

# RMD\_Digging Manual

## RMD\_Digging: A Toolkit for Information Mining and Analysis from ReaxFF Simulations

Qiang Liu

forliubinqiang@163.com

Institute of Nuclear Physics and Chemistry, Mianyang, Sichuan, China

### 1 How to cite RMD\_Digging Toolkit

✧ To cite:

Qiang Liu, RMD\_Digging ToolKit, version, [https://github.com/dadaoqiuzhi/RMD\\_Digging](https://github.com/dadaoqiuzhi/RMD_Digging),  
access date.

✧ Other literatures:

(1) Liu, Q.; Liu, S.; Lv, Y.; Hu, P.; Huang, Y.; Kong, M.; Li, G. Atomic-scale insight into the pyrolysis of polycarbonate by ReaxFF-based reactive molecular dynamics simulation. *Fuel* 2021, 287, 119484, DOI: <https://doi.org/10.1016/j.fuel.2020.119484>.

(2) Liu, Q.; Huang, W.; Liu, B.; Wang, P.-C.; Chen, H.-B. Gamma Radiation Chemistry of Polydimethylsiloxane Foam in Radiation-Thermal Environments: Experiments and Simulations. *ACS Appl. Mat. Interfaces* 2021, 13 (34), 41287-41302, DOI: <https://doi.org/10.1021/acsami.1c10765>.

(3) C. Li, Q. Liu, W. Gong, Z. Zhou, Z. Yao, X. Meng, Study on the atomic scale of thermal and thermo-oxidative degradation of polylactic acid via reactive molecular dynamics simulation, *Thermochim. Acta* 709 (2022) 179144.

(4) Liu, Q.; Huang, W.; Chen, H. Paving the Way to Simulate and Understand the Radiochemical Damage of Porous Polymer Foam. *ACS Materials Letters* 2023, 2174-2188.

## 2 Development history of RMD\_Digging

Prof. Adri C.T. van Duin of Pennsylvania State University and Prof. William A. Goddard, III develop the reactive force field simulation (ReaxFF) to model the reactive events and dynamic behavior of various systems like polymer materials, fuel, energetic materials, catalytic systems, metal materials, etc. ReaxFF simulation is mainly supported by the open-source Large-scale Atomic/Molecular Massively Parallel Simulator (LAMMPS), whose primary developers are from Sandia National Labs and Temple University. However, the pre- and post-processing as well as corresponding analysis are very hard for many beginners due to the lack of specialized knowledge. When I studied at Sichuan University for my doctorate during 2017-2020, I ever investigated the thermolysis of polymer materials by ReaxFF simulations. I also take advantage of this method to study the radiation effects of polymers now in the Institute of Nuclear Physics and Chemistry. I wrote some codes to process the raw data generated during simulations ever since 2019 based on MATLAB language. With the development of the codes, I gradually realize that I should devote myself to developing a toolkit, which is called RMD\_Digging. RMD is the abbreviation of Reactive Molecular Dynamics. Digging denotes the main functions of the toolkit, which is to dig out information from a wealth of data for further analysis. The first version of RMD\_Digging (V1.0) was finished in 2020. RMD\_Digging can be accessed on the code hoster of GitHub. Since then, I have also kept improving the functions of the toolkit by adding new functions, optimizing codes and debugging with the feedback from users, and this toolkit is updated to the second version (V2.0) in 2022 and V3.0 in 2023. Now the toolkit is renamed RMDigging when I decided to apply for its copyright in China. This toolkit (RMD\_Digging on the GitHub) is also an open-source utility code for researchers, and it can be downloaded and distributed free of charge and rewritten to meet a special need. I will also stick to its development in the future. I hope this toolkit can prosper in this field and contribute to one's study.

### **3 Requirements for MATLAB software version**

As the RMD\_Digging is developed by the R2020a version, the user is suggested to utilize MATLAB released not older than R2020a version. This will help to avoid much incompatible problem.

## 4 Main functions of RMD\_Digging

The main functions of RMD\_Digging can be divided into six modules (Figure 1). We will give a brief introduction and some examples.

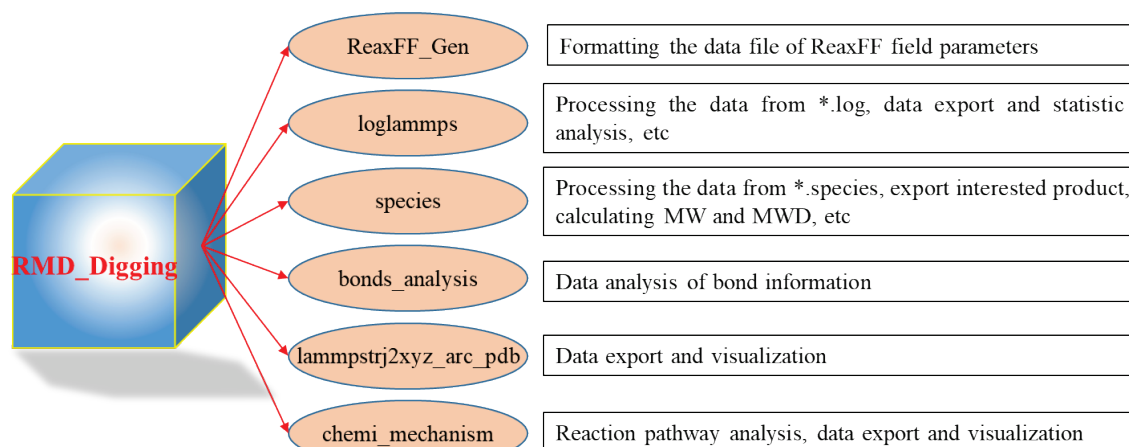


Figure 1 Six modules of the RMD\_Digging toolkit.

### 4.1 ReaxFF\_Gen module

This module is aimed to format the data file of ReaxFF parameters from the literature report. For picture files or PDF without editable permission, an online OCR tool is recommended to extract information first. Then the ReaxFF\_Gen module can be taken to process the files. The detailed instruction is as follows (Figure 2):

1. Make a copy of the data file containing the ReaxFF parameters, delete the title lines of each section in the duplicate and leave the numeric data with comments alone, each section should be separated by a blank line (eg. *ReaxFF\_Gen/input-example.txt*) .
2. Copy the processed content in the duplicate into the *input.txt* file in the ReaxFF\_Gen folder.
3. Start MATLAB and change the working directory to the ReaxFF\_Gen folder, and input ReaxFF\_Gen to invoke the corresponding program.
4. Input the chemical elements involved in the *input.txt* in the cell array. Note: the sort order of these chemical elements should be in line with the original

file. All will be done within several seconds at most.

5. After the normal termination of the program, the formatted parameters will be written in the *output.txt* file. For the case if the names of some chemical elements are two characters, one can manually refine the alignment of these lines.
6. Open the *output.txt* file and copy the removed title lines in Step 1 into the corresponding positions. Now one can copy the final file and rename it as they wish.

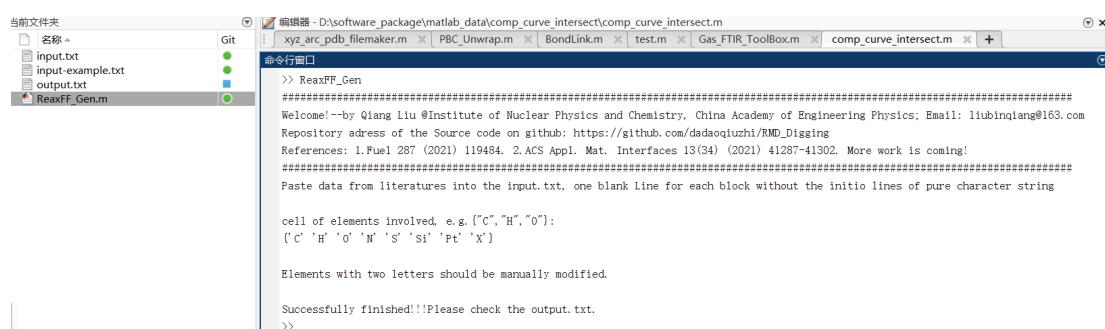


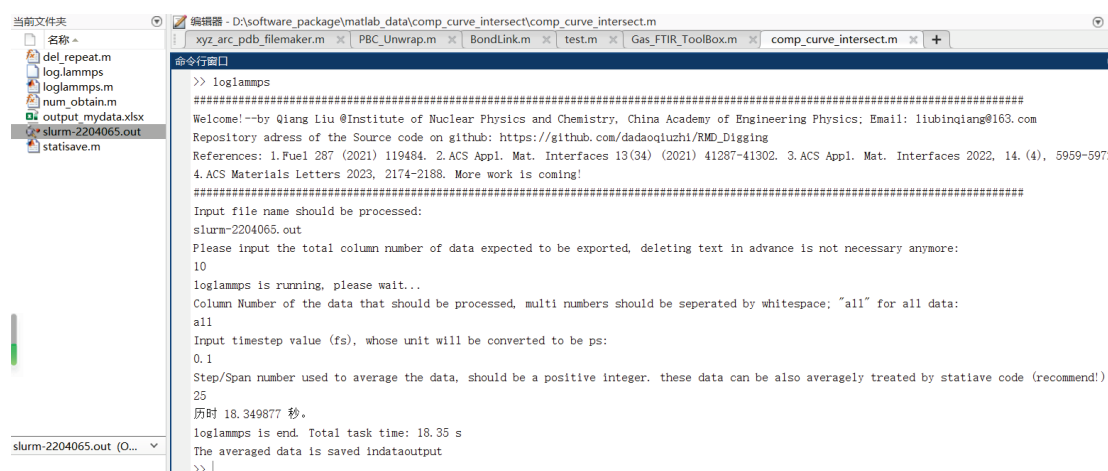
Figure 2 Illustration of running ReaxFF\_Gen module.

## 4.2 loglammps module

This functional module can process the data in the \*.log file outputted to the terminal by the thermos\_style command, which can screen out the interesting data and carry out statistical analysis. The running example is shown below (Figure 3):

1. Start MATLAB and change the working directory to the loglammps folder. Make a copy of the log.\* file, it is not necessary to remove all the non-numeric text for V3.0. It is also not necessary to copy the number block with the same size (same column) into the log.lammps file, and the user can control this by limit the column number and export data several times.
2. Input loglammps to invoke the corresponding program.
3. According to the prompt, you should input the filename (log.lammps) to be processed.

4. Input the total column number of the data expected to be exported, which limit the data object. Then input the column number of the interesting data, which should start from 1 and multi numbers should be separated by a space. “all” can be used to export all data.
5. Input the timestep of the simulation (default unit: fs) and the Step/Span number used to average the data (Note: the raw data are usually outputted every hundred steps).
6. The processed data is saved in the *dataoutput* matrix while the raw data is saved in the *datainput* matrix in the workspace. One can export these data in *dataoutput* matrix to the *output\_mydata.xlsx* file by uncommenting several lines.
7. One can further perform statistical analysis using *statisave* code and just input option one by one according to the prompt. The final processed data are assigned to the *datastatis* matrix.
8. Go to step 1 to process other data blocks if necessary.



```

>> loglammps
#####
Welcome!--by Qiang Liu @Institute of Nuclear Physics and Chemistry, China Academy of Engineering Physics; Email: liubinqiang@163.com
Repository adress of the Source code on github: https://github.com/dadaoqiuzhi/RMD_Digging
References: 1.Fuel 287 (2021) 119484. 2.ACS Appl. Mat. Interfaces 13(34) (2021) 41287-41302. 3.ACS Appl. Mat. Interfaces 2022, 14. (4), 5959-5972.
4.ACS Materials Letters 2023, 2174-2188. More work is coming!
#####
Input file name should be processed:
slurm-2204065.out
Please input the total column number of data expected to be exported, deleting text in advance is not necessary anymore:
10
loglammps is running, please wait...
Column Number of the data that should be processed, multi numbers should be seperated by whitespace; "all" for all data:
all
Input timestep value (fs), whose unit will be converted to be ps:
0.1
Step/Span number used to average the data, should be a positive integer. these data can be also averagely treated by statlave code (recommend!):
25
历时 18.349877 秒。
loglammps is end. Total task time: 18.35 s
The averaged data is saved indataoutput
>>

```

Figure 3 Illustration of running loglammps module.

## 4.3 species module

This module is aimed to process and generate data related to the *species.\** file,

For example, to obtain the evolution of specified products or reactants, to select and export special species and to calculate the molecular weight or molecular weight distribution of the limited species.

### 4.3.1 species\_analysis

The species\_analysis code will sort all the species according to the timestep. Note: this code cannot distinguish the isomers. Some instances are displayed here (Figure 4).

1. Make a copy of the *species.\** file (here is species-PDMS.out) to the species folder.
2. Start MATLAB and change the working directory to the species folder. Input species\_analysis to invoke this code.
3. Input the name of the *species.\** file according to the requirement.
4. The code will take some time to finish this task. The time cost depends on the size of the *species.\** file. A pop-up window will appear when the task is all normally finished.
5. The final output of the data can be found in the cell array named outputdata.

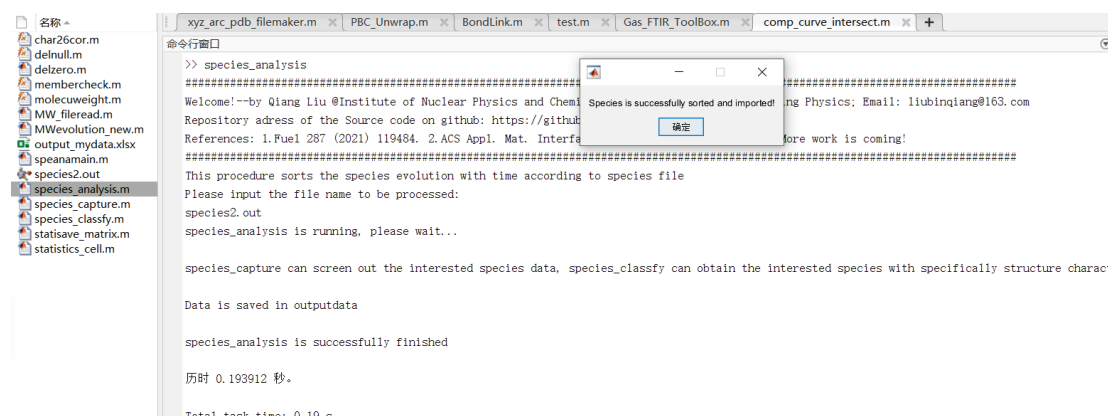


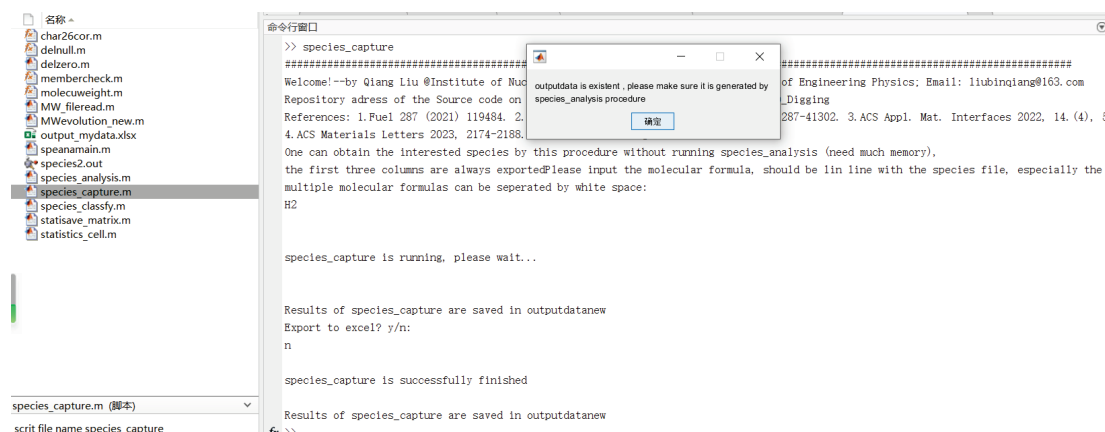
Figure 4 Diagram of the running of species\_analysis code in the species module.

### 4.3.2 species\_capture

The subroutine called species\_capture can find out the interesting species from

the generated cell array called `outputdata` according to the molecular formula given by the user (Figure 5).

1. Make sure the working directory is located in the `species` folder. Input `species_capture` to call this code.
2. Input the molecular formula expected to be exported for further analysis. For multiple molecular formulas, one can separate them by white space. Note: the molecular formula should be the same as that in the `species.*` file, and the order of these elements can also be confirmed by the `in.*` file.
3. Generally, several seconds is enough to complete this job and the processed data is saved in the `outputdata` new matrix.
4. The code will ask one if to export these data to `output_mydata.xlsx`. I suggest not to do so, one can just copy them from the `outputdata` new matrix.
5. From V3.0 on, running `species_analysis` program in advance is not necessary, which will import all products (need much time and memory for large file). Instead, run `species_capture` directly to import the interested products frame by frame.





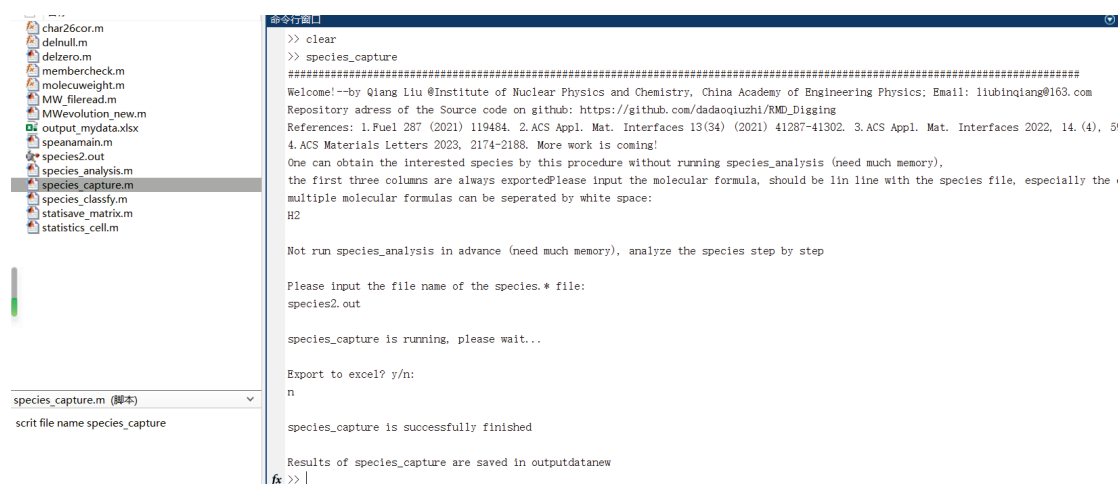


Figure 5 Example of running species species\_capture code in the species module.

### 4.3.3 species\_classfy

The species\_classfy code is used to retrieve species data from the outputdata cell array according to species limitations determined by some criteria (like restrictive conditions on molecular weight, elements and their content). One can repeatedly take this strategy to specify their target species (Figure 6):

1. Make sure the working directory is located in the species folder. Input species\_classfy to call this code.
2. Four options will be shown immediately in the interactive interface, and some examples are provided to explain their meaning.
3. Choose “a”, “b”, “c” or “d” according to your requirements. Meanwhile, the code needs to know if to sum up the filtered data. Then wait for the end of the subroutine.
4. The legal data will be extracted to the dataexport cell in the working space and the summing data is saved in sumdata.

```

>> species_classify

#####
Welcome!--by Qiang Liu @Institute of Nuclear Physics and Chemistry, China Academy of Engineering Physics: Email: liubinqiang@163.com
Repository address of the Source code on github: https://github.com/dadaoqiuzhi/RMD_Digging
References: 1.Fuel 287 (2021) 119484. 2.ACS Appl. Mat. Interfaces 13(34) (2021) 41287-41302. More work is coming!
#####

When species_analysis is executed, this procedure can obtain the interested species
(1)C20 means species with 20 C,C42+ denotes species with C number larger than 42, C100- is species with C less than 100
(2)M100 indicates species with Mw of 100, M125+ denotes Mw larger than 125, M400- is species less than 400
(3)eleC are species have C, eleCO are species have C and O
(4)eleonlyCH are species only have C and H, eleonlyCO are species only have C and O

Methods to filter out the interested species:
a:C1,C20,C42+,C100-,+ means >=,- means <
b:M100,M125+,M400-
c:eleC,eleCO
d:eleonlyC,eleonlyCO

Please select the option (a, b, c or d):
a
Sum the data? y/n:
n

if the exported data only have three column, not hit the interested species, please delet the irrelevant data in the work space
Please input the specific requirements according to the selected methods a or b, e.g. C100-, M100:
C5-

species_classify is running, please wait...
Results of nspecies_classify are saved in dataexport

species_classify is successfully finished
Sum of data are saved in sumdata
Results of species_classify are saved in dataexport, sum of data are saved in sumdata

More complex data abstraction can be performed by copying data in dataexport to outputdata and go on!
fx >>

```

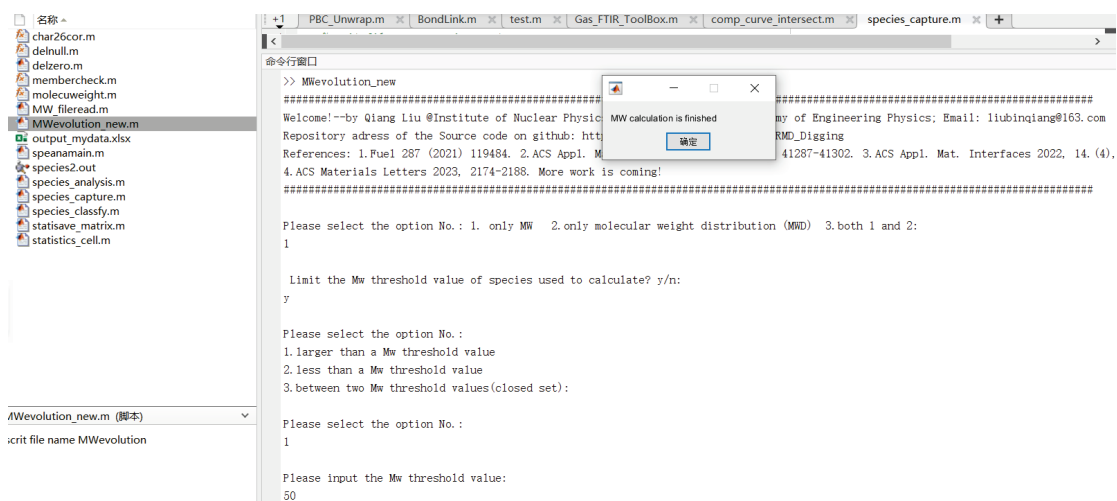
Figure 6 Instance of running species species\_classify code in the species module.

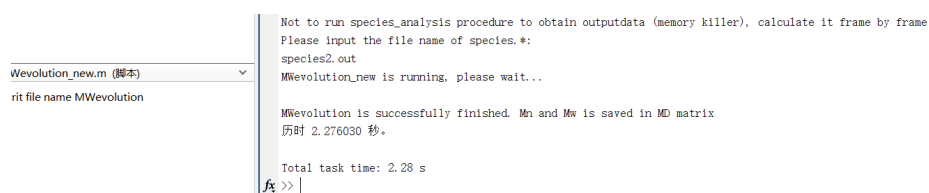
### 4.3.4 MWevolution\_new

The MWevolution\_new code is used to calculate the number-average and weight-average molecular weight as well as molecular weight distribution of the specified species (Figure 7 ). From V3.0 on, running species\_analysis program in advance is not necessary, which will import all products (need much time and memory for large file). Instead, run MWevolution\_new directly to import and process the interested products frame by frame:

1. Make sure the working directory is located in the species folder. Input MWevolution\_new to call this code.
2. Make option on calculation according to the requirements and hints. Option 1 means only calculating number-average and weight-average molecular weight. Option 2 indicates only calculating the molecular weight distribution of the specified frame. Option 3 is forced to calculate the molecular weight and molecular weight distribution of the specified frame.

3. If option 1 or 3 was chosen in the last step, the subroutine will ask the user whether to restrain the molecular weight of the species which will be considered to be calculated later. Three items are provided to limit the molecular weight of the interesting species to be larger than and less than one critical number as well as between a certain range. For item 3, a matrix should be given to limit the minimum and maximum molecular weight, eg. [200 50000].
4. If item 2 or 3 is chosen in the last step, the frame/timestep number is required for the molecular weight distribution. There are two ways to specify the frame/timestep number. Option 1 is to specify the frame/timestep number manually. Option 2 is to specify the frame/timestep number via arithmetic sequence. For option 2, the following information is required: the minimum timestep, maximum timestep, output frequency (LAMMPS), and common difference.
5. Then wait for the end of the running code until the corresponding pop-up window is shown. The molecular weight data are saved in the MD matrix while molecular weight distribution data is stored in the MWDdata matrix. The massive zero data can be removed by delzero code.





```
W evolution_new.m (脚本)
rit file name MWevolution

Not to run species_analysis procedure to obtain outputdata (memory killer), calculate it frame by frame
Please input the file name of species.*:
species2.out
MWevolution_new is running. please wait...

MWevolution is successfully finished. Mn and Mw is saved in MD matrix
历时 2.276030 秒。

Total task time: 2.28 s
fx >> |
```

Figure 7 illustration of running MWevolution\_new code in the species module.

## 4.4 bonds\_analysis module

This module can deal with the bond information, which is the key to identify products and perform topology analysis.

### 4.4.1 bonds\_analysis\_speedup

Both bonds\_analysis\_speedup and bonds\_analysis are the main routines to handle the bond information. We recommend bonds\_analysis\_speedup for practical use due to its faster operation (Figure 8).

1. Start MATLAB and change the working directory to the bonds\_analysis folder. Input bonds\_analysis\_speedup to invoke the corresponding program.
2. Input the name of the file to be processed.
3. Input the interesting frame/ timestep number.
4. Input the total atom number of the systems.
5. Wait for the code to run and finish successfully. The data is saved in the bondoutdata matrix.

```

>> bonds_analysis_speedup
#####
Welcome!--by Qiang Liu @Institute of Nuclear Physics and Chemistry, China Academy of Engineering Physics. Email: liuhingqiang@163.com
Repository adress of the Source code on github: https://github.com/dsdaoqiuzhi/RMD_Digging
References: 1.Fuel 287 (2021) 119484. 2.ACS Appl. Mat. Interfaces 13(34) (2021) 41287-41302. More work is coming!
#####
This program will read the BO information of a specified or all (not recommended) trajectories, the text in the 2-4 rows of each trajectory is omitted.
Please input the file name to be processed:
bonds.PDMS.Reax
Please input the output frequency of BO information (Positive integer):
400

Please input the timestep of the specified trajectory:
272400
Please input atom number:
770
bonds_analysis_speedup is running, please wait...

bonds_analysis is successfully finished, BO information is saved in bondoutdata
Export results? y/n?:
n

bonds_analysis is successfully finished. BO information is saved in bondoutdata.
fx >>

```

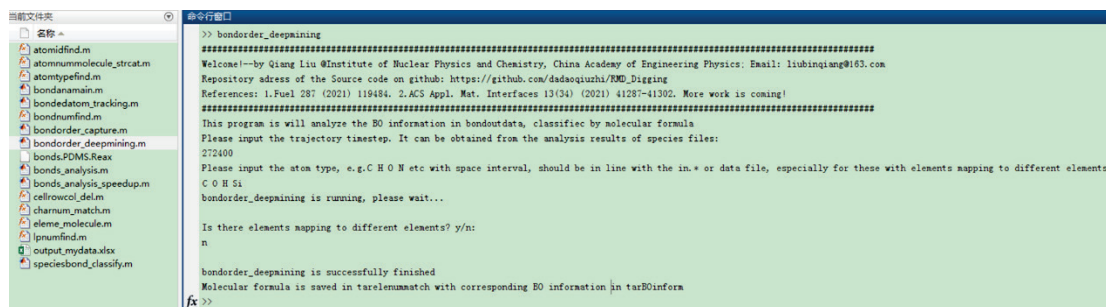
Figure 8 Examples of running bonds\_analysis\_speedup code of the bonds\_analysis module.

## 4.4.2 bondorder\_deepmining

The bondorder\_deepmining code can further treat the bond information in view of molecular formula or species compositions. The as-processed data block is separated by the pound sign (Figure 9).

1. Start MATLAB and change the working directory to the bonds\_analysis folder. Input bonds\_deepmining to invoke the corresponding program. The bonds\_analysis\_speedup code should be executed first.
2. Step by step, input the interesting frame/timestep number, the involved elements whose order should be in line with the original file as mentioned earlier.
3. The console output message that you should verify if you performed element mapping (eg, O is mapped to S). For general simulations, it is no (n).
4. When the subroutine exits, the handled bond data can be found in the tarBOinform cell while the relevant molecular formula or species compositions are stored in the tarelenummatch cell. Data in tarelenummatch are generally identical with the *species.\** file in almost all cases. The exception can be triggered if an abnormality occurs. The possible reason will be shown by the program and some solutions are given, where the most possible cause is the difference in the output frequency of bond data and

species data. One can ignore this warning to continue.



```
>> bondorder_deepmining
#####
Welcome!-by Qiang Liu @Institute of Nuclear Physics and Chemistry, China Academy of Engineering Physics. Email: liubinqiang@63.com
Repository address of the Source code on github: https://github.com/dadaoqunzhi/BMD_Deepmining
References: 1.Fuel 287 (2021) 119484. 2.ACS Appl. Mat. Interfaces 13(34) (2021) 41287-41302. More work is coming!
#####
This program is will analyze the BO information in bondoutdata, classified by molecular formula
Please input the trajectory timestep. It can be obtained from the analysis results of species files:
272400
Please input the atom type, e.g.C H O N etc with space interval, should be in line with the in.* or data file, especially for these with elements mapping to different elements.:
C O H Si
bondorder_deepmining is running, please wait...
Is there elements mapping to different elements? y/n:
n
bondorder_deepmining is successfully finished
Molecular formula is saved in tarBOinform with corresponding BO information in tarBOinform
>>
```

Figure 9 Examples of running bondorder\_deepmining code of the bonds\_analysis module.

## 4.5 lammpstrj2xyz\_arc\_pdb module

The purpose of this module is to export formatted files based on trajectory file (\*.lammpstrj) and bond data file (bonds.\*). The formatted files indicate standard xyz, arc or pdb files. The user can select one format according to their software resource and personal habit. Usually, Materials Studio or VMD are the most intended software. They can give pretty pictures for further analysis and paper publications or academic reports.

### 4.5.1 lammpstrj\_analysis

The function of the lammpstrj\_analysis code is to read and process topological information in the trajectory file (Figure 10).

1. Start MATLAB and change the working directory to the lammpstrj2xyz\_arc\_pdb folder. Make a copy of the \*.lammpstrj file (here is PLA.lammpstrj) to the lammpstrj2xyz\_arc\_pdb folder.
2. Input lammpstrj\_analysis to invoke this code.
3. Input the filename of the trajectory file, the output frequency of the trajectory file, the interesting frame/timestep number of the trajectory, and the total atom number according to the requirements.

4. After the normal exit of the code, one should choose whether to export the results to Excel (*output\_mydata.xlsx*). No is always recommended.
5. The processed data are saved in the trjdata matrix.

```
>> lammpstrj_analysis
#####
Welcome!--by Qiang Liu @Institute of Nuclear Physics and Chemistry, China Academy of Engineering Physics. Email: liubinqiang@163.com
Repository address of the Source code on github: https://github.com/dadaoqiuzhi/RMD_Digging
References: 1.Fuel 287 (2021) 119484. 2.ACS Appl. Mat. Interfaces 13(34) (2021) 41287-41302. More work is coming!
#####
This program is used to read a specified trajectory in *.lammpstrj

Filename name of *.lammpstrj file:
PLA.lammpstrj

Please input the output frequency of B0 information and trajectory file (Positive integer, see bonds or lammpstrj file):
400

Please input the timestep of the specified trajectory:
272400

Please input atom number:
770

lammpstrj_analysis is running, please wait...
\nlammpstrj_analysis is successfully finished.

Atomic coordination information of the specified trajectory is saved in trjdata

Export data to Excel? Much time is required for large data and the Excel should be closed.y/n:
n
>> |
```

Figure 10 Examples of running the lammpstrj\_analysis code of the lammpstrj2xyz\_arc\_pdb module.

#### 4.5.2 xyz\_arc\_pdb\_filemaker

The xyz\_arc\_pdb\_filemaker subroutine is aimed at exporting visual files by combining topological trajectory file and bond data file. Both single static diagrams and dynamic graphs can be generated via this code. The format of the exported file can be xyz, arc or pdb (Figure 11).

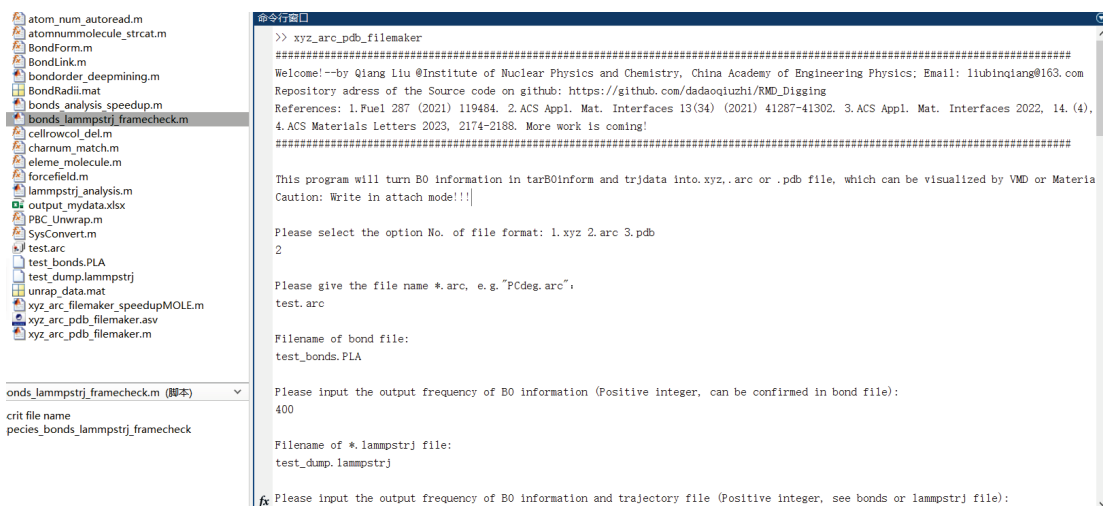
1. Make sure the working directory is situated in the lammpstrj2xyz\_arc\_pdb folder. Make a copy of the \*.lammpstrj file (here is *PLA.lammpstrj*) and *bonds.\** (*bonds.reax.PLA*) file to the lammpstrj2xyz\_arc\_pdb folder.
2. Input xyz\_arc\_pdb\_filemaker to call this code.
3. Select one option for the desired output format, namely xyz, arc or pdb.
4. Name the output file according to your last choice and the file suffix should be explicitly given.

5. Input the file name of *bonds.\** and its output frequency.
6. Input the file name of *\*.lammprtrj* and its output frequency.
7. Input the total atom number. From V3.0 on, the program will automatically read the total atom number.
8. The program asks the user to input the number for frame No. check of *bonds.\** and *\*.lammprtrj* files, so as to avoid mismatch induced error. One can choose no in the next step if you are self-confident about the non-existent of the possible problem.
9. User is asked to give answer about if to run check process for the consistency of frame No. in the *bonds.\** and *lammprtrj.\** (need extra time and memory for large files). If the residual check file in the workspace is found, rerun question will be triggered.
10. Input the involved elements whose order must be in line with the original file as mentioned earlier (check *in.\** or *species.\** files). The console output message that you should verify if you performed element mapping (eg, O is mapped to S). For general simulations, it is no (n).
11. If the mismatch number between *bonds.\** and *lammprtrj.\** files can not be ignored, the program will output the data to screen and let the user to choose corresponding methods to solve this problem. The program will automatically take care of the target frame No. input by user later (give correction).
12. Input the number system (binary, decimal and decimal, etc) for the task, which will be considered to encode the atom id. The selection of the number system is based on the total atom number and the limitation on the length of the name/serial number of any atom (id). A recommended value will be shown and the user should input a value not less than the recommended one.
13. Input the frame/timestep number of the interesting trajectory. There are two ways to specify the frame/timestep number. Option 1 is to specify the frame/timestep number manually. If there are several frames, these numbers should be given simultaneously and separated by white space. Option 2 is to



specify the frame/timestep number via arithmetic sequence. For option 2, the following information is required: the minimum timestep, maximum timestep, output frequency (LAMMPS), and common difference. Option 3 includes methods described in option 1 and 2. And the input sequence should abide by the rules: option 1 first and then option 2.

14. Input answers for a periodic structure. ON for periodic structure, OFF for the contrary.
15. Input answers for scaled coordinate. User should check the *in.\** file to know if the coordinate in the *\*.lammprj* file is scaled. The scaled coordinate will be converted into a real space coordinate. I recommend not to scale the coordinate.
16. The program asks whether to perform unwrap for atoms with coordinate affected by ghost position, causing discontinuity of structure. This can be managed by Materials studio via changing PBC display mode. However, this function here cannot realize this goal now.
17. For periodic structure, the cell parameters (angle, four decimal place) and space group (P1 for general case) should be provided according to the prompt message.
18. Finally, the code will run and export specified files. The detailed information of code invoking is output to show the task progress.



atom\_num\_autoread.m

atomnummolecule\_strcat.m

BondForm.m

BondLink.m

bondorder\_deepmining.m

BondRadii.mat

bonds\_analysis\_speedup.m

**bonds\_lammpstrj\_framecheck.m**

cellrowcol\_del.m

charnum\_match.m

elme\_molecule.m

forcefield.m

lammpstrj\_analysis.m

output\_mydata.xlsx

PBC\_Unwrap.m

SysConvert.m

test.arc

test\_bonds.PLA

test\_dump.lammpstrj

unrap\_data.mat

xyz\_arc\_filemaker\_speedupMOLE.m

xyz\_arc\_pdb\_filemaker.asv

xyz\_arc\_pdb\_filemaker.m

onds\_lammpstrj\_framecheck.m (脚本)

crit file name

pecies\_bonds\_lammpstrj\_framecheck

命令窗口

400  
Automatically read atom number from the file  
The atom number line is found, ready to read  
The atom number is found, which is 770

Please input the number for frame No. check of bonds.\* and \*.lammpstrj files compared with species file.  
avoiding mismatch induced error, >=5 is suggested:  
5  
If to run check process for the consistency of frame No. in the bonds.\* and lammpstrj.\* (need extra time and memory)y/n:  
y  
Check the consistency of frame No. between bonds.\* and lammpstrj.\* files.  
Sometimes energy minimization causes the inconsistent record problem  
species\_bonds\_lammpstrj\_framecheck is running, please wait...  
Firstly, check frame No. in bonds.\* file, please wait...  
The appointed frame No. in bonds.\* file (5 in total) has been read  
Then, check frame No. in lammpstrj.\* file, please wait...  
The appointed frame No. in lammpstrj.\* file (5 in total) has been read  
bonds\_lammpstrj\_framecheck is end  
5 frame No. in the bonds.\* and lammpstrj.\* are as follows:  

0	100000
400	100400
800	100800
1200	101200
1600	101600

  
As above, frame No. in the bonds.\* and lammpstrj.\* files are inconsistent,  
please treat the frame No. of lammpstrj.\* file by the following method:

命令窗口

Please treat the frame No. of lammpstrj.\* file (outputnew) by the following method:  
1. Subtract an integer 2. Add an integer 3. No action  
3  
Please input the unsigned number, eg. 8, 29, or 377, etc.:  
0  
  
Please input atom type like C,H,O,N, separated by white space, corresponding to 1,2,3,4...n (see \*.data or in.\*, especially for element mapping:  
C O H  
  
Does there exist element mapping?y/n:  
n  
  
Different coding system is adopted according to the atom number (ASCII)  
10 base coding system is recommended for atom id  
Please select coding system for atom id, positive integer and not less than the recommended value:  
10  
  
Please select the export method No.:  
1. Manually specify multi-trajectories  
2. Monotonically increasing frame(s) in arithmetic sequence(closed interval)  
3. Both method 1 and 2 (1 first, 2 last)  
1  
  
Trajectories in ascending order, separated by white space:  
100000 300000

命令窗口

Please input periodic boundary condition, ON/OFF:  
ON  
  
Does the coordinate is scaled in the \*.lammpstrj file, y/n:  
n  
If to perform unwrap for atoms with coordinate affected by ghost position, causing discontinuity of structure, y/n? (Unusable Now!):  
n  
Periodic boundary condition, alpha, four decimal digits:  
90.0000  
Periodic boundary condition, beta, four decimal digits:  
90.0000  
Periodic boundary condition, gamma, four decimal digits:  
90.0000  
  
Point group name, eg. "(P1)" for \*.arc and "P 1" for \*.pdb:  
(P1)  
  
xyz\_arc\_pdb\_filemaker is running, please wait...  
  
xyz\_arc\_pdb\_filemaker is searching for Group 1 trajectory: 100000,2 trajectories in total  
  
Step1:Group 1 trajectory 100000 is successfully processed by bonds\_analysis\_speedup and bondnumdata is generated,  
continue running bondorder\_deepmining program, please wait...  
  
Step2:Group 1 trajectory 100000 is successfully processed by bondorder\_deepmining, and tarB0inform is generated,  
continue running lammpstrj\_analysis, please wait...

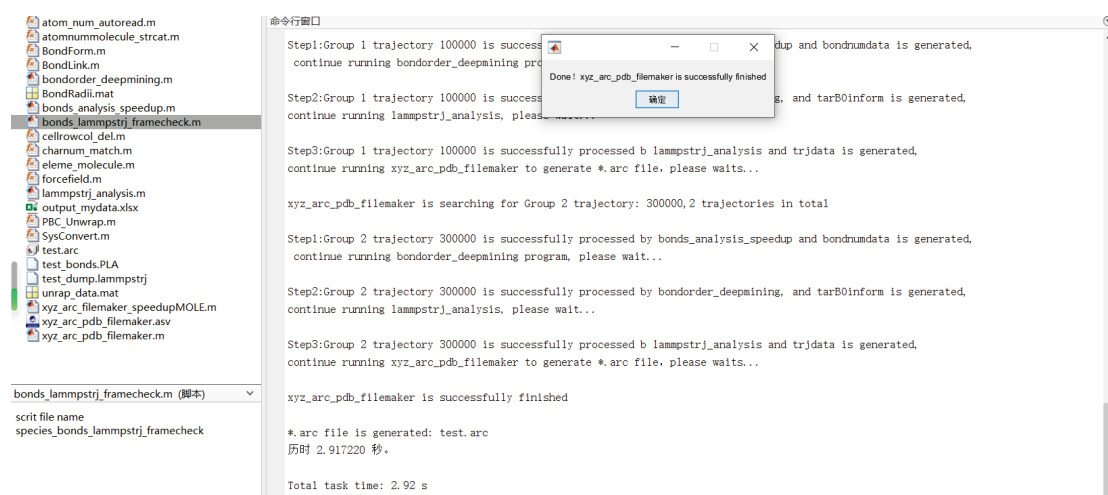


Figure 11 Examples of running xyz\_arc\_pdb\_filemaker code of the lammpstrj2xyz\_arc\_pdb module.

### 4.5.3 xyz\_arc\_filemaker\_speedupMOLE

The function of the xyz\_arc\_filemaker\_speedupMOLE code is the same as the xyz\_arc\_pdb\_filemaker, but it can only generate xyz or arc files. Users can try this subroutine on their own.

## 4.6 chemi\_mechanism

This module is used to study the reaction pathways and generate visualization files. Further analysis can be performed with the help of Materials Studio or VMD. The code will invoke most core subroutines in other modules.

### 4.6.1 chemi\_mechanism

The code named chemi\_mechanism is the core integrated code to analyze the reaction pathways and corresponding mechanisms, which relies on the most main code in other modules. The in-depth analysis also depends on the visualization analysis with the help of bond order data (Figure 12).

1. Start MATLAB and change the working directory to the chemi\_mechanism

folder. Make a copy of the *species.\** (*species-PLA.out*) file, *\*.lammprj* file (here is *PLA.lammprj*) and *bonds.\** (*bonds.reax.PLA*) file to the *chemi\_mechanism* folder.

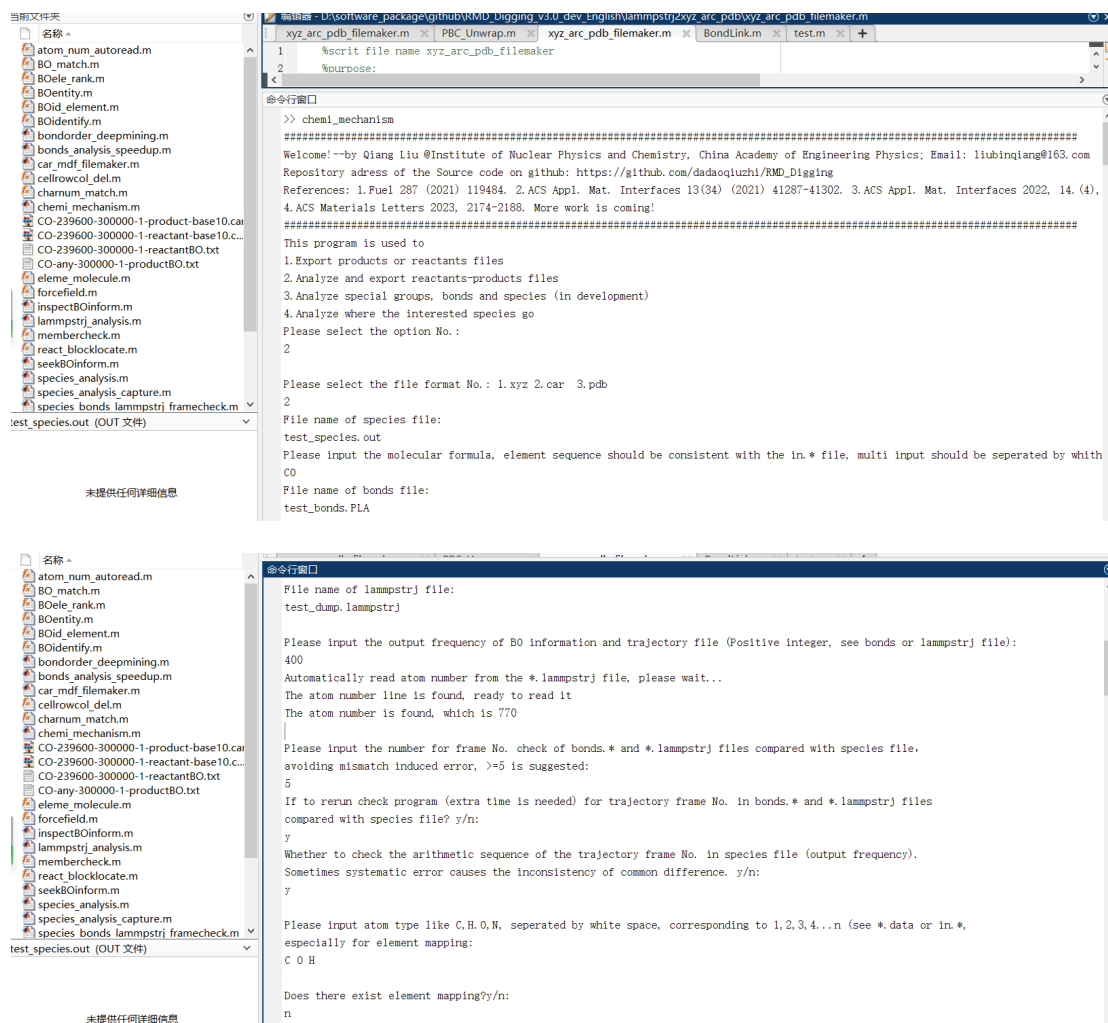
2. Input *chemi\_mechanism* to invoke this code.
3. The subroutine shows the four functions first and waits for feedback. These functions include: (1) Export products or reactants files; (2) Analyze and export reactants-products files; (3) Analyze special groups, bonds and species (in development); (4) Analyze where the interested species go. Users should select one item to continue.
4. Here we continue our tutor by selecting option 2. The main process of other options has little difference. One can finish other analysis task according to this tutor and corresponding explanatory text.
5. Input the file name of the *species.\** (*species-PLA.out*).
6. Input the molecular formula of the interesting product (CO<sub>2</sub>). Note: The molecular formula should be the same as that in the *species.\** file, and the order of these elements can also be confirmed by the *in.\** file.
7. Input the file name of *bonds.\** and *\*.lammprj*.
8. Input the output frequency of bond information and trajectory information or their least common multiple. Now the output frequency is deemed to be the same. In the future, I will deal with the case that the output frequency separately with a view to the actual situation. Besides, I consider to obtain these data automatically. If the user uses two different output frequencies for bond information and trajectory information, one should input the lowest common multiple to gain reasonable results.
9. Input the total atom number of the simulation. From V3.0 on, the program will automatically read the total atom number.
10. The program asks the user to input the number for frame No. check of *bonds.\** and *\*.lammprj* files compared with *species.\** file, so as to avoid mismatch induced error. One can choose no in the next step if you are self-confident about the non-existent of the possible problem. User is asked

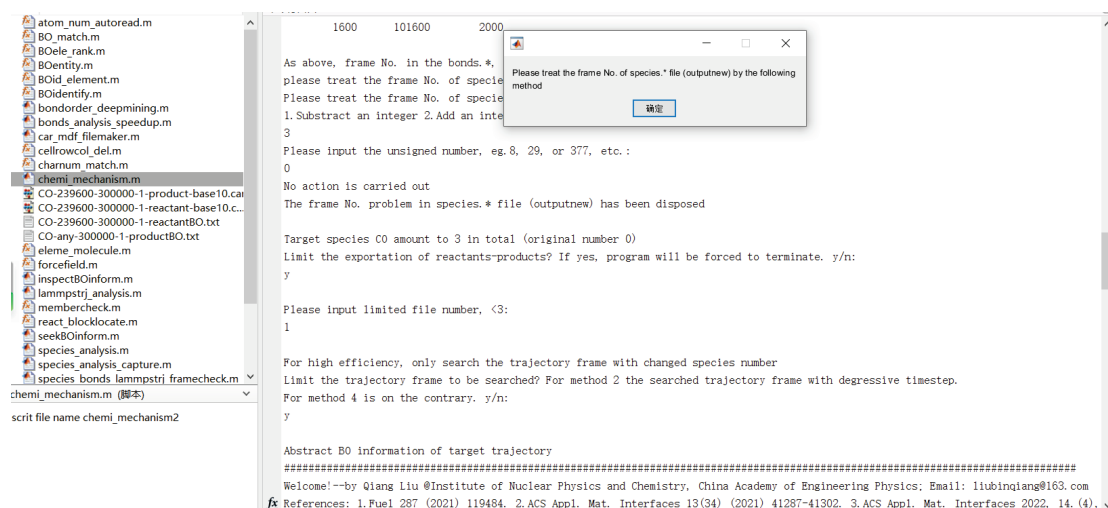
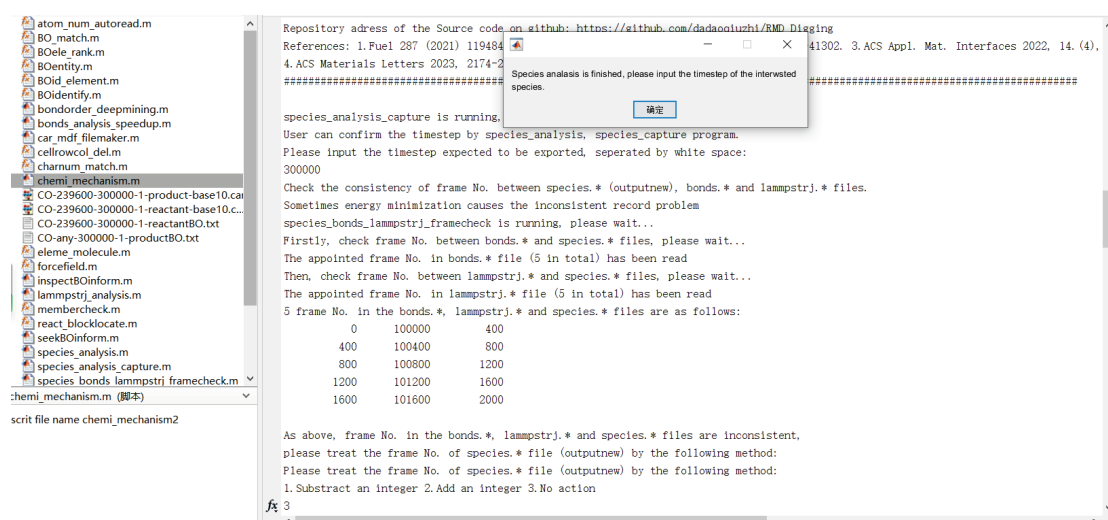
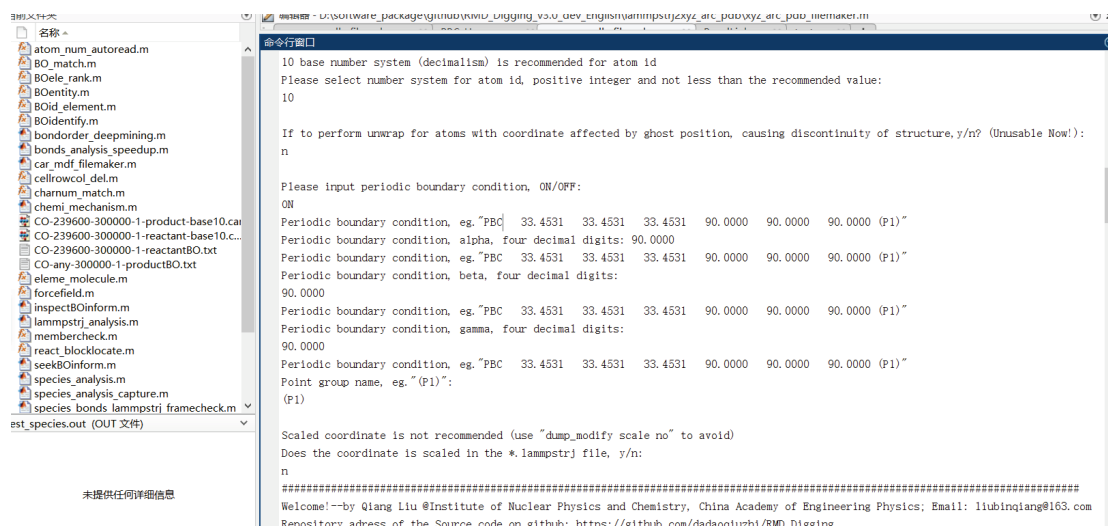
to give answer about if to run check process (need extra time and memory for large files). If the residual check file in the workspace is found, rerun question will be triggered.

11. 10. The program asks the user to check the frame No. in the species.\* file meets the arithmetic progression. Sometimes energy minimization will cause problem.
12. Input the atom type used in the simulation, which should be separated by white space for more than one element. And the order must be identical to the information in \*.data or in.\* files.
13. The console outputs message that you should verify if you performed element mapping (eg, O is mapped to S). For general simulations, it is no (n).
14. Input the number system (binary, decimal and decimal, etc) for the task, which will be considered to encode the atom id. The selection of the number system is based on the total atom number and the limitation on the length of the name/serial number of any atom (id). A recommended value will be shown and the user should input a value not less than the recommended one.
15. The program asks whether to perform unwrap for atoms with coordinate affected by ghost position, causing discontinuity of structure. This can be managed by Materials studio via changing PBC display mode. However, this function here cannot realize this goal now.
16. Input answers for a periodic structure. ON for periodic structure, OFF for the contrary.
17. For periodic structure, the cell parameters (angle, four decimal place) and space group (P1 for general case) should be provided according to the prompt message.
18. Input answers for scaled coordinate. User should check the in.\* file to know if the coordinate in the \*.lammprj file is scaled. The scaled coordinate will be converted into a real space coordinate. I recommend not to scale the coordinate.

19. The `chemi_mechanism` code invokes `species_analysis` code to read and process species data. Note: if there has been a variable related to the processed species data, a warning is triggered to ask if re-execute the `species_analysis` code. This function is aimed at several analysis requirements, which is time-saving for large data. Users can also execute `species_analysis` code individually to obtain basic analysis before using `chemi_mechanism` code.
20. Then `species_capture` is called to abstract the interesting products.
21. When the species information is ready and the species module exists, a pop-up window is shown.
22. The `chemi_mechanism` code asks for further input before the next action.
23. Input the frame/timestep number of the interesting products.
24. The program checks the total number of interesting products specified in the frame, which is printed on the screen. Meanwhile, the program asks whether to restrict the number of exportation times of the reaction pathway pairs (products->reactants). Because there are usually many reaction pathways, so the restriction on exportation times is always necessary. This will reduce the search range. Users should act according to circumstances.
25. If you expect to impose restriction on exportation times, an explicit number should be given. Otherwise, all pairs of reaction pathways will be searched and exported.
26. The program further asks if to only search the trajectory frame with varying number of products. This can save much time if users authorize the program to only search the trajectory frame with a changed product number. For option 2, the subroutine only searches the foregoing trajectory frame with decreasing product number. For option 4, the subroutine only searches the subsequent trajectory frame with the increasing product number.
27. When all inputs are accepted, the `chemi_mechanism` code calls the `bonds_analysis` module and `lammprj2xyz_arc_pdb` module to perform the appointed task. The completed progress is briefly shown in real-time.

28. Finally, the target files with the specified format are generated in the working folder. One can further analyze them by other software like Materials Studio or VMD. Moreover, these formatted files are in favor of searching, checking and analyzing other relevant information.







名称

atom\_num\_autoread.m

BO\_match.m

BOele\_rank.m

BOentity.m

BOId\_element.m

BOIdentify.m

bondorder\_deepmining.m

bonds\_analysis\_speedup.m

car\_mdf\_filemaker.m

cellrowcol\_del.m

channum\_match.m

chemi\_mechanism.m

CO-239600-300000-1-product-base10.c...

CO-239600-300000-1-reactant-base10.c...

CO-239600-300000-1-reactantBO.txt

CO-any-300000-1-productBO.txt

elme\_molecule.m

forcefield.m

inspectBOinform.m

lammprstr\_analysis.m

membercheck.m

react\_blocklocate.m

seekBOinform.m

species\_analysis.m

species\_analysis\_capture.m

species\_bonds\_lammprstr\_framecheck.m

chemi\_mechanism.m (脚本)

scri file name chemi\_mechanism2

命令窗口

References: 1.Fuel 287 (2021) 119484. 2.ACS Appl. Mat. Interfaces 13(34) (2021) 41287-41302. 3.ACS Appl. Mat. Interfaces 2022, 14, (4),  
4.ACS Materials Letters 2023, 2174-2188. More work is coming!  
#####  
  
bonds\_analysis\_speedup is running, please wait...  
bonds\_analysis\_speedup is successfully finished, BO information is saved in bondoutdata, search line number is recorded in readline  
#####  
Welcome!--by Qiang Liu @Institute of Nuclear Physics and Chemistry, China Academy of Engineering Physics; Email: liubinqiang@163.com  
github仓库地址: https://github.com/dadaoqiuzhi/RMD\_Digging  
References: 1.Fuel 287 (2021) 119484. 2.ACS Appl. Mat. Interfaces 13(34) (2021) 41287-41302. 3.ACS Appl. Mat. Interfaces 2022, 14, (4),  
4.ACS Materials Letters 2023, 2174-2188. More work is coming!  
#####  
  
species\_analysis is running, Please wait...  
bondorder\_deepmining is successfully finished  
Molecular formula is saved in tarelenummatch, corresponding BO information is saved in tarBOinform  
Target species number(3) is consistent with that of species recordation(3)  
Successfully delete the unexpected BO information in tarBOinform  
#####  
Welcome!--by Qiang Liu @Institute of Nuclear Physics and Chemistry, China Academy of Engineering Physics; Email: liubinqiang@163.com  
References: 1.Fuel 287 (2021) 119484. 2.ACS Appl. Mat. Interfaces 13(34) (2021) 41287-41302. 3.ACS Appl. Mat. Interfaces 2022, 14, (4),  
4.ACS Materials Letters 2023, 2174-2188. More work is coming!  
#####  
  
lammprstr\_analysis is running, please wait...  
lammprstr\_analysis is successfully finished  
Atom coordinate of the target trajectory is saved in trjdata  
New frame is found 289600, species number 2 is less than that of last frame 300000, which has 3#####

atom\_num\_autoread.m

BO\_match.m

BOele\_rank.m

BOentity.m

BOId\_element.m

BOIdentify.m

bondorder\_deepmining.m

bonds\_analysis\_speedup.m

car\_mdf\_filemaker.m

cellrowcol\_del.m

channum\_match.m

chemi\_mechanism.m

CO-239600-300000-1-product-base10.c...

CO-239600-300000-1-reactant-base10.c...

CO-239600-300000-1-reactantBO.txt

CO-any-300000-1-productBO.txt

elme\_molecule.m

forcefield.m

inspectBOinform.m

lammprstr\_analysis.m

membercheck.m

react\_blocklocate.m

seekBOinform.m

species\_analysis.m

species\_analysis\_capture.m

species\_bonds\_lammprstr\_framecheck.m

chemi\_mechanism.m (脚本)

scri file name chemi\_mechanism2

命令窗口

New frame is found 289600, species number 2 is less than that of last frame 300000, which has 3#####  
Welcome!--by Qiang Liu @Institute of Nuclear Physics and Chemistry, China Academy of Engineering Physics; Email: liubinqiang@163.com  
References: 1.Fuel 287 (2021) 119484. 2.ACS Appl. Mat. Interfaces 13(34) (2021) 41287-41302. 3.ACS Appl. Mat. Interfaces 2022, 14, (4),  
4.ACS Materials Letters 2023, 2174-2188. More work is coming!  
#####  
  
bonds\_analysis\_speedup is running, please wait...  
bonds\_analysis\_speedup is successfully finished, BO information is saved in bondoutdata, search line number is recorded in readline  
#####  
Welcome!--by Qiang Liu @Institute of Nuclear Physics and Chemistry, China Academy of Engineering Physics; Email: liubinqiang@163.com  
github仓库地址: https://github.com/dadaoqiuzhi/RMD\_Digging  
References: 1.Fuel 287 (2021) 119484. 2.ACS Appl. Mat. Interfaces 13(34) (2021) 41287-41302. 3.ACS Appl. Mat. Interfaces 2022, 14, (4),  
4.ACS Materials Letters 2023, 2174-2188. More work is coming!  
#####  
  
species\_analysis is running, Please wait...  
bondorder\_deepmining is successfully finished  
Molecular formula is saved in tarelenummatch, corresponding BO information is saved in tarBOinform  
  
Group 1 reactants(289600)-products(300000) hit nothing,already hit 0 groups, continue to search the next frame of reactants  
  
New frame is found 239600, species number 1 is less than that of last frame 289600, which has 2#####  
Welcome!--by Qiang Liu @Institute of Nuclear Physics and Chemistry, China Academy of Engineering Physics; Email: liubinqiang@163.com  
References: 1.Fuel 287 (2021) 119484. 2.ACS Appl. Mat. Interfaces 13(34) (2021) 41287-41302. 3.ACS Appl. Mat. Interfaces 2022, 14, (4),  
4.ACS Materials Letters 2023, 2174-2188. More work is coming!  
#####

atom\_num\_autoread.m

BO\_match.m

BOele\_rank.m

BOentity.m

BOId\_element.m

BOIdentify.m

bondorder\_deepmining.m

bonds\_analysis\_speedup.m

car\_mdf\_filemaker.m

cellrowcol\_del.m

channum\_match.m

chemi\_mechanism.m

CO-239600-300000-1-product-base10.c...

CO-239600-300000-1-reactant-base10.c...

CO-239600-300000-1-reactantBO.txt

CO-any-300000-1-productBO.txt

elme\_molecule.m

forcefield.m

inspectBOinform.m

lammprstr\_analysis.m

membercheck.m

react\_blocklocate.m

seekBOinform.m

species\_analysis.m

species\_analysis\_capture.m

species\_bonds\_lammprstr\_framecheck.m

chemi\_mechanism.m (脚本)

scri file name chemi\_mechanism2

命令窗口

#####  
  
bonds\_analysis\_speedup is running, please wait...  
bonds\_analysis\_speedup is successfully finished, BO information is saved in bondoutdata, search line number is recorded in readline  
#####  
Welcome!--by Qiang Liu @Institute of Nuclear Physics and Chemistry, China Academy of Engineering Physics; Email: liubinqiang@163.com  
github仓库地址: https://github.com/dadaoqiuzhi/RMD\_Digging  
References: 1.Fuel 287 (2021) 119484. 2.ACS Appl. Mat. Interfaces 13(34) (2021) 41287-41302. 3.ACS Appl. Mat. Interfaces 2022, 14, (4),  
4.ACS Materials Letters 2023, 2174-2188. More work is coming!  
#####  
  
species\_analysis is running, Please wait...  
bondorder\_deepmining is successfully finished  
Molecular formula is saved in tarelenummatch, corresponding BO information is saved in tarBOinform  
  
Now is searching group 2 reactants(239600)-products(300000), this is the hit group, already hits 1 groups(including this group),  
3 groups in total to be processed, limited to export 1 groups#####  
Welcome!--by Qiang Liu @Institute of Nuclear Physics and Chemistry, China Academy of Engineering Physics; Email: liubinqiang@163.com  
References: 1.Fuel 287 (2021) 119484. 2.ACS Appl. Mat. Interfaces 13(34) (2021) 41287-41302. 3.ACS Appl. Mat. Interfaces 2022, 14, (4),  
4.ACS Materials Letters 2023, 2174-2188. More work is coming!  
#####  
  
lammprstr\_analysis is running, please wait...  
lammprstr\_analysis is successfully finished  
Atom coordinate of the target trajectory is saved in trjdata#####  
Welcome!--by Qiang Liu @Institute of Nuclear Physics and Chemistry, China Academy of Engineering Physics; Email: liubinqiang@163.com

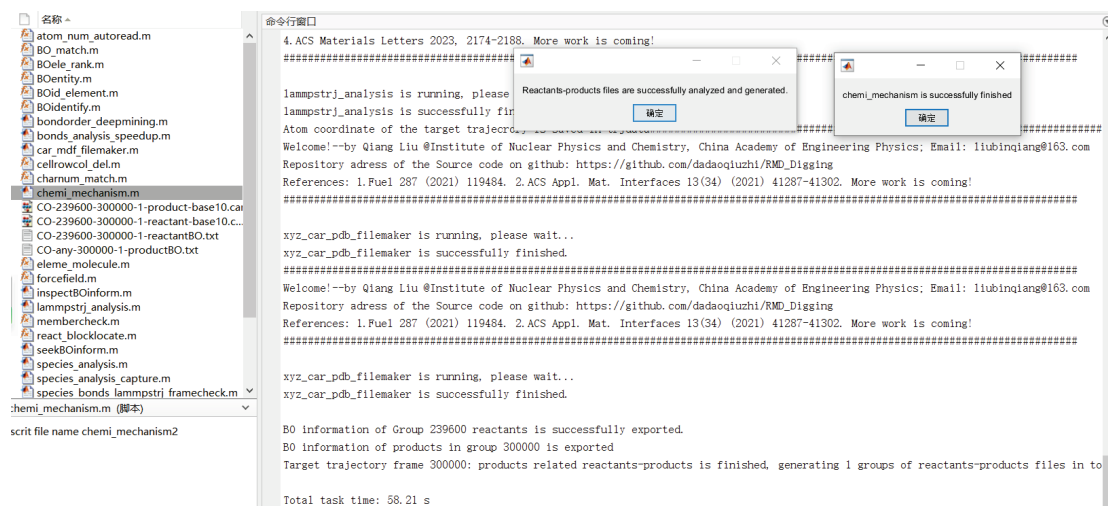


Figure 12 Examples of running chemi\_mechanism code of the chemi\_mechanism module.

Here a brief introduction on the further analysis and visualization of the generated files is given. The generated files in the last tutor are shown in Figure 13. We can see the CO<sub>2</sub> product in the 348000 frame can be formed by different reaction pathways, indicated by the reactants that occurred in the timestep of 262000, 299200, 314000, etc. The number after 348000 denotes the number of reactions in this frame that is related to the formation of the interesting product. The files with the file extension “car” can be visualized and analyzed by Materials Studio. For each product-reactant pair, the bond information is exported for subsequent use (see *\*-reactantBO.txt*). The bond information of the interesting product specified by the users in the very beginning is also exported, whose name has “any” character (see *\*-productBO.txt*). One can combine these files (*\*-reactantBO.txt*, *\*-productBO.txt* and *\*.car*) with the knowledge of software operation to analyze the detailed reaction pathways.

CO2-262000-348000-1-product.car	2022-4-26 9:32	Materials Studio 3D Atomistic Document	12 KB
CO2-262000-348000-1-reactant.car	2022-4-26 9:32	Materials Studio 3D Atomistic Document	12 KB
CO2-262000-348000-1-reactantBO.txt	2022-4-26 9:32	文本文档	12 KB
CO2-299200-348000-1-product.car	2022-4-26 9:32	Materials Studio 3D Atomistic Document	10 KB
CO2-299200-348000-1-reactant.car	2022-4-26 9:32	Materials Studio 3D Atomistic Document	10 KB
CO2-299200-348000-1-reactantBO.txt	2022-4-26 9:32	文本文档	10 KB
CO2-314000-348000-1-product.car	2022-4-26 9:31	Materials Studio 3D Atomistic Document	6 KB
CO2-314000-348000-1-reactant.car	2022-4-26 9:31	Materials Studio 3D Atomistic Document	6 KB
CO2-314000-348000-1-reactantBO.txt	2022-4-26 9:31	文本文档	7 KB
CO2-321600-348000-1-product.car	2022-4-26 9:31	Materials Studio 3D Atomistic Document	1 KB
CO2-321600-348000-1-reactant.car	2022-4-26 9:31	Materials Studio 3D Atomistic Document	1 KB
CO2-321600-348000-1-reactantBO.txt	2022-4-26 9:31	文本文档	1 KB
CO2-any-348000-1-productBO.txt	2022-4-26 9:32	文本文档	13 KB

Figure 13 Files generated by task 2 in the chemi\_mechanism code of the chemi\_mechanism module.

## 5 Reference<sup>1-7</sup>

- (1) Liu, Q.; Liu, S.; Lv, Y.; Hu, P.; Huang, Y.; Kong, M.; Li, G. Atomic-scale insight into the pyrolysis of polycarbonate by ReaxFF-based reactive molecular dynamics simulation. *Fuel* **2021**, 287, 119484, DOI: <https://doi.org/10.1016/j.fuel.2020.119484>.
- (2) Liu, Q.; Huang, W.; Liu, B.; Wang, P.-C.; Chen, H.-B. Gamma Radiation Chemistry of Polydimethylsiloxane Foam in Radiation-Thermal Environments: Experiments and Simulations. *ACS Appl. Mat. Interfaces* **2021**, 13 (34), 41287-41302, DOI: 10.1021/acsami.1c10765.
- (3) Chenoweth, K.; Cheung, S.; van Duin, A. C.; Kober, E. M. Simulations on the thermal decomposition of a poly(dimethylsiloxane) polymer using the ReaxFF reactive force field. *J. Am. Chem. Soc.* **2005**, 127 (19), 7192-202.
- (4) Vashisth, A.; Ashraf, C.; Zhang, W.; Bakis, C. E.; Duin, A. C. T. V. Accelerated ReaxFF Simulations for Describing the Reactive Cross-Linking of Polymers. *J. Phys. Chem. A* **122** (32), acs.jpca.8b03826-.
- (5) Yoon, K.; Rahnamoun, A.; Swett, J. L.; Iberi, V.; Cullen, D. A.; Vlassiounk, I. V.; Belianinov, A.; Jesse, S.; Sang, X.; Ovchinnikova, O. S.; Rondinone, A. J.; Unocic, R. R.; van Duin, A. C. T. Atomistic-Scale Simulations of Defect Formation in Graphene under Noble Gas Ion Irradiation. *ACS Nano* **2016**, 10 (9), 8376-8384, DOI: 10.1021/acs.nano.6b03036.
- (6) Liang, T.; Yun, K. S.; Cheng, Y. T.; Yilmaz, D. E.; Vishnu, K. G.; Verners, O.; Zou, C.; Phillpot, S. R.; Sinnott, S. B.; Duin, A. C. T. V. Reactive Potentials for Advanced Atomistic Simulations. *Annual Review of Materials Research* **2013**, 43 (43), 109-129.
- (7) Farah, K.; Müller-Plathe, F.; Böhm, M. C. Classical reactive molecular dynamics implementations: state of the art. *Chemphyschem* **2012**, 13 (5), 1127-1151.