# A guide to the Tight-Binding FITting (TBFIT) package

Hyun-Jung Kim [h.kim@fz-juelich.de, Infant@kias.re.kr]

March 2, 2021

This document is to provide explanation for the input file arguments of the TBFIT package.

# System Requirements and installation

The program has been written by modern Fortran2008 language. If you want to deactivate the use of some module interfaces written in Fortran2008 syntax, please remove -DF08 option in your option tag of the makefile.

LAPACK library should be properly linked in the makefile. For the eigenvalue solver with sparse matrix, Inspector-executor Sparse BLAS Routines and Extended Eigensolver Routines in the Intel Math Kernel Library (Intel MKL) are referred. If the system size is very big, you can calculate band structure with energy window constraint. This is available with EWINDOW tag and -DMKL\_SPARSE option. To use -DMKL\_SPARSE option, make sure that mkl\_spblas.f90 file is located in your \$MKLPATH/include folder.

To list up the space group information for the given geometry in the initial stages of the calculations, one can activate the use of space group library (Spglib). For this, put -DSPGLIB in your OPTION tag of the makefile, and provide appropriate library path in SPGLIB tag.

- How to install:
  - > tar -xvf TBFIT-master.zip
  - > cd TBFIT-master
  - > make tbfit.mpi (if MPI\_USE=YES)
  - > make tbfit.serial (if MPI\_USE=NO)
- How to run:

In the Example directory, you can run a test cases, for example:

- > cd TBFIT-master/Example/1H-MoS2/SOC
- > tbfit

Note that the output log will be written in TBFIT.out file by default. If you want to write the log with different name, please use -log option as follows:

```
> tbfit -log FNAME.out
```

Note also that the input arguments are read from INCAR-TB by default. However, if you want to provide your prefer file, you can specify by -i or -input option as follows:

```
> tbfit -i INCAR-TB_test
```

If you want to fit multiple target band, you can fit them in a single command line. For example, if you have define system A in INCAR-TB\_A and system B in INCAR-TB\_B, you can run them simultaneously as follows:

```
> tbfit -i INCAR-TB_A -i INCAR-TB_B
```

The output will be written in separate files with numbered index. Important note is that, for a fitting purpose, you should provide SAME parameter file for each system. This is because we assume a situation where the fitting parameter fit multiple system in a single run.

• python module: The TBFIT supports python module as of version 0.5.1. To utilize python module tbfitpy, compile python module and import tbfitpy which is located in PYTHON\_MODULE folder.

#### > make tbfitpy\_mod

After successful compilation, \_tbfitpy\_mod\_mpi.\*so file and tbfitpy\_mod\_mpi.py file will be generated (if MPI = NO in your make file, \_mpi suffix will be removed). Then, copy these files and tbfitpy.py into your \$PYTHONPATH. The prerequisites for the python modules are as follows:

⊙ f90wrap, NumPy, Matplotlib, PySwarms, tqdm, mpi4py

The example can be found in the "Example/Graphene/BAND\_FIT/ Step\_3.pso\_method/1.PSO/1.NPARTICLE\_50/python\_module" of your example folder.

# Part I. User's Guide

## 1. INPUT tags of the INCAR-TB

**TITLE** string Default: none The title of the system can be specified with this tag. Note that there should be no blank as this TITLE tag can be used in extension of the several output/input file name.

For example, one can specify as follows:

TITLE Graphene-1ML

Then, the output for band structure will be written in

band\_structure\_TBA.Graphene-1ML.dat

**GET\_BAND** logical Default: .TRUE. If .TRUE. TBFIT will perform tight-binding calculations for band structure evaluation.

**TBFIT** logical Default: .FALSE.

.TRUE.: Perform tight-binding parameter fitting which is defined in PFILE. After fitting is completed, whatever it is converged or not, additional tight binding calculations as defined in the INCAR-TB will be performed.

.FALSE. : Do not perform fitting procedures. In this case, regular tight binding calculations will be performed.

#### MITER integer Default: 100

Maximum number of iteration for the fitting procedures. If GA is set for LSTYPE, MITER represents the maximum number of generations.

#### **MXFIT** integer Default: 1

Maximum number of repeat for LMDIF procedures after MITER steps done. By checking convergence criteria defined by FDIF the variation of the fitness function, parameter constraints, i.e., upper or lower bounds defined by SET CONSTRAINT tags, TBFIT will determine to stop or not.

#### **LSTYPE** integer Default: LMDIF

Method for parameter fitting. Available tags are LMDIF and GA and PSO.

 ${\tt LMDIF}$  method: Levenberg-Marquardt method<br/>¹, ² using finite-difference for Jacobian.

<sup>&</sup>lt;sup>1</sup> Kenneth Levenberg, "A Method for the Solution of Certain Non-Linear Problems in Least Squares" Quarterly of Applied Mathematics 2, 164 (1944).

<sup>&</sup>lt;sup>2</sup>Donald Marquardt, "An Algorithm for Least-Squares Estimation of Nonlinear Parameters" SIAM Journal on Applied Mathematics 11, 431 (1963).

GA method: Genetic Algorithm<sup>3</sup> based on PIKAIA library<sup>4,5,6,7</sup>. To setup control parameters for the GA, see Sec. GA.

PSO method: Particle swarm optimization method<sup>8, 9</sup>. The control parameters to utilize PSO are: PSO\_NP, PSO\_OPT.

#### PTOL & FTOL real Default: 0.00001

Tolerence of iteration of the fitting procedures for LMDIF method. FTOL is a tolerence for the difference between target and calculated data from tight binding method. PTOL is as tolerence for the tight binding parameters. Normally, both values below 0.00001 is sufficient to reach a local minima.

#### FDIFF real Default: 0.001

Tolerence of iteration of the fitting procedures for LMDIF method. FDIFF is a tolerence for the difference between fitness function of previous and last steps. After finishing LMDIF iterations, fitting procedures might start again if CONSTRAINT and FDIFF condition does not fulfill, and will stops by MXFIT criteria.

#### **PSO\_NP** integer Default: 50

The number of particles in the swarm. This is critical value for the optimization result and should be larger if possible and one can have convergence test with this value.

#### **PSO\_OPT** real, real Default: $c_1$ =0.3, $c_2$ =0.4, and w=0.2

The optional parameter for parameter update policy. In PSO method, the *i*-th parameter set  $\mathbf{p}_i$  in the swarm is updated using its velocity vector  $\mathbf{v}_i^{t+1} = w\mathbf{v}_i^t + c_1r_1(\mathbf{p}_{best,i}^t - \mathbf{p}_i^t) + c_2r_2(\mathbf{p}_{best}^t - \mathbf{p}_i^t)$  so that the updated parameter  $\mathbf{p}_i^{t+1} = \mathbf{p}_i^t + \mathbf{v}_i^{t+1}$ . Here, w is the inertia weight and controls the momentum of the parameter update,  $c_1$  is a self-confidence factor (cognitive) and expresses how much the particle trust its own past experience ( $\mathbf{p}_{best,i}$ ) while  $c_2$  is a swarm confidence factor (social) and expresses how much it trusts the swarm's best experience ( $\mathbf{p}_{best}$ ).  $r_1$  and  $r_2$  is introduced to make swarm explore parameter space randomly searching for new best parameters.

PSO\_OPT 0.3 0.4 0.2 # c1=0.3, c2=0.4, w=0.2

<sup>&</sup>lt;sup>3</sup>D. E. Goldberg, "Genetic Algorithm in Search, Optimization, & Machine Learning" *Addison-Wesley* (1989).

<sup>&</sup>lt;sup>4</sup>P. Charbonneau and B. Knapp, "A user's guide to PIKAIA 1.0", (NCAR Technical Note 418+IA, 1995)

<sup>&</sup>lt;sup>5</sup>P. Charbonneau, "An introduction to genetic algorithm for numerical optimization" (NCAR Technical Note 450+IA, 2002)

<sup>&</sup>lt;sup>6</sup>P. Charbonneau, "Release notes for PIKAIA 1.2" (NCAR Technical Note 451+STR, 2002), http://www.hao.ucar.edu/modeling/pikaia/pikaia.php

<sup>&</sup>lt;sup>7</sup>Modern Fortran Edition of the Pikaia Genetic Algorithm. https://github.com/jacobwilliams/pikaia

<sup>&</sup>lt;sup>8</sup>J. Kennedy and R. Eberhart, "Particle Swarm Optimization" (IEEE, Piscataway, NJ, 1995) p. 1942.

<sup>&</sup>lt;sup>9</sup> R. Eberhart and J. Kennedy, "A new optimizer using particle swarm theory" (IEEE, New York, NY, 1995).

#### **FIT\_PLAIN** integer Default: .FALSE.

Logical flag whether we consider weight factor or not in cost function update in PSO fitting procedure (only available if LSTYPE= PSO).

.TRUE. : not including WEIGHT, .FALSE.: including WEIGHT

#### **NPAR** integer Default: 1

Number of parallel groups over PSO particles. If larger than 1, It divides CPU into NPAR groups in PSO routine. It is useful if number of k-points is less than number of total CPU.

#### **K\_UNIT** string Default: ANGSTROM

ANGSTROM: the unit of the k-point will be written in  $Å^{-1}$  unit.

RECIPROCAL : the unit of the k-point will be written in reciprocal unit (fractional).

#### **PFILE** string Default: PARAM\_FIT.dat

File name for tight-binding parameters. For the details, see Sec.4.

Note 1: After fitting procedure is over, the WEIGHT information is written in PFILE so that one can restart fitting. To use WEIGHT information written in PFILE, one can add USE\_WEIGHT tag as below:

PFILE PARAM\_FIT.dat USE\_WEIGHT

**POFILE** string Output file name for tight-binding parameters written after fitting procedures. Default: PARAM\_FIT.new.dat

#### **SGPLIB** logical

.TRUE.: Write space group information to the output log.

.FALSE. : Do not write space group information to the output log.

Note that this option is only applicable if you have put -DSPGLIB option in your makefile. See the details in System Requirements and installation section.

#### **EFILE** string, integer

File name for the target band structure for the fitting procedures. If the second  $integer\ n$  is followed by, TBFIT will read n-th column as a target band. Default is n=2.

#### EFILE DFT\_BANDSTRUCTURE.dat 2

```
# 1st eigen value
```

# k-path energy(eV)

0.00000 -12.36137

```
0.01693 -12.36162
  0.03386 -12.36118
    [\ldots]
  0.16932 -12.33324
  0.18625 -12.32696
  0.20319 -12.32014
# 2nd eigen value
# k-path energy(eV)
  0.00000 -12.36137
  0.01693 -12.36041
  0.03386 -12.35875
     [\ldots]
  0.16932 -12.32136
  0.18625 -12.31394
  0.20319 -12.30600
    [\ldots]
          _ EFILE DFT_BANDSTRUCTURE.out example _
```

If LORDER = .TRUE. and TBFIT = .TRUE., re-ordered target energy information should be provided in parallel with the original energy, as below:

```
# k-path
                   energy(eV)
                                   energy_ordered(eV) : 1st eig
   0.00000000
                   -14.75638384
                                     -14.75638384
   0.04796527
                   -14.74941045
                                     -14.74941045
        [\ldots]
   2.95861071
                   -14.74708715
                                     -14.74708715
   3.01399626
                   -14.75638384
                                     -14.75638384
        [...]
# k-path
                   energy(eV)
                                  energy_ordered(eV) : 3rd eig
   0.00000000
                    -2.27711578
                                      -2.27711578
   0.04796527
                    -2.27002886
                                      -2.27002886
         [\ldots]
   2.95861071
                    -2.26767593
                                      -0.44510940
   3.01399626
                    -2.27711578
                                      -0.44395430
        [\ldots]
           _ EFILE DFT_BANDSTRUCTURE_ORDERED.out example ___
```

If the second argument is VASP, TBFIT will read EFILE which is indicated as the first argument and recognize this file as a EIGENVAL file of VASP code. (note: In this case, LORDER should be .FALSE...) For example,

#### EFILE EIGENVAL VASP

#### EFILE\_EF real

The fermi level of EFILE. The energy will be shifted by the EFILE\_EF, i.e.,  $\epsilon_{n,k}^{target} = \epsilon_{n,k}^{DFT} - \text{EFILE\_EF}$ , where  $\epsilon_{n,k}^{DFT}$  is the eigenvalues provided by the EFILE.

#### **GFILE** string Default: POSCAR-TB

File name for the geometry and atomic orbital informations. The format is exactly same as POSCAR of VASP program. For the details of setting atomic orbitals, see Sec.3.

GFILE POSCAR-TB

The example file format of "GFILE" is as follows,

```
MoS2 # comment
  1.00000000000000  # scaling factor
    3.1716343
                 0.000000
                             0.00000 # lattice vector a1
                 2.746715
                             0.00000 # lattice vector a2
    1.5858171
    0.0000000
                 0.000000
                          15.00000 # lattice vector a3
 Mo S
                                      # atomic species
    1
                                   # number of atoms per species
           # coordinate type (direct or cartesian)
Direct
0.00000 0.00000 0.50000 dz2 dxy dx2 dyz dxz # coord, orbital
0.33333 0.33333 0.60645 s px py pz
0.33333 0.33333 0.39354 s px py pz
         \_ POSCAR-TB example: MoS_2 with Mo-d and S-sp \_
```

Note: TBFIT reads "GFILE" and reports it on to the screen in the parsing steps. If "PRINT\_GEOM .FALSE." tag is followed by, then program will not report the geometry informations, for the convenience.

GFILE POSCAR-TB PRINT\_GEOM .FALSE.

#### KFILE string Default: KPOINTS\_BAND

File name for the k-point setting. The format is exactly same as KPOINTS of VASP program.

```
k-points line mode example
40 ! intersections
Line-mode
Reciprocal
```

```
0.50000000 0.5000000 0 M

0.33333333 0.6666666 0 K

0.00000000 0.0000000 0 G

0.00000000 0.0000000 0 G

0.66666666 0.33333333 0 K'

KPOINTS_BAND line mode example
```

```
k-points grid mode example

0

GMonkhorst-Pack #'G'amma centered grid mode

4 4 1 # grid nk_1 nk_2 nk_3

0 0 0 # shift

KPOINTS_BAND grid mode example
```

Note1: In VASP, n k-points are generated along a k-path if n division is specified in the second line of KFILE. This will gives  $n \times m$  total k-points and n-1 divisions along each k-path, where m represents the total number of k-path specified. And at the end of each line segment of k-path, same k-points are marked twice. However, some program, such as FLEUR uses different strategy in generating k-path, that is, n+1 k-points along each path with only one k-points at the end of line segments. Hence, EFILE originated (or postprocessed) from FLEUR would have different format. This will gives  $n \times m + 1$  total k-points and n division of each k-path. You can read this kind of EFILE file as well, if you have also specify as below, letting TBFIT to recognize the type of generated k-path.

```
KFILE KPOINTS_BAND FLEUR
```

Note2: If your target band structure is originated from FHI-aims, the division between each k-path can be specified separately. For example, the typical way to specify band structure calculation tag in FHI-aims looks as follows:

```
output band 0.00
                   0.00
                         0.00
                                  0.50
                                         0.00
                                               0.00
                                                      70
                                                              Χ
                                                           G
output band 0.50
                                  0.50
                                                      20
                   0.00
                         0.00
                                         0.50
                                               0.00
                                                           Χ
                                                              Μ
output band 0.50
                   0.50
                         0.00
                                  0.00
                                         0.50
                                               0.00.70
                                                           М
                                                              Y
```

Here, each k-path is divided with 70, 20, and 70 divisions. In this case, you can specify each numbers (70 20 70) to your second line of KFILE.

```
k-points line mode example
70 20 70 ! intersections
Line-mode
Reciprocal
0.0 0.0 0.0 G
0.5 0.0 0.0 X

0.5 0.0 0.0 X

0.5 0.5 0.0 M

0.5 0.5 0.0 M

KPOINTS_BAND line mode (FHI-aims type)
```

#### **KREDUCE** integer Default: 1

For some cases, the target EFILE contains too many k-points with larger n of divisions between each line segment. In this case, you can reduce the number of k-points to be read by skipping k-points except every  $nk_{reduce}$ .

For example, if it has been divided with 10 k-points between each symmetry points in your EFILE, you can reduce by factor of 2 or 5 by specifying

KREDUCE 2 or KREDUCE 5 .

In this case, if KREDUCE is 2, then the total number of divisions will become 5, and this value should be same as number of k-point division which is specified in second line of your KFILE. Note that KREDUCE should be one of prime number of the k-point division of EFILE.

#### **LOCCHG** logical Default: .FALSE.

Setting tag for local potential. If .TRUE., one should give proper local potential parameter in your PFILE and should properly setup LOCAL.POT tag in your GFILE. For the details, see the explanation of LOCAL.POT in Sec.4.

#### **TYPMAG** string Default: NONMAG

Setting tag for magnetic moment: nonmagnetic, collinear, noncollinear If collinear and noncollinear tag is applied, MOMENT or MOMENT.C in the GFILE should be set up appropriately. For details, see MOMENT of the Sec.3.

#### **LSORB** logical Default: .FALSE.

Setting tag for spin-orbit coupling. If .TRUE.,  $lambda\_orb\_spec$  should be properly defined in the PFILE. For details, see Sec.4

#### **LORBIT** logical, string(optional), string(optional) Default: .TRUE.

Setting tag for orbital decomposed output. If .TRUE. the local orbital contribution will be printed out in bandstructure\_TBA.dat file. If you write rh or mx or my or mz next to the logical text with .TRUE., then, corresponding magnetization values, which represents the expectation value of pauli matrices  $\sigma_i$  where  $i=\{0,1,2,3\}$ , will be printed out. Here  $\sigma_0$  is  $2\times 2$  identity matrix to print out local orbital contribution. For example,

LORBIT .TRUE. mz

will print out <mz>. If you write re or im next to the logical text with .TRUE., then, real or imaginary part of the wavefunction coefficient will be printed out. Note that this option only applicable with LSORB .FALSE. in the current version.

LORBIT .TRUE. re

If you write wf next to the logical text with .TRUE., then, the wavefunction coefficient will be printed out. The real and imaginary part for each orbital basis is written. If LSORB .TRUE., the spinor-up and spinor-dn part will be written, so that four real values will construct wavefunction coefficient.

LORBIT .TRUE. wf

Note that the corresponding output file bandstructure\_TBA.dat file will be basically written by ascii (formatted) format. If you want to write in binary (unformatted) format, specify by bin (complex\*16; double precision) or bin4 (complex\*8; single precision) tag next<sup>10</sup>. For example,

LORBIT .TRUE. wf bin or LORBIT .TRUE. wf bin4

This tag will generate band\_structure\_TBA. (up/dn).bin file.

#### **PROJ\_BAND** logical, integers Default: .FALSE.

Setting tag for orbital/atom projected band structure output. The output will be written in separate file for each atom. The correct usage is as follows:

<sup>&</sup>lt;sup>10</sup>In the current version, bin or bin4 tag is only applicable when it is combined with wf option

```
PROJ_BAND .TRUE. 1:4 7
```

then you can get band\_structure\_atom.#.dat file

```
band_structure_atom.1.dat
band_structure_atom.2.dat
band_structure_atom.3.dat
band_structure_atom.4.dat
band_structure_atom.7.dat
```

and additionally, band\_structure\_atom.sum1.dat file will be printed out as well, where projected local DOS for those atoms ( $\{1,2,3,4,7\}$ ) are summed up in a single file.

If you write another PROJ\_BAND tag specifying different atom sets in the separate line, for example,

```
PROJ_BAND .TRUE. 1:4 7
PROJ_BAND .TRUE. 5:8
```

then, you can get another set of files (band\_structure\_atom.sum2.dat and band\_structure\_atom.{5..8}.dat) as follows:

```
band_structure_atom.1.dat
band_structure_atom.2.dat
band_structure_atom.3.dat
band_structure_atom.4.dat
band_structure_atom.5.dat
band_structure_atom.6.dat
band_structure_atom.7.dat
band_structure_atom.8.dat
band_structure_atom.sum1.dat (= 1+2+3+4+7)
band_structure_atom.sum2.dat (= 5+6+7+8)
```

Note1: The atom index should be written in ascending order and should not exceed total number of atoms of your system.

Note2: Since in the band\_structure\_atom.sum?.dat file the projected local DOS for each orbital basis of the specified atoms are summed up, one should make sure

that the specified atoms should have same orbital basis set. For instance, let's assume that you have atoms  $A=\{1:4,7\}$  with orbitals  $p_x,p_y,p_z$ , and atoms  $B=\{5:8\}$  with orbitals  $s,p_x,p_y,p_z$ . Then it will works fine with "PROJ\_BAND .TRUE. 1:4 7" tag, but not works fine with "PROJ\_BAND .TRUE. 1:4 5:8".

#### **CIRC\_DICHROISM** logical, integer, integer Default: .FALSE.

Calculate  $\vec{k}$ -resolved degree of polarization by circularly polarized light. Here, we compute the circular dichroism  $\eta_{mn}(\vec{k})$ , which demonstrate the optical chiral selection rule from state m to n of  $\vec{k}$ -point, using  $\eta_{mn}(\vec{k}) = \frac{|\mathcal{P}_{+}^{nm}(\vec{k})|^2 - |\mathcal{P}_{-}^{nm}(\vec{k})|^2}{|\mathcal{P}_{+}^{nm}(\vec{k})|^2 + |\mathcal{P}_{-}^{nm}(\vec{k})|^2}$ .

The matrix element  $\mathcal{P}_{\pm}^{nm}(\vec{k}) = \langle \psi_n(\vec{k}) | \mathcal{P}_{\pm}(\vec{k}) | \psi_m(\vec{k}) \rangle$  determines the  $\sigma_+$  and  $\sigma_-$  light adsorption, respectively, between band index n and m. Here, the interband matrix element  $\mathcal{P}_{\pm}(\vec{k})$  is given by  $\mathcal{P}_{\pm}(\vec{k}) = \mathcal{P}_x(\vec{k}) + i\mathcal{P}_y(\vec{k})$ , where  $\mathcal{P}_{\alpha}(\vec{k})$  is the canonical momentum operator  $(e/\hbar)\partial\hat{H}_{\vec{k}}/\partial k_{\alpha}$ . The usage is as follows:

```
CIRC_DICHROISM .TRUE. 18 19  # from band m=18 to n=19 CIRC_DICHROISM .TRUE. 18 20  # from band 18 to 20 ...
```

The results will be printed out in CIRC\_DICHROISM.EIG $_m-n$ .dat.

NOTE: this tag is only valid with orthogonal basis with USE\_OVERLAP = .FALSE. in PFILE.

#### **LOAD\_HOP** *logical*, *string* Default: .false.

If .true., one can load hopping file to read  $t_{ij}$  value. The following string should be the file name to be read. And the syntax of the file should be exactly same as the hopping.dat file, which is generated in the initial stages of the calculation. Hence, if you have pre-generated hopping.dat file (with LOAD\_HOP .FALSE.), you can copy it with a different name and modify the elements of  $t_{ij}$  column, and rerun the code with following tag (for example, if you have copied hopping.dat  $\rightarrow$  hopping\_modified.dat):

```
LOAD_HOP .TRUE. hopping_modified.dat
```

Below, you can see that the original hopping element can be modified by changing values of the t\_IJ(eV) column.

<sup>&</sup>lt;sup>11</sup>W. Yao, D. Xiao, and Q. Niu, "Valley-dependent optoelectronics from inversion symmetry breaking" Physical Review B 77, 235406 (2008)

```
1 1 -1.2 -0.7 0.0 ... s ... s ... -3.9 ...

1 2 -1.2 -0.7 0.0 ... s ... px ... 1.9 ...

...
hopping.dat example file
```

```
ORB_I ...
# Iatom Jatom
                    Rij
                                               ORB_J ...
                                                            t_IJ(eV)
                0.0 0.0 0.0 ...
    1
                                                             -2.0
           1
                                                                        . . .
           1
                0.0 0.0 0.0 ...
                                                              0.0
                                      S
                                           . . .
                                                 рх
                                                      . . .
              -1.2 -0.7 0.0 ...
                                                             -3.9
                                           . . .
                                      S
                                                 S
              -1.2 -0.7 0.0 ...
    1
                                                              1.9
                                      S
                                           . . .
                                                 рх
    . . .
                 hopping_modified.dat example file ____
```

#### **IBAND** integer Default: 1 (deprecated)

IBAND is the first eigenstate of the target data of EFILE. This value will be used in the WEIGHT SET section.

#### **FBAND** integer Default: NEIG (deprecated)

NEIG: number of orbital basis of the system. FBAND is the last eigenstate of the target data of EFILE. This value will be used in the WEIGHT SET section.

#### **SCISSOR** integer, real

If set, in the fitting procedures, target energy EDFT (n,k) will be shift by amound of the scissor operation. This operation works as follows:  $E'_{target}(n,k) = E_{target}(n,k) + e_{scissor}$  if  $n >= i_{scissor}$ . Note that this operation is only valied if TBFIT is .TRUE..

#### **NN\_MAX** integer Default: 3

Determine how many times the cell will be repeated in searching hopping pairs. If your system is sufficiently larger than the maximal value of hopping distances of your system, this can be reduced to 1, otherwise just use default value.

 $NN_MAX 3 3 3$ 

or

NN\_MAX 3

both settings will give  $3\times3\times3$  cell repeat.

#### **ERANGE** integer Default: 1 NEIG

If provided, the energy level between these energy window will be printed out in the bandstructure\_TBA.dat file.

ERANGE 4400 4700

Above example means that the energy level from  $4400^{th}$  to  $4700^{th}$  will be printed. This is particularly useful if you calculate very large systems. By setting ERANGE tag, you can save disk space a lot if LORBIT tag is turned on where orbital component information takes huge memory for larger systems.

#### **EWINDOW** real, integer Default: not activated

The eigenvalues within the energy window [emin:emax] will be calculated and stored. This option also useful in dealing with huge system. The usage for this tag is as follows:

EWINDOW -5.0:5.0 NE\_MAX 10

If provided, the energy level between these energy window will be printed out in the bandstructure\_TBA.dat file.

In the above setting, the eigenvalue ( $\{e\}$ ) within the energy window [-5.0:5.0] will be calculated and stored. The NE\_MAX represents the maximum number of eigenvalue to be searched within the window and usually should be larger than the number of actual eigenvalues (NE) within the range and should not exceed the total number of eigenvalue (NE\_TOT) of the system. The optimal values for NE\_MAX is about  $1.5 \times NE^{12}$ . Since the NE\_MAX is critical to the calculation speed, choosing the optimal values is essential. During the calculation, the program will find the optimal NE\_MAX and update in every k-point loop. <sup>13</sup>

Note 1: If the tag is specified in your input file, the Hamiltonian matrix will be constructed with the sparse matrix format rather than dense matrix format. The libraries to dealing with the sparse matrix is referred from Intel Math Kernel Library (MKL), please make sure that your library path is properly assigned. (suggest to use MKL version  $\geq 11.3$ )

Note 2: If NE\_MAX is not provided or exceeding NE\_TOT, i.e., NE\_MAX > NE\_TOT, NE\_MAX will be set to NE\_TOT by default.

 $<sup>^{12}{\</sup>rm Eric}$  Polizzi, "Density-matrix-based algorithm for solving eigenvalue problem" Physical~Review~B79, 115112 (2009)

<sup>&</sup>lt;sup>13</sup>Though, one need to provide reasonable NE\_MAX to save the memory, since NE\_MAX is used to reserve memory space for the eigenvector store internally.

#### PRTSEPK logical Default: .FALSE.

If .TRUE., band structure file, band\_structure\_TBA.dat, will be separated for each k-point, i.e., band\_structure\_TBA.kp\_1.dat, band\_structure\_TBA.kp\_2.dat, ..., etc. This tag is useful if you are dealing with very large system and many k-points, where due to the memory problem, calculation get failed to be finished.

#### **PRTHAMK** logical Default: .FALSE.

If .TRUE., hamiltonian matrix  $H_k$  will be written for each k-point into separate file Matrix.Hk\_Kik\_SPis.dat, where ik and is represents k-point and spin index, respectively.

#### **LORDER** *logical* Default: .FALSE.

If .TRUE., band structure will be re-ordered by maximizing the overlap between neighboring k-points. The overlap can be defined by the inner product of the wavefunctions  $\langle u_{n,k}|u_{m,k-1}\rangle$ . The re-ordered band structure will be written in band\_structure\_ordered.dat file. If OV\_CUT tag is also provided with real value, the cutoff for the overlap integral to be considered with the value. The bands above OV\_CUT will be swapped. The default value is  $\frac{\sqrt{2}}{2}$ .

LORDER .TRUE. OV\_CUT 0.7

NOTE: (important) If the LORDER = .TRUE. with TBFIT = .TRUE., then the fitting will be done with re-ordered band structure. Therefore, re-ordered target energy information should be provided in parallel with the original energy in EFILE. Please check EFILE tag, for the instruction how to provide re-ordered energy in this case. To get re-ordered energy: for VASP, one can use VaspBandUnfolding tools for re-ordering: for QE, one can use bands.x postprocessing program. In this case, WEIGHT tag should be carefully adjusted according to the re-ordered band index.

#### **LPHASE** logical Default: .TRUE.

If .TRUE., we construct Bloch basis functions  $\chi_i^{\vec{k}}(\vec{r}) = \sum_{\vec{R}} e^{i\vec{k}\cdot(\vec{R}+\vec{t}_i)} \phi_i(\vec{r}-\vec{R})$ . If .FALSE., the phase factor  $e^{i\vec{k}\cdot\vec{t}_j}$  is not included in the definition of Bloch functions.

#### **LTOTEN** logical Default: .FALSE.

If .TRUE., calculate total energy  $E_{TOT} = E_{band}$ , where  $E_{band} = \sum_{n,k} f_{n,k} e_{n,k}$ . Here,  $f_{n,k}$  and  $e_{n,k}$  is the Fermi-Dirac occupation function and eigen value of n-th level of k-th  $\vec{k}$ -point, respectively.

To calculate Fermi level, one should provide number of electrons and electronic temperature by NELECT and ELTEMP tag.

#### **NELECT** real Default: not specified

Total number of electrons of the tight binding models. In the collinear calculations,

i.e., TYPMAG is *collinear*, both spin-up and spin-down components should be specified together as follows:

NELECT 10 10 # for spin-up and spin-dn

#### **ELTEMP** real Default: 0.001/ $k_B$ $\sim$ 11.6045 K

Electronic temperature T (K), which will be used in Fermi-Dirac distribution function  $f(T) = \frac{1}{1 + \exp{(e - \mu)/k_B T}}$  to obtain Fermi level, where e is energy level (in eV),  $\mu$  is Fermi level (in eV), and  $k_B$  is Boltzman constant.

#### **SET** string

Setting tags for post processing, parameter constraints, and nearest neighbor setups, etc. Available list for the SET tags are as follows,

**GA**: for Genetic Algorithm setting

CONSTRAINT TBPARAM

NN\_CLASS

**RIBBON** 

BERRY\_CURVATURE

ZAK\_PHASE

WCC

Z2\_INDEX

PARITY\_CHECK

**EFIELD** 

WEIGHT

DOS

**EIGPLOT** 

STMPLOT

**EFFECTIVE** 

REPLOT

#### 2. Details of the SET

Each SET tag should be ended up by END tag.

**GA** Setting of control parameters for the **Genetic Algorithm** used in parameter fitting procedures. This setting is only effective when **LSTYPE** is set to **GA**. Below you can check the default settings for GA procedures. You can modify as your purpose or comment out to use default setup as a input.

```
SET GA
 MGEN 100 # maximum number of iterations. (default:500)
 NPOP 100 # population in each generation. (default:100)
 NGENE 6 # number of genes in chromosomal encoding.
           # should be in between 2 to 9. (default:6)
 PCROSS 0.85 # crossover probability. [min:max]=[0.0:1.0]
 RMUTMIN 0.0005 # minimum mutation rate. [0.0:1.0]
 RMUTMAX 0.25
                # maximum mutation rate. [0.0:1.0]
 RMUTINI 0.005 # initial mutation rate. [0.0:1.0]
 MUT_MOD 2
                 # mutation with 1: fixed rage
                 # mutation
                                  with 2: fitness dependent
                 # mutation
                                  with 3: distance dependent
                 # mutation+creep with 4: fixed rate
                 # mutation+creep with 5: fitness dependent
                 # mutation+creep with 6: distance dependent
                # relative fitness differential [0.0:1.0]
 FDIF 1.0
  IREP 3
           # reproduction plan 1: Full generational replacement
           #
                               2: Steady-state-replace-random
                               3: Steady-state-replace-worst
             # elitism 0: off, 1: on
             # Note that this tag applies only if IREP=1 or 2.
 VERBOSE 1
             # printed output 0/1/2=None/Minimal/Verbose
  CONVTOL 0.0001 # convergence tolerance (must be > 0.0).
 CONVWIN 20 # convergence window.
             # If CONVWIN consecutive solutions are found
             # convergence will be declaired.
             # Hence, give larger convergence window to reach minima.
  IGUESSF 0.1 # fraction of the initial population to set equal
              #to the initial guess. [0.0:1.0]
  ISEED 999 # random seed value (must be > 0).
END GA
                   _ GA default setup example ____
```

**STMPLOT** Setting of integrated eigen state wavefunction  $\Sigma |\psi_{nk}(r)|^2$  plot. Here, the summation runs over the eigen states within the energy window specified by STM\_ERANGE or equivalently STM\_WINDOW.

```
SET STMPLOT
 NGRID 40 40 80 # GRID for CHGCAR-STM output (default = 0.1 ang).
 STM_ERANGE -1.0:0.0 # energy window
 RCUT 6.0 # cut off radius(Å). Beyond this will not be calculated.
 REPEAT_CELL T T T # repeat orbital for each lattice vector?
     # this logical tag is especially useful if you only
     #consider center region of the very large cell.
     # If set "T T F", orbital contribution which is periodically
     # repeated in a3 direction wll not be considered to calculate.
     # Try this option if you have very large cell and you are
     # especially interested unitcell ceter.
END STMPLOT
                     _ STMPLOT setup example ____
```

**EIGPLOT** Setting of eigen state wavefunction  $\psi_{nk}(r)$  or charge density  $|\psi_{nk}(r)|^2$  plot.

```
SET EIGPLOT
  IEIG 3 5 # index(es) n of eigen state.
 IKPT 1 10 # index(es) k of k-point.
 NGRID 40 40 80 # GRID for CHGCAR output (default = 0.1 ang).
 RORIGIN 0.0 0.0 0.0 # shift of the origin of the cube file.
 WAVEPLOT .TRUE. # plot wavefunction (.true.) or charge density.
 RCUT 6.0 # cut off radius(Å). Beyond this will not be calculated.
END EIGPLOT
                 ____ EIGPLOT setup example __
```

**DOS** Setting of Density of states (DOS).

```
SET DOS
 GKGRID 100 100 1 # set Gamma centered Monkhorst-Pack grid
 KSHIFT 0.0 0.0 0.0 # shift of k-grid (k-offset)
 PRINT_KPTS .TRUE. IBZKPT-DOS_TB # print k-point to the file
 PRINT_EIG .TRUE. 1:2 3 # print specified energy surface
 PRINT_UNIT RECIPROCAL # k-point unit (or ANGSTROM 1/A)
 SMEARING 0.03 # gaussian smearing. Default = 0.025
                # number of grid points in energy window (erange)
 NEDOS 2000
 DOS_EWINDOW -20.0:10.0 # energy window to be plotted
 DOS_NRANGE 1:NEIG # energy window to be calculated (integer)
 DOS_SPARSE .TRUE. # or .FALSE. use sparse matrix? Default=.FALSE.
 DOS_FNAME DOS_TB_projected.dat # output file name for DOS output
 PRINT_LDOS .TRUE. 1:8 12 # Print local density of states for given
                         # atoms. Here, 1 to 8-th atoms and 12-th
                         # atoms will be resolved.
 LDOS_FNAME LDOS_TB_projected # header for LDOS file name.
```

Note1: NEIG variable of the DOS\_NRANGE tag indicates total number of states, i.e.,  $N\_ORB \times ISPINOR$ , where  $N\_ORB$  is total number of atomic orbitals and ISPINOR = 1 (LSORB = .FALSE.) or 2 (LSORB = .TRUE.). If you want to reduce calculation loads, you can adjust DOS\_NRANGE.

Note2: DOS\_SPARSE tag is only available if -DMKL\_SPARSE option is activated in the makefile. If set to .TRUE., DOS\_NRANGE should be as following:

```
DOS_NRANGE 1:NE_MAX or DOS_NRANGE NE_MAX
```

Here, NE\_MAX is integer value larger than zero and less equal than total number of states NEIG. This setting will reduce the resources required for hamiltonian matrix construction and time consuming for the eigenvalue problem by the energy window constraint in the help of sparse matrix eigen solver. See EWINDOW for more informations.

Note3: PRINT\_EIG is only applicable if DOS\_SPARSE = .FALSE.

#### **EFIELD** Setting of E-field.

```
SET EFIELD

EFIELD 0.0 0.0 0.1 # Efield along z direction

EF_ORIGIN 0.0 0.0 0.345690593 # (in fractional coordinate)

#EF_CORIGIN 0 0 0 # (in cartesian coordinate)

END EFIELD

EFIELD setup example
```

**WEIGHT** Setting of weight factor for the fitting procedures.

KRANGE integer: range of k-point where the weight factor is applied TBABND integer: range of eigen states of the tight binding calculation DFTBND integer: range of eigen states of the target energy bands

WEIGHT real: weighting factor

ORBT\_I ineteger: orbital index.  $n^{th}$  orbital states will get a penalty

SITE\_I ineteger: site index. ORBT\_I<sup>th</sup> orbital state at SITE\_I atom will get a penalty. This prohibit certain orbital character to be stabilized from the fitting procedures.

```
SET WEIGHT
  IBAND = 1
               # alias for initial band index
  FBAND = 20
               # alias for final band index
               # alias for valence band maximum of TB band
        = 20
               # alias for conduction band minimum of TB band
  CBMT
  VBMD
        = 17
               # alias for conduction band minimum of DFT band
  CBMD
       = 20
               # alias for conduction band minimum of DFT band
  KRANGE
                   TBABND
                                       DFTBND IBAND: FBAND WEIGHT 1
  KRANGE
                   TBABND VBMT: CBMT
                                       DFTBND VBMD:CBMT
                                                           WEIGHT 6
  KRANGE 20:60 X:M TBABND VBMT-2:CBMT DFTBND VBMD-2:CBMT WEIGHT 20
  KRANGE 20:60
                   TBABND 2:4
                                       DFTBND 2:4
                                                           WEIGHT 50
  KRANGE 1
               TBABND
                             ORBT_I 1 SITE_I Mo1 PENALTY 200
                       7
END WEIGHT
                       WEIGHT setup example _
```

Note 1: After fitting procedure is over, the WEIGHT information is written in PFILE so that one can restart fitting. To use WEIGHT information written in PFILE, one can add USE\_WEIGHT tag in your PFILE tag. Please find this information in PFILE tag.

Note 2: The k-point name also can be used to specify KRANGE. Above example show that one can set weight factor by specifying X:M which indicates the range from X point to M point. The k-point name can be defined in KFILE. Note that following syntax also works: X-10:M+10 or X-10:M, etc.

Note 3: You can also define valence/conduction band minimum/maximum for DFT or TBA bands by providing VBMT, CBMT, VBMD, CBMD. Once it has been defined, you can use this acronym in weight define as above WEIGHT setup example. You can also add or subtract some amount you wish to this value when specifying range, for example, VBMD-2:VBMD+2, which will be applying 15:19 if VBMD = 17.

**CONSTRAINT TBPARAM** Setting for parameter constraints for the fitting and calculation. The value of the specified two parameter will be kept same during the fitting and tight-binding calculations. If you are using GA method for the fitting procedures (LSTYPE), you are encouraged to give upper bound and lower bound for each parameters to minimize parameter search field in the randomize procedures of GA method. The default lower/upper bound for every parameter is -20.0/20.0.

```
SET CONSTRAINT

e_py_S = e_px_S # e_py_S is enforced to be same as e_px_S.

e_px_S <= 5.0 # upper bound for e_px_S (applied in GA)

e_px_S >= -5.0 # lower bound for e_px_S (applied in GA)

END CONSTRAINT

CONSTRAINT setup example
```

If the second argument '=' is replaced by '==' and the third argument is not

present, then this parameter will not be fitted and its initial guess as defined in PFILE will be fixed during the fitting procedures. Note that, exactly same effect can be achieved by putting 'FIXED' tag at the parameter specification line of the PFILE, and the detailed explanation can be found in Fixing parameter of Sec.4.

Note: Some special rule can be applied in parameter. For example, assuming that s and  $p_x$  orbital are in atom A and B. Then, one can consider hopping parameter such as  $sps\_1\_AB$ . In this case, usually, hopping between orbital s in A and orbital  $p_x$  in B is differ from hopping between orbital  $p_x$  in A and orbital s in B since atom A and atom B is different species. [see section 4.Details of the format of PFILE] However, in many reason, sometimes it is useful to distinguish A and B although they are same species, for example, S atom of 2H-MoS2 monolayer. Here, one can specify S atom in upper layer to Mo layer as Sa and S atom in lower layer to Mo layer as Sb. In this case, to correctly define hopping between s and p orbitals in Sa and Sb, one should provide following hopping parameter separately:  $sps\_1\_SaSb$ ,  $pss\_1\_SaSb$ . It is bothering to write same values twice with just changing sp to ps. And it should be specified in the constraint section that  $pss\_1\_SaSb = sps\_1\_SaSb$ . It can be avoided if one provide constraint statement that atom Sb and atom Sa are actually same species as follows:

```
SET CONSTRAINT
....
atom_Sb = atom_Sb
....
END CONSTRAINT
CONSTRAINT setup example
```

Now, it is enough to provide  $sps_1\_SaSb$  only for the s and p orbital hopping parameters between Sa and Sb atom.

#### **NN\_CLASS** Setting for nearest neighbor set up.

If the distance between two atomic species (For example, Mo and S) are 1st nearest type, and its upper limit is 3.2 angstrom (e.g., below this value will be regarded as the pair), thene we can set as follows,

```
Mo-S: 3.2 RO 3.171634
```

Here, number of dash '-' occurance between two atomic species indicates the distance class n, and the above example represents 1st nearest hopping between Mo and S. The following R0 tag defines optimal bonding distance between two neighbor pair. This value will be used in calling the scaling function to get the distance dependent hopping parameter.

```
SET NN_CLASS

Mo-Mo : 3.2 R0 3.171634

S-S : 3.28 R0 3.171634

S--S : 3.2 R0 3.193724

Mo-S : 2.5 R0 2.429624

END NN_CLASS

NN_CLASS setup example
```

**RIBBON** Setting for nanoribbon calculations.

At the initial stages of the calculations, TBFIT will generate GFILE-ribbon with the settings bellow.

NSLAB integer: multiplication of unitcell along each direction

VACUUM real: vacuum spacing along each direction.

KFILE\_R real: KFILE for ribbon band structure. Default: KFILE

PRINT\_ONLY\_R *logical*: if .TRUE. the geometry file will be generated with -ribbon suffix to the GFILE and the program will imediatly stops. Default: .FALSE.

```
SET RIBBON

NSLAB 1 20 1

VACUUM 0 20 0

KFILE_R KPOINTS_RIBBON

PRINT_ONLY_R .FALSE. or . TRUE.

END RIBBON

Ribbon calculation setup
```

**Z2\_INDEX** Automatic calculations for topological index  $[\nu_0 \ \nu_1, \nu_2, \nu_3]$  for 3D or  $\mathbb{Z}_2$  for 2D via WCC method. The output will be written at Z2.WCC.plane\_index.dat and Z2.GAP.plane\_index.dat. Here, plane\_index indicates one of six  $B_i$ - $B_j$  plane with  $B_k = 0$  or  $\pi$ . For example, if plane\_index = 0.0-B3.B1\_B2-PLANE, then it contains WCC information of B1-B2 plane with  $k_z = \pi$ .

```
SET Z2_INDEX

Z2_ERANGE 1:28 # upto occupied

Z2_DIMENSION 3D # or 2D:kz (2D WCC plane perpendicular to kz)

Z2_NKDIV 21 21 # k-grid for KPATH and k-direction for WCC

Z2_CHERN .TRUE. # 1st Chern number of given bands with ERANGE

END Z2_INDEX

Z2 index calculation using WCC method
```

**WCC** Wannier Charge Center calculation settings

```
SET WCC
WCC_ERANGE 1:28 # upto occupied
```

```
WCC_FNAME WCC.OUT.dat

WCC_FNAME_GAP WCC.GAP.dat # largest gap will be written

WCC_KPATH 0 0 0 1 0 0 # k_init -> k_end (ex, along b1)

WCC_KPATH_SHIFT 0 0 0.5 # kpoint shift along b3 direction

WCC_DIREC 2 #k-direction for WCC evolution (1:b1,2:b2,3:b3)

WCC_NKDIV 21 21 # k-grid for KPATH and k-direction(odd number)

WCC_CHERN .TRUE. # 1st Chern number of given bands with ERANGE

END WCC

Wannier charge center (WCC) setup: kz 0.5 (shift)
```

#### **ZAK\_PHASE** Setting for Zak phase calculations.

```
SET ZAK_PHASE

ZAK_ERANGE 1:28 # upto occupied

ZAK_FNAME ZAK_PHASE.OUT.dat

ZAK_KPATH 0 0 0 1 0 0 # k_init -> k_end (ex, along b1)

ZAK_DIREC 2 #k-direction for Zak phase evolution (1:b1,2:b2,3:b3)

ZAK_NKDIV 21 21 # k-grid for KPATH and k-direction

END ZAK_PHASE

Zak phase setup
```

#### **BERRY\_CURVATURE** Setting for Berry curvature calculations.

```
SET BERRY_CURVATURE

BERRYC_METHOD KUBO # .or. RESTA(not yet supported)

BERRYC_ERANGE 17:18

BERRYC_FNAME BERRYCURV.17-18 # output will be BERRYC_FNAME.dat

BERRYC_DIMENSION 2D:B3 # 2D plane perpendicular to kz)

END BERRY_CURVATURE

Berrycurvature setup
```

#### **PARITY\_CHECK** Setting for Parity eigenvalue calculations for given k-points.

```
SET PARITY_CHECK

PARITY_KP 0.0 0.0 0.0 G  # Gamma (reciprocal unit)

PARITY_KP 0.5 0.0 0.0 M1 # M1 (reciprocal unit)

PARITY_KP 0.0 0.5 0.0 M2 # M2 (reciprocal unit)

PARITY_KP 0.5 0.5 0.0 M3 # M3 (reciprocal unit)

PARITY_ORIGIN 0.0 0.0 0.0 # origin of the system (direct coord)

PARITY_OP1 -1 0 0 # Rotation matrix (R) for inversion

PARITY_OP2 0 -1 0 # => R*X=-X (invert coordinate)

PARITY_OP3 0 0 -1 # => X:direct coord; R: integer 3x3 array

NOCC 10 # 10 occupied states

PRINT_HAMILTONIAN .FALSE. # print out H(k) for each k-point

END PARITY_CHECK

Parity check setup
```

#### Note:

- You can add (or remove) PARITY\_KP tag if you want to get the parity information for another TRIM (time reversal invariant momenta: -k=k+G) point.
- To use this functionality and to get the meaningful results, your system should have inversion symmetry.
- The ROTATION tag is optional, the default is

$$R = \begin{bmatrix} PARITY\_OP1 \\ PARITY\_OP2 \\ PARITY\_OP3 \end{bmatrix} = \begin{bmatrix} -1 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & -1 \end{bmatrix}$$

**SYMMETRY\_EIG** Setting for Symmetry eigenvalue calculations for given k-points.

```
SET SYMMETRY_EIG

SYMMETRY_KP 0.0 0.0 0.0 G # Gamma (reciprocal unit)

PARITY_KP 0.6666667 0.6666667 0.0 M3 # M3 (reciprocal unit)

SYMMETRY_ORIGIN 0.0 0.0 0.0 # origin of the system (direct coord)

SYMMETRY_OP1 -1 -1 0 # Rotation matrix (R) for 2/3pi

SYMMETRY_OP2 0 1 0 #

SYMMETRY_OP3 0 0 1 #

ROT_ANGLE 120 # rotation angle counterclockwise

NOCC 10 # 10 occupied states

PRINT_HAMILTONIAN .FALSE. # print out H(k) for each k-point

END PARITY_CHECK

Parity check setup
```

**EFFECTIVE** Setting for evaluating effective hamiltonian and its eigenvalues. This job will be performed by the Löwdin downfolding technique<sup>14</sup>. Note that in the current version, only the energy-dependent effective Hamiltonian will be constructed, where the energy is representing one's interested region.

```
SET EFFECTIVE HAM

EFF_ORB C:pz # downfolding will be performed on these orbitals

EFF_EWINDOW -2:2 # energy window of interest (e_center=-2+2/2)

END EFFECTIVE HAM

Effective Hamiltonian construction
```

**REPLOT** Setting for replotting where DOS/LDOS/PBAND represents density of states (DOS), local density of states (LDOS), and atom-projected band structure (PBAND) by reading band\_structure.dat or band\_structure.up(dn).dat file, respectively.

```
SET REPLOT

FILE_FORMAT bin # let code know what kind of
```

P.-O. Löwdin, J. Chem. Phys. 19, 1396 (1951)
 E. Zurek, O. Jepsen, O. K. Anderson, Chem. Phys. Chem. 6, 1934 (2005)

```
# band_structure_TBA file to be read.
                 # if bin : band_structure_TBA.(up/dn).bin
                 # if dat : band_structure_TBA.(up/dn).dat
                 # will be loaded. Default: dat
REPLOT_BAND .TRUE. wf # replot band structure
   # with wavefunction coefficient (with tag wf). Instead of
   # wf tag, one can also request rh, mx, my, mz, where the
   # meaning of each tag can be found in LORBIT tag.
   # If no additional tag is specified, set to rh by default.
   # If no tag is specified, only energy band without orbital
   # information will be printed out. The output file will be
   # written in band_structure_TBA.replot_wf.(up/dn).dat file
   # with ascii format.
   # Note that to activate this tag, FILE_FORMAT tag should be
   # set to bin and corresponding file band_structure_TBA.replot
   # .(up/dn).bin should be exist in the folder.
REPLOT_PROJ_BAND .TRUE. 1:8 12 # replot projected band structure
                               # for given atoms and their sum
REPLOT_DOS .TRUE. # Recalculate DOS using pre-calculated
                # band_structure file, e.g.,
                # band_structure.dat (nonmagnetic or
                # non-collinear) or band_structure.up/dn.dat
                # (spin polarized case).
REPLOT_LDOS .TRUE. 1:8 12 # Print local density of states for
                         # given atoms. For example, here,
                         # 1st to 8-th atoms and 12-th
                         # atoms will be resolved.
REPLOT_DIDV .TRUE. # Print local density of states for given
                   # discrete energy level within DOS_EWINDOW.
                   # The outputs will be stored in
                   # ./didv/DIDV.replot.dat.#i where #i=[1:NEDOS]
SMEARING 0.03 # gaussian smearing. Default = 0.025
         2000 # number of grid points in energy window (erange)
DOS_EWINDOW -20.0:10.0 # energy window to be plotted
REPLOT_SLDOS .TRUE. # spatial LDOS within EWINDOW
REPEAT_CELL 20 20 1 # if REPLOT_SLDOS = .TRUE.,
                     # cell periodicity for visualization
RORIGIN 0.0 0.0 0.0 # shift of origin of atomic coordinates
                     # (fractional to unit vector a1, a2, a3,
                     # respectively).
BOND_CUT 1.8 # bond length <= bond_cut will not be written in
              # BOND.replot.dat Default: 3.0 (ang)
```

DOS setup example —

Note 1: This tag is useful if you want to get series of DOS/LDOS calculations with respect to the SMEARING and energy window. If this SET is activated, other calculations will be ignored and the program stops immediately after the calculation.

Note 2: If REPLOT\_DOS = .TRUE. or REPLOT\_LDOS = .TRUE. or REPLOT\_SLDOS = .TRUE., then it is encouraged to set SMEARING, NEDOS, DOS\_EWINDOW, otherwise the default values will be used.

Note 3: If REPLOT\_SLDOS = .TRUE., then, REPEAT\_CELL, RORIGIN, BOND\_CUT should be set together, otherwise the default values will be used.

Note 4: The output file names are as follows:

- ullet if REPLOT\_DOS = .TRUE. ightarrow DOS.replot.dat
- if REPLOT\_LDOS = .TRUE.  $\rightarrow$  LDOS.replot.ATOM\_INDEX.dat, where ATOM\_INDEX is the atom number. The example can be found in the "Example/Graphene/DENSITY\_OF\_STATE/replot" of your example folder.
- if REPLOT\_DOS = .TRUE. → SLDOS.replot.dat and BOND.replot.dat. The example can be found in the "Graphene/QSH/NANORIBBON/zigzag\_ribbon/SPATIAL\_LDOS\_REPLOT" of your example folder.

Note 5: The multiple declaration of REPLOT\_BAND, REPLOT\_PROJ\_BAND and REPLOT\_LDOS tag result in multiple band\_structure\_TBA.replot\_??.dat,

band\_structure\_TBA\_atom.sum#.dat and LDOS.replot.sum#.dat files, respectively, where ?? represents one of {rh, mx, my, mz, wf} tag you specified and # represents the order of declaration of the tag. The concept of multiple declaration is same as in PROJ\_BAND tag.

Note 6: The purpose of REPLOT\_BAND tag is clear as it only read .bin file and write .dat. That is, for converting binary file format into human readable ascii format. PROJ\_BAND.

Note 7: If you add bin or bin4 tag at the end of REPLOT\_BAND tag, for example, REPLOT\_BAND .TRUE. rh bin (or bin4)

then the result will be written with unformatted format with single precision. This is only available if your target file is already written in bin (or bin4; single precission) (see the LORBIT tag for the details). Using this tag, ascii formatted file can be saved in unformatted file, e.g., (band\_structure\_TBA.dat  $\rightarrow$  band\_structure\_TBA.bin).

### 3. Details of the format of GFILE

#### **Atomic orbital setup** *string*

Hydrogen-like atomic orbital can be specified for the orbital basis. The possible

orbital basises are 15:

```
s px py pz dz2 dxy dx2 dxz dx2
```

```
0 0.0 0.0 s px py pz # s, px, py, and pz orbitals at ATOM_A
0 0.0 0.5 s px py pz # s, px, py, and pz orbitals at ATOM_B

setup of atomic orbital basis in GFILE
```

#### **Custumized atomic orbital setup** string

If someone does not want to use Slater-Koster type interatomic hopping parameter, customized atomic orbital can be defined instead. In this case, distance and hopping pair dependent parameterization should be properly defined in the PFILE

Warning: This is experimental feature and on the development stages (do not use).

```
0 0.0 0.0 cp1 # cp1 orbital at ATOM_1
0 0.0 0.5 cp1 # cp1 orbital at ATOM_2
setup of custumized atomic orbital name cp1
```

#### **MOMENT** tag real

Magnetic moment for Each atomic orbital can be assigned as follows,

collinear case: 0.0

noncollinear case: 0.0 0.0 0.0 [M  $\theta$   $\phi$ ]

```
0 0.0 0.0 px py pz moment 0 0 1 # spin-up for pz
0 0.0 0.5 px py pz moment 0 0 -1 # spin-dn for pz

usage of moment tag in GFILE with collinear magnetism
```

```
0 0.0 0.0 px py pz moment 0 0 0 0 0 1 0 0 # spin-up for pz 0 0.0 0.5 px py pz moment 0 0 0 0 0 0 0 0 # spin-dn for pz
```

oxdot usage of moment tag in GFILE with noncollinear magnetism oxdot

#### MOMENT.C tag real

Similar to MOMENT but in noncollinear case, the  $1^{st}$ ,  $2^{nd}$ , and  $3^{rd}$  value represents,  $m_x$ ,  $m_y$ , and  $m_z$ , respectively. Here, x, y, and z represents the cartesian axis.

<sup>&</sup>lt;sup>15</sup>Please note that current version does not support the f orbitals. However, we will include f in the future release of TBFIT. For the Slater-Koster tables of f orbitals, please see [K. Lendi, Phys. Rev. B 9, 2433 (1974)].

noncollinear case: 0.0 0.0 0.0 [ $M_x$   $M_y$   $M_z$ ]

```
0 0.0 0.0 px py pz moment.c 0 0 0 0 0 0 0 1 # spin-up for pz 0 0.0 0.5 px py pz moment.c 0 0 0 0 0 0 0 0 -1 # spin-dn for pz usage of moment.c tag in GFILE with noncollinear magnetism
```

#### 4. Details of the format of PFILE

#### **ONSITE** parameters real

Onsite prameters for each atomic orbital should have the prefix e\_ and joint with the name of the orbital. The suffix should be the atomic species where the orbital placed.

$$e_dx2_Mo -0.34$$

#### **HOPPING** parameters type real

The type of tight binding hopping parameter used in the calculations.

case 1.) Slater-Koster type parameter The syntax is as follows:

hopping-type will have one of following prefix:  $\{ss, sp, sd, pp, pd, dd\}$ , and one of following suffix:  $\{s, p, d\}$ , which implies  $\sigma$ -,  $\pi$ -, and  $\delta$ -type inteaction. nn-class specifies the distance class. See NN\_CLASS for the details.

AB specifies the two atomic species (A and B atoms) where the orbital hopping take place. For example, for the  $dd\delta$  Slater-Koster parameter involved with the hopping process between the  $d_{z2}$  orbital in Mo atom and  $d_{yz}$  orbital in Mo, and they are  $2^{nd}$  neighbor pair, then the parameter should be the following form:

To activate the use of Slater-Koster type parameterization scheme, one should put the tag (default setting) in your PFILE:

IMPORTANT! If atom A and B is different species, special care should be taken in naming parameter. For example, assume that each atom A and B has orbital s,  $p_x$ . Then one can consider  $\sigma$ -type first nearest neighbour hopping between s orbital of A and  $p_x$  orbital of B. In this case, such hopping parameter should be written as:  $sps_1\_AB$  or  $pss_1\_BA$ . On the other hand, if we consider the hopping between

 $p_x$  orbital of A and x orbital of B, the hopping parameter for this case should be written as:  $sps_1BA$  or  $pss_1AB$ .

```
Example s and p orbital with different atomic species: atom A orbital s; atom B orbital px => sps_1_AB or pss_1_BA atom A orbital px; atom B orbital s => pss_1_AB or sps_1_BA
```

For the overlap integral parameters, one can add o\_ prefix, for example,

```
o_ddd_2_MoMo -0.2
```

Note that to activate the use of overlap integral, i.e., assuming the use of non-orthogonal basis, one should put USE\_OVERLAP .TRUE. tag into your PFILE.

To activate the use of scaling functions, i.e., distance dependent rescaling of parameters, one should set corresponding scaling parameters, for example,

```
s_ddd_2_MoMo -0.2
```

to rescale the ddd\_2\_MoMo parameter with respect to the distance and the directional cosine between orbitals. Note again that to apply same rule for the overlap integral parameters, one should put os\_ prefix instead, for example,

#### case2.) User defined parameters

Warning: This is experimental feature and on the development stages. In this case, the customized atomic orbital is assumed and the following scheme should be applied:

```
hopping-type_nn-class_AB
```

Here, the basic structure is same as case1.), however, the syntax of hopping-type is slightly different. That is: the prefix should have cc since this indicates customized hopping parameters. For the suffix, one should put user defined letter that characterize the hopping. For example,

represents the hopping between  $2^nd$  neighbor Bi atoms with the 'a' type of rule which characterizes hopping pair. If you want to setup the rule, you have to write the conditions to the source code: get\_cc\_param.f90.

Note that this mode work with following tag in your PFILE:

IS\_SK .FALSE.

#### **LOCAL POTENTIAL: LOCAL.POT** parameters *real*

If you want to apply local potential to the particular atomic site or particular orbital, then you can simply turn on LOCCHG (.TRUE.) and write local.pot tag together with the amount of local potential to be applied for each atomic orbitals in the GFILE. Next, you have to provide proper scaling parameter  $(U_{onsite}^i)$  for the local potential, since the local potential is applied on your Hamiltonian as:  $e_{onsite}^{i} = e_{onsite}^i + e_{loc.pot}^i \times U_{onsite}^i$ , i.e., it modifies onsite energy  $e_{onsite}^i$  to  $e_{onsite}^i$ . Here,  $U_{onsite}^i$  should be defined in your PFILE so that the syntax is local\_U\_orbital-type\_atom-name. orbital-type is one of s, p, or d type of orbital and atom-name is the name of atomic species you want to apply the local potential.

```
0 0.0 0.0 px py pz local.pot 1 1 1 # positive loc.pot
0 0.0 0.5 px py pz local.pot -1 -1 -1 # negative loc.pot

example of local.pot tag in GFILE
```

```
local_U_p_S 1.0 example of local.pot parameter in PFILE _____
```

**SOC** parameters real

 $case1.) \ensuremath{??}$  .TRUE.

Every spin-orbit coupling parameters in Slater-Koster method should have the prefix with lambda\_ and proper orbital information  $p_{-}$ (as a joinder, for example p orbital) and species information  $_{-}$ S(as a suffix, for example Sulpur atom) to precisely indicating the atomic orbital where the SOC effect will be applied.

case2.) ?? .FALSE.

In the case of user defined hopping parameter (orbital prefix start with c, see Sec.3 for the details) has been defined in the GFILE, SOC can be considered by setting up the Rashba and in-plane spin-orbit interaction. For Rashba type SOC, the prefix lrashba should be joint with nearest neighbor class n and hopping pair as follows.

lrashba\_c\_2\_BiBi 0.2

Above setting represents, Rashba type spin-orbit coupling between the custum type orbitals with c-prefix of the atom Bi and Bi.

#### **Fixing parmeter**

If one want some parameters not to be fitted during the fitting procedures, one can fix those parameters by adding FIXED or F. For example, if you want lambda\_p\_S to be kept as its initial value, then, set this parameter as follows,

lrashba\_c\_2\_BiBi 0.2 FIXED

#### **Example of PFILE**

IS_SK (or USE	.FALSE. # assume orthogonal basis sets _SKPARAM) .TRUE. # Slater-Koster type parameters 1 # SK scaling factor mode
	-0.34636955 -0.70447045
	-0.70447045 -0.17913534
<b>–</b> <i>J</i> –	-0.17913534
- <u>-</u> -	-2.96500556 -1.47877518

```
e_py_S
                  -1.47877518
      e_s_S
                 -10.51138070
 dds_1_MoMo
                  -1.04598377
 ddp_1_MoMo
                  0.44731993
 ddd_1_MoMo
                   0.10237760
   pps_1_SS
                   0.62323972
   ppp_1_SS
                  0.03251328
  pds_1_MoS
                  -2.32384045
  pdp_1_MoS
                  0.97229680
   sss_1_SS
                  -0.57287106
   sps_1_SS
                  -0.33278732
   sss_2_SS
                  -0.45573348
   sps_2_SS
                  -0.21906117
  sds_1_MoS
                   2.66111706
lambda_d_Mo
                   0.08014531 Fixed
 lambda_p_S
                   0.07567002 Fixed
 \_ example of PFILE: <code>PARAM_FIT.dat</code> for <code>MoS_2 (<code>IS_SK</code> .TRUE.) \_</code>
```

```
USE_OVERLAP
                       # assume orthogonal basis sets
                .FALSE.
IS_SK (or USE_SKPARAM) .FALSE. # Slater-Koster type parameters
       e_cp1_Bi
                -0.09222821
     ccb_1_BiBi
                  0.01723235
     cca_2_BiBi
                  0.13290800
               -0.0
     ccy_3_BiBi
     ccx_4_BiBi -0.03544401
               -0.01119142
lrashba_c_1_BiBi
lrashba_c_3_BiBi
lrashba_c_4_BiBi
                 -0.00636364
```

# Part II. Examples

## 1. Graphene with s and p orbitals

#### 1.1. Parameter fitting

#### 1.1.1. Prepare main control file: INCAR\_TB

TBFIT reads INCAR\_TB in the initial stages of the calculations to set up basic control parameters. In addition, GFILE, KFILE, and PFILE is mandatory for the normal run, and it should be predefined in your INCAR\_TB. For the fitting procedures, TBFIT should be .TRUE.. Then, one has to choose the fitting algorithm LSTYPE among LMDIF and GA. In this example, we will take LMDIF as a fitting algorithm. For the GA method, please find the example in "Example/Graphene/BAND\_FIT/Step\_1.genetic\_algorithm/" of your example folder.

The users must pay some time to "SET WEIGHT" which is quite important in fitting procedures. If you want to fit more tightly than the other regions of your target band structure, you can put much higher value for that particular region. In this example, we have 16 atomic orbitals (2 atoms  $\times$  4 orbitals per atom  $\{s, p_x, p_y, p_z\} \times 2$  spinors) in total. The target band structure, which is calculated by VASP, also have 16 band structure. In the INCAR\_TB file below, we give 5 weights.

The first line,

```
KRANGE: TBABND: DFTBND 1:16 WEIGHT 1
```

indicates that the all the k-range (:), and all the tight binding band structure (:), and first sixteen bands (1:16) of DFT target band structure will be weighted with "1".

And the second line,

```
KRANGE: TBABND 1:8 DFTBND 1:8 WEIGHT 7
```

indicates that the all the k-range (:), and first eight tight binding bands (1:8) will be targeted to the first eight (1:8) of DFT band structure with weight of "7". Hence, the valence bands will be much more tightly fitted.

And the third line,

```
KRANGE 10:51 TBABND 5:10 DFTBND 5:10 WEIGHT 27
```

indicates that the particular k-range (10:51, around M-K rigion), and 5-th to 10-th tight binding bands (5:10) will be targeted to those of DFT band structure with much larger weight of "27". The bands near the Fermi level and near the Dirac point will be fitted tightly. For the other lines, the same rules can be applied.

Next, the basic control tags should be informed. For example, MITER, PTOL, FTOL, PFILE, POFILE, EFILE ..., etc.

And then, one should specify whether your system is one of followings: nonmagnetic, noncollinear, collinear, which can be defined by TYPMAG, together with LSORB which is for the spin-orbit coupling effect. If you have put some LOCAL.POT tag in your PFILE to consider some local potential, you should turn on LOCCHG to .TRUE., and add some relevant parameters (see LOCAL.POT for the details) into the PFILE to manage the strength of the local potential. The tag ?? is to make sure that the hopping integral evaluation will follow the Slater and Koster's rule<sup>16</sup>. Note that in the current version, ?? = .FALSE. is experimental and under developing for the ease of use, so recommend to leave it to .TRUE..

Now you have to make some rule for the nearest neighbor (nn) hoppings by setting up "SET NN\_CLASS" which defines interatomic hopping. Once you set nn pair, the necessary hopping parameter for those hopping will be automatically determined and TBFIT will trying to find those parameters in your PFILE. The detailed syntax for the naming of hopping parameters can be found in Section-4: PFILE-detail.

Finally, with "SET CONSTRAINT TBPARAM" tag, one can add some constraint rule between parameters or restrict the magnitude of the parameters by applying upper/lower bound to them. For example, here, we have assumed that every p orbitals shall have same onsite energy. So by setting "e\_px\_C = e\_pz\_C", during the calculations, e\_px\_C value will be kept same as e\_pz\_C.

#### **INCAR\_TB** : control tags

```
TBFIT
        .TRUE.
        LMDIF
LSTYPE
        300
MITER
PTOL
        0.00000001
FTOL
        0.00000001
GFILE
        POSCAR-TB
KFILE
        KPOINTS_BAND
PFILE
        PARAM_FIT.dat
EFILE
        DFT_BAND.dat
POFILE
        PARAM_FIT.new.dat
LOCCHG
        .TRUE.
TYPMAG
        noncollinear
LSORB
       .TRUE.
LORBIT .FALSE.
```

<sup>16</sup> J. C. Slater and G. F. Koster, "Simplified LCAO Method for the Periodic Potential Problem", Phys. Rev. 94, 1498 (1954)

```
SET WEIGHT
   KRANGE:
                    TBABND:
                                  DFTBND 1:16 WEIGHT 1
   KRANGE :
                    TBABND 1:8
                                  DFTBND 1:8
                                              WEIGHT 7
   KRANGE 10:51
                    TBABND 5:10
                                  DFTBND 5:10 WEIGHT 27
   KRANGE 20:40
                    TBABND 1:4
                                  DFTBND 1:4
                                              WEIGHT 50
   KRANGE 1:8 53:80 TBABND 3:6 DFTBND 3:6
                                              WEIGHT 10
END WEIGHT
SET NN_CLASS
   C-C : 1.8 RO 1.4145
END NN_CLASS
SET CONSTRAINT TBPARAM
   e_px_C = e_pz_C
   e_py_C
           = e_pz_C
    e_s_C >= -10
           >= -10
   e_px_C
           >= -10
   e_py_C
           >= -10
   e_pz_C
 sss_1_CC
           >= -10
 sps_1_CC
            >= -10
 pps_1_CC
           >= -10
           >= -10
 ppp_1_CC
           >= -10
lambda_p_C
    e_s_C <= 10
   e_px_C <= 10
   e_py_C <= 10
   e_pz_C
           <= 10
 sss_1_CC
           <= 10
           <= 10
 sps_1_CC
 pps_1_CC
            <= 10
 ppp_1_CC
            <= 10
lambda_p_C
            <= 10
END CONSTRAINT TBPARAM
           _ example of INCAR-TB for Graphene _____
```

#### 1.1.2. Prepare parameter file: PARAM\_FIT

#### PARAM\_FIT.dat : PFILE

```
IS_SK .TRUE. # default
# index
             param_name
                              parameter
                              -5.50250547
1
                  e_s_C
2
                 e_px_C
                               0.27589107
                 e_py_C
3
                               0.27589107
4
                 e_pz_C
                              0.27589107
5
               sss_1_CC
                               4.14407741
6
                               5.06877443
               sps_1_CC
7
               pps_1_CC
                              -4.33297737
8
               ppp_1_CC
                              2.54569565
9
             o_sss_1_CC
                               0.02586060
10
             o_sps_1_CC
                              -0.04261519
11
             o_pps_1_CC
                               0.05063120
12
             o_ppp_1_CC
                               0.01711425
13
                             0.00000000
           lambda_p_C
              — example of PFILE for Graphene —
```

#### **POSCAR-TB**: GFILE

```
# Graphene with honeycomb lattice
  1.00000000000000
 2.45
                        0.0000000000000000
                                              0.0
 1.2250000000000000
                        2.1217622392718746
                                              0.0
 0.000000000000000
                        0.000000000000000
                                             15.0
 С
 2
 Direct coordinate
 0.16666666667 0.16666666667
                                 0.0
                                       s px py
                                                   pz
-0.16666666667 -0.16666666667
                                 0.0
                                       s px py
                                                  pz
               _{-} example of {	t GFILE} for {	t Graphene} .
```

Here, we have set s and  $p(p_x, p_y, \text{ and } p_z)$  orbital basis for each carbon atom.

#### **KPOINTS\_BAND**: KFILE

```
0.00000000 0.000000000 0 G

0.00000000 0.000000000 0 G

0.00000000 0.000000000 0.5 A

example of KFILE for Graphene
```

Note that the file syntax is exactly same as  $\mathtt{KPOINTS}$  file of  $\mathtt{VASP}$  program. See  $\mathtt{here}$  for the details.

#### **DFT\_BAND.dat** : **EFILE**

For the details of the EFILE, please go to EFILE section and also please find the example file in "Example/Graphene/BAND\_FIT/Step\_2.lmdif\_method/" of your example folder.

# Index

CIRC_DICHROISM, 13	PSO_NP, 5 PSO_OPT, 5
EFILE, 6 EFILE_EF, 8 ELTEMP, 17 ERANGE, 15 EWINDOW, 15	SCISSOR, 14 SET, 17 DOS, 19 EIGPLOT, 19
FBAND, 14 FDIFF, 5 FIT_PLAIN, 6 FTOL, 5	IEIG, 19 IKPT, 19 NGRID, 19 RCUT, 19 RORIGIN, 19 WAVEPLOT, 19
GET_BAND, 4 GFILE, 8	GA, 18 CONVTOL, 18
IBAND, 14	CONVWIN, 18 FDIF, 18
K_UNIT, 6 KFILE, 8 KREDUCE, 10	IELITE, 18 IGUESSF, 18 IREP, 18
LOAD_HOP, 13 LOCCHG, 10 LORBIT, 11 LORDER, 16 LPHASE, 16 LSORB, 10 LSTYPE, 4 LTOTEN, 16	ISEED, 18 MGEN, 18 MUT_MOD, 18 NGENE, 18 NPOP, 18 PCROSS, 18 RMUTINI, 18 RMUTMAX, 18 RMUTMIN, 18
MITER, 4 MXFIT, 4	VERBOSE, 18 STM, 18
NELECT, 16 NN_MAX, 14 NONMAG, 10 NPAR, 6	NGRID, 19 RCUT, 19 REPEAT_CELL, 19 STM_ERANGE, 19 SPGLIB, 6
PFILE, 6 POFILE, 6 PROJ_BAND, 11 PRTHAMK, 16 PRTSEPK, 16	TBFIT, 4 TITLE, 4