GLOBUS

Global and **L**ocal **O**ptimization of **B**ig **U**nordered **S**tructures

A program for the structure optimization, global minima search,

and potential parameter fitting

Version 1.5

User Manual

Nizhny Novgorod, 2018

Author, citation and copyrights

Prof. Stanislav K. Ignatov,

Department of Chemistry, Lobachevsky State University of Nizhny Novgrod, 23 Gagarin Avenue, 603950, Russia, *skignatov@gmail.com*

This software is released under the BSD License cited below and freely available for scientific or educational purposes. All publications describing work using this software should quote the following reference(s):

1. S.K. Ignatov, A.G. Razuvaev, A.E. Masunov. *Global structure optimization and properties of Ptn clusters (n=3-150) based on the DFT and DFT-calibrated empirical potentials and ReaxFF force field,* Book of Abstracts "16-th V.A. Fock meeting on Quantum, Theoretical and Computational Chemistry", Sochi 1-5.10.2018, p.10.

The GLOBUS code contains L-BFGS subprogram by Jorge Nocedal freely distributed under BSD license (Copyright(c) 1990, Jorge Nocedal), see http://users.iems.northwestern.edu/~nocedal/lbfgs.html Thus, all publications describing work using the GLOBUS software should also quote the L-BFGS reference:

2. D.C. Liu, J. Nocedal. *On the Limited Memory Method for Large Scale Optimization* (1989), Mathematical Programming B, 45, 3, pp. 503-528.

Program Features

GLOBUS is a computer program performing the structure optimization, global minimum search and potential parameters fitting for atomic structures including molecules, atomic and molecular clusters, and periodic systems with 1D, 2D, and 3D periodic boundary conditions.

- Global minimum search for geometry optimization and potential parameter fitting:
  - Artificial Bee Colony algorithm
  - Metropolis Monte Carlo Simulated Annealing
- Local optimization algorithms for geometry optimization and potential parameter fitting:
  - L-BFGS for unconstrained local optimization
  - SQP method for local optimization with equality and non-equality constraints
  - Conjugated gradients (CG) for unconstrained optimization
- Empirical potentials and force fields for energy calculations
  - LJ
  - Pairwise n-m potentials
  - Classical metallic potential of Gupta
  - Classical metallic potential of Sutton-Chen
  - n-component short-range Gaussian potential correction which can be applied to other potentials
  - Angular correction potential which can be applied to other potentials
  - ReaxFF force field
- Calculations with 1D, 2D, 3D periodic boundary conditions or without them
- Coordinates for the structure descriptions
  - Cartesian coordinates
  - Gaussian Z-matrix with or without variables and constants
  - Fractional coordinates and supercell description for periodic calculations
- Calculations of vibrational frequencies and normal modes
  - Can be performed automatically for all located LM
- IRC relaxation to closest LM in the case of imaginary frequencies detected
  - Can be performed automatically for all optimized structures
- "Cluster growth" mode
  - Automatic structure optimization of atomic clusters in the selected range of nuclearities
- Saving of all or selected LMs found
- Potential parameters fitting by experimental or DFT calculation data against
  - Cluster morphology comparison (algorithm of [Chevrot et al., JCP, 2011, 132, 084110])
  - Selected geometry parameters
  - Energy
  - Vibrational frequencies
  - Potential curve
- Ability for ReaxFF parameters fitting
- Parallelization
  - MPI for ABC algorithm in geometry and potential parameter optimization modes
- Interfaces to external server programs
  - LAMMPS (working in optimization mode with or without periodic boundary conditions)

Input file description

Common designations and abbreviations used below

Designation	Meaning
P/NP	Present or Not Present
Y/N	Yes or No
|	excluding OR (either… or…)
{…}	in the range of … values
[*parameters*]	not obligatory parameters
=*s*	string options (or list of values) should be provided
=*n*	integer option should be provided
=*x*	real options should be provided
$\begin{array}{|c|}\hline A \\ B \\ \hline\end{array}$	two or several mutually excluding options (either A or B should be provided)

Abbreviations used for the allowed values of keywords and their options

Abbreviation	Allowed values	Ranges	Example	Fortran variables
N	Positive integer	$[1, \infty)$	1,2,3,45	Integer
Z	Integer	$(-\infty, \infty)$	-5, -3, 0, 1, 12	Integer
Z_+	Nonnegative integer	$[0, \infty)$	0, 1, 2, 3,…	Integer
Q	Rational (written as decimal value)	$(-\infty, \infty)$	-23.65, 1.234, 0., 2	Real
Q_+	Nonnegative rational	$(0, \infty)$	1.234, 2.657, 2.,4	Real
{*a,b,c…*}	Set of allowed values	*(various)*	{-1, 0, 2, 5}	*(various)*

Structure of the input file (*.inp)

Each input file of GLOBUS is an ASCII text file containing several blocks which describe: (1) global commands and options; (2) potential type(s), their parameter values and their optimization options; (3) one or several groups of commands and options applied to the current molecular structure along with the initial geometry of this structure. All these blocks can be augmented by comment strings and set as follows:

! Comment strings marked by ! can be placed at any place of the *.inp-file
One or more lines marked by #. These lines contain global commands and options applied to all
    structures. The number of # strings is not limited.

* Block of 1 or 5 lines describing interatomic potential. First line of each block is marked with *.
    The exact syntax for each line describing the potential is described below (Section 4).
* …(*There can be up to MaxPot blocks describing potentials. MaxPot is set at compilation time* [default 100]*)*

@ Up to 10 lines marked by @ describes the 1ˢᵗ molecular structure
    Structure Geometry describing up to *MaxAt* atoms. (*MaxAt is set at compilation time* [default 1500]*)*
    Blank line – end of geometry

@ Up to 10 lines marked by @ describes the 2ⁿᵈ molecular structure
    Structure Geometry
    Blank line – end of geometry

@ … (*up to MaxStr structures. MaxStr is set at compilation time*[default 100]*)*

Keyword descriptions of the input files

1. Block marks

Keyword	Description
!	Comment string (no actions, commenting only)
#	Global command string (acts on all structures)
*	Potential description (acts on all structures)
@	Structure commands and structure description:

- These commands act only on the structure described under the current @-line (or several @-lines) before first blank string encountered (the only exception – blank string replacing *Variables* command in a Z-matrix).
- All the successive @-strings (not separated by blank lines) will act on the same structure.
- Up to 10 @-command lines can be set for each structure.
- Structure geometry should be provided after the last @-line of the given structure. Geometry can be described by Cartesian, Z-matrix, or fractional coordinates depending on the **Coord** command (see below)

Keyword	Description
<blank string>	Finish the current structure. Read next string.
Stop	Command **Stop** can be placed after any @-block. It terminates processing the *.inp file at this point. No data will be read in beyond this line. It can be used for the debug runs to process only one or several structures from of numerous structures described in the file. It is not required for the normal run.

2. Keywords on #-marked strings (global commands applied to all structures)

Keyword	Option	Conditions	Default	Description
ParOpt[=*s*]	s:{		SQP	Optimize potential parameters:
	SQP |			Use SQP method (local optimization)
	MC |			Use Metropolis Monte-Carlo Simulated Annealing (MCC-SA) method (global optimization)
	ABC }			Use Artificial Bee Colony (ABC) method (global optimization)
GeoOpt[=*s*]	s:{		LBFGS	Optimize molecular geometry
	LBFGS |			Use L-BFGS algorithm (local optimization)
	ABC |			Use Artificial Bee Colony (ABC) method (global optimization)
	MC }			Use Metropolis Monte Carlo Simulated Annealing method (global optimization of structure)
Growth[=(*s*)]	s:{			Global optimization of series of clusters of growing size
	{**ABC** | **MC**},		ABC	Method of optimization (ABC or MCC-SA)
	NBEG=*n*,		100	Starting nuclearity of clusters
	NEND=*n*,		100	Final nuclearity
	MaxCycIncr=*n*,		0	Increment of optimization cycles number. MaxCyc of the current cluster is *MaxCyc+MaxCycIncr*N**MaxCycExp*
	MaxCycExp=*n*}		1	Exponent of optimization cycles number
ABCoption				Options for ABC algorithm
	MaxCyc=*n*	$N, [1, \infty)$	1000	Number of optimization cycles
	Bees=*n*	$N, [1, \infty)$	100	Number of employed bees
	ScoutLimit=*n*	$N, [1, \infty)$	5	Scout limit
	Amplitude=*x*	$R, (0, \infty)$	5.	Amplitude of random atom position
	Ampl2=*x*	$R, (0, \infty)$	5.	Amplitude for scout bees
	MaxLM=*n*	$N, [1, \infty)$	10	Max number of LM to be saved
	LMFreq	Yes/No	No	Calculate frequencies at the located LM?
	Init=*s*	s:{**Random**, **Gauss**}	Random	Method for initial population generation: **Random** - random generation **Gauss** – generation with Gauss deviation from LMs specified in additional command ABCinit=(File1,File2...)
	ScoutMemo=*n*	$N, [0, MaxCyc]$	0	Scout memory about GM (cycles)
	Cluster=*s*	s:	Random	Generate atom positions for the cluster:

		{**Random** |		- randomly in cube
		Gauss |		- Gaussian deviation from the center
		FCC |		- random FCC lattice
		Spheric |		- randomly in sphere
		All}		- all methods randomly
	PrintIP	Yes/No	No	Print initial bee population?
MMCoption				Options for MCC-SA algorithm
	Temp	R, $(0, \infty)$	300.	Initial temperature, arbitrary units
	Cool	R, $(0, 1)$	0.999	Cooling rate. *Temp* at each next cycle is *Temp=Temp*Cool*
ABCinit=*(f_1, f_2, ...)*				Files to initialize ABC algorithm. It can be used for restart from the files *.lm of previous run. Use **Init=Gauss** option of the **ABCoption** command
AllFreq=*{Yes|No}*			No	Calculate frequencies for all the optimized structures independently whether Freq command is set on their @ lines
MaxCyc=*n*				
MaxGeoCyc=*n*				
ParBoundaryFactor=*x*		R, [1, 1000]	1.	Parameter Boundary Factor. If the optimization parameters boundaries are not set explicitly, they are defined as *Xmin(i)=X(i)* / ParBoundayFactor *Xmax(i)=X(i)* * ParBoundayFactor. X(i) is the initial value for the *i*-th optimization parameter.
PrintLevel=*n*		Z, {0,1,2,3}	1	Printing level
	0			Lower print
	1			Normal print
	2			Debug print
	3			Extended debug print
Reax=*(options)*				Additional options for ReaxFF calculations
Seed=*(n,m)*		*n*: N, $[1, \infty)$ *m*: N, $[1, \infty)$		Seed (two integer numbers) for random number generator
Analysis[=*(s)*]	s:{			Make additional analysis of located LM structures using the file *.lm (it additionally "purifies" the LMs located). The results are written to the file *.slm.
	File=*lmfile*	s		Name of LMs file *.lm to be analyzed
	Nlm=*n*	*n*: N, $[1, \infty)$	all LMs	Number of LMs in the file *.lm to be analyzed
	NoFineOpt	P/NP	NP	Do not perform additional fine optimization of LM structures
	NoFreq }	P/NP	NP	Do not calculate frequencies for LM structures
GenInp[=*(s)*]	s:{			Generate input files for external QC programs from all structures described in <FilInp>.inp . Specific QC commands can be set in the pattern file (<FilInp>.<ext0>, <ext0>=*gj0*, *nw0*, *pd0*, *ga0*). If pattern file is absent, it will be generated automatically.
	Beg |		Beg	Generate input files for initial geometry (before optimization)
	End |			Generate input files for geometry after optimization
	Both			Generate input files both for initial and optimized geometries
	Gau/		Gau	Gaussian input files (*.gjf)
	NW/			NWchem input files (*.nw)
	Gam/			GAMESS input files(*.gam)
	PDB }			PDB files (*.pdb)
Server=*(options)*				External server program making the calculations of structure, energy and other properties
	Prog	**LAMMPS**	NP	Use LAMMPS server program to make calculations.
OptVal			NP	Set initial optimization parameter values from input file. After this command, program reads one or several strings containing the values of potential parameters to be optimized. This allows restart using the "X:" string taken from out-file. The ending – blank line. Example: OptVal X: 1.435 2.657 2.345 10.768 1.0e-1 X: 21.35 12.67 21.3 1.68 1.0e-1 <blank line>
Stop				Finish processing the *.inp file at this point. No data will be

| | | | | read in beyond this line. |
| Exit | | | | The same as **Stop** |

3. *Keywords on @-marked strings (commands applied to the current structure)*

Keyword	Option	Conditions	Default	Description
SP				Single point calculation
SPG				Single point + gradient calculation
FOpt[=*s*]			FOpt	Full optimization
	s:{			
	CG	Present|Not	Not	Use CG algorithm instead of LBFGS for the given structure
	MaxCyc=*n*}	Z, [1, 1000]	1000	Maximum number of opt cycles for the given structure
Popt				Partial optimization (Constraints are set within Z-matrix)
Freq			NP	Calculate vibrational frequencies after geometry optimized
IRC				Use Intrinsic reaction Coordinate calculation
	Forward	Present/NP	NP	In forward direction only
	Reverse	Present/NP	NP	In reverse direction only
	Print	Present/NP	NP	Print coordinates during IRC
	MaxCyc=*n*	Z_+, $[0, \infty)$	1000	Maximum number of IRC steps
	Step=*x*	Q_+, $[0, \infty)$	0.02	Step length, Å
	Mode	N, [1, 3*Numat]		Normal mode number to follow by IRC
Coord={**Cartesian|GZ|FR**}			Cartesian	Coordinates describing the srtructure
	Cartesian			Structure is described by Cartesian coordinates
	GZ			Coordinates are in Gaussian Z-matrix format
	FR			Fractional coordinates for periodic calculations. **Cell** parameter should be present.
PBC[={**1|2|3**}]			3	Forces 1D, 2D or 3D periodic calculations
	1	P/NP	NP	1D-PBC
	2	P/NP	NP	2D-PBC
	3	P/NP	P	3D-PBC
Cell=(*a,b,c*)	*a*			Unit Cell Parameters in Å for crystal/slab/polymer system.
	b			Currently, only tetragonal cells are supported, so unit cell angles *alp=bet=gam*=90 deg.
	c			If *a,b* or *c* are negative, they will be considered as additional optimization variables and the unit cell will be subject of structure optimization.
SuperCell=(*na,nb,nc*)	*na*	Z+: $[0, \infty)$	1	Supercell sizes for periodic crystal/slab/polymer calculations.
	nb	Z+: $[0, \infty)$	1	This means that the system under calculation will consist in
	nc	Z+: $[0, \infty)$	1	*na* x *nb* x *nc* unit cells.
Box=(*xmin,xmax,ymin,ymax,zmin,zmax*)			NP	Box size for periodic calculations. This command is an alternative for **Cell** / **SuperCell** combination. **Box** suggests **SuperCell**=(1,1,1) and Cartesian coordinates, not fractional.
	xmin	Q: $(-\infty, \infty)$		lower box boundary coordinate in *x*-direction, Å
	xmax	Q: $(-\infty, \infty)$		higher box boundary coordinate in *x*-direction, Å
	ymin	Q: $(-\infty, \infty)$		lower box boundary coordinate in *y*-direction, Å
	ymax	Q: $(-\infty, \infty)$		higher box boundary coordinate in y-direction, Å
	zmin	Q: $(-\infty, \infty)$		lower box boundary coordinate in *z*-direction, Å
	ymax	Q: $(-\infty, \infty)$		higher box boundary coordinate in *z*-direction, Å
Save=(*comma-delimited list of values of formulas*)				Saves energy, gradients and other calculated parameters of the optimized structure for the further use as a named variable. These variables can be used in the further calculations, e.g. in the deviation calculations of the following structures. Prescribed values: E- optimized energy of the structure, G – its gradient norm, N – number of atoms. *Example:* **Save=(Extr2=E,deltaE=E2-E1,dEn=deltaE/N)** *Rules for formulas:*

1. Numeric values and saved variables can be used
2. Upper and lower cases do not differ
3. + - * / ^ operators can be used
4. ( ) parentheses (including nested) are allowed
5. Intrinsic functions are SQRT, SIN, COS, EXP, LOG
6. Order of computations as in the FORTRAN and C languages (from left to right)

Weights=(*w*, [*wg*, [*we*, [*wv*, [*wf*]]]])

	w	Weighting factor for the whole structure
	wg	Weighting factor for the geometry deviation from the reference. The deviation is calculated by the Kneller algorithm.
	we	Weighting factor for the energy deviation from the reference. The deviation is calculated by the rule described in the **Ediscrep** command
	wv	Weighting factor for the geometry variables deviation from the reference. The deviation is calculated by the rule described in the **Vdiscrep** command
	wf	Weighting factor for the frequencies deviation from the reference. The deviation is calculated by the rule described in the **Fdiscrep** command

Ediscrep=(*formula, RefValue*)

The rule describing the deviation of energy from the reference values. Deviation are calculated as
$SQRT(<Formula>{**}2 – <RefValue>{**}2)$
RefValue is a numeric value. Rules for formulas are described in **Save** command.

Vdiscrep=(*formula, RefValue*)

The rule describing the deviation of energy from the reference values. The rules are described in **Ediscrep** command

Fdiscrep=(*formula, RefValue*)

The rule describing the deviation of energy from the reference values. The rules are described in **Edsicrep** command.

Charge=*(options)*

Description of molecular, atomic and fragment charges as well as the scheme of charge calculations (for Reax and similar force fields)

	[**Total**=*Q*]	*Q:* $(-\infty, \infty)$	0.	Set atomic charges as described in section 5.1
	[**Atomic**]		NP	Set atomic charges as described in section 5.1
	[**Fragments**]		NP	Set fragment charges as described in section 5.1
	[**QEq|ACKS2**]		NP	Use QEq or ACKS2 methods of charge calculations instead of default EMM charge scheme.

Mult=*m*	*m*	$Z_+: [0, \infty)$	1	Electron spin multiplicity of molecule for QC calculations

4. *Potential description (applied to all structures)*

For all the potentials except ReaxFF force field, each potential acting between atoms A and B should be described by five lines, the first one of them should be marked with * sign:

Line 1:	*	Potential name	Atomic number of A	Atomic number of B	[*rCut*	[*rSwitch*]]

rCut – Cut-off radius of the potentail in Å. At the distances r> rCut the potential equals to zero.
rSwitch – Switching radius of the potential in Å. In the range (*rSwitch, rCut*], the potential is replaced by a switching function.

Line 2: N_P initial parameter values X_i, i=1…N_P. X_i: Q, $(-\infty,\infty)$. Number of parameters N_p is set in the program internally.

Line 3: N_P values {0|1} of parameter optimization flags: 0 – *i*-th parameter have not to be optimized, 1 – to be optimized.

Line 4: N_P values *XLB$_i$:* Q, $(-\infty,\infty)$ of lower boundary for the parameters to be optimized. If *i*-th parameter has not to be optimized, any value should be set at the *i-th* position.

Line 5: N_P values *XUB$_i$:* Q,$(-\infty,\infty)$ of upper boundary for the parameters to be optimized. If *i*-th parameter has not to be optimized, any value should be set at the *i-th* position.

Warning!

(1) Do not comment strings inside the potential description block (within lines 1-5). This can result in the improper potential parameter definition.
(2) Do not use exponential formats like 1.d-01, 2.5e-01 for potential parameter values and their boundaries. Only floating point numbers are readable.

For ReaxFF force field only Line 1 should be provided in the form:

Line 1:   *   **REAX**   *<ParameterFile>*

<ParameterFile> – either the *.rpm file of LAMMPS with original ReaxFF parameters, or the parameter file in the own SOPT format (see below). The format of *ParameterFile* is recognized automatically. The optimization of the ReaxFF parameters is possible only with the parameter file in SOPT format. Within this format, the optimization flags, XLB and XUB boundaries are described inside the *ParameterFile*.

Potential descriptions:

Potential name	*Potential*
LJ	Lennard-Jonnes potential
GP	Gupta potential
SC	Sutton-Chen potential
*n***G,** *n={1-10}*	Gauss potential containing *n* Gauss functions
REAX	ReaxFF force field
EAM	Splined EAM potential
Tersoff	Tersoff potential

5. *Structure geometry description (applied to all structures)*

For all structures in input file, the initial geometry parameters can be set in four different ways depending on the **Coord** keyword on the @-line of the current structure:

- Cartesian coordinates (default)
- Z-matrix (**Coord=GZ**)
- Z-matrix combined with Cartesians (**Coord=GZ**)
- Fractional coordinates of atoms inside the unit cell (**Coord=FR**)
- Coordinates of rigid molecular fragments (**Coord=XYZABG**). They describe the atomic/molecular complex or cluster consisted of one or several rigid fragments. Position of each fragment is determined by the six numbers $(x, y, z, \alpha, \beta, \gamma)$ describing the center position of the fragment and its Euler angles for its orientation relatively to the laboratory (fixed) frame.

5.1. Setting the structure geometry with Cartesian coordinates

Cartesian coordinate description contains one or multiple, each line describes single atom in the structure. Each line has a form:

AName | ANameN   X   Y   Z   [Frag   OptFlagX   OptFlagY   OptFlagZ   Charge]

Here:

AName - Atom name presented by atomic number, chemical symbol: C, 8, Pt

ANameN - chemical symbol combined with ordering number of atom in the structure: C1, O12, Pt25

X, Y, Z – values of Cartesian coordinates. They can be in Angstroems (default) or in a.u. if *Units=AU* is present at one of @-marked lines for the given structure

Frag – fragment description which the atom is belongs to. The fragments can be used to describe dissociation of the structure (see *Ediscrep*) or to set the atom charges of the electrically isolated parts of structure.

OptFlag1, OptFlag2, OptFlag3 – Integer 0|1 values used to mark the coordinates varied (OptFlag is 1) or fixed (OptGlag is 0) during the geometry optimization of *X, Y, Z* coordinates, *respectively*.

AtomCharge – Initial electric charge of atom used in some force fields or potentials. Their sum should be equal to the net charge of molecule (see **Charge** keyword). This also allows setting the fragment charges of the system. If **Charge** contain the option **Fragments**, the fragment charges (sums of atom charges for the corresponding fragments indicated by *Frag*) will be used in the further calculations.

5.2. Setting the structure geometry with Z-matrix

The Z-matrix description is similar to the Z-matrix description used in Gaussian programs, with only several differences. Z-matrix description contains one or multiple lines, each line describes single atom in the structure. The atom description is finished by the blank line, or the line containing keywords *Variables* or *Constants*:

	<Numat> lines describing atoms No. 1,2,3, ...X,....<Numat>						
Line1	*AName	ANameN*						
Line2	*AName	ANameN*	*RAbnd*	*Var1/Val1*				
Line3	*AName	ANameN*	*RAbnd*	*Var1/Val1*	*RAang*	*Var2/Val2*		
Line4	*AName	ANameN*	*RAbnd*	*Var1/Val1*	*RAang*	*Var2/Val2*	*RAdihed	-RAang2*	*Var3/Val3*
Lines 5...	*All remaining Numat-4 lines in the form of line 4 ...*						
	[*<blank line>*	**Variables**	**Constants**					
	Var1		*Val1*	[**s** *NumScanPoints* *ScanStep*]			
	Var2		*Val2*				
	...		*...*]				
	[*<blank line>*	**Constants**					
	Const1		*Val1*				
	Const2		*Val2*				
	...		*Val3*]				
	<blank line>						

Here:

Var1, Var2,… - literal variables describing the optimized geometry parameters

Val1, Val2, … - numeric values of bond lengths or angles in Z-matrix

s – mark of the variable to be scanned. 1, 2 or 3 variables can be scanned simultaneously. All remaining variables are optimized in a regular manner (relaxed scan). Each **s** mark should be followed by two numeric values: *NumScan Points* and *ScanStep.*

NumScanPoints - number of points during variable scanning.

ScanStep – scan step

Const1, Const2, … - literal constants describing the geometry parameters fixed during optimization

RAbnd, RAang, RAdihed, RAang2 – reference atoms of Z-matrix describing position of current atom *X*. They should be in the form of atom ordering numbers or in the form *ANameN (not AName!).* Reference atoms should already be described in Z-matrix before atom *X*.

Each atom *X* (except atoms 1, 2, and 3) are described by the bond *X–RAbnd* , valence angle *X-RAbnd–RAang* and dihedral angle *X–RAbnd–RAang–RAdihed*. Alternatively, the atom *X* can be described by the bond *X–RAbnd* and two valence angles: *X–RAbnd–RAang X–RAbnd–RAang2*. In contrast with Gaussian Z-matrix, *RAang2* is marked by leading minus sign (*–RAang2*), not by additional "1" at the end of line.

Atoms 1,2, and 3 are described by the shortened lines as shown above. This is because atom 1 is always placed at the coordinate system origin, atom 2 – on the axis OX, atom 3 – in the plane OXY.

5.3. *Setting the structure geometry with Z-matrix combined with cartesians*

The Z-matrix can be combined with Cartesian coordinates. In that case, some atoms are described by the regular Z-matrix. Other atoms are described by the lines in the form:

AName | ANameN   0   *X|Var1*   *Y|Var2*   *Z|Var3*

Here, 0 is the obligatory mark designating the Cartesian coordinates. In the following lines, this atom can be used as Reference Atoms (using its *ANameN*) for the remaining part of the structure.

X, Y, Z – numeric values of Cartesian coordinates considered as constants (fixed in optimization). They do not require any additional description in *Constants* or *Variables* sections.

Var1, Var2, Var3… - literal variables describing the optimized Cartesian coordinates. If present, they require setting the corresponding variable values in the *Variables* section of Z-matrix, just like regular Z-matrix variables.

5.4. *Setting the structure geometry with fractional coordinates*

If **Coord=FR**, the structure will be described with fractional coordinates of the unit cell, independently on the presence of **PBC** or **SuperCell** keywords. However, it requires setting of **Cell** keyword describing the *a, b, c* parameters of the unit cell (currently, only rectangular cells are supported). Once unit cell parameters are described, the fractional coordinates can be descried by *Numat* lines, each descring single atom in the unit cell:

AName | ANameN   FR1   FR2   FR3   [Frag   OptFlagX   OptFlagY   OptFlagZ   Charge]

Here:

FR1, FR2, FR3 – are the fractional coordinates, i.e. X/a, Y/b, Z/c.

The remaining options are similar to the Cartesian coordinates described above.

Examples of geometry description:

```
# GeoOpt=MC MaxCyc=1000 MaxGeoCyc=500
!# SEED=(1435589156,1494590021)
!# ABCoptions=(Bees=100,ScoutLimit=2,Amplitude=8.,MaxLM=10,MaxCycle=1000,Cluster=FCC)
# Server=(Prog=lammps)
# MMCoptions=(Temp=300.,Cool=0.9999)

!*reax ffield.reax.Au-Jarvi2008
*REAX ffield.reax.AuSCH-Jarvi2011

@au7  FOpt
Au1   0.0   0.0   0.0
Au2   2.0   0.0   0.0
Au3   1.2   1.0   0.0
Au4  -1.2   1.0   0.0
Au5  -2.0   0.0   0.0
Au6   1.2  -1.0   0.0
Au7  -1.2  -1.0   0.0
```

5.5. Setting the structure geometry with fragment coordinates

Fragment coordinates are N_{frag} sextets $(x_i, y_i, z_i, \alpha_i, \beta_i, \gamma_i)$, $i=1,2,...N_{frag}$ describing the molecule as a set of N_{frag} rigid fragments differing by the positions of their centers and Euler angles of their orientation relatively to the laboratory (fixed) coordinate system.

Each fragment is described by fragment name (in a form of *FragName* or *FragName_FragNumber*), six coordinates *x, y, z, alp, bet, gam* and six optimization flags (0 or 1), e.g.:

```
w_1  0.0   0.0   0.0   10.  30.  0.      0 0 0 1 1 1
```

Second part of fragment name (*_FragNumber*) is not necessary and used only for useful enumeration. Optimization flags f_{ik} , $k=1,...6$ determines whether the corresponding coordinate $c_{ik} = (x_i, y_i, z_i, \alpha_i, \beta_i, \gamma_i)$ are varied during the geometry optimization. Ending of geometry description is a blank line. After the geometry description, each unique fragment should be described at additional *Nfrag* groups of lines Each group of lines starts from a line

FragName   FragType

Here, *FragName* if the fragment name used in the geometry description. *FragType* can be: **TIPS, SPC**, **SPCE, TIP3P, TIP4P, TIP5P, user-defined**. For all fragment types except **user-defined**, no additional specifications are required. They describe the internally-stored fragments and potentials describing the well-known TIPS, SPC, SPC/E, TIP3P, TIP4P and TIP5P water models. User-defined

fragments should contain the description of all atoms in the fragment and all the potentials acting between these atoms. The format of the user-defined description:

> *AtomName(i,j)* x_{ij} y_{ij} z_{ij} q_{ij} ! (m_i lines describing coordinates and charge of the
>   ! atoms $j=1...m_i$ of Fragment i). Ending – blank line.
> *PotentialType* NA1 NA2 par1 [par2 ….]   ! Type of potentials acting between <u>chemical elements</u>
>   ! NA1 and NA2 in the fragment. Par1, par2 – are the
>   ! potential parameters of the given potential type. There
>   ! can be one or several potentials in the user-defined
>   ! fragment. Coulomb potential is automatically applied
>   ! for atom pairs with non-zero atomic charges q_{ij}.

At the moment, *PotentialType* can be **LJ** only (Lennard-Jones potential) in a form $U_{LJ}=A/r^{12}+B/r^6$ with two parameters *A* and *B*. Units: $[A]=$ (cal/mol Å^{12}), $[B]=$ (kcal/mol Å^6).

Example for the water hexamer consisted in TIP4P water molecules:

```
! (H2O)6                                         Comment
@w6 Coord=XyzAbg FOpt=(Xyzabg,MaxCycle=10000)    Command string
w_1  0.0   0.0   0.0   10.  30.   0.      0 0 0 1 1 1   First fragment
w_2  0.0   0.0   3.0   10.  170.  0.      1 1 1 1 1 1   Second fragment
w_3  3.0   1.4   1.0   20.  50.   60.     1 1 1 1 1 1   …
w_4  2.0   3.3   2.5   30.  60.   90.     1 1 1 1 1 1
w_5  3.2   2.4  -3.7   60.   0.  240.     1 1 1 1 1 1
w_6  1.9   1.9  -2.9   90.  20.  190.     1 1 1 1 1 1

w tip4p
```

The geometry optimization of fragment coordinates can be requested by the command **FOpt=XYZABG** (it is the default mode in the case of **Coord=XYZABG**). Otherwise, the geometry optimization can be performed in Cartesian coordinates (meaningless for the standard prescribed fragment types like SPC, TIP3P etc). The Numeric or analytic gradients can be requested during the geometry optimization using the regular keyword **Numeric** along with other optimization options, e.g.

FOpt=(XYZABG,Numeric,MaxCycle=10000)

Example for the OH radicals dimer using the rigid user-defined fragments with LJ potentials between oxygen atoms and Coulomb potentials between all atoms:

```
@oh2 Coord=XyzAbg FOpt=(Xyzabg,MaxCycle=1000)
oh  0.0   0.0   0.0   10.  30.  0.       1 1 1 1 1 1
oh  0.0   0.0   3.0   10.  170. 0.       1 1 1 1 1 1

oh user-defined
O 0.0 0.0 0.0          -0.82
H 0.0 0.0 1.0           0.82

LJ 8 8    629.4 625.5
```

The Euler angles are defined in a $Z(\alpha)Y(\beta)Z(\gamma)$ convention (see e.g. https://en.wikipedia.org/wiki/Euler_angles) which corresponds to [Korn's Handbook of Mathematics]. Other definition $Z(\alpha)X(\beta)Z(\gamma)$ can be applied after the program recompilation (subroutines `RotABG` and `GXYZ2GA` should be corrected).