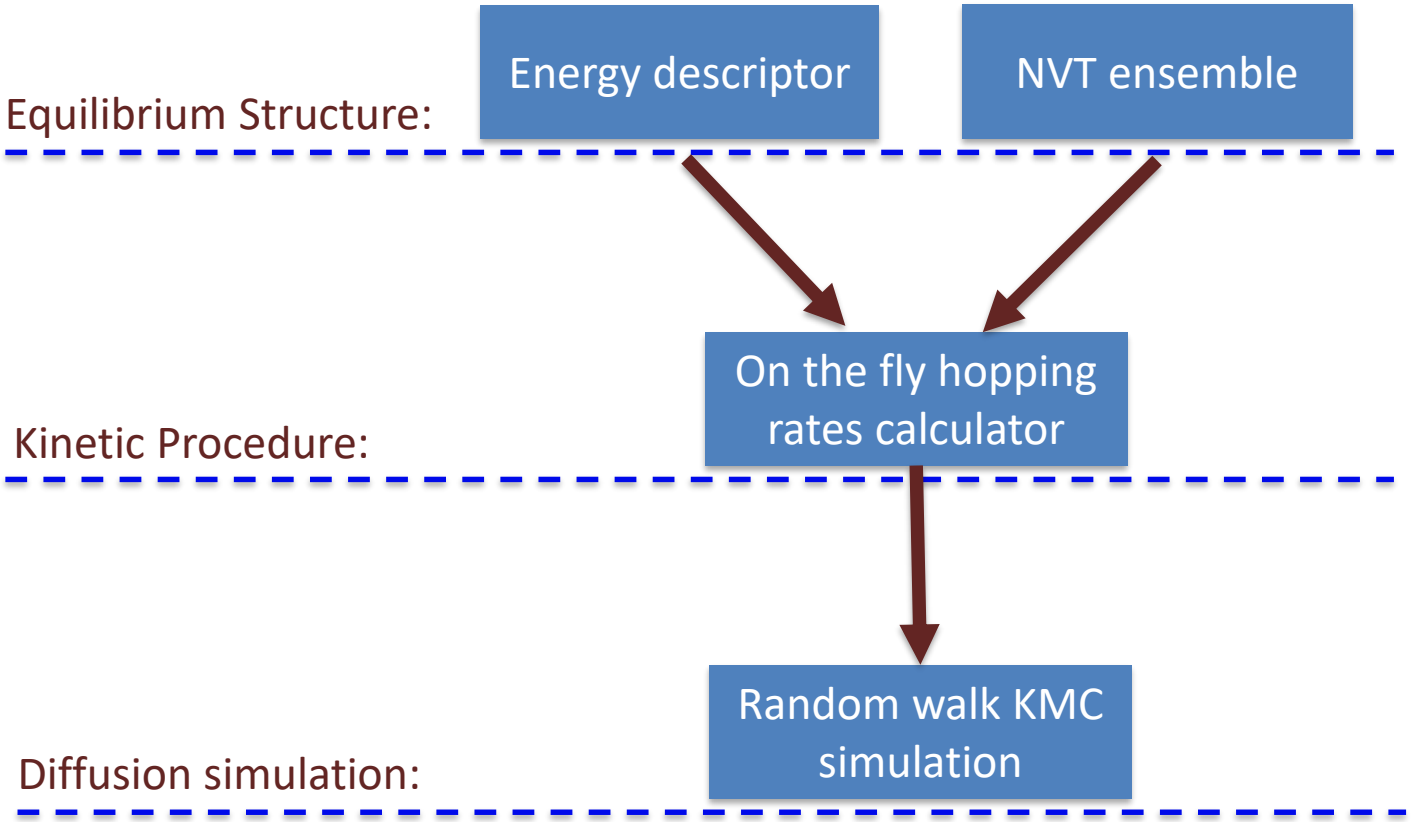
Clusterkit Manual

Description about initial version

Dr. Bin Ouyang
Code page: https://github.com/Jeff-oakley/Clusterkit/
Profile: http://bouyang.wixsite.com/road-to-mse
Email: jeffrey.oakley.ouyang@gmail.com
November 23, 2017

Computational framework



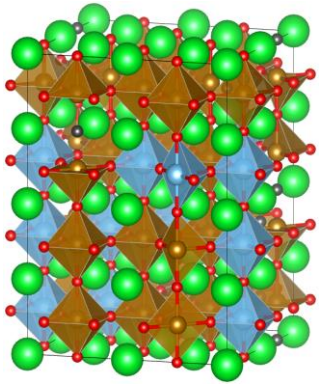
Implementation of cluster algorithms

Equilibrium Structure: **Energy descriptor** **NVT ensemble**

The idea of cluster expansion (CE):

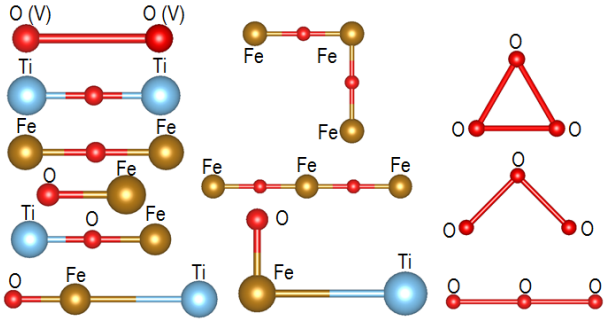
E_{total} of the system can be regarded as summation of atom clusters with 1 atom to N atoms

$$E = \sum_i n_i \varepsilon_i + \sum_{i>j} n_{ij} \varepsilon_{ij} + \sum_{i>j>k} n_{ijk} \varepsilon_{ijk} + \sum_{i>j>k>l} n_{ijkl} \varepsilon_{ijkl} \cdots\cdots$$



Made of clusters

Monomer: Sr, Fe, Ti, O
Dimer: Sr-Fe, Fe-Fe, Ti-Ti, O-O, Fe-O et al.
Trimer: Fe-Fe-Fe, O-O-O and Fe-Ti-O et al.
>> Fe-Ti-O: long chain, short chain, right triangle
>> Fe-Fe-Fe: triangle, chain
>> O-O-O: equilateral triangle, right triangle, chair
Tetramers and beyond.



The idea of Metropolis Monte Carlo (MC):

Random sample within configurational space with acceptance rate according to energy criterion

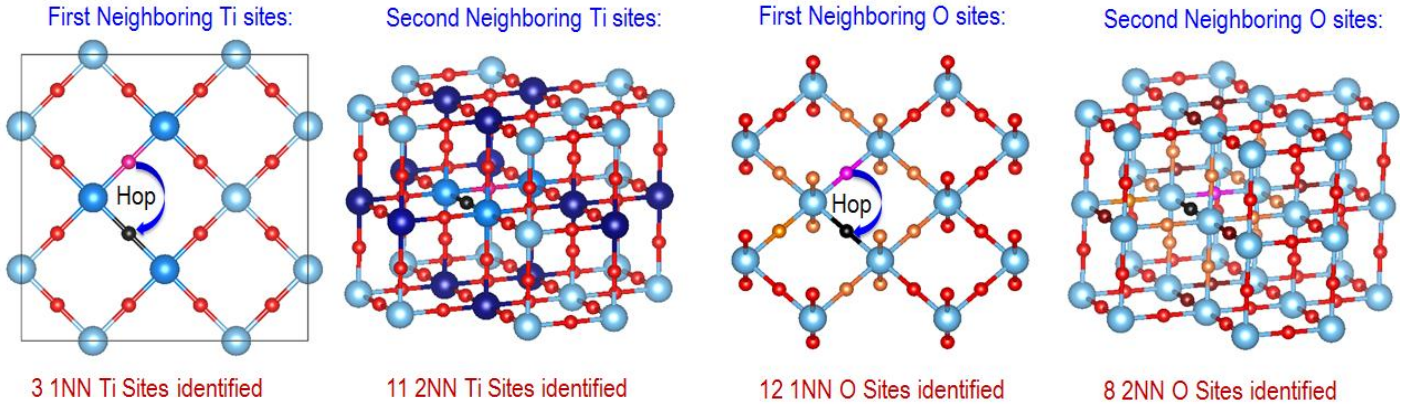
$$P = \exp(\frac{\Delta E}{k_B T})$$

Estimation of kinetic barrier

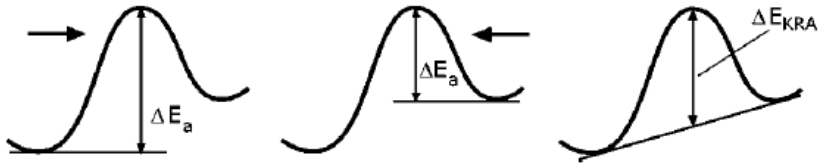
Kinetic Procedure:

On the fly hopping rates calculator

Finite sampling of local cluster interactions:



Kinetic cluster expansion:



$$\Delta E_{KRA} = E_{saddle} - \frac{1}{n}\sum_{j=1}^{n} E_{e_j}$$

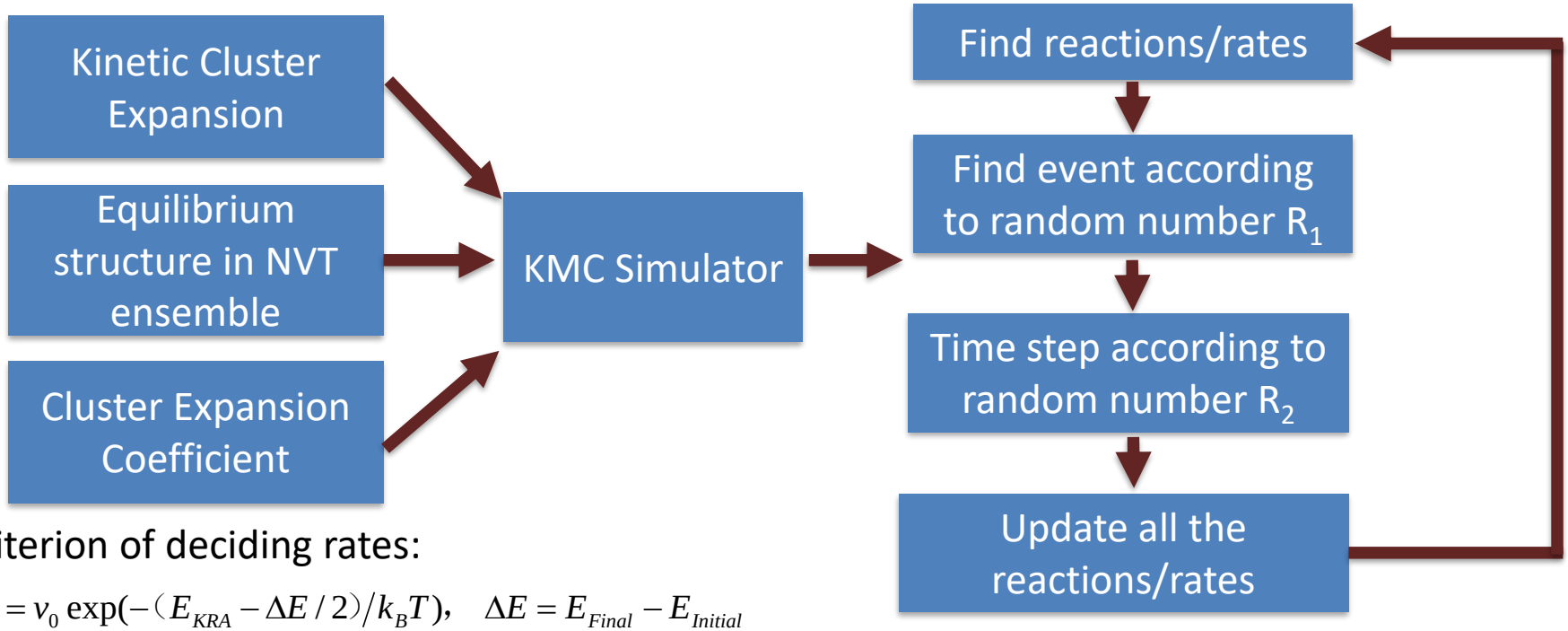
$$\Delta E_{KRA} = E_0 + \sum_{\alpha} n_\alpha \varepsilon_\alpha$$

$$\Delta E_a = \Delta E_{KRA} + \frac{1}{n}\left(\sum_{j=1}^{n} E_{e_j}\right) - E_{e_1}$$

KMC simulation for diffusion process

Diffusion simulation:

Random walk KMC simulation



Criterion of deciding rates:

$$\Gamma_m = \nu_0 \exp(-(E_{KRA} - \Delta E/2)/k_B T), \quad \Delta E = E_{Final} - E_{Initial}$$

Criterion of deciding jump phenomenon:

$$\frac{1}{\Gamma_{tot}} \sum_{m=1}^{k-1} \Gamma_m < R_1 \le \frac{1}{\Gamma_{tot}} \sum_{m=1}^{k} \Gamma_m$$

Criterion of updating diffusion time:

$$\Delta t = -\frac{1}{\Gamma_{tot}} \ln \xi$$

What can be done using Clusterkit?

1. Automatic workflow of cluster algorithms.
2. Cluster expansion description of solid solution system
3. Measurement of thermodynamic properties under NVT/NPT ensemble.
4. Construction of equilibrium structure at finite temperature.
5. Predicting of kinetic energy using kinetic cluster expansion.
6. Simulation of diffusion process with non-dilute concentration of diffusors and vacancies.

Installation

Download the source code at:

https://github.com/Jeff-oakley/Clusterkit/blob/bluewater/dist/Clusterkit-0.1.tar.gz or
https://github.com/Jeff-oakley/Clusterkit

Prerequisite:

numpy==1.11.1
pyyaml==3.12

And extract (if start with the tarball):

tar xvfz Clusterkit-0.1.tar.gz
cd Clusterkit-0.1/

Building using setup.py:

python setup install --user

Modify PYTHONPATH and PATH:

export PYTHONPATH=INSTALLDIR/lib/python2.7/site-packages:$PYTHONPATH
export PATH = INSTALLDIR/bin: $PATH

For automatic loading for each running:

Add the path variables in your .bashrc (or maybe .bash_profile), .zshenv,
or other script for the other shells.

Utilities

ClusterFind (Identify stoichiometry independent clusters):

usage: ClusterFind [-h] -Sub SUBLATT [--NCut NCUT] [--Ref REF]

Script for running CEMC.

optional arguments:
```
  -h, --help    show this help message and exit
  -Sub SUBLATT  Tell projection of sublattice with respect
                to reference lattice. We are expecting something like
                "[[0,1],[2,3],[4,5]]", which tells the code that we have two
                types of atoms for each of the three sublattices
  --NCut NCUT   The cutoff of cluster size (default: 3)
  --Ref REF     Reference lattice (default: POSCAR_Primary)
```

Utilities

CEExtract (extract cluster information from samples DFT calculations):

usage: CEExtract [-h] -Des CLUSTERDES -Comp COMP -Tag TAG [--RN RDNUM]
[--O OUTFILE] [--P PRECONF] [--Pos POSNAME]

Script for extracting CE data.

optional arguments:
-h, --help show this help message and exit
-Des CLUSTERDES Tell the description of each cluster
-Comp COMP Composition list used for making dir
-Tag TAG Tag list used for each composition
--RN RDNUM Number of random structures for each composition, can be one digit or list
--O OUTFILE Out put file name (default: CEDat.yaml)
--P PRECONF Prefix of each configuration name (default: POSCARRand)
--Pos POSNAME Name of the position file to read (default: POSCAR)

Utilities

CERegress (Performing regressing of cluster coefficient):

usage: CERegress [-h] [--DF CEDATAFILE] [--I CINDLST] [--O OUTFILE]

Script for performing regression.

optional arguments:
  -h, --help show this help message and exit
  --DF CEDATAFILE Tell yaml data file, usually generated from CEExtract (default: CEDat.yaml)
  --I CINDLST Effective indexes used in regressing
  --O OUTFILE

Utilities

CEMC (Running cluster expansion Monte Carlo simulations):

```
usage: CEMC [-h] -des CLUSTERDES -coef CLUSTERCOEF -t TEMP [--GrpLst GRPLST]
            [--N NLST] [--bias BIAS] [--ini INITCONF]

Script for running CEMC.

optional arguments:
  -h, --help        show this help message and exit
  -des CLUSTERDES    Tell the description of each cluster
  -coef CLUSTERCOEF  Tell the coefficient fitted for cluster expansion
  -t TEMP           Tell the temperature of the simulation
  --GrpLst GRPLST    group the sublattices with swap possibilities (default:
                     [[1,2],[3,4]])
  --N NLST          Tell the number of iterations required, 1st: outer loop,
                     2st: inner loop; 3st: step size of outputting structure
                     (default: [20,1500,500])
  --bias BIAS       Tell the bias parameter (default: 0.0)
  --ini INITCONF     Starting Configuration (default: POSCAR_Primary)
```

Utilities

KCEPre (Prepare NEB calculations given KCE descriptions):

```
usage: KCEPre [-h] -Sub SUBLATT -Ele ELENAME -Diff DIFFLATT [--NCut NCUT]
              [--DCut DCUT] [--N NBARRIER] [--C CONF] [--P PNTLST]
              [--O OUTFILE] [--MV MAXVAC]

Script for preparing KCE calculations.

optional arguments:
  -h, --help      show this help message and exit
  -Sub SUBLATT    Tell effective sublattice in reference cell
  -Ele ELENAME    List of elements with substitution
  -Diff DIFFLATT  Tell the diffuser sublattice, should be a integer in this version
  --NCut NCUT     The cutoff of cluster size (default: 2)
  --DCut DCUT     Cutoff distance of inteaction, default is half cell size (default: Half)
  --N NBARRIER    Number of random NEB calculations to be created
  --C CONF        Reference structures (default: POSCARRef)
  --P PNTLST      The StaInd and FinInd reference, should be a list of two index: StaInd and FinInd
(default: Center)
  --O OUTFILE     Output file for the KCE informations (default: KCEConf.yaml)
  --MV MAXVAC     Set maximum NVac to avoid significant thrinkage (default: 6)
```

Utilities

KCEExtract (Extract KCE information from NEB calculations):

```
usage: KCEExtract [-h] -Des KCLUSTERDES -Conf CONFLST -Map MAPDIC
                  [--Pre PREFIX] [--Ref POSREF] [--O OUTFILE]

Script for extracting KCE data.

optional arguments:
  -h, --help       show this help message and exit
  -Des KCLUSTERDES  Tell the description of each cluster
  -Conf CONFLST     Configuration list used for making dir
  -Map MAPDIC       Dictionary for position remapping
  --Pre PREFIX      Prefix of sample folders
  --Ref POSREF      Reference POSCAR with out doping
  --O OUTFILE       Out put file name (default: KCEDat.yaml)
```

Utilities

CVChk (Cross validation of CE and KCE regression):

Under developments…

The cross validation on clusters and samples are necessary, however this can be done conveniently using a extra script for specific system. A general utility will is under development to offer calculation of cross validations scores in one shot.

Utilities

KMC (Performance kinetic Monte Carlo simulations)

```
usage: KMC [-h] -des CLUSTERDES -coef CLUSTERCOEF -kdes KCLUSTERDES -kcoef
       KCLUSTERCOEF -t TEMP -sitelst SITELST -r RADIUS -fre FRE0
       [--ini INITCONF] [--ref REF] [--IsNew ISNEW]

Script for running KMC simulation.

optional arguments:
  -h, --help          show this help message and exit
  -des CLUSTERDES     Tell the description of each cluster
  -coef CLUSTERCOEF   Tell the coefficient fitted for cluster expansion
  -kdes KCLUSTERDES   Tell the description of each kinetic cluster
  -kcoef KCLUSTERCOEF Tell the coefficient fitted for kinetic cluster expansion
  -t TEMP             Tell the temperature of the simulation
  -sitelst SITELST    List of site information, 1st: sublattice list that contain vacancy2nd:
sublattice in Ref containing vacancy
  -r RADIUS           Radius that atomic jump can happen
  -fre FRE0           Attempt frequency of a certain atom
  --ini INITCONF      Starting Configuration (default: POSCAR_Primary)
  --ref REF           Reference configuration for vacancy identification (default: POSCAR_Ref)
  --IsNew ISNEW       If this is a new calculation (default: 1)
```

Utilities

KMCPost (Post processing of KMC simulations)

```
usage: KMCPost [-h] [--KF KMCOUTFILE] [--P POSFILE] [--O OUTFILE]

Script for extracting the DOSCAR.

optional arguments:
  -h, --help       show this help message and exit
  --KF KMCOUTFILE  Tell file of KMC simulation results (default:
Trace.dat)
  --P POSFILE      Initial configruation for KMC simulation (default:
POSCAR_Primary)
  --O OUTFILE      Output the processed diffusion results (default:
Diffusion.dat)
```

Input files

Primary lattice file in POSCAR format in VASP code for initialize Cluster expansion:

Currently we support only format of VASP with all methods implemented in Vasp.py. However this can be easily extended to other DFT code.

For instance, for the simulation of ABO_3, what we need is a POSCAR file with one specie in A and B site. A example is given below:

```
Sr1 Ti1 O3
1.000000
11.158513 0.000000 0.000000
0.000000 11.158513 0.000000
0.000000 0.000000 15.780520
Sr Ti O
32 32 96
Direct
0.000000 0.000000 0.000000
0.000000 0.000000 0.250000
0.000000 0.000000 0.500000
0.000000 0.000000 0.750000
0.000000 0.500000 0.000000
0.000000 0.500000 0.250000
0.000000 0.500000 0.500000
0.000000 0.500000 0.750000
0.250000 0.250000 0.000000
0.250000 0.250000 0.250000
0.250000 0.250000 0.500000
0.250000 0.250000 0.750000
0.250000 0.750000 0.000000
0.250000 0.750000 0.250000
```

Output files

Configuration database for cluster expansion: CeDat.yaml

This file with YAML format is used to consolidate the data generated from DFT training set. In this case we just need to keep on single file on base of hundreds/thousands of DFT calculations. Meanwhile, all information of cluster expansion would be stored in the same file.

```
A12B20:
  POSCARRand0:
    Cluster Expansion: [-1029.2888, 12, 48, 74, 32, 14, 74, 54, 14, 50, 42, 43, 171,
      19, 16, 26, 18, 182, 548, 274]
    Position File:
      AtomNum: [32, 12, 20, 86, 10]
      AtomSum: 160
      Base:
      - [11.272991, -0.008624, 0.00052]
      - [-0.008341, 11.251252, -0.005254]
      - [0.00073, -0.007529, 15.870263]
      CellName: 'Sr1 Ti1 O3 '
      EleName: [Sr, Ti, Fe, O, Vac]
      EleNum: 5
      IsSel: 0
      LatType: Direct
      LattConst: 1.0
      LattPnt:
      - [0.0, 0.0, 0.0]
      - [0.0, 0.0, 0.25]
      - [0.0, 0.0, 0.5]
      - [0.0, 0.0, 0.75]
      - [0.0, 0.5, 0.0]
      - [0.0, 0.5, 0.25]
      - [0.0, 0.5, 0.5]
```

OutputFiles

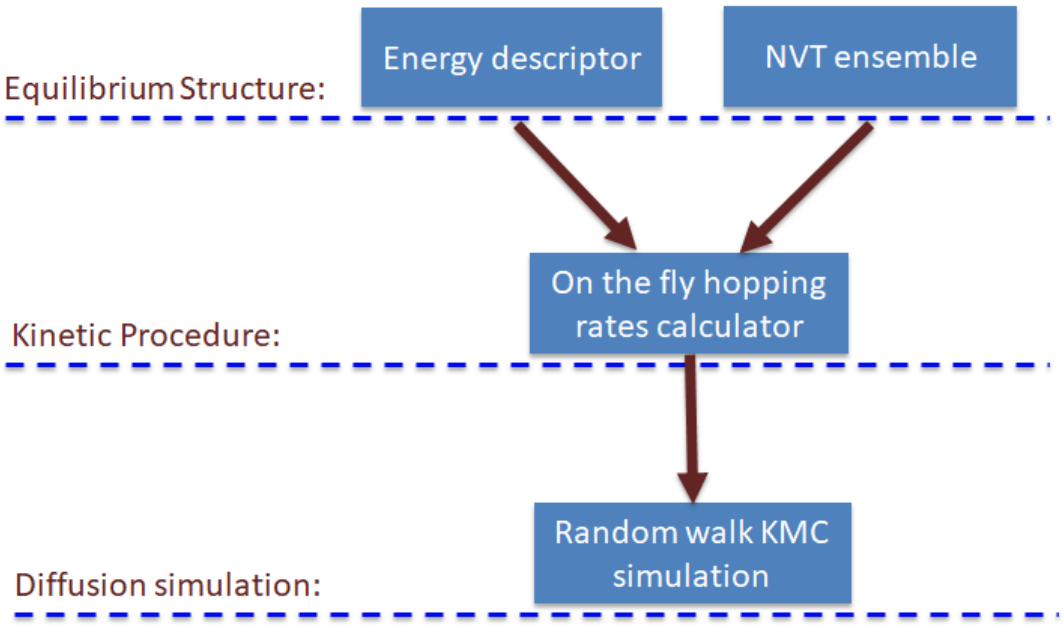
Configuration database for cluster expansion: KCEConf.yaml

This file with YAML format is used to describe the local atomic environment, diffusing atoms, effective local atoms for diffusions

```yaml
DCut: 5.879256499999999
EffPntLst:
- []
- [32, 33, 36, 37, 40, 41, 44, 45, 46, 48, 49, 50, 52, 53, 60, 61]
- [65, 66, 69, 70, 72, 73, 76, 77, 78, 80, 81, 82, 85, 89, 90, 92, 93, 95, 96, 99,
  101, 103, 104, 105, 107, 108, 111, 112, 113, 116, 117, 119, 120, 123, 124, 127]
KClusterDes:
- - [Sta, 1]
  - &id001 [5.029073801039524]
- - [Sta, 2]
  - *id001
- - [Sta, 1]
  - &id002 [4.184442368805207]
```

```yaml
NCut: 2
Pnt:
- [0.375, 0.375, 0.375]
- [0.25, 0.5, 0.25]
PntLst:
- !!python/object/apply:numpy.core.multiarray.scalar
  - &id006 !!python/object/apply:numpy.dtype
    args: [i8, 0, 1]
```

Summary



1. Automatic workflow of cluster algorithms.
2. (Kinetic) Cluster expansion description of solid solution system.
3. Statistic measurement of thermodynamic properties.
4. Diffusion simulation