The Defect Analysis Package manual

William D. Neilson* and Samuel T. Murphy*

Engineering Department, Lancaster University, Bailrigg, Lancaster, LA1 4YW, UK

E-mail: wdneilson@outlook.com; samuel.murphy@lancaster.ac.uk

Overview

The Defect Analysis Package (DefAP) exists to provide a detailed understanding of a materials defect chemistry. This document describes how to construct the input files necessary for the operation of DefAP and also details the output files that can be obtained. The input is highly customisable to the system being studied, with the operations that DefAP performs also tailored by the user.

Running DefAP

On command line: `python3 defap.py <filename>`, where `<filename>` is the name assigned to each input file.

Required libraries: NumPy, SciPy

Requires gnuplot (tested with gnuplot v5.2)

Requires a GV installation (only if using a Linux system and want to automatically view plots)

Input files

Summary

`filename.input` is the central input file of DefAP. It is the only input file required for all tasks that DefAP performs. There are a large number of parameters that can be used in this file, but depending on the system being studied (or tasks requested), many will not be required. Useful defaults also exist for many of these parameters that could mean the parameter does not need to be assigned in `filename.input`. `filename.defects` is the second core input file and details the defects being studied. It is only not required if DefAP is used to obtain the screened Madelung potential of a system. The additional input files are only necessary if particular DefAP capabilities are selected, as detailed in this manual. The flow diagram at the end of this manual is useful to help understand how DefAP uses the input files.

`filename.input`

Each line in this input file consists of a 'tag' or 'parameter' (i.e. a string), a colon sign ':', followed by the entered value. There are some exceptions to this rule, detailed in the list of optional parameters that follows:

Parameter	Explanation
`tasks`	Instruct what operation you want DefAP to perform. Can provide multiple tasks at once, simply separate with a space. Tasks: • `brouwer`: Calculate and plot defect concentrations as a function of the property defined with `loop`.

- `group`: Sum the defect concentrations of entries assigned the same 'group' in the `filename.defects` input file. Allows better presentation of plots.

- `energy`: Tabulate defect formation energies. Results recorded in `filename.output`.

- `form_plots`: Plot the formation energy of each defect as a function of the Fermi energy. A sub-folder (`filename_formation_plots`) will be created where several plots are made with different characteristics. *Requires* `energy` in tasks.

- `stability`: Perform checks on the thermodynamic stability of compounds not used in `constituents`.

- `madelung`: Calculate the Madelung potential for a periodic system with anisotropic dielectric properties. If additional `tasks` requested, the calculated Madelung potential is used to calculate a finite size error correction, applied to the defect formation energies.

- `autodisplay`: Automatically display defect concentration plot (`filename.eps`) at conclusion of tasks.

- `bibliography`: Write a bibliography, relevant to processes used, to `DefAP.bib`

`loop` Selection of property to find defect concentrations as a function of when `brouwer` selected as `task`:

- `0`: Volatile species partial pressure.
- `1`: Temperature.
- `2`: Dopant concentration.
- `3`: Artificial dopant concentration.

`min_value` Minimum volatile species partial pressure (10^(`min_value`) atm), temperature (K), or dopant concentration (10^(`min_value`) pfu), depending on loop chosen.

`max_value` Maximum volatile species partial pressure (10^(`max_value`) atm), temperature (K), or dopant concentration (10^(`max_value`) pfu), depending on loop chosen.

`iterator` Step taken between `min_value` and `max_value`.

`x_variable` Selection of x-axis in final defect concentration plots:

- `0`: Plot as function of the property defined with `loop` (default).
- `1`: Plot as a function of stoichiometry.

`Temperature` Desired temperature (K). Not used if `loop = 1`.

`Host` The system being studied. Elements separated by (-), number of atoms of each element separated by (_). Do not need to indicate number of atoms if just 1 atom.

`Host_energy` The DFT energy of one functional unit (eV).

`Host_supercell` The DFT energy of the defect-free supercell used to study the defects (eV).

`E_VBM` The energy of the valence band maximum (eV).

`chem_pot_method` Method used to obtain chemical potentials of each element in host.

- `Defined`: Each elements chemical potential user-defined.
- `Rich-Poor`: Study conditions containing an excess or deficit of elements.
- `Volatile`: Study binary compounds containing an element that is in the gas phase in its standard state. O, N, H, Cl and F are the current options available.
- `Volatile-Rich-Poor`: Study non-binary volatile compounds containing an excess or deficit of cation elements.

`real_gas` Choice of method to extrapolate chemical potential of volatile species as a function of temperature if `chem_pot_method` = `Volatile` or `Volatile-Rich-Poor`:

- `0`: Use ideal gas relations (default).
- `1`: Use real gas relations from the NIST Chemistry WebBook.
- `2`: Use real gas relations for oxygen from Johnston *et al.*[1]

`Constituents` Here provide table of information on lines immediately succeeding `Constituents`, dependent on `chem_pot_method` selected:

- `chem_pot_method` = `Defined`:

  Rows 1+: a b

  a : Element name.

  b : DFT energy of element (eV).

- `chem_pot_method` = `Rich-Poor`:

  Rows 1+: a b c

  c : Element fraction.

- `chem_pot_method` = `Volatile`:

  Row 1: d e

  Row 2: f g h i

  d : The volatile element in the host.

  e : The fixed volatile partial pressure (10^(`value`) atm). Place any number here if both `brouwer` is selected as a `task`, and `loop = 0` selected.

  f : The binary host compound. Entered in same format as `host`.

  g : The DFT energy of one functional unit (eV).

  h : The DFT energy of one cation atom (eV).

  i : Formation energy of binary compound under standard conditions, taken from literature (eV).

- `chem_pot_method` = `Volatile-Rich-Poor`:

  Row 1: d e j

  Rows 2+: k l g h i m

  j : Coefficient for volatile element in a balanced equation that makes host.

  k : A constituent binary compound of host. Entered in same format as `host`.

  l : Coefficient for this binary compound in a balanced equation that makes host.

  m : Fraction assigned to cation species in the binary compound.

`Electron_method` One of five options to calculate the concentrations of electrons must be selected:

- `Off`: If no charged defects in the system, select this option.
- `Fixed`: User specified concentration. Chosen concentration specified after `Fixed`, separated by a space.
- `Boltzmann`: Use Boltzmann statistics. Requires `Conductionband`.
- `Fermi-Dirac`: Use Fermi-dirac statistics. Requires `Conduction_band_limits` and `fu_unit_cell`. Also requires additional input file, `filename.dos`.
- `Effective_masses`: Density of states effective mass of electrons for chosen semiconductor is entered after `Effective_masses`, separated by a space. Alternatively, supply effective mass temperature dependence as a list, i.e. [[T_1,EM_1],[T_2,EM_2],..]. Requires `Volume_unit_cell` and `fu_unit_cell`.

`Hole_method` One of five options to calculate the concentrations of holes must be selected:

- `Off`: If no charged defects in the system, select this option.
- `Fixed`: User specified concentration. Chosen concentration specified after `Fixed`, separated by a space.
- `Boltzmann`: Use Boltzmann statistics. Requires `Valenceband`.
- `Fermi-Dirac`: Use Fermi-dirac statistics. Requires `Valence_band_limits` and `fu_unit_cell`. Also requires additional input file, `filename.dos`.
- `Effective_masses`: Density of states effective mass of holes for chosen semiconductor is entered after `Effective_masses`, separated by a space. Alternatively, supply effective mass temperature dependence as a list, i.e. [[T_1,EM_1],[T_2,EM_2],..]. Requires `Volume_unit_cell` and `fu_unit_cell`.

`Conductionband` The integral of the conduction band from the electronic density of states (must be in terms of per formula unit). *Only needed if `Electron_method = Boltzmann`.

`Valenceband` The integral of the valence band from the electronic density of states (must be in terms of per formula unit). *Only needed if `Hole_method = Boltzmann`.

`Conduction_band_limits` The limits of the conduction band in the electronic density of states supplied in `filename.dos`. Requires two values: The *lower-limit* followed by the *upper-limit*. *Only needed if `Electron_method = Fermi-Dirac`

`Valence_band_limits` The limits of the valence band in the electronic density of states supplied in `filename.dos`. Requires two values: The *lower-limit* followed by the *upper-limit*. *Only needed if `Hole_method = Fermi-Dirac`

`fu_unit_cell` The number of functional units in the unit cell. *Only needed if `Hole_method` and/or `Electron_method = Fermi-Dirac` or `Effective_masses`. Also required if `y_axis = 1`. **If Fermi-Dirac method used, this value should be the number of functional units in the density of states supplied in `filename.dos` (usually a unit cell is used to calculate the density of states).

`Volume_unit_cell` The volume of the unit cell ($Å^3$) *Only needed if `Hole_method` and/or `Electron_method = Effective_masses`. Also required if `y_axis = 1`.

`Bandgap` The electronic bandgap (eV).

`Stoichiometry` Selection of method to calculate stoichiometry in a host with a volatile element. Choice between `1` and `2` only relevant if dopants in the system.

- `0`: Don't calculate (default).
- `1`: Calculates stoichiometry with original cations, considers the cation/volatile species leaving the system in a substitution, but not the dopant added.
- `2`: Calculates a volatile to metal ratio, where any dopant added is treated as a metal.

`Defect_conc_method` Method to calculate the concentration of defects. Two options:

- `Boltzmann`: Boltzmann statistics applied (default).
- `Kasamatsu`: Kasamatsu statistics applied.

`Tab_correction` Add an externally calculated correction to the defect formation energies. The correction applied to each defect is requested in the input file `filename.defects`.

- `0`: Don't apply (default).
- `1`: Apply

`y_axis` Selection of y-axis units in final defect concentration plots:

- `0`: Units of concentration per host functional unit (default).
- `1`: Units of concentration per cm^{-3}. Requires `fu_unit_cell` and `Volume_unit_cell`.

`min_y_range` The minimum defect concentration plotted on the completed diagram when `brouwer` selected as `task`. Entered as logarithmic value. Default = -20: defect concentrations exceeding 10^{-20} per functional unit plotted.

`max_y_range` The maximum defect concentration plotted on the completed diagram when `brouwer` selected as `task`. Entered as logarithmic value. Default dependent upon selection of `y_axis`. If `y_axis = 0`, default = 0: defect concentrations less than 1 per functional unit plotted. If `y_axis = 1`, default = 20: defect concentrations less than 10^{20} per cm^{-3} plotted.

`Dopant_table` Provide the number of dopant elements added to the system (λ). Next, on the lines immediately succeeding `Dopant_table`, provide a row of information for each dopant added:

`Dopant_table` : λ

Rows 1+: a b c d e f

- a : Element name of dopant.
- b : A reference binary compound used to calculate the dopant chemical potential. Entered in same format as `host`. The reference can also be the dopant element itself, if preferred.
- c : The DFT energy of the chosen reference (eV pfu).
- d : Dopant concentration fitting option:
  - `0`: Don't fit.
  - `1`: Fit dopant concentration. Target given by e.
  - `2`: Fit dopant concentration. Target concentration varied; used if `loop = 2`.
  - DefAP uses Linear bisection to fit the concentration of one dopant, or Sequential Least Squares Programming (SLSQP) if fitting concentrations of multiple dopants is requested.
- e : Dopant target concentration (pfu), if fitting. Only required if $d =$ `1` or `2`. Enter any value here if `loop = 2`.
- f : Dopant range (eV): If fitting the dopant concentration, this defines the maximum allowable deviation in the dopant chemical potential. Only required if $d =$ `1` or `2`.

`Art_Dopant_Conc` The desired concentration (pfu) of an artificial defect.

`Art_Dopant_Chg` The charge on the artificial dopant concentration defined with `Art_Dopant_Conc`.

`Tolerance` If using SLSQP, this is the precision goal for the value of the function in the stopping criterion. Increase value to speed up optimisation, decrease if optimisation is not succeeding. (default = 1×10^{-10})

`max_iteration` If using SLSQP, this is the maximum number of iterations allowed to search for the dopant chemical potential. Decrease value to speed up optimisation, increase if optimisation is not succeeding. (default = 100)

`SLSQP_dial` Controls the the movement step in the dopant chemical potentials when the SLSQP optimiser relaunches - a relaunch is triggered when SLSQP exits without solution. Increasing value increases success rate at the expense of speed. (default = 10)

`Potential_convergence` The convergence criteria for the dopant concentration, if fitting: The difference between the logarithmic target concentration and the calculated concentration. (default = 1×10^{-3})

`Charge_convergence` The stopping criteria for the calculation of the Fermi level. Fermi level deemed satisfactory when the total charge does not exceed `Charge_convergence`. Increase to speed up calculation, may need to decrease in conditions of low defect concentrations. (default = 1×10^{-10})

`Stability_check` If `stability` selected as a `task`, the necessary information to perform checks on the compounds of interest are entered here. First, the number of compounds to be checked is to be entered (λ). Next, on the lines immediately succeeding `Stability_check`, provide a row of information for each compound studied:

`Stability_check` : λ

Rows 1+: a b

a : Compound of interest. Entered in same format as `host`.

b : DFT energy of the compound (eV pfu)

`Scheme` Selection of colour scheme for plots produced by DefAP:

- `0`: DefAP colour scheme (default).
- `1`: User customised scheme. Requires the input file, `filename.plot`.

`entropy` Indicates whether a vibrational entropy contribution is added:

- `0`: Do not add (default).
- `1`: Add contribution. Requires the input file, `filename.entropy`.

`entropy_units` The number of functional units the entropy values that are entered in `filename.entropy` represent (default = 1).

`Coulombic_correction` Automatically apply a point charge correction.

- `0`: Don't apply (default).
- `1`: Assume system is cubic, requires `Length` and `Dielectric_constant`.
- `2`: An anisotropic point charge correction, requiring `Screened_Madelung`.

`Dielectric_constant` The dielectric constant for the host system. *Only needed if `Coulombic_correction = 1`.

`Screened_Madelung` The screened Madelung potential. *Only needed if `Coulombic_correction = 2`. Can alternatively be calculated with `task = madelung`.

`Length` The length of supercell used (Å). *Only needed if `Coulombic_correction = 1`.

`Lattice` If `madelung` selected as a `task`, the lattice parameters are entered here. On the three lines succeeding `Lattice`, the three lattice vectors defining the unit cell of the system are given. i.e.:

`Lattice`

a_{11} a_{12} a_{13}

a_{21} a_{22} a_{23}

a_{31} a_{22} a_{33}

`Dielectric`	If `madelung` selected as a `task`, the dielectric constant tensor is entered here. On the three lines succeeding `Dielectric`, the tensor is given. i.e.: `Dielectric` ϵ_{11} ϵ_{12} ϵ_{13} ϵ_{21} ϵ_{22} ϵ_{23} ϵ_{31} ϵ_{22} ϵ_{33}
`Cutoff`	Specifies the cutoff for Madelung potential calculation if `madelung` selected as a `task`. (default = 30).
`Gamma`	Convergence parameter for Madelung potential calculation if `madelung` selected as a `task`. (default = 0.5).

`filename.defects`

In this input file, information about the defects being investigated is deposited. Each defect has its own row; the information required in each column is as follows:

Column 1 : The name of the defect.

Column 2 : The 'group' of the defect. Defects of the same 'group' should be listed adjacent to each other in this file.

Column 3 : The multiplicity of the defect.

Column 4 : The site of the defect, all defects which compete for the same site are given the same (magnitude insignificant) number, only necessary if using Kasamatsu statistics.

Column 5 : The charge of the defect.

Column 6 : The total DFT energy of the defect containing supercell.

Column 7 : Any correction you would like to add to the energy in Column 6. To apply, need to set `Tab_correction = 1` in `filename.input`. A number must be entered here

in all cases, but will be disregarded unless `Tab_correction = 1`.

Column 8+ : The final columns detail the elements that are added/removed to construct each defect. Each column is assigned its own element, corresponding to the number of unique elements in `host`. The order of the columns also corresponds to the order of the elements in `host`. Adding atoms requires a negative number, removal is a positive number. Further columns are added if studying dopants: the number and order of columns corresponds to the number and order of the rows succeeding `dopant_table`.

`filename.dos`

This file contains the coordinates of the electronic density of states for the system, required if `Fermi-Dirac` is requested to calculate the electron and hole concentrations. The first column contains the x coordinate (energy, eV) and the second column is the corresponding density of states at that energy. The valence band maximum must be positioned 0.0 eV. In `filename.input`, `fu_unit_cell` should be specified, which states the number of functional units your density of states represents.

`filename.plot`

This is a file used to alter the colours that DefAP assigns to different defects in the plots produced. Colours are assigned based on the 'group' a defect is assigned in `filename.defects`; all defects of the same group are assigned the same colour. Each defect in the same 'group' is then assigned a unique line style, with 9 unique line styles possible. The colour scheme that DefAP uses by default contains 14 unique colours, two of which are reserved for electrons and holes (if `brouwer` selected as a `task`). If the number of 'groups' exceeds the number of unique colours, the 'group' is assigned the colour black. `filename.plot` enables customisation of the colour assigned to each group, with no limit on the amount of 'groups' able to

have a unique colour. For the file to be read, `Scheme` must be set to `1` in `filename.input`. `filename.plot` contains four input parameters:

Parameter	Explanation
`concentration_colour`	A space separated list of desired colours applied to the `task = brouwer` plot. The order of the list corresponds to the order of 'groups' in `filename.defects`. The colours listed should be a hexadecimal color code, or a colour name recognised by gnuplot.
`electron_colour`	The name of the desired colour applied to the electron concentration in the `task = brouwer` plot. The colour listed should be a hexadecimal color code, or a colour name recognised by gnuplot.
`hole_colour`	The name of the desired colour applied to the hole concentration in the `task = brouwer` plot. Same rules apply as `electron_colour`.
`formation_colour`	A space separated list of desired colours applied to the `task = form_plots` plot. Same rules apply as `concentration_colour`.

`filename.entropy`

This optional input file is used if it is sought to include the impact of vibrational entropy when calculating the formation energy of each defect. To include the contribution, `entropy` should be set to `1` in `filename.input`. The externally calculated vibrational entropy values are compiled in `filename.entropy` with units: eV/K. If the values entered do not represent the entropy per functional unit, the parameter `entropy_units` in `filename.input` should be given the value of the number of functional units that the entered entropy values represent. The input file, `filename.entropy`, should be constructed as follows:

T	host	constituent	..	defect	..
T_1	$S_{T_1}^{host}$	$S_{T_1}^{constituent}$	..	$S_{T_1}^{defect}$	..
T_2	$S_{T_2}^{host}$	$S_{T_2}^{constituent}$	..	$S_{T_2}^{defect}$	..
:					

The first row is a space-separated list of headings. The heading names of the first two columns are fixed, the remaining must be changed. Beginning at the third column, a column is required for each constituent entered at `constituents`, with the exception of the volatile constituent. The order of the constituent columns in `filename.entropy` should correspond to the order of the rows entered at `constituents`. On row 1, enter the element/compound name, using the exact same format the name is entered in `constituents`. The input of constituent entropy's only applies if `chem_pot_method` = `Volatile-Rich-Poor`, otherwise skip to defects after 'host'. After all constituents have been entered, a column is required for each defect in `filename.defects`. On row 1, enter the name of the defect using the same spelling, format and order of the defect names in `filename.defects`.

Each row succeeding row 1 is where the entropy values are entered at a temperature entered in the first column (K). Temperatures should increase per row added, $T_1 < T_2$. For example, $S_{T_1}^{host}$ is the vibrational entropy of the host compound at a temperature, T_1. DefAP uses interpolation to obtain the vibrational entropy values at the temperature it requires. Testing shows that a minimum of four temperatures (4 rows) are required and the temperatures studied must fall within the range of temperature entered in `filename.entropy`.

Output files

`filename.output`

This file acts as a record of the tasks and processes DefAP is asked to perform. The parameters used to perform the calculations are catalogued, allowing the user to check that the

input files have been correctly set up. Additionally:

- If `energy` selected as a `task`, the calculated defect formation energies are printed here.
- If `brouwer` selected as a `task`, the calculated defect concentrations, chemical potentials for each element, and Fermi level are tabulated at each `increment`.
- If `stability` selected as a `task`, the results of the stability check are tabulated in this file. When `energy` is selected as the `task`, the stability check is performed at the conditions that the defect formation energies are calculated, with the results found after the defect formation table. When `brouwer` is selected as the `task`, the stability check is performed and printed at each `increment`. If DefAP finds any conditions of instability, a message will also be displayed in the main terminal.
- If `madelung` selected as a `task`, the calculated Madelung potential and calculation details are printed here.

`filename.res`

This is the main output data file when `brouwer` selected as a `task`, containing the log of the defect concentrations. The first column is the independent variable requested (volatile partial pressure, temperature, stoichiometry, etc.), the second is the final Fermi level at each condition. The third and fourth columns are the log of the electron and hole concentrations, respectively. After this, each column corresponds a defect contained within `filename.defects`. The order in which the defects appear in `filename.res` is the same as the order they appear in `filename.defects`. If `group` selected as a `task`, the sum of defects assigned the same group will instead be listed. The final column contains a calculated value for the stoichiometry (if `stoichiometry` $\neq$ `0`) of the host, unless plotting as a function of stoichiometry (`x_variable = 1`) in which case this column will be the independent variable selected.

`filename.p` and `filename.eps`

In order to allow easy visualisation of the defect concentrations, DefAP prepares a gnuplot input script (`filename.p`) based on the defect names given in the `filename.defects` input file. This is then automatically run through gnuplot to create an .eps file containing a Brouwer diagram showing the defect concentrations (`filename.eps`). If `autodisplay` is selected as a `task`, this will automatically be displayed on the screen when the calculation is complete (Runs the command 'gv filename.eps' or 'open filename.eps' if Linux or Mac OS system detected, respectively.). The values on the x axis of the plot are determined by `min_value` and `max_value` from the `filename.input` file, whilst the minimum concentration can be changed by altering `min_y_range` in the `filename.input` file. The plot can be further customised by modifying the `filename.p` file or using the data in `filename.res` to create your own plots. `filename.p` also contains the gnuplot instructions for the plotting of `filename_fermi.eps`

`filename.fermi` and `filename_fermi.eps`

`filename.fermi` is produced if studying a charged system and `brouwer` is selected as a `task`. The file contains two columns; the first is the independent variable selected (volatile partial pressure, temperature, etc.) and the second is the final Fermi energy calculated at this condition. `filename_fermi.eps` is the resultant plot.

`filename_formation_plots`

This is a directory, created if `form_plots` selected as a `task`, and holds the both the raw data and a collection of plots that show the variation in the defect formation energies with the Fermi level. Two data files exist in this directory:

- `filename.formation` : Contains the calculated formation energy of *every* defect in `filename.defects` as a function of the Fermi level.

- `filename.formation_grouped` : Contains the calculated formation energy of the defect with the lowest formation energy for each 'group', assigned in `filename.defects`, as a function of the Fermi level.

The directory contains a gnuplot script, `formation_plot.p`, that creates multiple plots using the two data files. A figure is created showing every defect (`formation.eps`) as well as a figure showing just the defect with the lowest formation energy for each 'group' (`formation_minimum.eps`). Two figures are also produced for each unique 'group' assigned in `filename.defects`. This allows for the study of an individual defect class: One plot shows every defect, the second displays only the defect with the lowest formation energy for the particular 'group'. The plots are clearly labeled.

References

(1) Johnston, K.; Castell, M. R.; Paxton, A. T.; Finnis, M. W. $SrTiO_3$ (001) (2×1) reconstructions: First-principles calculations of surface energy and atomic structure compared with scanning tunneling microscopy images. *Physical Review B* **2004**, *70*, 085415.

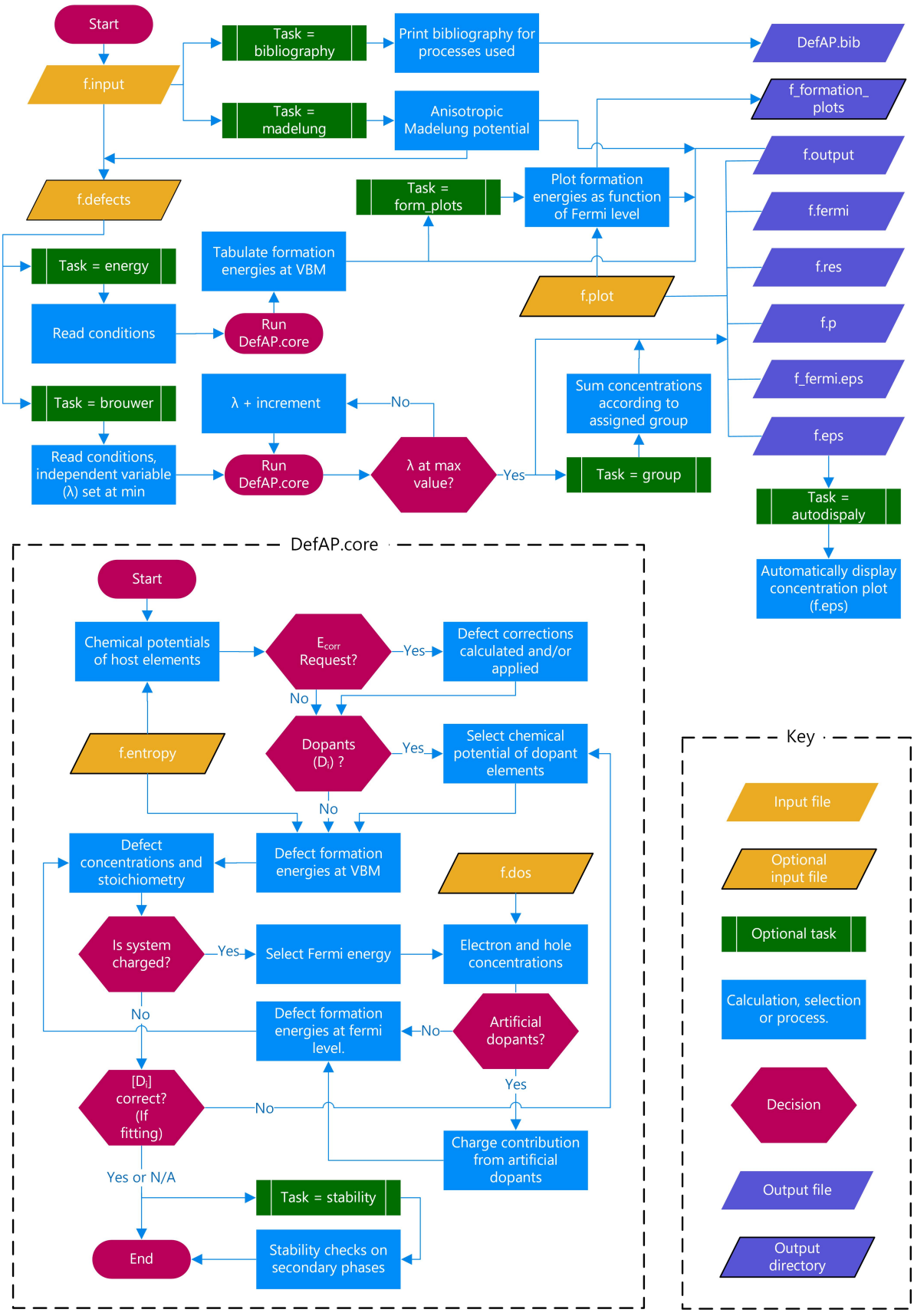

Figure 1: DefAP flow diagram