

## LASP 使用入门5 - 拟合全局神经网络势函数-准备数据

**SSW 轨迹：自动随机方向，不惧高能垒，轨迹连续**  
能够对势能面自动大范围采样

人工智能实现的基础，就是**大数据**。神经网络具有强大的拟合（函数）能力，能对大量势能面数据同时拟合。

通过把第一性原理计算（如**VASP**计算）和**SSW**结合，得到势能面大数据，再通过神经网络拟合这些大数据，有希望得到通用性强，精度高，具备反应预测能力的势函数（可参见相关理论论文）。

**LASP**提供了准备势能面大数据的计算平台

## LASP 使用入门5 - 拟合全局神经网络势函数-准备数据

**关键1:** 第一性原理SSW计算时，务必打开结构/能量/力输出选项 `SSW.printevery T`

`allstr.arc` 结构/能量  
`allfor.arc` 力/应力

从allstr.arc中，

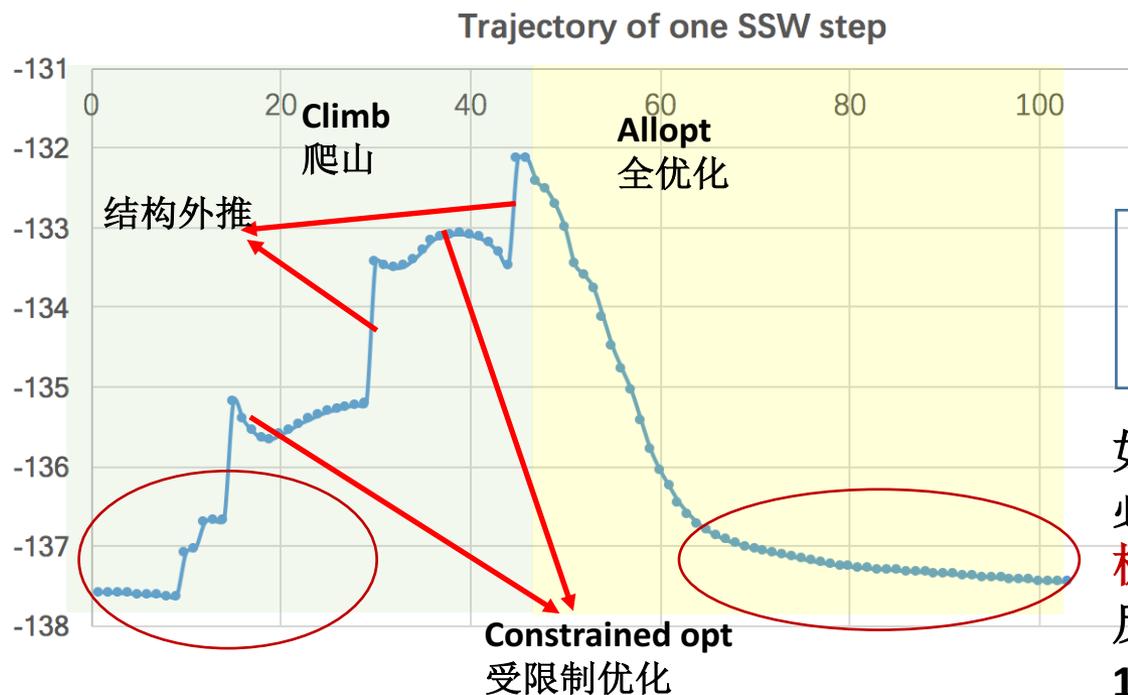
`grep Energy allstr.arc`

简单画出SSW轨迹（左图）

```
SSW.printevery T
SSW.printselect 1
SSW.printdelay -1
```

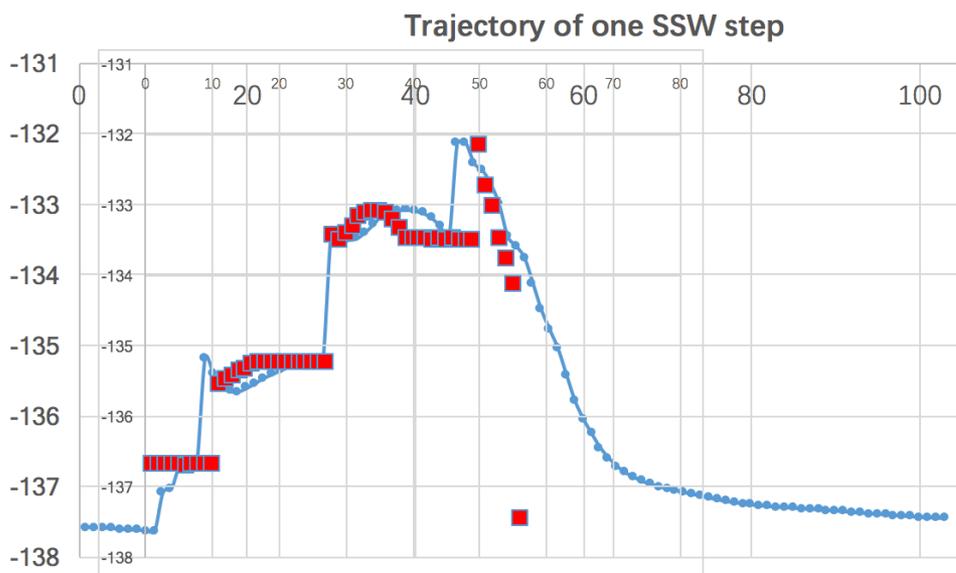
如果直接从所有SSW轨迹取点，必然会集中在极小点附近（图中两圆圈），反而

1. 能量高的区间取点很少
2. 结构外推处没有取点



## LASP 使用入门5 - 拟合全局神经网络势函数-准备数据

只有对势能面有效均匀采样, 势函数才能够具有更好可移植性和反应预测性, 建议选择 `SSW.printselect 6`



```
SSW.printevery T  
SSW.printselect 6  
SSW.printdelay -1
```

改变SSW.printselect, 会大大减少SSW轨迹取点

如图, 红色点, 相比蓝色点

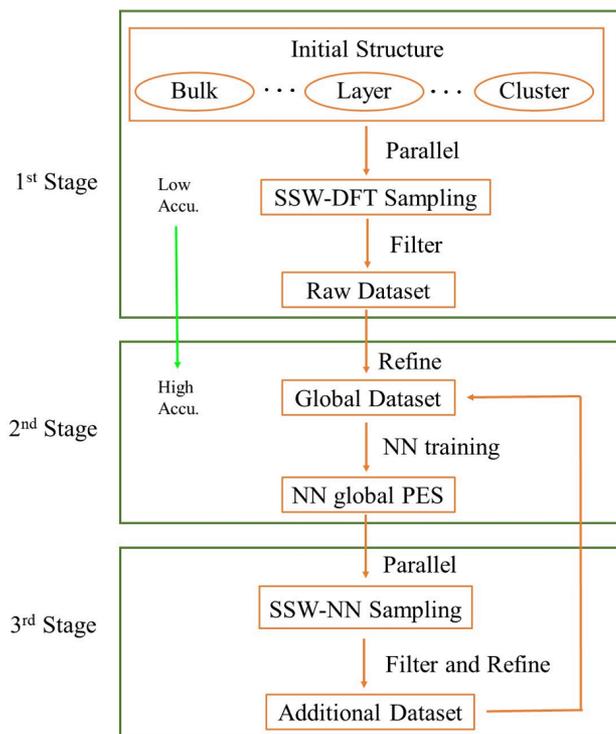
1. 去掉能量低区间点
2. 结构外推处插值取点 (由于是插值的, 能量没有真正计算, 看起来还是有能量跳跃)

## LASP 使用入门5 - 拟合全局神经网络势函数-准备数据

**关键2:** 尽量数据集结构多样性, 如从不同 (随机) 结构 (相) 进行SSW模拟

NN拟合过程如左图

第一性原理SSW计算是第一阶段, 最费时



- SSW取样不需要高精度

保证科学性下, 尽量降低Cutoff, K-mesh等

- 控制原子数 (每晶包) 一般 16~32 左右

不能太少 (晶胞太小就没液体结构了)

- 尽量并行多目录跑SSW

比如20个作业, 不一样的初始结构

最好其中一些接近目标体系

- 减少元素种类

目前NN架构, 三种元素以下代价小

## LASP 使用入门5 - 拟合全局神经网络势函数-准备数据

**关键3:** 拟合用数据集大小 ---- 尽量小（高度代表性）

从巨量的SSW轨迹结构中取十分之一到百分之一数据

按照20个SSW目录，每个目录100个SSWsteps

(SSW.SSWsteps 100)

预期产生  $20 * 100 * 200 * 0.4 = 16$  万数据

随机取其中5000-10000数据，进行（适当）高精度单点计算（同样打开SSW.printevery），基本就可以拟合出较合理的NN势函数

后续NN势函数还需要迭代-自学习，即第二阶段和第三阶段（P4），但时间较短，代价也小。

## LASP 使用入门5 - 拟合全局神经网络势函数-准备数据

体验LASP的 NN（和DFT比比速度，比比精度）：

Examples 里面提供了

1. C, H, O, N 四元素普遍适用的有机势函数
2. TiO<sub>2</sub> 固体势函数

目前大批NN势函数在审核中，GPU版本的LASP程序中...

- NN的拟合本身还有较多复杂技术，短期不对LASP用户开放；专业人士可自行用LASP取样的数据，进行（NN或其他）函数拟合。
- 对于软件购买用户的特定NN势函数需求，可以通过向LASP团队定制；或通过用户自行做前期数据（本教程），LASP团队帮助做第二/第三阶段数据，及NN拟合的联合模式，共同促进NN模拟技术在材料，化学中的应用。