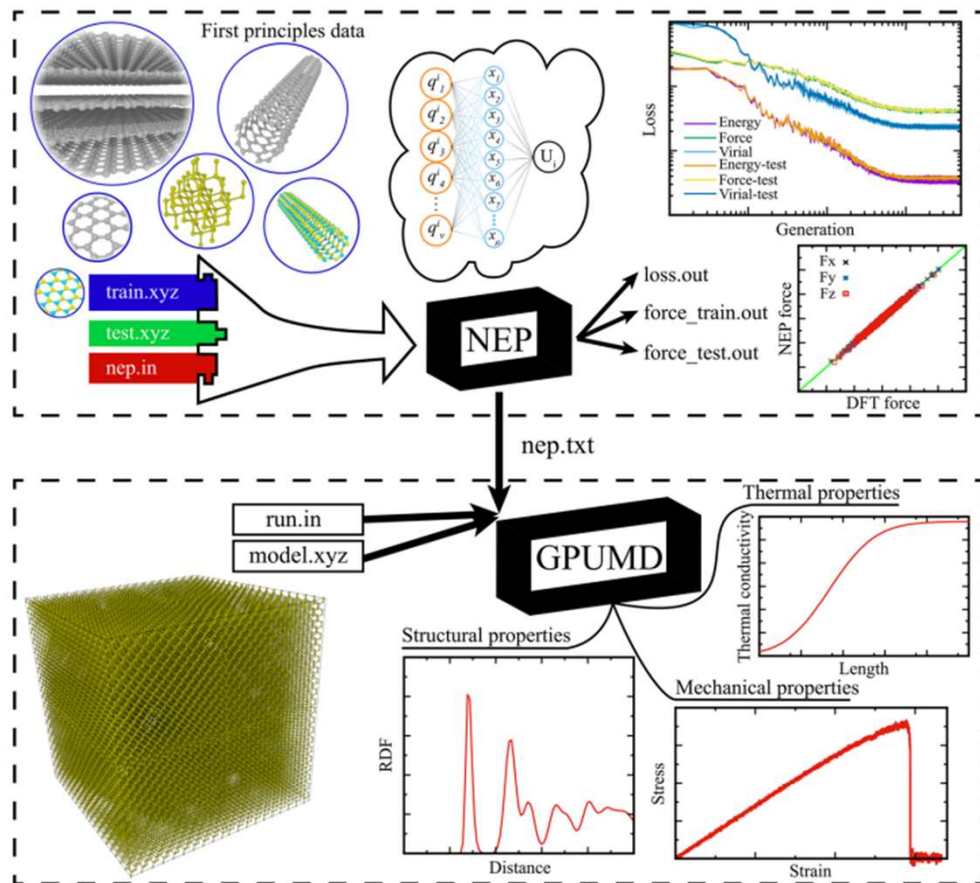


GPUMD的输入与输出

主讲人：应鹏华

第二届VASPKIT团队“机器学习”专题培训

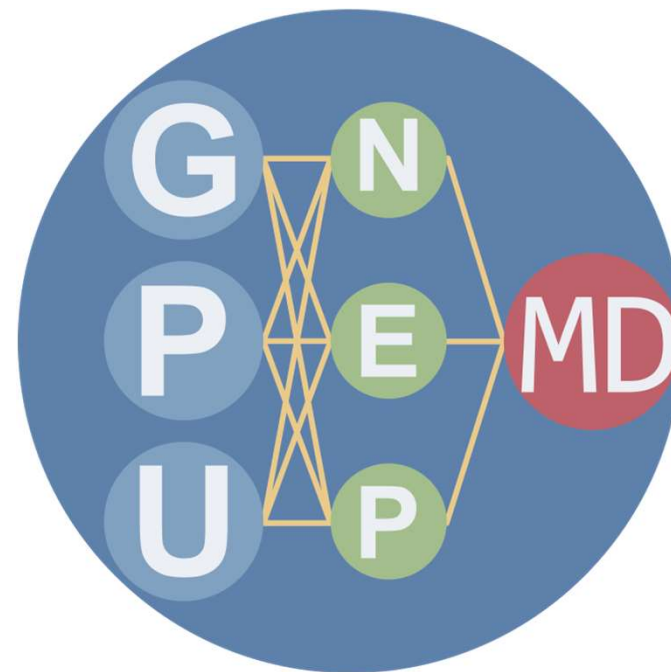
GPUMD&NEP的工作流



1. 准备训练集
2. 训练得到nep模型
3. 对得到的nep模型进行测试
4. 测试通过后, 采用gpumd进行MD模拟
5. 后处理MD模拟输出的轨迹和热力学数据
6. 分析测量物理量

下载和安装 GPUMD

1. 程序下载地址: <https://github.com/brucefan1983/GPUMD>
2. GPUMD 没有任何外部依赖, 程序仅使用标准的 C++和 CUDA
3. 在**windows**和**linux**下均可安装
4. 下载最新版本到本地, 然后解压缩
5. `find / -name nvcc` 找到nvcc执行文件的路径 (确认在cuda环境下)
6. 从终端进入 **src/** 文件夹, 然后敲 **make**
7. 编译成功后会在 **src/** 文件夹产生 **gpumd** 和 **nep** 两个可执行文件
8. 若有需要, 可以适当修改 **makefile**



实时演示task1: 在TEFS下安装最新版本的GPUMD

第二届VASPKIT团队“机器学习”专题培训

GPUMD&NEP社群

GPUMD仓库: <https://github.com/brucefan1983/GPUMD>

GPUMD例子: <https://github.com/brucefan1983/GPUMD/tree/master/examples>

GPUMD手册: <https://gpumd.org/>

GPUMD 中文QQ群: 887975816 (非常活跃, 有问必答)

NEP势函数开源模型: <https://gitlab.com/brucefan1983/nep-data/>;

<https://materialsmodeling.org/models-nep/>

基于GPUMD所发表的文章: <https://gpumd.org/publications.html>

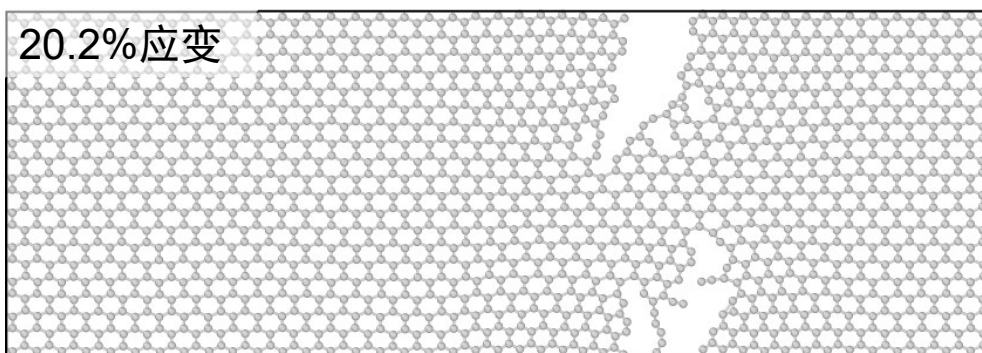
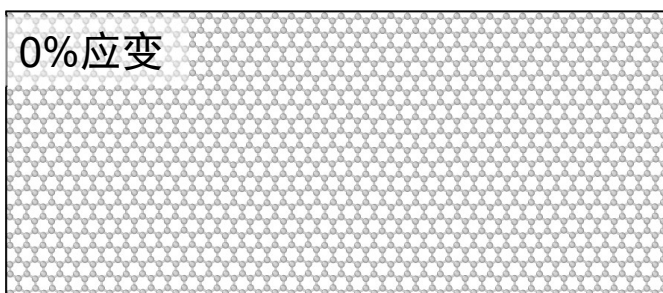
GPUMD&NEP微信公众号: GPUMD与NEP

GPUMD&NEP bilibili账号: GPUMD-NEP (<https://space.bilibili.com/3546727982303930>)

第二届VASPKIT团队“机器学习”专题培训

让我们运行第一个gpumd任务！

石墨烯单轴拉伸模拟



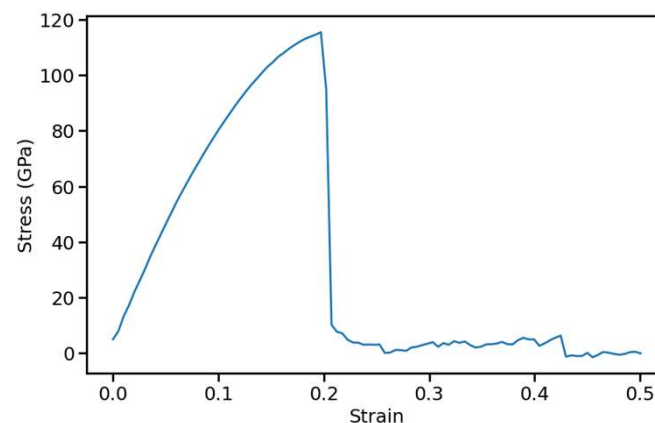
初始和断裂构型

实时演示task2: 在TEFS下运行石墨烯单轴拉伸的模拟

这些输入命令什么意思？让我们一起查阅gpumd.org。

```
1 potential      C_2024_NEP4.txt
2
3 time_step      1.0
4
5 velocity       300
6 ensemble       npt_ber 300 300 100 0 0 0 1000 1000 1000 1000
7 run            10000
8
9 ensemble       npt_scr 300 300 100 0 0 0 1000 1000 1000 1000
10 deform         5e-4 1 0 0
11 dump_thermo    1000
12 dump_position  1000
13 run            100000
```

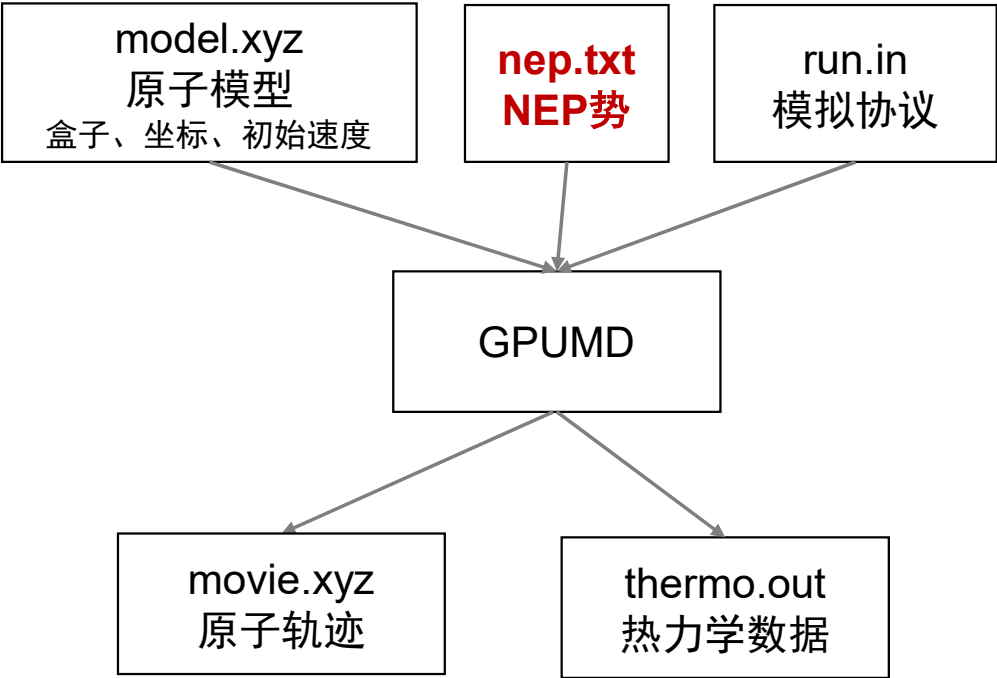
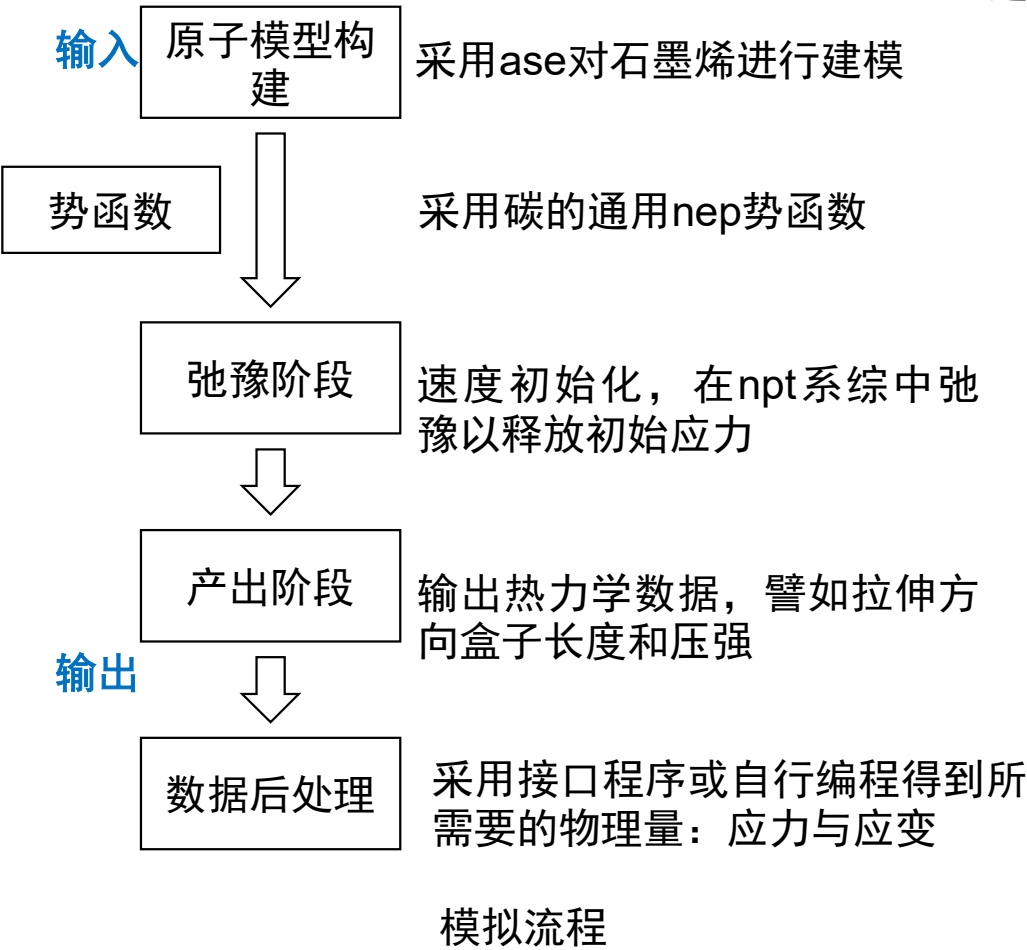
模拟的输入控制文件(run.in)



应力应变曲线

回顾一下，gpumd只是一个计算器！

这些输入输出文件更详细的解释，让我们一起查阅gpumd.org。



garbage in garbage out!
GPUMD的输入与输出

gpumd的输入与输出

gpumd输入文件：

- run.in

计算流程的输入文件，按照输入顺序读取原子模型和势函数，进行分子动力学模拟并输出需要的热力学数据。

- model.xyz

extended xyz格式的原子模型

- 势函数文件

nep.txt

gpumd输出文件：

- movie.xyz

动力学轨迹文件

- *out

根据不同的命令输出的热力学数据结果

第二届VASPKIT团队“机器学习”专题培训

gpumd的输入与输出

- **model.xyz**格式 (<https://github.com/libAtoms/extxyz>)

第1行：原子数目

第2行：几组keyword=value来定义盒子大小、边界条件。

properties=property_name:data_type:number_of_columns 来规定原子信息数组每一列的信息。

第3+行：与properties对应的列。

```
1 1600
2 Lattice="98.38048586991222 0.0 0.0 0.0 42.599999999999994 0.0 0.0 0.0 3.35" Properties=species:S:1:pos:R:3
  pbc="T T F"
3 C      96.53585176      0.71000000      1.67500000
4 C      97.76560783      1.42000000      1.67500000
5 C      97.76560783      2.84000000      1.67500000
6 C      96.53585176      3.55000000      1.67500000
7 C      96.53585176      4.97000000      1.67500000
8 C      97.76560783      5.68000000      1.67500000
9 C      97.76560783      7.10000000      1.67500000
10 C     96.53585176      7.81000000      1.67500000
```


gpumd的输入与输出

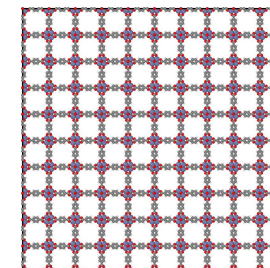
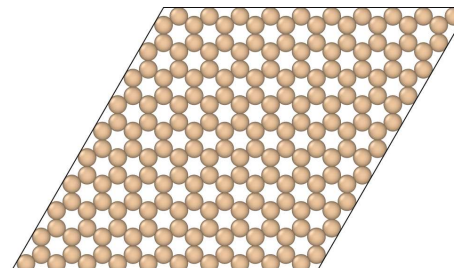
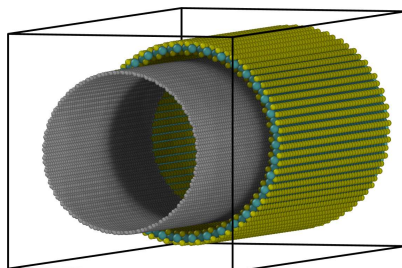
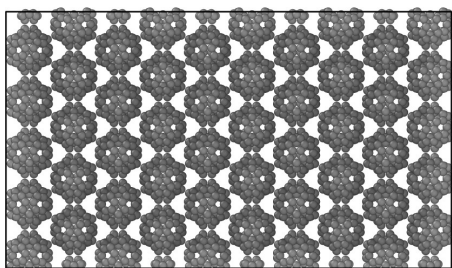
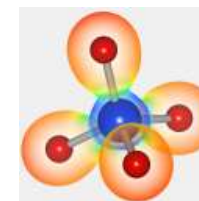
- 如何构建原子模型model.xyz

- 借助VESTA+OVITO从CIF文件转换为model.xyz

(<https://jp-minerals.org/vesta/en/>; <https://www.ovito.org/>)

- 借助ASE实现不同模型格式的转换 (<https://wiki.fysik.dtu.dk/ase/>)

- 编程生成model.xyz文件



实时演示task3: 创建model.xyz的几种方式

第二届VASPKIT团队“机器学习”专题培训

run.in中的输入参数: 设置

- **potential**

语法: *potential nep.txt*

功能: 读取nep势函数

- **time_step**

语法: *time_step 1.0*

功能: 时间步设为1fs

- **velocity**

语法: *velocity 300*

功能: 初始温度设为300K

- **deform**

语法: *deform 1e-4 1 0 0*

功能: 沿着x方向每步拉伸 $1e-4$ Å

- **change_box**

语法: *change_box 1*

功能: x、y、z方向盒子长度同时增加1 Å

语法: *change_box -1 0 1*

功能: x方向减小1 Å, y方向不变, z方向增加1 Å

语法: *change_box -1 0 1 0.01 0.02 -0.03*

功能: x方向减小1 Å, y方向不变, z方向增加1 Å; yz施加1%应变, xz施加2%应变, xy施加-3%应变。使用该命令时, 盒子必须为三斜。

- **fix**

语法: *fix 2*

功能: 固定group_label为2的所有原子

第二届VASPKIT团队“机器学习”专题培训

run.in中的输入参数：系综

○ ensemble (standard)

语法: *ensemble nve*

功能: NVE系综

提示: Berendsen方法适用于迅速获得平衡状态, 无法在动力学采样中获得物理量的正确涨落。

语法: *ensemble nvt_nhc 300 300 100*

功能: NVT系综Nose-Hoover热浴控温300K, 弛豫时间设为100个时间步

将*nvt_nhc*替换为*nvt_ber*、*nvt_bdp*、*nvt_lan*等可分别采用其它热浴方法控温。

语法: *ensemble npt_ber 10 500 100 0 100 1000*

功能: NPT系综Berendsen方法控温从10K升至500K, 三个方向耦合控压为0 GPa, 等效模量设为100 GPa, 控压弛豫时间设为1000个时间步。(适用于正交盒子, 三个方向均为周期性边界条件)

语法: *ensemble npt_ber 10 500 100 0 0 0 100 100 100 1000*

功能: 三个方向独立控压为0 GPa

语法: *ensemble npt_ber 10 500 100 0 0 0 0 0 100 100 100 100 100 100 1000*

功能: 包含剪切自由度六个方向独立控压为0 GPa (适用于三斜盒子)

将以上*npt_ber*替换为*npt_scr*可实现stochastic cell rescaling方法控压。

第二届VASPKIT团队“机器学习”专题培训

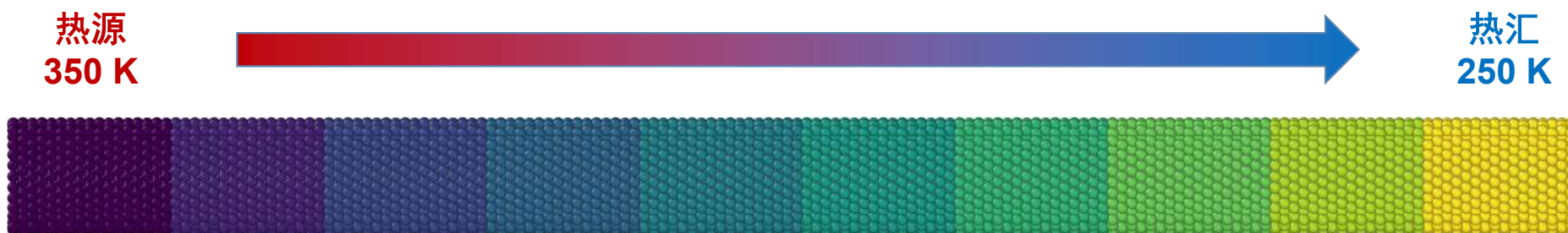
run.in中的输入参数：系综

○ ensemble (NEMD)

语法: `ensemble heat_lan 300 100 50 1 9`

功能:

- 采用朗之万热浴在热源和热汇区域控温以实现非平衡态分子动力学（NEMD）模拟。
- 平均温度 T 为300 K，弛豫时间为100步， ΔT 为50 K，因此热源温度为 $T + \Delta T = 350$ K，热汇温度为 $T - \Delta T = 250$ K。
- 热源施加在1号群组（`group_label=1`），热汇施加在9号群组。
- 将`heat_lan`替换为`heat_nhc`、`heat_bdp`等可分别实现其它热浴方法进行NEMD模拟。由于朗之万热浴拥有随机性和局域性，因此更适合在NEMD中使用 (J. Chem. Phys. 151, 234105 (2019))



实时演示task4: 一个非平衡态模拟计算热导率的小例子

第二届VASPKIT团队“机器学习”专题培训

run.in中的输入参数：系综

○ ensemble (mttk)

语法: `ensemble npt_mttk temp <T_1> <T_2> <direction> <p_1> <p_2> tperiod <tau_temp> pperiod <tau_press>`

功能: NPT系综, 与LAMMPS中的NPT系综等价 (https://docs.lammps.org/fix_nh.html)

语法: `ensemble npt_mttk temp 300 300 iso 10 10`

功能: NPT系综Nose-Hoover热浴控温300K, Parrinello-Rahman压浴各向同性控压10 GPa。

温度和压强弛豫时间默认分别设为100和1000个时间步。

如果将iso替换为aniso, 则盒子的三个方向独立控压。

如果将iso替换为tri, 则允许盒子发生剪切变形 (完全的静水压)。

语法: `ensemble npt_mttk temp 300 300 x 5 5 y 0 0 z 0 0`

功能: NPT系综Nose-Hoover热浴控温300K, Parrinello-Rahman压浴在x方向控压5 GPa, y方向和z方向控压0 GPa。

语法: `ensemble npt_mttk temp 300 300 x 5 5`

功能: NPT系综Nose-Hoover热浴控温300K, Parrinello-Rahman压浴在x方向独立控压为5 GPa, y和z方向固定。

run.in中的输入参数：三种控压方式

- 各向同性控压

```
ensemble npt_ber 300 300 100 0 50 1000
#ensemble npt_scr 300 300 100 0 50 1000
#ensemble npt_mttk temp 300 300 iso 0 0
```

- 三轴独立控压

```
ensemble npt_scr 300 300 100 10 0 0 50 50 50 1000
#ensemble npt_mttk temp 300 300 x 10 10 y 0 0 z 0 0
```

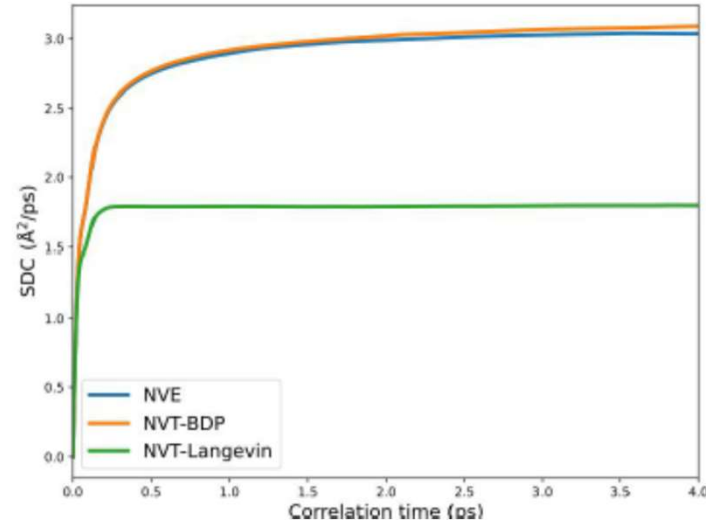
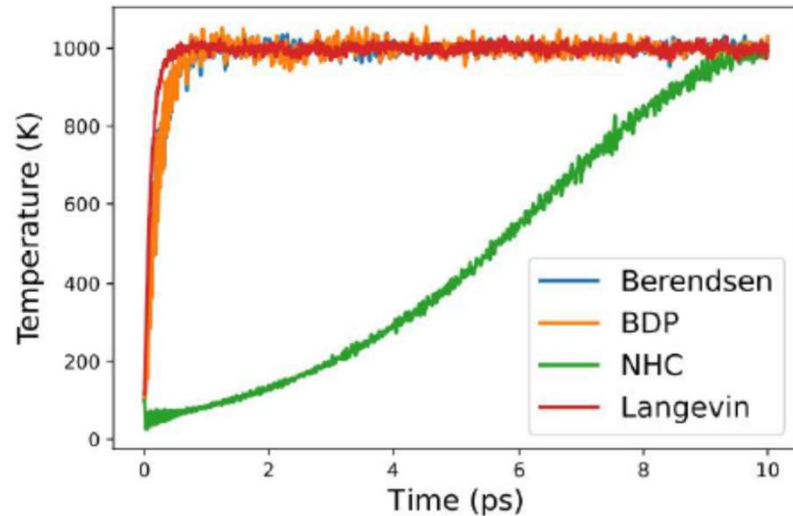
- 六自由度控压

```
ensemble npt_scr 300 300 100 0 0 0 0 0 10 50 50 50 50 50 50 1000
#ensemble npt_mttk temp 300 300 x 0 0 y 0 0 z 0 0 yz 0 0 xz 0 0 xy 10 10
```

第二届VASPKIT团队“机器学习”专题培训

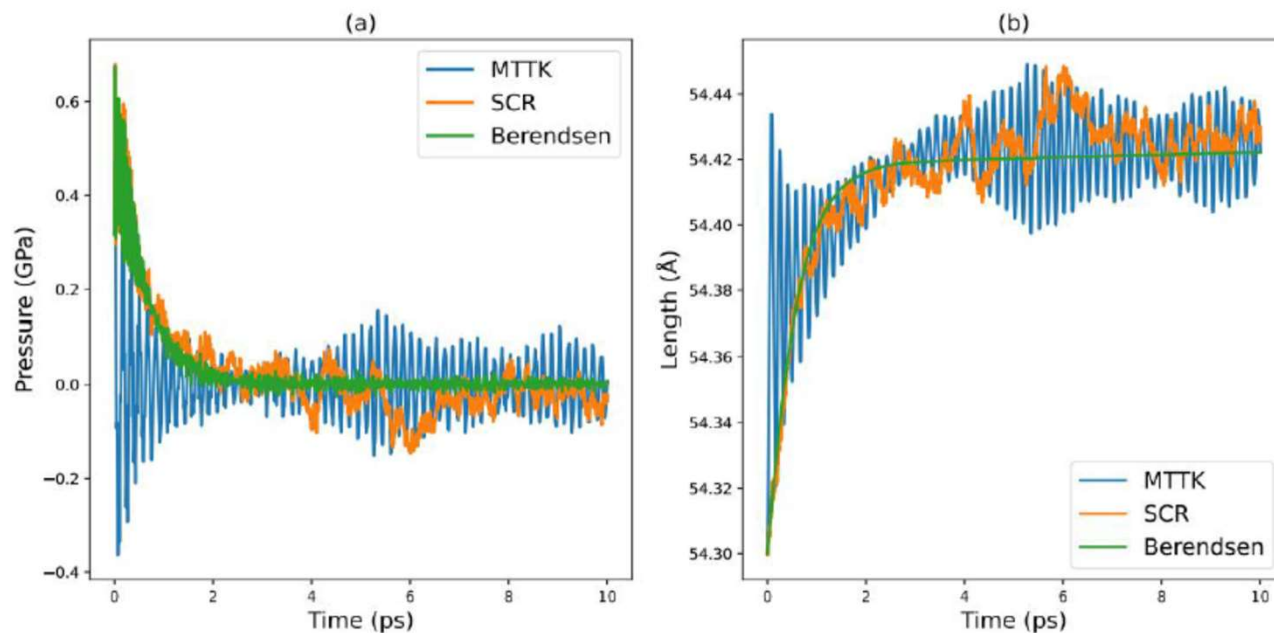
run.in中的输入参数：控温算法对比

控温算法	Berendsen	BDP	NHC	Langevin
是否产生正则系综	否	是	是	是
达到平衡态速度	快	快	慢	快
对动力学的影响	弱	弱	弱	强



第二届VASPKIT团队“机器学习”专题培训

run.in中的输入参数：控压算法对比



提示：Berendsen压浴无法产生物理的应变涨落，因而无法基于该压浴和应变涨落法计算材料的弹性模量。

$$\begin{aligned} S_{\alpha\beta\gamma\kappa} &= \frac{\langle V \rangle}{k_B T} \text{cov}(\epsilon_{\alpha\beta}(t), \epsilon_{\gamma\kappa}(t)) \\ &= \frac{\langle V \rangle}{k_B T} \langle (\epsilon_{\alpha\beta}(t) - \langle \epsilon_{\alpha\beta} \rangle) (\epsilon_{\gamma\kappa}(t) - \langle \epsilon_{\gamma\kappa} \rangle) \rangle \end{aligned}$$

M. Parrinello and A. Rahman, J. Chem. Phys., 76, 2662–2666 (1982). 16 / 30

第二届VASPKIT团队“机器学习”专题培训

run.in中的输入参数：运行

minimize

语法: *minimize sd 1.0e-6 10000*

功能: 采用最速下降法进行能量最小化, 终止标准为系统中原子最大力小于 1e-6 eV/\AA , 或者最大运行步数为10000步。

语法: *minimize fire -1 10000*

功能: 使用FIRE方法进行10000步的能量最小化。

run

语法: *run 1000000*

功能: 运行100万步。

一般来说, 在执行**run**命令前, 需要定义系综以进行分子动力学模拟。

例如:

“**ensemble** *nvt_nhc 300 300 100*

run *100000*”

这样, 就可以在NVT系综中控温300 K进行10万步的分子动力学模拟。

第二届VASPKIT团队“机器学习”专题培训

run.in中的输入参数：计算

compute

语法: `compute 0 10 100 temperature`

功能: 对group method为0的所有原子群计算温度, 每10步抽样一次平均100之后输出一次结果。结果写出至**compute.out**文件中。将temperature替换为potential, force, virial, jp和jk可以计算相应的势能、力、位力、热流的势能项和热流的动能项。在NEMD中计算温度时, 最后两列分别为热源、热汇和体系的能量交换。

compute_cohesive

语法: `compute_cohesive 0.9 1.2 301`

功能: 计算内聚能, 盒子缩放因子从0.9至1.2, 对中间301个构型分别进行能量最小化。结果写出到**cohesive.out**文件中, 第一列为缩放因子, 第二列为对应能量最小化后的势能。

compute_elastic

语法: `compute_elastic 0.01 cubic`

功能: 施加0.01应变对cubic晶体计算弹性常数, 结果写出至**elastic.out**文件中。

第二届VASPKIT团队“机器学习”专题培训

run.in中的输入参数：计算

compute_dos

语法: *compute_dos 5 200 400.0*

功能: 进行声子态密度计算, 每5步记录一次速度, 一共200个关联时间点, 最大角频率为400 THz。结果写出至**mvac.out**和**dos.out**, 前者包含速度自关联函数随关联时间演化的信息, 后者为态密度在角频率区间的分布。

compute_phonon

语法: *compute_phonon 11 0.01*

功能: 调用compute_phonon命令计算声子色散。参数11表示用11埃的截断距离计算力常数, 参数 0.01 表示有限差分法中的原子位移量为0.01 Å。结果写出至**omega2.out**和**D.out**文件, 分别包含色散信息和动力学矩阵。

compute_sdc

语法: *compute_sdc 1 1000*

功能: 计算自扩散系数 (self-diffusion coefficient), 每1步保存一次速度数据, 最大关联步为1000。结果写出至**sdc.out**文件, 包括速度自关联函数及自扩散系数随关联时间演化的信息。

第二届VASPKIT团队“机器学习”专题培训

run.in中的输入参数：计算-热导率相关

compute_hac

语法: *compute_hac 10 100000 10*

功能：在平衡态分子动力学(EMD)模拟中计算热流自关联函数(HAC)，每10步记录一次热流数据，最大关联步为1e5，输出的HAC每10个数据进行一次平均。结果写出至**hac.out**文件，包含HAC和跑动热导率的结果。

compute_hnemd

语法: *compute_hnemd 1000 0.00001 0 0*

功能：在齐性非平衡态分子动力学（HNEMD）模拟中计算热导率，每1000步的数据进行一次平均并输出，在x方向施加驱动力 $1\text{e-}5 \text{ \AA}^{-1}$ 。结果写出至**kappa.out**文件，包含面外面内跑动热导率的结果。

compute_shc

语法: *compute_shc 2 250 0 1000 400.0*

功能：在NEMD和HNEMD模拟中计算谱热流（热导率在声子频率空间的分布），每2步采集一次数据，最大关联步为250步，考虑x方向的谱分解，一共输出1000个频率点数据，最大角频率400 THz。结果写出至**shc.out**文件，包含非平衡态位力-速度关联函数和谱热流的结果。

第二届VASPKIT团队“机器学习”专题培训

run.in中的输入参数：输出

dump_position

语法：*dump_position 1000*

功能：每1000步输出一次轨迹文件。结果写出至**movie.xyz**，包含盒子信息及原子坐标。

dump_exyz

语法：*dump_exyz 1000 1 1*

功能：在dump_position基础上进一步输出原子**速度**和**力**的信息。结果写出至**dump.xyz**。

dump_thermo

语法：*dump_thermo 1000*

功能：每隔1000步输出一次系统的热力学信息。结果写出至**thermo.out**。

run.in文件的工作流

计算声子态密度

01.	potential	Graphene_Lindsay_2010_modified.txt	基本设置
02.			
03.	time_step	1	
04.			
05.	velocity	300	平衡阶段
06.	ensemble	npt_ber 300 300 100 0 0 0 100 100 100 1000	
07.	run	100000	
08.			
09.	ensemble	nve	产出阶段
10.	compute_dos	5 200 400.0	
11.	run	100000	

- 在平衡及产出阶段的区块一般包含以下三个部分：
- 系综设定 ensemble
- 计算或输出需要的物理量 compute* or dump*
- 运行分子动力学模拟 run

第二届VASPKIT团队“机器学习”专题培训

总结

- 通过时间演化，我们会得到模拟体系在相空间的轨迹。测量的目标就是从该轨迹中获取我们感兴趣的结果。
- 我们可以将轨迹输出后再测量，一般称为后处理。但是，GPUMD 的风格是尽量在轨迹产生的过程中就同时进行测量（所谓 on-the-fly 的处理方式），一般不需要输出轨迹进行后处理。
- 为此，在 GPUMD 中有很多计算“命令”，它们都会指示 GPUMD 一边演化轨迹，一边计算物理量，比如扩散系数、态密度、热导率等。当跑的步数完成，测量也就基本完成了。
- 当然，因为目前 GPUMD 中的“命令”个数有限，有些物理量的计算无法在程序内部完成，此时就只能输出轨迹进行后处理了。
- GPUMD支持的命令在不断丰富。如果你需要用到更多的命令，在下一讲中我们会介绍NEP与LAMMPS的接口，这样可以使用NEP势在LAMMPS进行模拟。