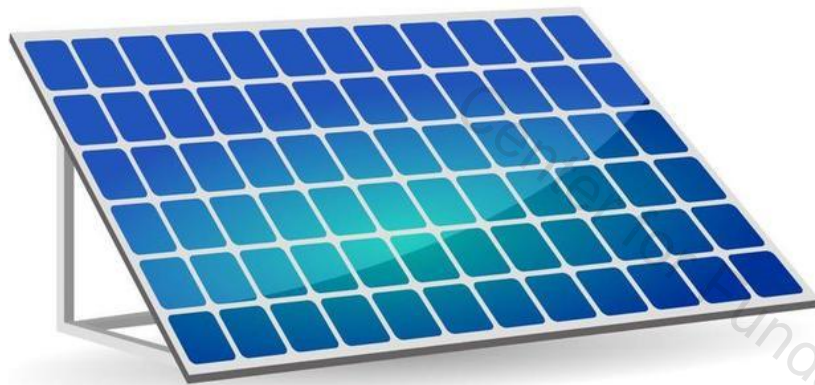
通过两个 B_{22} 共享一对原子形成

3.3 CuIn₅Se₈结构搜索 --- 背景介绍

Chalcopyrite结构:



太阳能电池材料:



CuInSe₂: 直接带隙、高吸收系数
使其作为合适的太阳能薄膜材料

空位组分:

实验中存在许多空位组分 → 较低形成能

CuInSe₂ → **CuIn₅Se₈** CuIn₃Se₅ Cu₂In₄Se₇

Cu: 1+

In: 3+

Se: 2-

1个In替换3个Cu, 产生2个空位

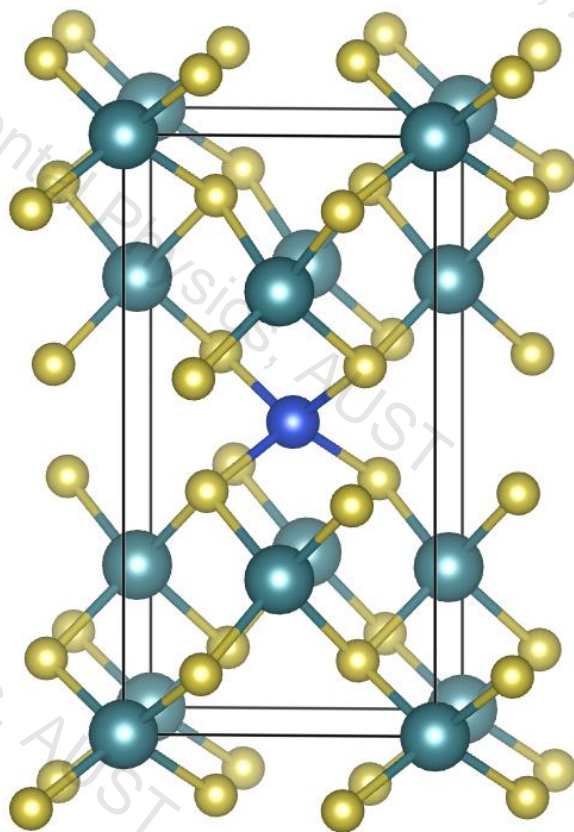
系统性的结构研究有助于理解空位化合物的形成

CuIn₅Se₈结构搜索 --- 低能结构

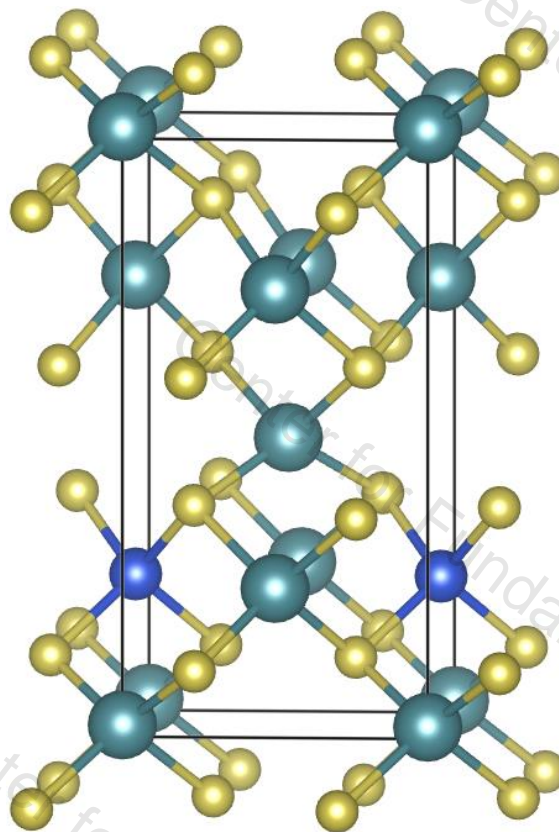
CuIn₅Se₈低能结构:

假定晶格不变进行低能结构搜索

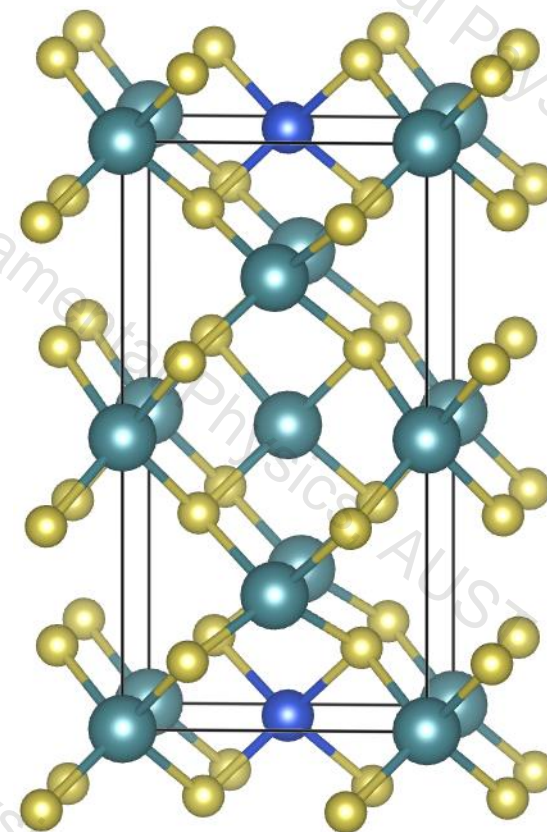
-3.6827 eV/atom



-3.6790 eV/atom



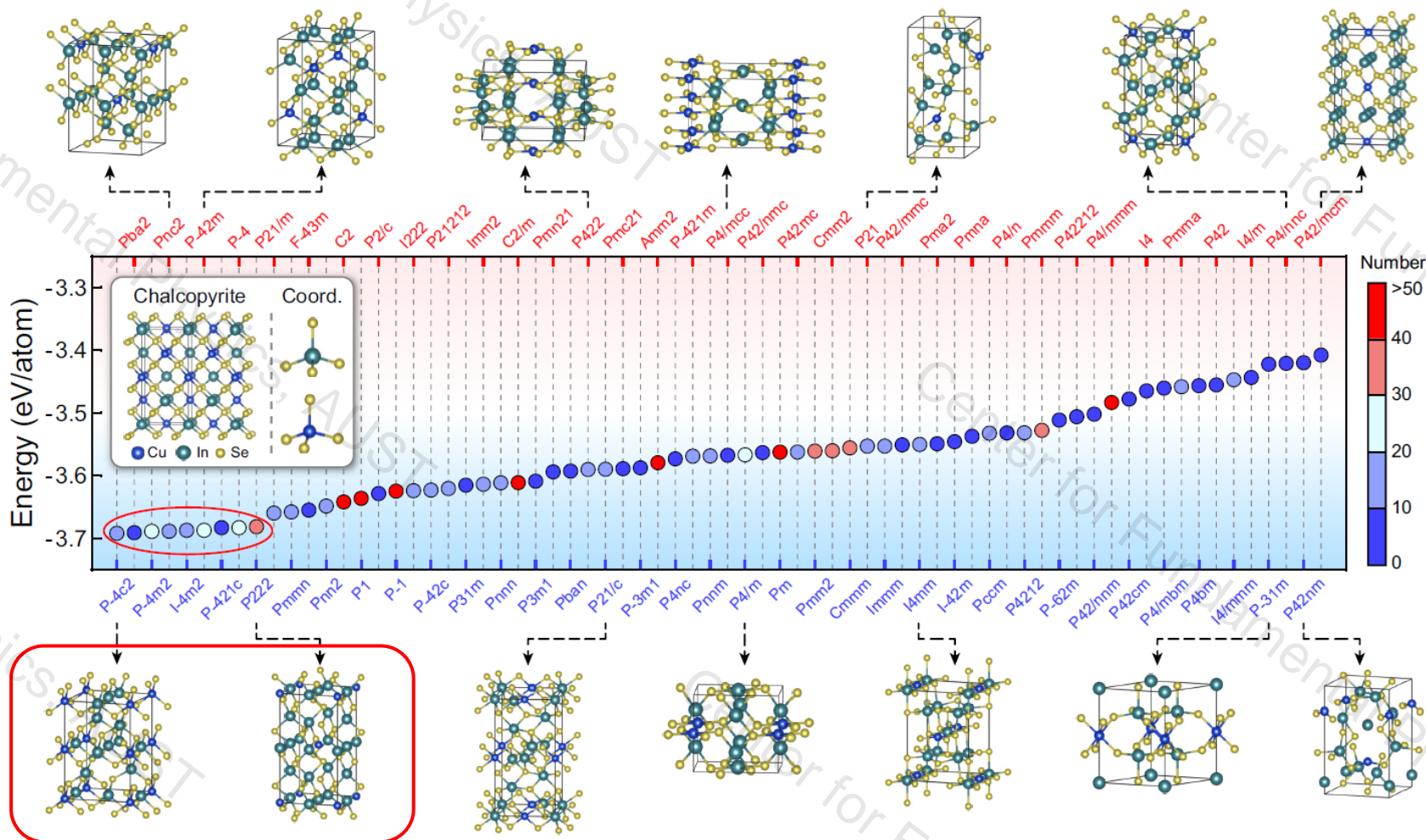
-3.6774 eV/atom



CuIn₅Se₈结构搜索 --- 全空间群搜索

SCCOP全空间群结构搜索:

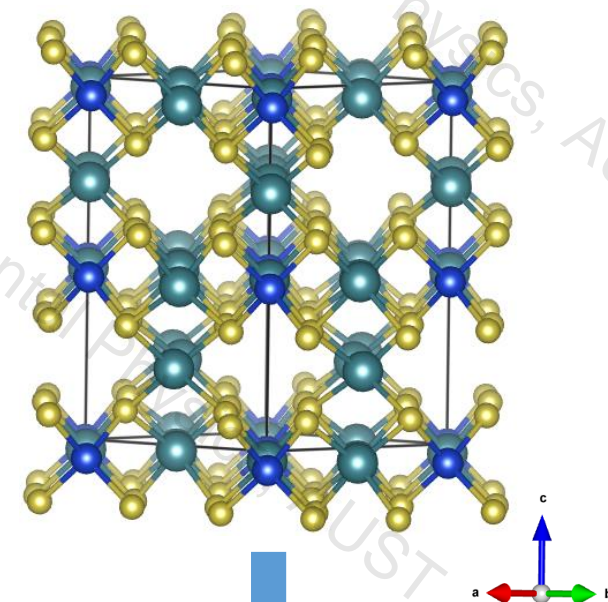
随着能量增加 \rightarrow Cu和In偏离四面体配位模式



低能构型:

空间群: P-4c2

原子数: 28

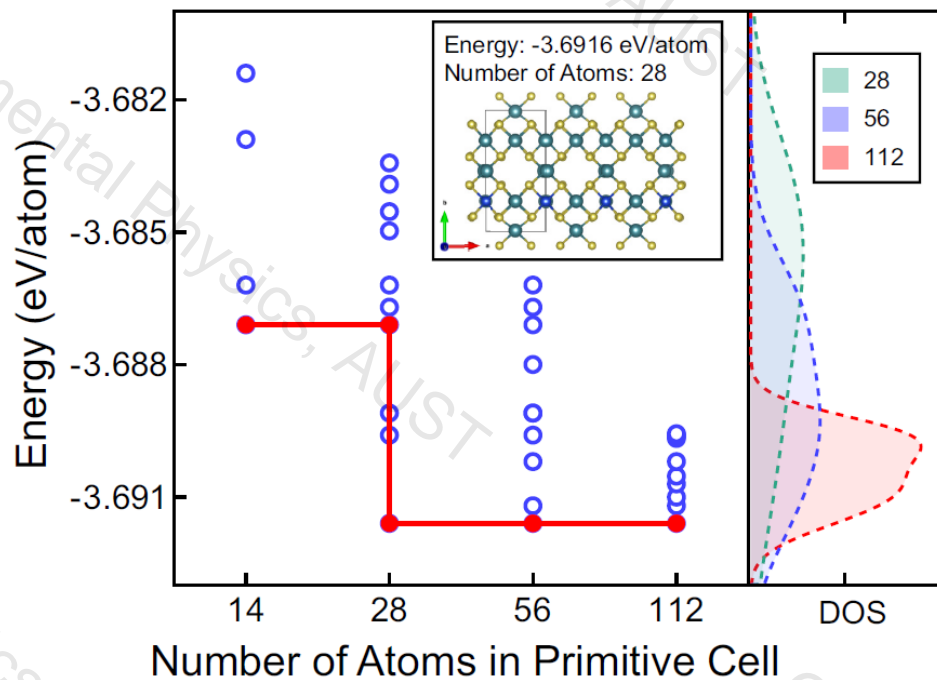


Cu和In在阳离子位点上排列

CuIn₅Se₈结构搜索 --- 固定位点搜索

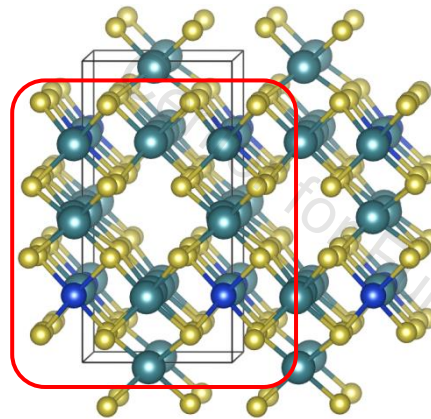
固定位点搜索过程:

原子数增加, 低能区插入更多点

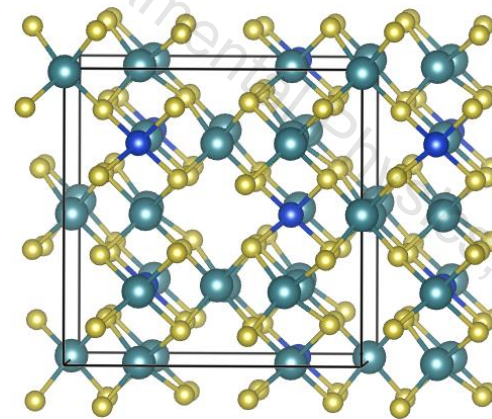


能量最低四个结构:

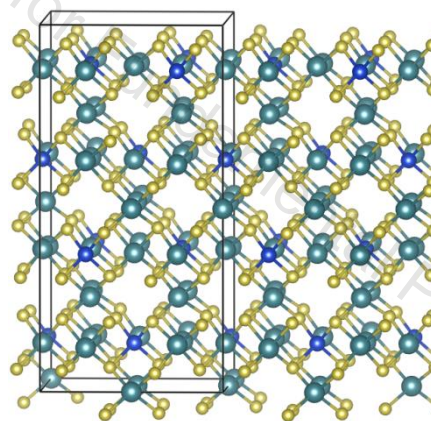
-3.6916 eV/atom (28)



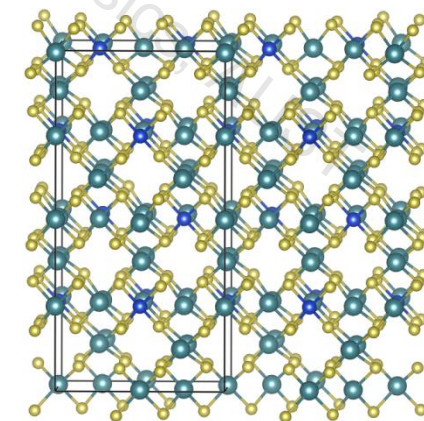
-3.6912 eV/atom (56)



-3.6910 eV/atom (112)



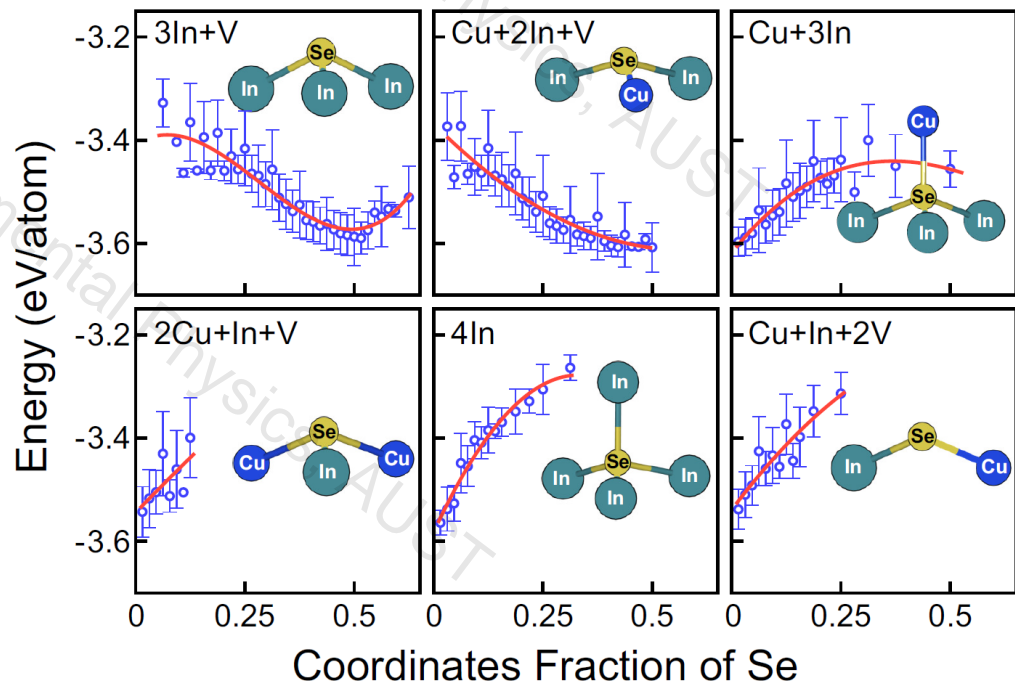
-3.6907 eV/atom (112)



CuIn₅Se₈结构搜索 --- 配位模式统计

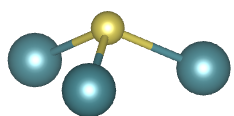
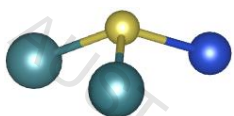
Se-中心的配位模式统计:

与x射线精细结构吸收谱结果一致



k=7: Cu+2In+V

k=9: 3In+V

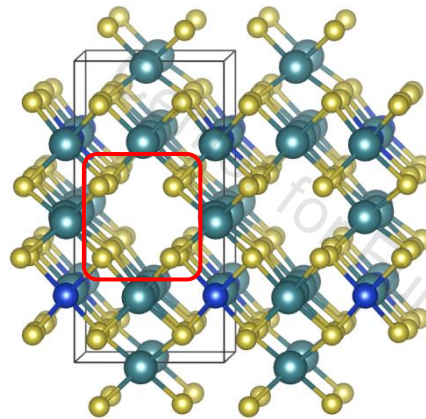


极小化库伦能
满足八电子规则

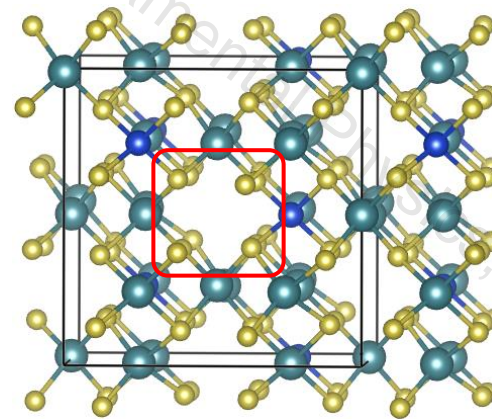
Se中心四面体的阳离子总价电子数

能量最低的几个结构:

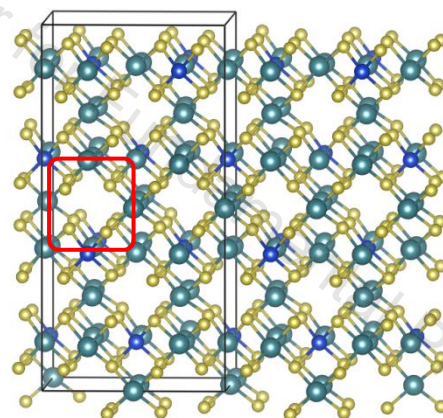
-3.6916 eV/atom (28)



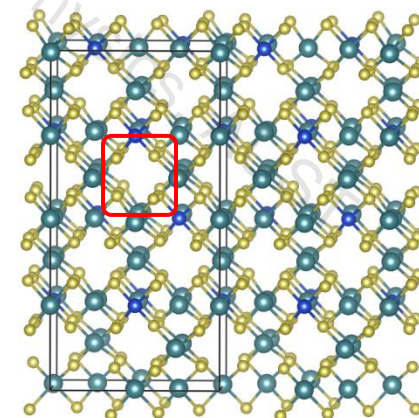
-3.6912 eV/atom (56)



-3.6910 eV/atom (112)



-3.6907 eV/atom (112)

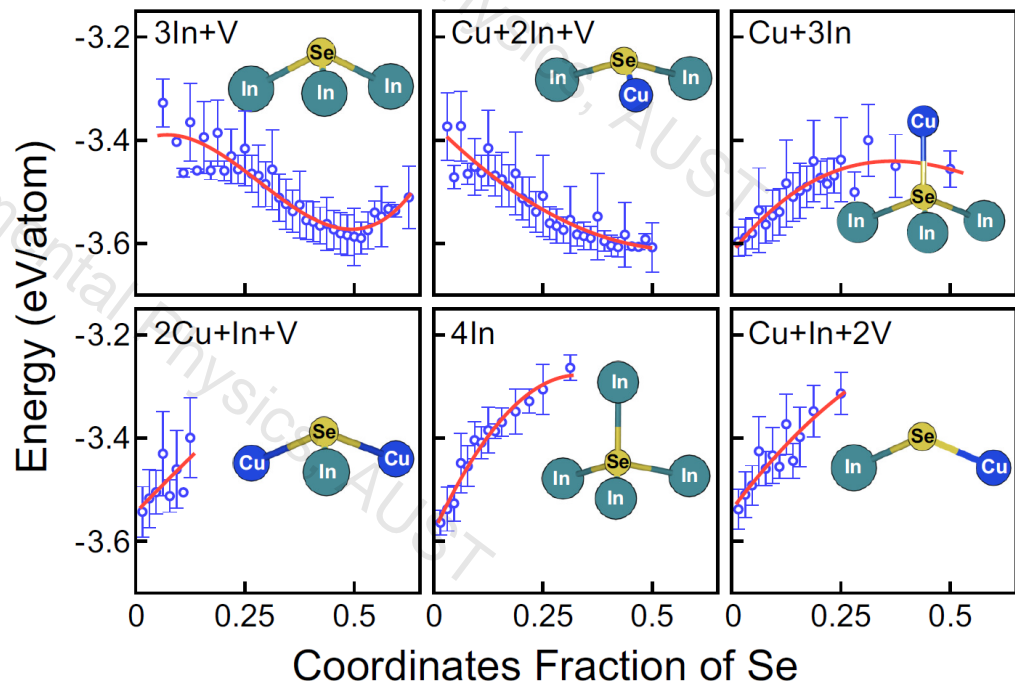


k=7和k=9各占50%配位模式

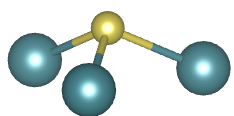
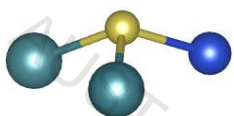
CuIn₅Se₈结构搜索 --- 配位模式统计

Se-中心的配位模式统计:

与x射线精细结构吸收谱结果一致



k=7: Cu+2In+V k=9: 3In+V

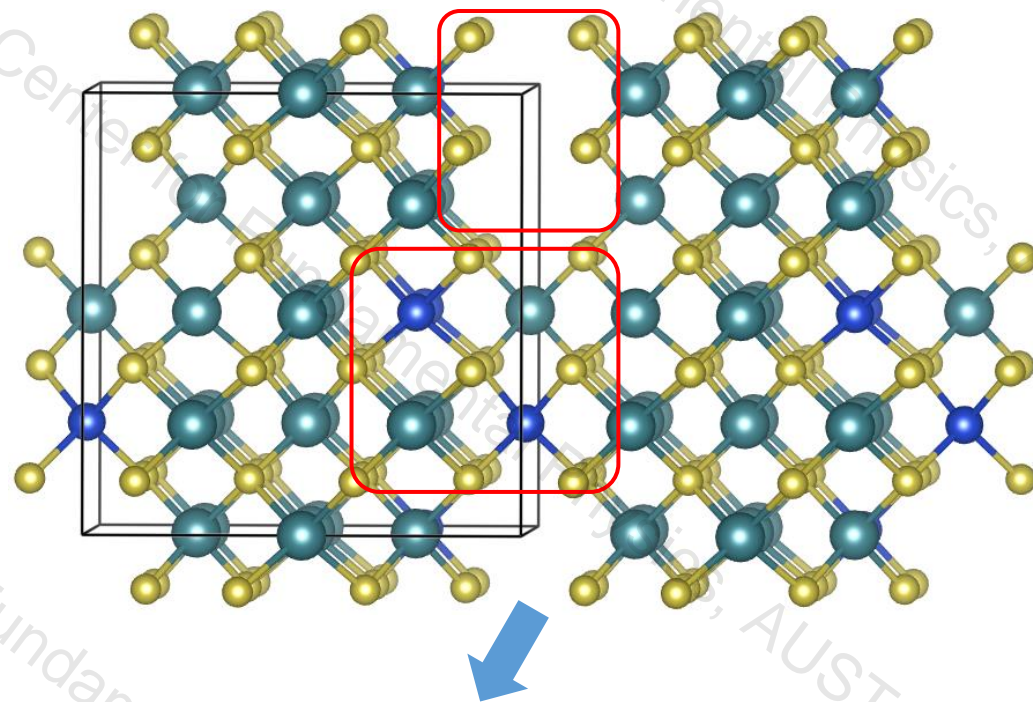


极小化库伦能
满足八电子规则

Se中心四面体的阳离子总价电子数

高能结构示例:

-3.2782 eV/atom (56)



- k=7,9不等于50%, 缺陷对无法完全配对;
- Cu原子距离过近导致库伦排斥增加。

3.4 SCCOP软件获取和使用：SCCOP代码开源 --- GitHub平台

SCCOP开源网页:

sccop Public

Edit Pins Watch 1 Fork 1 Starred 11

main 1 Branch 0 Tags

Go to file Add file Code

File/Folder	Update	Time
data	update	2 years ago
examples	update	2 years ago
images	update	8 hours ago
libs	update	8 hours ago
src	update	8 hours ago
.gitignore	Initial commit	3 years ago
LICENSE	Initial commit	3 years ago
README.md	Update README.md	4 hours ago

About

Symmetry crystal combinatorial optimization program for crystal prediction.

machine-learning crystal-structure simulated-annealing bayesian-optimization vasp-calculations crystal-prediction

Readme Apache-2.0 license Activity Custom properties 11 stars 1 watching 1 fork

Releases

No releases published [Create a new release](#)

Packages

No packages published [Publish your first package](#)

README Apache-2.0 license

Symmetry-based Combinatorial Crystal Optimization Program (SCCOP)

This software package implements the Symmetry-based Combinatorial Crystal Optimization Program (SCCOP) that predicts crystal structure of specific composition.

Table of Contents

- [How to cite](#)
- [Prerequisites](#)
- [Usage](#)
 - [Configuation](#)
 - [Customize initial search file](#)
 - [Submit SCCOP job on cluster](#)
 - [Successful example](#)
- [Data](#)
- [Authors](#)
- [License](#)

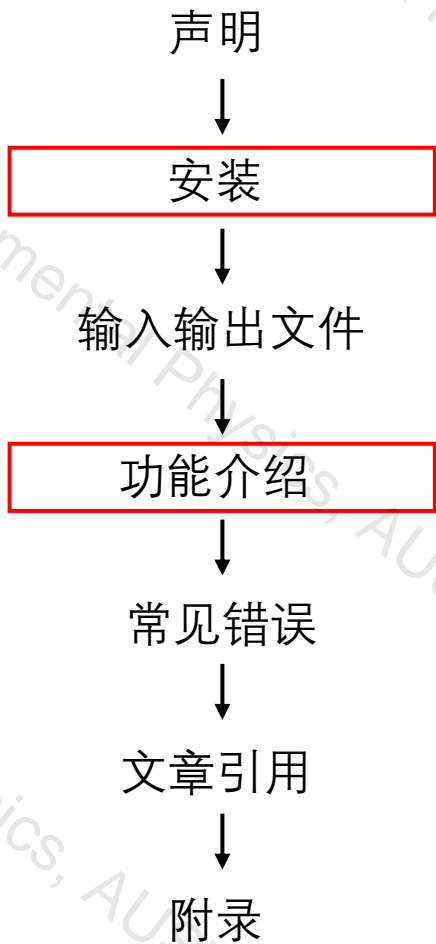
语言: Python + Shell

超过30,000行代码

依赖库均为免费开源

SCCOP帮助手册

手册目录:



目录

1 声明	1
1.1 Licence	1
2 安装	2
2.1 依赖环境	2
2.2 自主配置	2
2.2.1 软件下载	2
2.2.2 环境配置	2
2.3 安装包配置	3
2.3.1 软件下载	3
2.3.2 环境配置	3
2.4 SSH配置	3
3 输入与输出文件	4
3.1 输入文件	4
3.1.1 基本参数	4
3.1.2 模型参数	5
3.1.3 迭代参数	5
3.1.4 采样参数	6
3.1.5 搜索类型	6
3.1.6 搜索参数	8
3.1.7 样本筛选	8
3.1.8 DFT参数	9
3.1.9 结构特征	10
3.2 环境变量	10
3.2.1 任务管理系统	10
3.2.2 环境加载	11
3.2.3 VASP设置	11
3.3 输出文件	12
3.3.1 日志文件	12
3.3.2 结构文件	14
4 功能介绍	16
4.1 任务脚本	16
4.2 任务提交	17
4.3 按空间群搜索	17
4.3.1 输入文件	17
4.3.2 搜索实例	18
4.4 固定位点搜索	21
4.4.1 输入文件	21
4.4.2 搜索实例	22

Step 1

Step 3

Step 2

4.5 团簇搜索	24
4.5.1 输入文件	24
4.5.2 搜索实例	25
5 常见错误	27
5.1 节点通信错误	27
5.2 VASP启动异常	27
6 文章引用	28
6.1 引用方法	28
6.2 原理文章	28
6.3 应用文章	28
7 附录	29
7.1 软件目录结构	29
7.2 批量任务脚本	30
7.3 平面空间群	31
7.4 层状空间群	32
7.5 三维空间群	33

手册查阅顺序: 1. 安装
2. 功能介绍
3. 输出文件

一般来说, 默认参数适用于大部分情况, 程序调试成功后, 再根据自己的计算资源修改input.py

超算运行SCCOP示例

提交脚本(PBS):

```
shell > $ job_30.pbs
1  #!/bin/bash
2  #PBS -N sccop_pbs
3  #PBS -q main3
4  #PBS -l nodes=6:ppn=48
5  #PBS -l walltime=48:00:00
6
7  host=`hostname`
8  work=$PBS_O_WORKDIR
9
10 cd $work
11 mkdir output
12
13 ssh -T $host
14 cd /tmp
15 rm -r sccop search
16 rm sub_search_3d.sh
17 scp $work/sub_search_3d.sh .
18
19 echo work=$work > temp_file
20 cat sub_search_3d.sh >> temp_file
21 mv temp_file sub_search_3d.sh
22
23 for i in `seq 1 10`
24 do
25     sh sub_search_3d.sh
26     tar -zcf search_${i}.tar.gz search
27     scp search_${i}.tar.gz $work/output/.
28     rm search_${i}.tar.gz
29     rm -r search
30 done
```

天河2号, 个人服务器

提交脚本(SLURM):

```
shell > $ job_beijing.slurm
1  #!/bin/bash
2  #SBATCH -J sccop
3  #SBATCH -o sccop_%j.out
4  #SBATCH -e sccop_%j.err
5  #SBATCH -p v6_384
6  #SBATCH -N 2
7  #SBATCH -n 192
8
9  module load mpi/oneAPI/2022.1
10 export SLURM_OVERLAP=1
11
12 work=$SLURM_SUBMIT_DIR
13 host=`hostname`
14 ssh -T $host
15
16 cd /tmp
17 rm -r sccop
18 mkdir sccop
19 cd sccop
20 scp -r $work/data .
21 scp -r $work/libs .
22 scp -r $work/src .
23
24 python src/core/job_system.py
25 python src/main.py
26
27 cd ..
28 scp -r sccop/data/system.log $work/.
```

瀚海20, 22, 北龙超算

运行示例:

```
POSCAR-050-node136.out, True, -7.355353166666667
*WARNING* SinglePointEnergy is failed!
POSCAR-051-node136.out, False, 1000000.0
POSCAR-052-node136.out, True, -5.619335
POSCAR-053-node136.out, True, -5.0963294
POSCAR-054-node136.out, True, -5.2703275
POSCAR-055-node136.out, True, -4.4585394
POSCAR-056-node136.out, True, -6.203008375
POSCAR-057-node136.out, True, -7.1264536666666665
POSCAR-058-node136.out, True, -5.71548725
POSCAR-059-node136.out, True, -6.3693504
Energy file generated successfully!
Delete duplicates: 52
SCCOP optimize structures: 12
End Symmetry Combinatorial Crystal Optimization Program --- Recycle: 0
Start VASP Calculation --- Optimization
POSCAR-000-node131, -9.088195625
POSCAR-001-node131, -9.089726250
POSCAR-002-node132, -9.088203125
POSCAR-003-node132, -8.983441167
POSCAR-004-node133, -8.983651000
POSCAR-005-node133, -8.983535333
POSCAR-006-node134, -8.611571000
POSCAR-007-node134, -9.132847875
POSCAR-008-node135, -8.618795500
POSCAR-009-node135, -7.926331500
POSCAR-010-node136, -8.983409833
POSCAR-011-node136, -8.140675222
Delete same structures: 7
Energy file generated successfully!
All jobs are completed --- Optimization
[-9.13284787 -9.08972625 -8.983651 -8.6187955 -8.611571 -8.14067522]
Optimize configurations: 6
Start VASP Calculation --- Optimization
POSCAR-001-node132, -9.097467500
POSCAR-004-node133, -8.987320667
POSCAR-006-node135, -8.689496375
POSCAR-007-node131, -9.223258750
POSCAR-008-node134, -8.713237500
POSCAR-011-node136, -8.135457222
Energy file generated successfully!
All jobs are completed --- Optimization
Time: 218
```

申请6个CPU节点,
每个节点48核

测试组分C, 原胞内包含
5-10原子, 耗时: 3.6min

金刚石
石墨

4. SCCOP方法与应用总结

SCCOP总结:

方法总结

1. 提出了**基于直接不等价单元的约束性蒙特卡洛树搜索**结构采样方法;
2. 提出了**基于图神经网络的势能面划分策略**, 实现了样本的高效筛选;
3. 提出了**基于模拟退火的分区域搜索策略**, 极大减小了机器学习模型的不确定性;
4. 根据上述三点, 发展了**固定位点、空间群对称性、团簇**的结构预测程序SCCOP。

应用总结

1. 固定位点、二维材料, 三维材料结构预测, 进行了**全面方法测试和比较**;
2. 二维 $B_xC_yN_{1-x-y}$ 结构搜索中, 新发现了**28个动力学稳定的低能结构**;
3. 硼同素异形体的结构搜索中, 新发现了**7个硼团簇低能结构**, 以及**全新的 B_{22} 团簇单元**;
4. 通过系统性的结构搜索, **确认了有序空位组分 $CuIn_5Se_8$ 的低能构型**, 以及**Se原子的配位模式**。

结语：

AI 在物理计算各领域大有可为



中国科学技术大学

University of Science and Technology of China

感谢倾听!

欢迎批评指正

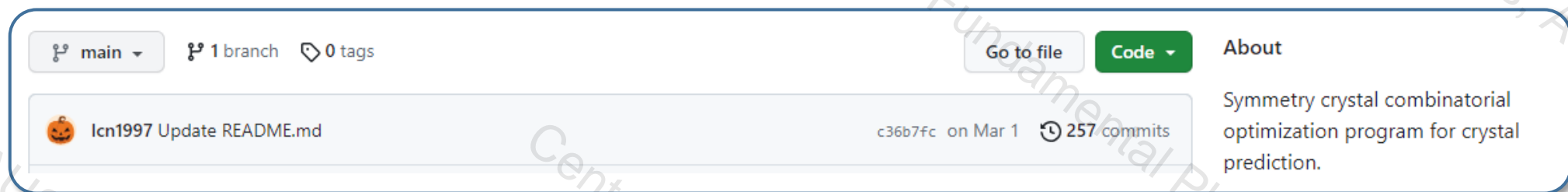
参考资料

论文:

1. Chuan-Nan Li, Han-Pu Liang, Xie Zhang, Zijing Lin and Su-Huai Wei. Graph deep learning accelerated efficient crystal structure search and feature extraction. *npj Comput. Mater.* **9**, 176 (2023)
2. Chuan-Nan Li, Han-Pu Liang, Yifeng Duan and Zijing Lin. Machine-learning accelerated annealing with fitting-search style for multicomponent alloy structure predictions. *Phys. Rev. Materials* **7**, 033802 (2023)
3. Chuan-Nan Li, Han-Pu Liang, Xie Zhang, Su-Huai Wei and Zijing Lin. Minimizing the training dataset for efficient crystal structure prediction by graph neural network [J]. *Nat. Comput. Sci.* (submitted)

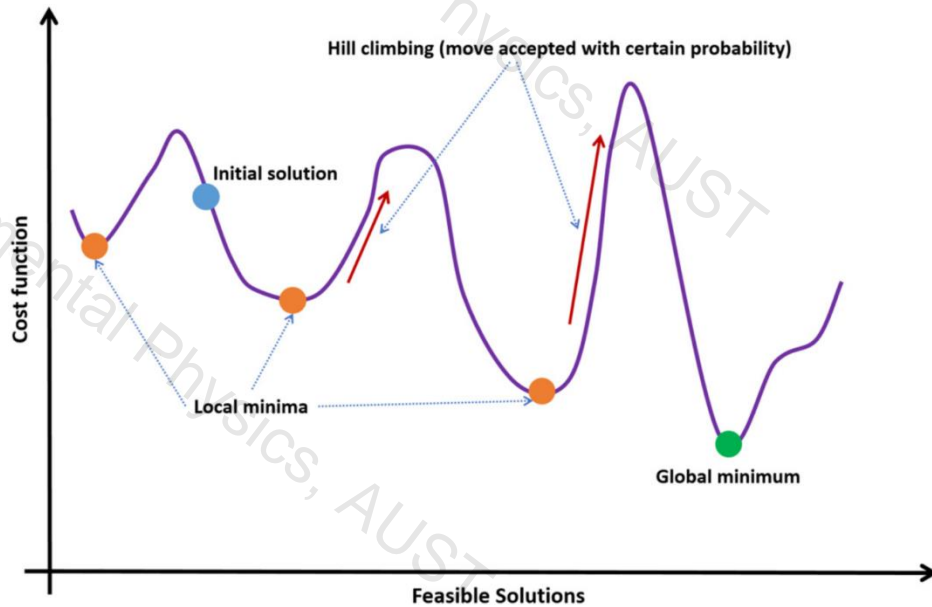
开源代码:

SCCOP: <https://github.com/TheCatOfHs/sccop>



SCCOP方法展望

无梯度优化:



模拟退火算法优化原子位置

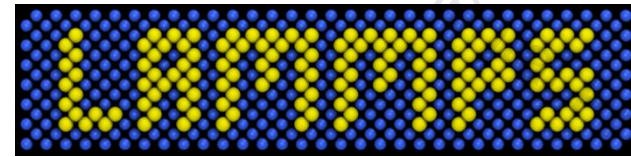
$$p(1 \rightarrow 2) = \begin{cases} 1, & E_1 > E_2 \\ \exp\left(-\frac{E_2 - E_1}{T}\right), & E_1 \leq E_2 \end{cases}$$

梯度优化:

(1)



(2)



通过接口调用

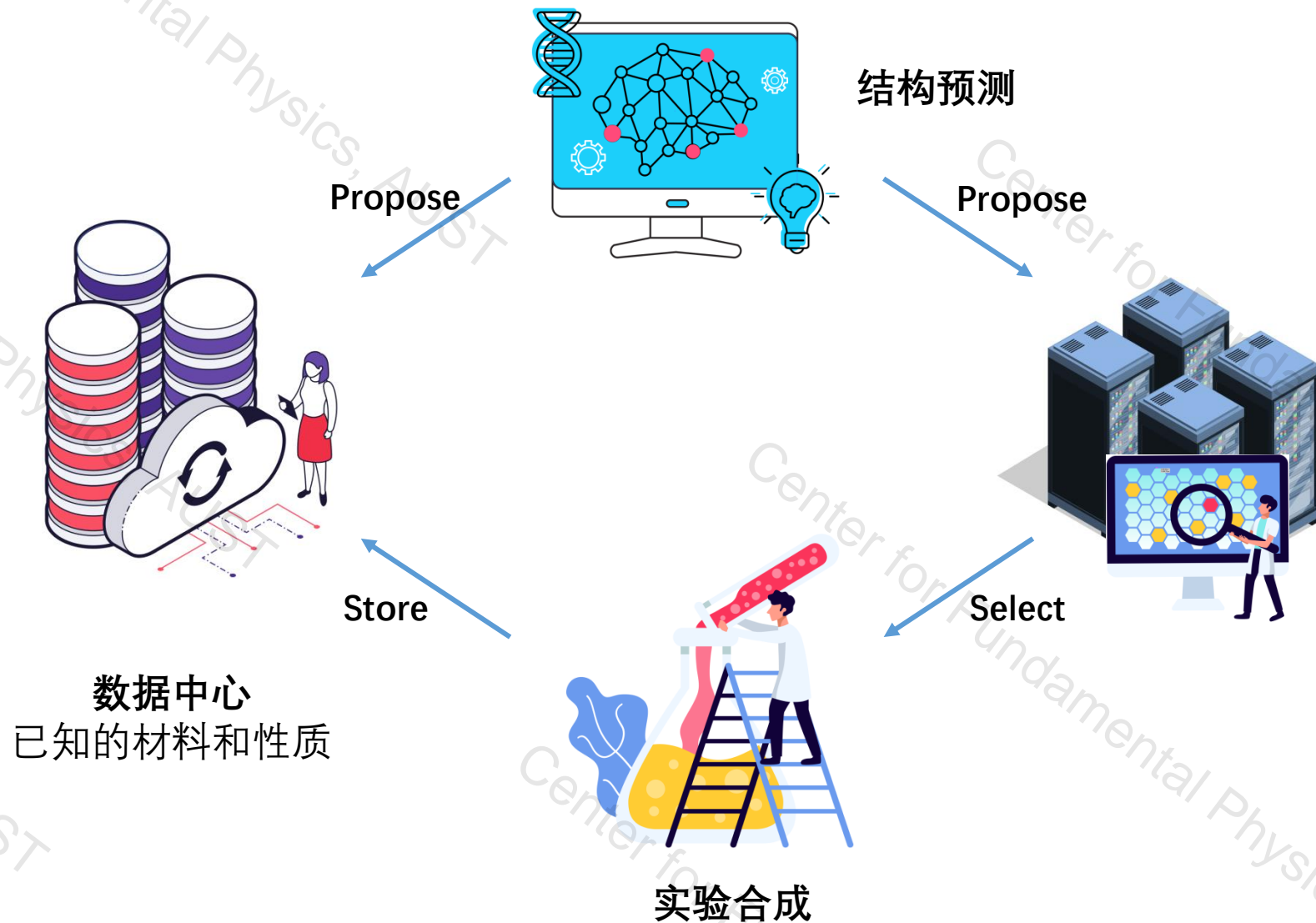
Conjugate Gradient(CG): 计算梯度的共轭方向

Broyden-Fletcher-Goldfarb-Shanno(BFGS): 近似Hessian矩阵, 需要计算前后两步的梯度差

SCCOP中如何计算势能面的梯度呢?

SCCOP应用展望

AI驱动的材料设计:



机器学习模型
独立于数据库

量化性质计算

1. 电池
2. 催化剂
3. 太阳能材料
4. 储氢材料
5.