README

Yang Huang*

<2023-02-22 Wed>

Contents

*yangh2@andrew.cmu.edu

1 Purpose

The purpose of this document is to explain how to calculate 3-body distribution function from a XDATCAR file (VASP output). You are free to modify and redistribute the code for general purpose.

The basic idea of this program is to divide a BIG XDATCAR file into a reasonable amount (2 or 3 times of cpus on a cluster) of samller XDATCAR files and calculate 3-body distribution functions for each individual job. The final distribution function is obtained by averaging over distribution functions from these jobs.

This program works best for small systems with long MD simulations and worst for very large systems. Meanwhile it requires a great amount of memory to allocate large arrays (depending on number of bins in your 3-body distribution function histogram) for each job.

2 List of tools

Tools that are implicitly required by others are labeled with [IMPLICIT].

2.1 init_bins.awk [IMPLICIT]

Generate input files. Required by init_bin.sh.

2.2 init_bins.sh [IMPLICIT]

Initialize basic settings for slurm_3bdfxdat runs. Required by slurm_3bdfxdat.

2.3 split_xdat.awk [IMPLICIT]

Split XDATCAR file into individual split files. Required by slurm_3bdfxdat.

2.4 3bdfxdat_bundle.sh [IMPLICIT]

Bundle a group of split files and calculate their 3-body distribution function. Required by slurm_3bdfxdat.

2.5 slurm_3bdfxdat

Submit a 3-body distribution function calculation job.

2.6 3bdfxdat_sum

Summarize all statistical results from a slurm_3bdfxdat sjob.

2.7 3bdf2s3

Calculate 3-body entropy based on results from 3bdfxdat_sum and generate supplementary files.

2.8 3bdf2s3_eps

Calculate 3-body entropy based on results from 3bdfxdat_sum and generate supplementary files. Leave out all the ambiguity about numerical integrals and interpolations caused by the shape of histograms. It is expected to give an upper boundary of S3.

2.9 Post-analysis tools

1. 3bdf_view_theta Print out angular distribution of bond R1 and bond R2. Usage: ./3bdf_view_theta fname R1 R2 output theta[0,PI] R3[A] f

2. 3bdfMsup [OBSOLETE] Print out g3-g2g2g2

3. 3bdf_print_sup [OBSOLETE] Print out g2g2g2

4. 3bdf2s3_interp.py Interpolate S3info and S3fluct and write the tot S3. output four columns to the STDOUT which are R[A], S3tot, S3Fluct, S3info.

3 How to

3.1 How to install

To compile, type "make cleanall" and "make" in the main directory. Then type "./install.sh" to link binaries to your "~/bin". All tools except slurm_3bdfxdat are now in your default bin directory.

3.2 How to run

Copy "slurm_3bdfxdat" to your run directory and modify the slurm script at your needs. Prepare all input files and submit the slurm script e.g. "sbatch -p n12 -n 12 slurm_3bdfxdat".

3.3 How to analyze

The slurm_3bdfxdat will provide separate "3bdf_bins.dat" files for a certain number samples. Use "3bdfxdat_sum" to gather all histograms from each "3bdf_bins.dat" file and write the overall histogram to an output file "3bdf_bins.dat-sum". Use "3bdf2s3" to normalize the histogram from " 3bdf_bins.dat-sum" which gives a 3-body distribution function and compute S3 from 3bdf. Use "3bdf_view_theta" and "xmgrace" to visualize your calculation.

3.4 Example: liquid Boron

Go to examples/data and submit slurm_3bdfxdat

4 Usage in details

4.1 slurm_3bdfxdat

4.1.1 Required files

- Trun This file contains in value which is the temperature of MD simulation.
- Mass This file contains three rows:
  1. type of species.
  2. atomic numbers.
  3. atomic masses [a.u.] in atomic units.
- POSCAR The POSCAR should be consistent with the XDATCAR file but not necessary to be one of the structures in the XDATCAR file.
- pdf.HH Pair distribution functions. You can generate those using "pdfxdat" command.
- XDATCAR This has to be a XDATCAR file from AIMD NVT simulation. Be careful about the different in format of XDATCAR from a NVT simulation and a NPT simulation.

4.1.2 Modify the slurm script at your needs

- Cut-off radius We use R1, R2 and THETA to control the shape of our histogram. By default, R1, R2 have the same range [rmin, rmax] and the range of theta depends on R1 and R2. By changing "rmin" and "rmax" in the slurm template, you can balance your calculation efficiency and accuracy. It is recommended to have "rmax" no longer than a quarter of the size of your box and use "rmin" to remove empty regime.

- Number of bins The number of bins for R1 and R2 are specified by dR where NR=CEIL[(rmax-rmin)/dR]. For THETA, it is more complex. See the [add a link to the ref] pdf for details.

- Distribute jobs In seeking of efficiency, you have to tell the script the number of individual 3bdfxdat runs depending on the number of cpus on your cluster. To do that, you can change "nsample", "njobs" and "nxdat" in the script which follow "nsamples=njobs*nxdat".

  - "nsampels": the number samples you want to gather from a XDATCAR file to calculate 3-body distribute function.
  - "njobs" : "nsampels" structures will be divided evenly into $njobs subroutines.
  - "nxdat" : each subroutines have to deal with "nxdat" single-structure XDATCAR files.

4.1.3 Submit jobs

Now you can submit the slurm_3bdfxdat. On euler, it is "sbatch -p n12 -n slurm_3bdfxdat".

4.1.4 Output

It will return a series of 3bdf_bins.dat_??? files where "???" is the index of first structures in the XDATCAR file. Each 3bdf_bins.dat_??? contains a statistical histogram after analyzing "nxdat" amount of files. In 3bdf_bins.dat_???, the first line is the number of analyzed xdatcar files "nxdat", the second line is the number of species, the third line is "rmin", "rmax" and "dR". The remaining of the file contains NRxNR lines and each line is an angular distribution histogram of bond R1 and bond R2 at their length intervals.

4.2 3bdfxdat_sum

To summarize these 3bdf_bins.dat_??? files, try "3bdfxdat_sum fout 3bdf_bins.dat_*". It will gather information from all 3bdf_bins.dat_??? files and write to the "fout". These 3bdf_bins.dat_??? files can have different number of samples. For example, the "3bdf_bins.dat_1" may have analyzed 100 samples while the "3bdf_bins.dat_2" only contains 20 samples. In this way, it is convenient to add additional information to existing results.

4.3 3bdf2s3

The statistical work is quite standard and should have less ambiguity in definitions. Yet the normalization of the 3-body histogram and 3-body entropy calculation are more subtle and involves detailed numerical realization and terminology definitions. It is recommended to check the code and test with a few examples before you make any conclusions. See the [add a link to the ref] pdf file for more description of implementation of normalization and integrals.

To run "3bdf2s3", try "3bdf2s3 fin dim", for instance, "3bdf2s3 3bdf_bins.dat-sum 3". This will normalize the histogram in "3bdf_bins.dat-sum" and compute the "S3Fluct" and "S3Info" based on given "dim". Because pair distribution function and 3-body distribution function are obtained with different programs and hence with different shapes of histograms (different rmin, rmax and dr), we use "dim" to navigate the integration based on a dimxdimxdim mesh grid where functions at each grid are interpolated from pair correlation functions.

The program generates three files: S3B.dat, S3B-ext.dat S3Fluct.dat and S3Info.dat. Each contains two columns: the Rc and fval. The mathematical expression are shown in the [add a link to the ref].

S3B-ext.dat only depends on two-body correlation function while S3B.dat is interpolated to the shape of 3bdf bins. In principle, these two agree when sizes of 3bdf bins equal to sizes of pdf bins. In general, S3B-ext is more accurate because pdf usually has finer bin settings.

4.4 3bdf2s3_eps [OBSOLETE]

This command tends to minimize the differences between superposition g2g2g2 and 3-body distribution histogram g3. In realization, the command defines a difference function (Eps=g3-g2g2g2), and for each binned g3, it will find a possible range [supp_min, supp_max] of g2g2g2 within the bin's volume. Three conditions may occur.

1. g3<supp_min: Eps=g3-supp_min;
2. g3>supp_max: Eps=g3-supp_max;
3. supp_min<g3<supp_max: Eps=0;

In the way, it estimates the upper bound of S3 and reduces the errors from numerical solutions and simulation noises.

To run "3bdf2s3_eps", try "3bdf2s3 fin", for instance, "3bdf2s3 3bdf_bins.dat-sum". This will normalize the histogram in "3bdf_bins.dat-sum" and compute the "S3Fluct" and "S3Info".

The program generates three files: S3B.dat, S3Fluct.dat and S3Info.dat. Each contains two columns: the Rc and fval.