



# A Machine Learning Model for Grain Boundary Properties Prediction Based on Kernel Density Estimation

## 基于核密度近似方法的晶界性质机器学习预测模型

Jincheng Zhang<sup>1</sup>, Zhihang Qian<sup>1</sup>, Shuyang Gu<sup>1</sup>, Yongjoong Shin<sup>1</sup>

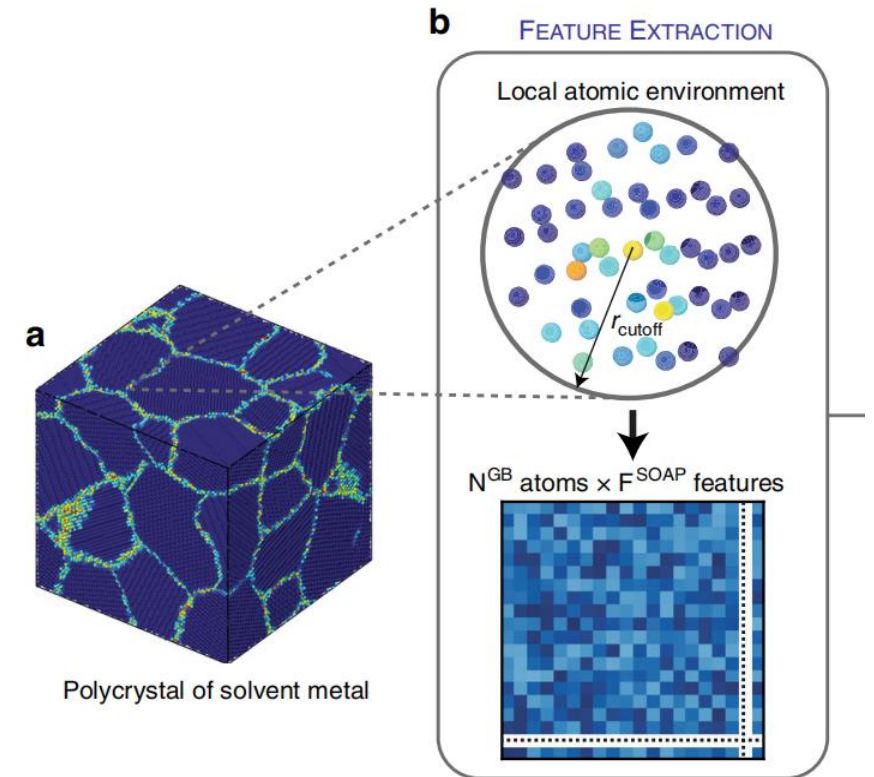
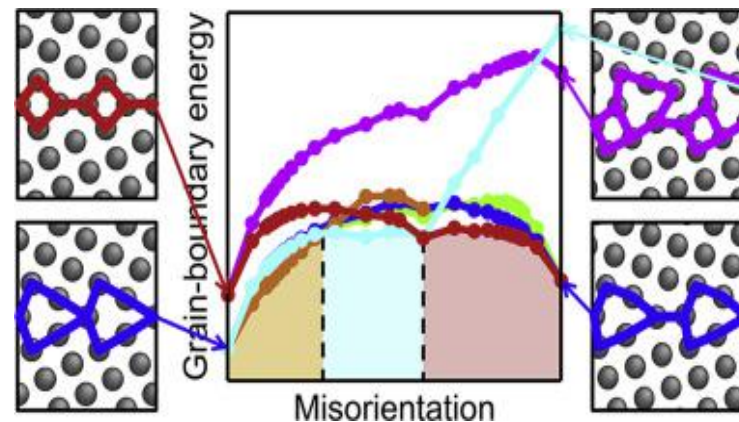
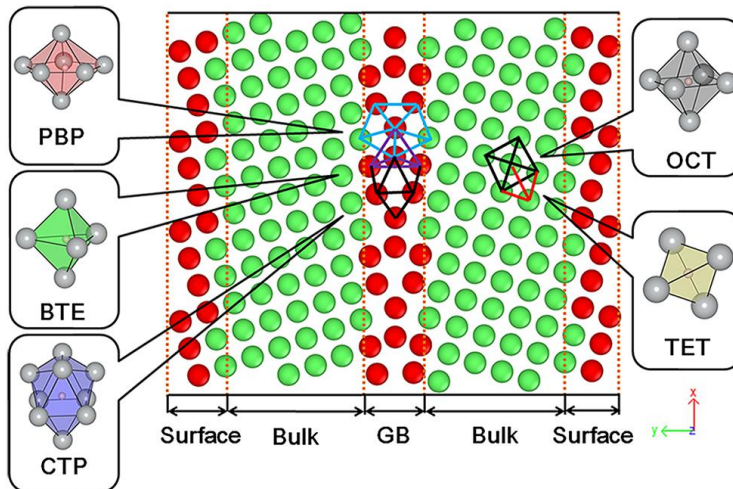
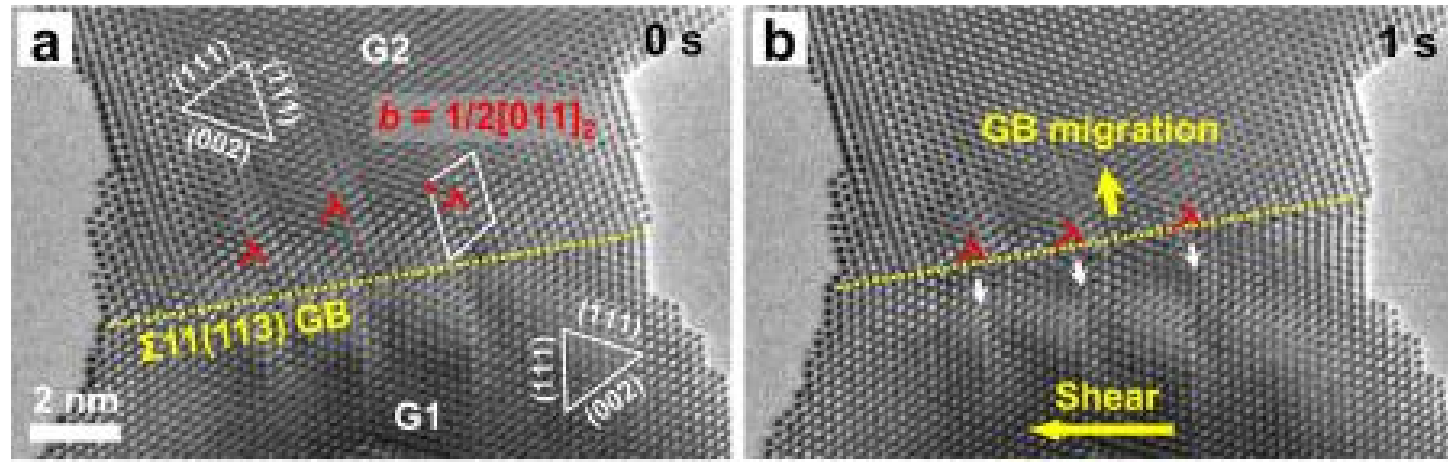
1. School of Materials Science and Engineering, Tsinghua University

2020.12.22

**Email: [jincheng18@mails.tsinghua.edu.cn](mailto:jincheng18@mails.tsinghua.edu.cn)**

# 1.1 Background —— Grain Boundary

晶界（GB）作为材料中广泛存在的重要缺陷，它的结构和行为很大程度上决定了多晶材料的物理、化学和力学性能。一个多世纪以来，晶界结构和行为的研究一直是材料科学的一个研究焦点。



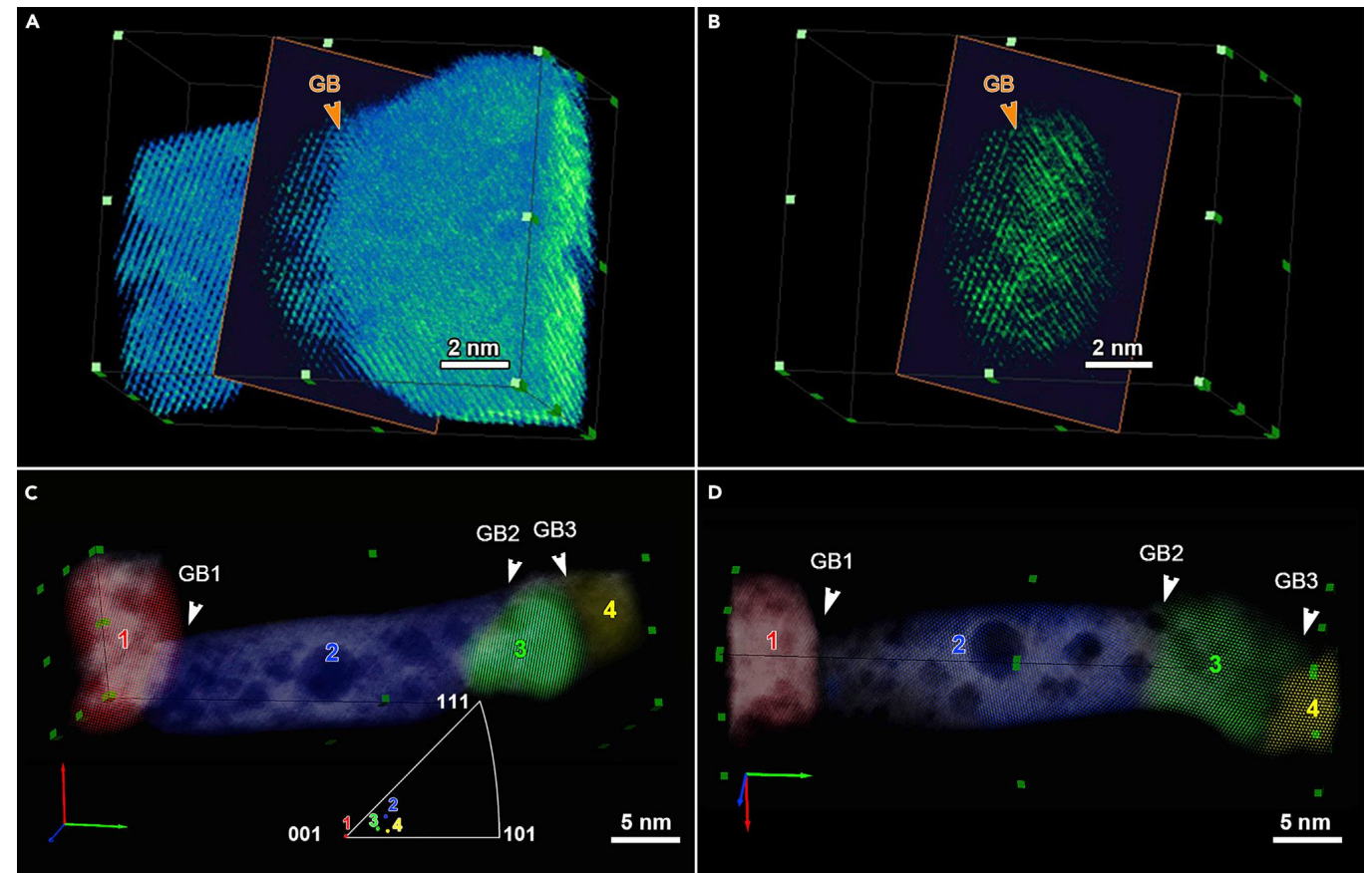
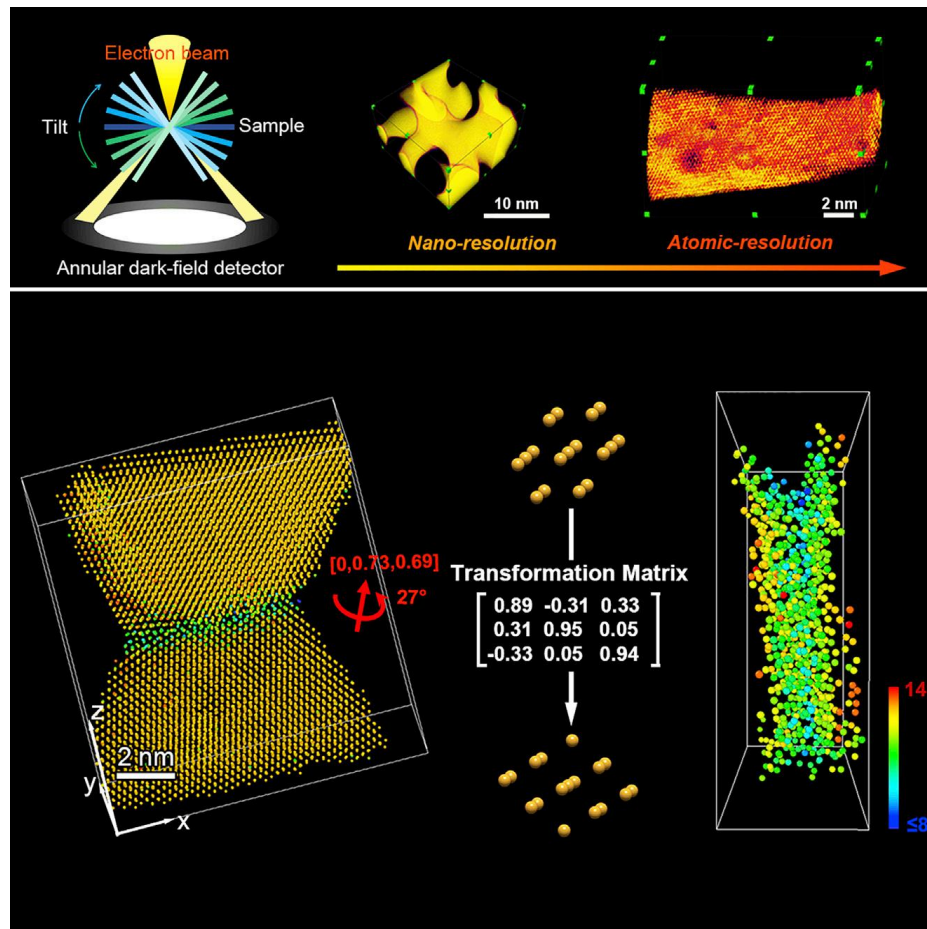
Xiao Zhou et al., Physical Review Letters, Volume 116, Issue 7, 2017

Malik Wagih et al., Nature Communication, Volume 11, Issue 1, 2020



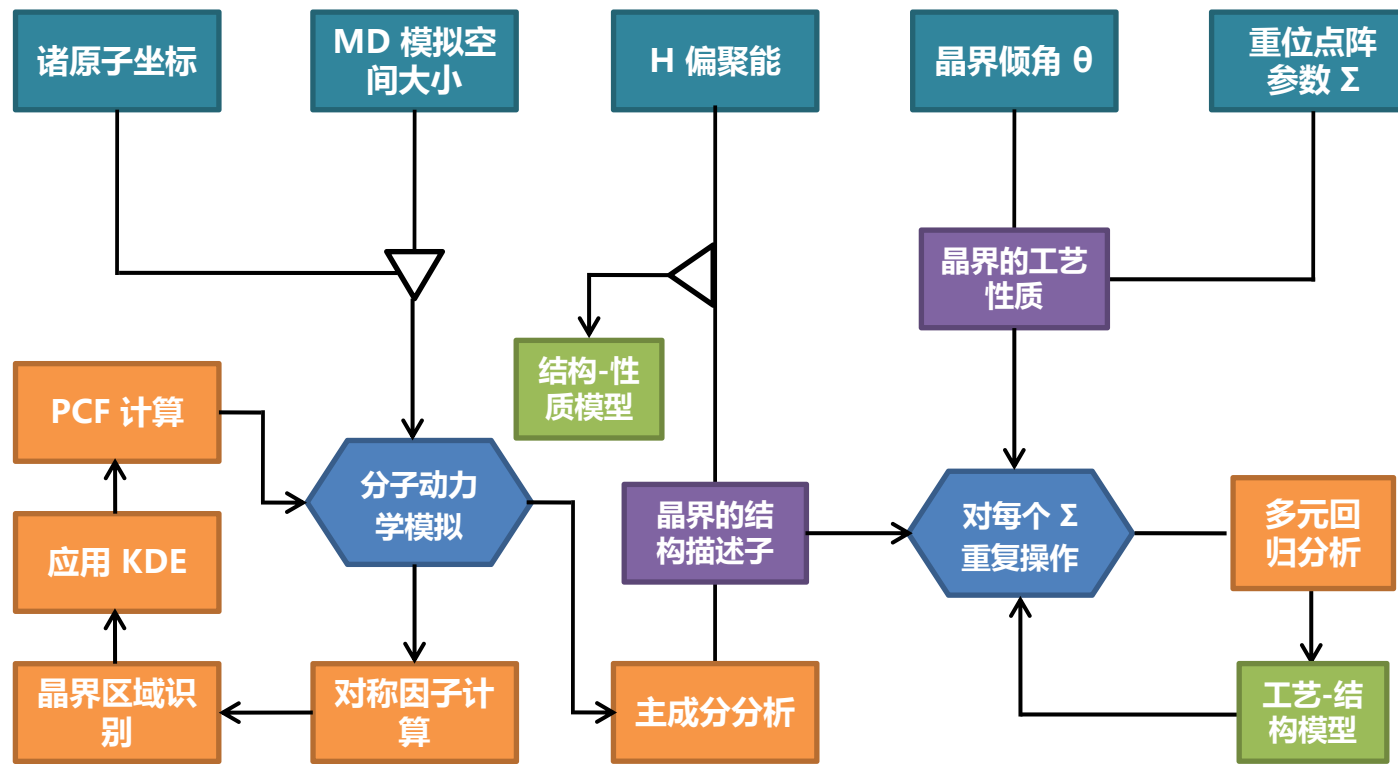
## 1.2 Background —— Recent Progress in Experiment

传统的透射电子显微技术（TEM）已经将材料研究推进到亚埃尺度。而日前发展出的电子层析三维重构技术（Electron Tomography）可以通过三维成像直接从纳米，甚至原子尺度解析材料的三维结构。为晶界结构的精确分析奠定了基础。

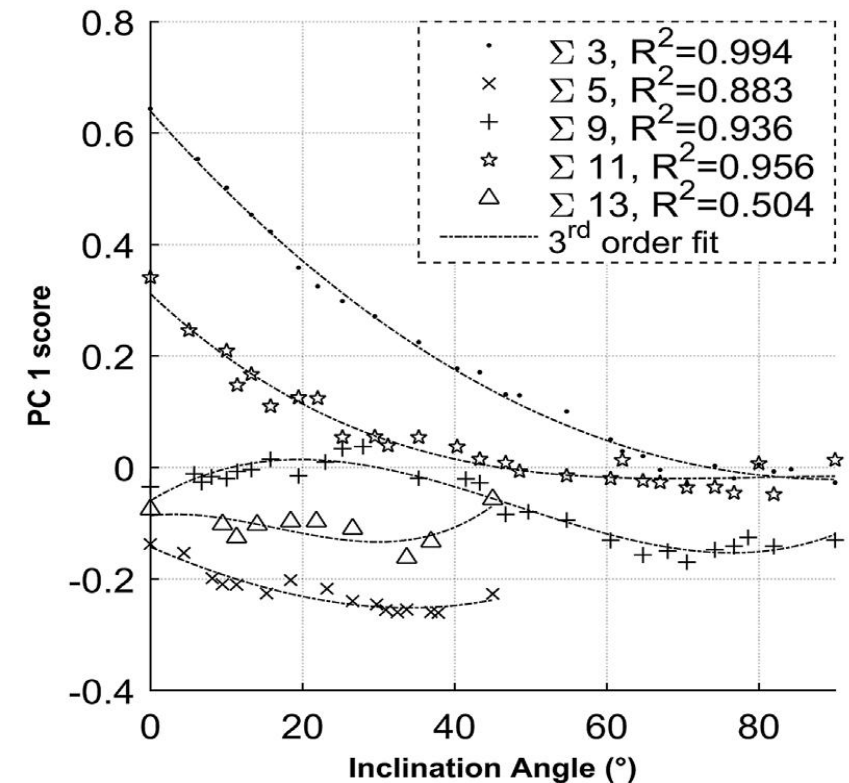


## 2.1 Motivation —— PSP models

**PSP 模型 (Process-Structure-Property)**，一方面可由材料的微观结构的描述子预测材料性质，另一方面可由所需的结构或性质反推需要的工艺参数（对晶界而言指它的倾角  $\theta$  和重位点阵参数  $\Sigma$  等）



**Fig.2.1** 本研究所采用的工作流，用于从晶界分子动力学模拟产生的数据中建立 PSP 关系模型



**Fig.2.1** 晶界工艺-结构关系（非宏观上的“工艺”）的多元回归分析



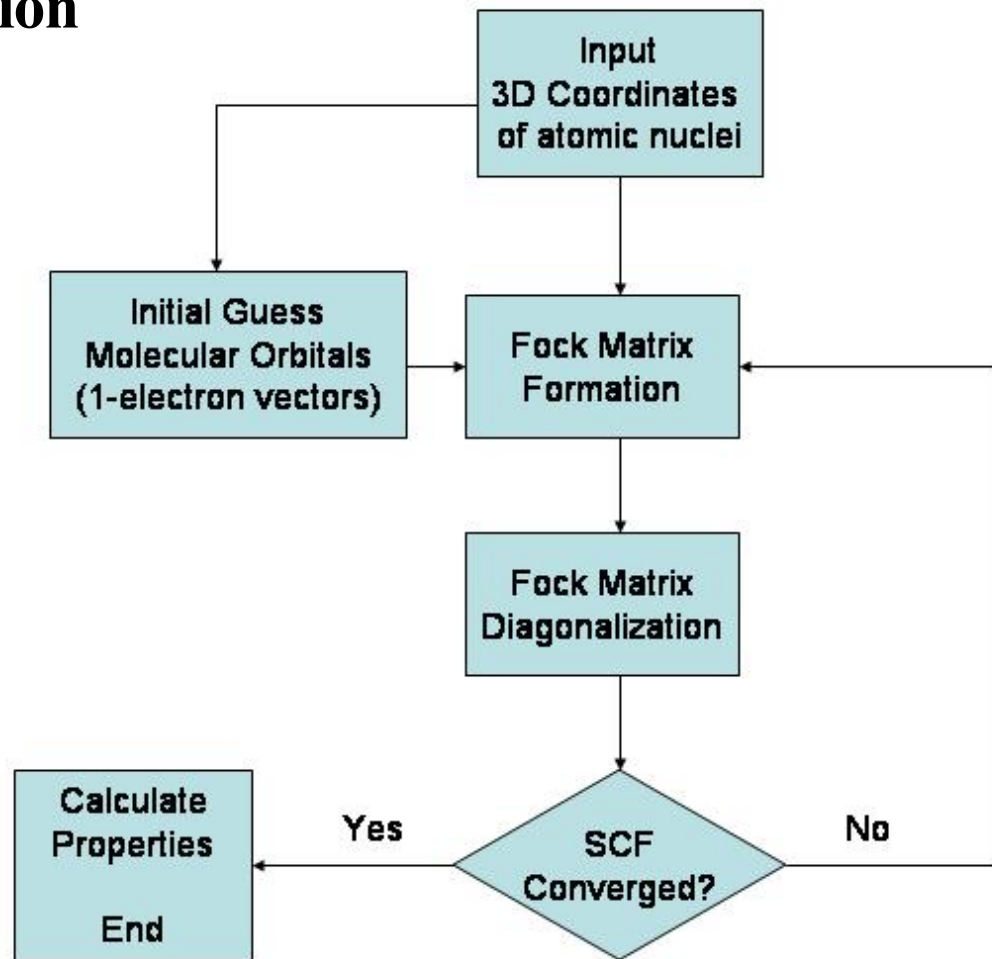
## 2.2 Motivation —— High Cost of DFT

### Estimated time duration of AB initio calculation

目前，几乎所有第一性原理的计算基于Hartree-Fock 方法或者基于算法源于 H-F 法，因此分析计算时间需要参考HF方法的计算流程以及实际计算时间。H-F法的计算流程如图：

理论上 Hartree-Fock 方法的计算时间与基底数有关，如果模型的基底数为  $N$  时，所需要的计算时间为  $N^4$ ，即基底数变为原来的 2 倍时，所需要的时间为原来的 16 倍。

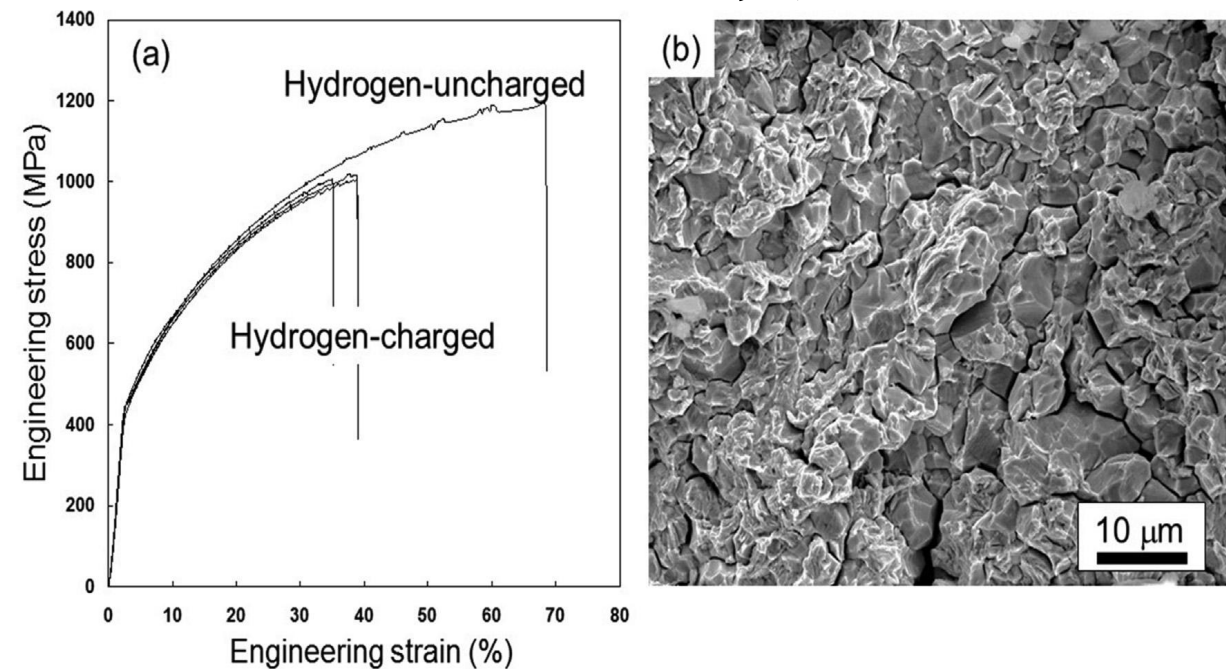
目前通过计算方法的改善，在计算过程中会有效排除接近于 0 的积分项，于是计算时间与接近于  $N^3$ 。而改进的方法会需要更长的计算时间，比如 Møller-Plesset perturbation theory, MP2 所需要时间为  $N^5$ , MP4 为  $N^6$ , 若是 Coupled cluster 方法(CC法)则会需要  $N^7$  的时间。





# 3.1 Case Study — Hydrogen Embrittlement (HE)

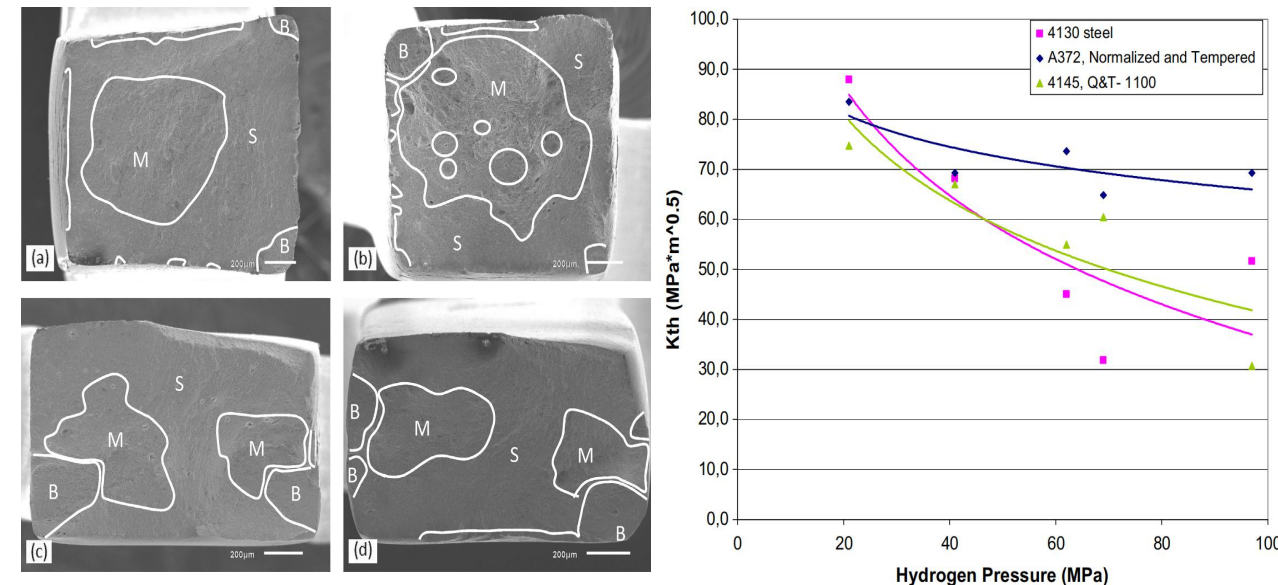
氢脆（hydrogen embrittlement, HE）是指金属材料在富氢环境中长期服役时，材料由于氢的作用而造成机械性能（强度）严重退化，发生脆断的现象



**Fig.3.1** (a) Ductility degradation in a hydrogen-charged Fe-18Mn-0.6C TWIP steel.

(b) SEM graph of an intergranularly fractured surface after HE.

氢脆敏感性一直是限制先进高强钢（AHSS）广泛应用的一个重要因素，在35 ~ 75MPa 商用高压储氢环境中使用的金属可能出现氢脆，引发脆性破坏

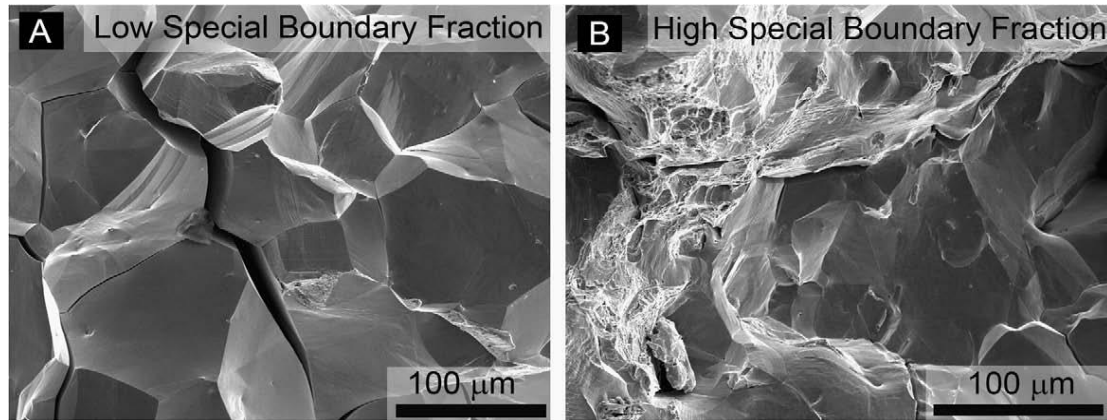


**Fig.3.2** Fracture appearance of (a) 980DP, (b)1200DP, (c) 980DP-650YS and (d)980DP-700YS

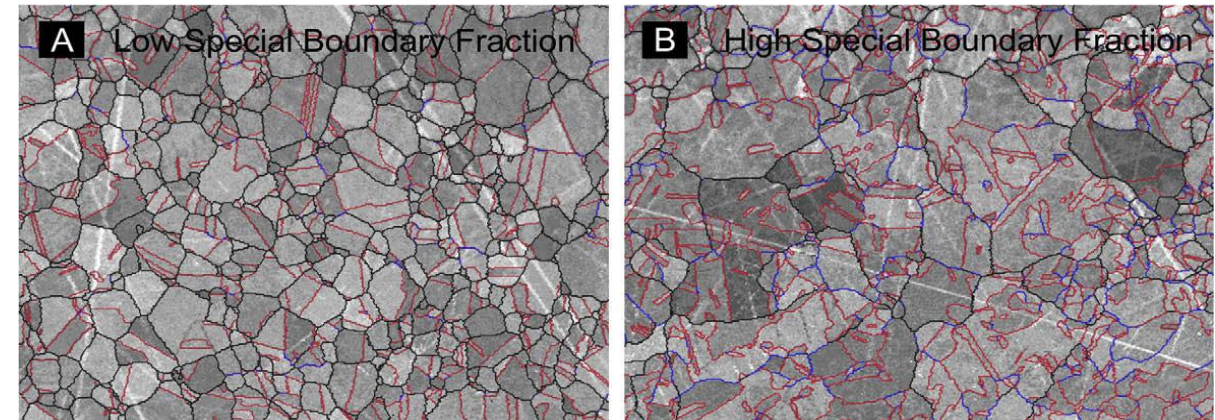
**Fig.3.3** Threshold stress intensity factors for several moderately affected carbon steels

## 3.2 Case Study — Anti-HE GBs Engineering

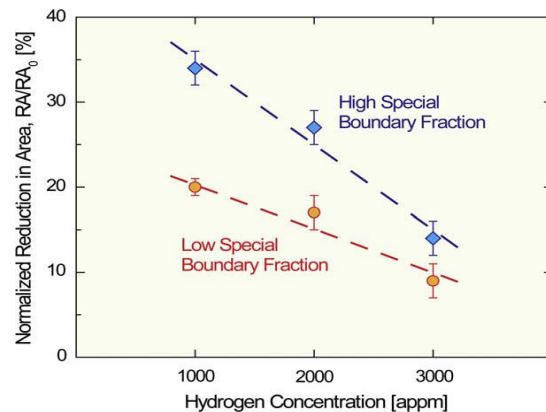
### ➤ 针对性地改变金属材料中的晶界（GBs）可能设计出低 HE 敏感性的材料



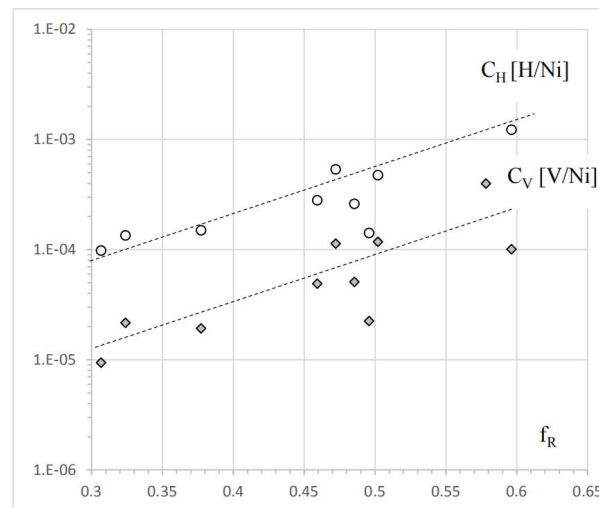
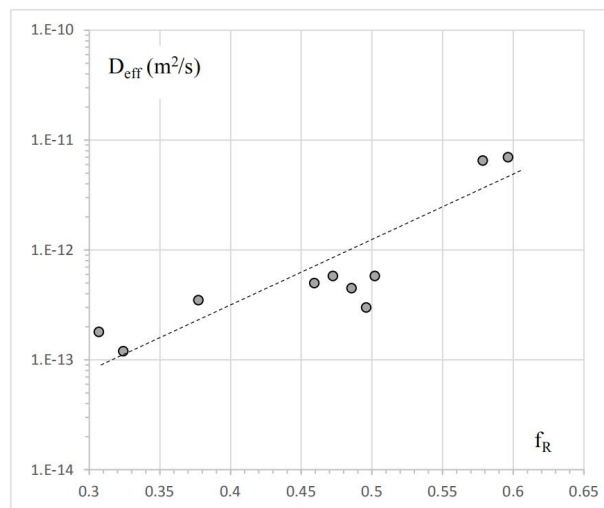
**Fig.3.4** 处理前后 Ni 的 HE 敏感性测试结果：(A)低特殊晶界含量组展示了100% 的穿晶断裂；(B)高特殊晶界含量组展示了更多的韧性断裂特征



**Fig.3.5** (A) 低比例(46%)和(B)高比例(75%)特殊晶界的 Ni-201 微结构的 EBSD 取向图。 $\Sigma 3$ 孪晶边界用红色表示， $\Sigma 3^n$ 特殊边界用蓝色表示



**Fig.3.6** 处理前后 Ni 的性能对比，数据：塑性变形区域占比



**Fig.3.7** (a) H 有效扩散系数 - 高角度晶界含量；(b) H 溶解度 & 空位集中度 - 高角度晶界含量

S. Bechtle et al. / Acta Materialia 57 (2009)  
A. Oudriss et al. / Scripta Materialia 66 (2012)



## 4.1 Methodology — Identifying Trapping Sites

固体中的晶界可视为晶格对称性破缺的位点，所以可以通过计算人为定义的对称因子来识别模拟体系中的晶界：

$$SP = \left| \sum_{i=1}^{12} \vec{R}_i \right|^2 \min_{\alpha(x_0), \beta(x_0), j=1, \dots, d} \sum_{i=1}^N K_{\lambda}(x_0, x_i) \left[ y_i - \alpha(x_0) - \sum_{j=1}^3 \beta_j(x_0) x_i^j \right]^2$$

上式右边为利用 Epanechnikov 核函数所作的局域回归，目的是让分界面的判断和 SP 的具体取值无关。

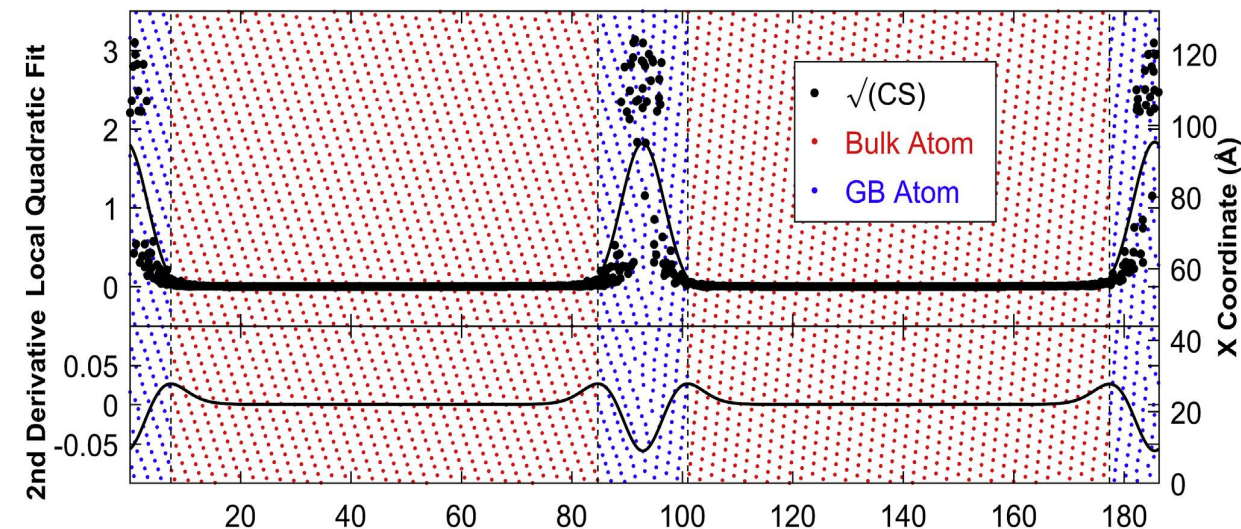


Fig.3.1 以对称因子局域回归的二次导数最大值点作为晶界原子和晶块原子的分界面

借用地理分析中常用的Voronoi多边形方法可确定固体中H偏聚位点的大致位置

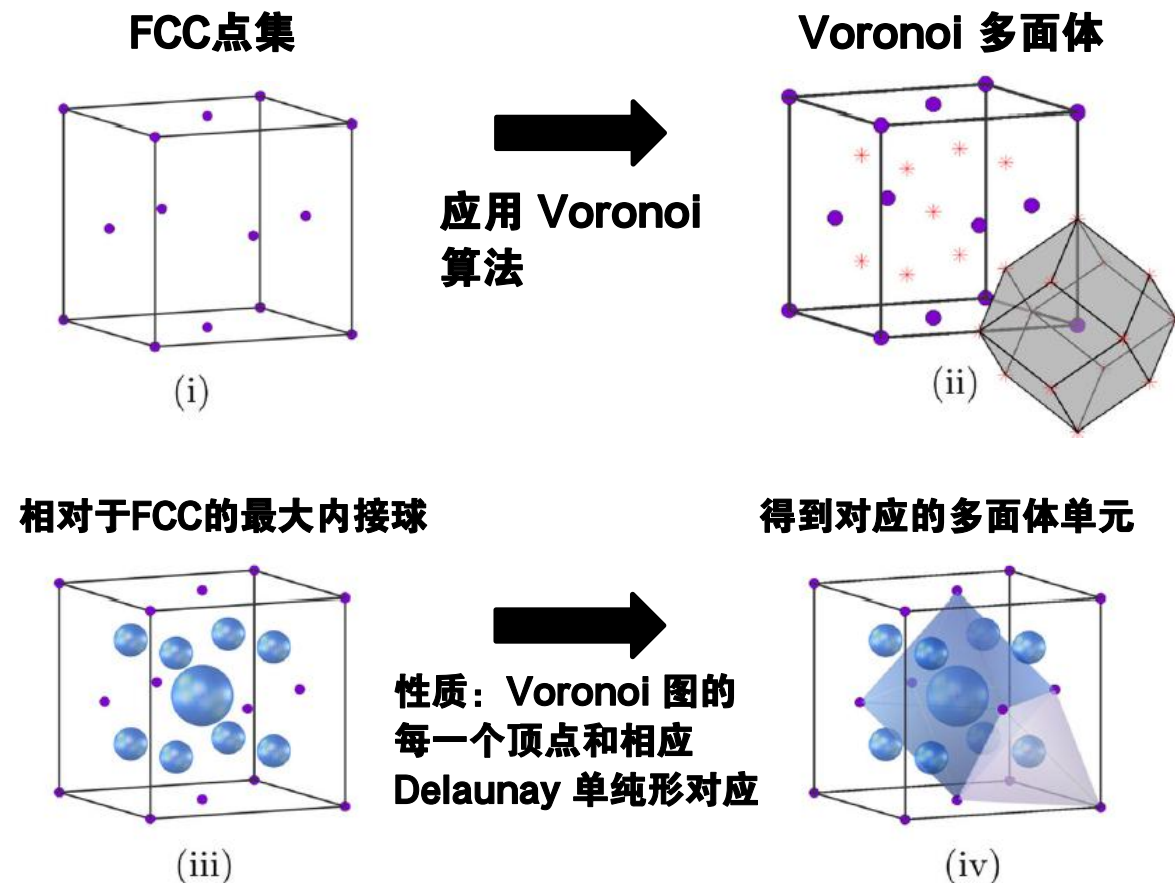


Fig.3.2 以完美 FCC 晶格为例，简单说明本研究所采用的 Voronoi 算法原理



## 4.2 Methodology — KDE & PCF Calculation

传统的 PCF 采用狄拉克  $\delta$  函数作为加和项：

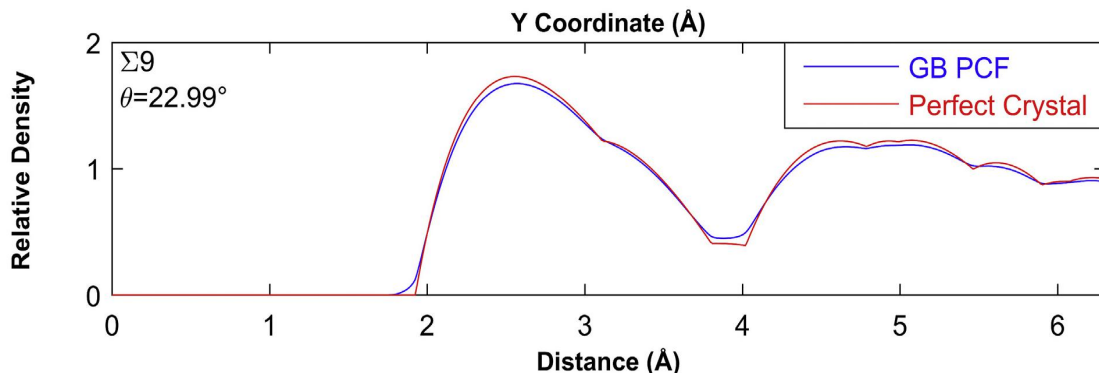
$$\psi'_i(r) = \frac{1}{N_{cutoff}} \sum_{k=1}^{N_{cutoff}} \delta(r - \|\vec{R}_{i,a}^{(k)}\|)$$

应用 KDE 改进后的 PCF 如下表示：

$$\psi_i(r) = \frac{1}{N_{cutoff}} \sum_{k=1}^{N_{cutoff}} k_e(r - \|\vec{R}_{i,a}^{(k)}\|, h_e)$$

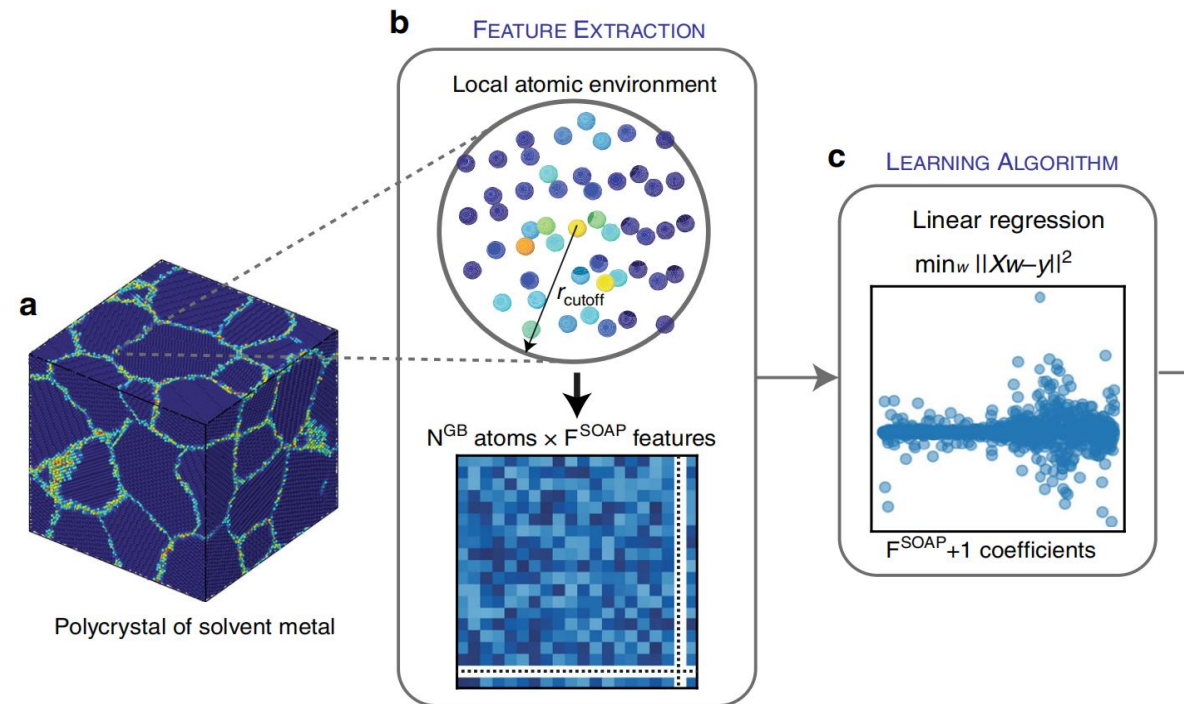
其中所采用的 Epanechnikov 核函数定义如下：

$$\kappa^e(u, h) = \begin{cases} \frac{3}{4\sqrt{5}h^2} \left(1 - \frac{u^2}{5h^2}\right) & \text{for } (u/h)^2 < 5 \\ 0 & \text{otherwise} \end{cases}$$



**Fig.4.1** 上图显示晶界的平均对相关函数（GB PCF）与理想 FCC 晶体的比较。

简单的说，径向分布函数 PCF 是用来描述某个所研究粒子周围其他粒子分布情况的物理量（描述子，descriptor），其积分代表平均配位数。对本研究而言，计算经过 KDE 改进的 PCF 函数可以提供一个大  $N^{GB} atoms \times N^{PCF} samples$  的高维数据数组以作为机器学习模型的输入



**Fig.4.2** 对于  $20 \times 20 \times 20 \text{ nm}^3$  退火的多晶 Al 体系 (原子用对称参数着色)，计算每个已识别的 GB 中偏聚位点的 PCF，转换为特征矩阵，作为 (c) 学习算法 (线性回归) 的输入

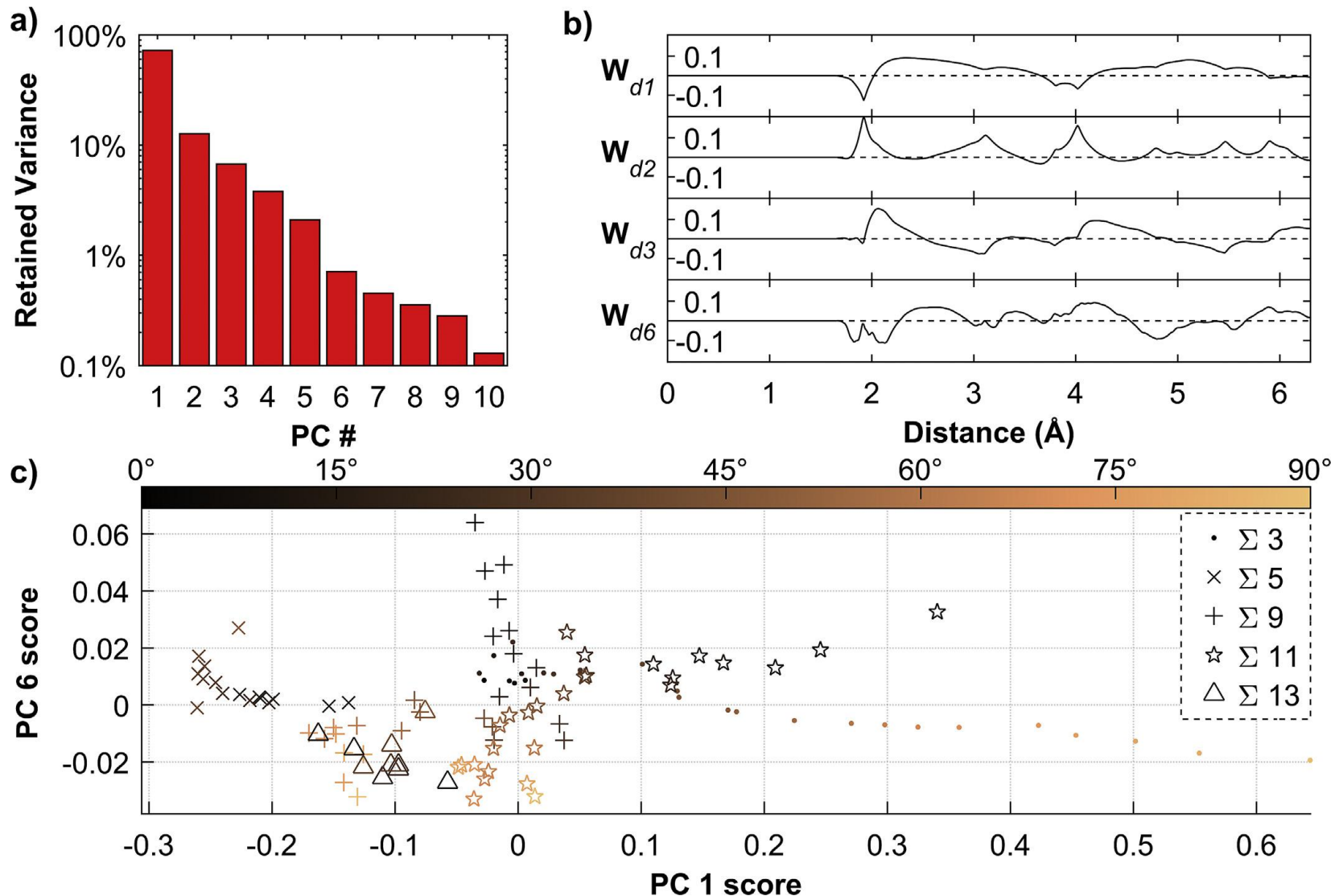
## 4.3 Methodology — PCA Analysis

PCF 代表了原子局域信息的详细量化，它是一种高维数据结构（相当于 PCF 采样点的数量）；

为了便于比较不同 H 偏聚位点的信息，并建立与 GB 倾角或 GB  $\Sigma$  的联系，可以使用维数缩减技术；

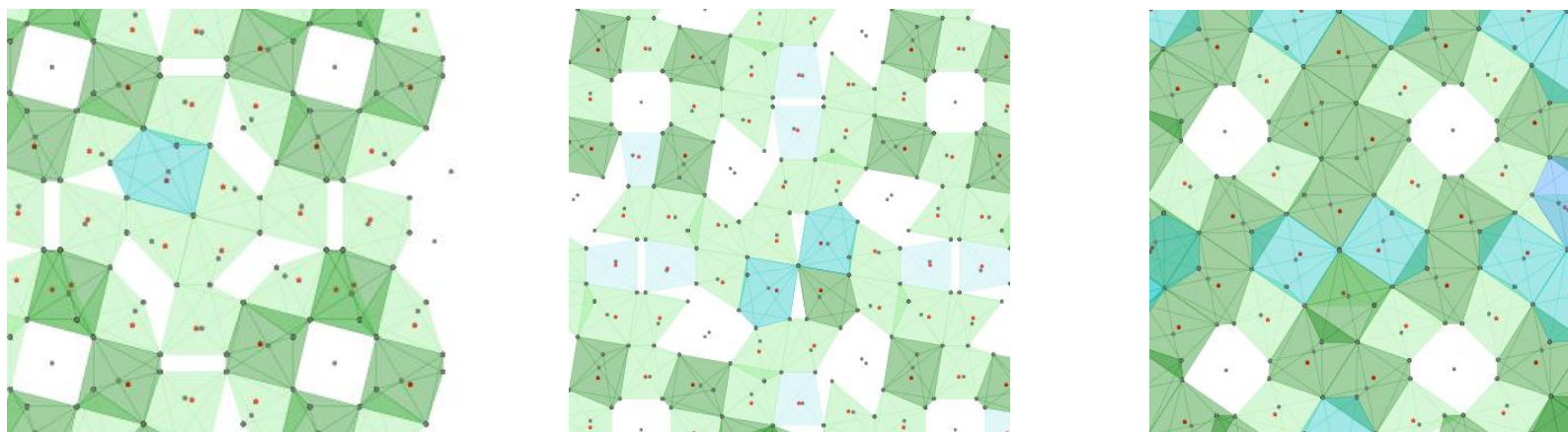
PCA 是用于降维的常用技术，它可以确定数据的一组正交基，其中第  $i$  个特征向量对应于具有第  $i$  个最大方差的方向，如图（a）所示；

从每个 H 偏聚位点的离散采样 PCF 中减去 GB 平均 PCF 后，通过奇异值分解进行主成分分析（PCA），得到 PCF 相互正交的一组基，如图（b）所示；

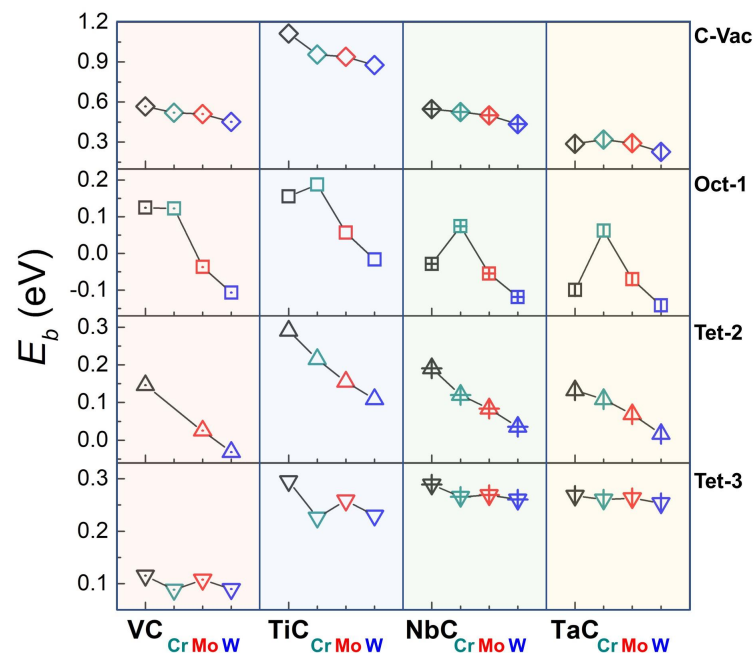




# 5.1 Results

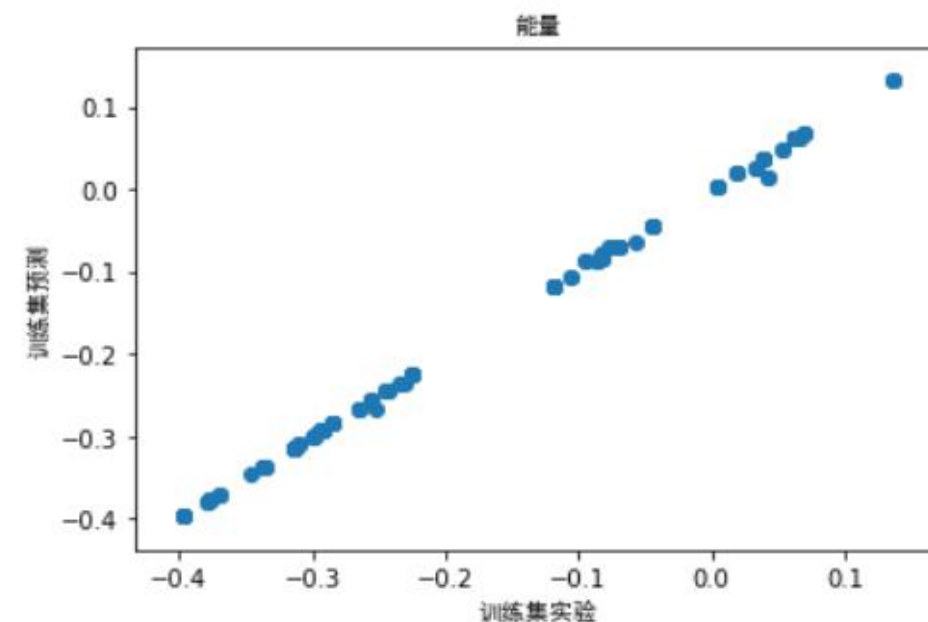


**Fig.5.1** 应用 Voronoi 算法对引入了点缺陷、线缺陷的模拟  $\alpha$ -Fe 实际体系中的不同晶界进行多面体单元识别 (a)  $\alpha$ -Fe twist  $\Sigma 17(001)$  (b)  $\alpha$ -Fe twist  $\Sigma 37(001)$  (c)  $\alpha$ -Fe twist  $\Sigma 13(001)$



**Fig.5.2** VC/BCC-Fe, TiC/BCC-Fe, NbC/BCC-Fe, TaC/BCC-Fe, TaC/BCC Fe-PBs不同间隙位的氢结合能。

**Fig.5.3** RF 算法结果交叉验证 (  $R^2$ : determination coefficient ) Testing set  $R^2$ : 0.998





欢迎大家批评指正！