

声明：本课程版权归华算科技所有，仅限个人学习，严禁任何形式的录制、传播和账号分享。一经发现，平台将依法保留追究权，情节严重者将承担法律责任。

# Python与机器学习

## ——机器学习前沿

华算科技 黄老师  
2022年2月25日



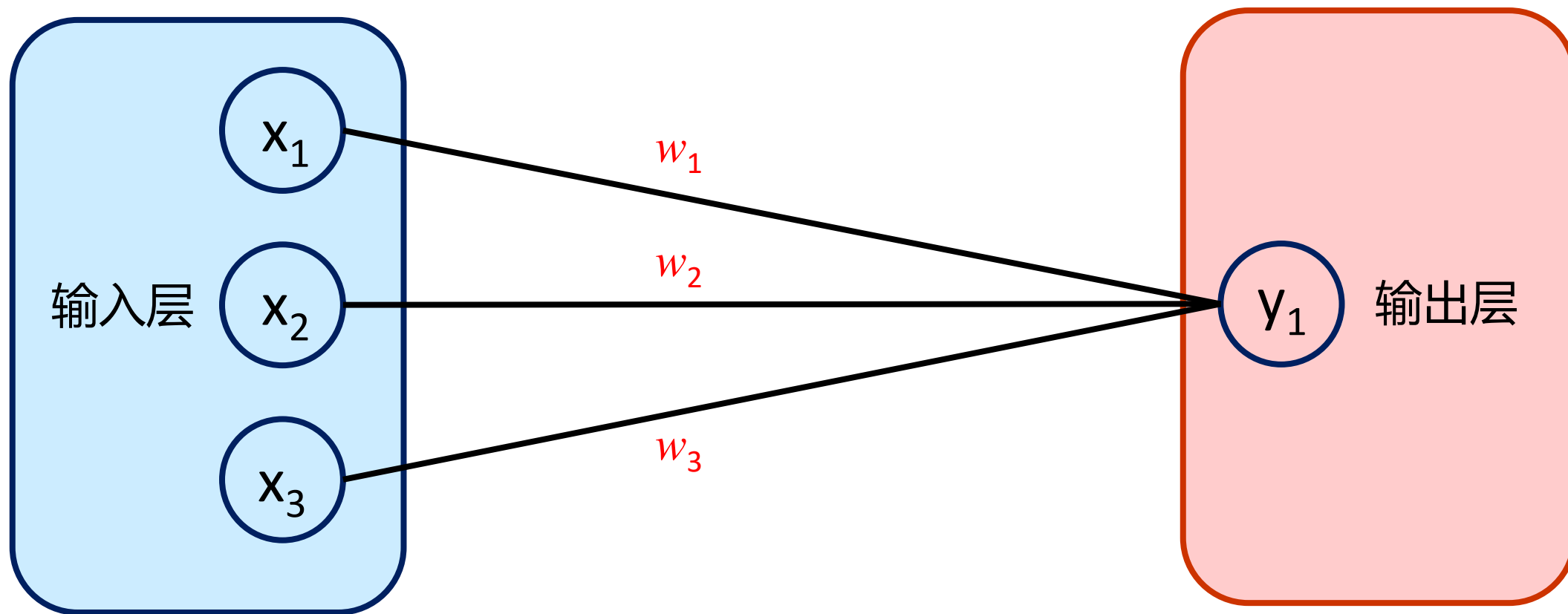
1. 神经网络
2. 态密度预测吸附能
3. 机器学习势
4. 实时从头算分子动力学
5. 新兴的机器学习计算程序

1. 神经网络
2. 态密度预测吸附能
3. 机器学习势
4. 实时从头算分子动力学
5. 新兴的机器学习计算程序

# 回归算法

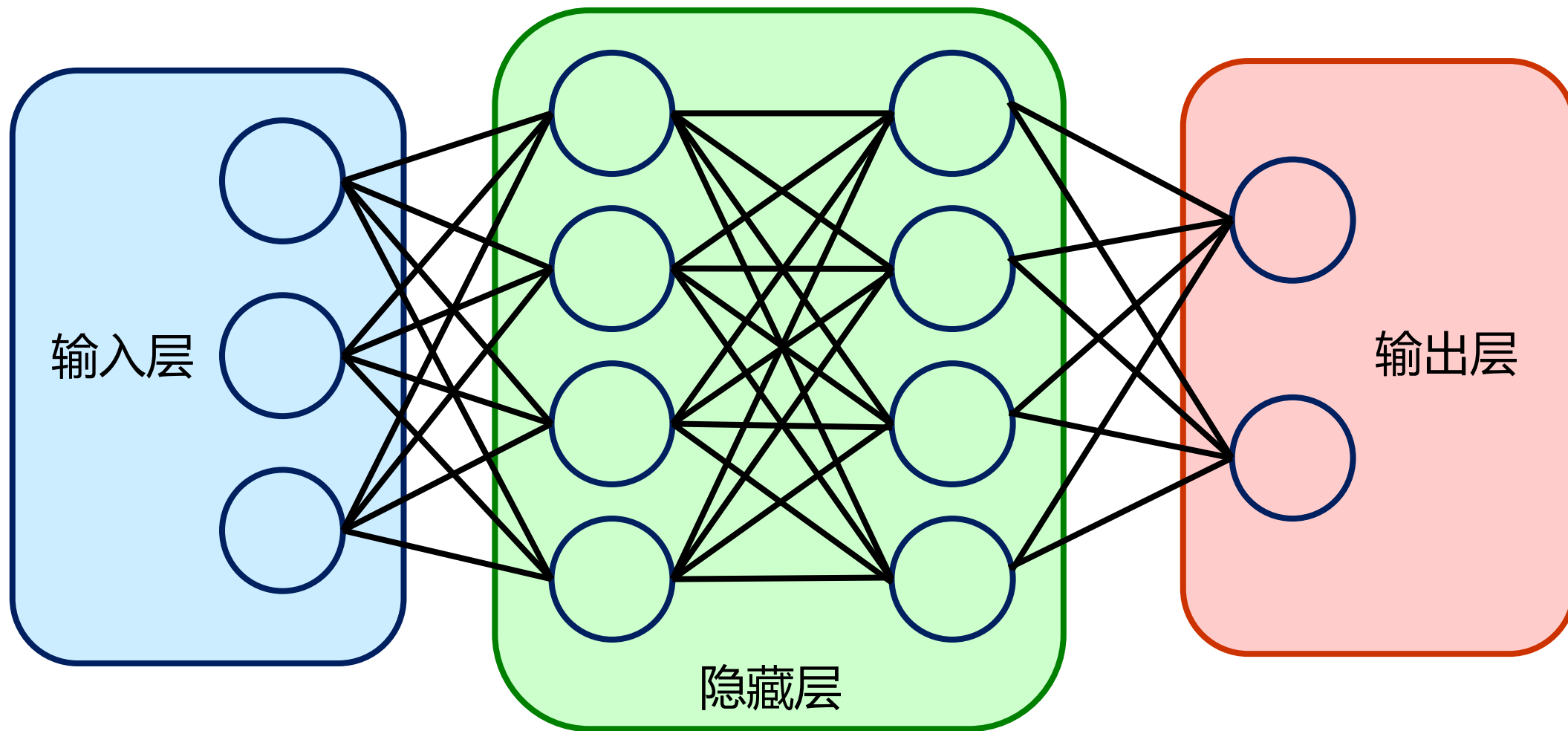
回归算法可以用下图进行简单描述

$$y_1 = \sum w_i x_i + b_1$$

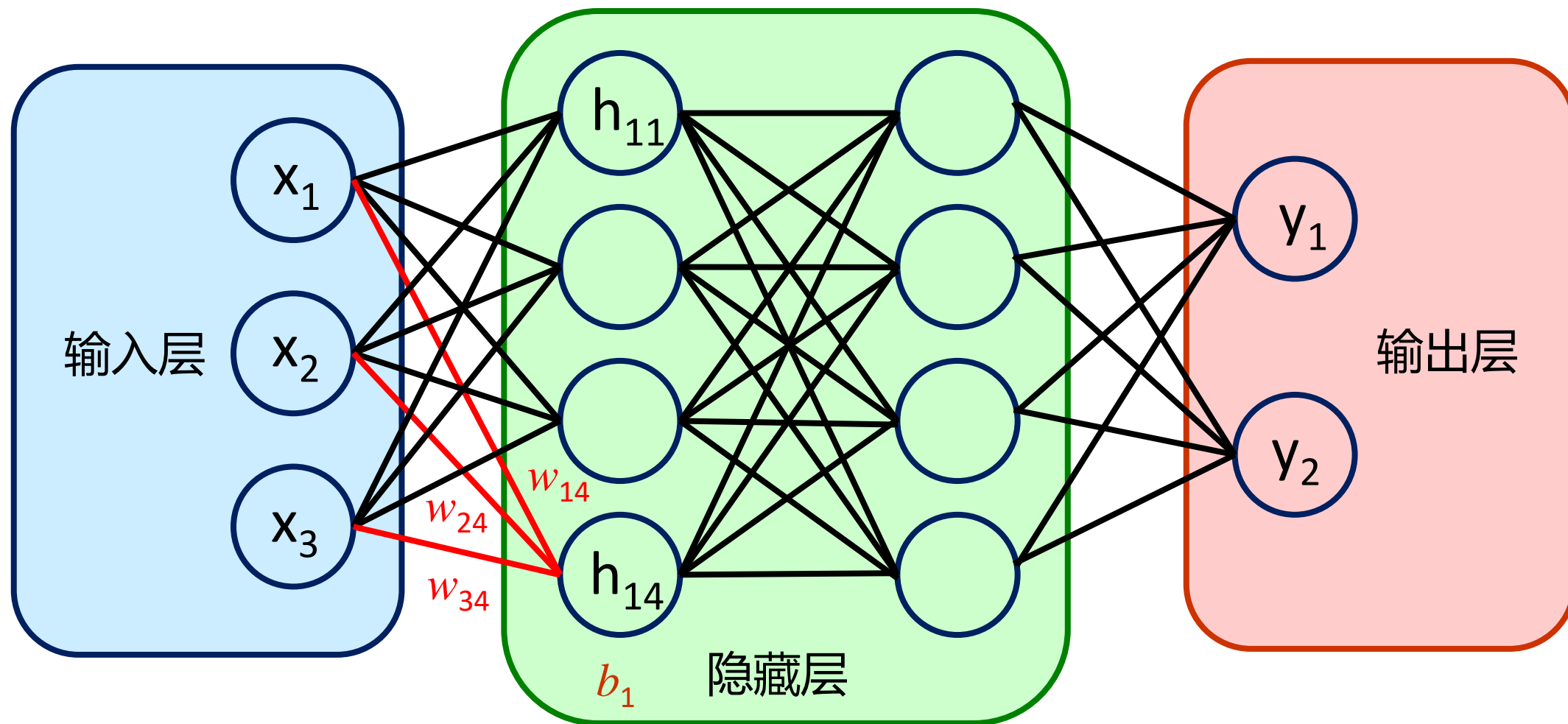


# 神经网络算法

由许多简单单元组成的广泛并行互连的网络，能够模拟生物神经系统对真实世界物体做出的交互反应



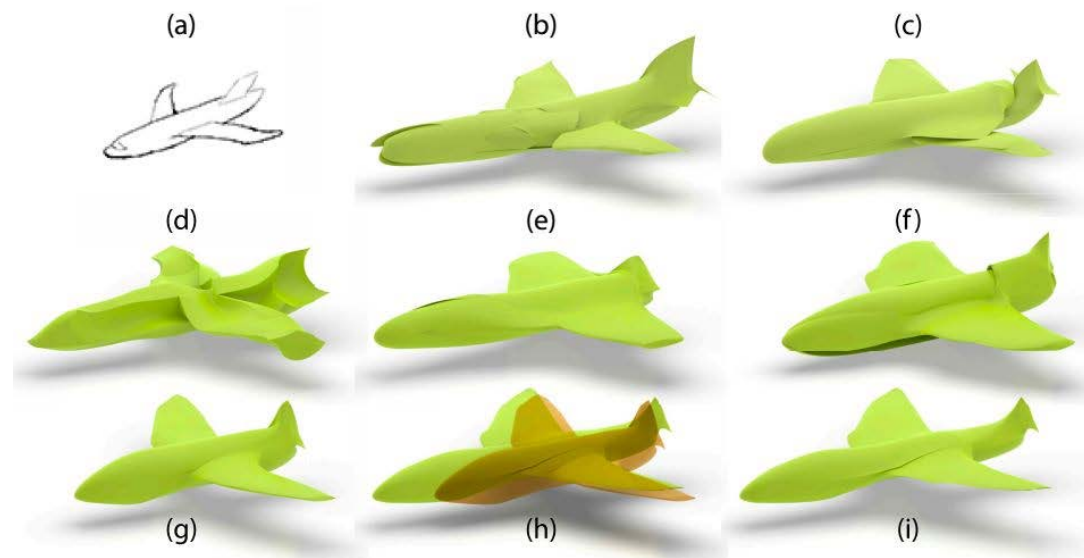
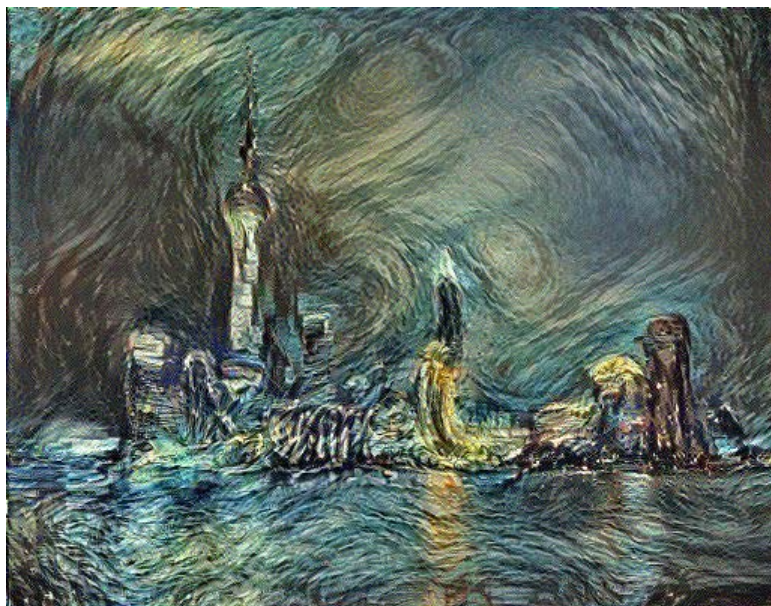
# 神经网络算法



$$h_{14} = f(\sum w_{i4}x_i + b_1) \quad b_1 \text{ 偏置} \quad f \text{ 激活函数}$$



# 神经网络算法的应用



1. 神经网络
2. 态密度预测吸附能
3. 机器学习势
4. 实时从头算分子动力学
5. 新兴的机器学习计算程序

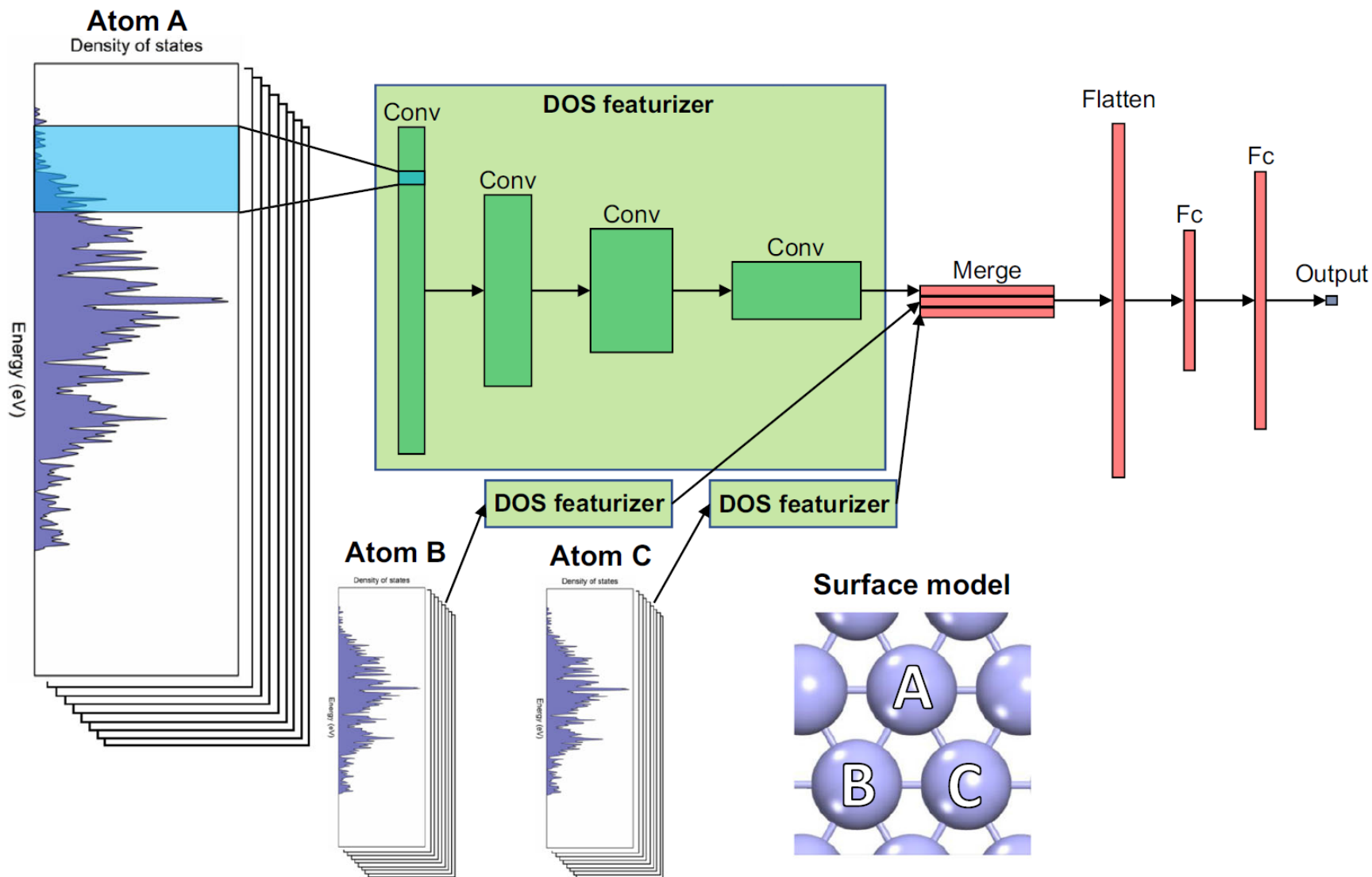


# 态密度预测吸附能

双金属合金表面吸  
附小分子、基团

双金属表面：  
2000个

吸附能数据：  
37000个

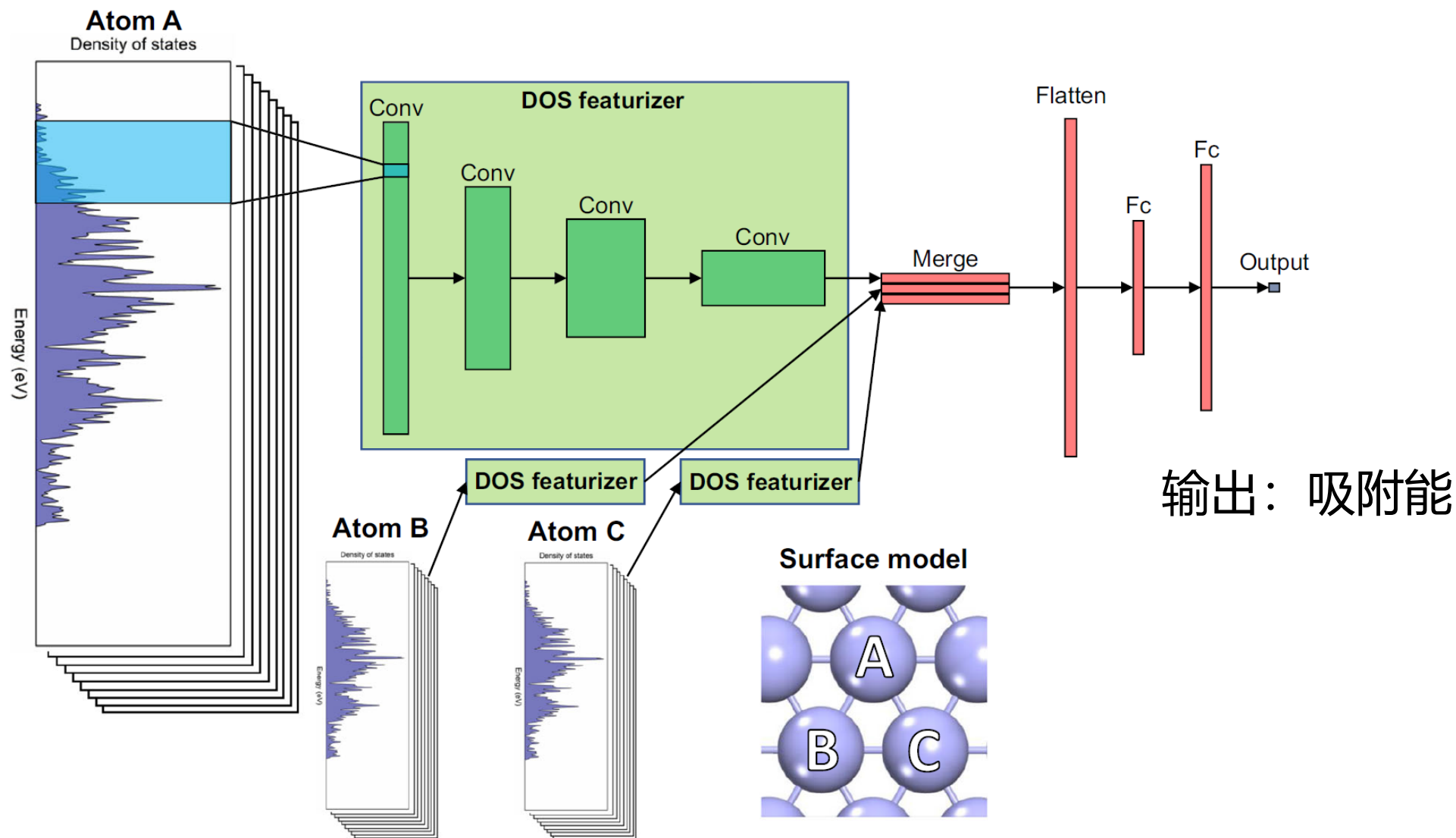


Fung, V., Hu, G., Ganesh, P. and Sumpter, B. G. *Nat. Commun.* **2021**, 12, 88.

# 态密度预测吸附能

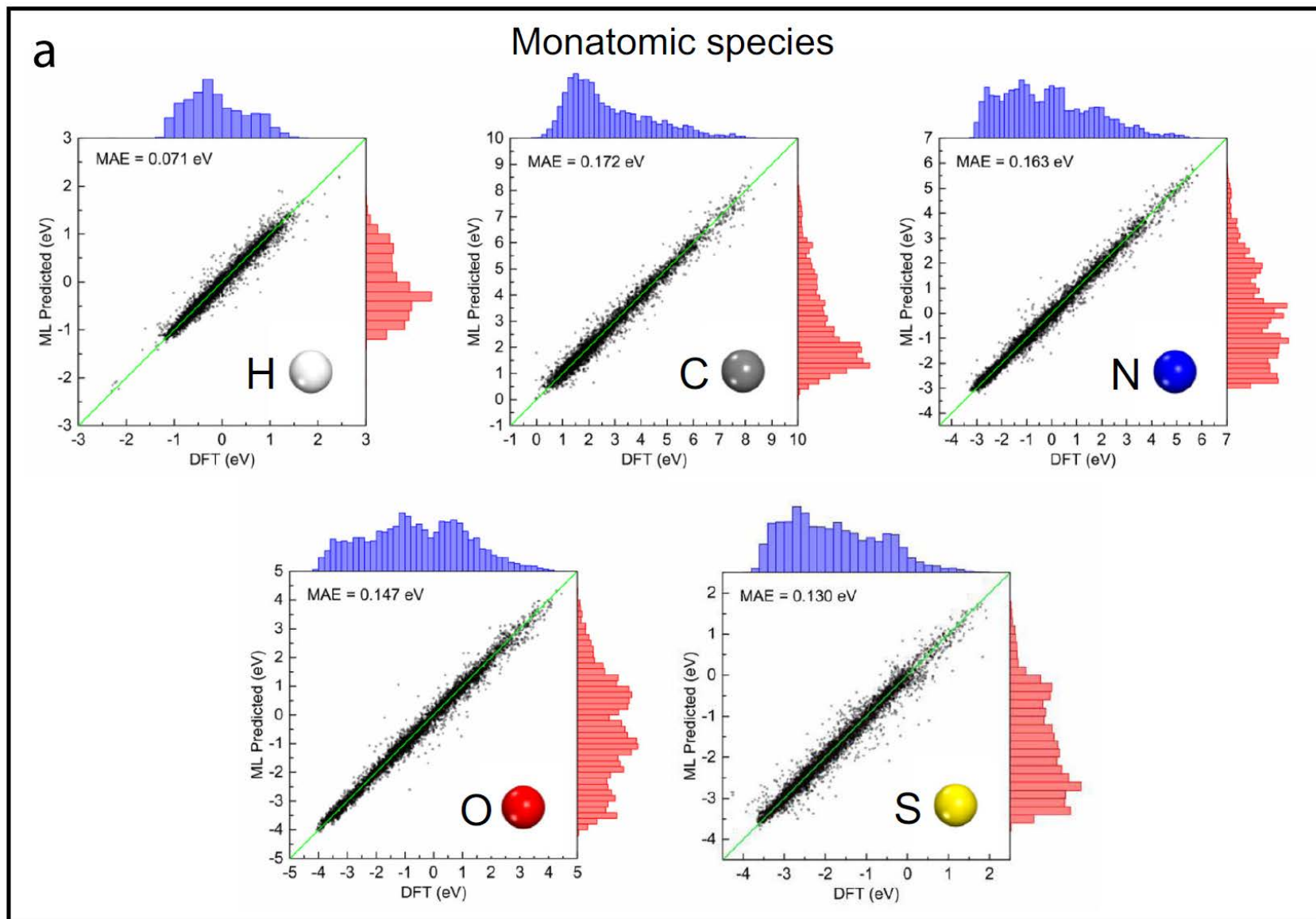
输入：  
各原子的PDOS

1个原子对应  
9个输入



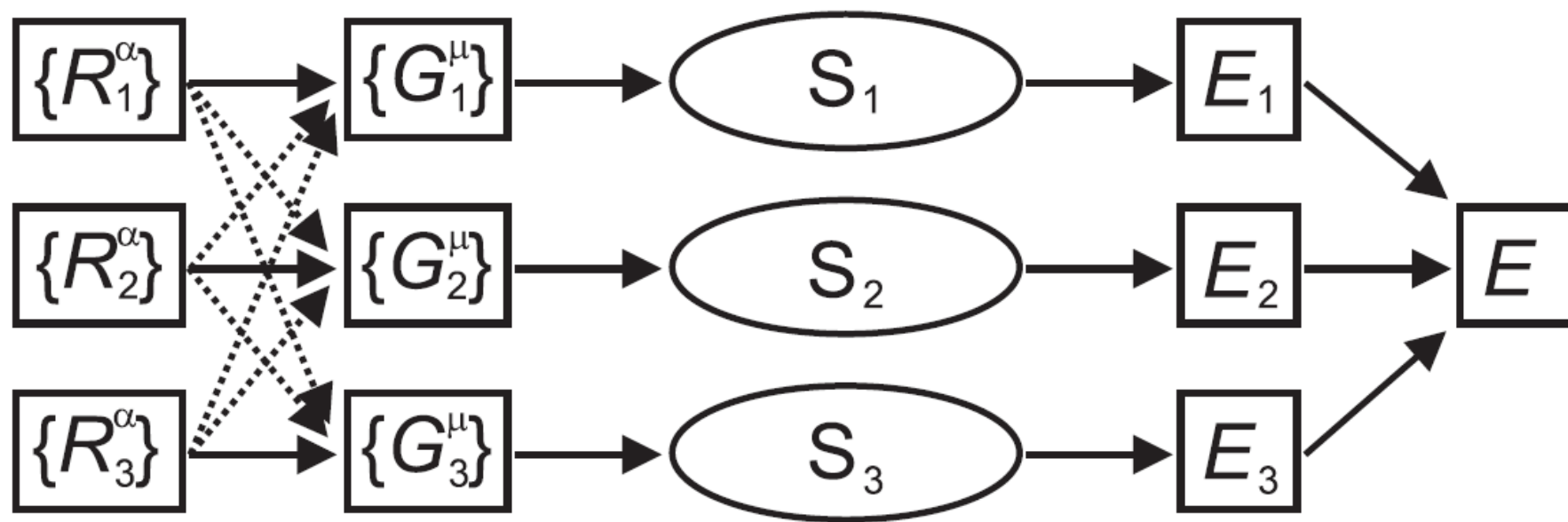
输出：吸附能

# 态密度预测吸附能

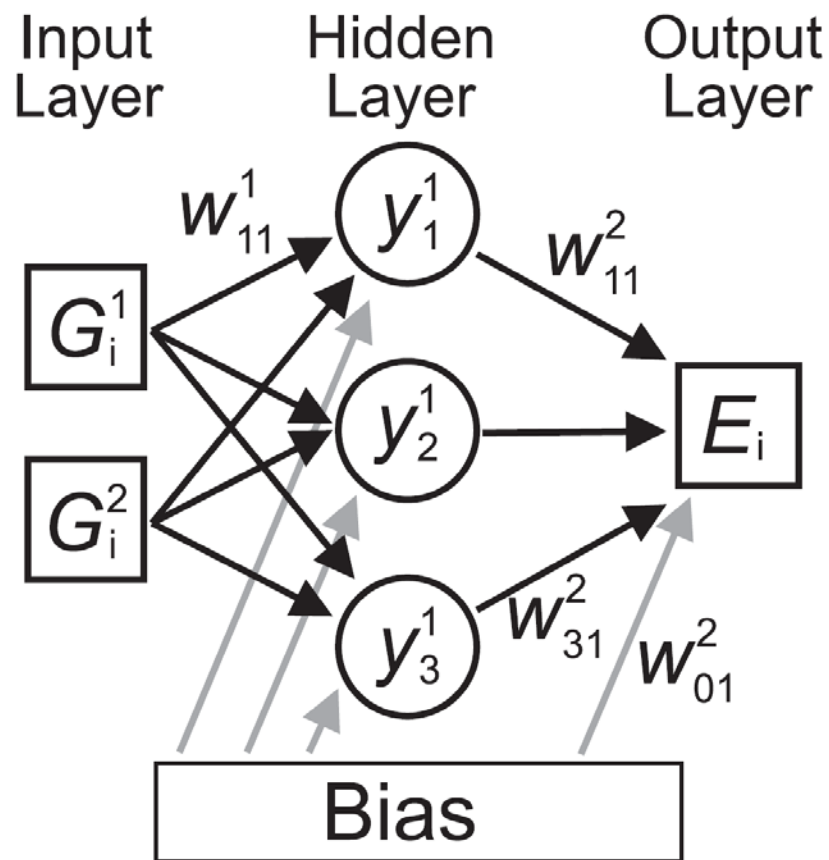


1. 神经网络
2. 态密度预测吸附能
3. 机器学习势
4. 实时从头算分子动力学
5. 新兴的机器学习计算程序

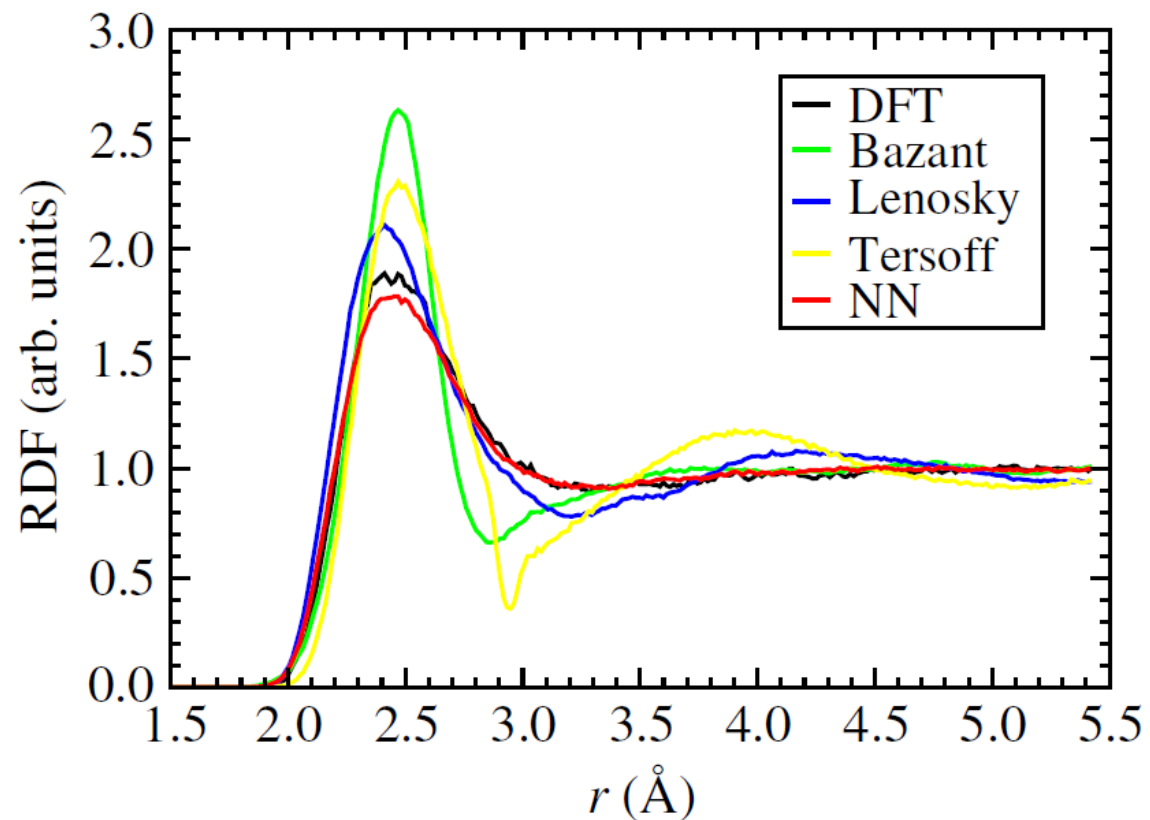
使用第一性原理计算的结果作为训练数据创建分子力场，并用于分子动力学计算



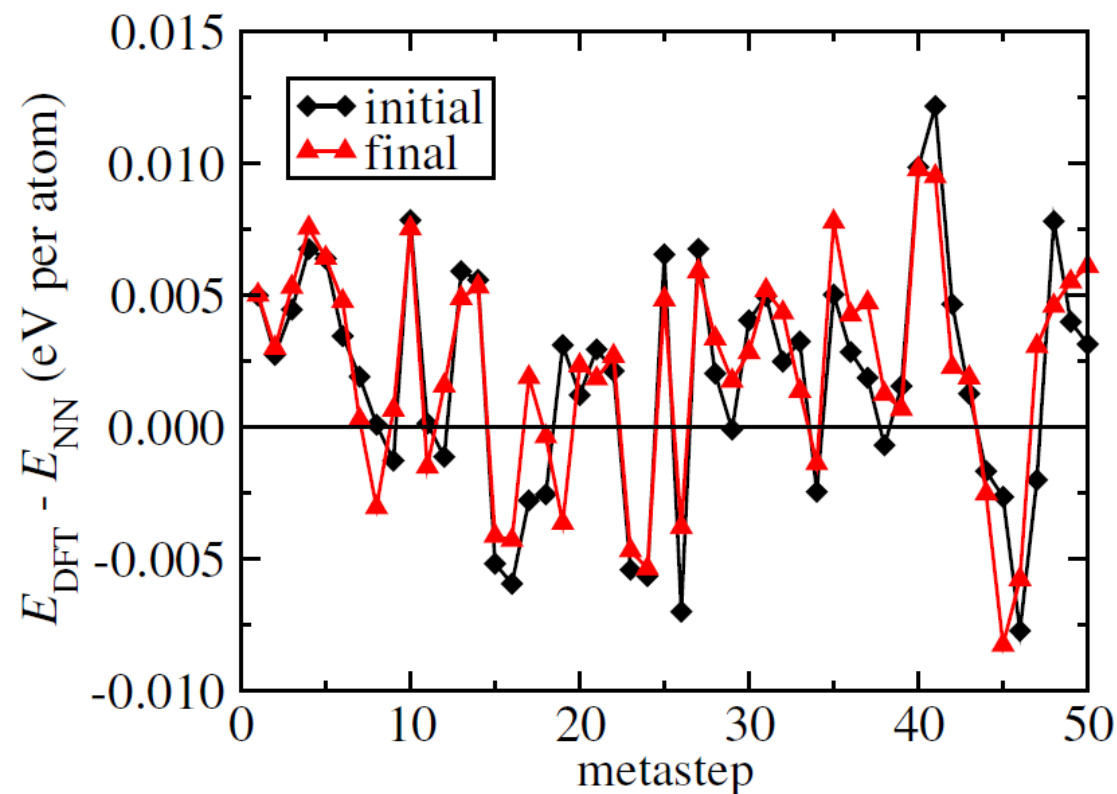
$$E = \sum_i E_i.$$



$$E_i = f_a^2 \left[ w_{01}^2 + \sum_{j=1}^3 w_{j1}^2 f_a^1 \left( w_{0j}^1 + \sum_{\mu=1}^2 w_{\mu j}^1 G_i^\mu \right) \right].$$

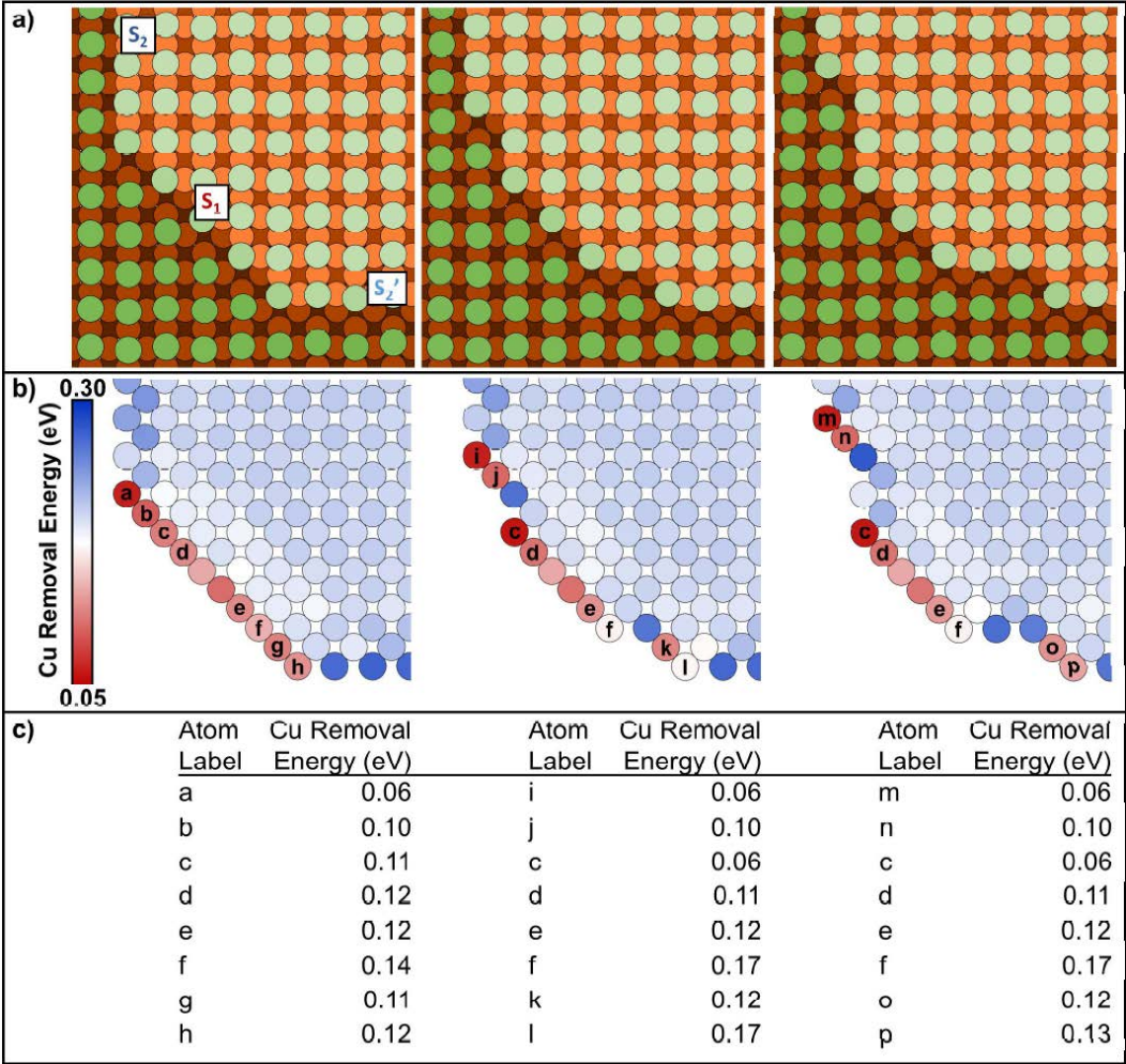


Si, 3000K, 20ps



Si, 300K, 元动力学  
(metadynamics), 2ps

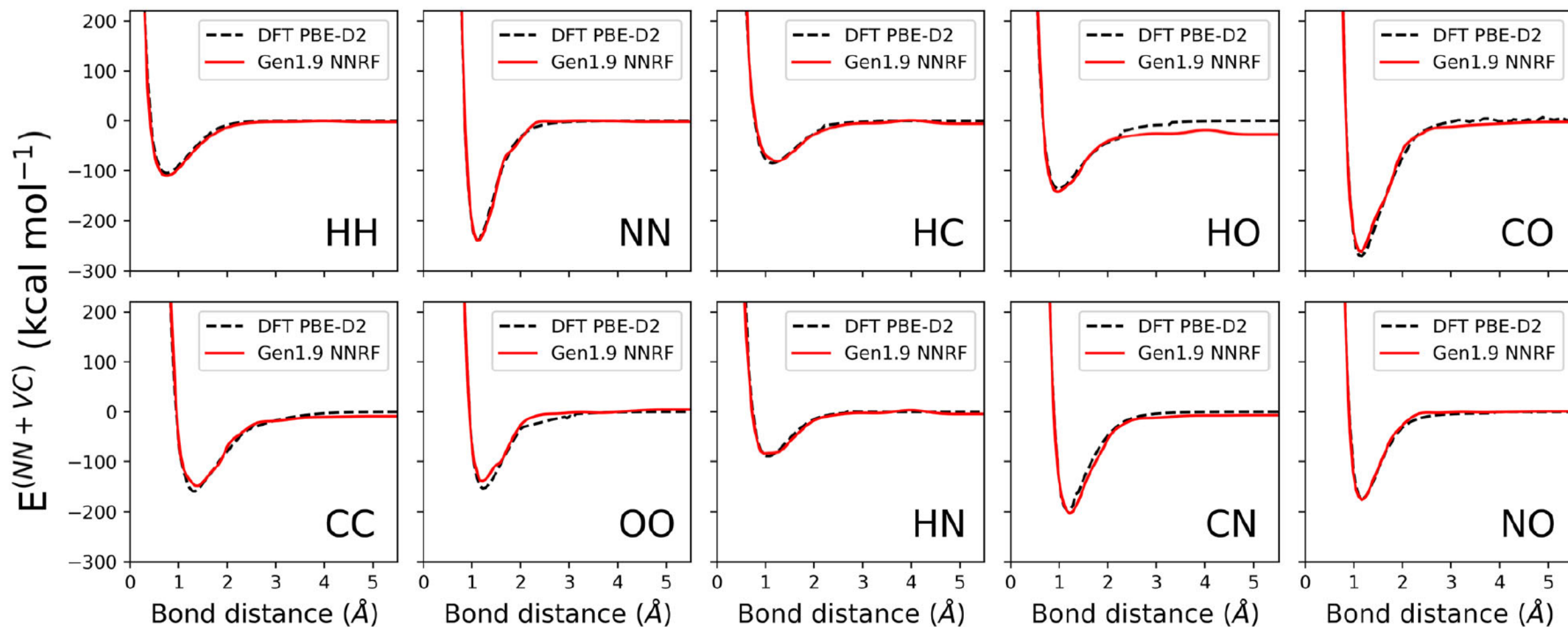




Cl/Cu(100)表面

M. C. Groenenboom, T. P. Moffat, and K. A. Schwarz. *J. Phys. Chem. C* **2020**, 124, 12359-12369.

3100分子, 11941构型    15凝聚体系, 32973构型



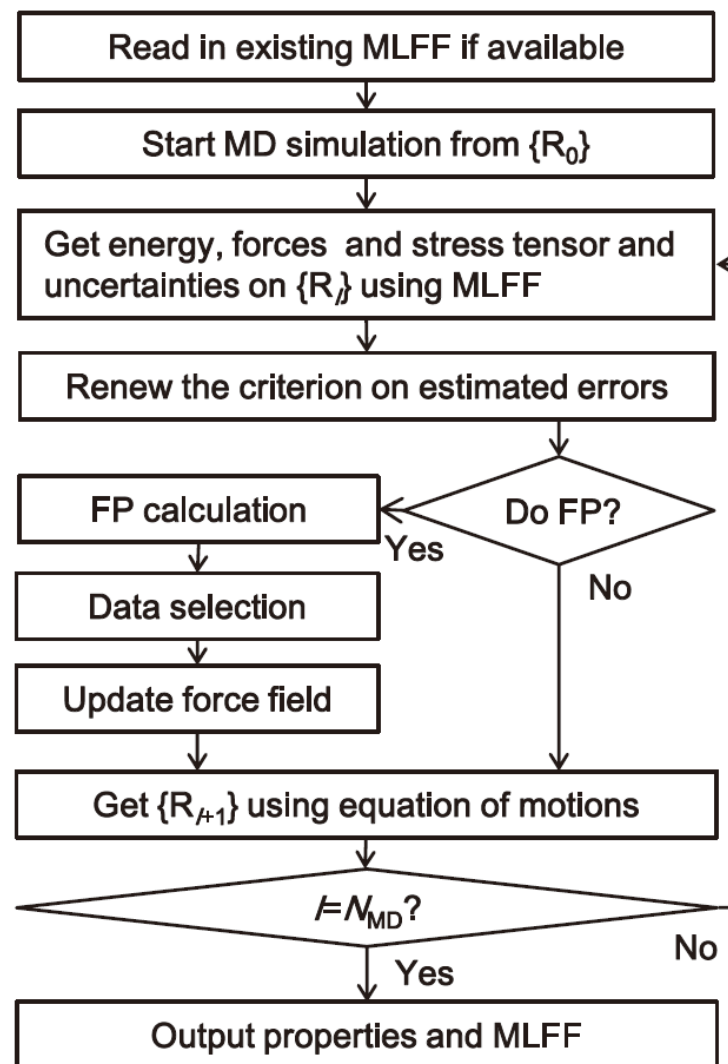
1. 神经网络
2. 态密度预测吸附能
3. 机器学习势
4. 实时从头算分子动力学
5. 新兴的机器学习计算程序

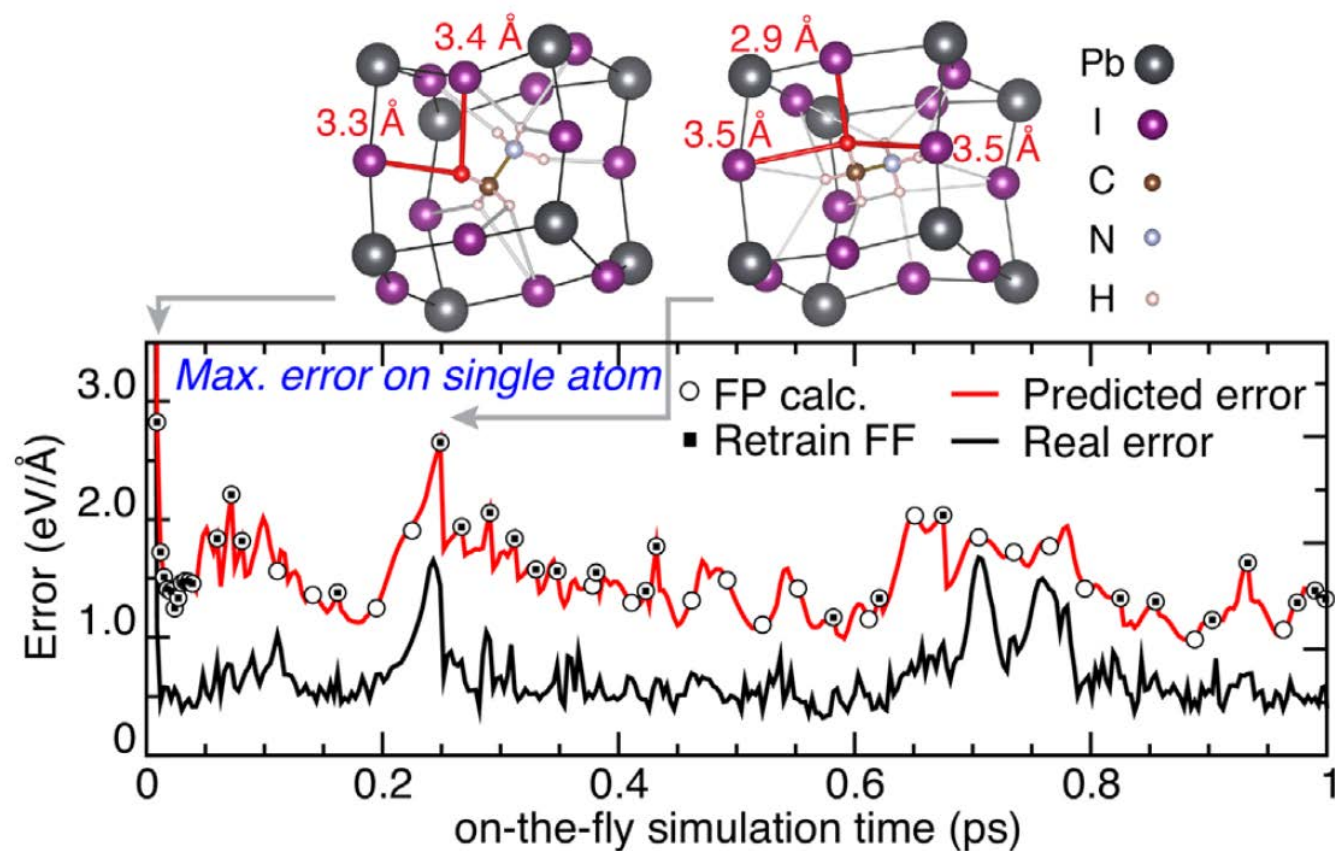
# 实时从头算分子动力学

机器学习势的缺陷：需要  
事先计算大量的 DFT 数  
据，并训练好力场

解决办法：“边算边用”

实时从头算分子动力学  
(on-the-fly AIMD)





(MA)PbI<sub>3</sub> 的模拟

1. 神经网络
2. 态密度预测吸附能
3. 机器学习势
4. 实时从头算分子动力学
5. 新兴的机器学习计算程序



Python Materials Genomics (pymatgen)  
用于材料分析的python库 (Python 3.7以上)

1. 高度灵活的类 (Element, Site, Molecule, Structure)
2. 支持各种输入输出 (VASP, ABINIT, CIF, Gaussian, XYZ等)
3. 分析工具 (生成相图, 扩散分析, 反应, 态密度等)
4. 结合Materials Project, Crystallography Open Database与其它数据库



```
In [1]: from pymatgen.ext.matproj import MPRester
```

```
In [2]: with MPRester("YOUR API KEY") as m:
        struct = m.get_data('Fe2O3')
        print(len(struct))
        print(struct[0])
        for i in struct:
            print(i['unit_cell_formula'], i['material_id'])
```

25

```
{'energy': -516.78134568,
 'sites': 80, 'unit_cell_formula': 'Fe2O3', 'e_above_hull': 0.2,
 'e': 'spglib', 'symbol': 'Fe2O3', 'material_id': 'mp-1844496'],
 'band_gap': 0.0,
 'en\ndata_Fe2O3\n_symmetry_47\n_cell_angle_alpha': 80.0}
```

```
In [4]: with MPRester("YOUR API KEY") as m:
        structure = m.get_structure_by_material_id("mp-1234")
        data = m.get_data('Li-Fe-O', prop = 'energy')
        print(structure)
        print(data)
```

Full Formula (Lu2 Al4)

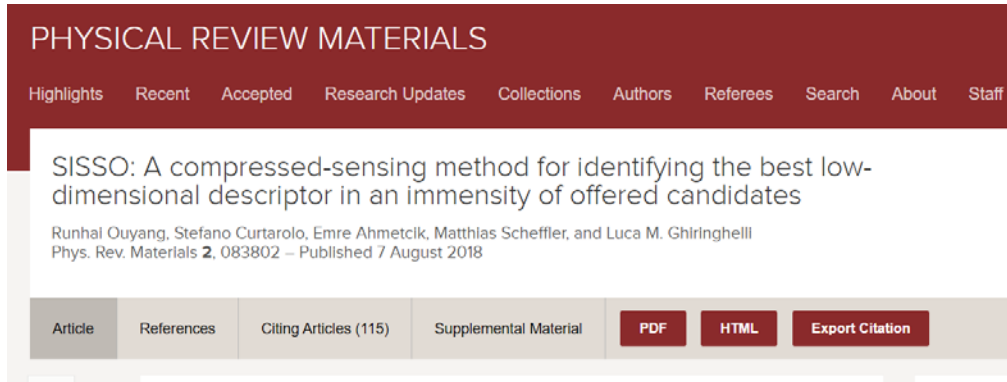
Reduced Formula: LuAl2

abc : 5.488740 5.488740 5.488740

angles: 60.000000 60.000000 60.000000

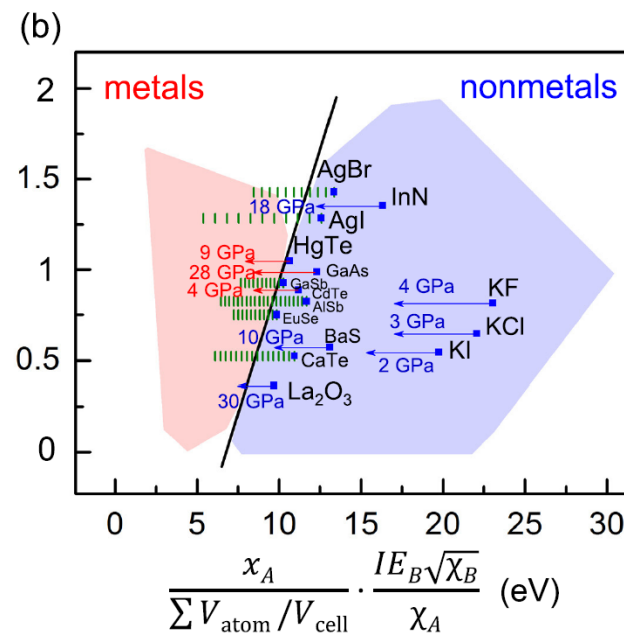
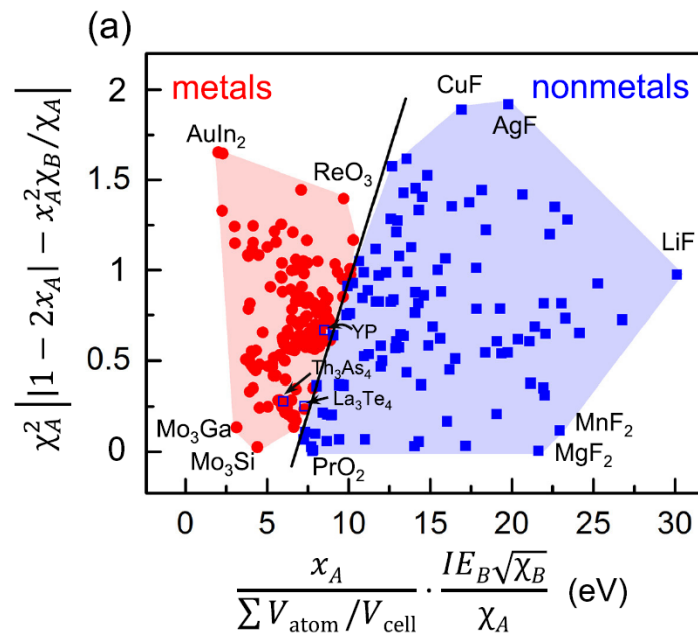
Sites (6)

#	SP	a	b	c	magmom
---	----	---	---	---	--------



## SISSO (Sure Independence Screening and Sparsifying Operator)

用于从高维描述符中识别出低维描述符



$\chi$  Pauling电负性

$IE$  电离能

$\chi$  原子组分

$V$  体积

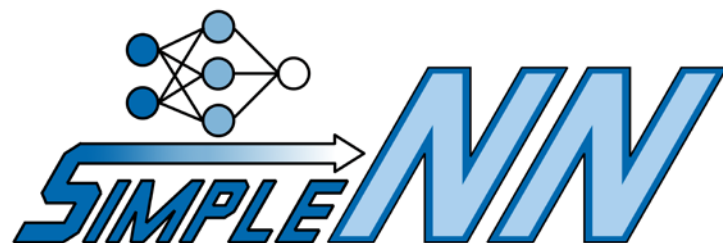
R. Ouyang, L. M. Ghiringhelli, *et. al. Phys. Rev. Mater.* **2018**, 2, 083802.

# MLFF相关软件



MAML

<https://github.com/materialsvirtuallab/maml>



SIMPLE-NN

<https://github.com/MDIL-SNU/SIMPLE-NN>

🏠 n2p2 - A neural network potential package

"v2.1.4-9-g72f58c0"

n2p2

<https://github.com/CompPhysVienna/n2p2>

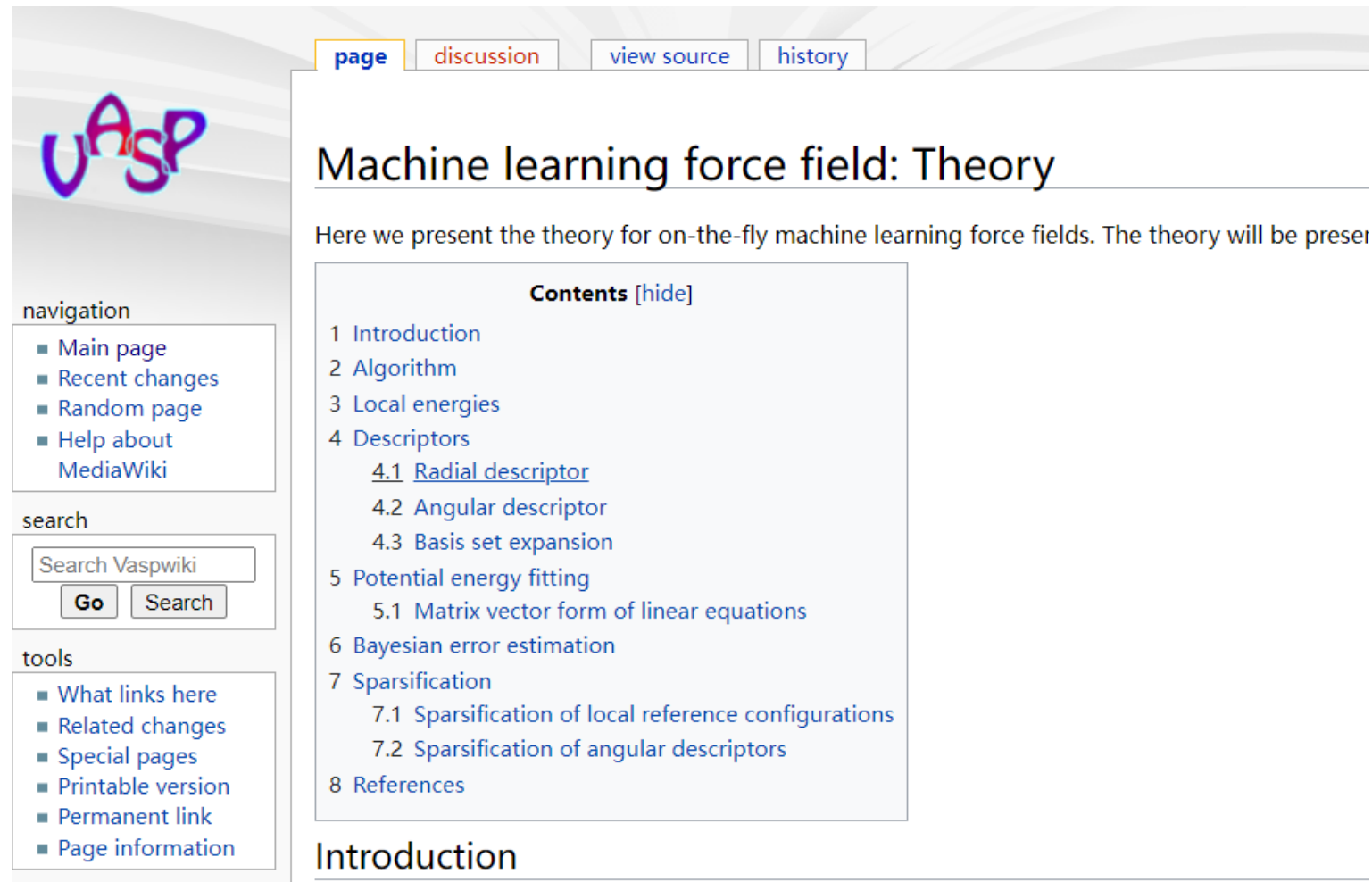


NequIP

<https://github.com/mir-group/nequip>

用于训练神经网络势，后续可使用 LAMMPS 或 ase进行MD模拟

## 从VASP 6.3开始，支持on-the-fly AIMD



The screenshot shows a MediaWiki page for 'Machine learning force field: Theory'. The page has a sidebar with navigation links (Main page, Recent changes, Random page, Help about MediaWiki), a search box, and tool links (What links here, Related changes, Special pages, Printable version, Permanent link, Page information). The main content area includes a title bar with 'page', 'discussion', 'view source', and 'history' tabs. The title is 'Machine learning force field: Theory'. Below the title is a paragraph: 'Here we present the theory for on-the-fly machine learning force fields. The theory will be preser'. A 'Contents [hide]' section lists the following topics: 1 Introduction, 2 Algorithm, 3 Local energies, 4 Descriptors (with sub-items 4.1 Radial descriptor, 4.2 Angular descriptor, 4.3 Basis set expansion), 5 Potential energy fitting (with sub-item 5.1 Matrix vector form of linear equations), 6 Bayesian error estimation, 7 Sparsification (with sub-items 7.1 Sparsification of local reference configurations, 7.2 Sparsification of angular descriptors), and 8 References. The 'Introduction' section is currently selected and expanded.

page discussion view source history

## Machine learning force field: Theory

Here we present the theory for on-the-fly machine learning force fields. The theory will be preser

**Contents** [hide]

- 1 Introduction
- 2 Algorithm
- 3 Local energies
- 4 Descriptors
  - 4.1 Radial descriptor
  - 4.2 Angular descriptor
  - 4.3 Basis set expansion
- 5 Potential energy fitting
  - 5.1 Matrix vector form of linear equations
- 6 Bayesian error estimation
- 7 Sparsification
  - 7.1 Sparsification of local reference configurations
  - 7.2 Sparsification of angular descriptors
- 8 References

### Introduction

**练习**永远是提升自己数据科学水平的最佳途径

三个推荐的网站



GitHub:  
<https://github.com>

LeetCode:  
<https://leetcode.com>  
<https://leetcode-cn.com>



kaggle

kaggle:  
<https://www.kaggle.com>