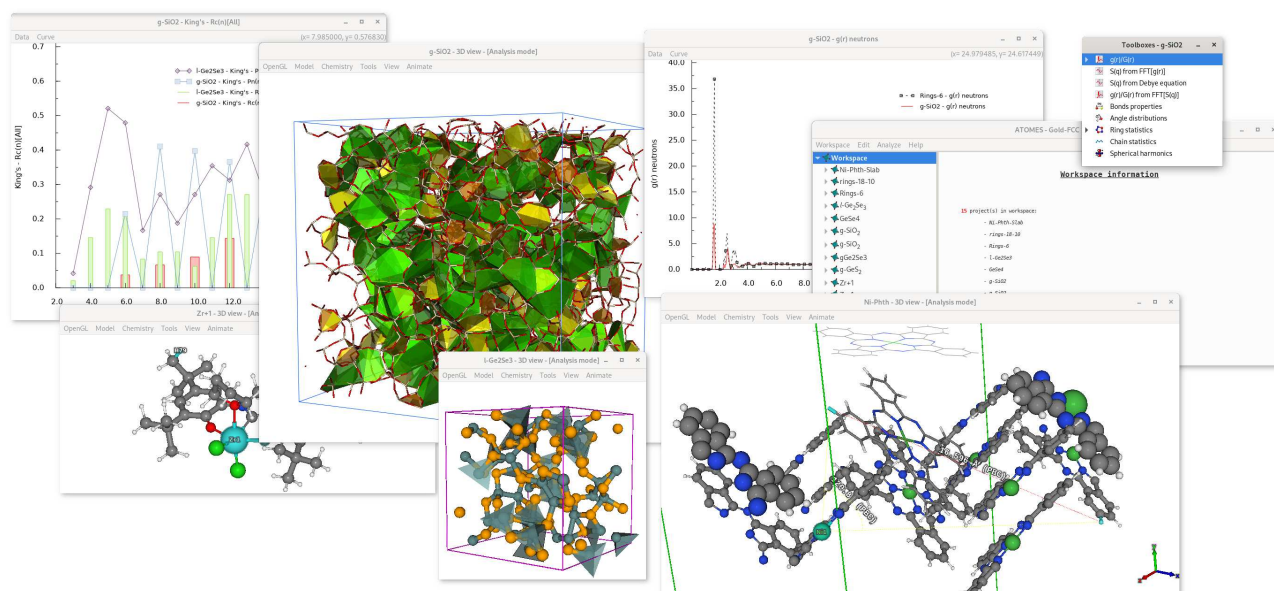


# Atomes

<https://atomes.ipcms.fr>



## User manual Atomes v1.0

Sébastien LE ROUX [sebastien.leroux@ipcms.unistra.fr](mailto:sebastien.leroux@ipcms.unistra.fr)



# Contents

<b>Contents</b>	<b>iii</b>
<b>List of figures</b>	<b>vii</b>
<b>List of tables</b>	<b>ix</b>
<b>1 Introduction</b>	<b>1</b>
<b>2 Programming framework</b>	<b>3</b>
2.1 Supported platforms . . . . .	3
2.2 Dependencies and requirements . . . . .	3
<b>3 Features</b>	<b>5</b>
3.1 Main window . . . . .	5
3.2 Workspace and project tree . . . . .	8
3.3 Files . . . . .	9
3.3.1 Importing atomic coordinates . . . . .	9
3.3.2 Reading <b>Atomes</b> project file(s) . . . . .	12
3.3.3 Reading <b>Atomes</b> workspace file(s) . . . . .	13
3.4 Analyzing models using <b>Atomes</b> . . . . .	14
3.5 Visual analysis using <b>Atomes</b> . . . . .	16
3.6 Visual edition and model creation using <b>Atomes</b> . . . . .	16
3.7 Preparing MD calculations in <b>Atomes</b> . . . . .	17
<b>4 Physico-chemical analysis in Atomes</b>	<b>19</b>
4.1 Visualisation of the results of the calculations . . . . .	20
4.2 Data and plot edition . . . . .	22
4.2.1 Data edition . . . . .	22
4.2.2 Plot edition . . . . .	23
4.3 Mouse interaction with the data plot . . . . .	25
4.3.1 Right button contextual menu . . . . .	25
4.3.2 Left button zoom in and out . . . . .	27
4.4 Saving the data . . . . .	29
4.5 Keyboard shortcuts . . . . .	29
<b>5 Visual analysis in Atomes</b>	<b>31</b>
5.1 Window top bar menu . . . . .	32
5.1.1 The "OpenGL" menu . . . . .	32
5.1.2 The "Model" menu . . . . .	39
5.1.3 The "Chemistry" menu . . . . .	42
5.1.4 The "Tools" menu . . . . .	49

5.1.5	The "View" menu . . . . .	53
5.1.6	The "Animate" menu . . . . .	54
5.2	Mouse interaction with the OpenGL window: visualization . . . . .	57
5.2.1	Object selection . . . . .	57
5.2.2	Object contextual menu . . . . .	58
5.3	Keyboard shortcuts . . . . .	65
<b>6</b>	<b>Visual edition in Atomes</b>	<b>67</b>
6.1	The "Edit" submenu . . . . .	67
6.1.1	The "Crystal builder" window . . . . .	67
6.1.2	The "Cell edition" window . . . . .	79
6.1.3	The "Model edition" window . . . . .	87
6.1.4	The "Extract/rebuild" buttons . . . . .	106
6.2	Mouse interaction with the OpenGL window: edition . . . . .	107
6.2.1	Atom selection . . . . .	108
6.2.2	The edition contextual menu . . . . .	110
6.2.3	The object edition contextual menu . . . . .	111
6.3	Keyboard shortcuts . . . . .	112
<b>7</b>	<b>Preparing molecular dynamics calculations in Atomes</b>	<b>115</b>
7.1	Classical MD . . . . .	115
7.1.1	DL-POLY v4.09 . . . . .	115
7.1.2	LAMMPS . . . . .	148
7.2	<i>Ab-initio</i> MD . . . . .	149
7.2.1	CPMD v4.3.0 . . . . .	149
7.2.2	CP2K . . . . .	155
7.3	Quantum Mechanics and Molecular Mechanics MD . . . . .	155
7.3.1	CPMD . . . . .	155
7.3.2	CP2K . . . . .	155
	<b>Appendix</b>	<b>I</b>
<b>A</b>	<b>Supported atomic coordinates file formats</b>	<b>III</b>
<b>B</b>	<b>Keyboard shortcuts and command line options</b>	<b>V</b>
B.1	Keyboard shortcuts . . . . .	V
B.2	Command line options . . . . .	IX
<b>C</b>	<b>The "Library" in Atomes</b>	<b>XI</b>
<b>D</b>	<b>Crystal building in Atomes</b>	<b>XV</b>
<b>E</b>	<b>The physics in Atomes</b>	<b>XXI</b>
E.1	The periodic boundary conditions . . . . .	XXI
E.2	Radial distribution functions fundamentals . . . . .	XXIII
E.3	Neutrons and X-rays scattering . . . . .	XXV
E.3.1	Total scattering - Debye approach . . . . .	XXV
E.3.2	Partial structure factors . . . . .	XXVIII
E.4	Local atomic coordination properties . . . . .	XXXI
E.4.1	Average first coordination numbers . . . . .	XXXI

E.4.2	Individual atomic neighbor analysis . . . . .	XXXI
E.4.3	Proportion of tetrahedral links and units in the structure model . . . . .	XXXII
E.4.4	Distribution of bond lengths for the first coordination spheres . . . . .	XXXIII
E.4.5	Angles distributions . . . . .	XXXIII
E.5	Ring statistics . . . . .	XXXIV
E.5.1	Size of the rings . . . . .	XXXIV
E.5.2	Definitions . . . . .	XXXIV
E.5.3	Description of a network using ring statistics - existing tools . . . . .	XXXIX
E.5.4	Rings and connectivity: the R.I.N.G.S. method implemented in <b>Atomes</b> .	XL
E.5.5	Bond defects in ring statistics . . . . .	XLVIII
E.5.6	Number of rings not found and that "potentially exist" . . . . .	XLIX
E.6	Chain statistics . . . . .	LI
E.7	Invariants of spherical harmonics as atomic order parameters . . . . .	LI
E.8	Mean square displacement of atoms . . . . .	LII
<b>F</b>	<b>The chemical properties database in Atomes</b>	<b>LV</b>
F.1	Atomic radii . . . . .	LV
F.1.1	Covalent radii . . . . .	LV
F.1.2	Ionic radii . . . . .	LVI
F.1.3	VDW radii . . . . .	LVI
F.1.4	Shannon radii in crystal . . . . .	LVII
F.2	Neutron scattering lengths . . . . .	LVII



# List of Figures

1.1	Overview of the <b>Atomes</b> program. . . . .	2
3.1	Main window of the <b>Atomes</b> program. . . . .	6
3.2	Project tree in the <b>Atomes</b> program. . . . .	8
3.3	Mouse left click menu in the workspace tree of the <b>Atomes</b> program. . . . .	9
3.4	The "Chemistry and physics" dialog in the <b>Atomes</b> program. . . . .	10
3.5	The "Box and periodicity" dialog in the <b>Atomes</b> program. . . . .	10
3.6	The "Bond cutoffs" dialog in the <b>Atomes</b> program. . . . .	11
3.7	The "Open Project File(s)" dialog in the <b>Atomes</b> program. . . . .	12
3.8	The "Open Workspace" dialog in the <b>Atomes</b> program. . . . .	13
4.1	Example workspace: physico-chemical analysis . . . . .	19
4.2	The "Toolboxes" dialog in the <b>Atomes</b> program. . . . .	20
4.3	Interaction with the "Toolboxes" dialog in the <b>Atomes</b> program. . . . .	21
4.4	"Data" and "Edit Data" menus, "Data Edition" window. . . . .	22
4.5	The "Curve" menu. . . . .	23
4.6	The data plot editing tool box in the <b>Atomes</b> program. . . . .	23
4.7	Examples of data plot edition in the <b>Atomes</b> program. . . . .	24
4.8	Mouse contextual menu over the graph window in the <b>Atomes</b> program. . . . .	25
4.9	The "Edit Data" a), "Add Data Set" b) and "Remove Data Set" c) menus. . . . .	26
4.10	The mouse left button zoom in/out of the graph window in the <b>Atomes</b> program. . . . .	28
5.1	Example workspace: visualization. . . . .	31
5.2	The "OpenGL" menu and the attached submenus. . . . .	32
5.3	Standard color maps in the <b>Atomes</b> program. . . . .	34
5.4	The "OpenGL material aspect and light settings" dialog in the <b>Atomes</b> program. . . . .	38
5.5	The "Model" menu and the attached submenus. . . . .	39
5.6	The "Atom(s) configuration" dialog in the <b>Atomes</b> program. . . . .	40
5.7	The concept of clones in the <b>Atomes</b> program. . . . .	41
5.8	The "Chemistry" menu and the attached submenus. . . . .	42
5.9	The "Tools" menu and the attached submenus in the <b>Atomes</b> program. . . . .	49
5.10	The "Measures" dialog in the <b>Atomes</b> program. . . . .	50
5.11	Some examples of measurements displayed in the OpenGL window. . . . .	51
5.12	The "View" menu and the attached submenus in the <b>Atomes</b> program. . . . .	53
5.13	The "Animates" menu of the OpenGL window in the <b>Atomes</b> program. . . . .	54
5.14	The "Spin" dialog of the OpenGL window in the <b>Atomes</b> program. . . . .	54
5.15	The "Sequencer" dialog of the OpenGL window in the <b>Atomes</b> program. . . . .	55
5.16	The "Recorder" dialog of the OpenGL window in the <b>Atomes</b> program. . . . .	55
5.17	The "Movie encoding" dialog of the OpenGL window in the <b>Atomes</b> program. . . . .	56
5.18	Illustration of the selection / label process using the mouse left button. . . . .	57
5.19	The object contextual menu: <b>a)</b> for an atom, <b>b)</b> for a chemical bond. . . . .	58
5.20	Construction of the submenus attached to the object contextual menu. . . . .	59
5.21	The object contextual menu and the "Style" submenu(s). . . . .	60
5.22	Multiple visual styles the <b>Atomes</b> program. . . . .	60
5.23	The object contextual menu and the "Copy" submenu(s). . . . .	61

5.24	The object contextual menu and the "Edit as new project" submenu(s).	62
6.1	Accessing the edition tools in the <b>Atomes</b> program.	67
6.2	Accessing the "Crystal builder" window in the <b>Atomes</b> program.	67
6.3	The "Crystal builder" window in the <b>Atomes</b> program.	68
6.4	The "space group info" dialog for the <b>R3c</b> group in the <b>Atomes</b> program.	70
6.5	Building a C-diamond crystal in the <b>Atomes</b> program.	71
6.6	Building a C-diamond like $C_{60}$ crystalline super-structure.	75
6.7	Building a C-diamond like, alternating $C_{60}$ and toluene molecules, crystalline super-structure.	76
6.8	Building a C-diamond like, toluene encapsulated $C_{240}$ , crystalline super-structure.	77
6.9	Accessing the "Cell edition" window in the <b>Atomes</b> program.	79
6.10	Model edition using the "Cut slab" tab.	88
6.11	Accessing the "Model edition" window in the <b>Atomes</b> program.	89
6.12	Modified selection process when the "Model edition" window is opened.	89
6.13	The atom(s) search tool.	90
6.14	Initial position with an isolated fragment selected in the model.	91
6.15	Final position of the fragment, after translation and rotation in the model.	94
6.16	The "Replace" tab with the option of replace objects "Normally" or "Randomly".	96
6.17	Search trees and "Select for ..." menu in the "Replace" tab.	97
6.18	Random substitution of 500 "Si" atoms by "water" molecules.	99
6.19	Normal removal of all "Ni-[ $N_4$ ]" coordination spheres from the model.	100
6.20	The "Insert" tab, with the options to choose the object to be inserted and the position of insertion.	101
6.21	Insertion of fullerene " $C_{240}$ " in the "Ni-Phth" project modified after the removal example.	102
6.22	The "Random move" tab.	103
6.23	Random rotation of "water" molecules in the modified " $SiO_2$ " project.	104
6.24	Illustration of the rebuild process during copy.	106
6.25	Activating the mouse "Edition" mode using the "Tools" menu.	107
6.26	Activating the "Measures" selection mode using the "Tools" menu.	108
6.27	The right button contextual menu on the background of the OpenGL window.	110
6.28	The object edition contextual menu: <b>a</b> ) for an atom, <b>b</b> ) for a chemical bond.	111
7.1	Initializing the force field in the DL-POLY calculation assistant.	116
7.2	The initialized force field in the DL-POLY calculation assistant.	117
7.3	The "CONTROL" file options in the DL-POLY calculation assistant.	118
7.4	"Select the component(s) of the force field" tab in the DL-POLY calculation assistant.	119
7.5	"Molecule(s)" tab in the DL-POLY calculation assistant.	120
7.6	Visualizing the "Molecule(s)" using the DL-POLY calculation assistant.	121
7.7	Visualizing the atom(s) number(s) within the "Molecule(s)".	122
7.8	Adding molecules to the force field using the DL-POLY calculation assistant.	123
7.9	Merging molecules from the force field using the DL-POLY calculation assistant.	124
7.10	The "Atom(s)" tab in the DL-POLY assistant.	125
7.11	Inserting a new atom description in the "Atom(s)" tab.	126
7.12	Merging two existing atom descriptions in the "Atom(s)" tab.	127
7.13	Editing an atom description in the "Atom(s)" tab.	128
7.14	Choosing a force field description for the atom type in the "Atom(s)" tab.	129
7.15	The "Update FIELD file with the force field parameters" dialog.	129
7.16	Updating parameters for the "FIELD" file using the "CB" atom(s) in the "Amber 99" force field.	130
7.17	"Core-shell unit(s)" tab in the "DL-POLY" assistant.	131
7.18	"Constrains bond(s)" tab in the "DL-POLY" assistant.	132
7.19	"Mean force potential(s)" tab in the "DL-POLY" assistant.	133
7.20	"Rigid unit(s)" tab in the "DL-POLY" assistant.	134
7.21	"Tethering potential(s)" tab in the "DL-POLY" assistant.	135
7.22	"Flexible chemical bond(s)" tab in the "DL-POLY" assistant.	136
7.23	"Flexible chemical bond(s)" edition dialog in the "DL-POLY" assistant.	137
7.24	"Bond angle(s)" tab in the "DL-POLY" assistant.	138
7.25	"Dihedral angle(s)" tab in the "DL-POLY" assistant.	139
7.26	"Improper angle(s)" tab in the "DL-POLY" assistant.	139



7.27	"Inversion angle(s)" tab in the "DL-POLY" assistant. . . . .	140
7.28	"van der Waals potential(s)" tab in the "DL-POLY" assistant. . . . .	141
7.29	"Metal potential(s)" tab in the "DL-POLY" assistant. . . . .	142
7.30	"Tersoff potential(s)" tab in the "DL-POLY" assistant. . . . .	143
7.31	"Three potential(s)" tab in the "DL-POLY" assistant. . . . .	144
7.32	"Four body potential(s)" tab in the "DL-POLY" assistant. . . . .	145
7.33	"External field(s)" tab in the "DL-POLY" assistant. . . . .	146
7.34	"DL-POLY files preview" dialog in the "DL-POLY" assistant. . . . .	147
7.35	"INFO section" tab in the "CPMD" assistant. . . . .	149
7.36	"CPMD section - Calculation options" tab in the "CPMD" assistant. . . . .	150
7.37	"CPMD section - Thermostat options" tab in the "CPMD" assistant. . . . .	151
7.38	"CPMD section - Restart options" tab in the "CPMD" assistant. . . . .	151
7.39	"DFT section" tab in the "CPMD" assistant. . . . .	152
7.40	"VDW section" tab in the "CPMD" assistant. . . . .	152
7.41	"PROP section" tab in the "CPMD" assistant. . . . .	153
7.42	"SYSTEM section" tab in the "CPMD" assistant. . . . .	153
7.43	"ATOMS section" tab in the "CPMD" assistant. . . . .	154
C.1	The "Library" dialog of the <b>Atomes</b> program. . . . .	XI
C.2	Using the "Library" dialog. . . . .	XII
E.1	Schematic representation of the idea of periodic boundary conditions. . . . .	XXII
E.2	Space discretization for the evaluation of the radial distribution function. . . . .	XXIII
E.3	Partial radial distribution functions of glassy GeS <sub>2</sub> at 300 K. . . . .	XXIV
E.4	Total neutron structure factor for glassy GeS <sub>2</sub> at 300 K. . . . .	XXVI
E.5	Example of various distribution functions neutron-weighted in glassy GeS <sub>2</sub> at 300 K. . . . .	XXVII
E.6	Partial structure factors of glassy GeS <sub>2</sub> at 300 K. . . . .	XXX
E.7	Coordination numbers. . . . .	XXXI
E.8	Illustration of several coordination spheres that can be found in glassy GeS <sub>2</sub> . . . . .	XXXI
E.9	Illustration of the presence of GeS <sub>4</sub> tetrahedra in the GeS <sub>2</sub> material's family. . . . .	XXXII
E.10	Corner sharing (left) and edge sharing (right) tetrahedra. . . . .	XXXII
E.11	Nearest neighbor distances distribution. . . . .	XXXIII
E.12	Bond angles (left) and dihedral angles (right). . . . .	XXXIII
E.13	King's criterion in the ring statistics. . . . .	XXXIV
E.14	Guttman's criterion in the ring statistics. . . . .	XXXV
E.15	Differences between the King and the Guttman ring statistics in an AB <sub>2</sub> system. . . . .	XXXVI
E.16	Theoretical maximum size of the rings for an AB <sub>2</sub> system. . . . .	XXXVII
E.17	Primitive rings in the ring statistics. . . . .	XXXVII
E.18	Strong rings in the ring statistics. . . . .	XXXVIII
E.19	A very simple network. . . . .	XXXIX
E.20	Two simple networks having very close compositions: 10 nodes and 7 links. . . . .	XL
E.21	The first comparison element: the total number of rings in the network. . . . .	XLI
E.22	The 9 different networks with 16 nodes, composed of 2 rings with 6 nodes and 1 ring with 4 nodes. . . . .	XLII
E.23	Schematic representation of cases g) → i) (1) and e) → f) (2) illustrated in figure E.22. . . . .	XLIII
E.24	Cluster of atoms with a bond defect isolated from an AB <sub>2</sub> amorphous network. . . . .	XLVIII
E.25	Illustration of the influence of homopolar bonds in ABAB rings. . . . .	XLIX
F.1	Covalent radii used in the <b>Atomes</b> program. . . . .	LV
F.2	Ionic radii used in the <b>Atomes</b> program. . . . .	LVI
F.3	Van Der Waals radii used in the <b>Atomes</b> program. . . . .	LVI
F.4	Shannon radii in crystals used in the <b>Atomes</b> program. . . . .	LVII
F.5	Neutron scattering lengths used in the <b>Atomes</b> program. . . . .	LVII



# List of Tables

2.1	Libraries used by <b>Atomes</b> . . . . .	3
A.1	Atomic coordinates file formats read by <b>Atomes</b> . . . . .	III
C.1	Example of "sml" file in XML format for caffeine. . . . .	XIII
D.1	Example of the "167-R-3c.sg1" file in XML format for the $R\bar{3}c$ space group. . . . .	XX
E.1	$R_N$ and $R_C$ for the network in figure E.19 . . . . .	XXXIX
E.2	$R_N$ and $R_C$ calculated for the networks illustrated in figure E.20. . . . .	XL
E.3	Number of rings in the simple networks represented in figure E.21. . . . .	XLI
E.4	Number of rings for the different networks presented in figure E.22. . . . .	XLII
E.5	Number of rings per node for the networks presented in figure E.22. . . . .	XLIII
E.6	General connectivity matrix. . . . .	XLIII
E.7	General connectivity matrix for the networks represented in figure E.22. . . . .	XLIV
E.8	Proportion of nodes at the origin of at least one ring of size $n$ for the networks in figure E.22. . . .	XLV
E.9	Connectivity profiles results of the ring statistics for the networks presented in figure E.22. . . .	XLVII



---

---

# Appendix

---



---

# Supported atomic coordinates file formats

---

The current version of **Atomes** can import atomic coordinates in the following formats:

- XYZ [30] \*
- Chem3D [31]
- CPMD trajectory [3] \* \*\*
- VASP trajectory [32] \* \*\*\*
- ISAACS Project File [33]
- PDB/ENT (Protein Data Bank) v3.30
- CIF (Crystallographic information)
- DL-POLY History file \*

**Table A.1** *Atomic coordinates file formats read by Atomes.*

\* including NPT trajectories.

\*\* atomic units are assumed in the case of CPMD trajectories.

\*\* and \*\*\* require to enter extra parameters through an interactive dialog box:

- Total number of atoms
- Number of chemical species
- Label and number of each atomic species

the input order, label and number of each atomic species will be the one assumed when reading the coordinates from the trajectory file.

**Atomes** can also export atomic coordinates in XYZ or Chem3D format.





---

# Keyboard shortcuts and command line options



---



## B.1 Keyboard shortcuts

- Main window

- Workspace:



 +  : open workspace

 +  : save workspace as



 +  : close workspace

- Project:



 +  : create new project



 +  : open project

- Misc:



 +  : show curve toolboxes

 +  : open periodic table



 +  : show about dialog



 +  : quit

- Curve window:

 +  : Autoscale

 +  : Close curve window

 +  : Open the data plot editing tool box [Fig. 4.6]

 +  : Export image

 +  : Save / export data

- OpenGL window:

- Single key shortcuts:


- \* Colors:


-  : change atom(s) colormap


-  : change polyhedra(ons) colormap


- \* Styles:


-  : change default style to "Ball and stick"

-  : change default style to "Cylinders"


-  : change default style to "Dots"

-  : change default style to "Spheres"

-  : change default style to "Spacefilled"

-  : change default style to "Wireframe"

- \* Measures:

-  : show all measures for the selection, if pressed:


- once: display inter-atomic distance(s)

- twice: display inter-atomic angles

- a third time: hide measures

- \* Misc:



-  : exit fullscreen mode

-  : pause / restart spinning



– Combined keys shortcuts:



\* Mouse mode:


 +  : enter mouse "Analysis" mode

 +  : enter mouse "Edition" mode



\* Selection:



 +  : select / unselect all atoms



 +  : copy all selected atom(s)



 +  : create new (empty project)



\* Misc:

 +  : label / unlabel all atoms

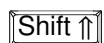

 +  : "Environments configuration" window [Sec. 5.1.3]

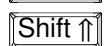

 +  : "Measures" dialog [Sec. 5.1.4]

 +  : "Recorder" dialog [Sec. 5.1.6]

 +  : enter / exit fullscreen mode

\* Camera motion:


 +  : zoom out


 +  : zoom in

• OpenGL window **Analysis mode only**

– Single key shortcuts:

\* Model rotation:

 : rotate right

 : rotate left


 : rotate up


 : rotate down

– Combined keys shortcuts:

\* Camera motion:




 +  : move camera right

 +  : move camera left



 +  : move camera up




 +  : move camera down



\* Spinning:

 +  +  : spin right / increase speed r. or reduce speed left

 +  +  : spin left / increase speed left or reduce speed right

 +  +  : spin up / increase speed up or reduce speed down


 +  +  : spin down / increase speed d. or reduce speed up

 +  : stop spinning


• OpenGL window: **Edition mode only**


– Single key shortcuts:

\* Atomic coordinates rotation:

 : rotate atomic coordinates right


 : rotate atomic coordinates left



 : rotate atomic coordinates up



 : rotate atomic coordinates down

– Combined keys shortcuts:

\* Atomic coordinates translation:

 +  : translate atomic coordinates right

 +  : translate atomic coordinates left

 +  : translate atomic coordinates up

 +  : translate atomic coordinates down

## B.2 Command line options

**Atomes** can be used from the command line, including using the following options:

- General options:

"-h" or -help: short help.

"-v" or -version: version information.

- File options:

"-awf filename": open **Atomes** workspace file. \* \*\*

"-apf filename": open **Atomes** project file. \*

"-xyz filename": open coordinates in XYZ format. \*

"-pdb filename": open coordinates in PDB format. \*

"-ent filename": open coordinates in PDB format. \*

"-c3d filename": open coordinates in Chem3D format. \*

"-cif filename": open coordinates in CIF format. \*

"-trj filename": open CPMD trajectory. \*

"-xdatcar filename": open VASP trajectory. \*

"-hist filename": open DL\_POLY history trajectory. \*

"-ipf filename": open ISAACS project file. \*

\* the structure of "filename" is expected to follow the structure introduced with the option (ex: "-xyz" for a file that contains coordinates in the XYZ format, but "filename" is not required to have the corresponding extension (ex: ".xyz")).

\*\* since **Atomes** has a single workspace only a single workspace can be opened by the command line.

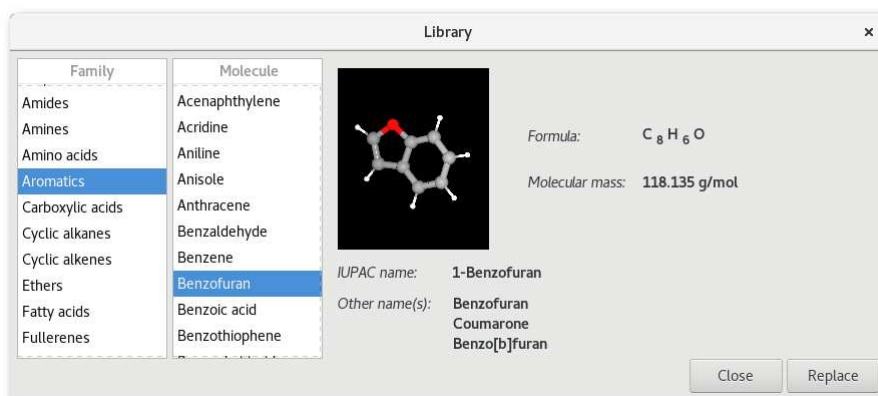


---

# The "Library" in Atomes

---

The **Atomes** program includes a library with a number of molecules to be used when building or editing atomistic models:

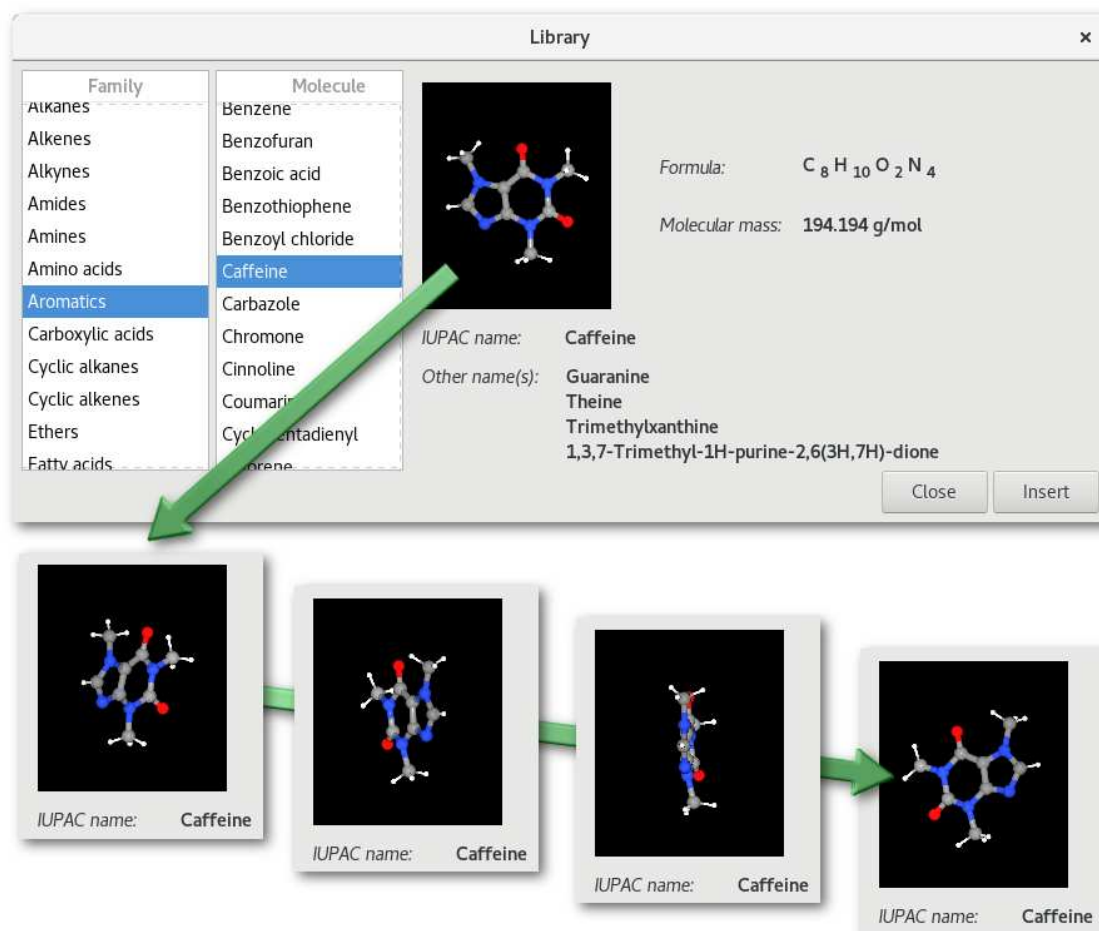


**Figure C.1** The "Library" dialog of the **Atomes** program.

Available molecules can be found in the library by sorted in the following families:

- Misc
- Alcohols
- Aldehydes
- Alkanes
- Alkenes
- Alkynes
- Amides
- Amines
- Amino acids
- Aromatics
- Carboxylic acids
- Cyclic alkanes
- Cyclic alkenes
- Ethers
- Fatty acids
- Fullerenes
- Heterocyclics
- Macrocycles
- Ketones
- Nitriles
- Nucleobases
- Steroids
- Sugars (Linears)
- Sugars (Cyclics)
- Sulfoxides
- Thiols

Each time a new molecule is selected in the tree view on the left side of the window (see figure C.1), then the right side of the "Library" dialog is refreshed with the data of the newly selected molecule. The data includes a 3D representation included in an active OpenGL window that allows to rotate the molecule:



**Figure C.2** Using the "Library" dialog.

The data files that contain the information regarding each molecule in the library are located in the "bin/library/molecules" directory.

To each molecule that appears in the "Library" tree view, a file with the ".sml" extension can be found in the corresponding family subfolder.

For the example in figure C.2:

- The family is: "Aromatics"
- The molecule is: "Caffeine"
- The corresponding file is: "Aromatics/caffeine.sml".

New files can be added to the library to be found by **Atomes** at startup, providing that the new files are added to one of the families of the library. These files must have the ".sml" extension, follow the "XML" encoding rules, and the following structure:



## The "SCL" "Simple chemical library XML file" format

```

<?xml version="1.0" encoding="UTF-8"?>
<!-- Simple chemical library XML file -->
<scl-xml>
  <class>Aromatics</class>
  <names>
    <library-name>Caffeine</library-name>
    <iupac-name>Caffeine</iupac-name>
    <other-names>
      <name>Guaranine</name>
      <name>Theine</name>
      <name>Trimethylxanthine</name>
      <name>1,3,7-Trimethyl-1H-purine-2,6(3H,7H)-dione</name>
    </other-names>
  </names>
  <chemistry>
    <atoms>24</atoms>
    <species number="4">
      <label id="0" num="8">C</label>
      <label id="1" num="10">H</label>
      <label id="2" num="2">O</label>
      <label id="3" num="4">N</label>
    </species>
  </chemistry>
  <coordinates>
    <atom id="1" sp="0" x="1.785021" y="-0.779129" z="-0.255949"/>
    <atom id="2" sp="0" x="0.401929" y="1.318216" z="-0.016004"/>
    <atom id="3" sp="0" x="-0.733464" y="0.413478" z="0.038347"/>
    <atom id="4" sp="0" x="-0.617909" y="-0.973998" z="-0.109571"/>
    <atom id="5" sp="0" x="-2.772439" y="-0.562269" z="0.209813"/>
    <atom id="6" sp="0" x="0.735608" y="-3.069196" z="-0.210904"/>
    <atom id="7" sp="0" x="-2.730788" y="1.961282" z="0.459758"/>
    <atom id="8" sp="0" x="2.908891" y="1.458603" z="-0.240195"/>
    <atom id="9" sp="1" x="-3.848274" y="-0.720508" z="0.332764"/>
    <atom id="10" sp="1" x="1.561100" y="-3.446348" z="-0.829847"/>
    <atom id="11" sp="1" x="-0.192018" y="-3.562176" z="-0.530166"/>
    <atom id="12" sp="1" x="0.935512" y="-3.340819" z="0.835372"/>
    <atom id="13" sp="1" x="-2.333498" y="2.706064" z="-0.245137"/>
    <atom id="14" sp="1" x="-2.525360" y="2.320312" z="1.478733"/>
    <atom id="15" sp="1" x="-3.818507" y="1.892617" z="0.326144"/>
    <atom id="16" sp="1" x="3.674681" y="0.991531" z="-0.874770"/>
    <atom id="17" sp="1" x="3.303682" y="1.537341" z="0.782105"/>
    <atom id="18" sp="1" x="2.721889" y="2.471199" z="-0.622647"/>
    <atom id="19" sp="2" x="2.892437" y="-1.297551" z="-0.212698"/>
    <atom id="20" sp="2" x="0.363733" y="2.532611" z="0.115710"/>
    <atom id="21" sp="3" x="1.659632" y="0.660920" z="-0.271271"/>
    <atom id="22" sp="3" x="0.612806" y="-1.608906" z="-0.393730"/>
    <atom id="23" sp="3" x="-2.109760" y="0.657267" z="0.234906"/>
    <atom id="24" sp="3" x="-1.874903" y="-1.560540" z="-0.000764"/>
  </coordinates>
</scl-xml>

```

**Table C.1** Example of "sml" file in XML format for caffeine.



---

# Crystal building in Atomes

---

The crystal building process in **Atomes** will be briefly presented in the next pages. 2 example cases will be used to illustrate the methodology implemented in the program:

1. The  $Fd\bar{3}m$  space group, N°227, with the first setting: origin 1.
2. The  $R\bar{3}c$  space group, N°167, with the second setting: rhombohedral parameters.

The terminology used in the **Atomes** program, and therefore used in this manual, is directly taken from the International tables for Crystallography Vol. A. [29]

In the following every single step depends on the considered space group, and corresponds to the way the process is implemented in **Atomes**:

1. Determine the origin of the coordinate system.
2. Calculate the inverse transformation matrix.
3. Select the space group extra site positions, if any.
4. Adjust the different symmetrical positions of the first Wyckoff position using 2.
5. Calculate the crystalline positions using 4 and 3.
6. Select the unique position(s) in 5 to build the crystal.
7. Calculate the object position in Cartesian coordinates using 6.
8. Insert object(s) at the appropriate coordinates based on occupancy.

## Origin of the coordinate system

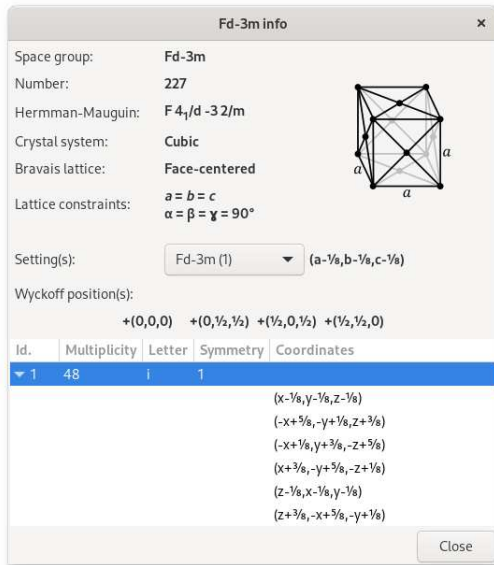
The origin of the coordinate system, if any particular, determine how the object fractional coordinates are modified to match the space group criteria before calculating the Cartesian coordinates.

The first step is to build the  $4 \times 3$  matrix that describe the operation of be performed, the required information is given by the space group setting, the easiest way to use it is on the form of the transformation matrix **P**:

Ex: (**a**, **b**, **c**) and can be converted in:

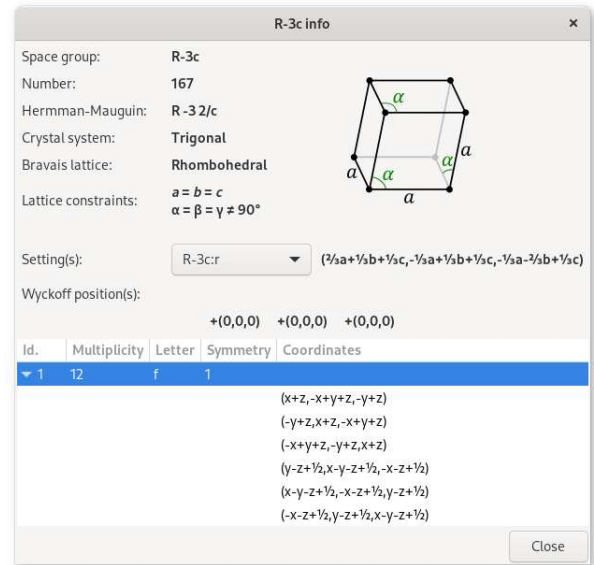
$$\begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} \quad (\text{D.1})$$

(1)  $\text{Fd}\bar{3}\text{m}$  space group, N°227, origin 1:



$$\left(a-\frac{1}{8}, b-\frac{1}{8}, c-\frac{1}{8}\right)$$

(2)  $\text{R}\bar{3}\text{c}$  space group, N°167, second setting:



$$\left(\frac{2}{3}\mathbf{a}+\frac{1}{3}\mathbf{b}+\frac{1}{3}\mathbf{c}, -\frac{1}{3}\mathbf{a}+\frac{1}{3}\mathbf{b}+\frac{1}{3}\mathbf{c}, -\frac{1}{3}\mathbf{a}-\frac{2}{3}\mathbf{b}+\frac{1}{3}\mathbf{c}\right)$$

The corresponding **P** matrix are:

$$\begin{bmatrix} 1 & 0 & 0 & -\frac{1}{8} \\ 0 & 1 & 0 & -\frac{1}{8} \\ 0 & 0 & 1 & -\frac{1}{8} \end{bmatrix}$$

$$\begin{bmatrix} \frac{2}{3} & -\frac{1}{3} & -\frac{1}{3} & 0 \\ \frac{1}{3} & \frac{1}{3} & -\frac{2}{3} & 0 \\ \frac{1}{3} & \frac{1}{3} & \frac{1}{3} & 0 \end{bmatrix}$$

## Inverse transformation matrix

To transform the coordinates to the appropriate system we use the corresponding inverse matrix  $\mathbf{P}^{-1}$ :

(1)  $\text{Fd}\bar{3}\text{m}$  space group, N°227, origin 1:

$$\begin{bmatrix} 1 & 0 & 0 & -\frac{1}{8} \\ 0 & 1 & 0 & -\frac{1}{8} \\ 0 & 0 & 1 & -\frac{1}{8} \end{bmatrix}$$

(2)  $\text{R}\bar{3}\text{c}$  space group, N°167, second setting:

$$\begin{bmatrix} 1 & 0 & 1 & 0 \\ -1 & 1 & 1 & 0 \\ 0 & -1 & 1 & 0 \end{bmatrix}$$

The  $\mathbf{P}^{-1}$  matrix allows to easily compute the coordinates  $(x', y', z')$  of a point  $(x, y, z)$  in the appropriate coordinate system using:

$$(x', y', z') = \mathbf{P}^{-1} \times (x, y, z) \quad (\text{D.2})$$

## Space group extra site position(s)

The space group extra site position(s) describe translations to be performed on the initial set of Wyckoff position to get the entire list of coordinates.

The list of extra position(s),  $E(e_x, e_y, e_z)$ , is imported from the space group data file, and is presented on top the Wyckoff positions table of the space group info window:

(1)  $Fd\bar{3}m$  space group, N°227, origin 1:

$$+(0,0,0) + (0, \frac{1}{2}, \frac{1}{2}) + (\frac{1}{2}, 0, \frac{1}{2}) + (\frac{1}{2}, \frac{1}{2}, 0)$$

(2)  $R\bar{3}c$  space group, N°167, second setting:

$$+(0,0,0) + (0,0,0) + (0,0,0)$$

## Symmetrical positions using the first Wyckoff position

The first step is to convert the list of coordinates from the first Wyckoff position on the form of matrix  $W$ :

(1)  $Fd\bar{3}m$  space group, N°227, origin 1:

$$(x, y, z) \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix}$$

(2)  $R\bar{3}c$  space group, N°167, second setting:

$$(x, y, z) \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix}$$

The  $W'$  matrix, suitable for the calculation, is calculated using the  $P^{-1}$  matrix for the corresponding space group:

$$W' = W \times P^{-1} \quad (D.3)$$

(1)  $Fd\bar{3}m$  space group, N°227, origin 1:

$$\begin{bmatrix} 1 & 0 & 0 & -\frac{1}{8} \\ 0 & 1 & 0 & -\frac{1}{8} \\ 0 & 0 & 1 & -\frac{1}{8} \end{bmatrix}$$

(2)  $R\bar{3}c$  space group, N°167, second setting:

$$\begin{bmatrix} 1 & 1 & 0 & 0 \\ -1 & 1 & 1 & 0 \\ 0 & -1 & 1 & 0 \end{bmatrix}$$

The results, converted back to standard coordinates, are presented in the Wyckoff positions table of the space group info window:

(1)  $Fd\bar{3}m$  space group, N°227, origin 1:

$$(x - \frac{1}{8}, y - \frac{1}{8}, z - \frac{1}{8})$$

...

(2)  $R\bar{3}c$  space group, N°167, second setting:

$$(x + z, -x + y + z, -y + z)$$

...

## Object position(s)

The first step in getting the final fractional coordinates  $F(f_x, f_y, f_z)$  of an object to insert with fractional coordinates given at input  $O(o_x, o_y, o_z)$  is to convert its coordinates to the proper coordinates system  $O'(o'_x, o'_y, o'_z)$ :

$$O'(o'_x, o'_y, o'_z) = P \times O(o_x, o_y, o_z) \quad (D.4)$$

Then for a particular Wyckoff position  $\mathbf{W}'$  and extra site position  $E(x, y, z)$ , the final fractional coordinates  $F(f_x, f_y, f_z)$  are calculated by:

$$F(f_x, f_y, f_z) = E(e_x, e_y, e_z) + \mathbf{W}' \times O'(o'_x, o'_y, o'_z) \quad (\text{D.5})$$

## Removing duplicate position(s)

Since one of the goal in **Atomes** is to build crystal structure of super-structure to work with, all crystalline positions cannot be conserved and duplicate positions (same positions using the periodic boundary conditions) must be removed.

When final positions  $F(f_x, f_y, f_z)$  are calculated duplicates are removed on the fly. The first final position with a unique set of coordinates  $F(f_x, f_y, f_z)$  is preserved while the others will be ignored. Duplicates are identical set of atomic coordinates, that can be found because of the symmetrical properties of the space group.

In **Atomes** a duplicate position  $F_b(b_x, b_y, b_z)$  will be removed if  $F_a(a_x, a_y, a_z)$  has already been found and providing that:

$$\begin{aligned} & \text{and} \quad \begin{bmatrix} a_x = b_x \text{ or } \text{abs}(a_x - b_x) = n, n \text{ integer} \\ a_y = b_y \text{ or } \text{abs}(a_y - b_y) = n, n \text{ integer} \\ a_z = b_z \text{ or } \text{abs}(a_z - b_z) = n, n \text{ integer} \end{bmatrix} \end{aligned}$$

## Calculate the object final position(s) in Cartesian coordinates

After the construction/selection process Cartesian coordinates  $C(c_x, c_y, c_z)$  of an object with fractional coordinates  $F(f_x, f_y, f_z)$  are calculated using:

$$C(c_x, c_y, c_z) = \mathbf{FtC} \times F(f_x, f_y, f_z) \quad (\text{D.6})$$

with  $\mathbf{FtC}$  the fractional to Cartesian conversion matrix is defined by:

$$\begin{bmatrix} \mathbf{a} & \mathbf{b} \cos \gamma & \mathbf{c} \cos \beta \\ 0 & \mathbf{b} \sin \gamma & \mathbf{c} \omega \\ 0 & 0 & \frac{\Omega}{\mathbf{a} \mathbf{b} \sin \gamma} \end{bmatrix} \quad (\text{D.7})$$

where  $\omega$  and  $\Omega$  are respectively:

$$\omega = \frac{\cos \alpha - \cos \beta \cos \gamma}{\sin \gamma} \quad (\text{D.8})$$

and:

$$\Omega = \mathbf{a} \mathbf{b} \mathbf{c} \sqrt{1 - \cos^2 \alpha - \cos^2 \beta - \cos^2 \gamma + 2 \cos \alpha \cos \beta \cos \gamma} \quad (\text{D.9})$$

$\mathbf{a}$ ,  $\mathbf{b}$ ,  $\mathbf{c}$  and  $\alpha$ ,  $\beta$ ,  $\gamma$  are the lattice parameters provided at input.

## Occupancy

The only thing left to do after that it to select which set of coordinates are to be occupied among the list prepared in the previous steps. Depending on the occupancy defined at input **Atomes** does the following:

- If the occupancy is equal to 1.0: all coordinates are to be occupied and no particular selection is performed.
- If the occupancy is  $< 1.0$ , then the following options are possible to handle occupancy:
  1. "Random for the initial cell only": sites are filled randomly in the initial cell only, then the initial cell is simply replicated.
  2. "Random cell by cell": sites are filled randomly for each cell, cell by cell separately.
  3. "Completely random": sites are filled randomly for the entire network, the final crystal is considered as whole.
  4. "Successively": sites are filled successively, all object(s) A are inserted (for the first  $n(A)$  positions), then all object(s) B are inserted (for the next  $n(B)$  positions) ...
  5. "Alternatively": sites are filled alternatively: object A is inserted on the first position, object B is inserted on the second position, object A on the third position, object B on the fourth position ... and so on.

For 1, 2, 4 and 5 the number of object(s) by cell is constant, but it can vary for 3. For 4 and 5 the order of the positions to be filled is the order they are found in during the previous steps.

## Final check: distance(s) between inserted objects

Finally the last calculation performed by **Atomes** when creating the crystal is distance check between the object(s) be inserted. Note that this check performed, if and only if, no *overlapping* is allowed.

If any distance  $d_{ab}$  between 2 objects  $a$  and  $b$ , is found to be *too small* the following can happen:

1. If  $d_{ab} < 0.5 \text{ \AA}$  then a warning message will pop-up and ask for the user confirmation to continue, or not, to build the crystal.
2. If the user decides to build the crystal anyway and if  $d_{ab} = 0.0$  then for fail safe reasons object  $b$  is removed from the list to build.

## The "SGL" "Space group info XML file" format

The data files that contain the information regarding each of the 230 space groups are located in the "bin/library/space\_groups" directory.

In this folder to each space group corresponds a file with a name following the construction:

"space group number"-"space group name".sgl

For the  $Fd\bar{3}m$  space group, the corresponding file is "227-Fd-3m.sgl".

These files must have the ".sgl" extension, follow the "XML" encoding rules, and the following structure:

```
<?xml version="1.0" encoding="UTF-8"?>
<!-- Space group info XML file -->
<sg-xml>
  <space-group>R-3c</space-group>
  <sg-num>167</sg-num>
  <hm-symbol>R -3 2/c</hm-symbol>
  <bravais>Trigonal</bravais>
  <settings num="2">
    <set name="R-3_c_c:h" x="a" y="b" z="c">
      <points num="3">
        <pt x="0" y="0" z="0"/>
        <pt x="2/3" y="1/3" z="1/3"/>
        <pt x="1/3" y="2/3" z="2/3"/>
      </points>
    </set>
    <set name="R-3_c_c:r" x="2/3a+1/3b+1/3c" y="-1/3a+1/3b+1/3c" z="-1/3a-2/3b+1/3c">
      <points num="3">
        <pt x="0" y="0" z="0"/>
        <pt x="0" y="0" z="0"/>
        <pt x="0" y="0" z="0"/>
      </points>
    </set>
  </settings>
  <wyckoff num="6">
    <wyck id="1" mul="12" let="f" site="1">
      <pos x="x" y="y" z="z"/>
      <pos x="-y" y="x-y" z="z"/>
      <pos x="-x+y" y="-x" z="z"/>
      <pos x="y" y="x" z="-z+1/2"/>
      <pos x="x-y" y="-y" z="-z+1/2"/>
      <pos x="-x" y="-x+y" z="-z+1/2"/>
      <pos x="-x" y="y" z="-z"/>
      <pos x="y" y="-x+y" z="-z"/>
      <pos x="x-y" y="x" z="-z"/>
      <pos x="-y" y="-x" z="z+1/2"/>
      <pos x="-x+y" y="y" z="z+1/2"/>
      <pos x="x" y="x-y" z="z+1/2"/>
    </wyck>
    <wyck id="2" mul="6" let="e" site=".2">
      <pos x="x" y="0" z="1/4"/>
      <pos x="0" y="x" z="1/4"/>
      <pos x="-x" y="-x" z="1/4"/>
      <pos x="-x" y="0" z="3/4"/>
      <pos x="0" y="-x" z="3/4"/>
      <pos x="x" y="x" z="3/4"/>
    </wyck>
    <wyck id="3" mul="6" let="d" site="-1">
      <pos x="1/2" y="0" z="0"/>
      <pos x="0" y="1/2" z="0"/>
      <pos x="1/2" y="1/2" z="0"/>
      <pos x="0" y="1/2" z="1/2"/>
      <pos x="1/2" y="0" z="1/2"/>
      <pos x="1/2" y="1/2" z="1/2"/>
    </wyck>
    <wyck id="4" mul="4" let="c" site="3.">
      <pos x="0" y="0" z="z"/>
      <pos x="0" y="0" z="-z+1/2"/>
      <pos x="0" y="0" z="-z"/>
      <pos x="0" y="0" z="z+1/2"/>
    </wyck>
    <wyck id="5" mul="2" let="b" site="-3.">
      <pos x="0" y="0" z="0"/>
      <pos x="0" y="0" z="1/2"/>
    </wyck>
    <wyck id="6" mul="2" let="a" site="32">
      <pos x="0" y="0" z="1/4"/>
      <pos x="0" y="0" z="3/4"/>
    </wyck>
  </wyckoff>
</sg-xml>
```

**Table D.1** Example of the "167-R-3c.sgl" file in XML format for the  $R\bar{3}c$  space group.



---

# The physics in Atomes

---

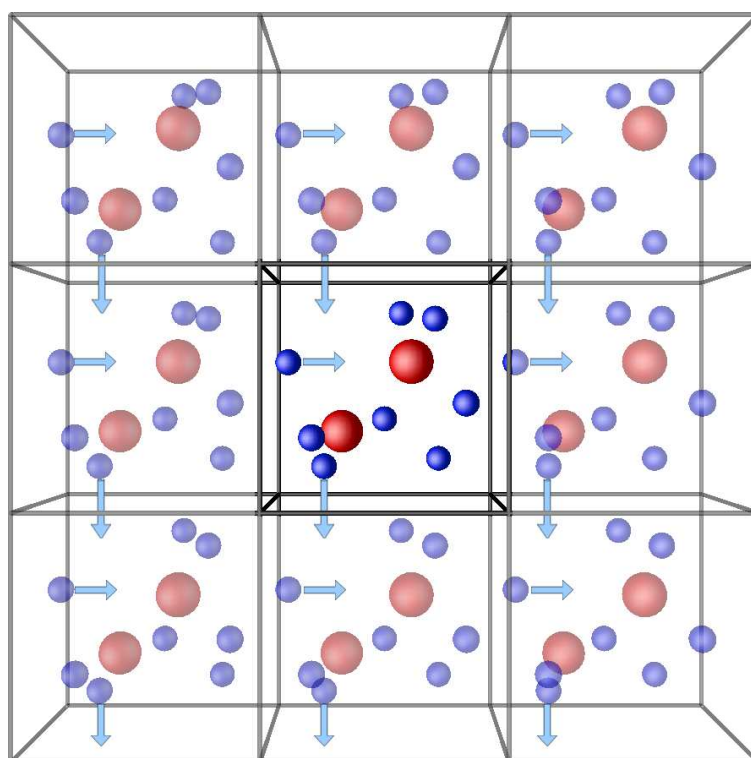
## E.1 The periodic boundary conditions

Taking into account the finite size of model/simulation box is crucial to compute correctly many of the structural characteristics (e.g. ring statistics) of the system being studied.

The importance of the finite size of model box can be illustrated using a 1  $dm^3$  edged cube of water (1 L) at room temperature. This cube contains approximately  $3.3 \times 10^{25}$  water molecules, each of them can be considered as a sphere having a diameter of 2.8 Å. Following this scheme surface interactions can affect up to 10 layers of spheres (water molecules) far from the surface of the model cubic box. In this case the number of water molecules exposed to the surface is about  $2 \times 10^{19}$ , which is a small fraction of the total number of molecules in the model.

Currently structure models often contain somewhere from 1 thousand to several thousands of molecules/atoms. As a result a very substantial fraction of them will be influenced by the finite size of the simulation/model box. The problems is solved by applying the so-called **Periodic Boundary Conditions** "PBC" which means surrounding the simulation box with its translational images in the 3 directions of space, as illustrated below.

Users of **Atomes** should take special care that their model boxes are inherently periodic so that when the periodic boundary conditions are applied the structural characteristics computed are not compromised.



**Figure E.1** *Schematic representation of the idea of periodic boundary conditions.*

Figure E.1 illustrates the principle of the periodic boundary conditions that can be used<sup>1</sup> in **Atomes**: a particle which goes out from the simulation box by one side is reintroduced in the box by the opposite side (in the 3 dimensions of space).

When PBC are used the maximum inter-atomic distance  $r_{max}$  which is taken into account in the calculations, depends on the lattice parameters:

$$r_{max} \approx \frac{L \times \sqrt{3}}{2} \quad \text{with } L = \text{box size} \quad (\text{E.1})$$

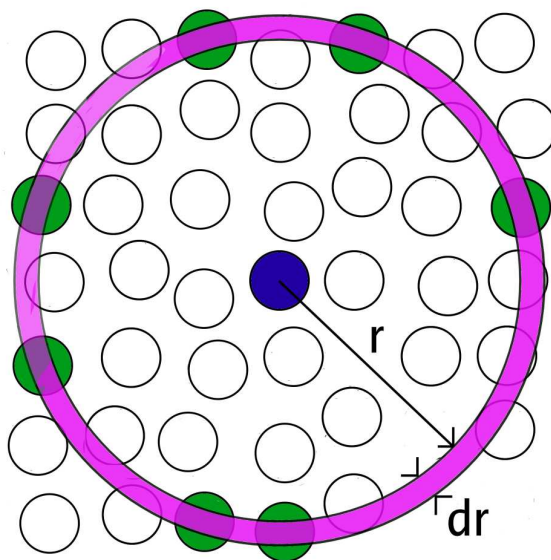
The surface/finite model size effects would therefore be small, if any. In general, the larger the simulation box and the number of molecules/atoms in it, the smaller the surface/size effects will be.

---

<sup>1</sup>Please note that the use of PBC is not mandatory, isolated molecules can be studied using **Atomes**

## E.2 Radial distribution functions fundamentals

The Radial Distribution Function, R.D.F. ,  $g(r)$ , also called pair distribution function or pair correlation function, is an important structural characteristic, therefore computed by **Atomes**.



**Figure E.2** Space discretization for the evaluation of the radial distribution function.

Considering a homogeneous distribution of the atoms/molecules in space, the  $g(r)$  represents the probability to find an atom in a shell  $dr$  at the distance  $r$  of another atom chosen as a reference point [Fig. E.2]. By dividing the physical space/model volume into shells  $dr$  [Fig. E.2] it is possible to compute the number of atoms  $dn(r)$  at a distance between  $r$  and  $r + dr$  from a given atom:

$$dn(r) = \frac{N}{V} g(r) 4\pi r^2 dr \quad (\text{E.2})$$

where  $N$  represents the total number of atoms,  $V$  the model volume and where  $g(r)$  is the radial distribution function. In this notation the volume of the shell of thickness  $dr$  is approximated:

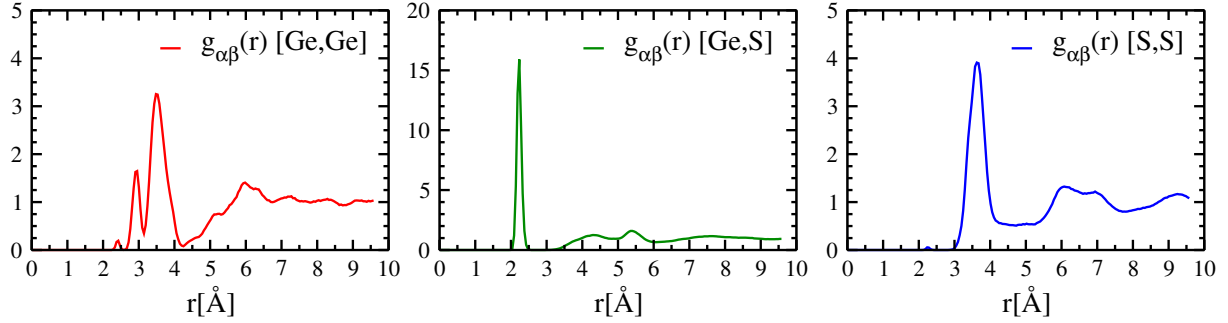
$$\left( V_{\text{shell}} = \frac{4}{3}\pi(r+dr)^3 - \frac{4}{3}\pi r^3 \simeq 4\pi r^2 dr \right) \quad (\text{E.3})$$

When more than one chemical species are present the so-called partial radial distribution functions  $g_{\alpha\beta}(r)$  may be computed :

$$g_{\alpha\beta}(r) = \frac{dn_{\alpha\beta}(r)}{4\pi r^2 dr \rho_\alpha} \quad \text{with} \quad \rho_\alpha = \frac{V}{N_\alpha} = \frac{V}{N \times c_\alpha} \quad (\text{E.4})$$

where  $c_\alpha$  represents the concentration of atomic species  $\alpha$ .

These functions give the density probability for an atom of the  $\alpha$  species to have a neighbor of the  $\beta$  species at a given distance  $r$ . The example features  $\text{GeS}_2$  glass.



**Figure E.3** *Partial radial distribution functions of glassy  $\text{GeS}_2$  at 300 K.*

Figure E.3 shows the partial radial distribution functions for  $\text{GeS}_2$  glass at 300 K. The total RDF of a system is a weighted sum of the respective partial RDFs, with the weights depend on the relative concentration and x-ray/neutron scattering amplitudes of the chemical species involved.

It is also possible to use the reduced  $\mathbf{G}_{\alpha\beta}(r)$  partial distribution functions defined as:

$$\mathbf{G}_{\alpha\beta}(r) = 4\pi r \rho_0 (g_{\alpha\beta}(r) - 1) \quad (\text{E.5})$$

**Atomes** gives access to:

- The partial  $g_{\alpha\beta}(r)$  and  $\mathbf{G}_{\alpha\beta}(r)$  distribution functions, and more see [Eq.E.14].
- The corresponding  $dn_{\alpha\beta}(r)$  integrated number of neighbors.

Also two methods are available to compute the radial distribution functions:

- The standard real space calculation typical to analyze 3-dimensional models
- The experiment-like calculation using the Fourier transform of the structure factor obtained using the Debye equation (see section E.3 for details).

## E.3 Neutrons and X-rays scattering

Model static structure factors  $S(q)$  may be compared to experimental scattering data and that is why are useful structural characteristics computed by **Atomes**. Thereafter we describe the theoretical background of  $S(q)$ s computed by **Atomes**.

### E.3.1 Total scattering - Debye approach

Neutron or X-ray scattering static structure factor  $S(q)$  is defined as:

$$S(q) = \frac{1}{N} \sum_{j,k} b_j b_k \langle e^{iq[\mathbf{r}_j - \mathbf{r}_k]} \rangle \quad (\text{E.6})$$

where  $b_j$  and  $\mathbf{r}_j$  represent respectively the neutron or X-ray scattering length, and the position of the atom  $j$ .  $N$  is the total number of atoms in the system studied.

To take into account the inherent/volume averaging of scattering experiments it is necessary to sum all possible orientations of the wave vector  $q$  compared to the vector  $\mathbf{r}_j - \mathbf{r}_k$ . This average on the orientations of the  $q$  vector leads to the famous Debye's equation:

$$S(q) = \frac{1}{N} \sum_{j,k} b_j b_k \frac{\sin(q|\mathbf{r}_j - \mathbf{r}_k|)}{q|\mathbf{r}_j - \mathbf{r}_k|} \quad (\text{E.7})$$

Nevertheless the instantaneous individual atomic contributions introduced by this equation E.7 are not easy to interpret. It is more interesting to express these contributions using the formalism of radial distribution functions [Sec. E.2].

In order to achieve this goal it is first necessary to split the self-atomic contribution ( $j = k$ ), from the contribution between distinct atoms:

$$S(q) = \sum_j c_j b_j^2 + \underbrace{\frac{1}{N} \sum_{j \neq k} b_j b_k \frac{\sin(q|\mathbf{r}_j - \mathbf{r}_k|)}{q|\mathbf{r}_j - \mathbf{r}_k|}}_{I(q)} \quad \text{with } c_j = \frac{N_j}{N} \quad (\text{E.8})$$

where  $4\pi \sum_j c_j b_j^2$  represents the total scattering cross section of the material.

The function  $I(q)$  which describes the interaction between distinct atoms is related to the radial distribution functions through a Fourier transformation:

$$I(q) = 4\pi\rho \int_0^\infty dr r^2 \frac{\sin qr}{qr} G(r) \quad (\text{E.9})$$

where the function  $G(r)$  is defined using the partial radial distribution functions [Eq. E.4]:

$$G(r) = \sum_{\alpha,\beta} c_\alpha b_\alpha c_\beta b_\beta (g_{\alpha\beta}(r) - 1) \quad (\text{E.10})$$

where  $c_\alpha = \frac{N_\alpha}{N}$  and  $b_\alpha$  represents the neutron or X-ray scattering length of species  $\alpha$ .

$G(r)$  approaches  $-\sum_{\alpha,\beta} c_\alpha b_\alpha c_\beta b_\beta$  for  $r = 0$ , and 0 for  $r \rightarrow \infty$ .

Usually the self-contributions are subtracted from equation E.8 and the structure factor is normalized using the relation:

$$S(q) - 1 = \frac{I(q)}{\langle b^2 \rangle} \quad \text{with} \quad \langle b^2 \rangle = \left( \sum_{\alpha} c_{\alpha} b_{\alpha} \right)^2 \quad (\text{E.11})$$

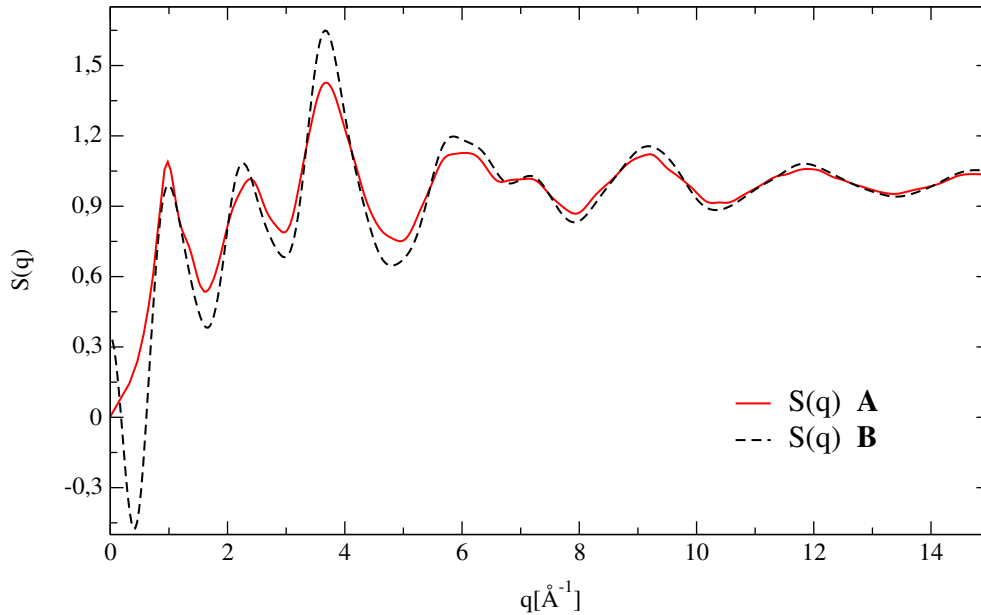
It is therefore possible to write the structure factor [Eq. E.7] in a more standard way:

$$S(q) = 1 + 4\pi\rho \int_0^{\infty} dr r^2 \frac{\sin qr}{qr} (\mathbf{g}(r) - 1) \quad (\text{E.12})$$

where  $\mathbf{g}(r)$  (the radial distribution function) is defined as:

$$\mathbf{g}(r) = \frac{\sum_{\alpha,\beta} c_{\alpha} b_{\alpha} c_{\beta} b_{\beta} g_{\alpha\beta}(r)}{\langle b^2 \rangle} \quad (\text{E.13})$$

In the case of a single atomic species system the normalization allows to obtain values of  $S(q)$  and  $\mathbf{g}(r)$  which are independent of the scattering factor/length and therefore independent of the measurement technique. In most cases, however, the total  $S(q)$  and  $\mathbf{g}(r)$  are combinations of the partial functions weighted using the scattering factor and therefore depend on the measurement technique (Neutron, X-rays ...) used or simulated.



**Figure E.4** Total neutron structure factor for glassy  $\text{GeS}_2$  at 300 K - **A** Evaluation using the atomic correlations [Eq. E.7], **B** Evaluation using the pair correlation functions [Eq. E.12].

Figure E.4 presents a comparison between the calculations of the total neutron structure factor done using the Debye relation [Eq. E.7] and the pair correlation functions

[Eq. E.12]. The material studied is a sample of glassy  $\text{GeS}_2$  at 300 K obtained using ab-initio molecular dynamics. In several cases the structure factor  $S(q)$  and the radial distribution function  $g(r)$  [Eq. E.13] can be compared to experimental data. To simplify the comparison **Atomes** computes several radial distribution functions used in practice such as  $G(r)$  defined [Eq. E.10], the differential correlation function  $D(r)$ ,  $\mathbf{G}(r)$ , and the total correlation function  $T(r)$  defined by:

$$\begin{aligned} D(r) &= 4\pi r \rho G(r) \\ \mathbf{G}(r) &= \frac{D(r)}{\langle b^2 \rangle} \\ T(r) &= D(r) + 4\pi r \rho \langle b^2 \rangle \end{aligned} \tag{E.14}$$

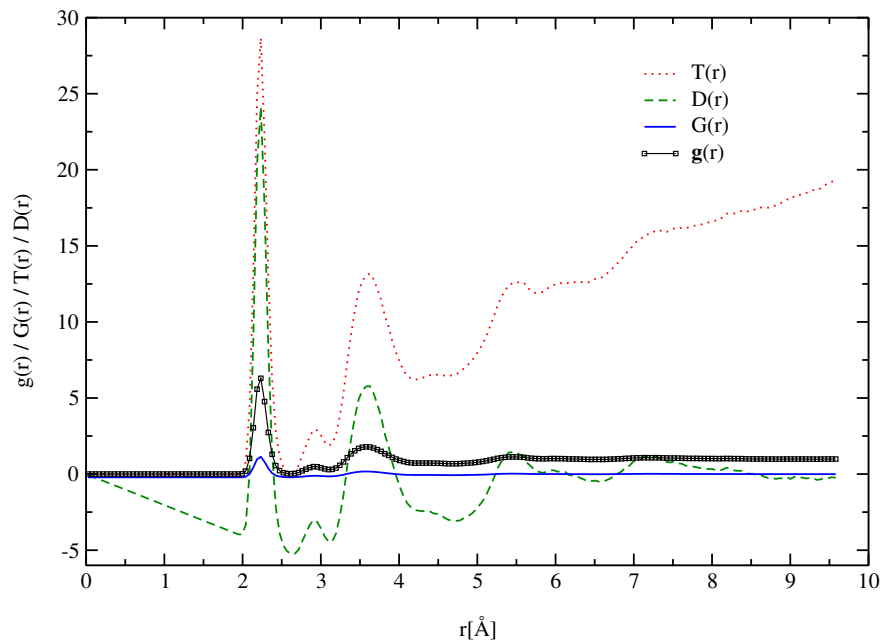
$g(r)$  equals zero for  $r = 0$  and approaches 1 for  $r \rightarrow \infty$ .

$D(r)$  equals zero for  $r = 0$  and approaches 0 for  $r \rightarrow \infty$ .

$\mathbf{G}(r)$  equals zero for  $r = 0$  and approaches 0 for  $r \rightarrow \infty$ .

$T(r)$  equals zero for  $r = 0$  and approaches  $\infty$  for  $r \rightarrow \infty$ .

This set of functions for a model of  $\text{GeS}_2$  glass (at 300 K) obtained using ab-initio molecular dynamics is presented in figure E.5.



**Figure E.5** *Example of various distribution functions neutron-weighted in glassy  $\text{GeS}_2$  at 300 K.*

**Atomes** can compute, for the case of **x-rays** and/or **neutrons**, the following functions:

- $S(q)$  and  $Q(q) = q[S(q) - 1.0]$  [12, 13] computed using the Debye equation
- $S(q)$  and  $Q(q) = q[S(q) - 1.0]$  [12, 13] computed using the Fourier transform of  $g(r)$
- $g(r)$ ,  $G(r)$ ,  $D(r)$  and  $T(r)$  computed using the standard real space calculation
- $g(r)$ ,  $G(r)$ ,  $D(r)$  and  $T(r)$  computed using the Fourier transform of Debye  $S(q)$

### E.3.2 Partial structure factors

There are a few, somewhat different definitions of partials  $S(q)$  used in practice, and computed by **Atomes**

#### E.3.2.1 Faber-Ziman definition/formalism

One way used to define the partial structure factors has been proposed by Faber and Ziman [14]. In this approach the structure factor is represented by the correlations between the different chemical species. To describe the correlation between the  $\alpha$  and the  $\beta$  chemical species the partial structure factor  $S_{\alpha\beta}^{FZ}(q)$  is defined by:

$$S_{\alpha\beta}^{FZ}(q) = 1 + 4\pi\rho \int_0^\infty dr r^2 \frac{\sin qr}{qr} (g_{\alpha\beta}(r) - 1) \quad (\text{E.15})$$

where the  $g_{\alpha\beta}(r)$  are the partial radial distribution functions [Eq. E.4].

The total structure factor is then obtained by the relation:

$$S(q) = \sum_{\alpha,\beta} c_\alpha b_\alpha c_\beta b_\beta [S_{\alpha\beta}^{FZ}(q) - 1] \quad (\text{E.16})$$

#### E.3.2.2 Ashcroft-Langreth definition/formalism

In a similar approach, based on the correlation between the chemical species, and developed by Ashcroft and Langreth [15, 16, 17], the partial structure factors  $S_{\alpha\beta}^{AL}(q)$  are defined by:

$$S_{\alpha\beta}^{AL}(q) = \delta_{\alpha\beta} + 4\pi\rho (c_\alpha c_\beta)^{1/2} \int_0^\infty dr r^2 \frac{\sin qr}{qr} (g_{\alpha\beta}(r) - 1) \quad (\text{E.17})$$

where  $\delta_{\alpha\beta}$  is the Kronecker delta,  $c_\alpha = \frac{N_\alpha}{N}$ , and the  $g_{\alpha\beta}(r)$  are the partial radial distribution functions [Eq. E.4].

Then the total structure factor can be calculated using:

$$S(q) = \frac{\sum_{\alpha,\beta} b_\alpha b_\beta (c_\alpha c_\beta)^{1/2} [S_{\alpha\beta}^{AL}(q) + 1]}{\sum_\alpha c_\alpha b_\alpha^2} \quad (\text{E.18})$$



### E.3.2.3 Bhatia-Thornton definition/formalism

In this approach, used in the case of binary systems  $AB_x$  [18] only, the total structure factor  $S(q)$  can be express as the weighted sum of 3 partial structure factors:

$$S(q) = \frac{\langle b \rangle^2 S_{NN}(q) + 2\langle b \rangle(b_A - b_B)S_{NC}(q) + (b_A - b_B)^2 S_{CC}(q) - (c_A b_A^2 + c_B b_B^2)}{\langle b \rangle^2} + 1 \quad (E.19)$$

where  $\langle b \rangle = c_A b_A + c_B b_B$ , with  $c_A$  and  $b_A$  representing respectively the concentration and the scattering length of species A.

$S_{NN}(q)$ ,  $S_{NC}(q)$  and  $S_{CC}(q)$  represent combinations of the partial structure factors calculated using the Faber-Ziman formalism and weighted using the concentrations of the 2 chemical species:

$$S_{NN}(q) = \sum_{A=1}^2 \sum_{B=1}^2 c_A c_B S_{AB}^{FZ}(q) \quad (E.20)$$

$$S_{NC}(q) = c_A c_B \times \left[ c_A \times (S_{AA}^{FZ}(q) - S_{AB}^{FZ}(q)) - c_B \times (S_{BB}^{FZ}(q) - S_{AB}^{FZ}(q)) \right] \quad (E.21)$$

$$S_{CC}(q) = c_A c_B \times \left[ 1 + c_A c_B \times \left[ \sum_{A=1}^2 \sum_{B \neq A}^2 (S_{AA}^{FZ}(q) - S_{AB}^{FZ}(q)) \right] \right] \quad (E.22)$$

- $S_{NN}(q)$  is the Number-Number partial structure factor.

Its Fourier transform allows to obtain a global description of the structure of the solid, ie. of the distribution of the experimental scattering centers, or atomic nuclei, positions. The nature of the chemical species spread in the scattering centers is not considered. Furthermore if  $b_A = b_B$  then  $S_{NN}(q) = S(q)$ .

- $S_{CC}(q)$  is the Concentration-Concentration partial structure factor.

Its Fourier transform allows to obtain an idea of the distribution of the chemical species over the scattering centers described using the  $S_{NN}(q)$ . Therefore the  $S_{CC}(q)$  describes the chemical order in the material. In the case of an ideal binary mixture of 2 chemical species A and B<sup>2</sup>,  $S_{CC}(q)$  is constant and equal to  $c_A c_B$ . In the case of an ordered chemical mixture (chemical species with distinct diameters, and with heteropolar and homopolar chemical bonds) it is possible to link the variations of the  $S_{CC}(q)$  to the product of the concentrations of the 2 chemical species of the mixture:

- $S_{CC}(q) = c_A c_B$ : random distribution.
- $S_{CC}(q) > c_A c_B$ : homopolar atomic correlations (A-A, B-B) preferred.
- $S_{CC}(q) < c_A c_B$ : heteropolar atomic correlations (A-B) preferred.
- $\langle b \rangle = 0$ :  $S_{CC}(q) = S(q)$ .

---

<sup>2</sup>Particles that can be described using spheres of the same diameter and occupying the same molar volume, subject to the same thermal constraints, in a mixture where the substitution energy of a particle by another is equal to zero.

- $S_{NC}(q)$  is the Number-Concentration partial structure factor.

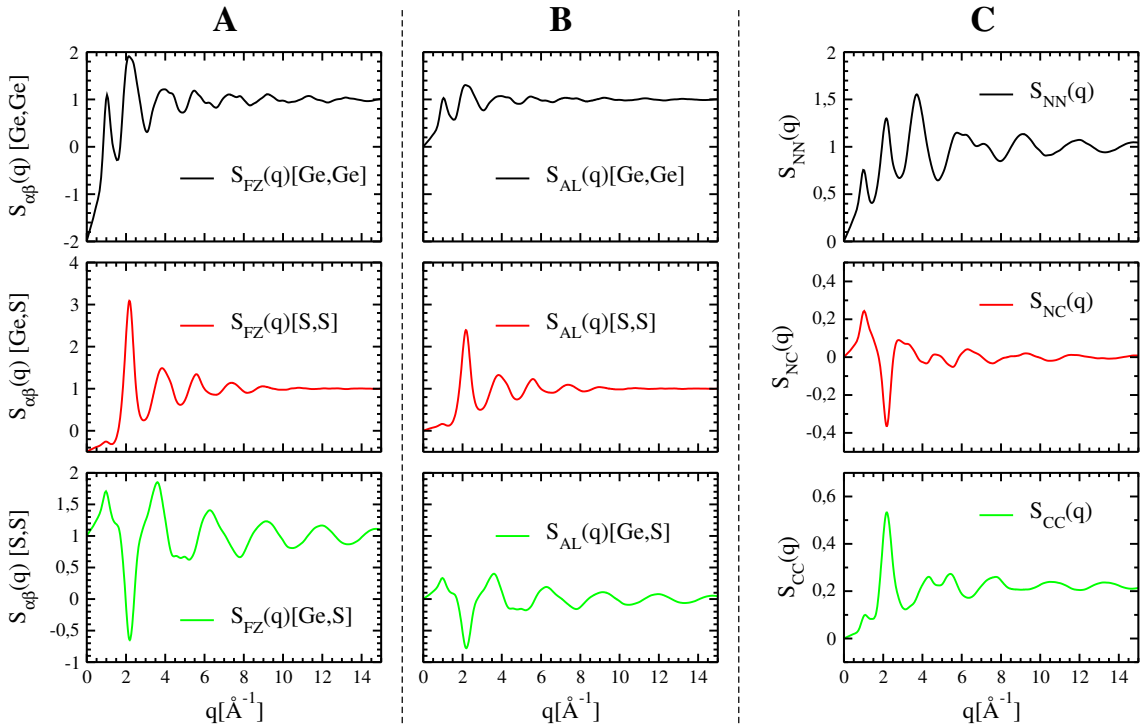
Its Fourier transform allows to obtain a correlation between the scattering centers and their occupation by a given chemical species. The more the chemical species related partial structure factors are different ( $S_{AA}(q) \neq S_{BB}(q)$ ) and the more the oscillations are important in the  $S_{NC}(q)$ . In the case of an ideal mixture  $S_{NC}(q) = 0$ , and all the information about the structure of the system is given by the  $S_{NN}(q)$ .

If we consider the binary mixture as an ionic mixture then it is possible to calculate the Charge-Charge  $S_{ZZ}(q)$  and the Number-Charge  $S_{NZ}(q)$  partial structure factors using the Concentration-Concentration  $S_{CC}(q)$  and the Number-Concentration  $S_{NC}(q)$ :

$$S_{ZZ}(q) = \frac{S_{CC}(q)}{c_A c_B} \quad \text{and} \quad S_{NZ}(q) = \frac{S_{NC}(q)}{c_B / Z_A} \quad (\text{E.23})$$

$c_A$  and  $Z_A$  represent the concentration and the charge of the chemical species A, the global neutrality of the system must be respected therefore  $c_A Z_A + c_B Z_B = 0$ .

Figure E.6 illustrates, and allows to compare, the partial structure factors of glassy  $\text{GeS}_2$  at 300 K calculated in the different formalism Faber-Ziman [14], Ashcroft-Langreth [15, 16, 17], and Bhatia-Thornton [18].



**Figure E.6** Partial structure factors of glassy  $\text{GeS}_2$  at 300 K. **A** Faber-Ziman [14], **B** Ashcroft-Langreth [15, 16, 17] and **C** Bhatia-Thornton [18].

**Atomes** can compute all types of partial structure factors: Faber-Ziman  $S_{\alpha\beta}^{FZ}(q)$ , Ashcroft-Langreth  $S_{\alpha\beta}^{AL}(q)$  and Bhatia-Thornton  $S_{NN}(q)$ ,  $S_{NC}(q)$ ,  $S_{CC}(q)$  and  $S_{ZZ}(q)$ .

## E.4 Local atomic coordination properties

Several properties related to the atomic bonds and angles between them can be computed using **Atomes**. The existence or the absence of a bond between two atoms  $i$  of species  $\alpha$  and  $j$  of species  $\beta$  is determined by the analysis of the partial  $g_{\alpha\beta}(r)$  and total  $g(r)$  radial distribution functions. Precisely **Atomes** will consider that a bond exists if the interatomic distance  $D_{ij}$  is smaller than both the cutoff given to describe the maximum distance for first neighbor atoms between the species  $\alpha$  and  $\beta$ ,  $Rcut_{\alpha\beta}$  (often the first minimum of  $g_{\alpha\beta}(r)$ ), and the first minimum of the total radial distribution function,  $Rcut_{tot}$ .

**Atomes** allows the user to specify both  $Rcut_{\alpha\beta}$  and  $Rcut_{tot}$  to choose an appropriate description of the atomic bonds in the system under study. When atomic bonds in a model are defined properly other structural characteristics can be evaluated, as follows:

### E.4.1 Average first coordination numbers

**Atomes** computes total as well as partials coordination numbers.

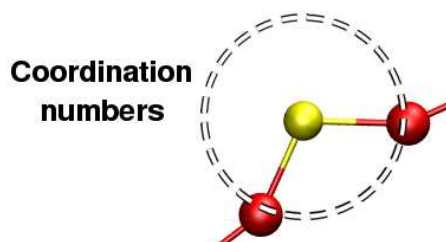


Figure E.7 Coordination numbers.

### E.4.2 Individual atomic neighbor analysis

**Atomes** computes the fraction of each type of coordination spheres in the model. The presence of structural defects can lead to a wide number of local environments, figure E.8 illustrates the different coordination spheres found in a  $\text{GeS}_2$  glass.

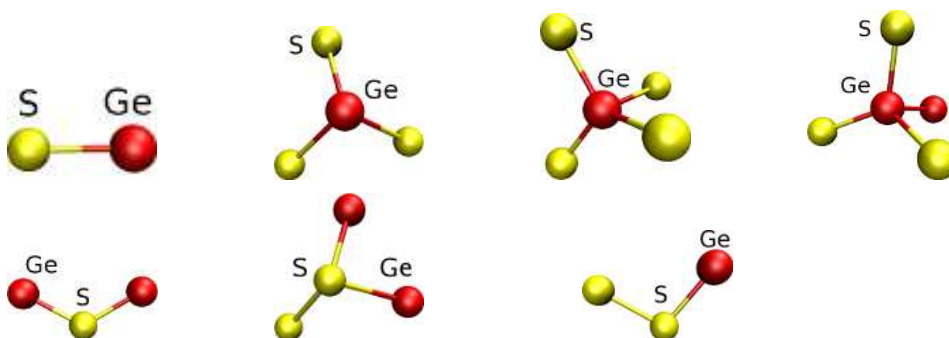
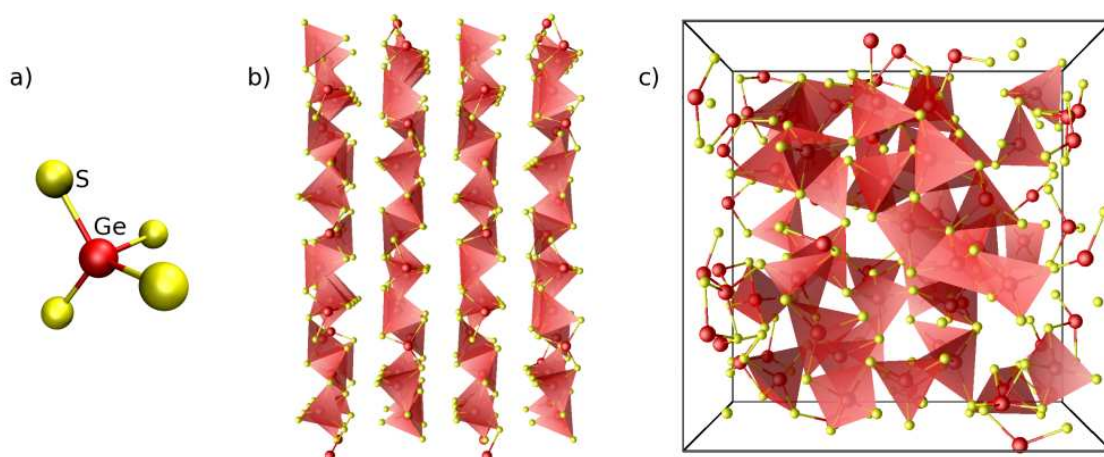


Figure E.8 Illustration of several coordination spheres that can be found in glassy  $\text{GeS}_2$ .

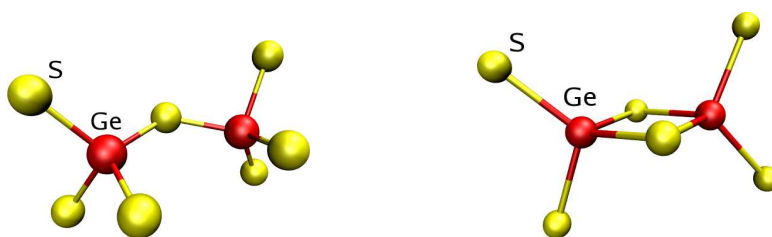
### E.4.3 Proportion of tetrahedral links and units in the structure model

Often the structure of a material is represented using building blocks. One of the most frequently occurring building blocks are tetrahedra. Figure E.9 shows a model of  $\text{GeS}_2$  materials using  $\text{GeS}_4$  tetrahedra as building blocks.



**Figure E.9** *Illustration of the presence of  $\text{GeS}_4$  tetrahedra in the  $\text{GeS}_2$  material's family. a)  $\text{GeS}_4$  tetrahedra, representations b) of the  $\alpha\text{-GeS}_2$  crystal and c) of the  $\text{GeS}_2$  glass using tetrahedra.*

**Atomes** computes the fraction of the different tetrahedra in materials, the distinction between these tetrahedra being made on the nature of the connection between each of them. Tetrahedra can be linked either by corners or edges [Fig. E.10], **Atomes** computes the fraction of atoms forming tetrahedra as well as to the fraction of linked tetrahedra.



**Figure E.10** *Corner sharing (left) and edge sharing (right) tetrahedra.*

### E.4.4 Distribution of bond lengths for the first coordination spheres

**Atomes** gives access to the bond length distribution between first neighbor atoms [Fig. E.11]:

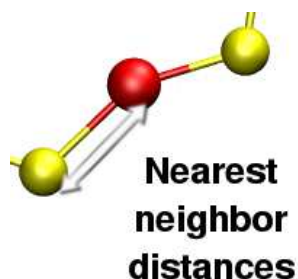


Figure E.11 *Nearest neighbor distances distribution.*

### E.4.5 Angles distributions

**Atomes** also computes the distributions of bond angles and dihedral angles [Fig. E.12]:

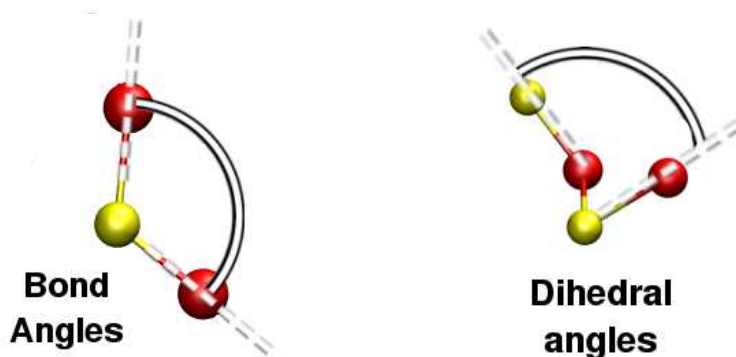


Figure E.12 *Bond angles (left) and diehdral angles (right).*

## E.5 Ring statistics

The analysis of the topology of network-type structure models (liquid, crystalline or amorphous systems) is often based on the part of the structural information which can be represented in the graph theory using nodes for the atoms and links for the bonds. The absence or the existence of a link between two nodes is determined by the analysis of the total and partial radial distribution functions of the system.

In such a network a series of nodes and links connected sequentially without overlap is called a path. Following this definition a ring is therefore simply a closed path. If we study thoroughly a specific node of this network we see that this node can be involved in numerous rings. Each of these rings is characterized by its size and can be classified based upon the relations between the nodes and the links which constitute it.

### E.5.1 Size of the rings

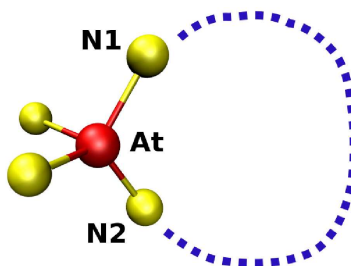
There are two possibilities for the numbering of rings. On the one hand, one can use the total number of nodes of the ring, therefore a N-membered ring is a ring containing N nodes. On the other hand, one can use the number of *network forming* nodes (ex: Si atoms in SiO<sub>2</sub> and Ge atoms in GeS<sub>2</sub> which are the atoms of highest coordination in these materials) an N-membered ring is therefore a ring containing 2×N nodes. For crystals and SiO<sub>2</sub>-like glasses the second definition is usually applied. Nevertheless the first method has to be used in the case of chalcogenide liquids and glasses in order to count rings with homopolar bonds (ex: Ge-Ge and S-S bonds in GeS<sub>2</sub>) - See section E.5.5 for further details.

From a theoretical point of view it is possible to obtain an estimate for the ring of maximum size that could exist in a network. This theoretical maximum size will depend on the properties of the system studied as well as on the definition of a ring.

### E.5.2 Definitions

#### E.5.2.1 King's shortest paths criterion

The first way to define a ring has been given by Shirley V. King [19] (and later by Franzblau [20]). In order to study the connectivity of glassy SiO<sub>2</sub> she defines a ring as the shortest path between two of the nearest neighbors of a given node [Fig. E.13].



**Figure E.13** *King's criterion in the ring statistics: a ring represents the shortest path between two of the nearest neighbors (N1 and N2) of a given node (At).*

In the case of the King's criterion one can calculate the maximum number of different ring sizes,  $NS_{max}(KSP)$ , which can be found using the atom **At** to initiate the search:

$$NS_{max}(KSP) = \frac{N_c(\mathbf{At}) \times (N_c(\mathbf{At}) - 1)}{2} \quad (\text{E.24})$$

where  $N_c(\mathbf{At})$  is the number of neighbors of atom **At**.  $NS_{max}(KSP)$  represents the number of ring sizes found if all couples of neighbors of atom **At** are connected together with paths of different sizes.

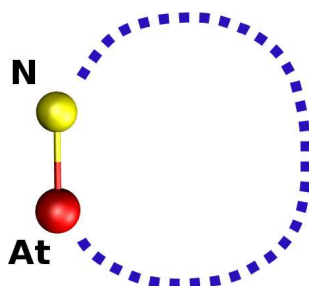
It is also possible to calculate the theoretical maximum size,  $TMS(KSP)$ , of a King's shortest path ring in the network using:

$$TMS(KSP) = 2 \times (D_{max} - 2) \times (N_{c_{max}} - 2) + 2 \times D_{max} \quad (\text{E.25})$$

where  $D_{max}$  is the longest distance, in number of chemical bonds, separating two atoms in the network, and  $N_{c_{max}}$  represents the average number of neighbors of the chemical species of higher coordination. If used when looking for rings, periodic boundary conditions have to be taken into account to calculate  $D_{max}$ . The relation [Eq. E.25] is illustrated in figure E.16-2).

### E.5.2.2 Guttman's shortest paths criterion

A later definition of ring was proposed by Guttman [21], who defines a ring as the shortest path which comes back to a given node (or atom) from one of its nearest neighbors [Fig. E.14].



**Figure E.14** Guttman's criterion in the ring statistics: a ring represents the shortest path which comes back to a given node (**At**) from one of its nearest neighbors (**N**).

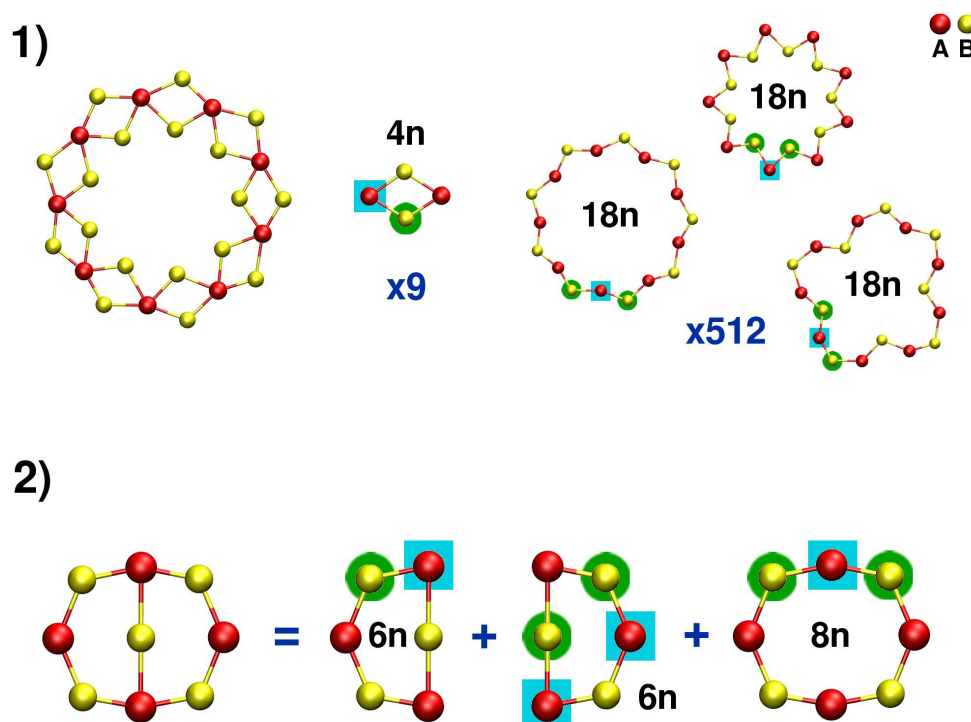
Differences between the King and the Guttman's shortest paths criteria are illustrated in figure E.15.

Like for the King's criterion, with the Guttman's criterion one can calculate the maximum number of different ring sizes,  $NS_{max}(GSP)$ , which can be found using the atom **At** to initiate the search:

$$NS_{max}(GSP) = N_c(\mathbf{At}) - 1 \quad (\text{E.26})$$

where  $N_c(\mathbf{At})$  is the number of neighbors of atom **At**.  $NS_{max}(GSP)$  represents the number of ring sizes found if the neighbors of atom **At** are connected together with paths of different sizes.





**Figure E.15** Differences between the King and the Guttman shortest paths criteria for the ring statistics in an  $AB_2$  system. In these two examples the search is initiated from chemical species A (blue square). The nearest neighbor(s) of chemical species B (green circles) are used to continue the analysis. 1) In the first example only rings with 4 nodes are found using the Guttman's criterion, whereas rings with 18 nodes are also found using the King's criterion ( $2^9$  rings with 18 nodes). 2) In the second example the King's shortest path criterion allows to find the ring with 8 nodes ignored by the Guttman's criterion which is only able to find the rings with 6 nodes.

It is also possible to calculate the Theoretical Maximum Size,  $TMS(GSP)$ , of a Guttman's ring in the network using:

$$TMS(GSP) = 2 \times D_{max} \quad (E.27)$$

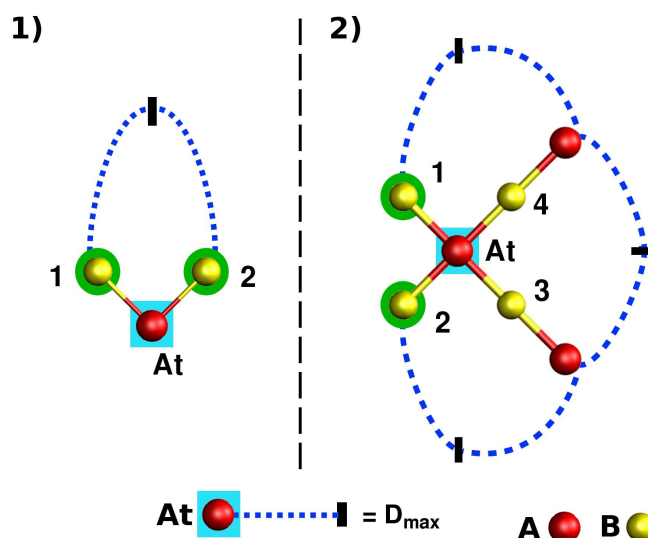
where  $D_{max}$  represents the longest distance, in number of chemical bonds, separating two atoms in the network. If used when looking for rings, periodic boundary conditions have to be taken into account to calculate  $D_{max}$ . The relation [Eq. E.27] is illustrated in figure E.16-1.

Since the introduction of the King's and the Guttman's criteria other definitions of rings have been proposed. These definitions are based on the properties of the rings to be decomposed into the sum of smaller rings.

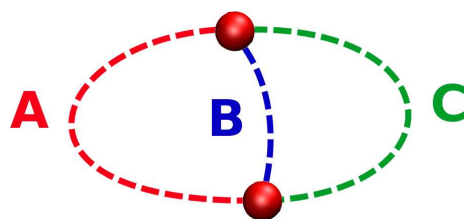
### E.5.2.3 The primitive rings criterion

A ring is primitive [22, 23] (or Irreducible [24]) if it can not be decomposed into two smaller rings [Fig. E.17].





**Figure E.16** Theoretical maximum size of the rings for an  $AB_2$  system ( $N_{c_{max}} = N_{c_A} = 4$ ) and using: 1) the Guttman's criterion, 2) the King's criterion. The theoretical maximum size represent the longest distance between two nearest neighbors 1 and 2 (green circles) of the atom **At** used to initiate the search (blue square).



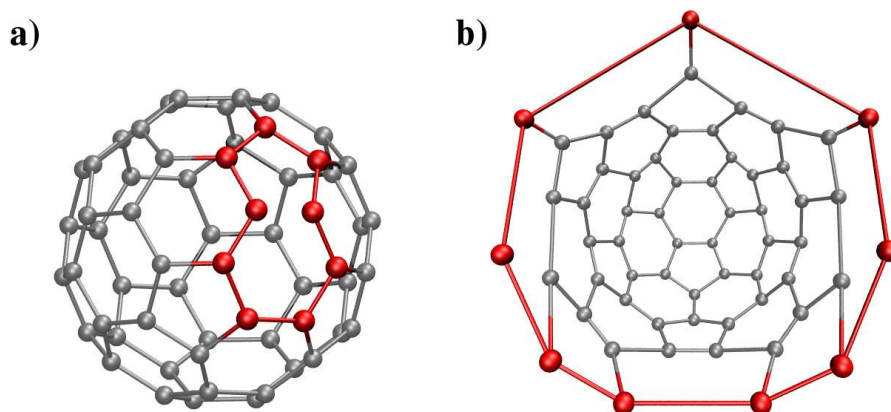
**Figure E.17** Primitive rings in the ring statistics: the 'AC' ring defined by the sum of the A and the C paths is primitive only if there is no B path shorter than A and shorter than C which allows to decompose the 'AC' ring into two smaller rings 'AB' and 'BC'.

The primitive rings analysis between the paths in figure E.17 may lead to 3 results depending on the relations between the paths A, B, and C:

- If paths A, B, and C have the same length:  $A = B = C$  then the rings 'AB', 'AC' and 'BC' are primitives.
- If the relation between the paths is like  $? = ? < ?$  (ex:  $A = B < C$ ) then 1 smaller ring ('AB') and 2 bigger rings ('AC' and 'BC') exist. None of these rings can be decomposed into the sum of two smaller rings therefore the 3 rings are again primitives.
- If the relation between the path is like  $? < ? = ?$  (ex:  $A < B = C$ ) or  $? < ? < ?$  (ex:  $A < B < C$ ) then a shortest path exists (A). It will be possible to decompose the ring ('BC') built without this shortest path into the sum of 2 smaller rings ('AB' and 'AC'), therefore this ring will not be primitive.

#### E.5.2.4 The strong rings criterion

The strong rings [22, 23] are defined by extending the definition of primitive rings. A ring is strong if it can not be decomposed into a sum of smaller rings whatever this sum is, ie. whatever the number of paths in the decomposition is.



**Figure E.18** *Strong rings in the ring statistics: a) the 9-carbon-atoms ring created after breaking a C-C bond in a Buckminster fullerene molecule is a counterexample of strong ring; b) the combination of shortest rings, 11 5-carbon-atoms rings and 19 6-carbon-atoms rings, appears easily after the deformation of the C<sub>60</sub> molecule.*

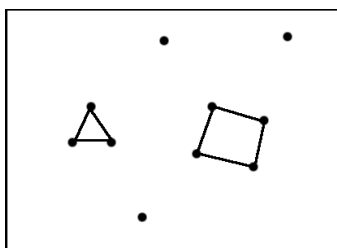
By definition the strong rings are also primitives, therefore to search for strong rings can be summed as to find the strong rings among the primitive rings. This technique is limited to relatively simple cases, like crystals or structures such as the one illustrated in figure E.18. On the one hand the CPU time needed to complete such an analysis for amorphous systems is very important. On the other hand it is not possible to search for strong rings using the same search depth than for other types of rings. The strong ring analysis is indeed diverging which makes it very complex to implement for amorphous materials.

In the case of primitive rings like in the case of strong rings, there is no theoretical maximum size of rings in the network.

### E.5.3 Description of a network using ring statistics - existing tools

Ring statistics are mainly used to obtain a snapshot of the connectivity of a network. Thereby the better the snapshot will be, the better the description and the understanding of the properties of the material will be. In the literature many papers present studies of materials using ring statistics. In these studies either the number of Rings per Node ' $R_N$ ' [34, 35] or the number of Rings per Cell ' $R_C$ ' [36, 37, 38] are given as a result of the analysis. The first ( $R_N$ ) is calculated for one node by counting all the rings corresponding to the property we are looking for (King's, Guttman's, primitive or strong ring criterion). The second ( $R_C$ ) is calculated by counting all the different rings corresponding at least once (at least for one node) to the property we are looking for (King's, shortest path, primitive or strong ring criterion). The values of  $R_N$  and  $R_C$  are often reduced to the number of nodes of the networks. Furthermore the results are presented according to each size of rings.

An example is proposed with a very simple network illustrated in figure E.19.



**Figure E.19** A very simple network.

This network is composed of 10 nodes, arbitrary of the same chemical species, and 7 bonds. Furthermore it is clear that in this network there are 1 ring with 3 nodes and 1 ring with 4 nodes.

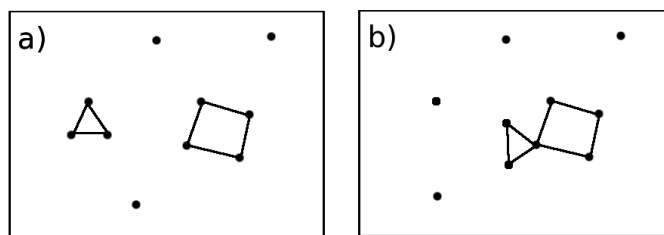
It is easy to calculate  $R_N$  and  $R_C$  for the network in figure E.19 ( $n$  = number of nodes):

$n$	$R_N(n)$	$n$	$R_C(n)$
3	1/10	3	3/10
4	1/10	4	4/10

**Table E.1**  $R_N$  and  $R_C$  for the network in figure E.19

In the literature the values of  $R_N$  and  $R_C$  are usually given separately [34, 35, 36, 37, 38]. Nevertheless these two properties are not sufficient in order to describe a network using rings. A simple example is proposed in figure E.20.

The two networks [Fig. E.20-a] and [Fig. E.20-b] do have very similar compositions with 10 nodes and 7 links but they are clearly different. Nevertheless the previous definitions of rings per cell and rings per node even taken together will lead to the same description for these two different networks [Tab. E.2].



**Figure E.20** Two simple networks having very close compositions: 10 nodes and 7 links.

$n$	$R_N(n)$	$n$	$R_C(n)$
3	1/10	3	3/10
4	1/10	4	4/10

**Table E.2**  $R_N$  and  $R_C$  calculated for the networks illustrated in figure E.20.

In both cases a) and b) there are 1 ring with 3 nodes and 1 ring with 4 nodes. It has to be noticed that these two rings have properties which correspond to each of the definitions introduced previously (King, Guttman, primitive and strong).

Thus none of these definitions is able to help to distinguish between these two networks. Therefore even though these simple networks are different, the previous definitions lead to the same description.

Thereby it is justified to wonder about the interpretation of the data presented in the literature for amorphous systems with a much higher complexity.

#### E.5.4 Rings and connectivity: the R.I.N.G.S. method implemented in **Atomes**

In the **Atomes** program the results of the ring statistics analysis are outputted following the new R.I.N.G.S. method [25, 39], this method is presented in the next pages.

The first goal of ring statistics is to give a faithful description of the connectivity of a network and to allow to compare this information with others obtained for already existing structures. It is therefore important to find a guideline which allows to establish a distinction and then a comparison between networks studied using ring statistics. We propose thereafter a new method to achieve this goal. First of all we noticed fundamental points that must be considered to get a reliable and transferable method:

1. The results must be reduced to the **total** number of nodes in the network.

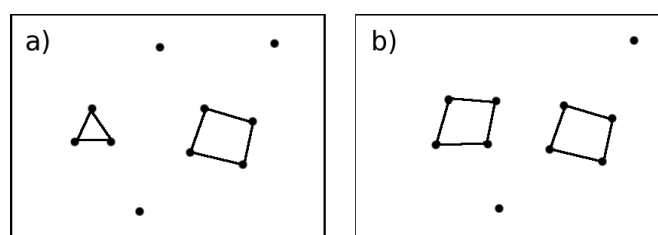
The nature of the nodes used to initiate the analysis when looking for rings will have a significant influence, therefore it is essential to reduce the results to a value for one node. Otherwise it would be impossible to compare the results to the ones obtained for systems made of nodes (particles) of different number and/or

nature.

2. *Different networks must be distinguishable whatever the method used to define a ring.*  
Indeed it is essential for the result of the analysis to be trustworthy independently of the method used to define a ring (King, Guttman, primitives, strong). Furthermore this will allow to compare the results of these different ring statistics.

#### E.5.4.1 Number of rings per cell ' $R_C$ '

We have already introduced this value, which is the first and the easiest way to compare networks using ring statistics.



**Figure E.21** *The first comparison element: the total number of rings in the network.*

$n$	$R_C(n)$	$n$	$R_C(n)$
3	1/10	3	0/10
4	1/10	4	2/10

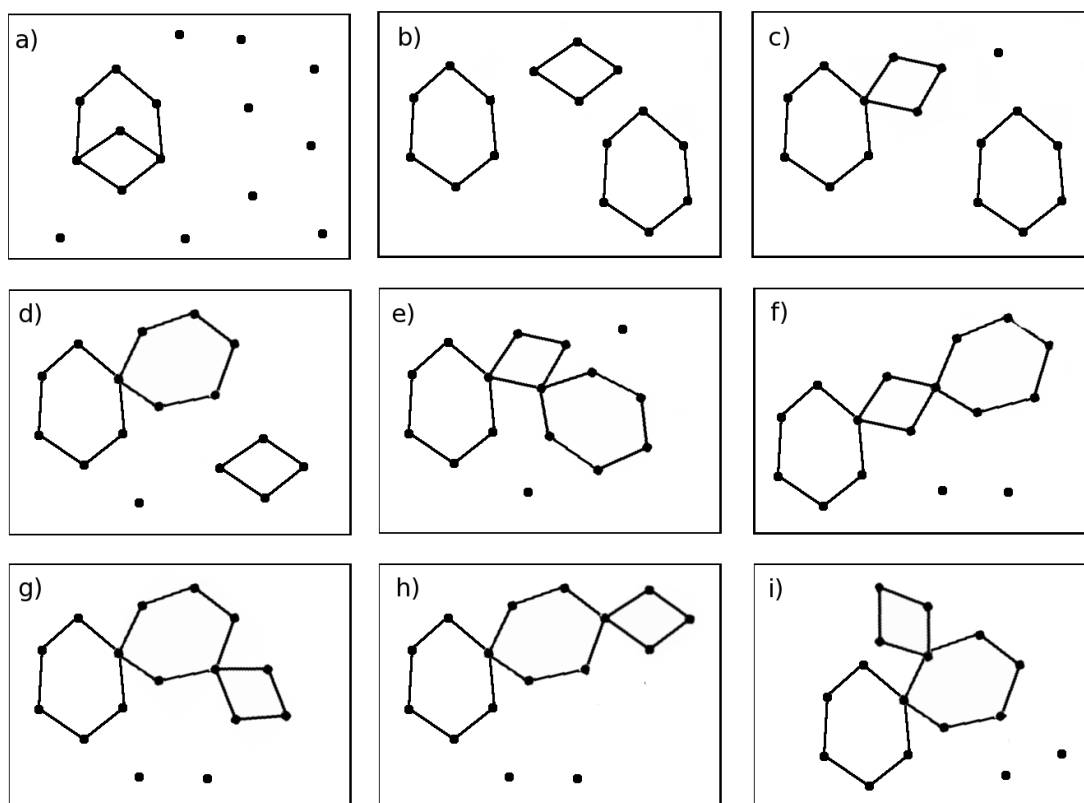
**Table E.3** *Number of rings in the simple networks represented in figure E.21.*

In the most simple cases, such as the one represented in figure E.21, the networks can be distinguished using only the number of rings [Tab. E.3]. Nevertheless in most of the cases other information are needed to describe accurately the connectivity of the networks.

#### E.5.4.2 Description of the connectivity: difference between rings and nodes

The second information needed to investigate the properties of a network using rings is the evaluation of the connectivity between rings. Indeed the distribution of the ring sizes gives a first information on the connectivity, nevertheless it can not be exactly evaluated unless one studies how the rings are connected. The impact of the relations between rings, already presented in figure E.20, has been illustrated in detail in figure E.22. Figure E.22 represents the different possibilities to combine 2 rings with 6 nodes and 1 ring with 4 nodes in a network composed of 16 nodes.

Among the 9 networks presented in figure E.22 none can be distinguished using the  $R_C$  value [Tab. E.4].



**Figure E.22** The 9 different networks with 16 nodes, composed of 2 rings with 6 nodes and 1 ring with 4 nodes.

$n$	$R_C(n)$
4	1/16
6	2/16

**Table E.4** Number of rings for the different networks presented in figure E.22.

Furthermore it is not possible to distinguish these networks using the  $R_N$  value. It seems possible to isolate the case a) [Tab. E.5] from the other cases b)  $\rightarrow$  i) [Tab. E.5]. Nevertheless the results obtained using the primitive rings criterion are similar for all cases a)  $\rightarrow$  i) [Tab. E.5], this is in contradiction with the second statement [2] proposed in our method.

Before introducing parameters able to distinguish the configurations presented in figure E.22 it is important to wonder about the number of cases to distinguish. From the point of view of the connectivity of the rings, configurations a), b), c) and d) are clearly different. Nevertheless following the same approach configurations e) and f) on the one hand and configurations g), h) and i) on the other hand are identical. A schematic representation [Fig. E.23] is sufficient to illustrate the similarity of the relations between these networks. The difference between each of these networks does

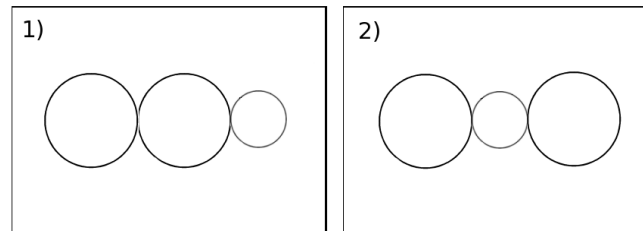
Case a)		$R_N(n)$	
$n$		King / Guttman.	Primitive / Strong.
4		4/16	<b>4/16</b>
6		10/16	<b>12/16</b>

Cases b) $\rightarrow$ i)		$R_N(n)$
$n$		All criteria.
4		<b>4/16</b>
6		<b>12/16</b>

**Table E.5** Number of rings per node for the networks presented in figure E.22.

not appear in the connectivity of the rings but in the connectivity of the particles.



**Figure E.23** Schematic representation of cases  $g) \rightarrow i)$  (1) and  $e) \rightarrow f)$  (2) illustrated in figure E.22.

Thus among the networks illustrated in figure E.22 six dispositions of the rings have to be distinguished (a, b, c, d, e, g). The proportions of particles involved, or not involved, in the construction of rings will become an important question.

The new tool defined in our method is able to describe accurately the information still missing on the connectivity. It is a square symmetric matrix of size  $(R - r + 1) \times (R - r + 1)$ , where  $R$  and  $r$  represent respectively the bigger and the smaller size of a ring found when analyzing the network: we have called this matrix the connectivity matrix [Tab. E.6].

$$C_{mat} = \begin{bmatrix} P_N(r) & P_N(r+1, r) & \cdots & P_N(R, r) \\ P_N(r, r+1) & \ddots & & P_N(R, r+1) \\ \vdots & & \ddots & \vdots \\ P_N(r, R) & \cdots & \cdots & P_N(R) \end{bmatrix}$$

**Table E.6** General connectivity matrix.

The diagonal elements  $P_N(i)$  of this matrix represent the **Proportion of Nodes** at the

origin of at least one ring of size  $i$ . And the non-diagonal elements  $P_N(i, j)$  represent the **Proportion of Nodes** at the origin of ring(s) of size  $i$  and  $j$ .

The matrix elements have a value ranging between 0 and 1. The lowest and non equal to 0 is of the form  $\frac{1}{Nn}$ , the highest and non equal to 1 is of the form  $\frac{Nn-1}{Nn}$ , where  $Nn$  represents the number of nodes in the network.

	King / Guttman.	Primitive / Strong.
Cas a)	$\begin{bmatrix} 4/16 & 2/16 \\ 2/16 & 5/16 \end{bmatrix}$	$\begin{bmatrix} 4/16 & 4/16 \\ 4/16 & 7/16 \end{bmatrix}$
	All criteria.	
Case b)	$\begin{bmatrix} 4/16 & 0/16 \\ 0/16 & 12/16 \end{bmatrix}$	
Case c)	$\begin{bmatrix} 4/16 & 1/16 \\ 1/16 & 12/16 \end{bmatrix}$	
Case d)	$\begin{bmatrix} 4/16 & 0/16 \\ 0/16 & 11/16 \end{bmatrix}$	
Case e) → f)	$\begin{bmatrix} 4/16 & 2/16 \\ 2/16 & 12/16 \end{bmatrix}$	
Case g) → i)	$\begin{bmatrix} 4/16 & 1/16 \\ 1/16 & 11/16 \end{bmatrix}$	
$n = \text{ring with } n \text{ nodes}$	$\begin{bmatrix} n4 & n6/n4 \\ n4/n6 & n6 \end{bmatrix}$	

**Table E.7** General connectivity matrix for the networks represented in figure E.22 and studied using the different definitions of rings.

The connectivity matrix of the configurations illustrated in figure E.22 are presented in table E.7. We see that this matrix allows to distinguish each network whatever the way used to define a ring is. This matrix remains simple for small systems (crystalline or amorphous) or when using a small maximum ring size for the analysis. Nevertheless its reading can be considerably altered when analysing amorphous systems with a high maximum ring size for the analysis.

To simplify the reading and the interpretation of the data contained in this matrix for more complex systems, we chose a similar approach to extract information on



the connectivity between the rings. As a first step we decided to evaluate only the diagonal elements  $P_N(n)$  of the general connectivity matrix. Indeed these values allow us to obtain a better view of the connectivity than the standard  $R_N$  value.

$P_N(n)$	
$n$	King / Guttman.    Primitive / Strong.
<b>Case a)</b>	
4	4/16

**Table E.8**  $P_N(n)$  - Proportion of nodes at the origin of at least one ring of size  $n$  for the networks in figure E.22.

It is clear [Tab. E.8] that using  $P_N(n)$  improves the separation between the networks illustrated in figure E.22. Nevertheless  $P_N(n)$  does not allow to distinguish each of them. We notice that the distinction between networks is improved [Tab. E.8] in particular when compared to the one obtain with  $R_N(n)$  [Tab. E.5].

Therefore in a second step we chose to calculate two properties whose definitions are very similar to the one of  $P_N(n)$ . The first, named  $P_{N_{\max}}(n)$ , represents the proportion of nodes for which the rings with  $n$  nodes are the longest closed paths found using these nodes to initiate the search. The second named,  $P_{N_{\min}}(n)$ , represents the proportion of nodes for which the rings with  $n$  nodes are the shortest closed paths found using these nodes to initiate the search.

The terms *longest* and *shortest path* must be considered carefully to avoid any confusion with the terms used in section E.5.2 to define the rings. For one node it is possible to find several rings whose properties correspond to the definitions proposed previously (King's, Guttman's, primitive or strong ring criterion). These rings are solutions found when looking for rings using this particular node to initiate the analysis. In order to calculate  $P_{N_{\max}}(n)$  and  $P_{N_{\min}}(n)$  the longest and the shortest path have to be determined among these different solutions.

$P_{N_{\max}}(n)$  and  $P_{N_{\min}}(n)$  have values ranging between 0 and  $P_N(n)$ . The lowest and non equal to 0 is of the form  $\frac{1}{Nn}$ , the highest and non equal to 1 is of the form  $\frac{Nn-1}{Nn}$ , where  $Nn$  represents the total number of nodes in the network. For the minimum ring size,  $s_{\min}$ , existing in the network or found during the search,  $P_{N_{\min}}(s_{\min}) = P_N(s_{\min})$ . In the same way for the maximum ring size,  $s_{\max}$ , existing in the network or found during the search,  $P_{N_{\max}}(s_{\max}) = P_N(s_{\max})$ .

To clarify these information it is possible to normalize  $P_{N_{\max}}(n)$  and  $P_{N_{\min}}(n)$  by  $P_N(n)$ . By reducing these values we obtain, for each size of rings, values independent of the total number of nodes  $Nn$  of the system. Then for a considered ring size the values only refer to the number of nodes where the search returns rings of this size:

$$P_{\max}(n) = \frac{P_{N_{\max}}(n)}{P_N(n)} \quad \text{and} \quad P_{\min}(n) = \frac{P_{N_{\min}}(n)}{P_N(n)}$$

The normalized terms  $P_{\max}(n)$  and  $P_{\min}(n)$  have values ranging between 0 and 1. The lowest and non equal to 0 is of the form  $\frac{1}{Nn}$ , the highest and non equal to 1 is of the form  $\frac{Nn-1}{Nn}$ . For the minimum ring size,  $s_{\min}$ , existing in the network or found during the search,  $P_{\min}(s_{\min}) = 1$ . In the same way for the maximum ring size,  $s_{\max}$ , existing in the network or found during the search,  $P_{\max}(s_{\max}) = 1$ .

$P_{\max}(n)$  and  $P_{\min}(n)$  give complementary information to the ones obtained with  $R_C(n)$  and  $P_N(n)$  in order to distinguish and compare networks using ring statistics. We can illustrate this result by presenting the complete information obtained with this method [Tab. E.9] for the networks represented in figure E.22.

$P_{\max}(n)$  and  $P_{\min}(n)$  give information about the connectivity of the rings with each other as a function of their size. If a ring of size  $n$  is found using a particular node to initiate the search,  $P_{\max}(n)$  gives the probability that this ring is the longest ring which can be found using this node to initiate the search. At the opposite,  $P_{\min}(n)$  gives the probability that this ring is the shortest ring which can be found using this node to initiate the search.

Thereafter we will use the terms 'connectivity profile' to designate the results of a ring statistics analysis. This profile is related to the definition of rings used in the search and is made of the 4 values defined in our method:  $R_C(n)$ ,  $P_N(n)$ ,  $P_{\max}(n)$  and  $P_{\min}(n)$ .

The **Atomes** program provides access to the connectivity profile of the system under

King / Guttman.				
$n$	$R_C(n)$	$P_N(n)$	$P_{\max}(n)$	$P_{\min}(n)$
<b>Case a)</b>				
4	1/16	4/16	0.5	1.0
6	2/16	5/16	1.0	0.6
Primitive / Strong.				
$n$	$R_C(n)$	$P_N(n)$	$P_{\max}(n)$	$P_{\min}(n)$
<b>Case a)</b>				
4	1/16	4/16	0.5	1.0
6	2/16	7/16	1.0	3/7
All criteria.				
$n$	$R_C(n)$	$P_N(n)$	$P_{\max}(n)$	$P_{\min}(n)$
<b>Case b)</b>				
4	1/16	4/16	1.0	1.0
6	2/16	12/16	1.0	1.0
<b>Case c)</b>				
4	1/16	4/16	0.75	1.0
6	2/16	12/16	1.0	11/12
<b>Case d)</b>				
4	1/16	4/16	1.0	1.0
6	2/16	11/16	1.0	1.0
<b>Case e) <math>\rightarrow</math> f)</b>				
4	1/16	4/16	0.5	1.0
6	2/16	12/16	1.0	10/12
<b>Case g) <math>\rightarrow</math> i)</b>				
4	1/16	4/16	0.75	1.0
6	2/16	11/16	1.0	10/11

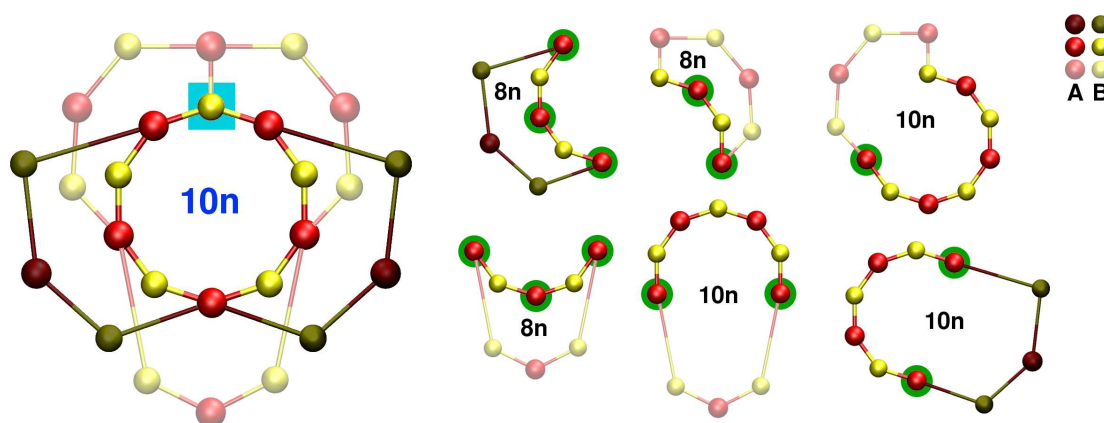
**Table E.9** Connectivity profiles results of the ring statistics for the networks presented in figure E.22.

study and allows to choose the study the connectivity using all the different methods used to define a ring. Thus King's rings, Guttman's rings, Primitive rings as well as Strong rings analysis are available.

## E.5.5 Bond defects in ring statistics

### E.5.5.1 ABAB and BABA rings

The ring statistics of amorphous networks are often focused on finding rings made of a succession of atoms with an alternation of chemical species, called ABAB rings. The most common examples come from the alternation of Si and O atoms (in silica polymorphs) or Ge and S (in GeS<sub>2</sub> polymorphs). These solids are usually built with tetrahedra (SiO<sub>4</sub> or GeS<sub>4</sub>) therefore we study the network distribution of tetrahedra. The ideal technique to setup the analysis of such systems is to choose the atoms of highest coordination to initiate the search, respectively Si in SiO<sub>2</sub> and Ge in GeS<sub>2</sub>. In most cases all rings can be found using this method. Nevertheless we can demonstrate that some solutions, so some rings, can be ignored by this analysis. This is highlighted in figure E.24 which represents a cluster of atoms isolated from an AB<sub>2</sub> amorphous network.



**Figure E.24** Cluster of atoms isolated from an AB<sub>2</sub> amorphous network. A bond defect is located on an atom of the chemical species B (blue square). When looking for King's shortest paths [S. E.5.2.1] using the chemical species A to initiate the search the central ring with 10 nodes is ignored. However among the solutions of the analysis (with the initial nodes circled in green) other rings with 10 nodes are found in the network.

We can see that this piece of network is characterized by a bond defect. An atom of the B species appears to be over-coordinated by three atoms of the A species. When looking for rings, using the King's criterion [S. E.5.2.1] and initiating the search using the A atoms, the central ring with 10 nodes is ignored. Nevertheless other rings with 10 nodes are found and stored as solutions of the analysis. In order to find the central ring the search has to be initiated from the overcoordinated B atom.

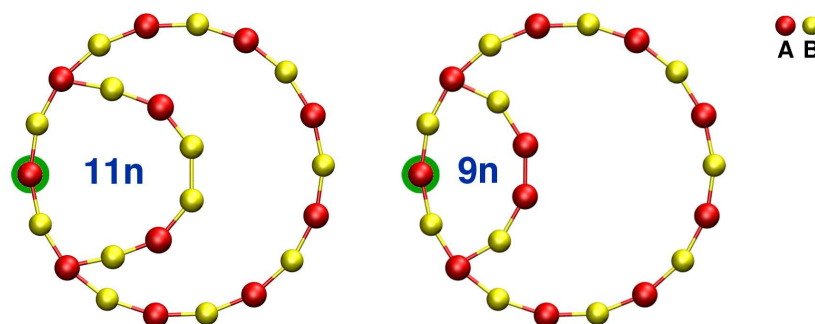
By analogy with the terminology ABAB this ring can be called a BABA ring. Indeed the alternation of chemical species is well respected. Therefore it is legitimate to question the relevance of the analysis without this result. In other words we have to check out if this BABA ring is, or not, an ABAB ring.

The properties of this ring meet the definition and can therefore improve the description

of the connectivity of the network. This kind of coordination defect [Fig. E.24] is uncommon in vitreous silica [34, 40], nevertheless it is frequent in chalcogenide glasses [41, 42].

### E.5.5.2 Homopolar bonds

In amorphous materials the homopolar bond defects can have a significant influence on the ring statistics. This is true in particular for  $AB_2$  chalcogenide glasses. Figure E.25 illustrates standard cases that may be encountered when looking for rings in an  $AB_2$  system which contains homopolar bonds.



**Figure E.25** *Illustration of the influence of homopolar bonds in ABAB rings: in both examples the smallest rings found when initiating the search using the circled nodes (green color) contain an homopolar bond A-A or B-B.*

The smallest rings found when initiating the search using the circled nodes (green color) are not ABAB rings. Therefore their size must be given using the total number of nodes. In figure E.25 the smallest rings are a ring with 9 nodes and a ring with 11 nodes containing respectively an A-A and a B-B homopolar bond. These rings are significantly smaller than the shortest ABAB ring with 18 nodes that may be found when looking for rings using the same green-circled nodes to initiate the analysis [Fig. E.25].

The **Atomes** program provides options to take into account or avoid A-B-A-B rings as well as homopolar bonds.

### E.5.6 Number of rings not found and that "potentially exist"

One of the first information it is possible to extract from ring statistics, except the number of rings, is the number of rings not found by the analysis. Indeed calculation times do strongly depend on the maximum search depth, ie. the maximum size of a ring. To carry out the analysis this value has to be chosen to get the best possible compromise between CPU time and quality of the description.

Nevertheless whatever this limiting value is, some rings of a size bigger than the maximum search depth may not be found by the analysis. In the King E.5.2.1 and the Guttman's criteria E.5.2.2 it is possible to evaluate the number of "potentially not

found" rings or rings that "potentially exist".

Thus for a given atom **At** we can consider that a closed path exists and is not found:

1. If the atom **At** has at least 2 nearest neighbors
2. If no closed path is found:
  - a- Starting from one neighbor to go back on the considered atom **At** (Guttman's criterion)
  - b- Between one couple of neighbors of the atom **At** (King's criterion)
3. If the 2 nearest neighbors of the atom **At** have at least 2 nearest neighbors (to avoid non bridging atoms)

Thus if during the analysis these 3 conditions are full filled (1, 2-a, 3 for the Guttman's criterion, and 1, 2-b, 3 for the King's criterion) then we can say that this analysis has potentially missed a ring between the neighbors of atom **At**. The smaller this number of "potentially" missed rings will be the better this analysis will be and the better the description of the connectivity of the material studied will be. The term "potentially" has been chosen because the method only allows to avoid first neighbor non bridging atoms.

Following this method **Atomes** gives access to the number of rings that "potentially exist" and not found during the analysis.

## E.6 Chain statistics

To get information on the connectivity of a material one can also rely on chain statistics. The idea of this calculation is to look for path between 2 atoms A and B, respecting the following rules:

- Total coordination for A ( $\alpha$ ) must be  $\neq 2$
- Total coordination for B ( $\beta$ ) must be  $\neq 2$
- Total coordination for all atom(s) between A and B must be equal to 2.

Chains are then litteraly succession of atoms isolated from the rest of the material.

**Atomes** offers several options to enforce specific definition of a chain for the search:

- Total coordination for A and B can be restricted to 1: searching for chains would mean searching for isolated 1 dimensional (on a coordination point of view) structures in the material.
- The chemistry of the atoms in the chain(s) can be considered:
  - Only searching for AAAA ( $\alpha\alpha\alpha\alpha$ ) chains (homopolar bonds exclusively).
  - Excluding homopolar bonds from the search (heteropolar bonds exclusively).
  - Only searching for ABAB ( $\alpha\beta\alpha\beta$ ) chains (perfect alternate of heteropolar bonds).

## E.7 Invariants of spherical harmonics as atomic order parameters

Invariants formed from bond spherical harmonics allow to obtain quantitative information on the local atomic symmetries in materials. The analysis starts by associating a set of spherical harmonics with every bond linking an atom to its nearest neighbors. For a given bond defined by a vector  $\vec{r}$  a spherical harmonic may be defined as:

$$Q_{lm}(\vec{r}) = Y_{lm}(\theta(\vec{r}), \psi(\vec{r})) \quad (\text{E.28})$$

where  $Y_{lm}(\theta, \psi)$  is the spherical harmonic associated to the bond,  $\theta$  and  $\psi$  are the angular components of the spherical coordinates of the bond which Cartesian coordinates are defined by  $\vec{r}$ .

Because the  $Q_{lm}$  for a given  $l$  can be scrambled by changing to a rotated coordinate system, it is important to consider rotational invariant combinations, such as [26, 43]:

$$Q_l = \left[ \frac{4\pi}{2l+1} \sum_{m=-l}^l |\bar{Q}_{lm}|^2 \right]^{1/2} \quad (\text{E.29})$$

where  $\bar{Q}_{lm}$  is defined by:

$$\bar{Q}_{lm} = \langle Q_{lm}(\vec{r}) \rangle \quad (\text{E.30})$$

and represents an average of the  $Y_{lm}(\theta, \psi)$  over all  $\vec{r}$  vectors in the system whether these vectors belong to the same atomic configuration or not. Just as the angular momentum quantum number,  $l$ , is a characteristic quantity of the 'shape' of an atomic orbital, the quantity  $Q_l$  is a rotationally invariant characteristic value of the shape/symmetry of a given local atomic configuration (if the average is not taken on all bonds of the system but only within a given configuration) or an average of such values for a set of configurations. Thus it is possible to compare  $Q_l$ 's computed for well known crystal structures (e.g. FCC, HFC ...) and some local atomic configurations in a material's model. The results of the comparison gives information for the presence/absence of a particular local atomic symmetry.

**Atomes** allows to compute the average  $Q_l$ 's for each chemical species as well as the average  $Q_l$ 's for a user specified local atomic coordination.

## E.8 Mean square displacement of atoms

Atoms in solids, liquids and gases move constantly at any given temperature, i.e. they are subject to a "thermal" displacement from their average positions. This displacement is particularly important in the case of a liquids. Atomic displacement does not follow a simple trajectory: "collisions" with other atoms render atomic trajectories quite complex shaped in space.

The trajectory followed by an atom in a liquid resembles that of a pedestrian random walk. Mathematically this represents a sequence of steps done one after another where each step follows a random direction which does not depend on the one of the previous step (Markov's chain of events).

In the case of a one-dimensional system (straight line) the displacement of the atom will therefore be either a forward step (+) or a backward step (-). Furthermore it will be impossible to predict one or the other direction (forward or backward) since they have an equal probability to occur.

One can conclude that the distance an atom may travel is close to zero. Nevertheless if we choose not to sum the displacements themselves (+/-) but the square of these displacements then we will end up with a non-zero, positive quantity of the total squared distance traveled. Consequently this allows to obtain a better evaluation of the real (square) distance traveled by an atom.

The **Mean Square Displacement** MSD is defined by the relation:

$$MSD(t) = \langle \mathbf{r}^2(t) \rangle = \langle |\mathbf{r}_i(t) - \mathbf{r}_i(0)|^2 \rangle \quad (\text{E.31})$$

where  $\mathbf{r}_i(t)$  is the position of the atom  $i$  at the time  $t$ , and the  $\langle \rangle$  represent an average on the time steps and/or the particles.

However, during the analysis of the results of molecular dynamics simulations it is important to subtract the drift of the center of mass of the simulation box:

$$MSD(t) = \langle |\mathbf{r}_i(t) - \mathbf{r}_i(0) - [\mathbf{r}_{cm}(t) - \mathbf{r}_{cm}(0)]|^2 \rangle \quad (\text{E.32})$$



where  $\mathbf{r}_{cm}(t)$  represents the position of the center of mass of the system at the time  $t$ . The MSD also contains information on the diffusion of atoms. If the system is solid (frozen) then MSD "saturate", and the kinetic energy is not sufficient enough to reach a diffusive behavior. Nevertheless if the system is not frozen (e.g. liquid) then the MSD will grow linearly in time. In such a case it is possible to investigate the behavior of the system looking at the slope of the MSD. The slope of the MSD or the so called diffusion constant  $D$  is defined by:

$$D = \lim_{t \rightarrow \infty} \frac{1}{6t} \langle \mathbf{r}^2(t) \rangle \quad (\text{E.33})$$

**Atomes** provides access to the several MSD related functions:

- MSD for each chemical species with autocorrelation on all the dynamics
- MSD for each chemical species without autocorrelation on all the dynamics (step by step)
- Directional MSD (x, y, z, xy, xz, yz) for each chemical species with autocorrelation on all the dynamics
- Directional MSD (x, y, z, xy, xz, yz) for each chemical species without autocorrelation on all the dynamics (step by step)
- Drift of the center of mass (x, y, z)
- Correction applied to correct the drift of the center of mass in equation [Eq. [E.32](#)] (x, y, z)



---

# The chemical properties database in **Atomes**

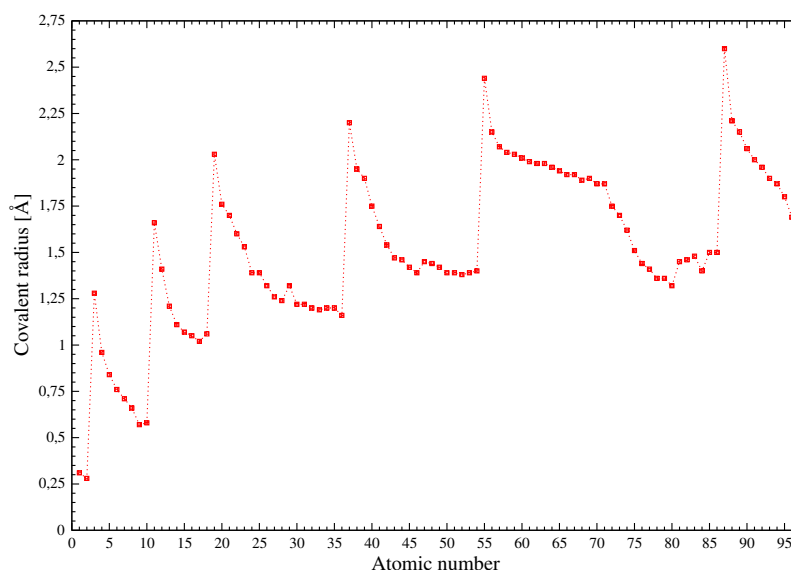
---

A database of chemical/physical properties is included in the **Atomes** program, this appendix presents these properties (atomic radii, x-ray and neutron scattering lengths) as well as the references of the articles from which this information was obtained. Note that the data presented in this appendix is available for download on the web site of the program:

<https://atomes.ipcms.fr>

## F.1 Atomic radii

### F.1.1 Covalent radii



**Figure F.1** Covalent radii used in the **Atomes** program.

Figure F.1 presents the covalent radii used in **Atomes** see [44] for details.

### F.1.2 Ionic radii

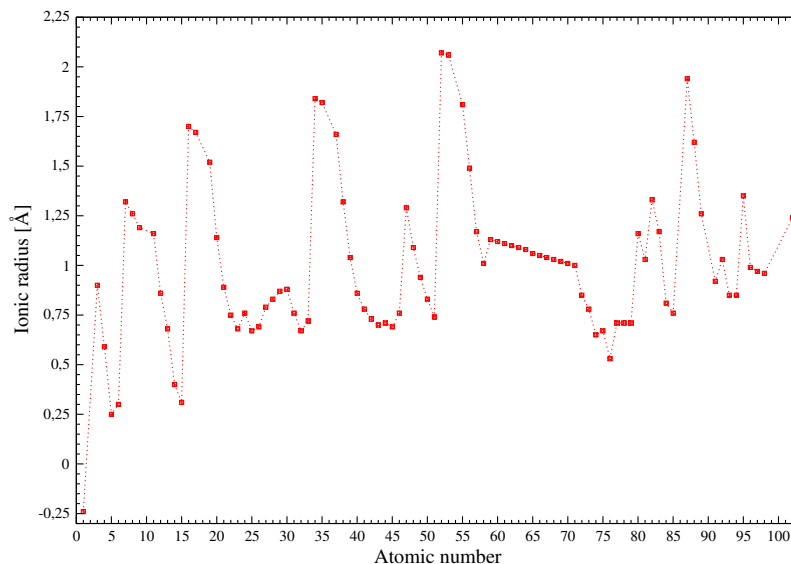


Figure F.2 Ionic radii used in the **Atomes** program.

Figure F.2 presents the ionic radii used in **Atomes** see [45] for details.

### F.1.3 VDW radii

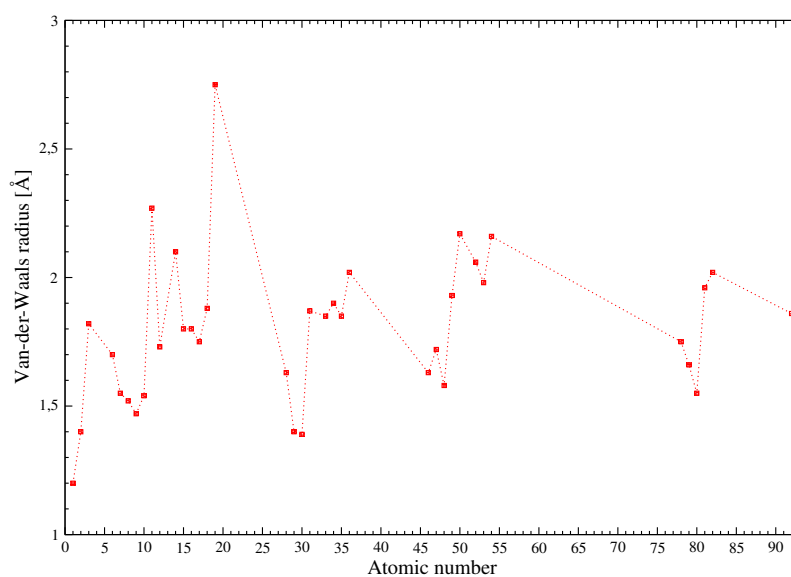
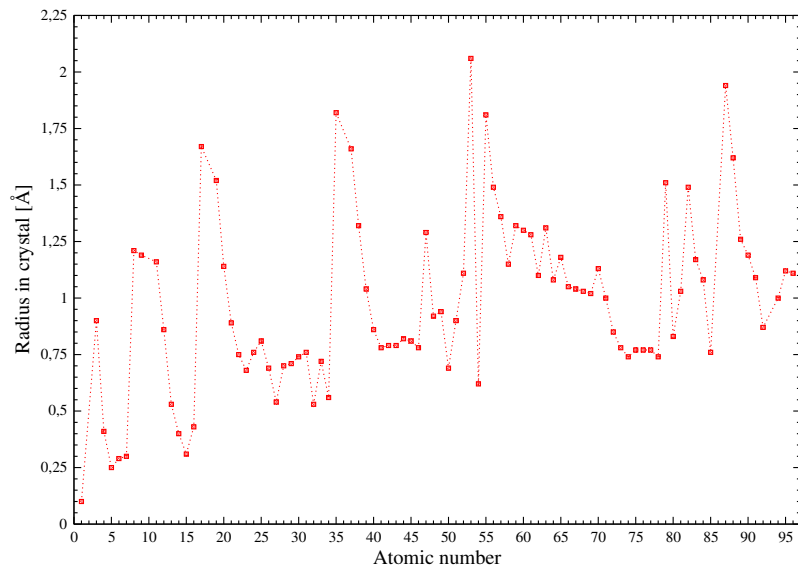


Figure F.3 Van Der Waals radii used in the **Atomes** program.

Figure F.3 presents the Van Der Waals radii used in **Atomes** see [46] for details.

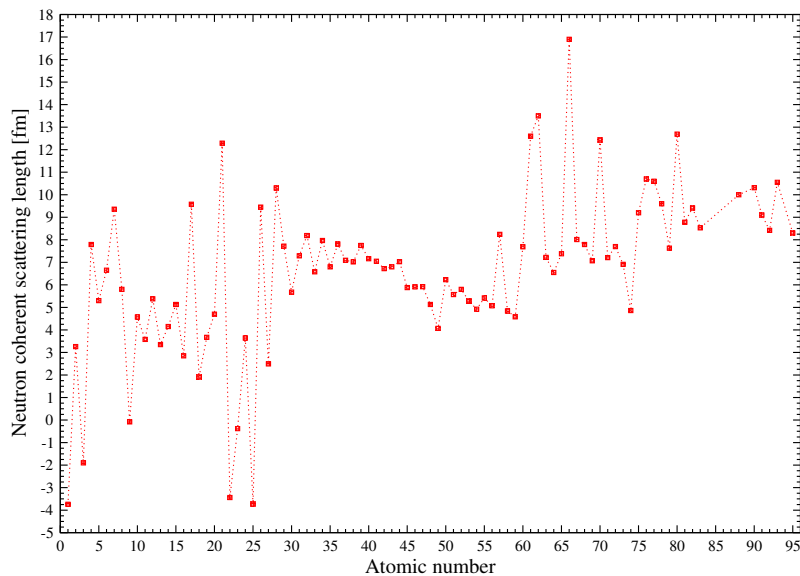
### F.1.4 Shannon radii in crystal



**Figure F.4** *Shannon radii in crystals used in the **Atomes** program.*

Figure F.4 presents the Shannon radii in crystals used in **Atomes** see [47, 48] for details.

## F.2 Neutron scattering lenghts



**Figure F.5** *Neutron scattering lengths used in the **Atomes** program.*

Figure F.5 presents the neutron scattering lengths used in **Atomes** see [49, 50] for details.



# Bibliography

- [1] [https://www.scd.stfc.ac.uk/pages/dl\\_poly.aspx](https://www.scd.stfc.ac.uk/pages/dl_poly.aspx). (Cité pages 1, 17 et 115.)
- [2] <https://lammmps.sandia.gov>. (Cité pages 1, 17 et 115.)
- [3] <http://www.cpmc.org>. (Cité pages 1, 17, 115 et III.)
- [4] <http://cp2k.berlios.de>. (Cité pages 1, 17 et 115.)
- [5] <http://www.gtk.org>. (Cité page 3.)
- [6] <https://www.openmp.org>. (Cité page 3.)
- [7] <https://www.opengl.org>. (Cité page 3.)
- [8] <https://ffmpeg.org>. (Cité page 3.)
- [9] <https://docs.gtk.org/pango>. (Cité page 9.)
- [10] M. P. Allen and D. J. Tildesley. *Computer simulation of liquids*. Oxford science publications (1987). (Cité page 14.)
- [11] P. S. Salmon. *J. Non-Cryst. Solids*, **353**:2959–2974 (2007). (Cité page 14.)
- [12] M. T. Dove, M. G. Tucker, and D. A. Keen. *Eur. Jour. Mat.*, **14**:331–348 (2002). (Cité pages 14 et XXVIII.)
- [13] B. Thijsse. *J. App. Cryst.*, **17**:61–76 (1984). (Cité pages 14 et XXVIII.)
- [14] T. E. Faber and Ziman J. M. *Phil. Mag.*, **11**(109):153–173 (1965). (Cité pages 14, XXVIII et XXX.)
- [15] N. W. Ashcroft and D. C. Langreth. *Phys. Rev.*, **156**(3):685–692 (1967). (Cité pages 14, XXVIII et XXX.)
- [16] N. W. Ashcroft and D. C. Langreth. *Phys. Rev.*, **159**(3):500–510 (1967). (Cité pages 14, XXVIII et XXX.)
- [17] N. W. Ashcroft and D. C. Langreth. *Phys. Rev.*, **166**(3):934 (1968). (Cité pages 14, XXVIII et XXX.)
- [18] A. B. Bhatia and D. E. Thornton. *Phys. Rev. B.*, **2**(8):3004–3012 (1970). (Cité pages 14, XXIX et XXX.)
- [19] S. V. King. *Nat.*, **213**:1112 (1967). (Cité pages 14 et XXXIV.)

- [20] D. S. Franzblau. *Phys. Rev. B.*, **44**(10):4925–4930 (1991). (Cité pages 14 et XXXIV.)
- [21] L. Guttman. *J. Non-Cryst. Solids*, **116**:145–147 (1990). (Cité pages 14 et XXXV.)
- [22] K. Goetzke and H. J. Klein. *J. Non-Cryst. Solids*, **127**:215–220 (1991). (Cité pages 14, XXXVI et XXXVIII.)
- [23] X. Yuan and A. N. Cormack. *Comp. Mat. Sci.*, **24**:343–360 (2002). (Cité pages 14, XXXVI et XXXVIII.)
- [24] F. Wooten. *Act. Cryst. A*, **58**(4):346–351 (2002). (Cité pages 14 et XXXVI.)
- [25] S. Le Roux and P. Jund. *Comp. Mat. Sci.*, **49**:70–83 (2010). (Cité pages 15 et XL.)
- [26] P. Steinhardt, D. R. Nelson, and M. Ronchetti. *Phys. Rev. B.*, **28**(2):784–805 (1983). (Cité pages 15 et LI.)
- [27] S. Le Roux, A. Bouzid, M. Boero, and C. Massobrio. *J. Chem. Phys.*, **138**(17):174505 (2013). (Cité page 19.)
- [28] Grace Development Team. *GGraphing, Advanced Computation and Exploration of data*, pages <http://plasma-gate.weizmann.ac.il/Grace/> (1996-2009). (Cité page 29.)
- [29] Space-group symmetry. *International Tables for Crystallography*, **A** (2016). (Cité pages 69 et XV.)
- [30] Minnesota Supercomputer Center. *Xmol*, page <http://www.hpc.unm.edu/chem/xmol/xmol.html> (1993). (Cité page III.)
- [31] CambridgeSoft. *Chem3D*, pages Cambridge Scientific Computing, Cambridge Massachusetts, USA (2009). (Cité page III.)
- [32] <http://cms.mpi.univie.ac.at/vasp>. (Cité page III.)
- [33] S. Le Roux and V. Petkov. *J. Appl. Cryst.*, **43**:181–185 (2010). (Cité page III.)
- [34] J. P. Rino, I. Ebbsjö, R. K. Kalia, A. Nakano, and P. Vashishta. *Phys. Rev. B.*, **47**(6):3053–3062 (1993). (Cité pages XXXIX et XLIX.)
- [35] R. M. Van Ginhoven, H. Jónsson, and L. R. Corrales. *Phys. Rev. B.*, **71**(2):024208 (2005). (Cité page XXXIX.)
- [36] M. Cobb, D. A. Drabold, and R. L. Cappelletti. *Phys. Rev. B.*, **54**(17):12162–12171 (1996). (Cité page XXXIX.)
- [37] X. Zhang and D. A. Drabold. *Phys. Rev. B.*, **62**(23):15695–15701 (2000). (Cité page XXXIX.)
- [38] D. N. Tafen and D. A. Drabold. *Phys. Rev. B.*, **71**(5):054206 (2005). (Cité page XXXIX.)
- [39] S. Le Roux and P. Jund. *Comp. Mat. Sci.*, **50**:1217 (2011). (Cité page XL.)
- [40] W. Jin, P. Vashishta, R. K. Kalia, and J. P. Rino. *Phys. Rev. B.*, **48**(13):9359–9368 (1993). (Cité page XLIX.)



- [41] S. Blaineau and P. Jund. *Phys. Rev. B.*, **69**(6):064201 (2004). (Cité page XLIX.)
- [42] S. Le Roux and P. Jund. *J. Phys. Cond. Mat.*, **19**(19):169102 (2007). (Cité page XLIX.)
- [43] A. Baranyai, A. Geiger, P. R. Gartrell-Mills, K. Heinzinger, R. McGreevy, G. Pálinkás, and I. Ruff. *Chem. Soc. Faraday Trans. 2*, **83**(8):1335–1365 (1987). (Cité page LI.)
- [44] B. Cordero. *Dalton Trans.*, pages 2832–2838 (2008). (Cité page LVI.)
- [45] J. C. Slater. *J. Chem. Phys.*, **41**(10):3199 (1964). (Cité page LVI.)
- [46] A. Bondi. *J. Phys. Chem.*, **68**(3):441 (1964). (Cité page LVI.)
- [47] R. D. Shannon and C. T. Prewitt. *Act. Cryst.c*, **B25**:925 (1969). (Cité page LVII.)
- [48] R. D. Shannon. *Act. Cryst.c*, **A23**:751–76A (1976). (Cité page LVII.)
- [49] <http://www.ncnr.nist.gov/resources/n-lenghs>. (Cité page LVII.)
- [50] V.F. Sears. *Neutrons News*, **3**(3):26–37 (1992). (Cité page LVII.)

This document has been prepared using the Linux operating system and free softwares:

The text editor	" <a href="#">gVim</a> "
The GNU image manipulation program	" <a href="#">The Gimp</a> "
The WYSIWYG plotting tool	" <a href="#">Grace</a> "
And the document preparation system	" <a href="#">L<sup>A</sup>T<sub>E</sub>X 2<sub>ε</sub></a> ".