THE DL_POLY_4 USER MANUAL

I.T. Todorov & W. Smith

STFC Daresbury Laboratory
Daresbury, Warrington WA4 4AD
Cheshire, England, United Kingdom

Version 4.07 – January 2015

ABOUT DL_POLY_4

DL_POLY_4 is a general purpose parallel molecular dynamics simulation package developed at Daresbury Laboratory by W. Smith and I.T. Todorov. The DL_POLY project was developed under the auspices of the Engineering and Physical Sciences Research Council (EPSRC) for the EPSRC's Collaborative Computational Project for the Computer Simulation of Condensed Phases (CCP5), the Computational Chemistry and Advanced Research Computing Groups (CCG & ARCG) at Daresbury Laboratory and the Natural Environment Research Council (NERC) for the NERC's *e*Science project *Computational Chemistry in the Environment* (*e*Minerals), directed by M.T. Dove.

DL_POLY_4 is the property of Daresbury Laboratory and is issued free **under licence** to academic institutions pursuing scientific research of a non-commercial nature. Commercial organisations may be permitted a licence to use the package after negotiation with the owners. Daresbury Laboratory is the sole centre for distribution of the package. Under no account is it to be redistributed to third parties without consent of the owners.

The purpose of the DL_POLY_4 package is to provide software for academic research that is inexpensive, accessible and free of commercial considerations. Users have direct access to source code for modification and inspection. In the spirit of the enterprise, contributions in the form of working code are welcome, provided the code is compatible with DL_POLY_4 in regard to its interfaces and programming style and it is adequately documented.

DISCLAIMER

ACKNOWLEDGEMENTS

DL_POLY_4 was developed at Daresbury Laboratory (DL - http://www.dl.ac.uk/), the Science and Technology Facilities Council (STFC - http://www.stfc.ac.uk/), UK, with support from the Engineering and Physical Sciences Research Council (EPSRC - http://www.epsrc.ac.uk/) and the Natural Environment Research Council (NERC - http://www.nerc.ac.uk/).

Advice, assistance and encouragement in the development of DL_POLY_4 has been given by many people. We gratefully acknowledge the following:

T.R. Forester, I.J. Bush, M. Leslie, M.F. Guest, R.J. Allan, D. Tildesley, M. Pinches, D. Rapaport, the UK's "Materials Chemistry Consortium" under C.R.A. Catlow and the *e*Minerals project under M.T. Dove.

`This document is produced with LaTeX & hdvipdfm`

Manual Notation

In the DL_POLY manuals specific fonts are used to convey specific meanings:

1. *directories* - indicates UNIX file directories
2. ROUTINES - indicates subroutines, functions and programs
3. *macros* - indicates a macro (file of UNIX commands)
4. **directive** - indicates directives or keywords
5. `variables` - indicates named variables and parameters
6. FILE - indicates filenames.

Contents

List of Tables

List of Figures

Chapter 0

Quick Word / INSTALL & RUN

For the experienced and quick minded this is a very brief resume of how to INSTALL & RUN DL_POLY_4 (which is no excuse for skipping the Introduction, Chapter 1!). For the rest of us it sketches out how to start running DL_POLY_4 jobs and where one should look to obtain more detailed information if need be.

If you have followed the procedure for obtaining and downloading the DL_POLY_4 package (see Obtaining the Source Code, Section 7), have successfully unpacked it and are ready to compile the source code, then jump to the Makefiles section of the DL_POLY_4 README Appendix E or alternatively view the README.TXT in *source*) and follow the instructions within.

If you have compiled successfully then a freshly date-stamped file, named DLPOLY.Z should appear in the listing of the *execute* subdirectory (the '`ls -haltr`' command issued on a Linux/Unix-like shell within *execute* will place the executable in the last row of the list). If **unsuccessful** then you should read the Compiling and Running DL_POLY_4 Section 5.2.1.

To run the code you first need to place the necessary input files within *execute*. TEST cases containing suitable input files, as well as examples of output files, can be obtained at the DL_POLY_4 FTP site, ftp://ftp.dl.ac.uk/ccp5/DL_POLY/. Examine the contents of *data*/`README.txt` and *bench*/`README.txt` for more information. To run the serial version you simply issue the command `DLPOLY.Z` (or `DLPOLY.Z.cu`) within the *execute* subdirectory. If you have compiled a parallel version and are running it on a parallel machine, then naturally you will need to familiarise yourself with the local procedures of how to run jobs on that machine. In general though, running a parallel job will usually require that you issue a necessary run command (e.g. `mpirun -n 8 DLPOLY.Z`) or submit a job script from within *execute*.

If you need to know more then search the manual and examine sections of interests. You may also wish to visit DL_POLY project web-page http://www.ccp5.ac.uk/DL_POLY/ and examine the useful links within (FAQ, User Forum, etc.).

If you are looking to gain more in depth experience, then regular training workshops are available. To find about upcoming workshops, subscribe to our mail list by following instructions in Section 1.7.

If you need one-to-one training, wish to collaborate scientifically and/or would like to become a contributor/developer then get in touch with me, Dr. I.T. Todorov, by emailing to ilian.todorov@stfc.ac.uk.

Best of luck!

Chapter 1

Introduction

Scope of Chapter

This chapter describes the concept, design and directory structure of DL_POLY_4 and how to obtain a copy of the source code.

1.1 The DL_POLY Package

DL_POLY [1] is a package of subroutines, programs and data files, designed to facilitate molecular dynamics simulations of macromolecules, polymers, ionic systems and solutions on a distributed memory parallel computer. It is available in two forms: DL_POLY_Classic (written by Bill Smith & Tim Forester, http://www.ccp5.ac.uk/DL_POLY_CLASSIC/) and DL_POLY_4 (written by Ilian Todorov & Bill Smith) [2, 3]. Both versions were originally written on behalf of CCP5, the UK's Collaborative Computational Project on Molecular Simulation, which has been in existence since 1980 ([4], http://www.ccp5.ac.uk/DL_POLY/).

The two forms of DL_POLY differ primarily in their method of exploiting parallelism. DL_POLY_Classic uses a Replicated Data (RD) strategy [5, 6, 7, 8] which works well simulations of up to 30,000 atoms on up to 100 processors. DL_POLY_4 is based on the Domain Decomposition (DD) strategy [2, 3, 9, 10, 5, 6], and is best suited for large molecular simulations from 10^3 to 10^9 atoms on large processor counts. The two packages are reasonably compatible, so that it is possible to scale up from a DL_POLY_Classic to a DL_POLY_4 simulation with little effort. It should be apparent from these comments that DL_POLY_4 is not intended as a replacement for DL_POLY_Classic.

Users are reminded that we are interested in hearing what other features could be usefully incorporated. We obviously have ideas of our own and CCP5 strongly influences developments, but other input would be welcome nevertheless. We also request that our users respect the integrity of DL_POLY_4 source and not pass it on to third parties. We require that all users of the package register with us, not least because we need to keep everyone abreast of new developments and discovered bugs. We have developed various forms of licence, which we hope will ward off litigation (from both sides), without denying access to genuine scientific users.

Further information on the DL_POLY packages may be obtained from the DL_POLY project website - http://www.ccp5.ac.uk/DL_POLY/ .

1.2 Functionality

The following is a list of the features DL_POLY_4 supports.

1.2.1 Molecular Systems

DL_POLY_4 will simulate the following molecular species:

- Simple atomic systems and mixtures, e.g. Ne, Ar, Kr, etc.
- Simple unpolarisable point ions, e.g. NaCl, KCl, etc.
- Polarisable point ions and molecules, e.g. MgO, H_2O, etc.
- Simple rigid molecules e.g. CCl_4, SF_6, Benzene, etc.
- Rigid molecular ions with point charges e.g. KNO_3, $(NH_4)_2SO_4$, etc.
- Polymers with rigid bonds, e.g. C_nH_{2n+2}
- Polymers with flexible and rigid bonds and point charges, e.g. proteins, macromolecules etc.
- Silicate glasses and zeolites
- Simple metals and metal alloys, e.g. Al, Ni, Cu, Cu_3Au, etc.
- Covalent systems as hydro-carbons and transition elements, e.g. C, Si, Ge, SiC, SiGe, ets.

1.2.2 Force Field

The DL_POLY_4 force field includes the following features:

1. All common forms of non-bonded atom-atom (van der Waals) potentials
2. Atom-atom (and site-site) coulombic potentials
3. Metal-metal (local density dependent) potentials [11, 12, 13, 14, 15, 16]
4. Tersoff (local density dependent) potentials (for hydro-carbons) [17]
5. Three-body valence angle and hydrogen bond potentials
6. Four-body inversion potentials
7. Ion core-shell polarasation
8. Tether potentials
9. Chemical bond potentials
10. Valence angle potentials
11. Dihedral angle (and improper dihedral angle) potentials
12. Inversion angle potentials
13. External field potentials.

The parameters describing these potentials may be obtained, for example, from the GROMOS [18], Dreiding [19] or AMBER [20] forcefield, which share functional forms. It is relatively easy to adapt DL_POLY_4 to user specific force fields.

1.2.3 Boundary Conditions

DL_POLY_4 will accommodate the following boundary conditions:

1. None, e.g. isolated molecules *in vacuo*
2. Cubic periodic boundaries
3. Orthorhombic periodic boundaries
4. Parallelepiped periodic boundaries
5. Slab (x,y periodic, z non-periodic).

These are described in detail in Appendix B. Note that periodic boundary conditions (PBC) 1 and 5 above require careful consideration to enable efficient load balancing on a parallel computer.

1.2.4 Java Graphical User Interface

The DL_POLY_4 Graphical User Interface (GUI) is the same one that also comes with DL_POLY_Classic, which is written in the Java®programming language from Sun®Microsystems. A major advantage of this is the free availability of the Java programming environment from Sun®, and also its portability across platforms. The compiled GUI may be run without recompiling on any Java®supported machine. The GUI is an integral component of the DL_POLY suites and is available on the same terms (see the GUI manual [21]).

1.2.5 Algorithms

1.2.5.1 Parallel Algorithms

DL_POLY_4 exclusively employs the Domain Decomposition parallelisation strategy [9, 10, 5, 6] (see Section 7.1.1).

1.2.5.2 Molecular Dynamics Algorithms

DL_POLY_4 offers a selection of MD integration algorithms couched in both Velocity Verlet (VV) and Leapfrog Verlet (LFV) manner [22]. These generate NVE, NVE$_{kin}$, NVT, NPT and N$\underline{\underline{\sigma}}$T ensembles with a selection of thermostats and barostats. Parallel versions of the RATTLE [23] and SHAKE [8] algorithms are used for solving bond constraints in the VV and LFV cast integrations respectively. The rotational motion of rigid bodies (RBs) is handled with Fincham's implicit quaternion algorithm (FIQA) [24] under the LFV scheme or with the "NOSQUISH" algorithm of Miller *et al* [25] under the VV integration.

The following MD algorithms are available:

1. Constant E algorithm
2. Evans constant E$_{kin}$ algorithm [26]
3. Langevin constant T algorithm [27]
4. Andersen constant T algorithm [28]
5. Berendsen constant T algorithm [29]
6. Nosé-Hoover constant T algorithm [30]
7. Langevin constant T,P algorithm [31]
8. Berendsen constant T,P algorithm [29]
9. Nosé-Hoover constant T,P algorithm [30]
10. Martyna, Tuckerman and Klein (MTK) constant T,P algorithm [32]
11. Langevin constant T,$\underline{\underline{\sigma}}$ algorithm [31]
12. Berendsen constant T,$\underline{\underline{\sigma}}$ algorithm [29]
13. Nosé-Hoover constant T,$\underline{\underline{\sigma}}$ algorithm [30]
14. Martyna, Tuckerman and Klein (MTK) constant T,$\underline{\underline{\sigma}}$ algorithm [32].

1.2.6 DL_POLY_Classic features incompatible or unavalable in DL_POLY_4

- Force field
  - Rigid bodies connected with constraint links are **not available**
  - Shell models specification is **solely** determined by the presence of mass on the shells
  - Dihedral potentials with more than three *original* parameters (see OPLS) have two artificially added parameters, defining the 1-4 electrostatic and van der Waals scaling factors, which **must** be placed at fourth and fifth position respectively, extending the original parameter list split by them

- Boundary conditions
  - Truncated octahedral periodic boundaries (`imcon` = 4) are **not available**
  - Rhombic dodecahedral periodic boundaries (`imcon` = 5) are **not available**
  - Hexagonal prism periodic boundaries (`imcon` = 7) are **not available**
- Electrostatics
  - Standard Ewald Summation is **not available**, but is **substituted** by Smoothed Particle Mesh Ewald (SPME) summation
  - Hautman-Klein Ewald Summation for 3D non-periodic but 2D periodic systems is **not available**
- Non-standard functionality
  - Temperature Accelerated Dynamics
  - Hyperdynamics
  - Solvation Energies

1.3 Programming Style

The programming style of DL_POLY_4 is intended to be as uniform as possible. The following stylistic rules apply throughout. Potential contributors of code are requested to note the stylistic convention.

1.3.1 Programming Language

DL_POLY_4 is written in free format FORTRAN90. In DL_POLY_4 we have adopted the convention of *explicit type declaration* i.e. we have used

```
Implicit None
```

in all subroutines. Thus all variables must be given an explicit type: `Integer, Real( Kind = wp)`, etc.

1.3.2 Modularisation and Intent

DL_POLY_4 exploits the full potential of the modularisation concept in FORTRAN90. Variables having in common description of certain feature or method in DL_POLY_4 are grouped in modules. This simplifies subroutines' calling sequences and decreases error-proneness in programming as subroutines must define what they use and from which module. To decrease error-proneness further, arguments that are passed in calling sequences of functions or subroutines have defined intent, i.e. whether they are to be:

- passed in only (`Intent (In)`) - the argument is not allowed to be changed by the routine
- passed out only (`Intent (Out)`) - the "coming in" value of the argument is unimportant
- passed in both directions in and out (`Intent (InOut)`) - the "coming in" value of the argument is important and the argument is allowed to be changed.

1.3.3 Memory Management

DL_POLY_4 exploits the dynamic array allocation features of FORTRAN90 to assign the necessary array dimensions.

1.3.4 Target Platforms

DL_POLY_4 is intended for distributed memory parallel computers.

Compilation of DL_POLY_4 in parallel mode requires **only** a FORTRAN90 compiler and Message Passing Interface (MPI) to handle communications. Compilation of DL_POLY_4 in serial mode is also possible and requires **only** a FORTRAN90 compiler.

1.3.5 Internal Documentation

All subroutines are supplied with a header block of FORTRAN90 comment (!) records giving:

1. The name of the author and/or modifying author
2. The version number or date of production
3. A brief description of the function of the subroutine
4. A copyright statement.

Elsewhere FORTRAN90 comment cards (!) are used liberally.

1.3.6 FORTRAN90 Parameters and Arithmetic Precision

All global parameters defined by the FORTRAN90 parameter statements are specified in the module file: SETUP_MODULE, which is included at compilation time in all subroutines requiring the parameters. All parameters specified in SETUP_MODULE are described by one or more comment cards.

One super-global parameter is defined at compilation time in the KINDS_F90 module file specifying the working precision (`wp`) by kind for real and complex variables and parameters. The default is 64-bit (double) precision, i.e. `Real(wp)`. Users wishing to compile the code with quadruple precision must ensure that their architecture and FORTRAN90 compiler can allow that and then change the default in KINDS_F90. Changing the precision to anything else that is allowed by the FORTRAN90 compiler and the machine architecture must also be compliant with the MPI working precision `mpi_wp` as defined in COMMS_MODULE (in such cases users must correct for that in there).

1.3.7 Units

Internally all DL_POLY_4 subroutines and functions assume the use of the following defined *molecular units*:

- The unit of time (t_o) is 1×10^{-12} seconds (i.e. picoseconds)
- The unit of length (ℓ_o) is 1×10^{-10} metres (i.e. Ångstroms)
- The unit of mass (m_o) is $1.6605402 \times 10^{-27}$ kilograms (i.e. Daltons - atomic mass units)
- The unit of charge (q_o) is $1.60217733 \times 10^{-19}$ Coulombs (i.e. electrons - units of proton charge)
- The unit of energy ($E_o = m_o(\ell_o/t_o)^2$) is $1.6605402 \times 10^{-23}$ Joules (10 J mol^{-1})
- The unit of pressure ($\mathcal{P}_o = E_o\ell_o^{-3}$) is 1.6605402×10^{7} Pascals (163.882576 atmospheres)
- Planck's constant ($\hbar$) which is $6.350780668 \times E_o t_o$.

In addition, the following conversion factors are used:

- The coulombic conversion factor (γ_o) is:

$$\gamma_o = \frac{1}{E_o}\left[\frac{q_o^2}{4\pi\epsilon_o\ell_o}\right] = 138935.4835 \quad ,$$

  such that:

$$U_{\mathtt{MKS}} = E_o\gamma_o U_{\mathtt{Internal}} \quad ,$$

  where U represents the configuration energy.

- The Boltzmann factor (k_B) is 0.831451115 E_o K^{-1}, such that:

$$T = E_{kin}/k_B$$

  represents the conversion from kinetic energy (in internal units) to temperature (in Kelvin).

Note: In the DL_POLY_4 OUTPUT file, the print out of pressure is in units of katms (kilo-atmospheres) at all times. The unit of energy is either DL_POLY units specified above, or in other units specified by the user at run time (see Section 6.1.3). The default is the DL_POLY unit.

Externally, DL_POLY_4 accepts information in its own specific formatting as described in Section 6.1. Irrespective of formatting rules, all values provided to define input entities are read in DL_POLY units (except otherwise specified as in the case of energy units) or their composite mixture representing the corresponding entity physically, i.e. velocities' components are in Ångsroms/picosecond.

1.3.8 Error Messages

All errors detected by DL_POLY_4 during run time initiate a call to the subroutine ERROR, which prints an error message in the standard output file and terminates the program. All terminations of the program are global (i.e. every node of the parallel computer will be informed of the termination condition and stop executing).

In addition to terminal error messages, DL_POLY_4 will sometimes print warning messages. These indicate that the code has detected something that is unusual or inconsistent. The detection is non-fatal, but the user should make sure that the warning does represent a harmless condition.

1.4 Directory Structure

The entire DL_POLY_4 package is stored in a UNIX directory structure. The topmost directory is named *dl_poly_4.nn*, where *nn* is a generation number. Beneath this directory are several sub-directories named: *source*, *utility*, *data*, *bench*, *execute*, *build*, *public*, and *java*.

Briefly, the content of each sub-directory is as follows:

sub-directory	contents
source	primary subroutines for the DL_POLY_4 package
utility	subroutines, programs and example data for all utilities
data	example input and output files for DL_POLY_4
bench	large test cases suitable for benchmarking
execute	the DL_POLY_4 run-time directory
build	makefiles to assemble and compile DL_POLY_4 programs
public	directory of routines donated by DL_POLY_4 users
java	directory of Java and FORTRAN routines for the Java GUI.

A more detailed description of each sub-directory follows.

1.4.1 The *source* Sub-directory

In this sub-directory all the essential source code for DL_POLY_4, excluding the utility software is stored. In keeping with the 'package' concept of DL_POLY_4, it does not contain any complete programs; these are assembled at compile time using an appropriate makefile. The subroutines in this sub-directory are documented in Chapter 7.

1.4.2 The *utility* Sub-directory

This sub-directory stores all the utility subroutines, functions and programs in DL_POLY_4, together with examples of data. Some of the various routines in this sub-directory are documented in the DL_POLY_Classic User Manual. Users who devise their own utilities are advised to store them in the *utility* sub-directory.

1.4.3 The *data* Sub-directory

This sub-directory contains examples of input and output files for testing the released version of DL_POLY_4. The examples of input data are copied into the *execute* sub-directory when a program is being tested. The test cases are documented in Chapter 8. Note that these are no longer within the distribution of any DL_POLY version but are made available on-line at the DL_POLY FTP - ftp://ftp.dl.ac.uk/ccp5/DL_POLY/ .

1.4.4 The *bench* Sub-directory

This directory contains examples of input and output data for DL_POLY_4 that are suitable for benchmarking DL_POLY_4 on large scale computers. These are described in Chapter 8. Note that these are no longer within the distribution of any DL_POLY version but are made available on-line at the DL_POLY FTP - ftp://ftp.dl.ac.uk/ccp5/DL_POLY/ .

1.4.5 The *execute* Sub-directory

In the supplied version of DL_POLY_4, this sub-directory contains only a few macros for copying and storing data from and to the *data* sub-directory and for submitting programs for execution (see Appendix C). However, when a DL_POLY_4 program is assembled by using the appropriate makefile, it will be placed in this sub-directory and will subsequently be executed from here. The output from the job will also appear here, so users will find it convenient to use this sub-directory if they wish to use DL_POLY_4 as intended. (The experienced user is not at all required to use DL_POLY_4 this way however.)

1.4.6 The *build* Sub-directory

This sub-directory contains the standard makefiles for the creation (i.e. compilation and linking) of the DL_POLY_4 simulation program. The makefiles supplied select the appropriate subroutines from the *source* sub-directory and deposit the executable program in the *execute* directory. The user is advised to copy the appropriate makefile into the *source* directory, in case any modifications are required. The copy in the *build* sub-directory will then serve as a backup.

1.4.7 The *public* Sub-directory

This sub-directory contains assorted routines donated by DL_POLY users. Potential users should note that these routines are **unsupported** and come **without any guarantee or liability whatsoever**. They should be regarded as potentially useful resources to be hacked into shape as needed by the user. This directory is available from the CCP5 Program Library by direct FTP(see below).

1.4.8 The *java* Sub-directory

The DL_POLY_4 Java Graphical User Interface (GUI) is based on the Java language developed by Sun. The Java source code for this GUI is to be found in this sub-directory. The source is complete and sufficient to create a working GUI, provided the user has installed the Java Development Kit, (1.4 or above) which is available free from Sun at http://java.sun.com/. The GUI, once compiled, may be executed on any machine where Java is installed [21].

1.5 Obtaining the Source Code

To obtain a copy of DL_POLY_4 it is necessary to have internet connection. Log on to the DL_POLY website - http://www.ccp5.ac.uk/DL_POLY/ , and follow the links to the DL_POLY_4 registration page, where you will firstly be shown the DL_POLY_4 academic software licence (see Appendix F), which details the terms and conditions under which the code will be supplied. **By proceeding further with the registration and download process you are signalling your acceptance of the terms of this licence.** Click the 'Registration' button to find the registration page, where you will be invited to enter your name, address and e-mail address. The code is supplied free of charge to **academic** users, but **commercial** users will be required to purchase a software licence.

Once the online registration has been completed, information on downloading the DL_POLY_4 source code will be sent by e-mail, so **it is therefore essential to supply a correct e-mail address**.

The *data* and *bench* subdirectories of DL_POLY_4 are not issued in the standard package, but can be downloaded directly from the FTP site (in the ccp5/DL_POLY/DL_POLY_4.0/ directory).

Note: Daresbury Laboratory is the **sole centre** for the distribution of DL_POLY_4 and copies obtained from elsewhere will be regarded as illegal and will not be supported.

1.6 OS and Hardware Specific Ports

Note that no support is offered for these highly specific developments!

1.7 Other Information

The DL_POLY website - http://www.ccp5.ac.uk/DL_POLY/ , provides additional information in the form of

1. Access to all documentation (including licences)
2. Frequently asked questions
3. Bug reports
4. Access to the DL_POLY online forum.

Daresbury Laboratory also maintains a DL_POLY_4 associated electronic mailing list, *dl_poly_4_news*, to which all registered DL_POLY_4 users are automatically subscribed. It is via this list that error reports and announcements of new versions are made. If you are a DL_POLY_4 user, but not on this list you may request to be added by sending a mail message to majordomo@dl.ac.uk with the one-line message: *subscribe dl_poly_4_news*.

The DL_POLY **Forum** is a web based centre for all DL_POLY users to exchange comments and queries. You may access the forum through the DL_POLY website. A registration (and vetting) process is required before you can use the forum, but it is open, in principle, to everyone.

Chapter 2

Force Field Interactions

Scope of Chapter

This chapter describes the variety of interaction potentials available in DL_POLY_4.

2.1 Introduction to the DL_POLY_4 Force Field

The force field is the set of functions needed to define the interactions in a molecular system. These may have a wide variety of analytical forms, with some basis in chemical physics, which must be parameterised to give the correct energy and forces. A huge variety of forms is possible and for this reason the DL_POLY_4 force field is designed to be agnostic and adaptable. While it is not supplied with its own force field parameters, many of the functions familiar to GROMOS [18], Dreiding [19] and AMBER [20] users have been coded in the package, as well as less familiar forms. In addition DL_POLY_4 retains the possibility of the user defining additional potentials.

In DL_POLY_4 the total configuration energy of a molecular system may be written as:

$$\begin{aligned}
U(\underline{r}_1,\underline{r}_2,\ldots,\underline{r}_N) \;=\; & \sum_{i_{shel}=1}^{N_{shel}} U_{shel}(i_{shel},\underline{r}_{core},\underline{r}_{shell}) \\
&+ \sum_{i_{teth}=1}^{N_{teth}} U_{teth}(i_{teth},\underline{r}_i^{\mathbf{t}=t},\underline{r}_i^{\mathbf{t}=0}) \\
&+ \sum_{i_{bond}=1}^{N_{bond}} U_{bond}(i_{bond},\underline{r}_a,\underline{r}_b) \\
&+ \sum_{i_{angl}=1}^{N_{angl}} U_{angl}(i_{angl},\underline{r}_a,\underline{r}_b,\underline{r}_c) \\
&+ \sum_{i_{dihd}=1}^{N_{dihd}} U_{dihd}(i_{dihd},\underline{r}_a,\underline{r}_b,\underline{r}_c,\underline{r}_d) \\
&+ \sum_{i_{inv}=1}^{N_{inv}} U_{inv}(i_{inv},\underline{r}_a,\underline{r}_b,\underline{r}_c,\underline{r}_d) \\
&+ \sum_{i=1}^{N-1}\sum_{j>i}^{N} U_{\text{2-body}}^{(metal,vdw,electostatics)}(i,j,|\underline{r}_i-\underline{r}_j|) \\
&+ \sum_{i=1}^{N}\sum_{j\neq i}^{N}\sum_{k\neq j}^{N} U_{tersoff}(i,j,k,\underline{r}_i,\underline{r}_j,\underline{r}_k) \\
&+ \sum_{i=1}^{N-2}\sum_{j>i}^{N-1}\sum_{k>j}^{N} U_{\text{3-body}}(i,j,k,\underline{r}_i,\underline{r}_j,\underline{r}_k) \\
&+ \sum_{i=1}^{N-3}\sum_{j>i}^{N-2}\sum_{k>j}^{N-1}\sum_{n>k}^{N} U_{\text{4-body}}(i,j,k,n,\underline{r}_i,\underline{r}_j,\underline{r}_k,\underline{r}_n) \\
&+ \sum_{i=1}^{N} U_{extn}(i,\underline{r}_i,\underline{v}_i) \quad ,
\end{aligned} \tag{2.1}$$

where U_{shel}, U_{teth}, U_{bond}, U_{angl}, U_{dihd}, U_{inv}, $U_{\text{2-body}}^{(metal)}$, $U_{tersoff}$, $U_{\text{3-body}}$ and $U_{\text{4-body}}$ are empirical interaction functions representing ion core-shell polarisation, tethered particles, chemical bonds, valence angles, dihedral (and improper dihedral angles), inversion angles, two-body, Tersoff, three-body and four-body forces respectively. The first six are regarded by DL_POLY_4 as *intra*-molecular interactions and the next four as *inter*-molecular interactions. The final term U_{extn} represents an *external field* potential. The position vectors $\underline{r}_a,\underline{r}_b,\underline{r}_c$ and $\underline{r}_d$ refer to the positions of the atoms specifically involved in a given interaction. (Almost universally, it is the *differences* in position that determine the interaction.) The numbers N_{shel}, N_{teth}, N_{bond}, N_{angl}, N_{dihd} and N_{inv} refer to the total numbers of these respective interactions present in the simulated system, and the indices i_{shel}, i_{teth}, i_{bond}, i_{angl}, i_{dihd} and i_{inv} uniquely specify an individual interaction of each type. It is important to note that there is no global specification of the intramolecular interactions in DL_POLY_4 - all core-shell units, tethered particles, chemical bonds, valence angles, dihedral angles and inversion angles must be individually cited. The same applies for bond constraints and PMF constraints.

The indices i, j (and k, n) appearing in the intermolecular interactions' (non-bonded) terms indicate the atoms involved in the interaction. There is normally a very large number of these and they are therefore specified globally according to the atom *types* involved rather than indices. In DL_POLY_4 it is assumed that the "pure" two-body terms arise from van der Waals interactions (regarded as short-ranged) and electrostatic interactions (coulombic, also regarded as long-ranged). Long-ranged forces require special techniques to evaluate accurately (see Section 2.4). The metal terms are many-body interactions which are functionally presented in an expansion of many two-body contributions augmented by a function of the local density, which again is derived from the two-body spatial distribution (and these are, therefore, evaluated in the two-body routines). In DL_POLY_4 the three-body terms are restricted to valence angle and H-bond forms.

Throughout this chapter the description of the force field assumes the simulated system is described as an assembly of atoms. This is for convenience only, and readers should understand that DL_POLY_4 does recognize molecular entities, defined through constraint bonds and rigid bodies. In the case of rigid bodies, the atomic forces are resolved into molecular forces and torques. These matters are discussed in greater detail in Sections 3.2 and 3.6.

2.2 The Intramolecular Potential Functions

In this section we catalogue and describe the forms of potential function available in DL_POLY_4. The **keywords** required to select potential forms are given in brackets () before each definition. The derivations of the atomic forces, virial and stress tensor are also outlined.

2.2.1 Bond Potentials

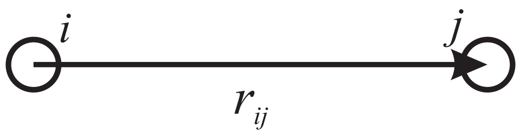

Figure 2.1: The interatomic bond vector

The bond potentials describe *explicit* chemical bonds between specified atoms. They are all functions of the interatomic distance. Only the coulomb potential makes an exception as it depends on the charges of the specified atoms. The potential functions available are as follows:

1. Harmonic bond: (**harm**)
$$U(r_{ij}) = \frac{1}{2}k(r_{ij} - r_o)^2 \tag{2.2}$$

2. Morse potential: (**mors**)
$$U(r_{ij}) = E_o[\{1 - \exp(-k(r_{ij} - r_o))\}^2 - 1] \tag{2.3}$$

3. 12-6 potential bond: (**12-6**)
$$U(r_{ij}) = \left(\frac{A}{r_{ij}^{12}}\right) - \left(\frac{B}{r_{ij}^{6}}\right) \tag{2.4}$$

4. Lennard-Jones potential: (**lj**)

$$U(r_{ij}) = 4\epsilon \left[\left(\frac{\sigma}{r_{ij}} \right)^{12} - \left(\frac{\sigma}{r_{ij}} \right)^{6} \right] \tag{2.5}$$

5. Restrained harmonic: (**rhrm**)

$$U(r_{ij}) = \begin{cases} \frac{1}{2}k(r_{ij} - r_o)^2 & : \quad |r_{ij} - r_o| \le r_c \\ \frac{1}{2}kr_c^2 + kr_c(|r_{ij} - r_o| - r_c) & : \quad |r_{ij} - r_o| > r_c \end{cases} \tag{2.6}$$

6. Quartic potential: (**quar**)

$$U(r_{ij}) = \frac{k}{2}(r_{ij} - r_o)^2 + \frac{k'}{3}(r_{ij} - r_o)^3 + \frac{k''}{4}(r_{ij} - r_o)^4 \tag{2.7}$$

7. Buckingham potential: (**buck**)

$$U(r_{ij}) = A \ \exp\left(-\frac{r_{ij}}{\rho}\right) - \frac{C}{r_{ij}^6} \tag{2.8}$$

8. Coulomb potential: (**coul**)

$$U(r_{ij}) = k \cdot U^{Electrostatics}(r_{ij}) \ \left(= \frac{k}{4\pi\epsilon_0\epsilon} \frac{q_i q_j}{r_{ij}} \right) \quad , \tag{2.9}$$

where q_ℓ is the charge on an atom labelled ℓ. It is worth noting that the Coulomb potential switches to the particular model of Electrostatics opted in CONTROL.

9. Shifted finitely extendible non-linear elastic (FENE) potential [33, 34, 35]: (**fene**)

$$U(r_{ij}) = \begin{cases} -0.5 \ k \ R_o^2 \ ln \left[1 - \left(\frac{r_{ij} - \Delta}{R_o} \right)^2 \right] & : \quad |r_{ij} - \Delta| < R_o \\ \infty & : \quad |r_{ij} - \Delta| \ge R_o \end{cases} \tag{2.10}$$

The FENE potential is used to maintain the distance between connected beads and to prevent chains from crossing each other. It is used in combination with the WCA, equation (2.94), potential to create a potential well for the flexible bonds of a molecule, that maintains the topology of the molecule. This implementation allows for a radius shift of up to half a R_o ($|\Delta| \le 0.5 \ R_o$) with a default of zero ($\Delta_{default} = 0$).

10. AMOEBA force-field bond potential [36]: (**amoe**)

$$U(r_{ij}) = k \ (r_{ij} - r_o)^2 \left[1 - 2.55 \ (r_{ij} - r_o) + (7/12) \ 2.55 \ (r_{ij} - r_o)^2 \right] \tag{2.11}$$

11. Tabulated potential: (**tab**). The potential is defined numerically in TABBND (see Section 4.3 and Section 6.1.8).

In these formulae r_{ij} is the distance between atoms labelled i and j:

$$r_{ij} = |\underline{r}_j - \underline{r}_i|^{1} \quad , \tag{2.12}$$

where $\underline{r}_\ell$ is the position vector of an atom labelled ℓ.

The force on the atom j arising from a bond potential is obtained using the general formula:

$$\underline{f}_j = -\frac{1}{r_{ij}} \left[\frac{\partial}{\partial r_{ij}} U(r_{ij}) \right] \underline{r}_{ij} \quad . \tag{2.13}$$

[1]**Note**: some DL_POLY_4 routines may use the convention that $\underline{r}_{ij} = \underline{r}_i - \underline{r}_j$.

The force $\underline{f}_i$ acting on atom i is the negative of this.

The contribution to be added to the atomic virial is given by

$$\mathcal{W} = -\underline{r}_{ij} \cdot \underline{f}_j \quad , \tag{2.14}$$

with only *one* such contribution from each bond.

The contribution to be added to the atomic stress tensor is given by

$$\sigma^{\alpha\beta} = r_{ij}^{\alpha} f_j^{\beta} \quad , \tag{2.15}$$

where α and β indicate the x, y, z components. The atomic stress tensor derived in this way is symmetric.

In DL_POLY_4 bond forces are handled by the routine BONDS_FORCES (and INTRA_COUL called within).

2.2.2 Distance Restraints

In DL_POLY_4 distance restraints, in which the separation between two atoms, is maintained around some preset value r_0 is handled as a special case of bond potentials. As a consequence, distance restraints may be applied only between atoms in the same molecule. Unlike with application of the "pure" bond potentials, the electrostatic and van der Waals interactions between the pair of atoms are still evaluated when distance restraints are applied. All the potential forms of the previous section are available as distance restraints, although they have different key words:

1. Harmonic potential: (**-hrm**)
2. Morse potential: (**-mrs**)
3. 12-6 potential bond: (**-126**)
4. Lennard-Jones potential: (**-lj**)
5. Restrained harmonic: (**-rhm**)
6. Quartic potential: (**-qur**)
7. Buckingham potential: (**-bck**)
8. Coulomb potential: (**-cul**)
9. FENE potential: (**-fne**)
10. AMOEBA force-field bond potential [36]: (**-amo**)
11. Tabulated potential: (**-tab**). The potential is defined numerically in TABBND (see Section 4.3 and Section 6.1.8).

In DL_POLY_4 distance restraints are handled by the routine BONDS_FORCES (and INTRA_COUL called within).

2.2.3 Valence Angle Potentials

The valence angle potentials describe the bond bending terms between the specified atoms. They should not be confused with the three-body potentials described later, which are defined by atom types rather than indices.

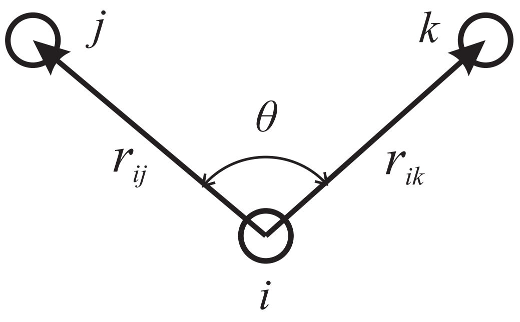

Figure 2.2: The valence angle and associated vectors

1. Harmonic: (**harm**)
$$U(\theta_{jik}) = \frac{k}{2}(\theta_{jik} - \theta_0)^2 \tag{2.16}$$

2. Quartic: (**quar**)
$$U(\theta_{jik}) = \frac{k}{2}(\theta_{jik} - \theta_0)^2 + \frac{k'}{3}(\theta_{jik} - \theta_0)^3 + \frac{k''}{4}(\theta_{jik} - \theta_0)^4 \tag{2.17}$$

3. Truncated harmonic: (**thrm**)
$$U(\theta_{jik}) = \frac{k}{2}(\theta_{jik} - \theta_0)^2 \exp[-(r_{ij}^8 + r_{ik}^8)/\rho^8] \tag{2.18}$$

4. Screened harmonic: (**shrm**)
$$U(\theta_{jik}) = \frac{k}{2}(\theta_{jik} - \theta_0)^2 \exp[-(r_{ij}/\rho_1 + r_{ik}/\rho_2)] \tag{2.19}$$

5. Screened Vessal [37]: (**bvs1**)
$$\begin{aligned} U(\theta_{jik}) \;&=\; \frac{k}{8(\theta_{jik} - \pi)^2}\left[(\theta_0 - \pi)^2 - (\theta_{jik} - \pi)^2\right]^2 \times \\ &\quad \exp[-(r_{ij}/\rho_1 + r_{ik}/\rho_2)] \end{aligned} \tag{2.20}$$

6. Truncated Vessal [38]: (**bvs2**)
$$\begin{aligned} U(\theta_{jik}) \;&=\; k\ (\theta_{jik} - \theta_0)^2\ [\ \theta_{jik}^a(\theta_{jik} + \theta_0 - 2\pi)^2 + \\ &\quad \frac{a}{2}\pi^{a-1}(\theta_0 - \pi)^3\ ]\ \ \exp[-(r_{ij}^8 + r_{ik}^8)/\rho^8] \end{aligned} \tag{2.21}$$

7. Harmonic cosine: (**hcos**)
$$U(\theta_{jik}) = \frac{k}{2}(\cos(\theta_{jik}) - \cos(\theta_0))^2 \tag{2.22}$$

8. Cosine: (**cos**)
$$U(\theta_{jik}) = A\ [1 + \cos(m\ \theta_{jik} - \delta)] \tag{2.23}$$

9. MM3 stretch-bend [39]: (**mmsb**)
$$U(\theta_{jik}) = A\ (\theta_{jik} - \theta_0)\ (r_{ij} - r_{ij}^o)\ (r_{ik} - r_{ik}^o) \tag{2.24}$$

10. Compass stretch-stretch [40]: (**stst**)
$$U(\theta_{jik}) = A\ (r_{ij} - r_{ij}^o)\ (r_{ik} - r_{ik}^o) \tag{2.25}$$

11. Compass stretch-bend [40]: (**stbe**)
$$U(\theta_{jik}) = A\ (\theta_{jik} - \theta_0)\ (r_{ij} - r_{ij}^o) \tag{2.26}$$

12. Compass all terms [40]: (**cmps**)
$$U(\theta_{jik}) = A\ (r_{ij} - r_{ij}^o)\ (r_{ik} - r_{ik}^o) + (\theta_{jik} - \theta_0)\ [B\ (r_{ij} - r_{ij}^o) + C\ (r_{ik} - r_{ik}^o)] \tag{2.27}$$

13. AMOEBA force-field angle [36]: (**amoe**)
$$\begin{aligned} U(\theta_{jik}) = k\ (\theta_{jik} - \theta_0)^2[1 \quad &- \quad 1.4 \cdot 10^{-2}(\theta_{jik} - \theta_0)\ \ + 5.6 \cdot 10^{-5}(\theta_{jik} - \theta_0)^2 \\ &- \quad 7.0 \cdot 10^{-7}(\theta_{jik} - \theta_0)^3 + 2.2 \cdot 10^{-8}(\theta_{jik} - \theta_0)^4] \end{aligned} \tag{2.28}$$

14. KKY [41]: (**kky**)
$$\begin{aligned} U(\theta_{jik}) &= f_k\ \sin\left[2(\theta_{jik} - \theta_0)\right] \cdot \sqrt{K_{ij} \cdot K_{ik}} \\ K_{ij} &= \frac{1}{\exp\left[g_r(r_{ij} - r_o)\right]} \end{aligned} \tag{2.29}$$

15. Tabulated potential: (**tab**). The potential is defined numerically in TABANG (see Section 4.3 and Section 6.1.8).

In these formulae θ_{jik} is the angle between bond vectors $\underline{r}_{ij}$ and $\underline{r}_{ik}$:
$$\theta_{jik} = cos^{-1}\left\{\frac{\underline{r}_{ij} \cdot \underline{r}_{ik}}{r_{ij} r_{ik}}\right\} \quad . \tag{2.30}$$

In DL_POLY_4 the most general form for the valence angle potentials can be written as:
$$U(\theta_{jik}, r_{ij}, r_{ik}) = A(\theta_{jik})\ S(r_{ij})\ S(r_{ik})\ S(r_{ik}) \quad , \tag{2.31}$$

where $A(\theta)$ is a purely angular function and $S(r)$ is a screening or truncation function. All the function arguments are scalars. With this reduction the force on an atom derived from the valence angle potential is given by:
$$f_\ell^\alpha = -\frac{\partial}{\partial r_\ell^\alpha} U(\theta_{jik}, r_{ij}, r_{ik}, r_{jk}) \quad , \tag{2.32}$$

with atomic label ℓ being one of i, j, k and α indicating the x, y, z component. The derivative is
$$\begin{aligned} -\frac{\partial}{\partial r_\ell^\alpha} U(\theta_{jik}, r_{ij}, r_{ik}, r_{jk}) \quad = \quad & -S(r_{ij})S(r_{ik})S(r_{jk})\frac{\partial}{\partial r_\ell^\alpha} A(\theta_{jik}) \\ & -A(\theta_{jik})S(r_{ik})S(r_{jk})(\delta_{\ell j} - \delta_{\ell i})\frac{r_{ij}^\alpha}{r_{ij}}\frac{\partial}{\partial r_{ij}} S(r_{ij}) \\ & -A(\theta_{jik})S(r_{ij})S(r_{jk})(\delta_{\ell k} - \delta_{\ell i})\frac{r_{ik}^\alpha}{r_{ik}}\frac{\partial}{\partial r_{ik}} S(r_{ik}) \\ & -A(\theta_{jik})S(r_{ij})S(r_{ik})(\delta_{\ell k} - \delta_{\ell j})\frac{r_{jk}^\alpha}{r_{jk}}\frac{\partial}{\partial r_{jk}} S(r_{jk}) \quad , \end{aligned} \tag{2.33}$$

with $\delta_{ab} = 1$ if $a = b$ and $\delta_{ab} = 0$ if $a \neq b$. In the absence of screening terms $S(r)$, this formula reduces to:
$$-\frac{\partial}{\partial r_\ell^\alpha} U(\theta_{jik}, r_{ij}, r_{ik}, r_{jk}) = -\frac{\partial}{\partial r_\ell^\alpha} A(\theta_{jik}) \quad . \tag{2.34}$$

The derivative of the angular function is

$$-\frac{\partial}{\partial r_\ell^\alpha} A(\theta_{jik}) = \left\{ \frac{1}{\sin(\theta_{jik})} \right\} \frac{\partial}{\partial \theta_{jik}} A(\theta_{jik}) \frac{\partial}{\partial r_\ell^\alpha} \left\{ \frac{\underline{r}_{ij} \cdot \underline{r}_{ik}}{r_{ij} r_{ik}} \right\} \quad , \tag{2.35}$$

with

$$\begin{aligned} \frac{\partial}{\partial r_\ell^\alpha} \left\{ \frac{\underline{r}_{ij} \cdot \underline{r}_{ik}}{r_{ij} r_{ik}} \right\} &= (\delta_{\ell j} - \delta_{\ell i}) \frac{r_{ik}^\alpha}{r_{ij} r_{ik}} + (\delta_{\ell k} - \delta_{\ell i}) \frac{r_{ij}^\alpha}{r_{ij} r_{ik}} - \\ & \cos(\theta_{jik}) \left\{ (\delta_{\ell j} - \delta_{\ell i}) \frac{r_{ij}^\alpha}{r_{ij}^2} + (\delta_{\ell k} - \delta_{\ell i}) \frac{r_{ik}^\alpha}{r_{ik}^2} \right\} \quad . \end{aligned} \tag{2.36}$$

The atomic forces are then completely specified by the derivatives of the particular functions $A(\theta)$ and $S(r)$. The contribution to be added to the atomic virial is given by

$$\mathcal{W} = -(\underline{r}_{ij} \cdot \underline{f}_j + \underline{r}_{ik} \cdot \underline{f}_k) \quad . \tag{2.37}$$

It is worth noting that in the absence of screening terms S(r), the virial is zero [42].

The contribution to be added to the atomic stress tensor is given by

$$\sigma^{\alpha\beta} = r_{ij}^\alpha f_j^\beta + r_{ik}^\alpha f_k^\beta \tag{2.38}$$

and the stress tensor is symmetric.

In DL_POLY_4 valence forces are handled by the routine ANGLES_FORCES.

2.2.4 Angular Restraints

In DL_POLY_4 angle restraints, in which the angle subtended by a triplet of atoms, is maintained around some preset value θ_0 is handled as a special case of angle potentials. As a consequence angle restraints may be applied only between atoms in the same molecule. Unlike with application of the "pure" angle potentials, the electrostatic and van der Waals interactions between the pair of atoms are still evaluated when distance restraints are applied. All the potential forms of the previous section are available as angular restraints, although they have different key words:

1. Harmonic: (**-hrm**)
2. Quartic: (**-qur**)
3. Truncated harmonic: (**-thm**)
4. Screened harmonic: (**-shm**)
5. Screened Vessal [37]: (**-bv1**)
6. Truncated Vessal [38]: (**-bv2**)
7. Harmonic cosine: (**-hcs**)
8. Cosine: (**-cos**)
9. MM3 stretch-bend [39]: (**-msb**)
10. Compass stretch-stretch [40]: (**-sts**)
11. Compass stretch-bend [40]: (**-stb**)

12. Compass all terms [40]: (**-cmp**)
13. AMOEBA force-field angle [36]: (**-amo**)
14. KKY [41]: (**-kky**)
15. Tabulated potential: (**-tab**). The potential is defined numerically in TABANG (see Section 4.3 and Section 6.1.8).

In DL_POLY_4 angular restraints are handled by the routine ANGLES_FORCES.

2.2.5 Dihedral Angle Potentials

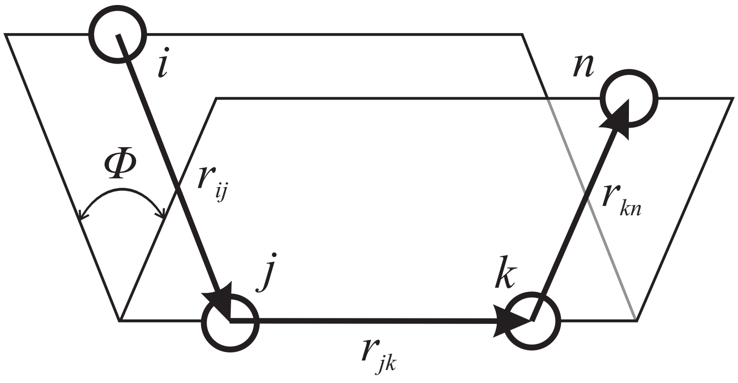

Figure 2.3: The dihedral angle and associated vectors

The dihedral angle potentials describe the interaction arising from torsional forces in molecules. (They are sometimes referred to as torsion potentials.) They require the specification of four atomic positions. The potential functions available in DL_POLY_4 are as follows:

1. Cosine potential: (**cos**)
$$U(\phi_{ijkn}) = A\ [1+\cos(m\phi_{ijkn}-\delta)] \tag{2.39}$$
2. Harmonic: (**harm**)
$$U(\phi_{ijkn}) = \frac{k}{2}\ (\phi_{ijkn}-\phi_0)^2 \tag{2.40}$$
3. Harmonic cosine: (**hcos**)
$$U(\phi_{ijkn}) = \frac{k}{2}\ (\cos(\phi_{ijkn})-\cos(\phi_0))^2 \tag{2.41}$$
4. Triple cosine: (**cos3**)
$$U(\phi) = \frac{1}{2}\ \{A_1\ (1+\cos(\phi)) + A_2\ (1-\cos(2\phi)) + A_3\ (1+\cos(3\phi))\} \tag{2.42}$$
5. Ryckaert-Bellemans [43] with fixed constants a-f: (**ryck**)
$$U(\phi) = A\ \{\ a + b\ \cos(\phi) + c\ \cos^2(\phi) + d\ \cos^3(\phi) + e\ \cos^4(\phi) + f\ \cos^5(\phi)\ \} \tag{2.43}$$
6. Fluorinated Ryckaert-Bellemans [44] with fixed constants a-h: (**rbf**)
$$U(\phi) = A\ \{\ a + b\ \cos(\phi) + c\ \cos^2(\phi) + d\ \cos^3(\phi) + e\ \cos^4(\phi) + f\ \cos^5(\phi) + g\ \exp(-h(\phi-\pi)^2))\ \} \tag{2.44}$$

7. OPLS torsion potential: (**opls**)

$$U(\phi) = A_0 + \frac{1}{2}\left\{A_1\ (1+\cos(\phi)) + A_2\ (1-\cos(2\phi)) + A_3\ (1+\cos(3\phi))\right\} \tag{2.45}$$

8. Tabulated potential: (**tab**). The potential is defined numerically in TABDIH (see Section 4.3 and Section 6.1.8).

In these formulae ϕ_{ijkn} is the dihedral angle defined by

$$\phi_{ijkn} = \cos^{-1}\{B(\underline{r}_{ij}, \underline{r}_{jk}, \underline{r}_{kn})\}\ \ , \tag{2.46}$$

with

$$B(\underline{r}_{ij}, \underline{r}_{jk}, \underline{r}_{kn}) = \left\{\frac{(\underline{r}_{ij}\times\underline{r}_{jk})\cdot(\underline{r}_{jk}\times\underline{r}_{kn})}{|\underline{r}_{ij}\times\underline{r}_{jk}||\underline{r}_{jk}\times\underline{r}_{kn}|}\right\}\ \ . \tag{2.47}$$

With this definition, the sign of the dihedral angle is positive if the vector product $(\underline{r}_{ij}\times\underline{r}_{jk})\times(\underline{r}_{jk}\times\underline{r}_{kn})$ is in the same direction as the bond vector $\underline{r}_{jk}$ and negative if in the opposite direction.

The force on an atom arising from the dihedral potential is given by

$$f_\ell^\alpha = -\frac{\partial}{\partial r_\ell^\alpha} U(\phi_{ijkn})\ \ , \tag{2.48}$$

with ℓ being one of i, j, k, n and α one of x, y, z. This may be expanded into

$$-\frac{\partial}{\partial r_\ell^\alpha} U(\phi_{ijkn}) = \left\{\frac{1}{\sin(\phi_{ijkn})}\right\}\frac{\partial}{\partial \phi_{ijkn}} U(\phi_{ijkn}) \frac{\partial}{\partial r_\ell^\alpha} B(\underline{r}_{ij}, \underline{r}_{jk}, \underline{r}_{kn})\ \ . \tag{2.49}$$

The derivative of the function $B(\underline{r}_{ij}, \underline{r}_{jk}, \underline{r}_{kn})$ is

$$\begin{aligned}\frac{\partial}{\partial r_\ell^\alpha} B(\underline{r}_{ij}, \underline{r}_{jk}, \underline{r}_{kn}) \ &= \ \frac{1}{|\underline{r}_{ij}\times\underline{r}_{jk}||\underline{r}_{jk}\times\underline{r}_{kn}|}\frac{\partial}{\partial r_\ell^\alpha}\{(\underline{r}_{ij}\times\underline{r}_{jk})\cdot(\underline{r}_{jk}\times\underline{r}_{kn})\} \ - \\ \frac{\cos(\phi_{ijkn})}{2}&\left\{\frac{1}{|\underline{r}_{ij}\times\underline{r}_{jk}|^2}\frac{\partial}{\partial r_\ell^\alpha}|\underline{r}_{ij}\times\underline{r}_{jk}|^2 + \frac{1}{|\underline{r}_{jk}\times\underline{r}_{kn}|^2}\frac{\partial}{\partial r_\ell^\alpha}|\underline{r}_{jk}\times\underline{r}_{kn}|^2\right\}\ \ ,\end{aligned} \tag{2.50}$$

with

$$\begin{aligned}\frac{\partial}{\partial r_\ell^\alpha}\{(\underline{r}_{ij}\times\underline{r}_{jk})\cdot(\underline{r}_{jk}\times\underline{r}_{kn})\} \ &= \ r_{ij}^\alpha([\underline{r}_{jk}\underline{r}_{jk}]_\alpha(\delta_{\ell k}-\delta_{\ell n}) + [\underline{r}_{jk}\underline{r}_{kn}]_\alpha(\delta_{\ell k}-\delta_{\ell j})) + \\ &\quad\ r_{jk}^\alpha([\underline{r}_{ij}\underline{r}_{jk}]_\alpha(\delta_{\ell n}-\delta_{\ell k}) + [\underline{r}_{jk}\underline{r}_{kn}]_\alpha(\delta_{\ell j}-\delta_{\ell i})) + \\ &\quad\ r_{kn}^\alpha([\underline{r}_{ij}\underline{r}_{jk}]_\alpha(\delta_{\ell k}-\delta_{\ell j}) + [\underline{r}_{jk}\underline{r}_{jk}]_\alpha(\delta_{\ell i}-\delta_{\ell j})) + \\ &\quad\ 2r_{jk}^\alpha[\underline{r}_{ij}\underline{r}_{kn}]_\alpha(\delta_{\ell j}-\delta_{\ell k})\ \ ,\end{aligned} \tag{2.51}$$

$$\begin{aligned}\frac{\partial}{\partial r_\ell^\alpha}|\underline{r}_{ij}\times\underline{r}_{jk}|^2 \ &= \ 2r_{ij}^\alpha([\underline{r}_{jk}\underline{r}_{jk}]_\alpha(\delta_{\ell j}-\delta_{\ell i}) + [\underline{r}_{ij}\underline{r}_{jk}]_\alpha(\delta_{\ell j}-\delta_{\ell k})) + \\ &\quad\ 2r_{jk}^\alpha([\underline{r}_{ij}\underline{r}_{ij}]_\alpha(\delta_{\ell k}-\delta_{\ell j}) + [\underline{r}_{ij}\underline{r}_{jk}]_\alpha(\delta_{\ell i}-\delta_{\ell j}))\ \ ,\end{aligned} \tag{2.52}$$

$$\begin{aligned}\frac{\partial}{\partial r_\ell^\alpha}|\underline{r}_{jk}\times\underline{r}_{kn}|^2 \ &= \ 2r_{kn}^\alpha([\underline{r}_{jk}\underline{r}_{jk}]_\alpha(\delta_{\ell n}-\delta_{\ell k}) + [\underline{r}_{jk}\underline{r}_{kn}]_\alpha(\delta_{\ell j}-\delta_{\ell k})) + \\ &\quad\ 2r_{jk}^\alpha([\underline{r}_{kn}\underline{r}_{kn}]_\alpha(\delta_{\ell k}-\delta_{\ell j}) + [\underline{r}_{jk}\underline{r}_{kn}]_\alpha(\delta_{\ell k}-\delta_{\ell n}))\ \ .\end{aligned} \tag{2.53}$$

Where we have used the following definition:

$$[\underline{a}\ \underline{b}]_\alpha = \sum_\beta (1-\delta_{\alpha\beta}) a^\beta b^\beta \quad . \tag{2.54}$$

Formally, the contribution to be added to the atomic virial is given by

$$\mathcal{W} = -\sum_{i=1}^{4} \underline{r}_i \cdot \underline{f}_i \quad . \tag{2.55}$$

However, it is possible to show (by tedious algebra using the above formulae, or more elegantly by thermodynamic arguments [42],) that the dihedral makes *no* contribution to the atomic virial.

The contribution to be added to the atomic stress tensor is given by

$$\begin{aligned} \sigma^{\alpha\beta} &= r_{ij}^\alpha p_i^\beta + r_{jk}^\alpha p_{jk}^\beta + r_{kn}^\alpha p_n^\beta \\ &\quad - \frac{\cos(\phi_{ijkn})}{2} \left\{ r_{ij}^\alpha g_i^\beta + r_{jk}^\alpha g_k^\beta + r_{jk}^\alpha h_j^\beta + r_{kn}^\alpha h_n^\beta \right\} \quad , \end{aligned} \tag{2.56}$$

with

$$\begin{aligned} p_i^\alpha &= (r_{jk}^\alpha [\underline{r}_{jk}\underline{r}_{kn}]_\alpha - r_{kn}^\alpha [\underline{r}_{jk}\underline{r}_{jk}]_\alpha)/(|\underline{r}_{ij}\times\underline{r}_{jk}||\underline{r}_{jk}\times\underline{r}_{kn}|) \\ p_n^\alpha &= (r_{jk}^\alpha [\underline{r}_{ij}\underline{r}_{jk}]_\alpha - r_{ij}^\alpha [\underline{r}_{jk}\underline{r}_{jk}]_\alpha)/(|\underline{r}_{ij}\times\underline{r}_{jk}||\underline{r}_{jk}\times\underline{r}_{kn}|) \\ p_{jk}^\alpha &= (r_{ij}^\alpha [\underline{r}_{jk}\underline{r}_{kn}]_\alpha + r_{kn}^\alpha [\underline{r}_{ij}\underline{r}_{jk}]_\alpha - 2r_{jk}^\alpha [\underline{r}_{ij}\underline{r}_{kn}]_\alpha)/(|\underline{r}_{ij}\times\underline{r}_{jk}||\underline{r}_{jk}\times\underline{r}_{kn}|) \\ g_i^\alpha &= 2(r_{ij}^\alpha [\underline{r}_{jk}\underline{r}_{jk}]_\alpha - r_{jk}^\alpha [\underline{r}_{ij}\underline{r}_{jk}]_\alpha)/|\underline{r}_{ij}\times\underline{r}_{jk}|^2 \\ g_k^\alpha &= 2(r_{jk}^\alpha [\underline{r}_{ij}\underline{r}_{ij}]_\alpha - r_{ij}^\alpha [\underline{r}_{ij}\underline{r}_{jk}]_\alpha)/|\underline{r}_{ij}\times\underline{r}_{jk}|^2 \\ h_j^\alpha &= 2(r_{jk}^\alpha [\underline{r}_{kn}\underline{r}_{kn}]_\alpha - r_{kn}^\alpha [\underline{r}_{jk}\underline{r}_{kn}]_\alpha)/|\underline{r}_{jk}\times\underline{r}_{kn}|^2 \\ h_n^\alpha &= 2(r_{kn}^\alpha [\underline{r}_{kn}\underline{r}_{kn}]_\alpha - r_{jk}^\alpha [\underline{r}_{jk}\underline{r}_{kn}]_\alpha)/|\underline{r}_{jk}\times\underline{r}_{kn}|^2 \quad . \end{aligned} \tag{2.57}$$

The sum of the diagonal elements of the stress tensor is zero (since the virial is zero) and the matrix is symmetric.

Lastly, it should be noted that the above description does not take into account the possible inclusion of distance-dependent 1-4 interactions, as permitted by some force fields. Such interactions are permissible in DL_POLY_4 and are described in the section on pair potentials below. DL_POLY_4 also permits scaling of the 1-4 van der Waals and Coulomb interactions by a numerical factor (see Table 6.10). **Note** that scaling is abandoned when the 1-4 members are also 1-3 members in a valence angle intercation (1-4 checks are performed in DIHEDRALS_14_CHECK routine). 1-4 interactions do, of course, contribute to the atomic virial.

In DL_POLY_4 dihedral forces are handled by the routine DIHEDRALS_FORCES (and INTRA_COUL and DIHEDRALS_14_VDW called within).

2.2.6 Improper Dihedral Angle Potentials

Improper dihedrals are used to restrict the geometry of molecules and as such need not have a simple relation to conventional chemical bonding. DL_POLY_4 makes no distinction between dihedral and improper dihedral angle functions (both are calculated by the same subroutines) and all the comments made in the preceding section apply.

An important example of the use of the improper dihedral is to conserve the structure of chiral centres in molecules modelled by united-atom centres. For example α-amino acids such as alanine ($CH_3CH(NH_2)COOH$), in which it is common to represent the CH_3 and CH groups as single centres. Conservation of the chirality of the α carbon is achieved by defining a harmonic improper dihedral angle potential with an equilibrium angle of 35.264^o. The angle is defined by vectors $\underline{r}_{12}$, $\underline{r}_{23}$ and $\underline{r}_{34}$, where the atoms 1,2,3 and 4 are shown in the following figure. The figure defines the D and L enantiomers consistent with the international (IUPAC) convention. When defining the dihedral, the atom indices are entered in DL_POLY_4 in the order 1-2-3-4.

In DL_POLY_4 improper dihedral forces are handled by the routine DIHEDRALS_FORCES.

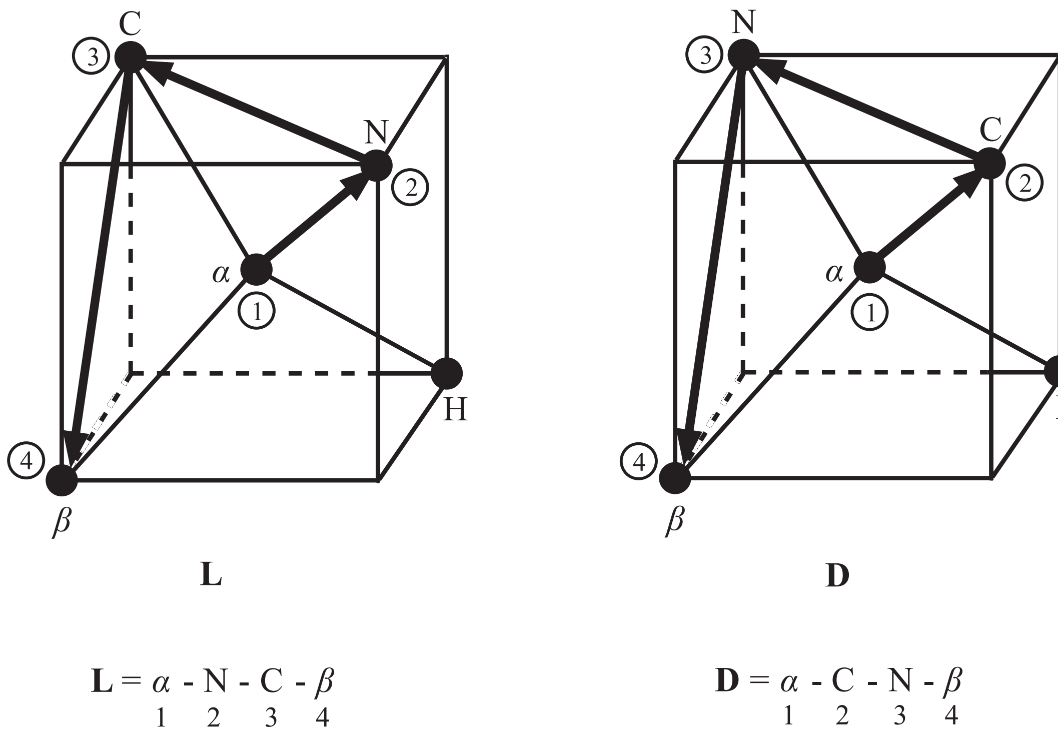

Figure 2.4: The L and D enantiomers and defining vectors

2.2.7 Torsional Restraints

In DL_POLY_4 the torsional restraints, in which the dihedral angle as defined by a quadruplet of atoms, is maintained around some preset value ϕ_0 is handled as a special case of dihedral potential. As a consequence angle restraints may be applied only between atoms in the same molecule. Unlike with application of the "pure" dihedral potentials, the electrostatic and van der Waals interactions between the pair of atoms are still evaluated when distance restraints are applied. All the potential forms of the previous section are available as torsional restraints, although they have different key words:

1. Cosine potential: (**-cos**)
2. Harmonic: (**-hrm**)
3. Harmonic cosine: (**-hcs**)
4. Triple cosine: (**-cs3**)
5. Ryckaert-Bellemans [43] with fixed constants a-f: (**-rck**)
6. Fluorinated Ryckaert-Bellemans [44] with fixed constants a-h: (**-rbf**)
7. OPLS torsion potential: (**-opl**)
8. Tabulated potential: (**-tab**). The potential is defined numerically in TABDIH (see Section 4.3 and Section 6.1.8).

In DL_POLY_4 torsional restraints are handled by the routine DIHEDRALS_FORCES.

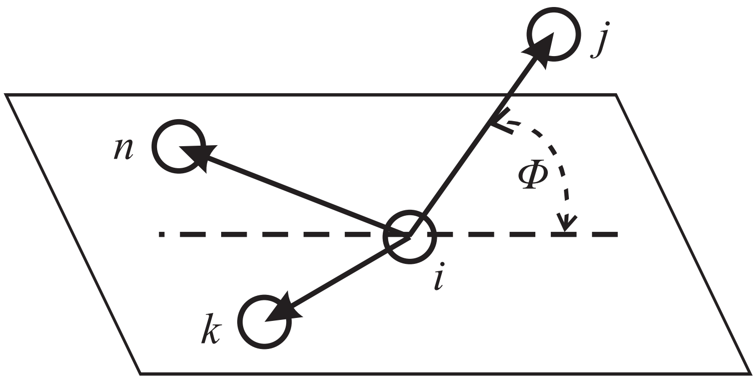

Figure 2.5: The inversion angle and associated vectors

2.2.8 Inversion Angle Potentials

The inversion angle potentials describe the interaction arising from a particular geometry of three atoms around a central atom. The best known example of this is the arrangement of hydrogen atoms around nitrogen in ammonia to form a trigonal pyramid. The hydrogens can 'flip' like an inverting umbrella to an alternative structure, which in this case is identical, but in principle causes a change in chirality. The force restraining the ammonia to one structure can be described as an inversion potential (though it is usually augmented by valence angle potentials also). The inversion angle is defined in the figure above - **note that the inversion angle potential is a sum of the three possible inversion angle terms**. It resembles a dihedral potential in that it requires the specification of four atomic positions.

The potential functions available in DL_POLY_4 are as follows:

1. Harmonic: (**harm**)
$$U(\phi_{ijkn}) = \frac{k}{2}\ (\phi_{ijkn} - \phi_0)^2 \tag{2.58}$$

2. Harmonic cosine: (**hcos**)
$$U(\phi_{ijkn}) = \frac{k}{2}\ (\cos(\phi_{ijkn}) - \cos(\phi_0))^2 \tag{2.59}$$

3. Planar potential: (**plan**)
$$U(\phi_{ijkn}) = A\ \left[1 - \cos(\phi_{ijkn})\right] \tag{2.60}$$

4. Extended planar potential: (**xpln**)
$$U(\phi_{ijkn}) = \frac{k}{2}\ \left[1 - \cos(m\ \phi_{ijkn} - \phi_0)\right] \tag{2.61}$$

5. Tabulated potential: (**tab**). The potential is defined numerically in TABINV (see Section 4.3 and Section 6.1.8).

In these formulae ϕ_{ijkn} is the inversion angle defined by

$$\phi_{ijkn} = \cos^{-1}\left\{\frac{\underline{r}_{ij} \cdot \underline{w}_{kn}}{r_{ij} w_{kn}}\right\}\quad , \tag{2.62}$$

with

$$\underline{w}_{kn} = (\underline{r}_{ij} \cdot \underline{\hat{u}}_{kn})\underline{\hat{u}}_{kn} + (\underline{r}_{ij} \cdot \underline{\hat{v}}_{kn})\underline{\hat{v}}_{kn} \tag{2.63}$$

and the unit vectors

$$\begin{aligned}\underline{\hat{u}}_{kn} &= (\underline{\hat{r}}_{ik}+\underline{\hat{r}}_{in})/|\underline{\hat{r}}_{ik}+\underline{\hat{r}}_{in}| \\ \underline{\hat{v}}_{kn} &= (\underline{\hat{r}}_{ik}-\underline{\hat{r}}_{in})/|\underline{\hat{r}}_{ik}-\underline{\hat{r}}_{in}| \quad .\end{aligned} \tag{2.64}$$

As usual, $\underline{r}_{ij} = \underline{r}_j - \underline{r}_i$ *etc.* and the hat $\underline{\hat{r}}$ indicates a *unit* vector in the direction of $\underline{r}$. The total inversion potential requires the calculation of three such angles, the formula being derived from the above using the cyclic permutation of the indices $j \to k \to n \to j$ *etc.*

Equivalently, the angle ϕ_{ijkn} may be written as

$$\phi_{ijkn} = \cos^{-1}\left\{\frac{[(\underline{r}_{ij}\cdot\underline{\hat{u}}_{kn})^2 + (\underline{r}_{ij}\cdot\underline{\hat{v}}_{kn})^2]^{1/2}}{r_{ij}}\right\} \quad . \tag{2.65}$$

Formally, the force on an atom arising from the inversion potential is given by

$$f_\ell^\alpha = -\frac{\partial}{\partial r_\ell^\alpha}U(\phi_{ijkn}) \quad , \tag{2.66}$$

with ℓ being one of i, j, k, n and α one of x, y, z. This may be expanded into

$$\begin{aligned}-\frac{\partial}{\partial r_\ell^\alpha}U(\phi_{ijkn}) &= \left\{\frac{1}{\sin(\phi_{ijkn})}\right\}\frac{\partial}{\partial\phi_{ijkn}}U(\phi_{ijkn}) \times \\ &\quad \frac{\partial}{\partial r_\ell^\alpha}\left\{\frac{[(\underline{r}_{ij}\cdot\underline{\hat{u}}_{kn})^2 + (\underline{r}_{ij}\cdot\underline{\hat{v}}_{kn})^2]^{1/2}}{r_{ij}}\right\} \quad .\end{aligned} \tag{2.67}$$

Following through, the (extremely tedious!) differentiation gives the result:

$$\begin{aligned}f_\ell^\alpha &= \left\{\frac{1}{\sin(\phi_{ijkn})}\right\}\frac{\partial}{\partial\phi_{ijkn}}U(\phi_{ijkn}) \times \\ &\left\{-(\delta_{\ell j}-\delta_{\ell i})\frac{\cos(\phi_{ijkn})}{r_{ij}^2}r_{ij}^\alpha + \frac{1}{r_{ij}w_{kn}}\left[(\delta_{\ell j}-\delta_{\ell i})\{(\underline{r}_{ij}\cdot\underline{\hat{u}}_{kn})\hat{u}_{kn}^\alpha + (\underline{r}_{ij}\cdot\underline{\hat{v}}_{kn})\hat{v}_{kn}^\alpha\}\right.\right. \\ &+(\delta_{\ell k}-\delta_{\ell i})\frac{\underline{r}_{ij}\cdot\underline{\hat{u}}_{kn}}{u_{kn}r_{ik}}\left\{r_{ij}^\alpha - (\underline{r}_{ij}\cdot\underline{\hat{u}}_{kn})\hat{u}_{kn}^\alpha - (\underline{r}_{ij}\cdot\underline{r}_{ik} - (\underline{r}_{ij}\cdot\underline{\hat{u}}_{kn})(\underline{r}_{ik}\cdot\underline{\hat{u}}_{kn}))\frac{r_{ik}^\alpha}{r_{ik}^2}\right\} \\ &+(\delta_{\ell k}-\delta_{\ell i})\frac{\underline{r}_{ij}\cdot\underline{\hat{v}}_{kn}}{v_{kn}r_{ik}}\left\{r_{ij}^\alpha - (\underline{r}_{ij}\cdot\underline{\hat{v}}_{kn})\hat{v}_{kn}^\alpha - (\underline{r}_{ij}\cdot\underline{r}_{ik} - (\underline{r}_{ij}\cdot\underline{\hat{v}}_{kn})(\underline{r}_{ik}\cdot\underline{\hat{v}}_{kn}))\frac{r_{ik}^\alpha}{r_{ik}^2}\right\} \\ &+(\delta_{\ell n}-\delta_{\ell i})\frac{\underline{r}_{ij}\cdot\underline{\hat{u}}_{kn}}{u_{kn}r_{in}}\left\{r_{ij}^\alpha - (\underline{r}_{ij}\cdot\underline{\hat{u}}_{kn})\hat{u}_{kn}^\alpha - (\underline{r}_{ij}\cdot\underline{r}_{in} - (\underline{r}_{ij}\cdot\underline{\hat{u}}_{kn})(\underline{r}_{in}\cdot\underline{\hat{u}}_{kn}))\frac{r_{in}^\alpha}{r_{in}^2}\right\} \\ &\left.\left.-(\delta_{\ell n}-\delta_{\ell i})\frac{\underline{r}_{ij}\cdot\underline{\hat{v}}_{kn}}{v_{kn}r_{in}}\left\{r_{ij}^\alpha - (\underline{r}_{ij}\cdot\underline{\hat{v}}_{kn})\hat{v}_{kn}^\alpha - (\underline{r}_{ij}\cdot\underline{r}_{in} - (\underline{r}_{ij}\cdot\underline{\hat{v}}_{kn})(\underline{r}_{in}\cdot\underline{\hat{v}}_{kn}))\frac{r_{in}^\alpha}{r_{in}^2}\right\}\right]\right\} \quad .\end{aligned} \tag{2.68}$$

This general formula applies to all atoms $\ell = i, j, k, n$. It must be remembered however, that these formulae apply to just one of the three contributing terms (i.e. one angle ϕ) of the full inversion potential: specifically the inversion angle pertaining to the out-of-plane vector $\underline{r}_{ij}$. The contributions arising from the other vectors $\underline{r}_{ik}$ and $\underline{r}_{in}$ are obtained by the cyclic permutation of the indices in the manner described above. All these force contributions must be added to the final atomic forces.

Formally, the contribution to be added to the atomic virial is given by

$$\mathcal{W} = -\sum_{i=1}^{4}\underline{r}_i\cdot\underline{f}_i \quad . \tag{2.69}$$

However, it is possible to show by thermodynamic arguments (*cf* [42],) or simply from the fact that the sum of forces on atoms j,k and n is equal and opposite to the force on atom i, that the inversion potential makes *no* contribution to the atomic virial.

If the force components f_ℓ^α for atoms $\ell = i, j, k, n$ are calculated using the above formulae, it is easily seen that the contribution to be added to the atomic stress tensor is given by

$$\sigma^{\alpha\beta} = r_{ij}^\alpha f_j^\beta + r_{ik}^\alpha f_k^\beta + r_{in}^\alpha f_n^\beta \quad . \tag{2.70}$$

The sum of the diagonal elements of the stress tensor is zero (since the virial is zero) and the matrix is symmetric.

In DL_POLY_4 inversion forces are handled by the routine INVERSIONS_FORCES.

2.2.9 The Calcite Four-Body Potential

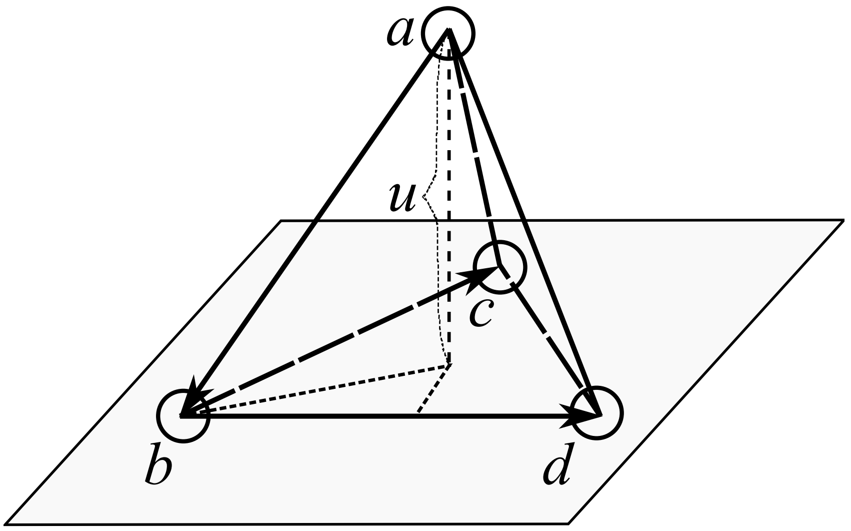

Figure 2.6: The vectors of the calcite potential

This potential [45, 46] is designed to help maintain the planar structure of the carbonate anion $[CO_3]^{2-}$ in a similar manner to the planar inversion potential described above. However, it is *not* an angular potential. It is dependent on the perpendicular displacement (u) of an atom a from a plane defined by three other atoms b, c, and d (see Figure 2.6) and has the form:

$$U_{abcd}(u) = Au^2 + Bu^4 \quad , \tag{2.71}$$

where the displacement u is given by

$$u = \frac{\underline{r}_{ab} \cdot \underline{r}_{bc} \times \underline{r}_{bd}}{|\underline{r}_{bc} \times \underline{r}_{bd}|} \quad . \tag{2.72}$$

Vectors $\underline{r}_{ab}$,$\underline{r}_{ac}$ and $\underline{r}_{ad}$ define bonds between the central atom a and the peripheral atoms b, c and d. Vectors $\underline{r}_{bc}$ and $\underline{r}_{bd}$ define the plane and are related to the bond vectors by

$$\begin{aligned} \underline{r}_{bc} &= \underline{r}_{ac} - \underline{r}_{ab} \\ \underline{r}_{bd} &= \underline{r}_{ad} - \underline{r}_{ab} \quad . \end{aligned} \tag{2.73}$$

In what follows it is convenient to define the vector product appearing in both the numerator and denominator of equation (2.72) as the vector $\underline{w}_{cd}$ *vis.*

$$\underline{w}_{cd} = \underline{r}_{bc} \times \underline{r}_{bd} \quad . \tag{2.74}$$

We also define the quantity $\gamma(u)$ as

$$\gamma(u) = -(2Au + 4Bu^3) \quad . \tag{2.75}$$

The forces on the individual atoms due to the calcite potential are then given by

$$\begin{aligned} \underline{f}_a &= -\gamma(u)\ \underline{\hat{w}}_{cd} \\ \underline{f}_c &= \quad \underline{r}_{bd} \times (\underline{r}_{ab} - u\underline{\hat{w}}_{cd})\ \gamma(u)/w_{cd} \\ \underline{f}_d &= -\underline{r}_{bc} \times (\underline{r}_{ab} - u\underline{\hat{w}}_{cd})\ \gamma(u)/w_{cd} \\ \underline{f}_b &= -(\underline{f}_a + \underline{f}_c + \underline{f}_d) \quad , \end{aligned} \tag{2.76}$$

where $w_{cd} = |\underline{w}_{cd}|$ and $\underline{\hat{w}}_{cd} = \underline{w}_{cd}/w_{cd}$. The virial contribution $\psi_{abcd}(u)$ is given by

$$\psi_{abcd}(u) = 2Au^2 + 4Bu^4 \tag{2.77}$$

and the stress tensor contribution $\sigma^{\alpha\beta}_{abcd}(u)$ by

$$\sigma^{\alpha\beta}_{abcd}(u) = \frac{u\ \gamma(u)}{w^2_{cd}}\ w^{\alpha}_{cd}\ w^{\beta}_{cd} \quad . \tag{2.78}$$

In DL_POLY_4 the calcite forces are handled by the routine INVERSIONS_FORCES, which is a convenient *intramolecular* four-body force routine. However, it is manifestly *not* an inversion potential as such.

2.2.10 Inversional Restraints

In DL_POLY_4 the inversional restraints, in which the inversion angle, as defined by a quadruplet of atoms, is maintained around some preset value ϕ_0, is handled as a special case of inversion potential. As a consequence angle restraints may be applied only between atoms in the same molecule. Unlike with application of the "pure" dihedral potentials, the electrostatic and van der Waals interactions between the pair of atoms are still evaluated when distance restraints are applied. All the potential forms of the previous section are available as torsional restraints, although they have different key words:

1. Harmonic: (**-hrm**)
2. Harmonic cosine: (**-hcs**)
3. Planar potential: (**-pln**)
4. Extended planar potential: (**-xpl**)
5. Tabulated potential: (**-tab**). The potential is defined numerically in TABINV (see Section 4.3 and Section 6.1.8).

In DL_POLY_4 inversional restraints are handled by the routine INVERSIONS_FORCES.

2.2.11 Tethering Forces

DL_POLY_4 also allows atomic sites to be tethered to a fixed point in space, $\vec{r_0}$, taken as their position at the beginning of the simulation (t = 0). This is also known as position restraining. The specification, which comes as part of the molecular description, requires a tether potential type and the associated interaction parameters.

Note, firstly, that application of tethering potentials means that the momentum will no longer be a conserved quantity of the simulation. Secondly, in constant pressure simulations, where the MD cell changes size or shape, the tethers' reference positions are scaled with the cell vectors.

The tethering potential functions available in DL_POLY_4 are as follows:

1. Harmonic: (**harm**)

$$U(r_{ij}) = \frac{1}{2} k (r_{i0})^2 \tag{2.79}$$

2. Restrained harmonic: (**rhrm**)

$$U(r_{ij}) = \begin{cases} \frac{1}{2} k (r_{i0})^2 & : \quad |r_{i0}| \le r_c \\ \frac{1}{2} k r_c^2 + k r_c (r_{i0} - r_c) & : \quad |r_{i0}| > r_c \end{cases} \tag{2.80}$$

3. Quartic potential: (**quar**)

$$U(r_{ij}) = \frac{k}{2} (r_{i0})^2 + \frac{k'}{3} (r_{i0})^3 + \frac{k''}{4} (r_{i0})^4 \tag{2.81}$$

as in each case r_{io} is the distance between the atom positions at moment $t = t1$ and $t = 0$.

The force on the atom i arising from a tether potential potential is obtained using the general formula:

$$\underline{f}_i = -\frac{1}{r_{i0}} \left[\frac{\partial}{\partial r_{i0}} U(r_{i0}) \right] \underline{r}_{i0} \quad . \tag{2.82}$$

The contribution to be added to the atomic virial is given by

$$\mathcal{W} = \underline{r}_{i0} \cdot \underline{f}_i \quad . \tag{2.83}$$

The contribution to be added to the atomic stress tensor is given by

$$\sigma^{\alpha\beta} = -r_{i0}^{\alpha} f_i^{\beta} \quad , \tag{2.84}$$

where α and β indicate the x, y, z components. The atomic stress tensor derived in this way is symmetric.

In DL_POLY_4 tether forces are handled by the routine TETHERS_FORCES.

2.3 The Intermolecular Potential Functions

In this section we outline the two-body, metal, Tersoff, three-body and four-body potential functions in DL_POLY_4. An important distinction between these and intramolecular (bond) forces in DL_POLY_4 is that they are specified by *atom types* rather than atom indices.

2.3.1 Short Ranged (van der Waals) Potentials

The short ranged pair forces available in DL_POLY_4 are as follows:

1. 12-6 potential: (**12-6**)

$$U(r_{ij}) = \left(\frac{A}{r_{ij}^{12}} \right) - \left(\frac{B}{r_{ij}^{6}} \right) \tag{2.85}$$

2. Lennard-Jones potential: (**lj**)

$$U(r_{ij}) = 4\epsilon \left[\left(\frac{\sigma}{r_{ij}} \right)^{12} - \left(\frac{\sigma}{r_{ij}} \right)^{6} \right] \tag{2.86}$$

3. n-m potential [47, 48]: (**nm**)

$$U(r_{ij}) = \frac{E_o}{(n-m)} \left[m \left(\frac{r_o}{r_{ij}} \right)^{n} - n \left(\frac{r_o}{r_{ij}} \right)^{m} \right] \tag{2.87}$$

4. Buckingham potential: (**buck**)
$$U(r_{ij}) = A\ \exp\left(-\frac{r_{ij}}{\rho}\right) - \frac{C}{r_{ij}^6} \tag{2.88}$$

5. Born-Huggins-Meyer potential: (**bhm**)
$$U(r_{ij}) = A\ \exp[B(\sigma - r_{ij})] - \frac{C}{r_{ij}^6} - \frac{D}{r_{ij}^8} \tag{2.89}$$

6. Hydrogen-bond (12-10) potential: (**hbnd**)
$$U(r_{ij}) = \left(\frac{A}{r_{ij}^{12}}\right) - \left(\frac{B}{r_{ij}^{10}}\right) \tag{2.90}$$

7. Shifted force n-m potential [47, 48]: (**snm**)
$$\begin{aligned} U(r_{ij}) \quad = \quad & \frac{\alpha E_o}{(n-m)}\left[m\beta^n\left\{\left(\frac{r_o}{r_{ij}}\right)^n - \left(\frac{1}{\gamma}\right)^n\right\} - n\beta^m\left\{\left(\frac{r_o}{r_{ij}}\right)^m - \left(\frac{1}{\gamma}\right)^m\right\}\right] + \\ & \frac{nm\alpha E_o}{(n-m)}\left(\frac{r_{ij}-\gamma r_o}{\gamma r_o}\right)\left\{\left(\frac{\beta}{\gamma}\right)^n - \left(\frac{\beta}{\gamma}\right)^m\right\} \quad , \end{aligned} \tag{2.91}$$
with
$$\begin{aligned} \alpha &= \frac{(n-m)}{[n\beta^m(1+(m/\gamma-m-1)/\gamma^m) - m\beta^n(1+(n/\gamma-n-1)/\gamma^n)]} \\ \beta &= \gamma\left(\frac{\gamma^{m+1}-1}{\gamma^{n+1}-1}\right)^{\frac{1}{n-m}} \\ \gamma &= \frac{r_{\rm cut}}{r_o} \quad . \end{aligned} \tag{2.92}$$
This peculiar form has the advantage over the standard shifted n-m potential in that both E_o and r_0 (well depth and location of minimum) retain their original values after the shifting process.

8. Morse potential: (**mors**)
$$U(r_{ij}) = E_o\ [\{1 - \exp(-k(r_{ij} - r_o))\}^2 - 1] \tag{2.93}$$

9. Shifted Weeks-Chandler-Anderson (WCA) potential [49]: (**wca**)
$$U(r_{ij}) = \begin{cases} 4\epsilon\left[\left(\frac{\sigma}{r_{ij}-\Delta}\right)^{12} - \left(\frac{\sigma}{r_{ij}-\Delta}\right)^6\right] + \epsilon & : \quad r_{ij} < 2^{\frac{1}{6}}\ \sigma + \Delta \\ 0 & : \quad r_{ij} \geq 2^{\frac{1}{6}}\ \sigma + \Delta \end{cases} \tag{2.94}$$
The WCA potential is the Lennard-Jones potential truncated at the position of the minimum and shifted to eliminate discontinuity (includes the effect of excluded volume). It is usually used in combination with the FENE, equation (2.10), bond potential. This implementation allows for a radius shift of up to half a σ ($|\Delta| \leq 0.5\ \sigma$) with a default of zero ($\Delta_{default} = 0$).

10. Standard DPD potential: (**dpd**)
$$U(r_{ij}) = \begin{cases} \frac{A}{2}\ r_c\ \left(1 - \frac{r_{ij}}{r_c}\right)^2 & : \quad r_{ij} < r_c \\ 0 & : \quad r_{ij} \geq r_c \end{cases} \tag{2.95}$$
It takes the Groot-Warren [50] form giving a soft and purely repulsive interaction.

11. AMOEBA force-field 14-7 pair potential [36]: (**amoe**)

$$U(r_{ij}) = \epsilon \left(\frac{1.07}{(r_{ij}/r_o) + 0.07} \right)^7 \left(\frac{1.12}{(r_{ij}/r_o)^7 + 0.12} - 2 \right) \tag{2.96}$$

12. Tabulation: (**tab**). The potential is defined numerically only.

The parameters defining these potentials are supplied to DL_POLY_4 at run time (see the description of the FIELD file in Section 6.1.3). Each atom type in the system is specified by a unique eight-character label defined by the user. The pair potential is then defined internally by the combination of two atom labels.

As well as the numerical parameters defining the potentials, DL_POLY_4 should also be provided with a cutoff radius, $r_{\rm vdw}$, which sets a range limit on the computation of the interactions. It is worth noting that some interaction come with a hard-wired cutoff in their parameter sets! Thus any provided cutoff radius, $r_{\rm vdw}$, will be reset if it is not equal or larger that the largest of these all. Together with the parameters, the cutoff is used by the subroutine VDW_GENERATE to construct an interpolation array `vvdw` for the potential function over the range 0 to $r_{\rm vdw}$. A second array `gvdw` is also calculated, which is related to the potential via the formula:

$$G(r_{ij}) = -r_{ij} \frac{\partial}{\partial r_{ij}} U(r_{ij}) \quad , \tag{2.97}$$

and is used in the calculation of the forces. Both arrays are tabulated in units of energy. The use of interpolation arrays, rather than the explicit formulae, makes the routines for calculating the potential energy and atomic forces very general, and enables the use of user defined pair potential functions. DL_POLY_4 also allows the user to read in the interpolation arrays directly from a file (implemented in the VDW_TABLE_READ routine) and the TABLE file (Section 6.1.6). This is particularly useful if the pair potential function has no simple analytical description (e.g. spline potentials).

The force on an atom j derived from one of these potentials is formally calculated with the standard formula:

$$\underline{f}_j = -\frac{1}{r_{ij}} \left[\frac{\partial}{\partial r_{ij}} U(r_{ij}) \right] \underline{r}_{ij} \quad , \tag{2.98}$$

where $\underline{r}_{ij} = \underline{r}_j - \underline{r}_i$. The force on atom i is the negative of this.

The contribution to be added to the atomic virial (for each pair interaction) is

$$\mathcal{W} = -\underline{r}_{ij} \cdot \underline{f}_j \quad . \tag{2.99}$$

The contribution to be added to the atomic stress tensor is given by

$$\sigma^{\alpha\beta} = r_{ij}^{\alpha} f_j^{\beta} \quad , \tag{2.100}$$

where α and β indicate the x, y, z components. The atomic stress tensor derived from the pair forces is symmetric.

Since the calculation of pair potentials assumes a spherical cutoff ($r_{\rm vdw}$) it is necessary to apply a *long-ranged correction* to the system potential energy and virial. Explicit formulae are needed for each case and are derived as follows. For two atom types a and b, the correction for the potential energy is calculated via the integral

$$U_{corr}^{ab} = 2\pi \frac{N_a N_b}{V} \int_{r_{\rm vdw}}^{\infty} g_{ab}(r) U_{ab}(r) r^2 dr \quad , \tag{2.101}$$

where N_a, N_b are the numbers of atoms of types a and b in the system, V is the system volume and $g_{ab}(r)$ and $U_{ab}(r)$ are the appropriate pair correlation function and pair potential respectively. It is usual to assume $g_{ab}(r) = 1$ for $r > r_{\rm vdw}$. DL_POLY_4 sometimes makes the additional assumption that the repulsive part of the short ranged potential is negligible beyond $r_{\rm vdw}$.

The correction for the system virial is

$$\mathcal{W}^{ab}_{corr} = -2\pi \frac{N_a N_b}{V} \int_{r_{\rm vdw}}^{\infty} g_{ab}(r) \frac{\partial}{\partial r} U_{ab}(r) r^3 dr \quad , \tag{2.102}$$

where the same approximations are applied.

Note that these formulae are based on the assumption that the system is reasonably isotropic beyond the cutoff.

DL_POLY_4 allows a short cut for mixing some of the explicitly specified pair interactions for single species of the same type so that cross-species interactions are generated if unspecified. This is only possible for the **12-6, lj, dpd, amoeba, & wca** types. The mixing is derived from the Lennard-Jones style characteristic paramteres for energy (ϵ) and distance (σ or r_0) terms. The available types of mixing within DL_POLY_4 are borrowed from [51]. The rules' names and formulae are as follows:

1. Lorentz-Berthelot

$$\epsilon_{ij} = \sqrt{\epsilon_i \; \epsilon_j} \quad ; \quad \sigma_{ij} = \frac{\sigma_i + \sigma_j}{2} \tag{2.103}$$

2. Fender-Halsey

$$\epsilon_{ij} = 2 \frac{\epsilon_i \; \epsilon_j}{\epsilon_i + \epsilon_j} \quad ; \quad \sigma_{ij} = \frac{\sigma_i + \sigma_j}{2} \tag{2.104}$$

3. Hogervorst (good hope)

$$\epsilon_{ij} = \sqrt{\epsilon_i \; \epsilon_j} \quad ; \quad \sigma_{ij} = \sqrt{\sigma_i \; \sigma_j} \tag{2.105}$$

4. Halgren HHG

$$\epsilon_{ij} = 4 \frac{\epsilon_i \; \epsilon_j}{\left(\epsilon_i^{1/2} + \epsilon_j^{1/2}\right)^2} \quad ; \quad \sigma_{ij} = \frac{\sigma_i^3 + \sigma_j^3}{\sigma_i^2 + \sigma_j^2} \tag{2.106}$$

5. Waldman-Hagler

$$\epsilon_{ij} = 2\sqrt{\epsilon_i \; \epsilon_j} \frac{(\sigma_i \; \sigma_j)^3}{\sigma_i^6 + \sigma_j^6} \quad ; \quad \sigma_{ij} = \left(\frac{\sigma_i^6 + \sigma_j^6}{2}\right)^{\frac{1}{6}} \tag{2.107}$$

6. Tang-Toennies

$$\epsilon_{ij}\sigma_{ij}^6 = \sqrt{\epsilon_i \sigma_i^6 \; \epsilon_j \sigma_j^6} \quad ; \quad \epsilon_{ij}\sigma_{ij}^{12} = \left[\frac{\left(\epsilon_i \sigma_i^{12}\right)^{13} + \left(\epsilon_j \sigma_j^{12}\right)^{13}}{2}\right]^{13} \tag{2.108}$$

7. Functional

$$\epsilon_{ij} = \frac{3 \; \sqrt{\epsilon_i \; \epsilon_j} \; (\sigma_i \; \sigma_j)^3}{\sum\limits_{L=0}^{2} \left[\frac{\left(\sigma_i^3 + \sigma_j^3\right)^2}{4 \; (\sigma_i \; \sigma_i)^L}\right]^{\frac{6}{6-2L}}} \quad ; \quad \sigma_{ij} = \frac{1}{3} \sum_{L=0}^{2} \left[\frac{\left(\sigma_i^3 + \sigma_j^3\right)^2}{4 \; (\sigma_i \; \sigma_i)^L}\right]^{\frac{1}{6-2L}} \tag{2.109}$$

It is woth noting that the i and j symbols in the equations for mixing denote atom types (species) and the indices for the same species interaction parameters are contracted to a single species index for simplicity.

In DL_POLY_4 the short ranged forces are calculated by the subroutine VDW_FORCES. The long-ranged corrections are calculated by routine VDW_LRC. The calculation makes use of the Verlet neighbour list (see above).

2.3.2 Metal Potentials

The metal potentials in DL_POLY_4 follow two similar but distinct formalisms. The first of these is the embedded atom model (EAM) [11, 12] and the second is the Finnis-Sinclair model (FS) [13]. Both are density dependent potentials derived from density functional theory (DFT) and describe the bonding of a metal atom ultimately in terms of the local electronic density. They are suitable for calculating the properties of metals and metal alloys. The extended EAM (EEAM) [52, 53] is a generalisation of the EAM formalism which can include both EAM and FS type of mixing rules (see below).

It is worth noting that the same formalism applies to the many-body perturbation component of the actinide oxide potentials as in [54]. Thus their many-body component description is included in this Section.

For single component metals the two main approaches, FS and EAM, are the same. **However**, they are subtly different in the way they are extended to handle alloys (see below). It follows that EAM and FS class potentials cannot be mixed in a single simulation. Furthermore, even for FS class potentials possessing different analytical forms there is no agreed procedure for mixing the parameters. Mixing EAM and EEAM potentials is only possible if the EAM ones are generalised to EEAM form (see below). The user is, therefore, strongly advised to be consistent in the choice of potential when modelling alloys.

The general form of the EAM and FS types of potentials is [55]

$$U_{metal} = \frac{1}{2}\sum_{i=1}^{N}\sum_{j\neq i}^{N} V_{ij}(r_{ij}) + \sum_{i=1}^{N} F(\rho_i) \quad , \tag{2.110}$$

where $F(\rho_i)$ is a functional describing the energy of embedding an atom in the bulk density, ρ_i, which is defined as

$$\rho_i = \sum_{j=1,j\neq i}^{N} \rho_{ij}(r_{ij}) \quad . \tag{2.111}$$

It should be noted that the density is determined by the coordination number of the atom defined by *pairs* of atoms. This makes the metal potential dependent on the local density (environmental). $V_{ij}(r_{ij})$ is a pair potential incorporating repulsive electrostatic and overlap interactions. N is the number of interacting particles in the MD box.

In DL_POLY_4 EAM and thus EEAM can be further generalised to include two-band (2B) densities [56, 57], for s- and d-bands,

$$F(\rho_i) = F^s(\rho_i^s) + F^d(\rho_i^d) \quad , \tag{2.112}$$

where

$$\rho_i^q = \sum_{j=1,j\neq i}^{N} \rho_{ij}^q(r_{ij}) \ , \quad q = s, d \quad , \tag{2.113}$$

instead of just the one, s, as in equations (2.110) and (2.111). These will be referred in the following text as 2BEAM and 2BEEAM. Mixing 2BEAM and EAM and alternatively 2BEEAM and EEAM potentials is only possible if the single band ones are generalised to 2B forms. The user is, again, reminded to be consistent in the choice of potential when modelling alloys.

The types of metal potentials available in DL_POLY_4 are as follows:

1. EAM potential: (**eam**) There are no explicit mathematical expressions for EAM potentials, so this potential type is read exclusively in the form of interpolation arrays from the TABEAM table file (as implemented in the METAL_TABLE_READ routine - Section 6.1.7.) The rules for combining the potentials from different metals to handle alloys are different from the FS class of potentials (see below).

2. EEAM potential (**eeam**) Similar to EAM above, it is given in the form of interpolation arrays from the TABEAM file, but the rules for combining the potentials from different metals are different from both EAM and FS classes (see below).

3. 2BEAM potential (**2beam**) Similar to EEAM for the s density terms and to EAM for the d ones. It is and given in the form of interpolation arrays from the TABEAM file, but the rules for combining the potentials from different metals are different from both EAM, EEAM and FS classes (see below).

4. 2BEEAM potential (**2beeam**) Similar to EEAM for both s and d density terms. It is and given in the form of interpolation arrays from the TABEAM file, but the rules for combining the potentials from different metals are different from both EAM, EEAM, 2BEAM and FS classes (see below).

5. Finnis-Sinclair potential [13]: (**fnsc**) Finnis-Sinclair potential is explicitly analytical. It has the following form:

$$\begin{aligned} V_{ij}(r_{ij}) &= \begin{cases} (r_{ij}-c)^2(c_0+c_1 r_{ij}+c_2 r_{ij}^2) & : \quad r_{ij}<c \\ 0 & : \quad r_{ij}>c \end{cases} \\ \rho_{ij}(r_{ij}) &= \begin{cases} (r_{ij}-d)^2+\beta\dfrac{(r_{ij}-d)^3}{d} & : \quad r_{ij}<d \\ 0 & : \quad r_{ij}>d \end{cases} \\ F(\rho_i) &= -A\sqrt{\rho_i} \quad , \end{aligned} \tag{2.114}$$

with parameters: c_0, c_1, c_2, c, A, d, β, both c and d are cutoffs. Since first being proposed a number of alternative analytical forms have been proposed, some of which are described below. The rules for combining different metal potentials to model alloys are different from the EAM potentials (see below).

6. Extended Finnis-Sinclair potential [58]: (**exfs**) It has the following form:

$$\begin{aligned} V_{ij}(r_{ij}) &= \begin{cases} (r_{ij}-c)^2(c_0+c_1 r_{ij}+c_2 r_{ij}^2+c_3 r_{ij}^3+c_4 r_{ij}^4) & : \quad r_{ij}<c \\ 0 & : \quad r_{ij}>c \end{cases} \\ \rho_{ij}(r_{ij}) &= \begin{cases} (r_{ij}-d)^2+B^2(r_{ij}-d)^4 & : \quad r_{ij}<d \\ 0 & : \quad r_{ij}>d \end{cases} \\ F(\rho_i) &= -A\sqrt{\rho_i} \quad , \end{aligned} \tag{2.115}$$

with parameters: c_0, c_1, c_2, c_3, c_4, c, A, d, B, both c and d are cutoffs.

7. Sutton-Chen potential [14, 15, 16]: (**stch**) The Sutton Chen potential is an analytical potential in the FS class. It has the form:

$$\begin{aligned} V_{ij}(r_{ij}) &= \epsilon\left(\frac{a}{r_{ij}}\right)^n \\ \rho_{ij}(r_{ij}) &= \left(\frac{a}{r_{ij}}\right)^m \\ F(\rho_i) &= -c\epsilon\sqrt{\rho_i} \quad , \end{aligned} \tag{2.116}$$

with parameters: ϵ, a, n, m, c. **Note** that the parameter c for the mixed potential in multi-component allys is irrelevant as outlined in [15]!

8. Gupta potential [59]: (**gupt**) The Gupta potential is another analytical potential in the FS class. It has the form:

$$\begin{aligned} V_{ij}(r_{ij}) &= 2A\exp\left(-p\frac{r_{ij}-r_0}{r_0}\right) \\ \rho_{ij}(r_{ij}) &= \exp\left(-2q_{ij}\frac{r_{ij}-r_0}{r_0}\right) \\ F(\rho_i) &= -B\sqrt{\rho_i} \quad , \end{aligned} \tag{2.117}$$

with parameters: A, r_0, p, B, q_{ij}.

9. Many body perturbation component potential [54]: (**mbpc**) This component is another analytical potential in the FS class which two body part may be defined by a matching van der Waals potential in the vdw section of the FIELD file. It has the form:

$$\begin{aligned} V_{ij}(r_{ij}) &= 0 \\ \rho_{ij}(r_{ij}) &= \left(\frac{a}{r_{ij}^m}\right) \frac{1}{2} \left[1 + \text{erf}\left(\alpha(r_{ij} - r_\text{o})\right)\right] \\ F(\rho_i) &= -\epsilon\sqrt{\rho_i} \ , \end{aligned} \tag{2.118}$$

with parameters: ϵ, a, m, α and r_o.

Note that the parameters α and r_o must be the same for all defined potentials of this type. DL_POLY_4 will set $\alpha = \text{Max}(0, \alpha_{pq})$ and $r_\text{o} = \text{Max}(0, r_{\text{o}\_pq})$ for all defined interactions of this type between species p and q. If after this any is left undefined, i.e. zero, the undefined entities will be set to their defaults: $\alpha = 20$ and $r_\text{o} = \text{Min}(1.5, 0.2\ r_\text{cut})$.

All of these metal potentials can be decomposed into pair contributions and thus fit within the general tabulation scheme of DL_POLY_4, where they are treated as pair interactions (though note that the metal cutoff, r_met has nothing to do with short ranged cutoff, r_vdw). DL_POLY_4 calculates this potential in two stages: the first calculates the local density, ρ_i, for each atom; and the second calculates the potential energy and forces. Interpolation arrays, `vmet`, `gmet` and `fmet` (METAL_GENERATE, METAL_TABLE_READ) are used in both these stages in the same spirit as in the van der Waals interaction calculations.

The total force $\underline{f}_k^{tot}$ on an atom k derived from this potential is calculated in the standard way:

$$\underline{f}_k^{tot} = -\underline{\nabla}_k U_{metal} \quad . \tag{2.119}$$

We rewrite the EAM/FS potential, equation (2.110), as

$$\begin{aligned} U_{metal} &= U_1 + U_2 \\ U_1 &= \frac{1}{2}\sum_{i=1}^{N}\sum_{j\neq i}^{N} V_{ij}(r_{ij}) \\ U_2 &= \sum_{i=1}^{N} F(\rho_i) \ , \end{aligned} \tag{2.120}$$

where $\underline{r}_{ij} = \underline{r}_j - \underline{r}_i$. The force on atom k is the sum of the derivatives of U_1 and U_2 with respect to $\underline{r}_k$, which is recognisable as a sum of pair forces:

$$\begin{aligned} -\frac{\partial U_1}{\partial \underline{r}_k} &= -\frac{1}{2}\sum_{i=1}^{N}\sum_{j\neq i}^{N} \frac{\partial V_{ij}(r_{ij})}{\partial r_{ij}} \frac{\partial r_{ij}}{\partial \underline{r}_k} = \sum_{j=1, j\neq k}^{N} \frac{\partial V_{kj}(r_{kj})}{\partial r_{kj}} \frac{\underline{\underline{r}}_{kj}}{r_{kj}} \\ -\frac{\partial U_2}{\partial \underline{r}_k} &= -\sum_{i=1}^{N} \frac{\partial F}{\partial \rho_i} \sum_{j\neq i}^{N} \frac{\partial \rho_{ij}(r_{ij})}{\partial r_{ij}} \frac{\partial r_{ij}}{\partial \underline{r}_k} \\ &= -\sum_{i=1, i\neq k}^{N} \frac{\partial F}{\partial \rho_i} \frac{\partial \rho_{ik}(r_{ik})}{\partial r_{ik}} \frac{\partial r_{ik}}{\partial \underline{r}_k} - \sum_{j=1, j\neq k}^{N} \frac{\partial F}{\partial \rho_k} \frac{\partial \rho_{kj}(r_{kj})}{\partial r_{kj}} \frac{\partial r_{kj}}{\partial \underline{r}_k} \\ &= \sum_{j=1, j\neq k}^{N} \left(\frac{\partial F}{\partial \rho_k} + \frac{\partial F}{\partial \rho_j}\right) \frac{\partial \rho_{kj}(r_{kj})}{\partial r_{kj}} \frac{\underline{\underline{r}}_{kj}}{r_{kj}} \quad . \end{aligned} \tag{2.121}$$

1. EAM force
The same as shown above. However, it is worth noting that the generation of the force arrays from tabulated data (implemented in the METAL_TABLE_DERIVATIVES routine) is done using a five point interpolation procedure.

2. EEAM force
   Information the same as that for EAM.

3. 2BEAM force
   Information the same as that for EAM. However, as there is a second embedding contribution from the extra band complexity: $U_2 = U_2^s + U_2^d$!

4. 2BEEAM force
   Information the same as that for EAM. However, as there is a second embedding contribution from the extra band complexity: $U_2 = U_2^s + U_2^d$!

5. Finnis-Sinclair force

$$\begin{aligned}
-\frac{\partial U_1}{\partial \underline{r_k}} &= \sum_{j=1,j\neq k}^{N} \left\{2(r_{kj}-c)(c_0+c_1 r_{kj}+c_2 r_{kj}^2)+(r_{kj}-c)^2(c_1+2c_2 r_{kj})\right\}\frac{\underline{\underline{r_{kj}}}}{r_{kj}} \\
-\frac{\partial U_2}{\partial \underline{r_k}} &= -\sum_{j=1,j\neq k}^{N} \frac{A}{2}\left(\frac{1}{\sqrt{\rho_k}}+\frac{1}{\sqrt{\rho_j}}\right)\left\{2(r_{kj}-d)+3\beta\frac{(r_{kj}-d)^2}{d}\right\}\frac{\underline{\underline{r_{kj}}}}{r_{kj}} \quad . \qquad (2.122)
\end{aligned}$$

6. Extended Finnis-Sinclair force

$$\begin{aligned}
-\frac{\partial U_1}{\partial \underline{r_k}} &= \sum_{j=1,j\neq k}^{N} \left\{2(r_{kj}-c)(c_0+c_1 r_{kj}+c_2 r_{kj}^2+c_3 r_{kj}^3+c_4 r_{kj}^4)+\right. \\
&\qquad \left.(r_{kj}-c)^2(c_1+2c_2 r_{kj}+3c_3 r_{kj}^2+4c_4 r_{kj}^3)\right\}\frac{\underline{\underline{r_{kj}}}}{r_{kj}} \qquad (2.123) \\
-\frac{\partial U_2}{\partial \underline{r_k}} &= -\sum_{j=1,j\neq k}^{N} \frac{A}{2}\left(\frac{1}{\sqrt{\rho_k}}+\frac{1}{\sqrt{\rho_j}}\right)\left\{2(r_{kj}-d)+4B^2(r_{kj}-d)^3\right\}\frac{\underline{\underline{r_{kj}}}}{r_{kj}} \quad .
\end{aligned}$$

7. Sutton-Chen force

$$\begin{aligned}
-\frac{\partial U_1}{\partial \underline{r_k}} &= -\sum_{j=1,j\neq k}^{N} n\epsilon\left(\frac{a}{r_{kj}}\right)^n \frac{\underline{\underline{r_{kj}}}}{r_{kj}} \\
-\frac{\partial U_2}{\partial \underline{r_k}} &= \sum_{j=1,j\neq k}^{N} \frac{mc\epsilon}{2}\left(\frac{1}{\sqrt{\rho_k}}+\frac{1}{\sqrt{\rho_j}}\right)\left(\frac{a}{r_{kj}}\right)^m \frac{\underline{\underline{r_{kj}}}}{r_{kj}} \quad . \qquad (2.124)
\end{aligned}$$

8. Gupta force

$$\begin{aligned}
-\frac{\partial U_1}{\partial \underline{r_k}} &= -\sum_{j=1,j\neq k}^{N} \frac{2Ap}{r_0}\exp\left(-p\frac{r_{kj}-r_0}{r_0}\right)\frac{\underline{\underline{r_{kj}}}}{r_{kj}} \\
-\frac{\partial U_2}{\partial \underline{r_k}} &= \sum_{j=1,j\neq k}^{N} \frac{Bq_{kj}}{r_0}\left(\frac{1}{\sqrt{\rho_k}}+\frac{1}{\sqrt{\rho_j}}\right)\exp\left(-2q_{kj}\frac{r_{kj}-r_0}{r_0}\right)\frac{\underline{\underline{r_{kj}}}}{r_{kj}} \quad . \qquad (2.125)
\end{aligned}$$

9. Many body perturbation component potential force

$$\begin{aligned}
-\frac{\partial U_1}{\partial \underline{r_k}} &= 0 \\
-\frac{\partial U_2}{\partial \underline{r_k}} &= \sum_{j=1,j\neq k}^{N} \frac{m\epsilon}{2}\left(\frac{1}{\sqrt{\rho_k}}+\frac{1}{\sqrt{\rho_j}}\right)\frac{a}{r_{kj}^m}\frac{\underline{\underline{r_{kj}}}}{r_{kj}} \quad . \qquad (2.126)
\end{aligned}$$

With the metal forces thus defined the contribution to be added to the atomic virial *from each atom pair* is then

$$\mathcal{W} = -\underline{r}_{ij} \cdot \underline{f}_{j} \quad , \tag{2.127}$$

which equates to:

$$\begin{aligned}
\Psi &= 3V\frac{\partial U}{\partial V} \\
\Psi &= \frac{3}{2}V\sum_{i=1}^{N}\sum_{j\neq i}^{N}\frac{\partial V_{ij}(r_{ij})}{\partial r_{ij}}\frac{\partial r_{ij}}{\partial V} + 3V\sum_{i=1}^{N}\frac{\partial F(\rho_i)}{\partial \rho_i}\frac{\partial \rho_i}{\partial V} = \Psi_1 + \Psi_2 \\
& \frac{\partial r_{ij}}{\partial V} = \frac{\partial V^{1/3}s_{ij}}{\partial V} = \frac{1}{3}V^{-2/3}s_{ij} = \frac{r_{ij}}{3V} \\
\Psi_1 &= \frac{1}{2}\sum_{i=1}^{N}\sum_{j\neq i}^{N}\frac{\partial V_{ij}(r_{ij})}{\partial r_{ij}}r_{ij} \\
& \frac{\partial \rho_i}{\partial V} = \frac{\partial}{\partial V}\sum_{j=1,j\neq i}^{N}\rho_{ij}(r_{ij}) = \sum_{j=1,j\neq i}^{N}\frac{\partial \rho_{ij}(r_{ij})}{\partial r_{ij}}\frac{\partial r_{ij}}{\partial V} = \frac{1}{3V}\sum_{j=1,j\neq i}^{N}\frac{\partial \rho_{ij}(r_{ij})}{\partial r_{ij}}r_{ij} \\
\Psi_2 &= \frac{1}{2}\sum_{i=1}^{N}\sum_{j\neq i}^{N}\left(\frac{\partial F(\rho_i)}{\partial \rho_i} + \frac{\partial F(\rho_j)}{\partial \rho_j}\right)\frac{\partial \rho_{ij}(r_{ij})}{\partial r_{ij}}r_{ij} \quad .
\end{aligned} \tag{2.128}$$

1. EAM virial
   The same as above.

2. EEAM virial
   The same as above.

3. 2BEAM virial
   The same as above but with a second embedding contribution from the extra band complexity: $\Psi_2 = \Psi_2^s + \Psi_2^d$!

4. 2BEEAM virial
   The same as above but with a second embedding contribution from the extra band complexity: $\Psi_2 = \Psi_2^s + \Psi_2^d$!

5. Finnis-Sinclair virial

$$\begin{aligned}
\Psi_1 &= \frac{1}{2}\sum_{i=1}^{N}\sum_{j\neq i}^{N}\left\{2(r_{ij}-c)(c_0+c_1r_{ij}+c_2r_{ij}^2)+(r_{ij}-c)^2(c_1+2c_2r_{ij})\right\}r_{ij} \\
\Psi_2 &= \frac{1}{2}\sum_{i=1}^{N}\sum_{j\neq i}^{N}\frac{A}{2}\left(\frac{1}{\sqrt{\rho_k}}+\frac{1}{\sqrt{\rho_j}}\right)\left\{2(r_{ij}-d)+3\beta\frac{(r_{ij}-d)^2}{d}\right\}r_{ij}a \quad .
\end{aligned} \tag{2.129}$$

6. Extended Finnis-Sinclair virial

$$\begin{aligned}
\Psi_1 &= \frac{1}{2}\sum_{i=1}^{N}\sum_{j\neq i}^{N}\left\{2(r_{ij}-c)(c_0+c_1r_{ij}+c_2r_{ij}^2+c_3r_{ij}^3+c_4r_{ij}^4)+\right. \\
&\qquad \left.(r_{ij}-c)^2(c_1+2c_2r_{ij}+3c_3r_{ij}^2+4c_4r_{ij}i^3)\right\}r_{ij} \\
\Psi_2 &= \frac{1}{2}\sum_{i=1}^{N}\sum_{j\neq i}^{N}\frac{A}{2}\left(\frac{1}{\sqrt{\rho_k}}+\frac{1}{\sqrt{\rho_j}}\right)\left\{2(r_{ij}-d)+4B^2(r_{ij}-d)^3\right\}r_{ij}a \quad .
\end{aligned} \tag{2.130}$$

7. Sutton-Chen virial

$$\begin{aligned}\Psi_1 &= -\frac{1}{2}\sum_{i=1}^{N}\sum_{j\neq i}^{N} n\epsilon \left(\frac{a}{r_{ij}}\right)^n \\ \Psi_2 &= \frac{1}{2}\sum_{i=1}^{N}\sum_{j\neq i}^{N} \frac{mc\epsilon}{2}\left(\frac{\partial F(\rho_i)}{\partial \rho_i} + \frac{\partial F(\rho_j)}{\partial \rho_j}\right)\left(\frac{a}{r_{ij}}\right)^m \quad .\end{aligned} \tag{2.131}$$

8. Gupta virial

$$\begin{aligned}\Psi_1 &= -\sum_{i=1}^{N}\sum_{j\neq i}^{N} \frac{Ap}{r_0}\exp\left(-p\frac{r_{ij}-r_0}{r_0}\right) r_{ij} \\ \Psi_2 &= \frac{1}{2}\sum_{i=1}^{N}\sum_{j\neq i}^{N} \frac{Bq_{ij}}{r_0}\left(\frac{1}{\sqrt{\rho_k}} + \frac{1}{\sqrt{\rho_j}}\right)\exp\left(-2q_{ij}\frac{r_{ij}-r_0}{r_0}\right) r_{ij} \quad .\end{aligned} \tag{2.132}$$

9. Many body perturbation component virial

$$\begin{aligned}\Psi_1 &= 0 \\ \Psi_2 &= \frac{1}{2}\sum_{i=1}^{N}\sum_{j\neq i}^{N} \frac{m\epsilon}{2}\left(\frac{\partial F(\rho_i)}{\partial \rho_i} + \frac{\partial F(\rho_j)}{\partial \rho_j}\right)\frac{a}{r_{ij}^m} \quad .\end{aligned} \tag{2.133}$$

The contribution to be added to the atomic stress tensor is given by

$$\sigma^{\alpha\beta} = r_{ij}^{\alpha} f_j^{\beta} \quad , \tag{2.134}$$

where α and β indicate the x, y, z components. The atomic stress tensor is symmetric.

The long-ranged correction for the DL_POLY_4 metal potential is in two parts. Firstly, by analogy with the short ranged potentials, the correction to the local density is

$$\begin{aligned}\rho_i &= \sum_{j=1,j\neq i}^{\infty} \rho_{ij}(r_{ij}) \\ \rho_i &= \sum_{j=1,j\neq i}^{r_{ij}<r_{\rm met}} \rho_{ij}(r_{ij}) + \sum_{j=1,j\neq i}^{r_{ij}\geq r_{\rm met}} \rho_{ij}(r_{ij}) = \rho_i^o + \delta\rho_i \\ \delta\rho_i &= 4\pi\bar{\rho}\int_{r_{\rm met}}^{\infty} \rho_{ij}(r)dr \quad ,\end{aligned} \tag{2.135}$$

where ρ_i^o is the uncorrected local density and $\bar{\rho}$ is the *mean particle density*. Evaluating the integral part of the above equation yields:

1. EAM density correction
   No long-ranged corrections apply beyond $r_{\rm met}$.
2. EEAM density correction
   No long-ranged corrections apply beyond $r_{\rm met}$.
3. 2BEAM density correction
   No long-ranged corrections apply beyond $r_{\rm met}$.
4. 2BEAM density correction
   No long-ranged corrections apply beyond $r_{\rm met}$.
5. Finnis-Sinclair density correction
   No long-ranged corrections apply beyond cutoffs c and d.

6. Extended Finnis-Sinclair density correction
   No long-ranged corrections apply beyond cutoffs c and d.

7. Sutton-Chen density correction
$$\delta\rho_i = \frac{4\pi\bar{\rho}a^3}{(m-3)}\left(\frac{a}{r_{\rm met}}\right)^{m-3} \quad . \tag{2.136}$$

8. Gupta density correction
$$\delta\rho_i = \frac{2\pi\bar{\rho}r_0}{q_{ij}}\left[r_{\rm met}^2 + 2r_{\rm met}\left(\frac{r_0}{q_{ij}}\right) + 2\left(\frac{r_0}{q_{ij}}\right)^2\right]\exp\left(-2q_{ij}\frac{r_{\rm met}-r_0}{r_0}\right) \quad . \tag{2.137}$$

9. Many body perturbation component density correction
$$\delta\rho_i = \frac{4\pi\bar{\rho}}{(m-3)}\frac{a}{r_{\rm met}^{m-3}} \quad . \tag{2.138}$$

The density correction is applied immediately after the local density is calculated. The pair term correction is obtained by analogy with the short ranged potentials and is

$$\begin{aligned}
U_1 &= \frac{1}{2}\sum_{i=1}^{N}\sum_{j\neq i}^{\infty}V_{ij}(r_{ij}) \\
U_1 &= \frac{1}{2}\sum_{i=1}^{N}\sum_{j\neq i}^{r_{ij}<r_{\rm met}}V_{ij}(r_{ij}) + \frac{1}{2}\sum_{i=1}^{N}\sum_{j\neq i}^{r_{ij}\geq r_{\rm met}}V_{ij}(r_{ij}) = U_1^o + \delta U_1 \\
\delta U_1 &= 2\pi N\bar{\rho}\int_{r_{\rm met}}^{\infty}V_{ij}(r)r^2dr \\
U_2 &= \sum_{i=1}^{N}F(\rho_i^0 + \delta\rho_i) \\
U_2 &= \sum_{i=1}^{N}F(\rho_i^0) + \sum_{i=1}^{N}\frac{\partial F(\rho_i)_0}{\partial\rho_i}\delta\rho_i = U_2^0 + \delta U_2 \\
\delta U_2 &= 4\pi\bar{\rho}\sum_{i=1}^{N}\frac{\partial F(\rho_i)_0}{\partial\rho_i}\int_{r_{\rm met}}^{\infty}\rho_{ij}(r)r^2dr \quad .
\end{aligned} \tag{2.139}$$

Note: that $\delta U2$ is not required if ρ_i has already been corrected. Evaluating the integral part of the above equations yields:

1. EAM energy correction
   No long-ranged corrections apply beyond $r_{\rm met}$.

2. EEAM energy correction
   No long-ranged corrections apply beyond $r_{\rm met}$.

3. 2BEAM energy correction
   No long-ranged corrections apply beyond $r_{\rm met}$.

4. 2BEEAM energy correction
   No long-ranged corrections apply beyond $r_{\rm met}$.

5. Finnis-Sinclair energy correction
   No long-ranged corrections apply beyond cutoffs c and d.

6. Extended Finnis-Sinclair energy correction
   No long-ranged corrections apply beyond cutoffs c and d.

7. Sutton-Chen energy correction

$$\begin{aligned}\delta U_1 &= \frac{2\pi N\bar{\rho}\epsilon a^3}{(n-3)}\left(\frac{a}{r_{\rm met}}\right)^{n-3} \\ \delta U_2 &= -\frac{4\pi\bar{\rho}a^3}{(m-3)}\left(\frac{a}{r_{\rm met}}\right)^{m-3}\left\langle\frac{Nc\epsilon}{2\sqrt{\rho_i^0}}\right\rangle \quad .\end{aligned} \tag{2.140}$$

8. Gupta energy correction

$$\begin{aligned}\delta U_1 &= \frac{4\pi N\bar{\rho}Ar_0}{p}\left[r_{\rm met}^2+2r_{\rm met}\left(\frac{r_0}{p}\right)+2\left(\frac{r_0}{p}\right)^2\right]\times \\ &\quad \exp\left(-p\frac{r_{\rm met}-r_0}{r_0}\right) \\ \delta U_2 &= -\frac{2\pi\bar{\rho}r_0}{q_{ij}}\left[r_{\rm met}^2+2r_{\rm met}\left(\frac{r_0}{q_{ij}}\right)+2\left(\frac{r_0}{q_{ij}}\right)^2\right]\times \\ &\quad \exp\left(-2q_{ij}\frac{r_{\rm met}-r_0}{r_0}\right)\left\langle\frac{NB}{2\sqrt{\rho_i^0}}\right\rangle \quad .\end{aligned} \tag{2.141}$$

9. Many body perturbation component energy correction

$$\begin{aligned}\delta U_1 &= 0 \\ \delta U_2 &= -\frac{4\pi\bar{\rho}}{(m-3)}\frac{a}{r_{\rm met}^{m-3}}\left\langle\frac{N\epsilon}{2\sqrt{\rho_i^0}}\right\rangle \quad .\end{aligned} \tag{2.142}$$

To estimate the virial correction we assume the corrected local densities are constants (i.e. independent of distance - at least beyond the range $r_{\rm met}$). This allows the virial correction to be computed by the methods used in the short ranged potentials:

$$\begin{aligned}\Psi_1 &= \frac{1}{2}\sum_{i=1}^{N}\sum_{j\neq i}^{\infty}\frac{\partial V_{ij}(r_{ij})}{\partial r_{ij}}r_{ij} \\ \Psi_1 &= \frac{1}{2}\sum_{i=1}^{N}\sum_{j\neq i}^{r_{ij}<r_{\rm met}}\frac{\partial V_{ij}(r_{ij})}{\partial r_{ij}}r_{ij}+\frac{1}{2}\sum_{i=1}^{N}\sum_{j\neq i}^{r_{ij}\geq r_{\rm met}}\frac{\partial V_{ij}(r_{ij})}{\partial r_{ij}}r_{ij}=\Psi_1^0+\delta\Psi_1 \\ \delta\Psi_1 &= 2\pi N\bar{\rho}\int_{r_{\rm met}}^{\infty}\frac{\partial V_{ij}(r)}{\partial r_{ij}}r^3dr \\ \Psi_2 &= \sum_{i=1}^{N}\frac{\partial F(\rho_i)}{\partial\rho_i}\sum_{j\neq i}^{\infty}\frac{\partial\rho_{ij}(r_{ij})}{\partial r_{ij}}r_{ij} \\ \Psi_2 &= \sum_{i=1}^{N}\frac{\partial F(\rho_i)}{\partial\rho_i}\sum_{j\neq i}^{r_{ij}<r_{\rm met}}\frac{\partial\rho_{ij}(r_{ij})}{\partial r_{ij}}r_{ij}+\sum_{i=1}^{N}\frac{\partial F(\rho_i)}{\partial\rho_i}\sum_{j\neq i}^{r_{ij}\geq r_{\rm met}}\frac{\partial\rho_{ij}(r_{ij})}{\partial r_{ij}}r_{ij}=\Psi_2^0+\delta\Psi_2 \\ \delta\Psi_2 &= 4\pi\bar{\rho}\sum_{i=1}^{N}\frac{\partial F(\rho_i)}{\partial\rho_i}\int_{r_{\rm met}}^{\infty}\frac{\partial\rho_{ij}(r)}{\partial r}r^3dr \quad .\end{aligned} \tag{2.143}$$

Evaluating the integral part of the above equations yields:

1. EAM virial correction
   No long-ranged corrections apply beyond $r_{\rm met}$.

2. EEAM virial correction
   No long-ranged corrections apply beyond $r_{\rm met}$.

3. 2BEAM virial correction
   No long-ranged corrections apply beyond $r_{\rm met}$.

4. 2BEEAM virial correction
   No long-ranged corrections apply beyond $r_{\rm met}$.

5. Finnis-Sinclair virial correction
   No long-ranged corrections apply beyond cutoffs c and d.

6. Extended Finnis-Sinclair virial correction
   No long-ranged corrections apply beyond cutoffs c and d.

7. Sutton-Chen virial correction

$$\begin{aligned} \delta\Psi_1 &= -n\frac{2\pi N\bar{\rho}\epsilon a^3}{(n-3)}\left(\frac{a}{r_{\rm met}}\right)^{n-3} \\ \delta\Psi_2 &= m\frac{4\pi\bar{\rho}a^3}{(m-3)}\left(\frac{a}{r_{\rm met}}\right)^{m-3}\left\langle\frac{Nc\epsilon}{2\sqrt{\rho_i^0}}\right\rangle \quad . \end{aligned} \tag{2.144}$$

8. Gupta virial correction

$$\begin{aligned} \delta\Psi_1 &= -\frac{p}{r_0}\frac{4\pi N\bar{\rho}Ar_0}{p}\left[r_{\rm met}^3+3r_{\rm met}^2\left(\frac{r_0}{p}\right)+6r_{\rm met}\left(\frac{r_0}{p}\right)^2+6\left(\frac{r_0}{p}\right)^3\right]\times \\ & \quad \exp\left(-p\frac{r_{\rm met}-r_0}{r_0}\right) \\ \delta\Psi_2 &= \frac{q_{ij}}{r_0}\frac{2\pi\bar{\rho}r_0}{q_{ij}}\left[r_{\rm met}^3+3r_{\rm met}^2\left(\frac{r_0}{q_{ij}}\right)+6r_{\rm met}\left(\frac{r_0}{q_{ij}}\right)^2+6\left(\frac{r_0}{q_{ij}}\right)^3\right]\times \\ & \quad \exp\left(-2q_{ij}\frac{r_{\rm met}-r_0}{r_0}\right)\left\langle\frac{NB}{2\sqrt{\rho_i^0}}\right\rangle \quad . \end{aligned} \tag{2.145}$$

9. Many body perturbation component virial correction

$$\begin{aligned} \delta\Psi_1 &= 0 \\ \delta\Psi_2 &= m\frac{4\pi\bar{\rho}}{(m-3)}\frac{a}{r_{\rm met}^{m-3}}\left\langle\frac{N\epsilon}{2\sqrt{\rho_i^0}}\right\rangle \quad . \end{aligned} \tag{2.146}$$

In the energy and virial corrections we have used the approximation:

$$\sum_i^N \rho_i^{-1/2} = \frac{N}{<\rho_i^{1/2}>} \quad , \tag{2.147}$$

where $<\rho_i^{1/2}>$ is regarded as a constant of the system.

In DL_POLY_4 the metal forces are handled by the routine METAL_FORCES. The local density is calculated by the routines METAL_LD_COLLECT_EAM, METAL_LD_COLLECT_FST, METAL_LD_COMPUTE, METAL_LD_SET_HALO and METAL_LD_EXPORT. The long-ranged corrections are calculated by METAL_LRC. Reading and generation of EAM table data from TABEAM is handled by METAL_TABLE_READ and METAL_TABLE_DERIVATIVES.

Notes on the Treatment of Alloys

The distinction to be made between EAM and FS potentials with regard to alloys concerns the mixing rules for unlike interactions. Starting with equations (2.110) and (2.111), it is clear that we require mixing rules for terms $V_{ij}(r_{ij})$ and $\rho_{ij}(r_{ij})$ when atoms i and j are of different kinds. Thus two different metals A and B we can distinguish 4 possible variants of each:

$$V_{ij}^{AA}(r_{ij}),\ V_{ij}^{BB}(r_{ij}),\ V_{ij}^{AB}(r_{ij}),\ V_{ij}^{BA}(r_{ij})$$

and

$$\rho_{ij}^{AA}(r_{ij}),\ \rho_{ij}^{BB}(r_{ij}),\ \rho_{ij}^{AB}(r_{ij}),\ \rho_{ij}^{BA}(r_{ij})\quad .$$

These forms recognise that the contribution of a type A atom to the potential of a type B atom may be different from the contribution of a type B atom to the potential of a type A atom. In both EAM [60] and FS [15] cases it turns out that

$$V_{ij}^{BA}(r_{ij}) = V_{ij}^{BA}(r_{ij})\quad , \tag{2.148}$$

though the mixing rules are different in each case (**beware!**). This has the following implications to densities of mixtures for different potential frameworks:

- EAM case - it is required that [60]:

$$\begin{aligned}\rho_{ij}^{AB}(r_{ij}) &= \rho_{ij}^{BB}(r_{ij})\\ \rho_{ij}^{BA}(r_{ij}) &= \rho_{ij}^{AA}(r_{ij})\quad ,\end{aligned} \tag{2.149}$$

  which means that an atom of type A contributes the same density to the environment of an atom of type B as it does to an atom of type A, and *vice versa*.

- EEAM case - all densities can be different [52, 53]:

$$\rho_{ij}^{AA}(r_{ij}) \neq \rho_{ij}^{BB}(r_{ij}) \neq \rho_{ij}^{AB}(r_{ij}) \neq \rho_{ij}^{BA}(r_{ij})\quad ! \tag{2.150}$$

- 2BEAM case - similarly to the EAM case it is required that:

$$\begin{aligned}\rho_{ij}^{d\ AB}(r_{ij}) &= \rho_{ij}^{d\ BB}(r_{ij})\\ \rho_{ij}^{d\ BA}(r_{ij}) &= \rho_{ij}^{d\ AA}(r_{ij})\quad ,\end{aligned} \tag{2.151}$$

  for the d-band densities, whereas for the s-band ones:

$$\rho_{ij}^{s\ BA}(r_{ij}) = \rho_{ij}^{s\ AB}(r_{ij})\quad , \tag{2.152}$$

  which means that an atom of type A contributes the same s density to the environment of an atom of type B as an atom of type B to an environment of an atom of type A. However, in general:

$$\rho_{ij}^{s\ AA}(r_{ij}) \neq \rho_{ij}^{s\ BB}(r_{ij}) \neq \rho_{ij}^{s\ AB}(r_{ij})\quad . \tag{2.153}$$

- 2BEEAM case - similarly to the EEAM case all s and d densities can be different:

$$\begin{aligned}\rho_{ij}^{s\ AA}(r_{ij}) \neq \rho_{ij}^{s\ BB}(r_{ij}) \neq \rho_{ij}^{s\ AB}(r_{ij}) \neq \rho_{ij}^{s\ BA}(r_{ij})\\ \rho_{ij}^{d\ AA}(r_{ij}) \neq \rho_{ij}^{d\ BB}(r_{ij}) \neq \rho_{ij}^{d\ AB}(r_{ij}) \neq \rho_{ij}^{d\ BA}(r_{ij})\quad .\end{aligned} \tag{2.154}$$

- FS case - here a different rule applies [15]:

$$\rho_{ij}^{AB}(r_{ij}) = (\rho_{ij}^{AA}(r_{ij})\ \rho_{ij}^{BB}(r_{ij}))^{1/2} \tag{2.155}$$

  so that atoms of type A and B contribute the same densities to each other, but not to atoms of the same type.

The above rules have the following consequences to the specifications of these potentials in the DL_POLY_4 FIELD file for an alloy composed of n different metal atom types both the EAM types and FS types of potentials require the specification of $n(n+1)/2$ pair functions $V_{ij}^{AB}(r_{ij})$. However, the its only the simple EAM type together with all the FS types that require only n density functions $\rho_{ij}^{AA}(r_{ij})$, whereas the EEAM class requires all the cross functions $\rho_{ij}^{AB}(r_{ij})$ possible or n^2 in total! In addition to the $n(n+1)/2$ pair functions and n or n^2 density functions both the EAM and EEAM potentials require further specification of n functional forms of the density dependence (i.e. the embedding function $F(\rho_i)$ in equation (2.110)). The matter is further complicated when the 2BEAM type of potential is used with the extra specification of n embedding functions and $n(n+1)/2$ density functions for the s-band. Similarly, in the 2BEEAM an extra n embedding functions and n^2 density functions for the s-band are required.

It is worth noting that in the 2BEAM and 2BEEAM the s-band contribution is usually only for the alloy component, so that local concentrations of a single element revert to the standard EAM or EEAM! In such case, the densities functions must be zeroed in the DL_POLY_4 TABEAM file.

For EAM, EEAM, 2BEAM and 2BEEAM potentials all the functions are supplied in tabular form via the table file TABEAM (see section 6.1.7) to which DL_POLY_4 is redirected by the FIELD file data. The FS potentials are defined via the necessary parameters in the FIELD file.

2.3.3 Tersoff Potentials

The Tersoff [17] potential is is a bond-order potential, developed to be used in multi-component covalent systems by an effective coupling of two-body and higher many-body correlations into one model. The central idea is that in real systems, the strength of each bond depends on the local environment, i.e. an atom with many neighbors forms weaker bonds than an atom with few neighbors. Effectively, it is a pair potential the strength of which depends on the environment. At the present there are two versions of this potential available in DL_POLY_4: **ters** and **kihs**. In these particular implementations **ters** has 11 atomic and 2 bi-atomic parameters whereas **kihs** [61] has 16 atomic parameters. The energy is modelled as a sum of pair-like interactions, where the coefficient of the attractive term in the pair-like potential (which plays the role of a bond order) depends on the local environment giving a many-body potential.

The form of the Tersoff potential is: (**ters**)

$$U_{ij} = f_C(r_{ij})\ [f_R(r_{ij}) - \gamma_{ij}\ f_A(r_{ij})] \ \ , \tag{2.156}$$

where f_R and f_A are the repulsive and attractive pair potential respectively:

$$f_R(r_{ij}) = A_{ij}\ \exp(-a_{ij}\ r_{ij}) \ \ , \ \ f_A(r_{ij}) = B_{ij}\ \exp(-b_{ij}\ r_{ij}) \tag{2.157}$$

and f_C is a smooth cutoff function with parameters R and S so chosen that to include the first-neighbor shell:

- **ters**:

$$f_C(r_{ij}) = \begin{cases} 1 & : \ r_{ij} < R_{ij} \\ \frac{1}{2} + \frac{1}{2}\cos\left[\pi\ \frac{r_{ij}-R_{ij}}{S_{ij}-R_{ij}}\right] & : \ R_{ij} < r_{ij} < S_{ij} \\ 0 & : \ r_{ij} > S_{ij} \end{cases} \tag{2.158}$$

- **kihs** - here f_C is modified to a have continuous second-order differential:

$$f_C(r_{ij}) = \begin{cases} 1 & : \ r_{ij} < R_{ij} \\ \frac{1}{2} + \frac{9}{16}\cos\left[\pi\ \frac{r_{ij}-R_{ij}}{S_{ij}-R_{ij}}\right] - \frac{1}{16}\cos\left[3\pi\ \frac{r_{ij}-R_{ij}}{S_{ij}-R_{ij}}\right] & : \ R_{ij} < r_{ij} < S_{ij} \\ 0 & : \ r_{ij} > S_{ij} \ \ . \end{cases} \tag{2.159}$$

γ_{ij} expresses a dependence that can accentuate or diminish the attractive force relative to the repulsive force, according to the local environment, such that:

- **ters**:

$$\begin{aligned}
\gamma_{ij} &= \chi_{ij}\ (1 + \beta_i^{\eta_i}\ \mathcal{L}_{ij}^{\eta_i})^{-\frac{1}{2\eta_i}} \\
\mathcal{L}_{ij} &= \sum_{k\neq i,j} f_C(r_{ik})\ \omega_{ik}\ g(\theta_{ijk}) \\
g(\theta_{ijk}) &= 1 + \frac{c_i^2}{d_i^2} - \frac{c_i^2}{d_i^2 + (h_i - \cos\theta_{ijk})^2}
\end{aligned} \tag{2.160}$$

- **kihs**:

$$\begin{aligned}
\gamma_{ij} &= (1 + \mathcal{L}_{ij}^{\eta_i})^{-\delta_i} \\
\mathcal{L}_{ij} &= \sum_{k\neq i,j} f_C(r_{ik})\ g(\theta_{ijk})\ \overbrace{\exp\left[\alpha_i (r_{ij} - r_{ik})^{\beta_i}\right]}^{\omega_{ik}} \\
g(\theta_{ijk}) &= c_{1i} + g_o(\theta_{ijk})\ g_a(\theta_{ijk}) \\
g_o(\theta_{ijk}) &= \frac{c_{2i}\ (h_i - \cos\theta_{ijk})^2}{c_{3i} + (h_i - \cos\theta_{ijk})^2} \\
g_a(\theta_{ijk}) &= 1 + c_{4i}\ \exp\left[-c_{5i}\ (h_i - \cos\theta_{ijk})^2\right] \quad ,
\end{aligned} \tag{2.161}$$

where the term $\mathcal{L}_{ij}$ defines the effective coordination number of atom i i.e. the number of nearest neighbors, taking into account the relative distance of the two neighbors, i and k, $r_{ij} - r_{ik}$, and the bond angle, θ_{ijk}, between them with respect to the central atom i. The function $g(\theta)$ has a minimum for $h_i = \cos(\theta_{ijk})$, the parameter d_i in **ters** and c_{3i} in **kihs** determines how sharp the dependence on angle is, whereas the rest express the strength of the angular effect. Further mixed parameters are defined as:

$$\begin{aligned}
a_{ij} = (a_i + a_j)/2 \quad &, \quad b_{ij} = (b_i + b_j)/2 \\
A_{ij} = (A_i A_j)^{1/2} \quad &, \quad B_{ij} = (B_i B_j)^{1/2} \\
R_{ij} = (R_i R_j)^{1/2} \quad &, \quad S_{ij} = (S_i S_j)^{1/2} \quad .
\end{aligned} \tag{2.162}$$

Singly subscripted parameters, such as a_i and η_i, depend only on the type of atom.

For **ters** the chemistry between different atom types is locked in the two sets bi-atomic parameters χ_{ij} and ω_{ij}:

$$\begin{aligned}
\chi_{ii} &= 1 \quad , \quad \chi_{ij} = \chi_{ji} \\
\omega_{ii} &= 1 \quad , \quad \omega_{ij} = \omega_{ji} \quad ,
\end{aligned} \tag{2.163}$$

which define only one independent parameter each per pair of atom types. The χ parameter is used to strengthen or weaken the heteropolar bonds, relative to the value obtained by simple interpolation. The ω parameter is used to permit greater flexibility when dealing with more drastically different types of atoms.

The force on an atom ℓ derived from this potential is formally calculated with the formula:

$$f_\ell^\alpha = -\frac{\partial}{\partial r_\ell^\alpha} E_{\texttt{tersoff}} = \frac{1}{2}\sum_i \sum_{j\neq i} -\frac{\partial}{\partial r_\ell^\alpha} U_{ij} \quad , \tag{2.164}$$

with atomic label ℓ being one of i, j, k and α indicating the x, y, z component. The derivative after the summation is worked out as

$$-\frac{\partial U_{ij}}{\partial r_\ell^\alpha} = -\frac{\partial}{\partial r_\ell^\alpha} f_C(r_{ij}) f_R(r_{ij}) + \gamma_{ij} \frac{\partial}{\partial r_\ell^\alpha} f_C(r_{ij}) f_A(r_{ij}) + f_C(r_{ij}) f_A(r_{ij}) \frac{\partial}{\partial r_\ell^\alpha} \gamma_{ij} \quad , \tag{2.165}$$

with the contributions from the first two terms being:

$$-\frac{\partial}{\partial r_\ell^\alpha} f_C(r_{ij}) f_R(r_{ij}) = -\left\{ f_C(r_{ij}) \frac{\partial}{\partial r_{ij}} f_R(r_{ij}) + f_R(r_{ij}) \frac{\partial}{\partial r_{ij}} f_C(r_{ij}) \right\} \times \left\{ \delta_{j\ell} \frac{r_{i\ell}^\alpha}{r_{i\ell}} - \delta_{i\ell} \frac{r_{\ell j}^\alpha}{r_{\ell j}} \right\} \tag{2.166}$$

$$\gamma_{ij} \frac{\partial}{\partial r_\ell^\alpha} f_C(r_{ij}) f_A(r_{ij}) = \gamma_{ij} \left\{ f_C(r_{ij}) \frac{\partial}{\partial r_{ij}} f_A(r_{ij}) + f_A(r_{ij}) \frac{\partial}{\partial r_{ij}} f_C(r_{ij}) \right\} \times \left\{ \delta_{j\ell} \frac{r_{i\ell}^\alpha}{r_{i\ell}} - \delta_{i\ell} \frac{r_{\ell j}^\alpha}{r_{\ell j}} \right\} \quad , \tag{2.167}$$

and from the third (angular) term:

- **ters**:

$$f_C(r_{ij}) f_A(r_{ij}) \frac{\partial}{\partial r_\ell^\alpha} \gamma_{ij} = f_C(r_{ij}) f_A(r_{ij}) \; \chi_{ij} \quad \times \left(-\frac{1}{2}\right) \left(1 + {\beta_i}^{\eta_i} \; \mathcal{L}_{ij}^{\eta_i}\right)^{-\frac{1}{2\eta_i}-1} {\beta_i}^{\eta_i} \; \mathcal{L}_{ij}^{\eta_i-1} \frac{\partial}{\partial r_\ell^\alpha} \mathcal{L}_{ij} \quad , \tag{2.168}$$

  where

$$\frac{\partial}{\partial r_\ell^\alpha} \mathcal{L}_{ij} = \frac{\partial}{\partial r_\ell^\alpha} \sum_{k \neq i,j} \omega_{ik} \; f_C(r_{ik}) \; g(\theta_{ijk}) \quad . \tag{2.169}$$

  The angular term can have three different contributions depending on the index of the particle participating in the interaction:

$$\begin{aligned} \ell = i \quad &: \quad \frac{\partial}{\partial r_i^\alpha} \mathcal{L}_{ij} = \sum_{k \neq i,j} \omega_{ik} \left[g(\theta_{ijk}) \frac{\partial}{\partial r_i^\alpha} f_C(r_{ik}) + f_C(r_{ik}) \frac{\partial}{\partial r_i^\alpha} g(\theta_{ijk}) \right] \\ \ell = j \quad &: \quad \frac{\partial}{\partial r_j^\alpha} \mathcal{L}_{ij} = \sum_{k \neq i,j} \omega_{ik} \; f_C(r_{ik}) \frac{\partial}{\partial r_j^\alpha} g(\theta_{ijk}) \\ \ell \neq i,j \quad &: \quad \frac{\partial}{\partial r_\ell^\alpha} \mathcal{L}_{ij} = \omega_{i\ell} \left[g(\theta_{ij\ell}) \frac{\partial}{\partial r_\ell^\alpha} f_C(r_{i\ell}) + f_C(r_{i\ell}) \frac{\partial}{\partial r_\ell^\alpha} g(\theta_{ij\ell}) \right] \quad , \end{aligned} \tag{2.170}$$

- **kihs**:

$$f_C(r_{ij}) f_A(r_{ij}) \frac{\partial}{\partial r_\ell^\alpha} \gamma_{ij} = f_C(r_{ij}) f_A(r_{ij}) \; \times (-\delta_i \; \eta_i) \left(1 + \mathcal{L}_{ij}^{\eta_i}\right)^{-\delta_i - 1} \mathcal{L}_{ij}^{\eta_i-1} \frac{\partial}{\partial r_\ell^\alpha} \mathcal{L}_{ij} \quad , \tag{2.171}$$

  where

$$\frac{\partial}{\partial r_\ell^\alpha} \mathcal{L}_{ij} = \frac{\partial}{\partial r_\ell^\alpha} \sum_{k \neq i,j} \omega_{ik}(r_{ij}, r_{ik}) \; f_C(r_{ik}) \; g(\theta_{ijk}) \quad . \tag{2.172}$$

  It is worth noting that the derivative of ω_{ik}:

$$\frac{\partial}{\partial r_\ell^\alpha} \omega_{ik} = \alpha_i \; \beta_i \; (r_{ij} - r_{ik})^{\beta_i - 1} \; \omega_{ik} \; \left\{ (\delta_{\ell j} - \delta_{\ell i}) \frac{r_{ij}^\alpha}{r_{ij}} - (\delta_{\ell k} - \delta_{\ell i}) \frac{r_{ik}^\alpha}{r_{ik}} \right\} \quad , \tag{2.173}$$

  now has three different contributions depending on the index of the particle participating in the interaction! Hence the angular term's derivative is more elaborate to express than the one in the **ters** case.

The derivative of $g(\theta_{ijk})$ is worked out in the following manner:

$$\frac{\partial}{\partial r_\ell^\alpha} g(\theta_{ijk}) = \frac{\partial g(\theta_{ijk})}{\partial \theta_{ijk}} \; \frac{-1}{\sin\theta_{ijk}} \; \frac{\partial}{\partial r_\ell^\alpha} \left\{ \frac{\underline{r}_{ij} \cdot \underline{r}_{ik}}{r_{ij}\; r_{ik}} \right\} \quad , \tag{2.174}$$

where

$$\frac{\partial g(\theta_{ijk})}{\partial \theta_{ijk}} = \frac{2\; c_i^2 (h_i - \cos\theta_{ijk})\; \sin\theta_{ijk}}{[d_i^2 + (h_i - \cos\theta_{ijk})^2]^2} \tag{2.175}$$

$$\begin{aligned} \frac{\partial}{\partial r_\ell^\alpha} \left\{ \frac{\underline{r}_{ij} \cdot \underline{r}_{ik}}{r_{ij} r_{ik}} \right\} &= (\delta_{\ell j} - \delta_{\ell i}) \frac{r_{ik}^\alpha}{r_{ij} r_{ik}} + (\delta_{\ell k} - \delta_{\ell i}) \frac{r_{ij}^\alpha}{r_{ij} r_{ik}} - \\ & \cos(\theta_{jik}) \left\{ (\delta_{\ell j} - \delta_{\ell i}) \frac{r_{ij}^\alpha}{r_{ij}^2} + (\delta_{\ell k} - \delta_{\ell i}) \frac{r_{ik}^\alpha}{r_{ik}^2} \right\} \quad . \end{aligned} \tag{2.176}$$

The contribution to be added to the atomic virial can be derived as

$$\begin{aligned} \mathcal{W} &= 3V \frac{\partial E_{\texttt{tersoff}}}{\partial V} = \frac{3\;V}{2} \sum_i \sum_{j \neq i} \frac{\partial U_{ij}}{\partial V} = \sum_i \underline{r_i} \cdot \underline{f_i} \\ &= \frac{1}{2} \sum_i \sum_{j \neq i} - \left(\underline{r_{ij}} \cdot \underline{f_{ij}} + \underline{r_{ik}} \cdot \underline{f_{ik}} \right) = \frac{1}{2} \sum_i \sum_{j \neq i} \left(\frac{\partial U_{ij}}{\partial r_{ij}} \cdot \underline{r_{ij}} + \frac{\partial U_{ik}}{\partial r_{ik}} \cdot \underline{r_{ik}} \right) \end{aligned} \tag{2.177}$$

- **ters**:

$$\begin{aligned} \mathcal{W} = \;& \frac{1}{2} \sum_i \sum_{j \neq i} \left\{ \left[\frac{\partial}{\partial r_{ij}} f_C(r_{ij}) f_R(r_{ij}) - \gamma_{ij} \frac{\partial}{\partial r_{ij}} f_C(r_{ij}) f_A(r_{ij}) \right] r_{ij} - \right. \\ & \left(-\frac{1}{2} \right) f_C(r_{ij}) f_A(r_{ij})\; \chi_{ij} \left(1 + {\beta_i}^{\eta_i}\; \mathcal{L}_{ij}^{\eta_i} \right)^{-\frac{1}{2\eta_i} - 1} \; {\beta_i}^{\eta_i}\; \mathcal{L}_{ij}^{\eta_i - 1} \times \\ & \left. \sum_{k \neq i,j} \omega_{ik}\; g(\theta_{ijk}) \left[\frac{\partial}{\partial r_{ik}} f_C(r_{ik}) \right] r_{ik} \right\} \quad , \end{aligned} \tag{2.178}$$

- **hiks**:

$$\begin{aligned} \mathcal{W} = \;& \frac{1}{2} \sum_i \sum_{j \neq i} \left\{ \left[\frac{\partial}{\partial r_{ij}} f_C(r_{ij}) f_R(r_{ij}) - \gamma_{ij} \frac{\partial}{\partial r_{ij}} f_C(r_{ij}) f_A(r_{ij}) \right] r_{ij} - \right. \\ & (-\delta_i\; \eta_i)\; f_C(r_{ij}) f_A(r_{ij})\; \chi_{ij} \left(1 + \mathcal{L}_{ij}^{\eta_i} \right)^{-\delta_i - 1} \mathcal{L}_{ij}^{\eta_i - 1} \times \\ & \left. \sum_{k \neq i,j} \omega_{ik} \left[r_{ik}\; g(\theta_{ijk}) \frac{\partial}{\partial r_{ik}} f_C(r_{ik}) + \alpha_i\; \beta_i\; (r_{ij} - r_{ik})^{\beta_i} f_C(r_{ik}) \right] \right\} \quad . \end{aligned} \tag{2.179}$$

The contribution to be added to the atomic stress tensor is given by

$$\sigma^{\alpha\beta} = -r_i^\alpha f_i^\beta \quad , \tag{2.180}$$

where α and β indicate the x, y, z components. The stress tensor is symmetric.

Interpolation arrays, `vter` and `gter` (set up in TERSOFF_GENERATE) - similar to those in van der Waals interactions (Section 2.3.1), are used in the calculation of the Tersoff forces, virial and stress.

The Tersoff potentials are very short ranged, typically of order 3 Å. This property, plus the fact that Tersoff potentials (two- and three-body contributions) scale as N^3, where N is the number of particles, makes it essential that these terms are calculated by the link-cell method [62].

DL_POLY_4 applies no long-ranged corrections to the Tersoff potentials. In DL_POLY_4 Tersoff forces are handled by the routine TERSOFF_FORCES.

2.3.4 Three-Body Potentials

The three-body potentials in DL_POLY_4 are mostly valence angle forms. (They are primarily included to permit simulation of amorphous materials e.g. silicate glasses.) However, these have been extended to include the Dreiding [19] hydrogen bond. The potential forms available are as follows:

1. Harmonic: (**harm**)
$$U(\theta_{jik}) = \frac{k}{2}(\theta_{jik} - \theta_0)^2 \tag{2.181}$$

2. Truncated harmonic: (**thrm**)
$$U(\theta_{jik}) = \frac{k}{2}(\theta_{jik} - \theta_0)^2 \exp[-(r_{ij}^8 + r_{ik}^8)/\rho^8] \tag{2.182}$$

3. Screened Harmonic: (**shrm**)
$$U(\theta_{jik}) = \frac{k}{2}(\theta_{jik} - \theta_0)^2 \exp[-(r_{ij}/\rho_1 + r_{ik}/\rho_2)] \tag{2.183}$$

4. Screened Vessal [37]: (**bvs1**)
$$\begin{aligned} U(\theta_{jik}) &= \frac{k}{8(\theta_{jik} - \pi)^2}\left\{\left[(\theta_0 - \pi)^2 - (\theta_{jik} - \pi)^2\right]^2\right\} \times \\ &\quad \exp[-(r_{ij}/\rho_1 + r_{ik}/\rho_2)] \end{aligned} \tag{2.184}$$

5. Truncated Vessal [38]: (**bvs2**)
$$\begin{aligned} U(\theta_{jik}) &= k\ [\theta_{jik}^a(\theta_{jik} - \theta_0)^2(\theta_{jik} + \theta_0 - 2\pi)^2 - \\ &\quad \frac{a}{2}\pi^{a-1}(\theta_{jik} - \theta_0)^2(\pi - \theta_0)^3]\ \exp[-(r_{ij}^8 + r_{ik}^8)/\rho^8] \end{aligned} \tag{2.185}$$

6. Dreiding hydrogen bond [19]: (**hbnd**)
$$U(\theta_{jik}) = D_{hb}\ \cos^4(\theta_{jik})\ [5(R_{hb}/r_{jk})^{12} - 6(R_{hb}/r_{jk})^{10}] \tag{2.186}$$

Note that for the hydrogen bond, the hydrogen atom *must* be the central atom. Several of these functions are identical to those appearing in the *intra*-molecular valence angle descriptions above. There are significant differences in implementation however, arising from the fact that the three-body potentials are regarded as *inter*-molecular. Firstly, the atoms involved are defined by atom types, not specific indices. Secondly, there are *no* excluded atoms arising from the three-body terms. (The inclusion of other potentials, for example pair potentials, may in fact be essential to maintain the structure of the system.)

The three-body potentials are very short ranged, typically of order 3 Å. This property, plus the fact that three-body potentials scale as N^4, where N is the number of particles, makes it essential that these terms are calculated by the link-cell method [62].

The calculation of the forces, virial and stress tensor as described in the section valence angle potentials above.

DL_POLY_4 applies no long-ranged corrections to the three-body potentials. The three-body forces are calculated by the routine THREE_BODY_FORCES.

2.3.5 Four-Body Potentials

The four-body potentials in DL_POLY_4 are entirely inversion angle forms, primarily included to permit simulation of amorphous materials (particularly borate glasses). The potential forms available in DL_POLY_4 are as follows:

1. Harmonic: (**harm**)
$$U(\phi_{ijkn}) = \frac{k}{2}\;(\phi_{ijkn} - \phi_0)^2 \tag{2.187}$$
2. Harmonic cosine: (**hcos**)
$$U(\phi_{ijkn}) = \frac{k}{2}\;(\cos(\phi_{ijkn}) - \cos(\phi_0))^2 \tag{2.188}$$
3. Planar potential: (**plan**)
$$U(\phi_{ijkn}) = A\;[1 - \cos(\phi_{ijkn})] \tag{2.189}$$

These functions are identical to those appearing in the *intra*-molecular inversion angle descriptions above. There are significant differences in implementation however, arising from the fact that the four-body potentials are regarded as *inter*-molecular. Firstly, the atoms involved are defined by atom types, not specific indices. Secondly, there are *no* excluded atoms arising from the four-body terms. (The inclusion of other potentials, for example pair potentials, may in fact be essential to maintain the structure of the system.)

The four-body potentials are very short ranged, typically of order 3 Å. This property, plus the fact that four-body potentials scale as N^4, where N is the number of particles, makes it essential that these terms are calculated by the link-cell method [62].

The calculation of the forces, virial and stress tensor described in the section on inversion angle potentials above.

DL_POLY_4 applies no long-ranged corrections to the four body potentials. The four-body forces are calculated by the routine FOUR_BODY_FORCES.

2.4 Long Ranged Electrostatic (coulombic) Potentials

DL_POLY_4 incorporates several techniques for dealing with long-ranged electrostatic potentials[2]. These are as follows:

1. Direct Coulomb sum
2. Force-shifted Coulomb sum
3. Coulomb sum with distance dependent dielectric
4. Reaction field
5. Smoothed Particle Mesh Ewald (SPME)

All of these can be used in conjunction with the shell model technique used to account for ions polarisation.

The SPME technique is restricted to periodic systems only. (Users must exercise care when using pseudo-periodic boundary conditions.) The other techniques can be used with either periodic or non-periodic systems safely, although in the case of the direct Coulomb sum there are likely to be problems with convergence.

[2]Unlike the other elements of the force field, the electrostatic forces are NOT specified in the input FIELD file, but by setting appropriate directives in the CONTROL file. See Section 6.1.1.

DL_POLY_4 will correctly handle the electrostatics of both molecular and atomic species. However, it is assumed that the system is electrically neutral. A warning message is printed if the system is found to be charged, but otherwise the simulation proceeds as normal.

Note that DL_POLY_4 does not use the basic Ewald method, which is an option in DL_POLY_Classic, on account of it being too slow for large scale systems. The SPME method is the standard Ewald method in DL_POLY_4.

2.4.1 Direct Coulomb Sum

Use of the direct Coulomb sum is sometimes necessary for accurate simulation of isolated (non-periodic) systems. It is *not* recommended for periodic systems.

The interaction potential for two charged ions is

$$U(r_{ij}) = \frac{1}{4\pi\epsilon_0\epsilon}\frac{q_i q_j}{r_{ij}} \quad , \tag{2.190}$$

with q_ℓ the charge on an atom labelled ℓ, and r_{ij} the magnitude of the separation vector $\underline{r}_{ij} = \underline{r}_j - \underline{r}_i$.

The force on an atom j derived from this force is

$$\underline{f}_j = \frac{1}{4\pi\epsilon_0\epsilon}\frac{q_i q_j}{r_{ij}^3}\underline{r}_{ij} \quad , \tag{2.191}$$

with the force on atom i the negative of this.

The contribution to the atomic virial is

$$\mathcal{W} = -\frac{1}{4\pi\epsilon_0\epsilon}\frac{q_i q_j}{r_{ij}} \quad , \tag{2.192}$$

which is simply the negative of the potential term.

The contribution to be added to the atomic stress tensor is

$$\sigma^{\alpha\beta} = r_{ij}^{\alpha} f_j^{\beta} \quad , \tag{2.193}$$

where α, β are x, y, z components. The atomic stress tensor is symmetric.

In DL_POLY_4 these forces are handled by the subroutine COUL_CP_FORCES.

2.4.2 Force-Shifted Coulomb Sum

This form of the Coulomb sum has the advantage that it drastically reduces the range of electrostatic interactions, without giving rise to a violent step in the potential energy at the cutoff. Its main use is for preliminary preparation of systems and it is not recommended for realistic models.

The form of the simple truncated and shifted potential function is

$$U(r_{ij}) = \frac{q_i q_j}{4\pi\epsilon_0\epsilon}\left\{\frac{1}{r_{ij}} - \frac{1}{r_{\rm cut}}\right\} \quad , \tag{2.194}$$

with q_ℓ the charge on an atom labelled ℓ, $r_{\rm cut}$ the cutoff radius and r_{ij} the magnitude of the separation vector $\underline{r}_{ij} = \underline{r}_j - \underline{r}_i$.

A further refinement of this approach is to truncate the $1/r$ potential at $r_{\rm cut}$ and add a linear term to the potential in order to make both the energy and the force zero at the cutoff. This removes the heating effects that arise from the discontinuity in the forces at the cutoff in the simple truncated and shifted potential (the formula above). (The physics of this potential, however, is little better. It is only recommended for very crude structure optimizations.)

The force-shifted potential is thus

$$U(r_{ij}) = \frac{q_i q_j}{4\pi\epsilon_0\epsilon}\left[\left\{\frac{1}{r_{ij}} + \frac{1}{r_{\rm cut}^2}\ r_{ij}\right\} - \left\{\frac{1}{r_{\rm cut}} + \frac{1}{r_{\rm cut}^2}\ r_{\rm cut}\right\}\right] = \frac{q_i q_j}{4\pi\epsilon_0\epsilon}\left[\frac{1}{r_{ij}} + \frac{r_{ij}}{r_{\rm cut}^2} - \frac{2}{r_{\rm cut}}\right] \quad , \tag{2.195}$$

with the force on an atom j given by

$$\underline{f}_j = \frac{q_i q_j}{4\pi\epsilon_0\epsilon}\left[\frac{1}{r_{ij}^3} - \frac{1}{r_{ij} r_{\rm cut}^2}\right]\underline{r}_{ij} \quad , \tag{2.196}$$

with the force on atom i the negative of this.

The force-shifted Coulomb potential can be elegantly extended to emulate long-range ordering by including distance depending damping function erfc($\alpha\ r_{ij}$) (identical to that seen in the real-space portion of the Ewald sum) and thus mirror the effective charge screening [63] as shown below

$$\begin{aligned} U(r_{ij}) = \frac{q_i q_j}{4\pi\epsilon_0\epsilon} & \left[\left\{\frac{\mathrm{erfc}(\alpha\ r_{ij})}{r_{ij}} + \left(\frac{\mathrm{erfc}(\alpha\ r_{\rm cut})}{r_{\rm cut}^2} + \frac{2\alpha}{\sqrt{\pi}}\ \frac{\exp(-\alpha^2\ r_{\rm cut}^2)}{r_{\rm cut}}\right)\ r_{ij}\right\} - \right. \\ & \left.\left\{\frac{\mathrm{erfc}(\alpha\ r_{\rm cut})}{r_{\rm cut}} + \left(\frac{\mathrm{erfc}(\alpha\ r_{\rm cut})}{r_{\rm cut}^2} + \frac{2\alpha}{\sqrt{\pi}}\ \frac{\exp(-\alpha^2\ r_{\rm cut}^2)}{r_{\rm cut}}\right)\ r_{\rm cut}\right\}\right] \quad , \end{aligned} \tag{2.197}$$

with the force on an atom j given by

$$\begin{aligned} \underline{f}_j = \frac{q_i q_j}{4\pi\epsilon_0\epsilon} & \left[\left(\frac{\mathrm{erfc}(\alpha\ r_{ij})}{r_{ij}^2} + \frac{2\alpha}{\sqrt{\pi}}\ \frac{\exp(-\alpha^2\ r_{ij}^2)}{r_{ij}}\right) - \right. \\ & \left.\left(\frac{\mathrm{erfc}(\alpha\ r_{\rm cut})}{r_{\rm cut}^2} + \frac{2\alpha}{\sqrt{\pi}}\ \frac{\exp(-\alpha^2\ r_{\rm cut}^2)}{r_{\rm cut}}\right)\right]\ \frac{\underline{r}_{ij}}{r_{ij}} \quad , \end{aligned} \tag{2.198}$$

with the force on atom i the negative of this.

It is worth noting that, as discussed in [63] and references therein, this is only an approximation of the Ewald sum and its accuracy and effectiveness become better when the cutoff is large (> 10 preferably 12 Å).

The contribution to the atomic virial is

$$\mathcal{W} = -\underline{r}_{ij} \cdot \underline{f}_j \quad , \tag{2.199}$$

which is *not* the negative of the potential term in this case.

The contribution to be added to the atomic stress tensor is given by

$$\sigma^{\alpha\beta} = r_{ij}^{\alpha} f_j^{\beta} \quad , \tag{2.200}$$

where α, β are x, y, z components. The atomic stress tensor is symmetric.

In DL_POLY_4 these forces are handled by the routine COUL_FSCP_FORCES.

2.4.3 Coulomb Sum with Distance Dependent Dielectric

This potential attempts to address the difficulties of applying the direct Coulomb sum, without the brutal truncation of the previous case. It hinges on the assumption that the electrostatic forces are effectively 'screened' in real systems - an effect which is approximated by introducing a dielectric term that increases with distance.

The interatomic potential for two charged ions is

$$U(r_{ij}) = \frac{1}{4\pi\epsilon_0\epsilon(r_{ij})}\frac{q_i q_j}{r_{ij}} \quad , \tag{2.201}$$

with q_ℓ the charge on an atom labelled ℓ, and r_{ij} the magnitude of the separation vector $\underline{r}_{ij} = \underline{r}_j - \underline{r}_i$. $\epsilon(r)$ is the distance dependent dielectric function. In DL_POLY_4 it is assumed that this function has the form

$$\epsilon(r) \;=\; \epsilon\, r \quad , \tag{2.202}$$

where ϵ is a constant. Inclusion of this term effectively accelerates the rate of convergence of the Coulomb sum.

The force on an atom j derived from this potential is

$$\underline{f}_j = \frac{1}{2\pi\epsilon_0\epsilon}\frac{q_i q_j}{r_{ij}^4}\underline{r}_{ij} \quad , \tag{2.203}$$

with the force on atom i the negative of this.

The contribution to the atomic virial is

$$\mathcal{W} = -\underline{r}_{ij} \cdot \underline{f}_j \quad , \tag{2.204}$$

which is -2 times the potential term.

The contribution to be added to the atomic stress tensor is given by

$$\sigma^{\alpha\beta} = r_{ij}^{\alpha} f_j^{\beta} \quad , \tag{2.205}$$

where α, β are x, y, z components. The atomic stress tensor is symmetric.

In DL_POLY_4 these forces are handled by the routine COUL_DDDP_FORCES.

2.4.4 Reaction Field

In the reaction field method it is assumed that any given molecule is surrounded by a spherical cavity of finite radius within which the electrostatic interactions are calculated explicitly. Outside the cavity the system is treated as a dielectric continuum. The occurrence of any net dipole within the cavity induces a polarisation in the dielectric, which in turn interacts with the given molecule. The model allows the replacement of the infinite Coulomb sum by a finite sum plus the reaction field.

The reaction field model coded into DL_POLY_4 is the implementation of Neumann based on charge-charge interactions [64]. In this model, the total coulombic potential is given by

$$U_c = \frac{1}{4\pi\epsilon_0\epsilon}\sum_{j<n} q_j q_n \left[\frac{1}{r_{nj}} + \frac{B_0 r_{nj}^2}{2R_c^3}\right] \quad , \tag{2.206}$$

where the second term on the right is the reaction field correction to the explicit sum, with R_c the radius of the cavity. The constant B_0 is defined as

$$B_0 = \frac{2(\epsilon_1 - 1)}{(2\epsilon_1 + 1)} \quad , \tag{2.207}$$

with ϵ_1 the dielectric constant outside the cavity. The effective pair potential is therefore

$$U(r_{ij}) = \frac{1}{4\pi\epsilon_0\epsilon} q_i q_j \left[\frac{1}{r_{ij}} + \frac{B_0 r_{ij}^2}{2R_c^3}\right] \quad . \tag{2.208}$$

This expression unfortunately leads to large fluctuations in the system coulombic energy, due to the large 'step' in the function at the cavity boundary. In DL_POLY_4 this is countered by subtracting the value of the potential at the cavity boundary from each pair contribution. The term subtracted is

$$\frac{1}{4\pi\epsilon_0\epsilon}\frac{q_i q_j}{R_c}\left[1+\frac{B_0}{2}\right] \quad . \tag{2.209}$$

The effective pair force on an atom j arising from another atom n within the cavity is given by

$$\underline{f}_j = \frac{q_i q_j}{4\pi\epsilon_0\epsilon}\left[\frac{1}{r_{ij}^3}-\frac{B_0}{R_c^3}\right]\underline{r}_{ij} \quad . \tag{2.210}$$

In DL_POLY_4 the reaction field is optionally extended to emulate long-range ordering in a force-shifted manner by countering the reaction term and using a distance depending damping function erfc($\alpha\ r_{ij}$) (identical to that seen in the real-space portion of the Ewald sum) and thus mirror the effective charge screening [63]:

$$\begin{aligned} U(r_{ij}) = \frac{q_i q_j}{4\pi\epsilon_0\epsilon}\left[\left\{\frac{\mathrm{erfc}(\alpha\ r_{ij})}{r_{ij}}+\left(\frac{\mathrm{erfc}(\alpha\ r_{\mathrm{cut}})}{r_{\mathrm{cut}}^2}+\frac{2\alpha}{\sqrt{\pi}}\ \frac{\exp(-\alpha^2\ r_{\mathrm{cut}}^2)}{r_{\mathrm{cut}}}\right)\ r_{ij}\right\}-\right. \\ \left.\left\{\frac{\mathrm{erfc}(\alpha\ r_{\mathrm{cut}})}{r_{\mathrm{cut}}}+\left(\frac{\mathrm{erfc}(\alpha\ r_{\mathrm{cut}})}{r_{\mathrm{cut}}^2}+\frac{2\alpha}{\sqrt{\pi}}\ \frac{\exp(-\alpha^2\ r_{\mathrm{cut}}^2)}{r_{\mathrm{cut}}}\right)\ r_{\mathrm{cut}}\right\}+\frac{B_0(r_{ij}^2-r_{\mathrm{cut}}^2)}{2r_{\mathrm{cut}}^3}\right], \end{aligned} \tag{2.211}$$

with the force on an atom j given by

$$\begin{aligned} \underline{f}_j = \frac{q_i q_j}{4\pi\epsilon_0\epsilon}\left[\left(\frac{\mathrm{erfc}(\alpha\ r_{ij})}{r_{ij}^2}+\frac{2\alpha}{\sqrt{\pi}}\ \frac{\exp(-\alpha^2\ r_{ij}^2)}{r_{ij}}\right)-\right. \\ \left.\left(\frac{\mathrm{erfc}(\alpha\ r_{\mathrm{cut}})}{r_{\mathrm{cut}}^2}+\frac{2\alpha}{\sqrt{\pi}}\ \frac{\exp(-\alpha^2\ r_{\mathrm{cut}}^2)}{r_{\mathrm{cut}}}\right)-\frac{B_0 r_{ij}}{r_{\mathrm{cut}}^3}\right]\frac{\underline{r}_{ij}}{r_{ij}} \quad , \end{aligned} \tag{2.212}$$

with the force on atom i the negative of this.

It is worth noting that, as discussed in [63] and references therein, this is only an approximation of the Ewald sum and its accuracy and effectiveness become better when the cutoff is large (> 10 preferably 12 Å).

The contribution of each effective pair interaction to the atomic virial is

$$\mathcal{W} = -\underline{r}_{ij}\cdot\underline{f}_j \tag{2.213}$$

and the contribution to the atomic stress tensor is

$$\sigma^{\alpha\beta} = r_{ij}^{\alpha} f_j^{\beta} \quad , \tag{2.214}$$

where α, β are x, y, z components. The atomic stress tensor is symmetric.

In DL_POLY_4 the reaction field is handled by the subroutine COUL_RFP_FORCES.

2.4.5 Smoothed Particle Mesh Ewald

The Ewald sum [22] is the best technique for calculating electrostatic interactions in a periodic (or pseudo-periodic) system.

The basic model for a neutral periodic system is a system of charged point ions mutually interacting via the Coulomb potential. The Ewald method makes two amendments to this simple model. Firstly each ion is effectively neutralised (at long-ranged) by the superposition of a spherical Gaussian cloud of opposite charge centred on the ion. The combined assembly of point ions and Gaussian charges becomes the *Real*

Space part of the Ewald sum, which is now short ranged and treatable by the methods described above (Section 2)[3]. The second modification is to superimpose a second set of Gaussian charges, this time with the same charges as the original point ions and again centred on the point ions (so nullifying the effect of the first set of Gaussians). The potential due to these Gaussians is obtained from Poisson's equation and is solved as a Fourier series in *Reciprocal Space*. The complete Ewald sum requires an additional correction, known as the self energy correction, which arises from a Gaussian acting on its own site, and is constant. Ewald's method, therefore, replaces a potentially infinite sum in real space by two finite sums: one in real space and one in reciprocal space; and the self energy correction.

For molecular systems, as opposed to systems comprised simply of point ions, additional modifications EWALD_EXCL_FORCES are necessary to correct for the excluded (intra-molecular) coulombic interactions. In the real space sum these are simply omitted. In reciprocal space however, the effects of individual Gaussian charges cannot easily be extracted, and the correction is made in real space. It amounts to removing terms corresponding to the potential energy of an ion ℓ due to the Gaussian charge on a neighbouring charge m (or *vice versa*). This correction appears as the final term in the full Ewald formula below. The distinction between the error function erf and the more usual complementary error function erfc found in the real space sum, should be noted.

The same considerations and modifications EWALD_FRZN_FORCES are taken into account for frozen atoms, which mutual coulombic interaction must be excluded.

The total electrostatic energy is given by the following formula.

$$\begin{aligned} U_c &= \frac{1}{2V_o\epsilon_0\epsilon}\sum_{\underline{k}\neq\underline{0}}^{\infty}\frac{\exp(-k^2/4\alpha^2)}{k^2}|\sum_j^N q_j\exp(-i\underline{k}\cdot\underline{r}_j)|^2+\frac{1}{4\pi\epsilon_0\epsilon}\sum_{n<j}^{N^*}\frac{q_jq_n}{r_{nj}}\mathrm{erfc}(\alpha r_{nj})- \\ &\frac{1}{4\pi\epsilon_0\epsilon}\sum_{molecules}\sum_{\ell\leq m}^{M^*}q_\ell q_m\left\{\delta_{\ell m}\frac{\alpha}{\sqrt{\pi}}+\frac{\mathrm{erf}(\alpha r_{\ell m})}{r_{\ell m}^{1-\delta_{\ell m}}}\right\}- \\ &\frac{1}{4\pi\epsilon_0\epsilon}\sum_{\ell\leq m}^{F^*}q_\ell q_m\left\{\delta_{\ell m}\frac{\alpha}{\sqrt{\pi}}+\frac{\mathrm{erf}(\alpha r_{\ell m})}{r_{\ell m}^{1-\delta_{\ell m}}}\right\}-\frac{1}{4\pi\epsilon_0\epsilon}\frac{\pi}{2V_o\alpha^2}\left\{\sum_j^N q_j\right\}^2 \quad , \end{aligned} \tag{2.215}$$

where N is the number of ions in the system and N^* the same number discounting any excluded (intramolecular and frozen) interactions. M^* represents the number of excluded atoms in a given molecule. F^* represents the number of frozen atoms in the MD cell. V_o is the simulation cell volume and $\underline{k}$ is a reciprocal lattice vector defined by

$$\underline{k}=\ell\underline{u}+m\underline{v}+n\underline{w} \quad , \tag{2.216}$$

where ℓ, m, n are integers and $\underline{u}, \underline{v}, \underline{w}$ are the *reciprocal space* basis vectors. Both V_o and $\underline{u}, \underline{v}, \underline{w}$ are derived from the vectors $(\underline{a}, \underline{b}, \underline{c})$ defining the simulation cell. Thus

$$V_o=|\underline{a}\cdot\underline{b}\times\underline{c}| \tag{2.217}$$

and

$$\begin{aligned} \underline{u} &= 2\pi\frac{\underline{b}\times\underline{c}}{\underline{a}\cdot\underline{b}\times\underline{c}} \\ \underline{v} &= 2\pi\frac{\underline{c}\times\underline{a}}{\underline{a}\cdot\underline{b}\times\underline{c}} \\ \underline{w} &= 2\pi\frac{\underline{a}\times\underline{b}}{\underline{a}\cdot\underline{b}\times\underline{c}} \quad . \end{aligned} \tag{2.218}$$

With these definitions, the Ewald formula above is applicable to general periodic systems. The last term in the Ewald formula above is the Fuchs correction [65] for electrically non-neutral MD cells which prevents the build-up of a charged background and the introduction of extra pressure due to it.

[3]Strictly speaking, the real space sum ranges over all periodic images of the simulation cell, but in the DL_POLY_4 implementation, the parameters are chosen to restrict the sum to the simulation cell and its nearest neighbours i.e. the *minimum images* of the cell contents.

In practice the convergence of the Ewald sum is controlled by three variables: the real space cutoff $r_{\rm cut}$; the convergence parameter α and the largest reciprocal space vector $\underline{k}_{max}$ used in the reciprocal space sum. These are discussed more fully in Section 5.3.5. DL_POLY_4 can provide estimates if requested (see CONTROL file description 6.1.1).

As its name implies the Smoothed Particle Mesh Ewald (SPME) method is a modification of the standard Ewald method. DL_POLY_4 implements the SPME method of Essmann *et al.* [66]. Formally, this method is capable of treating van der Waals forces also, but in DL_POLY_4 it is confined to electrostatic forces only. The main difference from the standard Ewald method is in its treatment of the reciprocal space terms. By means of an interpolation procedure involving (complex) B-splines, the sum in reciprocal space is represented on a three dimensional rectangular grid. In this form the Fast Fourier Transform (FFT) may be used to perform the primary mathematical operation, which is a 3D convolution. The efficiency of these procedures greatly reduces the cost of the reciprocal space sum when the range of $\underline{k}$ vectors is large. The method (briefly) is as follows (for full details see [66]):

1. Interpolation of the $\exp(-i\ \underline{k}\cdot\underline{r}_j)$ terms (given here for one dimension):

$$\exp(2\pi i\ u_j k/L) \approx b(k) \sum_{\ell=-\infty}^{\infty} M_n(u_j-\ell)\ \exp(2\pi i\ k\ell/K)\ \ , \tag{2.219}$$

   in which k is the integer index of the $\underline{k}$ vector in a principal direction, K is the total number of grid points in the same direction and u_j is the fractional coordinate of ion j scaled by a factor K (i.e. $u_j = Ks_j^x$) . **Note** that the definition of the B-splines implies a dependence on the integer K, which limits the formally infinite sum over ℓ. The coefficients $M_n(u)$ are B-splines of order n and the factor $b(k)$ is a constant computable from the formula:

$$b(k) = \exp(2\pi i\ (n-1)k/K) \left[\sum_{\ell=0}^{n-2} M_n(\ell+1)\ \exp(2\pi i\ k\ell/K)\right]^{-1}\ \ . \tag{2.220}$$

2. Approximation of the structure factor $S(\underline{k})$:

$$S(\underline{k}) \approx b_1(k_1)\ b_2(k_2)\ b_3(k_3)\ Q^\dagger(k_1,k_2,k_3)\ \ , \tag{2.221}$$

   where $Q^\dagger(k_1,k_2,k_3)$ is the discrete Fourier transform of the *charge array* $Q(\ell_1,\ell_2,\ell_3)$ defined as

$$\begin{aligned} Q(\ell_1,\ell_2,\ell_3) &= \sum_{j=1}^{N} q_j \sum_{n_1,n_2,n_3} M_n(u_{1j}-\ell_1-n_1L_1)\ \times\ M_n(u_{2j}-\ell_2-n_2L_2)\ \times \\ &\qquad M_n(u_{3j}-\ell_3-n_3L_3)\ \ , \end{aligned} \tag{2.222}$$

   in which the sums over $n_{1,2,3}$ etc are required to capture contributions from all relevant periodic cell images (which in practice means the nearest images).

3. Approximating the reciprocal space energy U_{recip}:

$$U_{recip} = \frac{1}{2V_o\epsilon_0\epsilon} \sum_{k_1,k_2,k_3} G^\dagger(k_1,k_2,k_3)\ Q(k_1,k_2,k_3)\ \ , \tag{2.223}$$

   where $G^\dagger$ is the discrete Fourier transform of the function

$$G(k_1,k_2,k_3) = \frac{\exp(-k^2/4\alpha^2)}{k^2}\ B(k_1,k_2,k_3)\ (Q^\dagger(k_1,k_2,k_3))^*\ \ , \tag{2.224}$$

   in which $(Q^\dagger(k_1,k_2,k_3))^*$ is the complex conjugate of $Q^\dagger(k_1,k_2,k_3)$ and

$$B(k_1,k_2,k_3) = |b_1(k_1)|^2\ |b_2(k_2)|^2\ |b_3(k_3)|^2\ \ . \tag{2.225}$$

   The function $G(k_1,k_2,k_3)$ is thus a relatively simple product of the Gaussian screening term appearing in the conventional Ewald sum, the function $B(k_1,k_2,k_3)$ and the discrete Fourier transform of $Q(k_1,k_2,k_3)$.

4. Calculating the atomic forces, which are given formally by:

$$f_j^\alpha = -\frac{\partial U_{recip}}{\partial r_j^\alpha} = -\frac{1}{V_o \epsilon_0 \epsilon} \sum_{k_1,k_2,k_3} G^\dagger(k_1,k_2,k_3)\, \frac{\partial Q(k_1,k_2,k_3)}{\partial r_j^\alpha} \quad . \tag{2.226}$$

Fortunately, due to the recursive properties of the B-splines, these formulae are easily evaluated.

The virial and the stress tensor are calculated in the same manner as for the conventional Ewald sum.

The DL_POLY_4 subroutines required to calculate the SPME contributions are:

1. SPME_CONTAINER containing
   (a) BSPGEN, which calculates the B-splines
   (b) BSPCOE, which calculates B-spline coefficients
   (c) SPL_CEXP, which calculates the FFT and B-spline complex exponentials
2. PARALLEL_FFT and GPFA_MODULE (native DL_POLY_4 subroutines that respect the domain decomposition concept) which calculate the 3D complex fast Fourier transforms
3. EWALD_SPME_FORCES, which calculates the reciprocal space contributions (uncorrected)
4. EWALD_REAL_FORCES, which calculates the real space contributions (corrected)
5. EWALD_EXCL_FORCES, which calculates the reciprocal space corrections due to the coulombic exclusions in intramolecular interactions
6. EWALD_FRZN_FORCES, which calculates the reciprocal space corrections due to the exclusion interactions between frozen atoms
7. TWO_BODY_FORCES, in which all of the above subroutines are called sequentially and also the Fuchs correction [65] for electrically non-neutral MD cells is applied if needed.

2.5 Polarisation Shell Models

An atom or ion is polarisable if it develops a dipole moment when placed in an electric field. It is commonly expressed by the equation

$$\underline{\mu} = \alpha \underline{E} \quad , \tag{2.227}$$

where $\underline{\mu}$ is the induced dipole and $\underline{E}$ is the electric field. The constant α is the polarisability.

In the *static* shell model a polarisable atom is represented by a massive core and massless shell, connected by a harmonic spring, hereafter called the core-shell unit. The core and shell carry different electric charges, the sum of which equals the charge on the original atom. There is no electrostatic interaction (i.e. self interaction) between the core and shell of the same atom. Non-coulombic interactions arise from the shell alone.

The core-shell interaction is described by a harmonic spring potential of the form:

$$U_{spring}(r_{ij}) = \frac{1}{2} k_2 r_{ij}^2 \quad , \tag{2.228}$$

However, sometimes an anharmonic spring is used, described by a quartic form:

$$U_{spring}(r_{ij}) = \frac{1}{2} k_2 r_{ij}^2 + \frac{1}{4} k_4 r_{ij}^4 . \tag{2.229}$$

Normally, in practice, k_2 is much larger than k_4.

The effect of an external electric field, $\underline{E}$ is to separate the core and shell by a distance

$$\underline{d} = q_s \underline{E}/k_2 \quad , \tag{2.230}$$

giving rise to a *polarisation* dipole

$$\underline{\mu} = q_s \underline{d} \quad . \tag{2.231}$$

The condition of static equilibrium then gives the polarisability as:

$$\alpha = q_s^2/k_2 \quad , \tag{2.232}$$

where q_s is the shell charge and k_2 is the force constant of the harmonic spring.

The calculation of the forces, virial and stress tensor in this model is based on that for a diatomic molecule with charged atoms. The part coming from the spring potential is similar in spirit as for chemical bonds, equations (2.13-2.15), while the electrostatics is as described in the above section. The relationship between the kinetic energy and the temperature is different however, as the core-shell unit is permitted only three translational degrees of freedom, and the degrees of freedom corresponding to rotation and vibration of the unit are discounted as if the kinetic energy of these is regarded as zero (equation 3.11).

2.5.1 Dynamical (Adiabatic Shells) Shell Model

The dynamical shell model is a method of incorporating polarisability into a molecular dynamics simulation. The method used in DL_POLY_4 is that devised by Fincham *et al* [67] and is known as the adiabatic shell model.

In the adiabatic method, a fraction of the atomic mass is assigned to the shell to permit a dynamical description. The fraction of mass, x, is chosen to ensure that the natural frequency of vibration $\nu_{\texttt{core-shell}}$ of the harmonic spring (which depends on the reduced mass, i.e.

$$\nu_{\texttt{core-shell}} = \frac{1}{2\pi} \left[\frac{k_2}{x(1-x)m} \right]^{1/2} \quad , \tag{2.233}$$

with m the rigid ion atomic mass) is well above the frequency of vibration of the whole atom in the bulk system. Dynamically, the core-shell unit resembles a diatomic molecule with a harmonic bond, however, the high vibrational frequency of the bond prevents effective exchange of kinetic energy between the core-shell unit and the remaining system. Therefore, from an initial condition in which the core-shell units have negligible internal vibrational energy, the units will remain close to this condition throughout the simulation. This is essential if the core-shell unit is to maintain a net polarisation. (In practice, there is a slow leakage of kinetic energy into the core-shell units, but this should should not amount to more than a few percent of the total kinetic energy. To determine safe shell masses in practice, first a rigid ion simulation is performed in order to gather the velocity autocorrelation functions, VAF, of the ions of interest to polarise. Each VAF is then Fast Fourier transformed to find their highest frequency of interaction, $\nu_{\texttt{rigid-ion}}$. It is, then, a safe choice to assign a shell mass, $x\ m$, so that $\nu_{\texttt{core-shell}} \geq 3\ \nu_{\texttt{rigid-ion}}$. The user must make sure to assign the correct mass, $(1-x)\ m$, to the core!)

2.5.2 Relaxed (Massless Shells) Model

The relaxed shell model is presented in [68], where shells have no mass and as such their motion is not governed by the usual Newtonian equation, whereas their cores' motion is. Because of that, shells respond instantaneously to the motion of the cores: for any set of core positions, the positions of the shells are such that the force on every shell is zero. The energy is thus a minimum with respect to the shell positions. This represents the physical fact that the system is always in the ground state with respect to the electronic degrees of freedom.

Relaxation of the shells is carried out at each time step and involves a search in the multidimensional space of shell configurations. The search in DL_POLY_4 is based on the powerful conjugate-gradients technique [69] in an adaptation as shown in [68]. Each time step a few iterations (10÷30) are needed to achieve convergence to zero net force.

2.5.3 Breathing Shell Model Extension

While for low-symmetry structures, the conventional, dipolar rigid shell model (RSM) is sufficient to absorb most of the effects of partial covalency/ionic polarisation, for some high-symmetry systems, a breathing shell model (BSM) [70] is used as a refinement to represent the contribution of higher-order charge deformations (of oxide species). This is done by the inclusion of non-central ion interaction to account for a finite ion shell radius, r_i^o, which is allowed to deform isotropically under its environment. However, all short-range repulsion potentials, i.e. **vdw**, a BSM ion interacts by with its environment, must act upon the radius of the ion, $U = U(r_i - r_i^o)$, rather than the nuclear position, $U = U(r_i)$! A further constraining potential is then added to represent the self-energy of the ion's breathing shell. This most commonly uses the same shape as the one of the harmonic bond, see Section 2.2.1:

$$U(r_{ij}) = \frac{1}{2}k(r_{ij}^{BSM} - r_{ij}^o)^2 \quad , \tag{2.234}$$

where i and j are the intramolecular index of the ion's core and shell respectively. Hence, to employ the BSM, the user needs to specify an extra bonded interaction for each BSM'ed core-shell pair in the relevant **bonds** sections in the FIELD file for all molecules that contain BSM ions. It is worth noting that the BSM energy and virial are thus part of the bonds' energy and virial and their calculation as part of the bonds forces routine BONDS_FORCES.

The most significant consequence of the introduction of the BSM is that, by coupling the repulsive interactions via common shell radii, it creates a many-body effect that is able to reproduce the Cauchy violation ($C_{44} \neq C_{12}$) for rock salt structured materials.

2.5.4 Further Notes

In DL_POLY_4 the core-shell forces of the rigid shell model are handled by the routine CORE_SHELL_FORCES. In case of the adiabatic shell model the kinetic energy is calculated by CORE_SHELL_KINETIC and temperature scaling applied by routine CORE_SHELL_QUENCH. In case of the relaxed shell model shell are relaxed to zero force by CORE_SHELL_RELAXED.

All shell models can be used in conjunction with the methods for long-ranged forces described above.

Note that DL_POLY_4 determines which shell model to use by scanning shell weights provided the FIELD file (see Section 6.1.3). If all shells have zero weight the DL_POLY_4 will choose the relaxed shell model. If no shell has zero weight then DL_POLY_4 will choose the dynamical one. In case when some shells are massless and some are not DL_POLY_4 will terminate execution controllably and provide information about the error and possible possible choices of action in the OUTPUT file (see Section 6.2.6).

2.6 External Fields

In addition to the molecular force field, DL_POLY_4 allows the use of an *external* force field. Examples of fields available include:

1. Electric field: (**elec**)

$$\underline{F_i} = \underline{F_i} + q_i \cdot \underline{E} \tag{2.235}$$

2. Oscillating shear: (**oshr**)
$$\underline{F}_x = A\cos(2n\pi \cdot z/L_z) \tag{2.236}$$
3. Continuous shear: (**shrx**)
$$\underline{v}_x = \frac{1}{2}A\frac{|z|}{z} \qquad : |z| > z_0 \tag{2.237}$$
4. Gravitational field: (**grav**)
$$\underline{F_i} = \underline{F_i} + m_i \cdot \underline{G} \tag{2.238}$$
5. Magnetic field: (**magn**)
$$\underline{F_i} = \underline{F_i} + q_i \cdot (\underline{v_i} \times \underline{H}) \tag{2.239}$$
6. Containing sphere: (**sphr**)
$$\underline{F} = A\ (R_0 - r)^{-n} \qquad : r > R_{\rm cut} \tag{2.240}$$
7. Repulsive wall: (**zbnd**)
$$\underline{F}_z = A\ (z_o - z) \qquad : p \cdot z > p \cdot z_o \tag{2.241}$$
8. X-Piston: (**xpis**)
$$\underline{F}_x = \frac{m_k}{\sum_{k=i}^{j} m_k} P \cdot \underline{\rm Area}(\perp \text{ X-}direction) \qquad : \forall\ k = i, .., j \quad . \tag{2.242}$$
9. Harmonic restraint zone in z-direction: (**zres**)
$$\underline{F}_z = \begin{cases} A\ (z_{com} - z_{max}) & : \quad z_{com} > z_{max} \\ A\ (z_{min} - z_{com}) & : \quad z_{com} < z_{min} \quad , \end{cases} \tag{2.243}$$
where z_{com} is the chosen molecule centre of mass.
10. Harmonic restraint zone in z-direction - (pull out): (**zrs-**)
$$\underline{F}_z = \begin{cases} A\ (z - z_{max}) & : \quad z \geq (z_{max} + z_{min})/2 \\ A\ (z_{min} - z) & : \quad z < (z_{max} + z_{min})/2 \end{cases} \tag{2.244}$$
11. Harmonic restraint zone in z-direction + (pull in): (**zrs+**)
$$\underline{F}_z = \begin{cases} A\ (z - z_{max}) & : \quad z > z_{max} \\ A\ (z_{min} - z) & : \quad z < z_{min} \end{cases} \tag{2.245}$$
12. Oscillating electric field: (**osel**)
$$\underline{F_i} = \underline{F_i} + q_i \cdot \underline{E} \cdot \sin(2\pi\omega t) \ , \tag{2.246}$$
where t is the simulated time.

It is recommended that the use of an external field should be accompanied by a thermostat (this does not apply to examples 6 and 7, since these are conservative fields). The "Oscillating shear" and "X-piston" fields may only be used with orthorhombic cell geometry (imcon=1,2) and "Continuous shear" field with slab cell geometry (imcon=6).

In the case of the "X-piston" field it is strongly advised that the number of piston particles is chosen to be very small in comparison with the rest of the system (< 5%) and that the piston contains its own set of whole molecules (i.e. there are no molecules partially mapped on the piston), which do not include any core-shell, CB, PMF or RB units! The field releases the system's centre of mass to move unconstrained and gain momentum. This makes any temperature control options control the full kinetic energy of the system and thus the only ensemble valid under this conditions and possible within DL_POLY_4 at the present is the micro-canonical (NVE)!

The user is advised to be careful with the parameters' units! For more insight, do examine Table 6.17 and the example at equation (6.9) in Section 6.1.3.

In DL_POLY_4 external field forces are handled by the routines EXTERNAL_FIELD_APPLY and EXTERNAL_FIELD_CORRECT.

2.7 Treatment of Frozen Atoms, Rigid Body and Core-Shell Units

Frozen atoms, core-shell units and rigid body units are treated in a manner similar to that of the **intra**-molecular interactions due to their "by site" definition.

DL_POLY_4 allows for atoms to be completely immobilized (*i.e.* "frozen" at a fixed point in the MD cell). This is achieved by setting all forces and velocities associated with that atom to zero during each MD timestep. Frozen atoms are signalled by assigning an atom a non-zero value for the freeze parameter in the FIELD file. DL_POLY_4 does not calculate contributions to the virial or the stress tensor arising from the constraints required to freeze atomic positions. Neither does it calculate contributions from *intra*- and *inter*-molecular interactions between frozen atoms. As with the tethering potential, the reference position of a frozen site is scaled with the cell vectors in constant pressure simulations. In the case of frozen rigid bodies, their "centre of mass" is scaled with the cell vectors in constant pressure simulations and the positions of their constituent sites are then moved accordingly.

In DL_POLY_4 the frozen atom option is handled by the subroutine FREEZE_ATOMS.

The rigid body dynamics (see Section 3.6) is resolved by solving the Eulerian equations of rotational motion. However, their statics includes calculation of the individual contributions of each RB's centre of mass stress and virial due to the action of the resolved forces on sites/atoms constituting it. These contribute towards the total system stress and pressure.

As seen in Section 2.5 core-shell units are dealt with (i) kinetically by the adiabatic shell model or (ii) statically by the dynamic shell model. Both contribute to the total system stress (pressure) but in different manner. The former does it via the kinetic stress (energy) and atomic stress (potential energy) due to the core-shell spring. The latter via atomic stress (potential energy) due to the shells move to minimised configuration.

2.8 Tabulation and interpolation in the treatment of intermolecular interactions

By default DL_POLY_4 tabulates in memory most of the intermolecular interactions keeping values of the potential and the negative of its first derivative times the distance (or virial) over an equidistant grid. This is done for reasons of speed as due to the large variety of potential forms, some could be quite expensive to evaluate if run unoptimised. The memory tabulation could be overridden for non-tabulated interactions upon user specified options such as **metal direct** for metal interactions and **vdw direct** for van der Waals interactions.

When energy and force are calculated for tabulated interactions a 3-point interpolation scheme of our own is used to interpolate the value for the requested distance.

A 5-point interpolation is used for finding the numerical derivatives of (2B)(E)EAM type potentials which are supplied in TABEAM by the user. For this a Lagrange formula is used, which can be found in any textbook on numerical methods.

2.9 Open Knowledgebase of Interatomic Models - openKIM

DL_POLY_4 force-field allows for full model interactions specification by using an openKIM model - http://OpenKIM.org/. A KIM model contains all necessary interactions (bonded and non-bonded of any kind) and their parameters for a specific model within a designated container. Thus a KIM model can be used as a force-field container when made available to DL_POLY_4 at run time (see the description of the FIELD file in Section 6.1.3) upon a specification within FIELD together with a matching molecular description for the model system, also specified in FIELD. Due to the history of the openKIM initiative and the constraints of the logic of the DL_POLY_4 FIELD file the designated place for a KIM model specification is in the VDW section. In order to use openKIM functionality within DL_POLY_4 one must further ensure that DL_POLY_4 is cross-compiled with openKIM. For guidance on this together with testing, please refer to README_KIM.TXT in *source* and follow instructions within.

Chapter 3

Integration Algorithms

Scope of Chapter

This chapter describes the integration algorithms coded into DL_POLY_4.

3.1 Introduction

As a default the DL_POLY_4 integration algorithms are based on the Velocity Verlet (VV) scheme, which is both simple and time reversible [22]. It generates trajectories in the microcanonical (NVE) ensemble in which the total energy (kinetic plus potential) is conserved. If this property drifts or fluctuates excessively in the course of a simulation it indicates that the timestep is too large or the potential cutoffs too small (relative r.m.s. fluctuations in the total energy of 10^{-5} are typical with this algorithm).

The VV algorithm has two stages (VV1 and VV2). At the first stage it requires values of position ($\underline{r}$), velocity ($\underline{v}$) and force ($\underline{f}$) at time t. The first stage is to advance the velocities to $t+(1/2)\Delta t$ by integration of the force and then to advance the positions to a full step $t+\Delta t$ using the new half-step velocities:

1. VV1:
$$\underline{v}(t+\frac{1}{2}\Delta t) \leftarrow \underline{v}(t) + \frac{\Delta t}{2}\,\frac{\underline{f}(t)}{m} \quad , \tag{3.1}$$
where m is the mass of a site and Δt is the timestep
$$\underline{r}(t+\Delta t) \leftarrow \underline{r}(t) + \Delta t\;\underline{v}(t+\frac{1}{2}\Delta t) \tag{3.2}$$

2. FF:
Between the first and the second stage a recalculation of the force at time $t+\Delta t$ is required since the positions have changed
$$\underline{f}(t+\Delta t) \leftarrow \underline{f}(t) \tag{3.3}$$

3. VV2:
In the second stage the half-step velocities are advanced to to a full step using the new force
$$\underline{v}(t+\Delta t) \leftarrow \underline{v}(t+\frac{1}{2}\Delta t) + \frac{\Delta t}{2}\,\frac{\underline{f}(t+\Delta t)}{m} \tag{3.4}$$

DL_POLY_4 also offers integration algorithms based on the leapfrog Verlet (LFV) scheme [22]. Although LFV scheme is somewhat simpler and numerically faster than the VV scheme, it is not time reversible and does not offer the numerical stability the VV scheme does. Furthermore, all kinetic related properties have approximate estimators due to the half a step out of phase between velocity and position.

The LFV algorithm is one staged. It requires values of position ($\underline{r}$) and force ($\underline{f}$) at time t and velocity ($\underline{v}$) at half a timestep behind: $t-(1/2)\Delta t$. Firstly, the forces are recalculated afresh at time t (from time $t-\Delta t$) since the positions have changed from the last step:

1. FF:
$$\underline{f}(t) \leftarrow \underline{f}(t-\Delta t) \quad , \tag{3.5}$$
where Δt is the timestep

2. LFV:
The velocities are advanced by a timestep to $t+(1/2)\Delta t$ by integration of the new force
$$\underline{v}(t+\frac{1}{2}\Delta t) \leftarrow \underline{v}(t-\frac{1}{2}\Delta t) + \Delta t\,\frac{\underline{f}(t)}{m} \quad , \tag{3.6}$$
where m is the mass of a site, and then the positions are advanced to a full step $t+\Delta t$ using the new half-step velocities
$$\underline{r}(t+\Delta t) \leftarrow \underline{r}(t) + \Delta t\;\underline{v}(t+\frac{1}{2}\Delta t) \tag{3.7}$$

Molecular dynamics simulations normally require properties that depend on position and velocity *at the same time* (such as the sum of potential and kinetic energy). The velocity at time t is obtained from the average of the velocities half a timestep either side of timestep t:

$$\underline{v}(t) \leftarrow \frac{1}{2}\left[\underline{v}(t-\frac{1}{2}\Delta t)+\underline{v}(t+\frac{1}{2}\Delta t)\right] \quad . \tag{3.8}$$

The instantaneous kinetic energy, for example, can then be obtained from the atomic velocities as

$$E_{kin}(t) = \frac{1}{2}\sum_{1}^{\mathcal{N}} m_i v_i^2(t) \quad , \tag{3.9}$$

and assuming the system has no net momentum the instantaneous temperature is

$$\mathcal{T}(t) = \frac{2}{k_B f} E_{kin}(t) \quad , \tag{3.10}$$

where i labels particles (that can be free atoms or rigid bodies), $\mathcal{N}$ the number of particles (free atoms and rigid bodies) in the system, k_B the Boltzmann's constant and f the number of degrees of freedom in the system.

$$f = 3\mathcal{N} - 3\mathcal{N}_{frozen} - 3\mathcal{N}_{shells} - \mathcal{N}_{constraints} - 3 - p \quad . \tag{3.11}$$

Here $\mathcal{N}_{frozen}$ indicates the number of frozen atoms in the system, $\mathcal{N}_{shells}$ number of core-shell units and $\mathcal{N}_{constraints}$ number of bond and PMF constraints. Three degrees of freedom are subtracted for the centre of mass zero net momentum (which we impose) and p is zero for periodic or three for non-periodic systems, where it accounts for fixing angular momentum about origin (which we impose).

In the case of rigid bodies (see Section 3.6) the first part of equation (3.11)

$$f\prime = 3\mathcal{N} - 3\mathcal{N}_{frozen} \tag{3.12}$$

splits into

$$f\prime = \left(3\mathcal{N}^{FP} - 3\mathcal{N}^{FP}_{frozen}\right) + \left(3\mathcal{N}^{RB(\mathbf{tra})} - 3\mathcal{N}^{RB(\mathbf{tra})}_{frozen}\right) + \left(3\mathcal{N}^{RB(\mathbf{rot})} - 3\mathcal{N}^{RB(\mathbf{rot})}_{frozen}\right) \tag{3.13}$$

or

$$f\prime = f^{FP} + f^{RB(\mathbf{tra})} + f^{RB(\mathbf{rot})} \quad . \tag{3.14}$$

Here FP stands for a free particle, i.e. a particle not participating in the constitution of a rigid body, and RB for a rigid body. In general a rigid body has 3 translational (**tra**) degrees of freedom, corresponding to its centre of mass being allowed to move in the 3 general direction of space, and 3 rotational (**rot**), corresponding to the RB being allowed to rotate around the 3 general axis in space. It is not far removed to see that for a not fully frozen rigid body one must assign 0 translational degrees of freedom but depending on the "frozenness" of the RB one may assign 1 rotational degrees of freedom when all the frozen sites are in line (i.e. rotation around one axis only) or 3 when just one site is frozen.

The routines NVE_0_VV and NVE_0_LFV implement the Verlet algorithm in velocity and leapfrog flavours respectively for free particles and calculate the instantaneous temperature. Whereas the routines NVE_1_VV and NVE_1_LFV implement the same for systems also containing rigid bodies. The conserved quantity is the total energy of the system

$$\mathcal{H}_{\rm NVE} = U + E_{kin} \quad , \tag{3.15}$$

where U is the potential energy of the system and E_{kin} the kinetic energy at time t.

The full selection of integration algorithms, indicating both VV and LFV cast integration, within DL_POLY_4 is as follows:

NVE_0_VV	NVE_0_LFV	Constant E algorithm
NVE_1_VV	NVE_1_LFV	The same as the above but also incorporating RB integration
DPD_THERMOSTAT	N/A	Constant T algorithm (DPD [71])
NVT_E0_VV	NVT_E0_LFV	Constant E_{kin} algorithm (Evans [26])
NVT_E1_VV	NVT_E1_LFV	The same as the above but also incorporating RB integration
NVT_L0_VV	NVT_L0_LFV	Constant T algorithm (Langevin [27])
NVT_L1_VV	NVT_L1_LFV	The same as the above but also incorporating RB integration
NVT_A0_VV	NVT_A0_LFV	Constant T algorithm (Andersen [28])
NVT_A1_VV	NVT_A1_LFV	The same as the above but also incorporating RB integration
NVT_B0_VV	NVT_B0_LFV	Constant T algorithm (Berendsen [29])
NVT_B1_VV	NVT_B1_LFV	The same as the above but also incorporating RB integration
NVT_H0_VV	NVT_H0_LFV	Constant T algorithm (Hoover [30])
NVT_H1_VV	NVT_H1_LFV	The same as the above but also incorporating RB integration
NVT_G0_VV	NVT_G0_LFV	Constant T algorithm (GST [72])
NVT_G1_VV	NVT_G1_LFV	The same as the above but also incorporating RB integration
NPT_L0_VV	NPT_L0_LFV	Constant T,P algorithm (Langevin [31])
NPT_L1_VV	NPT_L1_LFV	The same as the above but also incorporating RB integration
NPT_B0_VV	NPT_B0_LFV	Constant T,P algorithm (Berendsen [29])
NPT_B1_VV	NPT_B1_LFV	The same as the above but also incorporating RB integration
NPT_H0_VV	NPT_H0_LFV	Constant T,P algorithm (Hoover [30])
NPT_H1_VV	NPT_H1_LFV	The same as the above but also incorporating RB integration
NPT_M0_VV	NPT_M0_LFV	Constant T,P algorithm (Martyna-Tuckerman-Klein [32])
NPT_M1_VV	NPT_M1_LFV	The same as the above but also incorporating RB integration
NPT_L0_VV	NPT_L0_LFV	Constant T,$\underline{\underline{\sigma}}$ algorithm (Langevin [31])
NPT_L1_VV	NPT_L1_LFV	The same as the above but also incorporating RB integration
NST_B0_VV	NST_B0_LFV	Constant T,$\underline{\underline{\sigma}}$ algorithm (Berendsen [29])
NST_B1_VV	NST_B1_LFV	The same as the above but also incorporating RB integration
NST_H0_VV	NST_H0_LFV	Constant T,$\underline{\underline{\sigma}}$ algorithm (Hoover [30])
NST_H1_VV	NST_H1_LFV	The same as the above but also incorporating RB integration
NST_M0_VV	NST_M0_LFV	Constant T,$\underline{\underline{\sigma}}$ algorithm (Martyna-Tuckerman-Klein [32])
NST_M0_VV	NST_M0_LFV	The same as the above but also incorporating RB integration

It is worth noting that the last four ensembles are also optionally available in an extended from to constant normal pressure and constant surface area, NP$_n$AT, or constant surface tension, NP$_n\gamma$T [73].

3.2 Bond Constraints

The SHAKE algorithm for bond constraints was devised by Ryckaert *et al.* [74] and is widely used in molecular simulation. It is a two stage algorithm based on the leapfrog Verlet integration scheme [22]. In the first stage the LFV algorithm calculates the motion of the atoms in the system assuming a complete absence of the rigid bond forces. The positions of the atoms at the end of this stage do not conserve the distance constraint required by the rigid bond and a correction is necessary. In the second stage the deviation in the length of a given rigid bond is used retrospectively to compute the constraint force needed to conserve the bondlength. It is relatively simple to show that the constraint force has the form:

$$\underline{G}_{ij} \approx \frac{1}{2}\frac{\mu_{ij}}{\Delta t^2}\frac{(d_{ij}^2 - d_{ij}'^2)}{\underline{d}_{ij}^o \cdot \underline{d}_{ij}'}\,\underline{d}_{ij}^o \quad , \tag{3.16}$$

where: μ_{ij} is the reduced mass of the two atoms connected by the bond; $\underline{d}_{ij}^o$ and $\underline{d}_{ij}'$ are the original and intermediate bond vectors; d_{ij} is the constrained bondlength; and Δt is the Verlet integration time step. It should be noted that this formula is an approximation only.

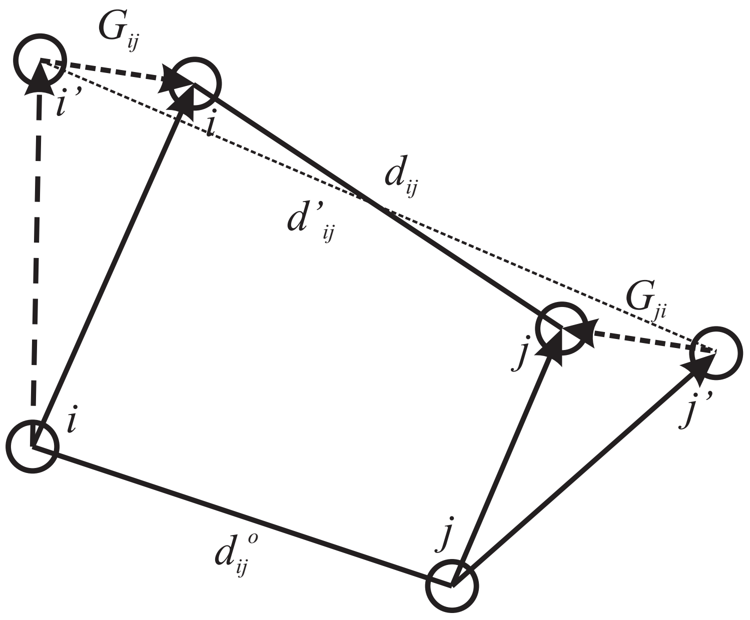

Figure 3.1: The SHAKE (RATTLE_VV1) schematics and associated vectors. The algorithm calculates the constraint force $\underline{G}_{ij} = -\underline{G}_{ji}$ that conserves the bondlength d_{ij} between atoms i and j, following the initial movement to positions $i\prime$ and $j\prime$ under the unconstrained forces $\underline{F}_i$ and $\underline{F}_j$ and velocities $\underline{v}_i$ and $\underline{v}_j$.

The RATTLE algorithm was devised by Andersen [23] and it fits within the concept of the Velocity Verlet integration scheme. It consists of two parts RATTLE_VV1 and RATTLE_VV2 applied respectively in stages one and two of Velocity Verlet algorithm. RATTLE_VV1 is similar to the SHAKE algorithm as described above and handles the bond length constraint. However, due to the difference in the velocity update between VV (VV1) and LFV schemes, the constraint force generated to conserve the bondlength in RATTLE_VV1 has the form as in (3.16) but missing the factor of a half:

$$\underline{G}_{ij} \approx \frac{\mu_{ij}}{\Delta t^2} \frac{(d_{ij}^2 - d_{ij}'^2)}{\underline{d}_{ij}^o \cdot \underline{d}_{ij}'} \, \underline{d}_{ij}^o \quad . \tag{3.17}$$

The constraint force in RATTLE_VV2 imposes a new condition of rigidity on constraint bonded atom velocities. RATTLE_VV2 is also a two stage algorithm. In the first stage, the VV2 algorithm calculates the velocities of the atoms in the system assuming a complete absence of the rigid bond forces (since forces have just been recalculated afresh after VV1). The relative velocity of atom i with respect to atom j (or vice versa) constituting the rigid bond ij may not be perpendicular to the bond - i.e. may have a non-zero component along the bond. However, by the stricter definition of rigidity this is is required to be zero as it will otherwise lead to a change in the rigid bond length during the consequent timestepping. In the second stage the deviation from zero of the scalar product $\underline{d}_{ij} \cdot (\underline{v}_j - \underline{v}_i)$ is used retrospectively to compute the constraint force needed to keep the bond rigid over the length of the timestep Δt. It is relatively simple to show that the constraint force has the form:

$$\underline{B}_{ij} \approx \frac{\mu_{ij}}{\Delta t} \frac{\underline{d}_{ij} \cdot (\underline{v}_j - \underline{v}_i)}{d_{ij}^2} \, \underline{d}_{ij} \quad . \tag{3.18}$$

The velocity corrections can therefore be written as

$$\underline{v}_i^{corr} = \Delta t \, \frac{\underline{B}_{ij}}{m_i} = \frac{\mu_{ij}}{m_i} \frac{\underline{d}_{ij} \cdot (\underline{v}_j - \underline{v}_i)}{d_{ij}^2} \, \underline{d}_{ij} \quad . \tag{3.19}$$

For a system of simple diatomic molecules, computation of the constraint force will, in principle, allow the correct atomic positions to be calculated in one pass. However, in the general polyatomic case this correction is merely an interim adjustment, not only because the above formula is approximate, but the successive correction of other bonds in a molecule has the effect of perturbing previously corrected bonds. Either part of the RATTLE algorithm is therefore iterative, with the correction cycle being repeated for all bonds until: each has converged to the correct length, within a given tolerance for RATTLE_VV1 (SHAKE) and the relative bond velocities are perpendicular to their respective bonds within a given tolerance for RATTLE_VV2 (RATTLE). The tolerance may be of the order 10^{-4} Å to 10^{-8} Å depending on the precision desired.

The SHAKE procedure may be summarised as follows:

1. All atoms in the system are moved using the LFV algorithm, assuming an absence of rigid bonds (constraint forces). (This is stage 1 of the SHAKE algorithm.)
2. The deviation in each bondlength is used to calculate the corresponding constraint force, equation (3.16), that (retrospectively) 'corrects' the bond length.
3. After the correction, equation (3.16), has been applied to all bonds, every bondlength is checked. If the largest deviation found exceeds the desired tolerance, the correction calculation is repeated.
4. Steps 2 and 3 are repeated until all bondlengths satisfy the convergence criterion (this iteration constitutes stage 2 of the SHAKE algorithm).

The RATTLE procedures may be summarised as follows:

1. RATTLE stage 1:
   (a) All atoms in the system are moved using the VV algorithm, assuming an absence of rigid bonds (constraint forces). (This is stage 1 of the RATTLE_VV1 algorithm.)
   (b) The deviation in each bondlength is used to calculate the corresponding constraint force, equation (3.17), that (retrospectively) 'corrects' the bond length.
   (c) After the correction, equation (3.17), has been applied to all bonds, every bondlength is checked. If the largest deviation found exceeds the desired tolerance, the correction calculation is repeated.
   (d) Steps (b) and (c) are repeated until all bondlengths satisfy the convergence criterion (this iteration constitutes stage 2 of the RATTLE_VV1 algorithm).
2. Forces calculated afresh.
3. RATTLE stage 2:
   (a) All atom velocities are updated to a full step, assuming an absence of rigid bonds. (This is stage 1 of the RATTLE_VV2 algorithm.)
   (b) The deviation of $\underline{d}_{ij} \cdot (\underline{v}_j - \underline{d}_i)$ in each bond is used to calculate the corresponding constraint force that (retrospectively) 'corrects' the bond velocities.
   (c) After the correction, equation (3.18), has been applied to all bonds, every bond velocity is checked against the above condition. If the largest deviation found exceeds the desired tolerance, the correction calculation is repeated.
   (d) Steps (b) and (c) are repeated until all bonds satisfy the convergence criterion (this iteration constitutes stage 2 of the RATTLE_VV2 algorithm).

The parallel version of the RATTLE algorithm, as implemented in DL_POLY_4, is derived from the RD_SHAKE algorithm [8] although its implementation in the Domain Decomposition framework requires no *global merging* operations and is consequently significantly more efficient. The routine CONSTRAINTS_SHAKE is called to apply corrections to the atomic positions and the routine CONSTRAINTS_RATTLE to apply corrections to the atomic velocities of constrained particles.

It should be noted that the fully converged constraint forces G_{ij} make a contribution to the system virial and the stress tensor.

The contribution to be added to the atomic virial (for each constrained bond) is

$$\mathcal{W} = -\underline{d}_{ij} \cdot \underline{G}_{ij} \quad . \tag{3.20}$$

The contribution to be added to the atomic stress tensor (for each constrained bond) is given by

$$\sigma^{\alpha\beta} = d_{ij}^{\alpha} G_{ij}^{\beta} \quad , \tag{3.21}$$

where α and β indicate the x, y, z components. The atomic stress tensor derived from the pair forces is symmetric.

3.3 Potential of Mean Force (PMF) Constraints and the Evaluation of Free Energy

A generalization of bond constraints can be made to constrain a system to some point along a reaction coordinate. A simple example of such a reaction coordinate would be the distance between two ions in solution. If a number of simulations are conducted with the system constrained to different points along the reaction coordinate then the mean constraint force may be plotted as a function of reaction coordinate and the function integrated to obtain the free energy for the overall process [75]. The PMF constraint force, virial and contributions to the stress tensor are obtained in a manner analogous to that for a bond constraint (see previous section). The only difference is that the constraint is now applied between the centres of two groups which need not be atoms alone. DL_POLY_4 reports the PMF constraint virial, $\mathcal{W}_{PMF}$, for each simulation. Users can convert this to the PMF constraint force from

$$G_{PMF} = \frac{\mathcal{W}_{PMF}}{d_{PMF}} \quad , \tag{3.22}$$

where is d_{PMF} the constraint distance between the two groups used to define the reaction coordinate.

The routines PMF_SHAKE and PMF_RATTLE are called to apply corrections to the atomic positions and respectively the atomic velocities of all particles constituting PMF units.

In presence of both bond constraints and PMF constraints. The constraint procedures, i.e. SHAKE or RATTLE, for both types of constraints are applied iteratively in order bonds-PMFs until convergence of $\mathcal{W}_{PMF}$ reached. The number of iteration cycles is limited by the same limit as for the bond constraints' procedures (SHAKE/RATTLE).

3.4 Thermostats

The system may be coupled to a heat bath to ensure that the average system temperature is maintained close to the requested temperature, $T_{\rm ext}$. When this is done the equations of motion are modified and the system no longer samples the microcanonical ensemble. Instead trajectories in the canonical (NVT) ensemble, or something close to it are generated. DL_POLY_4 comes with seven different thermostats: Evans (Gaussian constraints) [26], Langevin [27, 76], Andersen [28], Berendsen [29], Nosé-Hoover (N-H) [30] and the gentle stochastic thermostat (GST) [72, 77] as well as Dissipative Particle Dynamics (DPD) method [78, 79, 50, 71], Of these, only the Langevin, N-H, GST and DPD algorithms generate *true* trajectories in the canonical (NVT) ensemble. The rest will produce properties that typically differ from canonical averages by $\mathcal{O}(1/\mathcal{N})$ [22] (where $\mathcal{N}$ is the number of particles in the system), as the Evans algorithm generates trajectories in the (NVE$_{kin}$) ensemble.

3.4.1 Evans Thermostat (Gaussian Constraints)

Kinetic temperature can be made a constant of the equations of motion by imposing an additional constraint on the system. If one writes the equations of motion as:

$$\begin{aligned} \frac{d\underline{r}(t)}{dt} &= \underline{v}(t) \\ \frac{d\underline{v}(t)}{dt} &= \frac{\underline{f}(t)}{m} - \chi(t)\ \underline{v}(t) \ \ , \end{aligned} \tag{3.23}$$

the kinetic temperature constraint χ can be found as follows:

$$\begin{aligned} \frac{d}{dt}\mathcal{T} &\propto \frac{d}{dt}\left(\frac{1}{2}\sum_i m_i \underline{v}_i^2\right) = \sum_i m_i \underline{v}_i \cdot \frac{d}{dt}\underline{v}_i = 0 \\ & \sum_i m_i \underline{v}_i(t) \cdot \left\{\frac{\underline{f}_i(t)}{m_i} - \chi(t)\ \underline{v}_i(t)\right\} = 0 \\ \chi(t) &= \frac{\sum_i \underline{v}_i(t) \cdot \underline{f}_i(t)}{\sum_i m_i \underline{v}_i^2(t)} \ \ , \end{aligned} \tag{3.24}$$

where $\mathcal{T}$ is the instantaneous temperature defined in equation (3.10).

The VV implementation of the Evans algorithm is straight forward. The conventional VV1 and VV2 steps are carried out as before the start of VV1 and after the end of VV2 there is an application of thermal constraining. This involves the calculation of $\chi(t)$ before the VV1 stage and $\chi(t+\Delta t)$ after the VV2 stage with consecutive thermalisation on the unthermostated velocities for half a timestep at each stage in the following manner:

1. Thermostat VV1

$$\begin{aligned} \chi(t) &\leftarrow \frac{\sum_i \underline{v}_i(t) \cdot \underline{f}_i(t)}{2\ E_{kin}(t)} \\ \underline{v}(t) &\leftarrow \underline{v}(t)\ \exp\left(-\chi(t)\frac{\Delta t}{2}\right) \ \ . \end{aligned} \tag{3.25}$$

2. VV1:

$$\begin{aligned} \underline{v}(t+\frac{1}{2}\Delta t) &\leftarrow \underline{v}(t) + \frac{\Delta t}{2}\ \frac{\underline{f}(t)}{m} \\ \underline{r}(t+\Delta t) &\leftarrow \underline{r}(t) + \Delta t\ \underline{v}(t+\frac{1}{2}\Delta t) \end{aligned} \tag{3.26}$$

3. RATTLE_VV1
4. FF:

$$\underline{f}(t+\Delta t) \leftarrow \underline{f}(t) \tag{3.27}$$

5. VV2:

$$\underline{v}(t+\Delta t) \leftarrow \underline{v}(t+\frac{1}{2}\Delta t) + \frac{\Delta t}{2}\left[\frac{\underline{f}(t+\Delta t)}{m}\right] \tag{3.28}$$

6. RATTLE_VV2
7. Thermostat VV2

$$\begin{aligned} \chi(t+\Delta t) &\leftarrow \frac{\sum_i \underline{v}_i(t+\Delta t) \cdot \underline{f}_i(t+\Delta t)}{2\ E_{kin}(t+\Delta t)} \\ \underline{v}(t+\Delta t) &\leftarrow \underline{v}(t+\Delta t)\ \exp\left(-\chi(t+\Delta t)\frac{\Delta t}{2}\right) \ \ . \end{aligned} \tag{3.29}$$

The algorithm is self-consistent and requires no iterations.

The LFV implementation of the Evans algorithm is iterative as an initial estimate of $\chi(t)$ at full step is calculated using an unconstrained estimate of the velocity at full step, $\underline{v}(t)$). The iterative part is as follows:

1. FF:
$$\underline{f}(t) \leftarrow \underline{f}(t-\Delta t) \tag{3.30}$$

2. LFV: The iterative part is as follows:
$$\begin{aligned} \texttt{scale} = 1 + \chi(t)\ \frac{\Delta t}{2} \quad &, \quad \texttt{scale\_v} = \frac{2}{\texttt{scale}} - 1\ ,\ \texttt{scale\_f} = \frac{\Delta t}{\texttt{scale}} \\ \underline{v}(t+\frac{1}{2}\Delta t) \quad &\leftarrow \quad \texttt{scale\_v}\ \underline{v}(t-\frac{1}{2}\Delta t) + \texttt{scale\_f}\ \frac{\underline{f}(t)}{m} \\ \underline{r}(t+\Delta t) \quad &\leftarrow \quad \underline{r}(t) + \Delta t\ \underline{v}(t+\frac{1}{2}\Delta t) \end{aligned} \tag{3.31}$$

3. SHAKE

4. Full step velocity:
$$\underline{v}(t) \leftarrow \frac{1}{2}\left[\underline{v}(t-\frac{1}{2}\Delta t) + \underline{v}(t+\frac{1}{2}\Delta t)\right] \tag{3.32}$$

5. Thermostat:
$$\chi(t) \leftarrow \frac{\sum_i \underline{v}_i(t)\cdot \underline{f}_i(t)}{2\ E_{kin}(t)} \quad . \tag{3.33}$$

Several iterations are required to obtain self consistency. In DL_POLY_4 the number of iterations is set to 8 (9 if bond constraints are present).

The conserved quantity by these algorithms is the system kinetic energy.

The VV and LFV flavours of the Gaussian constraints algorithm are implemented in the DL_POLY_4 routines NVT_E0_VV and NVT_E0_LFV respectively. The routines NVT_E1_VV and NVT_E1_LFV implement the same but also incorporate RB dynamics.

3.4.2 Langevin Thermostat

The Langevin thermostat works by coupling every particle to a viscous background and a stochastic heath bath (Brownian dynamics) such that

$$\begin{aligned} \frac{d\underline{r}_i(t)}{dt} &= \underline{v}_i(t) \\ \frac{d\underline{v}_i(t)}{dt} &= \frac{\underline{f}_i(t) + \underline{R}_i(t)}{m_i} - \chi\ \underline{v}_i(t) \quad , \end{aligned} \tag{3.34}$$

where χ is the user defined *constant* (positive, in units of ps^{-1}) specifying the thermostat friction parameter and $R(t)$ is stochastic force with zero mean that satisfies the fluctuation- dissipation theorem:

$$\left\langle R_i^\alpha(t)\ R_j^\beta(t')\right\rangle = 2\ \chi\ m_i\ k_B T\ \delta_{ij}\ \delta_{\alpha\beta}\ \delta(t-t') \quad , \tag{3.35}$$

where superscripts denote Cartesian indices, subscripts particle indices, k_B is the Boltzmann constant, T the target temperature and m_i the particle's mass. The Stokes-Einstein relation for the diffusion coefficient can then be used to show that the average value of $R_i(t)$ over a time step (in thermal equilibrium) should be a random deviate drawn from a Gaussian distribution of zero mean and unit variance, $\texttt{Gauss}(0,1)$, scaled by $\sqrt{\frac{2\ \chi\ m_i\ k_B T}{\Delta t}}$.

The effect of this algorithm is thermostat the system on a local scale. Particles that are too "cold" are given more energy by the noise term and particles that are too "hot" are slowed down by the friction. Numerical instabilities, which usually arise from inaccurate calculation of a local collision-like process, are thus efficiently kept under control and cannot propagate.

The generation of random forces is implemented in the routine LANGEVIN_FORCES.

The VV implementation of the algorithm is tailored in a Langevin Impulse (LI) manner [76]:

1. VV1:

$$\begin{aligned} \underline{v}(t+\epsilon) &\leftarrow \underline{v}(t) + \frac{\Delta t}{2}\frac{\underline{f}(t)}{m} \\ \underline{v}(t+\frac{1}{2}\Delta t-\epsilon) &\leftarrow \exp(-\chi\;\Delta t)\;\underline{v}(t+\epsilon) + \frac{\sqrt{2\;\chi\;m\;k_B T}}{m}\;\underline{Z}_1(\chi,\Delta t) \\ \underline{r}(t+\Delta t) &\leftarrow \underline{r}(t) + \frac{1-\exp(-\chi\;\Delta t)}{\chi}\;\underline{v}(t+\epsilon) + \frac{\sqrt{2\;\chi\;m\;k_B T}}{\chi\;m}\;\underline{Z}_2(\chi,\Delta t)\quad, \end{aligned} \tag{3.36}$$

   where $\underline{Z}_1(\chi,\Delta t)$ and $\underline{Z}_2(\chi,\Delta t)$ are joint Gaussian random variables of zero mean, sampling from a *bivariate* Gaussian distribution [76]:

$$\begin{bmatrix} \underline{Z}_1 \\ \underline{Z}_2 \end{bmatrix} = \begin{bmatrix} \sigma_2^{1/2} & 0 \\ (\sigma_1-\sigma_2)\sigma_2^{-1/2} & (\Delta t - \sigma_1^2\sigma_2^{-1})^{1/2} \end{bmatrix} \begin{bmatrix} \underline{R}_1 \\ \underline{R}_2 \end{bmatrix} \tag{3.37}$$

   with

$$\sigma_k = \frac{1-\exp(-k\;\chi\;\Delta t)}{k\;\chi}\quad,\quad k=1,2 \tag{3.38}$$

   and $\underline{R}_k$ vectors of independent standard Gaussian random numbers of zero mean and unit variance, `Gauss`$(0,1)$, - easily related to the Langevin random forces as defined in equation (3.35).

2. RATTLE_VV1

3. FF:

$$\underline{f}(t+\Delta t) \leftarrow \underline{f}(t) \tag{3.39}$$

4. VV2:

$$\underline{v}(t+\Delta t) \leftarrow \underline{v}(t+\frac{1}{2}\Delta t-\epsilon) + \frac{\Delta t}{2}\frac{\underline{f}(t+\Delta t)}{m} \tag{3.40}$$

5. RATTLE_VV2 .

The algorithm is self-consistent and requires no iterations. It is worth noting that the integration is conditional upon the Cholesky factorisation which is impossible when $\Delta t < \sigma_1^2/\sigma_2$. That is why a safety check is put in place which when failed iteratively increases the integration timestep to a safe value with respect to the above condition.

The LFV implementation of the Langevin thermostat is straightforward:

1. FF:

$$\begin{aligned} \underline{f}(t) &\leftarrow \underline{f}(t-\Delta t) \\ \underline{R}(t) &\leftarrow \underline{R}(t-\Delta t) \end{aligned} \tag{3.41}$$

$$\tag{3.42}$$

2. LFV and Thermostat:

$$\begin{aligned}
\texttt{scale} = 1 + \chi \, \frac{\Delta t}{2} \quad &, \quad \texttt{scale\_v} = \frac{2}{\texttt{scale}} - 1 \ , \ \texttt{scale\_f} = \frac{\Delta t}{\texttt{scale}} \\
\underline{v}(t + \frac{1}{2}\Delta t) \quad &\leftarrow \quad \texttt{scale\_v} \ \underline{v}(t - \frac{1}{2}\Delta t) + \texttt{scale\_f} \ \frac{\underline{f}(t) + \underline{R}(t)}{m} \\
\underline{r}(t + \Delta t) \quad &\leftarrow \quad \underline{r}(t) + \Delta t \ \underline{v}(t + \frac{1}{2}\Delta t) \quad ,
\end{aligned} \tag{3.43}$$

where $\underline{R}(t)$ are the Langevin random forces as defined in equation (3.35).

3. SHAKE

4. Full step velocity:

$$\underline{v}(t) \leftarrow \frac{1}{2}\left[\underline{v}(t - \frac{1}{2}\Delta t) + \underline{v}(t + \frac{1}{2}\Delta t)\right] \quad . \tag{3.44}$$

Note that by the nature of the ensemble the centre of mass will not be stationary although the ensemble average warrants its proximity to the its original position, i.e. the COM momentum accumulation ensemble average will tend towards zero. By default this accumulation is removed and thus the correct application of stochastic dynamics the user is advised to use in the **no vom** option in the CONTROL file (see Section 6.1.1). If the option is not applied then the dynamics will lead to peculiar thermalisation of different atomic species to mass- and system size-dependent temperatures.

The VV and LFV flavours of the Langevin thermostat are implemented in the DL_POLY_4 routines NVT_L0_VV and NVT_L0_LFV respectively. The routines NVT_L1_VV and NVT_L1_LFV implement the same but also incorporate RB dynamics.

3.4.3 Andersen Thermostat

This thermostat assumes the idea that the system, or some subset of the system, has an instantaneous interaction with some fictional particles and exchanges energy. Practically, this interaction amounts to replacing the momentum of some atoms with a new momentum drawn from the correct Boltzmann distribution at the desired temperature. The strength of the thermostat can be adjusted by setting the average time interval over which the interactions occur, and by setting the magnitude of the interaction. The collisions are best described as a random (Poisson) process so that the probability that a collision occurs in a time step Δt is

$$P_{\texttt{collision}}(t) = 1 - \exp\left(-\frac{\Delta t}{\tau_T}\right) \quad , \tag{3.45}$$

where τ_T is the thermostat relaxation time. The hardest collision is to completely reset the momentum of the Poisson selected atoms in the system, with a new one selected from the Boltzmann distribution

$$F(\underline{v}_i) = \sqrt{\left(\frac{m_i}{2\pi k_B T_{\texttt{ext}}}\right)^3} \exp\left(-\frac{m_i \ \underline{v}_i^2}{2k_B T_{\texttt{ext}}}\right) = \sqrt{\frac{k_B T_{\texttt{ext}}}{2m_i}} \ \texttt{Gauss}(0,1) \quad . \tag{3.46}$$

where subscripts denote particle indices, k_B is the Boltzmann constant, $T_{\texttt{ext}}$ the target temperature and m_i the particle's mass. The thermostat can be made softer by mixing the new momentum $\underline{v}_i^{\texttt{new}}$ drawn from $F(\underline{v}_i)$ with the old momentum $\underline{v}_i^{\texttt{old}}$

$$\underline{v}_i = \alpha \ \underline{v}_i^{\texttt{old}} + \sqrt{1 - \alpha^2} \ \underline{v}_i^{\texttt{new}} \quad , \tag{3.47}$$

where α $(0 \leq \alpha \leq 1)$ is the softness of the thermostat. In practice, a uniform distribution random number, $\texttt{uni}(i)$, is generated for each particle in the system, which is compared to the collision probability. If $\texttt{uni}(i) \leq 1 - \exp\left(-\frac{\Delta t}{\tau_T}\right)$ the particle momentum is changed as described above.

The VV implementation of the Andersen algorithm is as follows:

1. VV1:

$$\begin{aligned}\underline{v}(t+\frac{1}{2}\Delta t) &\leftarrow \underline{v}(t)+\frac{\Delta t}{2}\,\frac{\underline{f}(t)}{m} \\ \underline{r}(t+\Delta t) &\leftarrow \underline{r}(t)+\Delta t\;\underline{v}(t+\frac{1}{2}\Delta t)\end{aligned} \tag{3.48}$$

2. RATTLE_VV1
3. FF:

$$\underline{f}(t+\Delta t) \leftarrow \underline{f}(t) \tag{3.49}$$

4. VV2:

$$\underline{v}(t+\Delta t) \leftarrow \underline{v}(t+\frac{1}{2}\Delta t)+\frac{\Delta t}{2}\left[\frac{\underline{f}(t+\Delta t)}{m}\right] \tag{3.50}$$

5. RATTLE_VV2
6. Thermostat: Note that the MD cell centre of mass momentum must not change!

$$\begin{aligned}\texttt{If} \quad & \left(\texttt{uni}(i) \le 1-\exp\left(-\frac{\Delta t}{\tau_T}\right)\right)\ \texttt{Then} \\ \underline{v}_i^{\texttt{new}}(t+\Delta t) &\leftarrow \sqrt{\frac{k_B T}{2m_i}}\ \texttt{Gauss}(0,1) \\ \underline{v}_i(t+\Delta t) &\leftarrow \alpha\;\underline{v}_i(t+\Delta t)+\sqrt{1-\alpha^2}\;\underline{v}_i^{\texttt{new}}(t+\Delta t) \\ \texttt{End} \quad & \texttt{If}\ .\end{aligned} \tag{3.51}$$

The algorithm is self-consistent and requires no iterations.

The LFV implementation of the Andersen algorithm is as follows:

1. FF:

$$\underline{f}(t) \leftarrow \underline{f}(t-\Delta t) \tag{3.52}$$

2. LFV:

$$\begin{aligned}\underline{v}(t+\frac{1}{2}\Delta t) &\leftarrow \underline{v}(t-\frac{1}{2}\Delta t)+\Delta t\;\frac{\underline{f}(t)}{m} \\ \underline{r}(t+\Delta t) &\leftarrow \underline{r}(t)+\Delta t\;\underline{v}(t+\frac{1}{2}\Delta t)\end{aligned} \tag{3.53}$$

3. Full step velocity:

$$\underline{v}(t) \leftarrow \frac{1}{2}\left[\underline{v}(t-\frac{1}{2}\Delta t)+\underline{v}(t+\frac{1}{2}\Delta t)\right] \tag{3.54}$$

4. Thermostat: Note that the MD cell centre of mass momentum must not change!

$$\begin{aligned}\texttt{If} \quad & \left(\texttt{uni}(i) \le 1-\exp\left(-\frac{\Delta t}{\tau_T}\right)\right)\ \texttt{Then} \\ \underline{v}_i^{\texttt{new}}(t+\frac{1}{2}\Delta t) &\leftarrow \sqrt{\frac{k_B T}{2m_i}}\ \texttt{Gauss}(0,1) \\ \underline{v}_i(t+\frac{1}{2}\Delta t) &\leftarrow \alpha\;\underline{v}_i(t+\frac{1}{2}\Delta t)+\sqrt{1-\alpha^2}\;\underline{v}_i^{\texttt{new}}(t+\frac{1}{2}\Delta t) \\ \underline{v}_i(t) &\leftarrow \underline{v}_i(t+\frac{1}{2}\Delta t) \\ \texttt{End} \quad & \texttt{If}\ .\end{aligned} \tag{3.55}$$

5. SHAKE

The algorithm is self-consistent and requires no iterations.

The VV and LFV flavours of the Andersen thermostat are implemented in the DL_POLY_4 routines NVT_A0_VV and NVT_A0_LFV respectively. The routines NVT_A1_VV and NVT_A1_LFV implement the same but also incorporate RB dynamics.

3.4.4 Berendsen Thermostat

In the Berendsen algorithm the instantaneous temperature is pushed towards the desired temperature $T_{\rm ext}$ by scaling the velocities at each step by

$$\chi(t) = \left[1 + \frac{\Delta t}{\tau_T}\left(\frac{\sigma}{E_{kin}(t)} - 1\right)\right]^{1/2} \quad , \tag{3.56}$$

where

$$\sigma = \frac{f}{2}\; k_B\; T_{\rm ext} \tag{3.57}$$

is the target thermostat energy (depending on the external temperature and the system total degrees of freedom, f - equation (3.11)) and τ_T a specified time constant for temperature fluctuations (normally in the range [0.5, 2] ps).

The VV implementation of the Berendsen algorithm is straight forward. A conventional VV1 and VV2 (thermally unconstrained) steps are carried out. At the end of VV2 velocities are scaled by a factor of χ in the following manner

1. VV1:
$$\begin{aligned} \underline{v}(t+\frac{1}{2}\Delta t) &\leftarrow \underline{v}(t) + \frac{\Delta t}{2}\,\frac{\underline{f}(t)}{m} \\ \underline{r}(t+\Delta t) &\leftarrow \underline{r}(t) + \Delta t\;\underline{v}(t+\frac{1}{2}\Delta t) \end{aligned} \tag{3.58}$$
2. RATTLE_VV1
3. FF:
$$\underline{f}(t+\Delta t) \leftarrow \underline{f}(t) \tag{3.59}$$
4. VV2:
$$\underline{v}(t+\Delta t) \leftarrow \underline{v}(t+\frac{1}{2}\Delta t) + \frac{\Delta t}{2}\,\frac{\underline{f}(t+\Delta t)}{m} \tag{3.60}$$
5. RATTLE_VV2
6. Thermostat:
$$\begin{aligned} \chi(t+\Delta t) &\leftarrow \left[1 + \frac{\Delta t}{\tau_T}\left(\frac{\sigma}{E_{kin}(t+\Delta t)} - 1\right)\right]^{1/2} \\ \underline{v}(t+\Delta t) &\leftarrow \underline{v}(t+\Delta t)\;\chi \quad . \end{aligned} \tag{3.61}$$

The LFV implementation of the Berendsen algorithm is iterative as an initial estimate of $\chi(t)$ at full step is calculated using an unconstrained estimate of the velocity at full step, $\underline{v}(t)$.

1. FF:
$$\underline{f}(t) \leftarrow \underline{f}(t-\Delta t) \tag{3.62}$$

2. LFV: The iterative part is as follows:

$$\begin{aligned} \underline{v}(t+\frac{1}{2}\Delta t) &\leftarrow \left[\underline{v}(t-\frac{1}{2}\Delta t)+\Delta t\;\frac{\underline{f}(t)}{m}\right]\;\chi(t) \\ \underline{r}(t+\Delta t) &\leftarrow \underline{r}(t)+\Delta t\;\underline{v}(t+\frac{1}{2}\Delta t) \end{aligned} \tag{3.63}$$

3. SHAKE

4. Full step velocity:

$$\underline{v}(t) \leftarrow \frac{1}{2}\left[\underline{v}(t-\frac{1}{2}\Delta t)+\underline{v}(t+\frac{1}{2}\Delta t)\right] \tag{3.64}$$

5. Thermostat:

$$\chi(t) \leftarrow \left[1+\frac{\Delta t}{\tau_T}\left(\frac{\sigma}{E_{kin}(t)}-1\right)\right]^{1/2}\;. \tag{3.65}$$

Several iterations are required to obtain self consistency. In DL_POLY_4 the number of iterations is set to 3 (4 if bond constraints are present).

Note that the MD cell's centre of mass momentum is removed at the end of the integration algorithms.

The Berendsen algorithms conserve total momentum but not energy.

The VV and LFV flavours of the Berendsen thermostat are implemented in the DL_POLY_4 routines NVT_B0_VV and NVT_B0_LFV respectively. The routines NVT_B1_VV and NVT_B1_LFV implement the same but also incorporate RB dynamics.

3.4.5 Nosé-Hoover Thermostat

In the Nosé-Hoover algorithm [30] Newton's equations of motion are modified to read:

$$\begin{aligned} \frac{d\underline{r}(t)}{dt} &= \underline{v}(t) \\ \frac{d\underline{v}(t)}{dt} &= \frac{\underline{f}(t)}{m}-\chi(t)\;\underline{v}(t) \end{aligned} \tag{3.66}$$

The friction coefficient, χ, is controlled by the first order differential equation

$$\frac{d\chi(t)}{dt}=\frac{2E_{kin}(t)-2\sigma}{q_{mass}} \tag{3.67}$$

where σ is the target thermostat energy, equation (3.57), and

$$q_{mass}=2\;\sigma\;\tau_T^2 \tag{3.68}$$

is the thermostat mass, which depends on a specified time constant τ_T (for temperature fluctuations normally in the range [0.5, 2] ps).

The VV implementation of the Nosé-Hoover algorithm takes place in a symplectic manner as follows:

1. Thermostat: Note $E_{kin}(t)$ changes inside

$$\begin{aligned} \chi(t+\frac{1}{4}\Delta t) &\leftarrow \chi(t)+\frac{\Delta t}{4}\;\frac{2E_{kin}(t)-2\sigma}{q_{mass}} \\ \underline{v}(t) &\leftarrow \underline{v}(t)\;\exp\left(-\chi(t+\frac{1}{4}\Delta t)\;\frac{\Delta t}{2}\right) \qquad (3.69) \\ \chi(t+\frac{1}{2}\Delta t) &\leftarrow \chi(t+\frac{1}{4}\Delta t)+\frac{\Delta t}{4}\;\frac{2E_{kin}(t)-2\sigma}{q_{mass}} \end{aligned} \tag{3.70}$$

2. VV1:

$$\begin{aligned} \underline{v}(t+\frac{1}{2}\Delta t) &\leftarrow \underline{v}(t)+\frac{\Delta t}{2}\,\frac{\underline{f}(t)}{m} \\ \underline{r}(t+\Delta t) &\leftarrow \underline{r}(t)+\Delta t\,\underline{v}(t+\frac{1}{2}\Delta t) \end{aligned} \tag{3.71}$$

3. RATTLE_VV1

4. FF:

$$\underline{f}(t+\Delta t) \leftarrow \underline{f}(t) \tag{3.72}$$

5. VV2:

$$\underline{v}(t+\Delta t) \leftarrow \underline{v}(t+\frac{1}{2}\Delta t)+\frac{\Delta t}{2}\,\frac{\underline{f}(t+\Delta t)}{m} \tag{3.73}$$

6. RATTLE_VV2

7. Thermostat: Note $E_{kin}(t+\Delta t)$ changes inside

$$\begin{aligned} \chi(t+\frac{3}{4}\Delta t) &\leftarrow \chi(t+\frac{1}{2}\Delta t)+\frac{\Delta t}{4}\,\frac{2E_{kin}(t+\Delta t)-2\sigma}{q_{mass}} \\ \underline{v}(t+\Delta t) &\leftarrow \underline{v}(t+\Delta t)\,\exp\left(-\chi(t+\frac{3}{4}\Delta t)\,\frac{\Delta t}{2}\right) \\ \chi(t+\Delta t) &\leftarrow \chi(t+\frac{3}{4}\Delta t)+\frac{\Delta t}{4}\,\frac{2E_{kin}(t+\Delta t)-2\sigma}{q_{mass}} \quad . \end{aligned} \tag{3.74}$$

The algorithm is self-consistent and requires no iterations.

The LFV implementation of the Nosé-Hoover algorithm is iterative as an initial estimate of $\chi(t)$ at full step is calculated using an unconstrained estimate of the velocity at full step, $\underline{v}(t)$.

1. FF:

$$\underline{f}(t) \leftarrow \underline{f}(t-\Delta t) \tag{3.75}$$

2. LFV: The iterative part is as follows:

$$\begin{aligned} \underline{v}(t+\frac{1}{2}\Delta t) &\leftarrow \underline{v}(t-\frac{1}{2}\Delta t)+\Delta t\,\left[\frac{\underline{f}(t)}{m}-\chi(t)\,\underline{v}(t)\right] \\ \underline{r}(t+\Delta t) &\leftarrow \underline{r}(t)+\Delta t\,\underline{v}(t+\frac{1}{2}\Delta t) \end{aligned} \tag{3.76}$$

3. SHAKE

4. Full step velocity:

$$\underline{v}(t) \leftarrow \frac{1}{2}\left[\underline{v}(t-\frac{1}{2}\Delta t)+\underline{v}(t+\frac{1}{2}\Delta t)\right] \tag{3.77}$$

5. Thermostat:

$$\begin{aligned} \chi(t+\frac{1}{2}\Delta t) &\leftarrow \chi(t-\frac{1}{2}\Delta t)+\Delta t\,\frac{2E_{kin}(t)-2\sigma}{q_{mass}} \\ \chi(t) &\leftarrow \frac{1}{2}\left[\chi(t-\frac{1}{2}\Delta t)+\chi(t+\frac{1}{2}\Delta t)\right] \quad . \end{aligned} \tag{3.78}$$

Several iterations are required to obtain self consistency. In DL_POLY_4 the number of iterations is set to 2 (3 if bond constraints are present).

The conserved quantity is derived from the extended Hamiltonian for the system which, to within a constant, is the Helmholtz free energy:

$$\mathcal{H}_{\mathrm{NVT}} = \mathcal{H}_{\mathrm{NVE}} + \frac{q_{mass}\ \chi(t)^2}{2} + f\ k_B\ T_{\mathrm{ext}} \int_o^t \chi(s) ds \quad , \tag{3.79}$$

where f is the system's degrees of freedom - equation (3.11).

The VV and LFV flavours of the Nosé-Hoover thermostat are implemented in the DL_POLY_4 routines NVT_H0_VV and NVT_H0_LFV respectively. The routines NVT_H1_VV and NVT_H1_LFV implement the same but also incorporate RB dynamics.

3.4.6 Gentle Stochastic Thermostat

The Gentle Stochastic Thermostat [72, 77] is an extension of the Nosé-Hoover algorithm [30]

$$\begin{aligned} \frac{d\underline{r}(t)}{dt} &= \underline{v}(t) \\ \frac{d\underline{v}(t)}{dt} &= \frac{\underline{f}(t)}{m} - \chi(t)\ \underline{v}(t) \end{aligned} \tag{3.80}$$

in which the thermostat friction, χ, has its own Brownian dynamics:

$$\frac{d\chi(t)}{dt} = \frac{2E_{kin}(t) - 2\sigma}{q_{mass}} - \gamma\ \chi(t) + \frac{\sqrt{2\ \gamma\ k_B\ T_{\mathrm{ext}}\ q_{mass}}}{q_{mass}} \frac{d\omega(t)}{dt} \quad , \tag{3.81}$$

governed by the Langevin friction γ (positive, in units of ps^{-1}), where $\omega(t)$ is the standard Brownian motion (Wiener process - `Gauss`(0,1)), σ is the target thermostat energy, as in equation (3.57).

$$q_{mass} = 2\ \sigma\ \tau_T^2 \tag{3.82}$$

is the thermostat mass, which depends on a specified time constant τ_T (for temperature fluctuations normally in the range [0.5, 2] ps).

It is worth noting that equation (3.81) similar to the Ornstein-Uhlenbeck equation:

$$\frac{d\chi}{dt} = -\frac{\alpha\sigma^2}{2}\chi + \sigma \frac{d\omega}{dt} \quad , \tag{3.83}$$

which for a given realization of the Wiener process $\omega(t)$ has an exact solution:

$$\chi_{n+1} = e^{-\epsilon t} \left(\chi_n + \sigma \sqrt{\frac{e^{2\ \epsilon t} - 1}{2\epsilon}} \Delta\omega \right) \quad , \tag{3.84}$$

where $\epsilon = \alpha\sigma^2/2$ and $\Delta\omega \sim \mathcal{N}(0,1)$. The VV implementation of the Gentle Stochastic Thermostat algorithm takes place in a symplectic manner as follows:

1. Thermostat: Note $E_{kin}(t)$ changes inside and $R_g(t)$, drown from `Gauss`$(0,1)$, is carried over from the previous half-timestep

$$\begin{aligned} \chi(t + \frac{1}{4}\Delta t) \quad \leftarrow \quad & \chi(t)\ \exp\left[-\gamma\ \frac{\Delta t}{4}\right] + \sqrt{\frac{k_B\ T_{\mathrm{ext}}}{q_{mass}} \left(1 - \exp^2\left[-\gamma\ \frac{\Delta t}{4}\right]\right)}\ R_g(t) \\ & + \frac{\Delta t}{4}\ \frac{2E_{kin}(t) - 2\sigma}{q_{mass}} \end{aligned}$$

$$\begin{aligned}
\underline{v}(t) &\leftarrow \underline{v}(t)\ \exp\left(-\chi(t+\frac{1}{4}\Delta t)\ \frac{\Delta t}{2}\right) \\
\chi(t+\frac{1}{2}\Delta t) &\leftarrow \chi(t+\frac{1}{4}\Delta t)\ \exp\left[-\gamma\ \frac{\Delta t}{4}\right] + \sqrt{\frac{k_B\ T_{\text{ext}}}{q_{mass}}\left(1-\exp^2\left[-\gamma\ \frac{\Delta t}{4}\right]\right)}\ R_g(t) \\
&\quad + \frac{\Delta t}{4}\ \frac{2E_{kin}(t)-2\sigma}{q_{mass}}
\end{aligned} \tag{3.85}$$

2. VV1:

$$\begin{aligned}
\underline{v}(t+\frac{1}{2}\Delta t) &\leftarrow \underline{v}(t) + \frac{\Delta t}{2}\ \frac{\underline{f}(t)}{m} \\
\underline{r}(t+\Delta t) &\leftarrow \underline{r}(t) + \Delta t\ \underline{v}(t+\frac{1}{2}\Delta t)
\end{aligned} \tag{3.86}$$

3. RATTLE_VV1

4. FF:

$$\underline{f}(t+\Delta t) \leftarrow \underline{f}(t) \tag{3.87}$$

5. VV2:

$$\underline{v}(t+\Delta t) \leftarrow \underline{v}(t+\frac{1}{2}\Delta t) + \frac{\Delta t}{2}\ \frac{\underline{f}(t+\Delta t)}{m} \tag{3.88}$$

6. RATTLE_VV2

7. Thermostat: Note $E_{kin}(t+\Delta t)$ changes inside and $R_g(t+\Delta t)$ is drown anew from `Gauss`$(0,1)$, and is to be carried over the next half-timestep

$$\begin{aligned}
\chi(t+\frac{3}{4}\Delta t) &\leftarrow \chi(t+\frac{1}{2}\Delta t)\ \exp\left[-\gamma\ \frac{\Delta t}{4}\right] + \sqrt{\frac{k_B\ T_{\text{ext}}}{q_{mass}}\left(1-\exp^2\left[-\gamma\ \frac{\Delta t}{4}\right]\right)}\ R_g(t+\Delta t) \\
&\quad + \frac{\Delta t}{4}\ \frac{2E_{kin}(t)-2\sigma}{q_{mass}} \\
\underline{v}(t+\Delta t) &\leftarrow \underline{v}(t+\Delta t)\ \exp\left(-\chi(t+\frac{3}{4}\Delta t)\ \frac{\Delta t}{2}\right) \\
\chi(t+\Delta t) &\leftarrow \chi(t+\frac{3}{4}\Delta t)\ \exp\left[-\gamma\ \frac{\Delta t}{4}\right] + \sqrt{\frac{k_B\ T_{\text{ext}}}{q_{mass}}\left(1-\exp^2\left[-\gamma\ \frac{\Delta t}{4}\right]\right)}\ R_g(t+\Delta t) \\
&\quad + \frac{\Delta t}{4}\ \frac{2E_{kin}(t)-2\sigma}{q_{mass}}
\end{aligned} \tag{3.89}$$

The algorithm is self-consistent and requires no iterations.

The LFV implementation of the Gentle Stochastic Thermostat algorithm is iterative as an initial estimate of $\chi(t)$ at full step is calculated using an unconstrained estimate of the velocity at full step, $\underline{v}(t)$.

1. FF:

$$\underline{f}(t) \leftarrow \underline{f}(t-\Delta t) \tag{3.90}$$

2. LFV: The iterative part is as follows:

$$\begin{aligned}
\underline{v}(t+\frac{1}{2}\Delta t) &\leftarrow \underline{v}(t-\frac{1}{2}\Delta t) + \Delta t\ \left[\frac{\underline{f}(t)}{m} - \chi(t)\ \underline{v}(t)\right] \\
\underline{r}(t+\Delta t) &\leftarrow \underline{r}(t) + \Delta t\ \underline{v}(t+\frac{1}{2}\Delta t)
\end{aligned} \tag{3.91}$$

3. SHAKE

4. Full step velocity:

$$\underline{v}(t) \leftarrow \frac{1}{2}\left[\underline{v}(t-\frac{1}{2}\Delta t)+\underline{v}(t+\frac{1}{2}\Delta t)\right] \tag{3.92}$$

5. Thermostat: Note $R_g(t)$ is drown from `Gauss`$(0,1)$ just once per timestep

$$\begin{aligned}\chi(t+\frac{1}{2}\Delta t) &\leftarrow \chi(t-\frac{1}{2}\Delta t)\ \exp\left[-\gamma\ \frac{\Delta t}{4}\right]+\sqrt{\frac{k_B\ T_{\rm ext}}{q_{mass}}\left(1-\exp^2\left[-\gamma\ \frac{\Delta t}{4}\right]\right)}\ R_g(t) \\ &\quad + \Delta t\ \frac{2E_{kin}(t)-2\sigma}{q_{mass}} \\ \chi(t) &\leftarrow \frac{1}{2}\left[\chi(t-\frac{1}{2}\Delta t)+\chi(t+\frac{1}{2}\Delta t)\right]\ .\end{aligned} \tag{3.93}$$

Several iterations are required to obtain self consistency. In DL_POLY_4 the number of iterations is set to 3 (4 if bond constraints are present).

The conserved quantity is derived from the extended Hamiltonian for the system which, to within a constant, is the Helmholtz free energy:

$$\mathcal{H}_{\rm NVT}=\mathcal{H}_{\rm NVE}+\frac{q_{mass}\ \chi(t)^2}{2}+f\ k_B\ T_{\rm ext}\int_o^t \chi(s)ds\ \ , \tag{3.94}$$

where f is the system's degrees of freedom - equation (3.11).

The VV and LFV flavours of the Gentle Stochastic Thermostat are implemented in the DL_POLY_4 routines NVT_G0_VV and NVT_G0_LFV respectively. The routines NVT_G1_VV and NVT_G1_LFV implement the same but also incorporate RB dynamics.

3.4.7 Dissipative Particle Dynamics Thermostat

An elegant way to integrate the DPD equations of motions, as shown in Appendix A, is introduced by Shardlow [71]. By applying ideas commonly used in solving differential equations to the case of integrating the equations of motion in DPD, the integration process is factorised by splitting the conservative forces calculation from that of the dissipative and random terms. In this way the conservative part can be solved using traditional molecular dynamics methods, while the fluctuation-dissipation part is solved separately as a stochastic differential (Langevin) equation. There are two Shardlow integrators, called S1 (**dpds1**) and S2 (**dpds2**), based on splitting the equations of motion up to first and second order, respectively, using Suzuki-Trotter(Strang) expansion of the Lioville evolution operator and thus warranting the integrators' symplectic.

To describe the two integrators we define the algorithmic sequence

$$\mathrm{S}(\Delta t) = \left\{\begin{array}{l} \text{For all pairs of particles for which } r_{ij}<r_c: \\ (1): \quad \underline{v}_i \leftarrow \underline{v}_i-\frac{1}{2m_i}\left\{\gamma_{ij}w^2_{(r_{ij})}(\underline{v}_{ij}\cdot\underline{e}_{ij})\underline{e}_{ij}\Delta t+\sqrt{2\gamma_{ij}k_BT}w_{(r_{ij})}\zeta_{ij}\underline{e}_{ij}\sqrt{\Delta t}\right\} \\ (2): \quad \underline{v}_j \leftarrow \underline{v}_j+\frac{1}{2m_i}\left\{\gamma_{ij}w^2_{(r_{ij})}(\underline{v}_{ij}\cdot\underline{e}_{ij})\underline{e}_{ij}\Delta t-\sqrt{2\gamma_{ij}k_BT}w_{(r_{ij})}\zeta_{ij}\underline{e}_{ij}\sqrt{\Delta t}\right\} \\ (3): \quad \underline{v}_i \leftarrow \underline{v}_i+\frac{1}{2m_i}\left\{\sqrt{2\gamma_{ij}k_BT}w_{(r_{ij})}\zeta_{ij}\underline{e}_{ij}\sqrt{\Delta t}-\frac{\gamma_{ij}w^2(r_{ij})\Delta t}{1+\gamma_{ij}w^2(r_{ij})\Delta t}\times\right. \\ \qquad\qquad\qquad \left.\left[(\underline{v}_{ij}\cdot\underline{e}_{ij})\underline{e}_{ij}+\sqrt{2\gamma_{ij}k_BT}w_{(r_{ij})}\zeta_{ij}\underline{e}_{ij}\sqrt{\Delta t}\right]\right\} \\ (4): \quad \underline{v}_j \leftarrow \underline{v}_j-\frac{1}{2m_i}\left\{\sqrt{2\gamma_{ij}k_BT}w_{(r_{ij})}\zeta_{ij}\underline{e}_{ij}\sqrt{\Delta t}+\frac{\gamma_{ij}w^2(r_{ij})\Delta t}{1+\gamma_{ij}w^2(r_{ij})\Delta t}\times\right. \\ \qquad\qquad\qquad \left.\left[(\underline{v}_{ij}\cdot\underline{e}_{ij})\underline{e}_{ij}+\sqrt{2\gamma_{ij}k_BT}w_{(r_{ij})}\zeta_{ij}\underline{e}_{ij}\sqrt{\Delta t}\right]\right\} \end{array}\right. \tag{3.95}$$

as the Shardlow operator, where ζ_{ij} is the random number with zero mean and unit variance, unique for every unique pair $\{ij\}$ in the system, and $w_{(r_{ij})}=w^C(r_{ij})$ is the DPD conservative force switching function in equation (A.4), γ_{ij} is the drag coefficient between the types of particles i and j, $\underline{v}_{ij} = \underline{v}_j - \underline{v}_i$ is the inter-particle relative velocity and $\underline{e}_{ij} = \underline{r}_{ij}/r_{ij}$ is the inter-particle unit vector. k_B and T are the Boltzmann constant and the system target temperature.

If we define the velocity Verlet microcsannonical (NVE) ensemble sequence as NVE$(t+\Delta t)$ then Shardlow's first ($\mathcal{S}1$) and second ($\mathcal{S}2$) order splittings can be written algorithmically as the following sequential operators applications:

$$\begin{aligned} \mathcal{S}1 &:: \ \mathrm{S}(\ \Delta\mathrm{t}\ ) \rightarrow \mathrm{NVE}(t+\Delta t) \\ \mathcal{S}2 &:: \ \mathrm{S}(\Delta t/2) \rightarrow \mathrm{NVE}(t+\Delta t) \rightarrow \mathrm{S}(\Delta t/2) \ . \end{aligned} \tag{3.96}$$

The application of these DPD thermostats are implemented in the DL_POLY_4 routine DPD_THERMOSTAT and which only applies as a perturbation around the NVE integrator incorporating both particle and RB dynamics.

3.5 Barostats

The size and shape of the simulation cell may be dynamically adjusted by coupling the system to a barostat in order to obtain a desired average pressure ($P_{\rm ext}$) and/or isotropic stress tensor ($\underline{\underline{\sigma}}$). DL_POLY_4 has four such algorithms: the Langevin type barostat [31], the Berendsen barostat [29], the Nosé-Hoover type barostat [30] and the Martyna-Tuckerman-Klein (MTK) barsotat [32]. Only the Berendsen barostat does not have defined conserved quantity.

Note that the MD cell's centre of mass momentum is removed at the end of the integration algorithms with barostats.

3.5.1 Instantaneous pressure and stress

The instantaneous pressure in a system,

$$\mathcal{P}(t) = \frac{[2E_{kin}(t) - \mathcal{W}_{\rm atomic}(t) - \mathcal{W}_{\rm constrain}(t-\Delta t) - \mathcal{W}_{\rm PMF}(t-\Delta t)]}{3V(t)} \ , \tag{3.97}$$

is a function of the system volume, kinetic energy and virial, $\mathcal{W}$.
Note that when bond constraints or/and PMF constraints are present in the system $\mathcal{P}$ will not converge to the exact value of $P_{\rm ext}$ during equilibration in NPT and NσT simulations. This is due to iterative nature of the constrained motion, in which the virials $\mathcal{W}_{\rm constrain}$ and $\mathcal{W}_{\rm PMF}$ are calculated retrospectively to the forcefield virial $\mathcal{W}_{\rm atomic}$.

The instantaneous stress tensor in a system,

$$\underline{\underline{\sigma}}(t) = \underline{\underline{\sigma}}_{kin}(t) + \underline{\underline{\sigma}}_{\rm atomic}(t) + \underline{\underline{\sigma}}_{\rm constrain}(t-\Delta t) + \underline{\underline{\sigma}}_{\rm PMF}(t-\Delta t) \ , \tag{3.98}$$

is a sum of the forcefield, $\underline{\underline{\sigma}}_{\rm atomic}$, constrain, $\underline{\underline{\sigma}}_{\rm constrains}$, and PMF, $\underline{\underline{\sigma}}_{\rm PMF}$, stresses.
Note that when bond constraints or/and PMF constraints are present in the system, the quantity $\frac{\mathtt{Tr}[\underline{\underline{\sigma}}]}{3V}$ will not converge to the exact value of $P_{\rm ext}$ during equilibration in NPT and NσT simulations. This is due to iterative nature of the constrained motion in which the constraint and PMF stresses are calculated retrospectively to the forcefield stress.

3.5.2 Langevin Barostat

DL_POLY_4 implements a Langevin barostat [31] for isotropic and anisotropic cell fluctuations.

Cell size variations

For isotropic fluctuations the equations of motion are:

$$\begin{aligned}
\frac{d}{dt}\underline{r}(t) &= \underline{v}(t) + \eta(t)\ \underline{r}(t) \\
\frac{d}{dt}\underline{v}(t) &= \frac{\underline{f}(t) + \underline{R}(t)}{m} - \left[\chi + \left(1 + \frac{3}{f}\right)\eta(t)\right]\underline{v}(t) \\
\frac{d}{dt}\eta(t) &= 3V(t)\frac{\mathcal{P}(t) - P_{\rm ext}}{p_{mass}} + 3\frac{2E_{kin}(t)}{f}\frac{1}{p_{mass}} - \chi_p\ \eta(t) + \frac{R_p}{p_{mass}} \\
p_{mass} &= \frac{(f+3)\ k_B\ T_{\rm ext}}{(2\pi\ \chi_p)^2} \\
\frac{d}{dt}\underline{\underline{\mathbf{H}}}(t) &= \eta(t)\ \underline{\underline{\mathbf{H}}}(t) \\
\frac{d}{dt}V(t) &= [3\eta(t)]\ V(t)\quad ,
\end{aligned} \tag{3.99}$$

where χ and χ_p are the user defined *constants* (positive, in units of ps^{-1}), specifying the thermostat and barostat friction parameters, $R(t)$ is the Langevin stochastic force (see equation (3.35)), $\mathcal{P}$ the instantaneous pressure (equation (3.97)) and R_p is the stochastic (Langevin) pressure variable

$$\langle R_p(t)\ R_p(t')\rangle = 2\ \chi_p\ p_{mass}\ k_B T\ \delta(t - t')\quad , \tag{3.100}$$

which is drawn from Gaussian distribution of zero mean and unit variance, `Gauss`$(0,1)$, scaled by $\sqrt{\frac{2\ \chi_p\ p_{mass}\ k_B T}{\Delta t}}$. k_B is the Boltzmann constant, T the target temperature and p_{mass} the barostat mass. $\underline{\underline{\mathbf{H}}}$ is the cell matrix whose columns are the three cell vectors $\underline{a}, \underline{b}, \underline{c}$.

The conserved quantity these generate is:

$$\mathcal{H}_{\rm NPT} = \mathcal{H}_{\rm NVE} + \frac{p_{mass}\ \eta(t)^2}{2} + P_{\rm ext}V(t)\quad . \tag{3.101}$$

The VV implementation of the Langevin algorithm only requires iterations if bond or PMF constraints are present (4 until satisfactory convergence of the constraint forces is achieved). These are with respect to the pressure (i.e. $\eta(t)$) in the first part, VV1+RATTLE_VV1. The second part is conventional, VV2+RATTLE_VV2, as at the end the velocities are scaled by a factor of χ.

1. Thermostat: Note $2E_{kin}(t)$ changes inside

$$\underline{v}(t) \leftarrow \exp\left(-\chi\ \frac{\Delta t}{4}\right)\ \underline{v}(t) \tag{3.102}$$

2. Barostat: Note $E_{kin}(t)$ and $\mathcal{P}(t)$ have changed and change inside

$$\begin{aligned}
\eta(t) &\leftarrow \exp\left(-\chi_p\ \frac{\Delta t}{8}\right)\ \eta(t) \\
\eta(t + \frac{1}{4}\Delta t) &\leftarrow \eta(t) + \frac{\Delta t}{4}\ \left[3V(t)\frac{\mathcal{P}(t) - P_{\rm ext}}{p_{mass}} + \right. \\
&\qquad\qquad \left. 3\frac{2E_{kin}(t)}{f}\frac{1}{p_{mass}} + \frac{R_p(t)}{p_{mass}}\right] \\
\eta(t + \frac{1}{4}\Delta t) &\leftarrow \exp\left(-\chi_p\ \frac{\Delta t}{8}\right)\ \eta(t + \frac{1}{4}\Delta t) \\
\underline{v}(t) &\leftarrow \exp\left[-\left(1 + \frac{3}{f}\right)\eta(t + \frac{1}{4}\Delta t)\ \frac{\Delta t}{2}\right]\ \underline{v}(t) \\
\eta(t + \frac{1}{4}\Delta t) &\leftarrow \exp\left(-\chi_p\ \frac{\Delta t}{8}\right)\ \eta(t + \frac{1}{4}\Delta t)
\end{aligned} \tag{3.103}$$

$$\begin{aligned}
\eta(t+\frac{1}{2}\Delta t) &\leftarrow \eta(t+\frac{1}{4}\Delta t)+\frac{\Delta t}{4}\ \Big[3V(t)\frac{\mathcal{P}(t)-P_{\rm ext}}{p_{mass}}+\\
&\qquad 3\frac{2E_{kin}(t)}{f}\frac{1}{p_{mass}}+\frac{R_p(t)}{p_{mass}}\Big]\\
\eta(t+\frac{1}{2}\Delta t) &\leftarrow \exp\left(-\chi_p\ \frac{\Delta t}{8}\right)\ \eta(t+\frac{1}{2}\Delta t)
\end{aligned}$$

3. Thermostat: Note $E_{kin}(t)$ has changed and changes inside

$$\underline{v}(t)\leftarrow \exp\left(-\chi\ \frac{\Delta t}{4}\right)\ \underline{v}(t) \tag{3.104}$$

4. VV1:

$$\begin{aligned}
\underline{v}(t+\frac{1}{2}\Delta t) &\leftarrow \underline{v}(t)+\frac{\Delta t}{2}\ \frac{\underline{f}(t)+\underline{R}(t)}{m}\\
\underline{\underline{\mathbf{H}}}(t+\Delta t) &\leftarrow \exp\left[\eta(t+\frac{1}{2}\Delta t)\ \Delta t\right]\ \underline{\underline{\mathbf{H}}}(t)\\
V(t+\Delta t) &\leftarrow \exp\left[3\eta(t+\frac{1}{2}\Delta t)\ \Delta t\right]\ V(t)\\
\underline{r}(t+\Delta t) &\leftarrow \exp\left[\eta(t+\frac{1}{2}\Delta t)\ \Delta t\right]\ \underline{r}(t)+\Delta t\ \underline{v}(t+\frac{1}{2}\Delta t)
\end{aligned} \tag{3.105}$$

5. RATTLE_VV1

6. FF:

$$\begin{aligned}
\underline{f}(t+\Delta t) &\leftarrow \underline{f}(t)\\
\underline{R}(t+\Delta t) &\leftarrow \underline{R}(t)\\
R_p(t+\Delta t) &\leftarrow R_p(t)
\end{aligned} \tag{3.106}$$

7. VV2:

$$\underline{v}(t+\Delta t) \leftarrow \underline{v}(t+\frac{\Delta t}{2})+\frac{\Delta t}{2}\ \frac{\underline{f}(t)+\underline{R}(t)}{m} \tag{3.107}$$

8. RATTLE_VV2

9. Thermostat: Note $E_{kin}(t+\Delta t)$ has changed and changes inside

$$\underline{v}(t+\Delta t)\leftarrow \exp\left(-\chi\ \frac{\Delta t}{4}\right)\ \underline{v}(t+\Delta t) \tag{3.108}$$

10. Barostat: Note $E_{kin}(t+\Delta t)$ and $\mathcal{P}(t+\Delta t)$ have changed and change inside

$$\begin{aligned}
\eta(t+\frac{1}{2}\Delta t) &\leftarrow \exp\left(-\chi_p\ \frac{\Delta t}{8}\right)\ \eta(t+\frac{1}{2}\Delta t)\\
\eta(t+\frac{3}{4}\Delta t) &\leftarrow \eta(t+\frac{1}{2}\Delta t)+\frac{\Delta t}{4}\ \Big[3V(t+\Delta t)\frac{\mathcal{P}(t+\Delta t)-P_{\rm ext}}{p_{mass}}+\\
&\qquad 3\frac{2E_{kin}(t+\Delta t)}{f}\frac{1}{p_{mass}}+\frac{R_p(t)}{p_{mass}}\Big]\\
\eta(t+\frac{3}{4}\Delta t) &\leftarrow \exp\left(-\chi_p\ \frac{\Delta t}{8}\right)\ \eta(t+\frac{3}{4}\Delta t)\\
\underline{v}(t+\Delta t) &\leftarrow \exp\left[-\left(1+\frac{3}{f}\right)\eta(t+\frac{3}{4}\Delta t)\ \frac{\Delta t}{2}\right]\ \underline{v}(t+\Delta t)
\end{aligned} \tag{3.109}$$

$$\begin{aligned}
\eta(t+\frac{3}{4}\Delta t) &\leftarrow \exp\left(-\chi_p\,\frac{\Delta t}{8}\right)\,\eta(t+\frac{3}{4}\Delta t) \\
\eta(t+\Delta t) &\leftarrow \eta(t+\frac{3}{4}\Delta t)+\frac{\Delta t}{4}\,\left[3V(t+\Delta t)\frac{\mathcal{P}(t+\Delta t)-P_{\rm ext}}{p_{mass}}+\right. \\
&\qquad\qquad \left. 3\frac{2E_{kin}(t+\Delta t)}{f}\frac{1}{p_{mass}}+\frac{R_p(t)}{p_{mass}}\right] \\
\eta(t+\Delta t) &\leftarrow \exp\left(-\chi_p\,\frac{\Delta t}{8}\right)\,\eta(t+\Delta t)
\end{aligned}$$

11. Thermostat: Note $E_{kin}(t+\Delta t)$ has changed and changes inside

$$\underline{v}(t+\Delta t) \leftarrow \exp\left(-\chi\,\frac{\Delta t}{4}\right)\,\underline{v}(t+\Delta t)\quad , \tag{3.110}$$

The LFV implementation of the Langevin algorithm is iterative, until self consistency in the full step velocity, $\underline{v}(t)$, is obtained. Initial estimate of $\eta(t)$ at full step are calculated using an unconstrained estimate of the velocity at full step, $\underline{v}(t)$. Also calculated is an unconstrained estimate of the half step position $\underline{r}(t+\frac{1}{2}\Delta t)$.

1. FF:

$$\begin{aligned}
\underline{f}(t) &\leftarrow \underline{f}(t-\Delta t) \\
\underline{R}(t) &\leftarrow \underline{R}(t-\Delta t) \\
R_p(t) &\leftarrow R_p(t-\Delta t)
\end{aligned} \tag{3.111}$$

2. LFV: The iterative part is as follows:

$$\begin{aligned}
\texttt{scale} &= 1+\left[\chi+\left(1+\frac{3}{f}\right)\eta(t)\right]\frac{\Delta t}{2} \\
\texttt{scale\_v} &= \frac{2}{\texttt{scale}}-1 \\
\texttt{scale\_f} &= \frac{\Delta t}{\texttt{scale}} \\
\underline{v}(t+\frac{1}{2}\Delta t) &\leftarrow \texttt{scale\_v}\;\underline{v}(t-\frac{1}{2}\Delta t)+\texttt{scale\_f}\;\frac{\underline{f}(t)+\underline{R}(t)}{m} \\
\underline{r}(t+\Delta t) &\leftarrow \underline{r}(t)+\Delta t\,\left\{\underline{v}(t+\frac{1}{2}\Delta t)+\eta(t+\frac{1}{2}\Delta t)\underline{r}(t+\frac{1}{2}\Delta t)\right\} \\
\underline{\underline{\mathbf{H}}}(t+\Delta t) &\leftarrow \exp\left[\eta(t+\frac{1}{2}\Delta t)\;\Delta t\right]\;\underline{\underline{\mathbf{H}}}(t) \\
V(t+\Delta t) &\leftarrow \exp\left[3\eta(t+\frac{1}{2}\Delta t)\;\Delta t\right]\;V(t)
\end{aligned} \tag{3.112}$$

3. SHAKE

4. Full step velocity and half step position:

$$\begin{aligned}
\underline{v}(t) &\leftarrow \frac{1}{2}\left[\underline{v}(t-\frac{1}{2}\Delta t)+\underline{v}(t+\frac{1}{2}\Delta t)\right] \\
\underline{r}(t+\frac{1}{2}\Delta t) &\leftarrow \frac{\underline{r}(t)+\underline{r}(t+\Delta t)}{2}
\end{aligned} \tag{3.113}$$

5. Thermostat and Barostat:

$$\begin{aligned}
\eta(t+\frac{1}{2}\Delta t) &\leftarrow \exp\left(-\chi_p\;\Delta t\right)\;\eta(t-\frac{1}{2}\Delta t)+ \\
&\quad \Delta t\;\left[3V(t+\Delta t)\frac{\mathcal{P}(t+\Delta t)-P_{\rm ext}}{p_{mass}}+3\frac{2E_{kin}(t+\Delta t)}{f}\frac{1}{p_{mass}}\right] \\
\eta(t) &\leftarrow \frac{1}{2}\left[\eta(t-\frac{1}{2}\Delta t)+\eta(t+\frac{1}{2}\Delta t)\right]\quad .
\end{aligned} \tag{3.114}$$

Several iterations are required to obtain self consistency. In DL_POLY_4 the number of iterations is set to 7 (8 if bond constraints are present). Note also that the change in box size requires the SHAKE algorithm to be called each iteration.

The VV and LFV flavours of the langevin barostat (and Nosé-Hoover thermostat) are implemented in the DL_POLY_4 routines NPT_L0_VV and NPT_L0_LFV respectively. Both VV and LFV implementations make use of the DL_POLY_4 module LANGEVIN_MODULE. The routines NPT_L1_VV and NPT_L1_LFV implement the same but also incorporate RB dynamics.

Cell size and shape variations

The isotropic algorithms (VV and LFV) may be extended to allowing the cell shape to vary by defining η as a tensor, $\underline{\underline{\eta}}$ and extending the Langevin pressure variable R_p to a stochastic (Langevin) tensor $\underline{\underline{\mathbf{R_p}}}$:

$$\langle R_{p,i}(t)\ R_{p,j}(t')\rangle = 2\ \chi_p\ p_{mass}\ k_B T\ \delta_{ij}\ \delta(t-t')\quad , \tag{3.115}$$

which is drawn from Gaussian distribution of zero mean and unit variance, $\texttt{Gauss}(0,1)$, scaled by $\sqrt{\frac{2\ \chi_p\ p_{mass}\ k_B T}{\Delta t}}$. k_B is the Boltzmann constant, T the target temperature and p_{mass} the barostat mass. **Note** that $\underline{\underline{\mathbf{R_p}}}$ has to be symmetric and only 6 independent components must be generated each timestep.

The equations of motion are written in the same fashion as is in the isotropic algorithm with slight modifications (as now the equations with η are extended to matrix forms)

$$\begin{aligned}
\frac{d}{dt}\underline{r}(t) &= \underline{v}(t) + \underline{\underline{\eta(\mathbf{t})}}\cdot\underline{r}(t) \\
\frac{d}{dt}\underline{v}(t) &= \frac{\underline{f}(t)+\underline{R}(t)}{m} - \left[\chi\ \underline{\underline{\mathbf{1}}} + \underline{\underline{\eta}}(t) + \frac{\texttt{Tr}\left[\underline{\underline{\eta}}(t)\right]}{f}\ \underline{\underline{\mathbf{1}}}\right]\cdot\underline{v}(t) \\
\frac{d}{dt}\underline{\underline{\eta}}(t) &= \frac{\underline{\underline{\sigma}}(t) - P_{\rm ext}\ V(t)\ \underline{\underline{\mathbf{1}}}}{p_{mass}} + \frac{2E_{kin}(t)}{f}\frac{\underline{\underline{\mathbf{1}}}}{p_{mass}} - \chi_p\underline{\underline{\eta}}(t) + \frac{\underline{\underline{\mathbf{R_p}}}}{p_{mass}} \\
p_{mass} &= \frac{(f+3)}{3}\ \frac{k_B\ T_{\rm ext}}{(2\pi\ \chi_P)^2} \\
\frac{d}{dt}\underline{\underline{\mathbf{H}}}(t) &= \underline{\underline{\eta}}(t)\cdot\underline{\underline{\mathbf{H}}}(t) \\
\frac{d}{dt}V(t) &= \texttt{Tr}[\underline{\underline{\eta}}(t)]\ V(t)\quad .
\end{aligned} \tag{3.116}$$

where $\underline{\underline{\sigma}}$ is the stress tensor (equation (3.98)) and $\underline{\underline{\mathbf{1}}}$ is the identity matrix.

The conserved quantity these generate is:

$$\mathcal{H}_{\rm N\underline{\underline{\sigma}}T} = \mathcal{H}_{\rm NVE} + \frac{p_{mass}\ \texttt{Tr}[\underline{\underline{\eta}}\cdot\underline{\underline{\eta}}^T]}{2} + P_{\rm ext}V(t)\quad . \tag{3.117}$$

the VV and LFV algorithmic equations are, therefore, written in the same fashion as in the isotropic case with slight modifications. For the VV couched algorithm these are of the following sort

$$\begin{aligned}
\underline{\underline{\eta}}(t) &\leftarrow \exp\left(-\chi_p\ \frac{\Delta t}{8}\right)\ \underline{\underline{\eta}}(t) \\
\underline{\underline{\eta}}(t+\frac{1}{4}\Delta t) &\leftarrow \underline{\underline{\eta}}(t) + \\
&\qquad \frac{\Delta t}{4}\left[\frac{\underline{\underline{\sigma}}(t) - P_{\rm ext}\ V(t)\ \underline{\underline{\mathbf{1}}}}{p_{mass}} + \frac{2E_{kin}(t)}{f}\frac{\underline{\underline{\mathbf{1}}}}{p_{mass}} + \frac{\underline{\underline{\mathbf{R_p}}}(t)}{p_{mass}}\right] \\
\underline{v}(t) &\leftarrow \exp\left[-\left(\underline{\underline{\eta}}(t+\frac{1}{4}\Delta t) + \frac{1}{f}\texttt{Tr}\left[\underline{\underline{\eta}}(t+\frac{1}{4}\Delta t)\right]\right)\ \frac{\Delta t}{2}\right]\cdot\underline{v}(t) \\
\underline{r}(t+\Delta t) &\leftarrow \exp\left[\underline{\underline{\eta}}(t+\frac{1}{2}\Delta t)\ \Delta t\right]\cdot\underline{r}(t) + \Delta t\ \underline{v}(t+\frac{1}{2}\Delta t)
\end{aligned} \tag{3.118}$$

Similarly, for the LFV couched algorithms these are

$$\begin{aligned}
\underline{\underline{\eta}}(t+\frac{1}{2}\Delta t) &\leftarrow \exp\left(-\chi_p(t)\Delta t\right)\ \underline{\underline{\eta}}(t-\frac{\Delta t}{2})+ \\
&\quad \Delta t\ \frac{\underline{\underline{\sigma}}(t)-P_{\rm ext}\ V(t)\ \underline{\underline{\mathbf{1}}}}{p_{mass}}+\frac{2E_{kin}(t)}{f}\frac{\underline{\underline{\mathbf{1}}}}{p_{mass}}+\frac{\underline{\underline{\mathbf{R_p}}}(t)}{p_{mass}} \\
\underline{\underline{\texttt{scale}}} &= \underline{\underline{\mathbf{1}}}+\left[\chi\ \underline{\underline{\mathbf{1}}}+\left(1+\frac{3}{f}\right)\underline{\underline{\eta}}(t)\right]\frac{\Delta t}{2} \\
\underline{\underline{\texttt{scale\_v}}} &= \frac{2}{\underline{\underline{\texttt{scale}}}}-\underline{\underline{\mathbf{1}}} \qquad (3.119)\\
\underline{\underline{\texttt{scale\_f}}} &= \frac{\Delta t}{\underline{\underline{\texttt{scale}}}} \\
\underline{v}(t+\frac{1}{2}\Delta t) &\leftarrow \underline{\underline{\texttt{scale\_v}}}\cdot\underline{v}(t-\frac{1}{2}\Delta t)+\underline{\underline{\texttt{scale\_f}}}\cdot\frac{\underline{f}(t)+\underline{R}(t)}{m} \\
\underline{r}(t+\Delta t) &\leftarrow \underline{r}(t)+\Delta t\ \left\{\underline{v}(t+\frac{1}{2}\Delta t)+\underline{\underline{\eta}}(t+\frac{1}{2}\Delta t)\cdot\underline{r}(t+\frac{1}{2}\Delta t)\right\}
\end{aligned}$$

It is worth noting DL_POLY_4 uses Taylor expansion truncated to the quadratic term to approximate exponentials of tensorial terms.

This ensemble is optionally extending to constant normal pressure and constant surface area, NP$_n$AT [73], by semi-isotropic constraining of the barostat equation of motion to:

$$\frac{d}{dt}\eta_{\alpha\beta}(t)=\begin{cases}\frac{\sigma_{zz}(t)-P_{\rm ext}\ V(t)}{p_{mass}}+\frac{2E_{kin}(t)}{f\ p_{mass}}-\chi_p\eta_{zz}(t)+\frac{R_{p,zz}(t)}{p_{mass}} & :\ (\alpha=\beta)=z \\ 0\quad ;\quad \eta_{\alpha\beta}(0)=0 & :\ (\alpha,\beta)\neq z\ .\end{cases} \qquad (3.120)$$

Similarly, this ensemble is optionally extending to constant normal pressure and constant surface tension, NP$_n\gamma$T [73], by semi-isotropic constraining of the barostat equation of motion to:

$$\frac{d}{dt}\eta_{\alpha\beta}(t)=\begin{cases}\frac{\sigma_{\alpha\alpha}(t)-[P_{\rm ext}-\gamma_{\rm ext}/h_z(t)]\ V(t)}{p_{mass}}+\frac{2E_{kin}(t)}{f\ p_{mass}}-\chi_p\eta_{\alpha\alpha}(t)+\frac{R_{p,\alpha\alpha}(t)}{p_{mass}} & :\ (\alpha=\beta)=x,y \\ & : \\ \frac{\sigma_{zz}(t)-P_{\rm ext}\ V(t)}{p_{mass}}+\frac{2E_{kin}(t)}{f}\frac{1}{p_{mass}}-\chi_p\eta_{zz}(t)+\frac{R_{p,zz}(t)}{p_{mass}} & :\ (\alpha=\beta)=z \\ 0\quad ;\quad \eta_{\alpha\beta}(0)=0 & :\ (\alpha\neq\beta)=x,y,z\end{cases}\ , \qquad (3.121)$$

where $\gamma_{\rm ext}$ is the user defined external surface tension and $h_z(t)=V(t)/A_{xy}(t)$ is the instantaneous hight of the MD box (or MD box volume over area). The instnatneous surface tension is defined as

$$\gamma_\alpha(t)=-h_z(t)\left[\sigma_{\alpha\alpha}(t)-P_{\rm ext}\right]\ . \qquad (3.122)$$

The case $\gamma_{\rm ext}=0$ generates the NPT anisotropic ensemble for the orthorhombic cell (imcon=2 in CONFIG, see Appendix B). This can be considered as an ”orthorhombic” constraint on the NσT ensemble. The constraint can be strengthened further, to a ”semi-orthorhombic” one, by imposing that the MD cell change isotropically in the (x,y) plane which leads to the following modification in the N$P_n\gamma$T set of equatons

$$\begin{aligned}\frac{d}{dt}\eta_{\alpha\alpha}(t) &= \frac{\left[\sigma_{xx}(t)+\sigma_{yy}(t)\right]/2-\left[P_{\rm ext}-\gamma_{\rm ext}/h_z(t)\right]\ V(t)}{p_{mass}}+\frac{2\ E_{kin}(t)}{f\ p_{mass}}- \\ &\quad -\chi_p\eta_{\alpha\alpha}(t)+\frac{R_{p,xx}(t)+R_{p,yy}(t)}{2\ p_{mass}}\quad :\ (\alpha=\beta)=x,y\ .\end{aligned} \qquad (3.123)$$

The VV and LFV flavours of the non-isotropic Langevin barostat (and Nosé-Hoover thermostat) are implemented in the DL_POLY_4 routines NST_L0_VV and NST_L0_LFV respectively. Both make use of the DL_POLY_4 module LANGEVIN_MODULE. The routines NST_L1_VV and NST_L1_LFV implement the same but also incorporate RB dynamics.

3.5.3 Berendsen Barostat

With the Berendsen barostat the system is made to obey the equation of motion at the beginning of each step

$$\frac{d\mathcal{P}(t)}{dt} = \frac{P_{\text{ext}} - \mathcal{P}(t)}{\tau_P} \quad , \tag{3.124}$$

where $\mathcal{P}$ is the instantaneous pressure (equation (3.97)) and τ_P is the barostat relaxation time constant.

Cell size variations

In the isotropic implementation, at each step the MD cell volume is scaled by a factor η, and the coordinates and cell vectors by $\eta^{1/3}$,

$$\eta(t) = 1 - \frac{\beta \Delta t}{\tau_P} \left(P_{\text{ext}} - \mathcal{P}(t)\right) \tag{3.125}$$

where β is the isothermal compressibility of the system. In practice β is a specified constant which DL_POLY_4 takes to be the isothermal compressibility of liquid water. The exact value is not critical to the algorithm as it relies on the ratio τ_P/β. τ_P is a specified time constant for pressure fluctuations, supplied by the user.

It is worth noting that the barostat and the thermostat are independent and fully separable.

The VV implementation of the Berendsen algorithm only requires iterations if bond or PMF constraints are present (13 until satisfactory convergence of the constraint forces is achieved). These are with respect to the pressure (i.e. $\eta(t)$) in the first part, VV1+RATTLE_VV1. The second part is conventional, VV2+RATTLE_VV2, as at the end the velocities are scaled by a factor of χ.

1. VV1:

$$\begin{aligned} \underline{v}(t + \frac{1}{2}\Delta t) &\leftarrow \underline{v}(t) + \frac{\Delta t}{2} \frac{\underline{f}(t)}{m} \\ \underline{r}(t + \Delta t) &\leftarrow \eta(t)^{1/3}\, \underline{r}(t) + \Delta t\, \underline{v}(t + \frac{1}{2}\Delta t) \\ \underline{\underline{\mathbf{H}}}(t + \Delta t) &\leftarrow \eta(t)^{1/3}\, \underline{\underline{\mathbf{H}}}(t) \\ V(t + \Delta t) &\leftarrow \eta(t)\, V(t) \end{aligned} \tag{3.126}$$

2. RATTLE_VV1

3. Barostat:

$$\eta(t) = 1 - \frac{\beta \Delta t}{\tau_P} \left(P_{\text{ext}} - \mathcal{P}(t)\right) \tag{3.127}$$

4. FF:

$$\underline{f}(t + \Delta t) \leftarrow \underline{f}(t) \tag{3.128}$$

5. VV2:

$$\underline{v}(t + \Delta t) \leftarrow \underline{v}(t + \frac{1}{2}\Delta t) + \frac{\Delta t}{2} \frac{\underline{f}(t + \Delta t)}{m} \tag{3.129}$$

6. RATTLE_VV2

7. Thermostat:

$$\begin{aligned} \chi(t + \Delta t) &\leftarrow \left[1 + \frac{\Delta t}{\tau_T}\left(\frac{\sigma}{E_{kin}(t + \Delta t)} - 1\right)\right]^{1/2} \\ \underline{v}(t + \Delta t) &\leftarrow \underline{v}(t + \Delta t)\, \chi \quad . \end{aligned} \tag{3.130}$$

where $\underline{\underline{\mathbf{H}}}$ is the cell matrix whose columns are the three cell vectors $\underline{a}, \underline{b}, \underline{c}$.

The LFV implementation of the Berendsen algorithm is iterative, until self consistency in the full step velocity, $\underline{v}(t)$, is obtained. Initial estimates of $\chi(t)$ and $\eta(t)$ at full step are calculated using an unconstrained estimate of the velocity at full step, $\underline{v}(t)$.

1. FF:
$$\underline{f}(t) \leftarrow \underline{f}(t-\Delta t) \tag{3.131}$$

2. LFV: The iterative part is as follows:
$$\begin{aligned} \underline{v}(t+\frac{1}{2}\Delta t) &\leftarrow \left[\underline{v}(t-\frac{1}{2}\Delta t) + \Delta t\,\frac{\underline{f}(t)}{m}\right]\,\chi(t) \\ \underline{r}(t+\Delta t) &\leftarrow \eta(t)^{1/3}\,\underline{r}(t) + \Delta t\,\underline{v}(t+\frac{1}{2}\Delta t) \\ \underline{\underline{\mathbf{H}}}(t+\Delta t) &\leftarrow \eta(t)^{1/3}\,\underline{\underline{\mathbf{H}}}(t) \\ V(t+\Delta t) &\leftarrow \eta(t)\,V(t) \end{aligned} \tag{3.132}$$

3. SHAKE

4. Full step velocity:
$$\underline{v}(t) \leftarrow \frac{1}{2}\left[\underline{v}(t-\frac{1}{2}\Delta t) + \underline{v}(t+\frac{1}{2}\Delta t)\right] \tag{3.133}$$

5. Thermostat:
$$\chi(t) \leftarrow \left[1 + \frac{\Delta t}{\tau_T}\left(\frac{\sigma}{E_{kin}(t)} - 1\right)\right]^{1/2} \tag{3.134}$$

6. Barostat:
$$\eta(t) = 1 - \frac{\beta\Delta t}{\tau_P}\,(P_{\rm ext} - \mathcal{P}(t)) \quad . \tag{3.135}$$

Several iterations are required to obtain self consistency. In DL_POLY_4 the number of iterations is set to 7 (8 if bond constraints are present). Note also that the change in box size requires the SHAKE algorithm to be called each iteration.

The Berendsen algorithms conserve total momentum but not energy.

The VV and LFV flavours of the Berendsen barostat (and thermostat) are implemented in the DL_POLY_4 routines NPT_B0_VV and NPT_B0_LFV respectively. The routines NPT_B1_VV and NPT_B1_LFV implement the same but also incorporate RB dynamics.

Cell size and shape variations

The extension of the isotropic algorithm to anisotropic cell variations is straightforward. A tensor $\underline{\underline{\eta}}$ is defined as
$$\underline{\underline{\eta}}(t) = \underline{\underline{\mathbf{1}}} - \frac{\beta\Delta t}{\tau_P}(P_{\rm ext}\,\underline{\underline{\mathbf{1}}} - \underline{\underline{\sigma}}(t)/V(t)) \quad , \tag{3.136}$$

where where $\underline{\underline{\sigma}}$ is the stress tensor (equation (3.98)) and $\underline{\underline{\mathbf{1}}}$ is the identity matrix. Then new cell vectors and volume are given by
$$\begin{aligned} \underline{\underline{\mathbf{H}}}(t+\Delta t) &\leftarrow \underline{\underline{\eta}}(t)\cdot\underline{\underline{\mathbf{H}}}(t) \\ V(t+\Delta t) &\leftarrow \mathtt{Tr}[\underline{\underline{\eta}}(t)]\;V(t) \quad . \end{aligned} \tag{3.137}$$

and the velocity updates for VV and LFV algorithms as

$$\begin{aligned}
\texttt{VV1}: \quad \underline{r}(t+\Delta t) &\leftarrow \underline{\underline{\eta}}(t)\cdot\underline{r}(t)+\Delta t\;\underline{v}(t+\frac{1}{2}\Delta t) \\
\texttt{LFV}: \quad \underline{r}(t+\Delta t) &\leftarrow \underline{\underline{\eta}}(t)\cdot\underline{r}(t)+\Delta t\;\underline{v}(t+\frac{1}{2}\Delta t)\quad . \qquad (3.138)
\end{aligned}$$

This ensemble is optionally extending to constant normal pressure and constant surface area, NP$_n$AT [73], by semi-isotropic constraining of the barostat equation of motion to:

$$\eta_{\alpha\delta}(t)=\begin{cases} 1-\frac{\beta\Delta t}{\tau_P}\left[P_{\rm ext}-\sigma_{zz}(t)/V(t)\right] & : \quad (\alpha=\delta)=z \\ 1 & : \quad (\alpha=\delta)=x,y \\ 0 & : \quad (\alpha\neq\delta)\quad . \end{cases} \qquad (3.139)$$

Similarly, this ensemble is optionally extending to constant normal pressure and constant surface tension, NP$_n\gamma$T [73], by semi-isotropic constraining of the barostat equation of motion to:

$$\eta_{\alpha\delta}(t)=\begin{cases} 1-\frac{\beta\Delta t}{\tau_P}\left[P_{\rm ext}-\gamma_{\rm ext}\;V(t)/h_z(t)-\sigma_{\alpha\alpha}(t)/V(t)\right] & : \quad (\alpha=\delta)=x,y \\ & : \\ 1-\frac{\beta\Delta t}{\tau_P}\left[P_{\rm ext}-\sigma_{zz}(t)/V(t)\right] & : \quad (\alpha=\delta)=z \\ 0 & : \quad (\alpha\neq\delta)\quad , \end{cases} \qquad (3.140)$$

where $\gamma_{\rm ext}$ is the user defined external surface tension and $h_z(t)=V(t)/A_{xy}(t)$ is the instantaneous hight of the MD box (or MD box volume over area). One defines the instantaneous surface tension as given in equation (3.122). The case $\gamma_{\rm ext}=0$ generates the NPT anisotropic ensemble for the orthorhombic cell (imcon=2 in CONFIG, see Appendix B). This can be considered as an "orthorhombic" constraint on the NσT ensemble. The constraint can be strengthened further, to a "semi-orthorhombic" one, by imposing that the MD cell change isotropically in the (x,y) plane which leads to the following change in the equations above

$$\eta_{\alpha\alpha}(t)=1-\frac{\beta\Delta t}{\tau_P}\left[P_{\rm ext}-\gamma_{\rm ext}\;\frac{V(t)}{h_z(t)}-\frac{\sigma_{xx}(t)+\sigma_{yy}(t)}{2\;V(t)}\right] \quad : \quad (\alpha=\delta)=x,y\quad . \qquad (3.141)$$

The VV and LFV flavours of the non-isotropic Berendsen barostat (and thermostat) are implemented in the DL_POLY_4 routines NST_B0_VV and NST_B0_LFV respectively. The routines NST_B1_VV and NST_B1_LFV implement the same but also incorporate RB dynamics.

3.5.4 Nosé-Hoover Barostat

DL_POLY_4 uses the Melchionna modification of the Nosé-Hoover algorithm [80] in which the equations of motion involve a Nosé-Hoover thermostat and a barostat in the same spirit. Additionally, as shown in [81], a modification allowing for coupling between the thermostat and barostat is also introduced.

Cell size variation

For isotropic fluctuations the equations of motion are:

$$\begin{aligned}
\frac{d}{dt}\underline{r}(t) &= \underline{v}(t)+\eta(t)\;\left(\underline{r}(t)-\underline{R}_0(t)\right) \\
\frac{d}{dt}\underline{v}(t) &= \frac{\underline{f}(t)}{m}-\left[\chi(t)+\eta(t)\right]\underline{v}(t) \\
\frac{d}{dt}\chi(t) &= \frac{2E_{kin}(t)+p_{mass}\;\eta(t)^2-2\sigma-k_B\;T_{\rm ext}}{q_{mass}} \\
q_{mass} &= 2\;\sigma\;\tau_T^2 \qquad (3.142)
\end{aligned}$$

$$
\begin{aligned}
\frac{d}{dt}\eta(t) &= 3V(t)\frac{\mathcal{P}(t)-P_{\rm ext}}{p_{mass}} - \chi(t)\eta(t) \\
p_{mass} &= (f+3)\ k_B\ T_{\rm ext}\ \tau_P^2 \\
\frac{d}{dt}\underline{\underline{\mathbf{H}}}(t) &= \eta(t)\ \underline{\underline{\mathbf{H}}}(t) \\
\frac{d}{dt}V(t) &= [3\eta(t)]\ V(t)\quad ,
\end{aligned}
$$

where η is the barostat friction coefficient, $\underline{R}_0(t)$ the system centre of mass at time t, q_{mass} the thermostat mass, τ_T a specified time constant for temperature fluctuations, σ the target thermostat energy (equation (3.57)), p_{mass} the barostat mass, τ_P a specified time constant for pressure fluctuations, $\mathcal{P}$ the instantaneous pressure (equation (3.97)) and V the system volume. $\underline{\underline{\mathbf{H}}}$ is the cell matrix whose columns are the three cell vectors $\underline{a}, \underline{b}, \underline{c}$.

The conserved quantity is, to within a constant, the Gibbs free energy of the system:

$$
\mathcal{H}_{\rm NPT} = \mathcal{H}_{\rm NVE} + \frac{q_{mass}\ \chi(t)^2}{2} + \frac{p_{mass}\ \eta(t)^2}{2} + P_{\rm ext}V(t) + (f+1)\ k_B\ T_{\rm ext}\int_o^t \chi(s)ds\quad , \tag{3.143}
$$

where f is the system's degrees of freedom - equation (3.11).

The VV implementation of the Nosé-Hoover algorithm only requires iterations if bond or PMF constraints are present (5 until satisfactory convergence of the constraint forces is achieved). These are with respect to the pressure (i.e. $\eta(t)$) in the first part, VV1+RATTLE_VV1. The second part is conventional, VV2+RATTLE_VV2, as at the end the velocities are scaled by a factor of χ.

1. Thermostat: Note $2E_{kin}(t)$ changes inside

$$
\begin{aligned}
\chi(t+\frac{1}{8}\Delta t) &\leftarrow \chi(t) + \frac{\Delta t}{8}\ \frac{2E_{kin}(t) + p_{mass}\ \eta(t)^2 - 2\sigma - k_B\ T_{\rm ext}}{q_{mass}} \\
\underline{v}(t) &\leftarrow \exp\left(-\chi(t+\frac{1}{8}\Delta t)\ \frac{\Delta t}{4}\right)\ \underline{v}(t) \\
\chi(t+\frac{1}{4}\Delta t) &\leftarrow \chi(t+\frac{1}{8}\Delta t) + \frac{\Delta t}{8}\ \frac{2E_{kin}(t) + p_{mass}\ \eta(t)^2 - 2\sigma - k_B\ T_{\rm ext}}{q_{mass}}
\end{aligned} \tag{3.144}
$$

2. Barostat: Note $E_{kin}(t)$ and $\mathcal{P}(t)$ have changed and change inside

$$
\begin{aligned}
\eta(t) &\leftarrow \exp\left(-\chi(t+\frac{1}{4}\Delta t)\ \frac{\Delta t}{8}\right)\ \eta(t) \\
\eta(t+\frac{1}{4}\Delta t) &\leftarrow \eta(t) + \frac{\Delta t}{4}\ \frac{3\left[\mathcal{P}(t) - P_{\rm ext}\right]V(t)}{p_{mass}} \\
\eta(t+\frac{1}{4}\Delta t) &\leftarrow \exp\left(-\chi(t+\frac{1}{4}\Delta t)\ \frac{\Delta t}{8}\right)\ \eta(t+\frac{1}{4}\Delta t) \\
\underline{v}(t) &\leftarrow \exp\left[-\eta(t+\frac{1}{4}\Delta t)\ \frac{\Delta t}{2}\right]\ \underline{v}(t) \\
\eta(t+\frac{1}{4}\Delta t) &\leftarrow \exp\left(-\chi(t+\frac{1}{4}\Delta t)\ \frac{\Delta t}{8}\right)\ \eta(t+\frac{1}{4}\Delta t) \\
\eta(t+\frac{1}{2}\Delta t) &\leftarrow \eta(t+\frac{1}{4}\Delta t) + \frac{\Delta t}{4}\ \frac{3\left[\mathcal{P}(t) - P_{\rm ext}\right]V(t)}{p_{mass}} \\
\eta(t+\frac{1}{2}\Delta t) &\leftarrow \exp\left(-\chi(t+\frac{1}{4}\Delta t)\ \frac{\Delta t}{8}\right)\ \eta(t+\frac{1}{2}\Delta t)
\end{aligned} \tag{3.145}
$$

3. Thermostat: Note $E_{kin}(t)$ has changed and changes inside

$$
\chi(t+\frac{3}{8}\Delta t) \leftarrow \chi(t+\frac{1}{4}\Delta t) + \frac{\Delta t}{8}\ \frac{2E_{kin}(t) + p_{mass}\ \eta(t+\frac{1}{2}\Delta t)^2 - 2\sigma - k_B\ T_{\rm ext}}{q_{mass}}
$$

$$\begin{aligned}
\underline{v}(t) &\leftarrow \exp\left(-\chi(t+\frac{3}{8}\Delta t)\ \frac{\Delta t}{4}\right)\ \underline{v}(t) \\
\chi(t+\frac{1}{2}\Delta t) &\leftarrow \chi(t+\frac{3}{8}\Delta t) + \frac{\Delta t}{8}\ \frac{2E_{kin}(t) + p_{mass}\ \eta(t+\frac{1}{2}\Delta t)^2 - 2\sigma - k_B\ T_{\mathrm{ext}}}{q_{mass}}
\end{aligned} \tag{3.146}$$

4. VV1:

$$\begin{aligned}
\underline{v}(t+\frac{1}{2}\Delta t) &\leftarrow \underline{v}(t) + \frac{\Delta t}{2}\ \frac{\underline{f}(t)}{m} \\
\underline{\underline{\mathbf{H}}}(t+\Delta t) &\leftarrow \exp\left[\eta(t+\frac{1}{2}\Delta t)\ \Delta t\right]\ \underline{\underline{\mathbf{H}}}(t) \\
V(t+\Delta t) &\leftarrow \exp\left[3\eta(t+\frac{1}{2}\Delta t)\ \Delta t\right]\ V(t) \\
\underline{r}(t+\Delta t) &\leftarrow \exp\left[\eta(t+\frac{1}{2}\Delta t)\ \Delta t\right]\ (\underline{r}(t) - \underline{R}_0(t)) + \Delta t\ \underline{v}(t+\frac{1}{2}\Delta t) + \underline{R}_0(t)
\end{aligned} \tag{3.147}$$

5. RATTLE_VV1

6. FF:

$$\underline{f}(t+\Delta t) \leftarrow \underline{f}(t) \tag{3.148}$$

7. VV2:

$$\underline{v}(t+\Delta t) \leftarrow \underline{v}(t+\frac{\Delta t}{2}) + \frac{\Delta t}{2}\ \frac{\underline{f}(t)}{m} \tag{3.149}$$

8. RATTLE_VV2

9. Thermostat: Note $E_{kin}(t+\Delta t)$ has changed and changes inside

$$\begin{aligned}
\chi(t+\frac{5}{8}\Delta t) &\leftarrow \chi(t+\frac{1}{2}\Delta t) + \frac{\Delta t}{8}\ \frac{2E_{kin}(t+\Delta t) + p_{mass}\ \eta(t+\frac{1}{2}\Delta t)^2 - 2\sigma - k_B\ T_{\mathrm{ext}}}{q_{mass}} \\
\underline{v}(t+\Delta t) &\leftarrow \exp\left(-\chi(t+\frac{5}{8}\Delta t)\ \frac{\Delta t}{4}\right)\ \underline{v}(t+\Delta t) \\
\chi(t+\frac{3}{4}\Delta t) &\leftarrow \chi(t+\frac{5}{8}\Delta t) + \frac{\Delta t}{8}\ \frac{2E_{kin}(t+\Delta t) + p_{mass}\ \eta(t+\frac{1}{2}\Delta t)^2 - 2\sigma - k_B\ T_{\mathrm{ext}}}{q_{mass}}
\end{aligned} \tag{3.150}$$

10. Barostat: Note $E_{kin}(t+\Delta t)$ and $\mathcal{P}(t+\Delta t)$ have changed and change inside

$$\begin{aligned}
\eta(t+\frac{1}{2}\Delta t) &\leftarrow \exp\left(-\chi(t+\frac{3}{4}\Delta t)\ \frac{\Delta t}{8}\right)\ \eta(t+\frac{1}{2}\Delta t) \\
\eta(t+\frac{3}{4}\Delta t) &\leftarrow \eta(t+\frac{1}{2}\Delta t) + \frac{\Delta t}{4}\ \frac{3\left[\mathcal{P}(t+\Delta t) - P_{\mathrm{ext}}\right] V(t+\Delta t)}{p_{mass}} \\
\eta(t+\frac{3}{4}\Delta t) &\leftarrow \exp\left(-\chi(t+\frac{3}{4}\Delta t)\ \frac{\Delta t}{8}\right)\ \eta(t+\frac{3}{4}\Delta t) \\
\underline{v}(t+\Delta t) &\leftarrow \exp\left[-\eta(t+\frac{3}{4}\Delta t)\ \frac{\Delta t}{2}\right]\ \underline{v}(t+\Delta t) \\
\eta(t+\frac{3}{4}\Delta t) &\leftarrow \exp\left(-\chi(t+\frac{3}{4}\Delta t)\ \frac{\Delta t}{8}\right)\ \eta(t+\frac{3}{4}\Delta t) \\
\eta(t+\Delta t) &\leftarrow \eta(t+\frac{3}{4}\Delta t) + \frac{\Delta t}{4}\ \frac{3\left[\mathcal{P}(t+\Delta t) - P_{\mathrm{ext}}\right] V(t+\Delta t)}{p_{mass}} \\
\eta(t+\Delta t) &\leftarrow \exp\left(-\chi(t+\frac{3}{4}\Delta t)\ \frac{\Delta t}{8}\right)\ \eta(t+\Delta t)
\end{aligned} \tag{3.151}$$

11. Thermostat: Note $E_{kin}(t+\Delta t)$ has changed and changes inside

$$\begin{aligned}
\chi(t+\frac{7}{8}\Delta t) &\leftarrow \chi(t+\frac{3}{4}\Delta t)+\frac{\Delta t}{8}\,\frac{2E_{kin}(t+\Delta t)+p_{mass}\;\eta(t+\Delta t)^2-2\sigma-k_B\;T_{\rm ext}}{q_{mass}} \\
\underline{v}(t+\Delta t) &\leftarrow \exp\left(-\chi(t+\frac{7}{8}\Delta t)\;\frac{\Delta t}{4}\right)\;\underline{v}(t+\Delta t) \\
\chi(t+\Delta t) &\leftarrow \chi(t+\frac{7}{8}\Delta t)+\frac{\Delta t}{8}\,\frac{2E_{kin}(t+\Delta t)+p_{mass}\;\eta(t+\Delta t)^2-2\sigma-k_B\;T_{\rm ext}}{q_{mass}} \\
\underline{v}(t+\Delta t) &\leftarrow \underline{v}(t+\Delta t)-\underline{V}_0(t+\Delta t)\quad,
\end{aligned} \tag{3.152}$$

where $\underline{V}_0(t+\Delta t)$ is the c.o.m. velocity at timestep $t+\Delta t$ and $\underline{\underline{\mathbf{H}}}$ is the cell matrix whose columns are the three cell vectors $\underline{a},\underline{b},\underline{c}$.

The LFV implementation of the Nosé-Hoover algorithm is iterative, until self consistency in the full step velocity, $\underline{v}(t)$, is obtained. Initial estimates of $\chi(t)$ and $\eta(t)$ at full step are calculated using an unconstrained estimate of the velocity at full step, $\underline{v}(t)$. Also calculated is an unconstrained estimate of the half step position $\underline{r}(t+\frac{1}{2}\Delta t)$.

1. FF:

$$\underline{f}(t)\leftarrow\underline{f}(t-\Delta t) \tag{3.153}$$

2. LFV: The iterative part is as follows:

$$\begin{aligned}
\underline{v}(t+\frac{1}{2}\Delta t) &\leftarrow \underline{v}(t-\frac{1}{2}\Delta t)+\Delta t\;\left[\frac{\underline{f}(t)}{m}-(\chi(t)+\eta(t))\;\underline{v}(t)\right] \\
\underline{r}(t+\Delta t) &\leftarrow \underline{r}(t)+\Delta t\;\left\{\underline{v}(t+\frac{1}{2}\Delta t)+\eta(t+\frac{1}{2}\Delta t)\left[\underline{r}(t+\frac{1}{2}\Delta t)-\underline{R}_0(t)\right]\right\} \\
\underline{\underline{\mathbf{H}}}(t+\Delta t) &\leftarrow \exp\left[\eta(t+\frac{1}{2}\Delta t)\;\Delta t\right]\;\underline{\underline{\mathbf{H}}}(t) \\
V(t+\Delta t) &\leftarrow \exp\left[3\eta(t+\frac{1}{2}\Delta t)\;\Delta t\right]\;V(t)
\end{aligned} \tag{3.154}$$

3. SHAKE

4. Full step velocity and half step position:

$$\begin{aligned}
\underline{v}(t) &\leftarrow \frac{1}{2}\left[\underline{v}(t-\frac{1}{2}\Delta t)+\underline{v}(t+\frac{1}{2}\Delta t)\right] \\
\underline{r}(t+\frac{1}{2}\Delta t) &\leftarrow \frac{\underline{r}(t)+\underline{r}(t+\Delta t)}{2}
\end{aligned} \tag{3.155}$$

5. Thermostat and Barostat:

$$\begin{aligned}
\chi(t+\frac{1}{2}\Delta t) &\leftarrow \chi(t-\frac{1}{2}\Delta t)+\Delta t\;\frac{2E_{kin}(t)+p_{mass}\;\eta(t)^2-2\sigma-k_B\;T_{\rm ext}}{q_{mass}} \\
\eta(t+\frac{1}{2}\Delta t) &\leftarrow \exp\left(-\chi(t)\Delta t\right)\;\eta(t-\frac{1}{2}\Delta t)+\Delta t\;\frac{3\left[\mathcal{P}(t)-P_{\rm ext}\right]V(t)}{p_{mass}} \\
\chi(t) &\leftarrow \frac{1}{2}\left[\chi(t-\frac{1}{2}\Delta t)+\chi(t+\frac{1}{2}\Delta t)\right] \\
\eta(t) &\leftarrow \frac{1}{2}\left[\eta(t-\frac{1}{2}\Delta t)+\eta(t+\frac{1}{2}\Delta t)\right]\quad.
\end{aligned} \tag{3.156}$$

Several iterations are required to obtain self consistency. In DL_POLY_4 the number of iterations is set to 7 (8 if bond constraints are present). Note also that the change in box size requires the SHAKE algorithm to be called each iteration.

The VV and LFV flavours of the Nosé-Hoover barostat (and thermostat) are implemented in the DL_POLY_4 routines NPT_H0_VV and NPT_H0_LFV respectively. The routines NPT_H1_VV and NPT_H1_LFV implement the same but also incorporate RB dynamics.

Cell size and shape variation

The isotropic algorithms (VV and LFV) may be extended to allowing the cell shape to vary by defining η as a tensor, $\underline{\underline{\eta}}$. The equations of motion are written in the same fashion as is in the isotropic algorithm with slight modifications (as now the equations with η are extended to matrix forms)

$$\begin{aligned}
\frac{d}{dt}\underline{r}(t) &= \underline{v}(t) + \underline{\underline{\eta}}(t) \cdot (\underline{r}(t) - \underline{R}_0(t)) \\
\frac{d}{dt}\underline{v}(t) &= \frac{\underline{f}(t)}{m} - \left[\chi(t)\ \underline{\underline{\mathbf{1}}} + \underline{\underline{\eta}}(t)\right] \cdot \underline{v}(t) \\
\frac{d}{dt}\chi(t) &= \frac{2E_{kin}(t) + p_{mass}\ \mathtt{Tr}[\underline{\underline{\eta}}(t) \cdot \underline{\underline{\eta}}(t)^T] - 2\sigma - 3^2\ k_B\ T_{\rm ext}}{q_{mass}} \\
q_{mass} &= 2\ \sigma\ \tau_T^2 \\
\frac{d}{dt}\underline{\underline{\eta}}(t) &= \frac{\underline{\underline{\sigma}}(t) - P_{\rm ext}\ V(t)\ \underline{\underline{\mathbf{1}}}}{p_{mass}} - \chi(t)\underline{\underline{\eta}}(t) \\
p_{mass} &= \frac{(f+3)}{3}\ k_B\ T_{\rm ext}\ \tau_P^2 \\
\frac{d}{dt}\underline{\underline{\mathbf{H}}}(t) &= \underline{\underline{\eta}}(t) \cdot \underline{\underline{\mathbf{H}}}(t) \\
\frac{d}{dt}V(t) &= \mathtt{Tr}[\underline{\underline{\eta}}(t)]\ V(t) \quad ,
\end{aligned} \tag{3.157}$$

where $\underline{\underline{\sigma}}$ is the stress tensor (equation (3.98)) and $\underline{\underline{\mathbf{1}}}$ is the identity matrix. The VV and LFV algorithmic equations are, therefore, written in the same fashion as above with slight modifications in (i) the equations for the thermostat and barostat frictions, and (ii) the equations for the system volume and cell parameters. The modifications in (i) for the VV couched algorithm are of the following sort

$$\begin{aligned}
\chi(t + \frac{1}{8}\Delta t) &\leftarrow \chi(t) + \frac{\Delta t}{8}\ \frac{2E_{kin}(t) + p_{mass}\ \mathtt{Tr}[\underline{\underline{\eta}}(t) \cdot \underline{\underline{\eta}}(t)^T] - 2\sigma - 3^2\ k_B\ T_{\rm ext}}{q_{mass}} \\
\underline{v}(t) &\leftarrow \exp\left[-\underline{\underline{\eta}}(t + \frac{1}{4}\Delta t)\ \frac{\Delta t}{2}\right] \cdot \underline{v}(t) \\
\underline{\underline{\eta}}(t + \frac{1}{4}\Delta t) &\leftarrow \underline{\underline{\eta}}(t) + \frac{\Delta t}{4}\ \frac{\underline{\underline{\sigma}}(t) - P_{\rm ext}\ V(t)\ \underline{\underline{\mathbf{1}}}}{p_{mass}} \quad ,
\end{aligned} \tag{3.158}$$

whereas for the LFV couched algorithm they are

$$\begin{aligned}
\chi(t + \frac{1}{2}\Delta t) &\leftarrow \chi(t - \frac{1}{2}\Delta t) + \Delta t\ \frac{2E_{kin}(t) + p_{mass}\ \mathtt{Tr}[\underline{\underline{\eta}}(t) \cdot \underline{\underline{\eta}}(t)^T] - 2\sigma - 3^2\ k_B\ T_{\rm ext}}{q_{mass}} \\
\underline{v}(t + \frac{1}{2}\Delta t) &\leftarrow \underline{v}(t - \frac{1}{2}\Delta t) + \Delta t\ \left[\frac{\underline{f}(t)}{m} - \left\{\chi(t)\ \underline{\underline{\mathbf{1}}} + \underline{\underline{\eta}}(t)\right\} \cdot \underline{v}(t)\right] \\
\underline{\underline{\eta}}(t + \frac{1}{2}\Delta t) &\leftarrow \exp\left(-\chi(t)\Delta t\right)\ \underline{\underline{\eta}}(t - \frac{1}{2}\Delta t) + \Delta t\ \frac{\underline{\underline{\sigma}}(t) - P_{\rm ext}\ V(t)\ \underline{\underline{\mathbf{1}}}}{p_{mass}} \quad .
\end{aligned} \tag{3.159}$$

The modifications in (ii) are the same for both the VV and LFV couched algorithms

$$\underline{\underline{\mathbf{H}}}(t + \Delta t) \leftarrow \exp\left(\underline{\underline{\eta}}(t + \frac{1}{2}\Delta t)\ \Delta t\right) \cdot \underline{\underline{\mathbf{H}}}(t)$$

$$V(t+\Delta t) \quad \leftarrow \quad \exp\left(\mathtt{Tr}\left[\underline{\underline{\eta}}(t+\frac{1}{2}\Delta t)\right]\ \Delta t\right)\ V(t) \quad . \tag{3.160}$$

It is worth noting DL_POLY_4 uses Taylor expansion truncated to the quadratic term to approximate exponentials of tensorial terms.

The conserved quantity is, to within a constant, the Gibbs free energy of the system:

$$\mathcal{H}_{\mathrm{N}\underline{\underline{\sigma}}\mathrm{T}} = \mathcal{H}_{\mathrm{NVE}} + \frac{q_{mass}\ \chi(t)^2}{2} + \frac{p_{mass}\ \mathtt{Tr}[\underline{\underline{\eta}}\cdot\underline{\underline{\eta}}^T]}{2} + P_{\mathrm{ext}}V(t) + (f+3^2)\ k_B\ T_{\mathrm{ext}}\int_o^t \chi(s)ds \quad , \tag{3.161}$$

where f is the system's degrees of freedom - equation (3.11).

This ensemble is optionally extending to constant normal pressure and constant surface area, NP$_n$AT [73], by semi-isotropic constraining of the barostat equation of motion and slight amending the thermostat equation of motion and the conserved quantity to:

$$\begin{aligned}
\frac{d}{dt}\eta_{\alpha\beta}(t) &= \begin{cases} \frac{\sigma_{zz}(t)-P_{\mathrm{ext}}\ V(t)}{p_{mass}} - \chi(t)\eta_{zz}(t) & : \quad (\alpha=\beta)=z \\ 0 \quad ; \quad \eta_{\alpha\beta}(0)=0 & : \quad (\alpha,\beta)\neq z \end{cases} \\
\frac{d}{dt}\chi(t) &= \frac{2E_{kin}(t) + p_{mass}\ \mathtt{Tr}[\underline{\underline{\eta}}(t)\cdot\underline{\underline{\eta}}(t)^T] - 2\sigma - k_B\ T_{\mathrm{ext}}}{q_{mass}} \\
\mathcal{H}_{\mathrm{NP_nAT}} &= \mathcal{H}_{\mathrm{NVE}} + \frac{q_{mass}\ \chi(t)^2}{2} + \frac{p_{mass}\ \mathtt{Tr}[\underline{\underline{\eta}}\cdot\underline{\underline{\eta}}^T]}{2} + P_{\mathrm{ext}}V(t) + (f+1)\ k_B\ T_{\mathrm{ext}}\int_o^t \chi(s)ds \quad .
\end{aligned} \tag{3.162}$$

Similarly, this ensemble is optionally extending to constant normal pressure and constant surface tension, NP$_n\gamma$T [73], by semi-isotropic constraining of the barostat equation of motion and slight amending the thermostat equation of motion and the conserved quantity to:

$$\begin{aligned}
\frac{d}{dt}\eta_{\alpha\beta}(t) &= \begin{cases} \frac{\sigma_{\alpha\alpha}(t)-[P_{\mathrm{ext}}-\gamma_{\mathrm{ext}}/h_z(t)]\ V(t)}{p_{mass}} - \chi(t)\eta_{\alpha\alpha}(t) & : \quad (\alpha=\beta)=x,y \\ & : \\ \frac{\sigma_{zz}(t)-P_{\mathrm{ext}}\ V(t)}{p_{mass}} - \chi(t)\eta_{zz}(t) & : \quad (\alpha=\beta)=z \\ 0 \quad ; \quad \eta_{\alpha\beta}(0)=0 & : \quad (\alpha\neq\beta)=x,y,z \end{cases} \\
\frac{d}{dt}\chi(t) &= \frac{2E_{kin}(t) + p_{mass}\ \mathtt{Tr}[\underline{\underline{\eta}}(t)\cdot\underline{\underline{\eta}}(t)^T] - 2\sigma - 3\ k_B\ T_{\mathrm{ext}}}{q_{mass}} \\
\mathcal{H}_{\mathrm{NP_n\gamma T}} &= \mathcal{H}_{\mathrm{NVE}} + \frac{q_{mass}\ \chi(t)^2}{2} + \frac{p_{mass}\ \mathtt{Tr}[\underline{\underline{\eta}}\cdot\underline{\underline{\eta}}^T]}{2} + P_{\mathrm{ext}}V(t) + (f+3)\ k_B\ T_{\mathrm{ext}}\int_o^t \chi(s)ds \quad .
\end{aligned} \tag{3.163}$$

where γ_{ext} is the user defined external surface tension and $h_z(t) = V(t)/A_{xy}(t)$ is the instantaneous hight of the MD box (or MD box volume over area). One defines the instantaneous surface tension as given in equation (3.122). The case $\gamma_{\mathrm{ext}} = 0$ generates the NPT anisotropic ensemble for the orthorhombic cell (imcon=2 in CONFIG, see Appendix B). This can be considered as an "orthorhombic" constraint on the NσT ensemble. The constraint can be strengthened further, to a "semi-orthorhombic" one, by imposing that the MD cell change isotropically in the (x,y) plane which leads to the following changes in the equations above

$$\begin{aligned}
\frac{d}{dt}\eta_{\alpha\alpha}(t) &= \frac{[\sigma_{xx}(t)+\sigma_{yy}(t)]/2 - [P_{\mathrm{ext}}-\gamma_{\mathrm{ext}}/h_z(t)]\ \ V(t)}{p_{mass}} - \chi(t)\eta_{\alpha\alpha}(t) \quad : \quad (\alpha=\beta)=x,y \\
\mathcal{H}_{\mathrm{NP_n\gamma=0T}} &= \mathcal{H}_{\mathrm{NVE}} + \frac{q_{mass}\ \chi(t)^2}{2} + \frac{p_{mass}\ \mathtt{Tr}[\underline{\underline{\eta}}\cdot\underline{\underline{\eta}}^T]}{2} + P_{\mathrm{ext}}V(t) + (f+2)\ k_B\ T_{\mathrm{ext}}\int_o^t \chi(s)ds \quad .
\end{aligned} \tag{3.164}$$

The VV and LFV flavours of the non-isotropic Nosé-Hoover barostat (and thermostat) are implemented in the DL_POLY_4 routines NST_H0_VV and NST_H0_LFV respectively. The routines NST_H1_VV and NST_H1_LFV implement the same but also incorporate RB dynamics.

3.5.5 Martyna-Tuckerman-Klein Barostat

DL_POLY_4 includes the Martyna-Tuckerman-Klein (MTK) interpretation of the VV flavoured Nosé-Hoover algorithms [32] for isotropic and anisotropic cell fluctuations in which the equations of motion are only slightly augmented with respect to those for the coupled Nosé-Hoover thermostat and barostat. Compare the isotropic cell changes case, equations (3.142), to

$$\begin{aligned}
\frac{d}{dt}\underline{r}(t) &= \underline{v}(t) + \eta(t)\;\underline{r}(t) \\
\frac{d}{dt}\underline{v}(t) &= \frac{\underline{f}(t)}{m} - \left[\chi(t) + \left(1+\frac{3}{f}\right)\eta(t)\right]\underline{v}(t) \\
\frac{d}{dt}\chi(t) &= \frac{2E_{kin}(t) + p_{mass}\;\eta(t)^2 - 2\sigma - k_B\;T_{\mathrm{ext}}}{q_{mass}} \\
q_{mass} &= 2\;\sigma\;\tau_T^2 \\
\frac{d}{dt}\eta(t) &= 3V(t)\frac{\mathcal{P}(t)-P_{\mathrm{ext}}}{p_{mass}} + 3\frac{2E_{kin}(t)}{f}\frac{1}{p_{mass}} - \chi(t)\eta(t) \\
p_{mass} &= (f+3)\;k_B\;T_{\mathrm{ext}}\;\tau_P^2 \\
\frac{d}{dt}\underline{\underline{\mathbf{H}}}(t) &= \eta(t)\;\underline{\underline{\mathbf{H}}}(t) \\
\frac{d}{dt}V(t) &= [3\eta(t)]\;V(t)\quad,
\end{aligned} \tag{3.165}$$

and the anisotropic cell change case, equations (3.157), to

$$\begin{aligned}
\frac{d}{dt}\underline{r}(t) &= \underline{v}(t) + \underline{\underline{\eta}}(t)\cdot\underline{r}(t) \\
\frac{d}{dt}\underline{v}(t) &= \frac{\underline{f}(t)}{m} - \left[\chi(t)\;\underline{\underline{\mathbf{1}}} + \underline{\underline{\eta}}(t) + \frac{\mathtt{Tr}\left[\underline{\underline{\eta}}(t)\right]}{f}\;\underline{\underline{\mathbf{1}}}\right]\cdot\underline{v}(t) \\
\frac{d}{dt}\chi(t) &= \frac{2E_{kin}(t) + p_{mass}\;\mathtt{Tr}[\underline{\underline{\eta}}(t)\cdot\underline{\underline{\eta}}(t)^T] - 2\sigma - 3^2\;k_B\;T_{\mathrm{ext}}}{q_{mass}} \\
q_{mass} &= 2\;\sigma\;\tau_T^2 \\
\frac{d}{dt}\underline{\underline{\eta}}(t) &= \frac{\underline{\underline{\sigma}}(t) - P_{\mathrm{ext}}\;V(t)\;\underline{\underline{\mathbf{1}}}}{p_{mass}} + \frac{2E_{kin}(t)}{f}\frac{\underline{\underline{\mathbf{1}}}}{p_{mass}} - \chi(t)\underline{\underline{\eta}}(t) \\
p_{mass} &= \frac{(f+3)}{3}\;k_B\;T_{\mathrm{ext}}\;\tau_P^2 \\
\frac{d}{dt}\underline{\underline{\mathbf{H}}}(t) &= \underline{\underline{\eta}}(t)\cdot\underline{\underline{\mathbf{H}}}(t) \\
\frac{d}{dt}V(t) &= \mathtt{Tr}[\underline{\underline{\eta}}(t)]\;V(t)\quad.
\end{aligned} \tag{3.166}$$

The changes include one extra dependence to the velocity and barostat equations and removal of the centre of mass variable $\underline{R}_0(t)$ dependence in the position equation.

The modifications in for the VV couched algorithms are of the following sort

$$\begin{aligned}
\eta(t+\frac{1}{4}\Delta t) &\leftarrow \eta(t) + \frac{\Delta t}{4}\left[3V(t)\frac{\mathcal{P}(t)-P_{\mathrm{ext}}}{p_{mass}} + 3\frac{2E_{kin}(t)}{f}\frac{1}{p_{mass}}\right] \\
\underline{v}(t) &\leftarrow \exp\left[-\left(1+\frac{3}{f}\right)\eta(t+\frac{1}{4}\Delta t)\;\frac{\Delta t}{2}\right]\underline{v}(t) \\
\underline{r}(t+\Delta t) &\leftarrow \exp\left[\eta(t+\frac{1}{2}\Delta t)\;\Delta t\right]\underline{r}(t) + \Delta t\;\underline{v}(t+\frac{1}{2}\Delta t)
\end{aligned} \tag{3.167}$$

for the isotropic cell fluctuations case and

$$\begin{aligned}
\underline{\underline{\eta}}(t+\frac{1}{4}\Delta t) &\leftarrow \underline{\underline{\eta}}(t)+\frac{\Delta t}{4}\left[\frac{\underline{\underline{\sigma}}(t)-P_{\rm ext}\,V(t)\,\underline{\underline{\mathbf{1}}}}{p_{mass}}+\frac{2E_{kin}(t)}{f}\frac{\underline{\underline{\mathbf{1}}}}{p_{mass}}\right] \\
\underline{v}(t) &\leftarrow \exp\left[-\left(\underline{\underline{\eta}}(t+\frac{1}{4}\Delta t)+\frac{1}{f}\texttt{Tr}\left[\underline{\underline{\eta}}(t+\frac{1}{4}\Delta t)\right]\right)\frac{\Delta t}{2}\right]\cdot\underline{v}(t) \\
\underline{r}(t+\Delta t) &\leftarrow \exp\left[\underline{\underline{\eta}}(t+\frac{1}{2}\Delta t)\;\Delta t\right]\cdot\underline{r}(t)+\Delta t\;\underline{v}(t+\frac{1}{2}\Delta t)
\end{aligned} \tag{3.168}$$

for the anisotropic cell fluctuations case. Similarly, for the LFV couched algorithms these are

$$\begin{aligned}
\eta(t+\frac{1}{2}\Delta t) &\leftarrow \exp\left(-\chi(t)\Delta t\right)\;\eta(t-\frac{1}{2}\Delta t)+ \\
&\quad\;\Delta t\;\left[3V(t)\frac{\mathcal{P}(t)-P_{\rm ext}}{p_{mass}}+3\frac{2E_{kin}(t)}{f}\frac{1}{p_{mass}}\right] \\
\underline{v}(t+\frac{1}{2}\Delta t) &\leftarrow \underline{v}(t-\frac{1}{2}\Delta t)+\Delta t\;\left[\frac{\underline{f}(t)}{m}-\left\{\chi(t)+\left(1+\frac{3}{f}\right)\eta(t)\right\}\;\underline{v}(t)\right] \\
\underline{r}(t+\Delta t) &\leftarrow \underline{r}(t)+\Delta t\;\left\{\underline{v}(t+\frac{1}{2}\Delta t)-\eta(t+\frac{1}{2}\Delta t)\;\underline{r}(t+\frac{1}{2}\Delta t)\right\}
\end{aligned} \tag{3.169}$$

for the isotropic cell fluctuations case and

$$\begin{aligned}
\underline{\underline{\eta}}(t+\frac{1}{2}\Delta t) &\leftarrow \exp\left(-\chi(t)\Delta t\right)\;\underline{\underline{\eta}}(t-\frac{\Delta t}{2})+ \\
&\quad\;\Delta t\;\left[\frac{\underline{\underline{\sigma}}(t)-P_{\rm ext}\;V(t)\;\underline{\underline{\mathbf{1}}}}{p_{mass}}+\frac{2E_{kin}(t)}{f}\frac{\underline{\underline{\mathbf{1}}}}{p_{mass}}\right] \\
\underline{v}(t+\frac{1}{2}\Delta t) &\leftarrow \underline{v}(t-\frac{1}{2}\Delta t)+\Delta t\;\left[\frac{\underline{f}(t)}{m}-\left\{\chi(t)\;\underline{\underline{\mathbf{1}}}+\underline{\underline{\eta}}(t)+\frac{\texttt{Tr}\left[\underline{\underline{\eta}}(t)\right]}{f}\;\underline{\underline{\mathbf{1}}}\right\}\cdot\underline{v}(t)\right] \\
\underline{r}(t+\Delta t) &\leftarrow \underline{r}(t)+\Delta t\;\left\{\underline{v}(t+\frac{1}{2}\Delta t)+\underline{\underline{\eta}}(t+\frac{1}{2}\Delta t)\cdot\underline{r}(t+\frac{1}{2}\Delta t)\right\}
\end{aligned} \tag{3.170}$$

for the anisotropic cell fluctuations case.

This ensemble is optionally extending to constant normal pressure and constant surface area, NP$_n$AT [73], by semi-isotropic constraining of the barostat equation of motion and slight amending the thermostat equation of motion and the conserved quantity to:

$$\begin{aligned}
\frac{d}{dt}\eta_{\alpha\beta}(t) &= \begin{cases}\frac{\sigma_{zz}(t)-P_{\rm ext}\;V(t)}{p_{mass}}+\frac{2E_{kin}(t)}{f\;p_{mass}}-\chi(t)\eta_{zz}(t) & : \;\;(\alpha=\beta)=z \\ 0\;\;\;;\;\;\;\eta_{\alpha\beta}(0)=0 & : \;\;(\alpha,\beta)\neq z\end{cases} \\
\frac{d}{dt}\chi(t) &= \frac{2E_{kin}(t)+p_{mass}\;\texttt{Tr}[\underline{\underline{\eta}}(t)\cdot\underline{\underline{\eta}}(t)^T]-2\sigma-k_B\;T_{\rm ext}}{q_{mass}} \\
\mathcal{H}_{\rm NP_nAT} &= \mathcal{H}_{\rm NVE}+\frac{q_{mass}\;\chi(t)^2}{2}+\frac{p_{mass}\;\texttt{Tr}[\underline{\underline{\eta}}\cdot\underline{\underline{\eta}}^T]}{2}+P_{\rm ext}V(t)+(f+1)\;k_B\;T_{\rm ext}\;\int_o^t\chi(s)ds\;\;.
\end{aligned} \tag{3.171}$$

Similarly, this ensemble is optionally extending to constant normal pressure and constant surface tension, NP$_n\gamma$T [73], by semi-isotropic constraining of the barostat equation of motion and slight amending the thermostat equation of motion and the conserved quantity to:

$$\frac{d}{dt}\eta_{\alpha\beta}(t) = \begin{cases}\frac{\sigma_{\alpha\alpha}(t)-[P_{\rm ext}-\gamma_{\rm ext}/h_z(t)]\;V(t)}{p_{mass}}+\frac{2E_{kin}(t)}{f}\frac{1}{p_{mass}}-\chi(t)\eta_{\alpha\alpha}(t) & : \;\;(\alpha=\beta)=x,y \\ & : \\ \frac{\sigma_{zz}(t)-P_{\rm ext}\;V(t)}{p_{mass}}+\frac{2E_{kin}(t)}{f\;p_{mass}}-\chi(t)\eta_{zz}(t) & : \;\;(\alpha=\beta)=z \\ 0\;\;\;;\;\;\;\eta_{\alpha\beta}(0)=0 & : \;\;(\alpha\neq\beta)=x,y,z\end{cases}$$

$$\begin{aligned} \frac{d}{dt}\chi(t) &= \frac{2E_{kin}(t) + p_{mass}\ \mathtt{Tr}[\underline{\underline{\eta}}(t)\cdot\underline{\underline{\eta}}(t)^T] - 2\sigma - 3\ k_B\ T_{\rm ext}}{q_{mass}} \\ \mathcal{H}_{\rm NP_n\gamma T} &= \mathcal{H}_{\rm NVE} + \frac{q_{mass}\ \chi(t)^2}{2} + \frac{p_{mass}\ \mathtt{Tr}[\underline{\underline{\eta}}\cdot\underline{\underline{\eta}}^T]}{2} + P_{\rm ext}V(t) + (f+3)\ k_B\ T_{\rm ext}\int_o^t \chi(s)ds\ , \end{aligned} \tag{3.172}$$

where $\gamma_{\rm ext}$ is the user defined external surface tension and $h_z(t) = V(t)/A_{xy}(t)$ is the instantaneous hight of the MD box (or MD box volume over area). One defines the instantaneous surface tension as given in equation (3.122). The case $\gamma_{\rm ext} = 0$ generates the NPT anisotropic ensemble for the orthorhombic cell (imcon=2 in CONFIG, see Appendix B). This can be considered as an "orthorhombic" constraint on the NσT ensemble. The constraint can be strengthened further, to a "semi-orthorhombic" one, by imposing that the MD cell change isotropically in the (x, y) plane which leads to the following changes in the equations above

$$\begin{aligned} \frac{d}{dt}\eta_{\alpha\alpha}(t) &= \frac{[\sigma_{xx}(t)+\sigma_{yy}(t)]/2 - [P_{\rm ext} - \gamma_{\rm ext}/h_z(t)]\ V(t)}{p_{mass}} + \frac{2\ E_{kin}(t)}{f\ p_{mass}} - \chi(t)\eta_{\alpha\alpha}(t)\ : \\ &\qquad\qquad : \ (\alpha = \beta) = x, y \\ \mathcal{H}_{\rm NP_n\gamma=0T} &= \mathcal{H}_{\rm NVE} + \frac{q_{mass}\ \chi(t)^2}{2} + \frac{p_{mass}\ \mathtt{Tr}[\underline{\underline{\eta}}\cdot\underline{\underline{\eta}}^T]}{2} + P_{\rm ext}V(t) + (f+2)\ k_B\ T_{\rm ext}\int_o^t \chi(s)ds\ . \end{aligned} \tag{3.173}$$

Although the Martyna-Tuckerman-Klein equations of motion have same conserved quantities as the Nosé-Hoover's ones they are proven to generate ensembles that conserve the phase space volume and thus have well defined conserved quantities even in presence of forces external to the system [81], which is not the case for Nosé-Hoover NPT and N$\underline{\underline{\sigma}}$T ensembles.

The NPT and N$\underline{\underline{\sigma}}$T versions of the MTK ensemble are implemented in the DL_POLY_4 routines NPT_M0_VV and NST_M0_VV - in VV flavour, and NPT_M0_LFV and NST_M0_LFV - LFV flavour, respectively. The corresponding routines incorporating RB dynamics are NPT_M1_VV and NPT_M1_LFV, and NST_M1_VV and NST_M1_LFV.

3.6 Rigid Bodies and Rotational Integration Algorithms

3.6.1 Description of Rigid Body Units

A rigid body unit is a collection of point entities whose local geometry is time invariant. One way to enforce this in a simulation is to impose a sufficient number of bond constraints between the atoms in the unit. However, in many cases this may be either problematic or impossible. Examples in which it is impossible to specify sufficient bond constraints are

1. linear molecules with more than 2 atoms (e.g. CO_2)
2. planar molecules with more than three atoms (e.g. benzene).

Even when the structure *can* be defined by bond constraints the network of bonds produced may be problematic. Normally, they make the iterative SHAKE in the LFV integration (or RATTLE in the VV integration) procedure slow, particularly if a ring of constraints is involved (as occurs when one defines water as a constrained triangle). It is also possible, inadvertently, to over constrain a molecule (e.g. by defining a methane tetrahedron to have 10 rather than 9 bond constraints) in which case the SHAKE/RATTLE procedure will become unstable. In addition, massless sites (e.g. charge sites) cannot be included in a simple constraint approach making modelling with potentials such as TIP4P water impossible.

All these problems may be circumvented by defining rigid body units, the dynamics of which may be described in terms of the translational motion of the centre of mass (COM) and rotation about the COM.

To do this we need to define the appropriate variables describing the position, orientation and inertia of a rigid body, and the rigid body equations of motion[1].

The mass of a rigid unit M is the sum of the atomic masses in that unit:

$$M = \sum_{j=1}^{N_{sites}} m_j \quad , \tag{3.174}$$

where m_j is the mass of an atom and the sum includes all sites (N_{sites}) in the body. The position of the rigid unit is defined as the location of its centre of mass $\underline{R}$:

$$\underline{R} = \frac{1}{M} \sum_{j=1}^{N_{sites}} m_j \underline{r}_j \quad , \tag{3.175}$$

where $\underline{r}_j$ is the position vector of atom j. The rigid body translational velocity $\underline{V}$ is defined by:

$$\underline{V} = \frac{1}{M} \sum_{j=1}^{N_{sites}} m_j \underline{v}_j \tag{3.176}$$

and its angular momentum $\underline{J}$ can then be defined by the expression:

$$\underline{J} = \sum_{j=1}^{N_{sites}} m_j \left(\underline{d}_j \times \left[\underline{v}_j - \underline{V} \right] \right) \quad , \tag{3.177}$$

where $\underline{v}_j$ is the velocity of atom j and $\underline{d}_j$ is the displacement vector of the atom j from the COM, is given by:

$$\underline{d}_j = \underline{r}_j - \underline{R} \quad . \tag{3.178}$$

The net translational force $\underline{F}$ acting on the rigid body unit is the vector sum of the forces acting on the atoms of the body:

$$\underline{F} = \sum_{j=1}^{N_{sites}} \underline{f}_j \tag{3.179}$$

and the torque vector $\underline{\tau}$ acting on the body in the universal frame of reference is given by:

$$\underline{\tau} = \sum_{j=1}^{N_{sites}} \underline{d}_j \times \underline{f}_j \quad , \tag{3.180}$$

where $\underline{f}_j$ is the force on a rigid unit site.

A rigid body also has associated with it a rotational inertia matrix $\underline{\underline{\mathbf{I}}}$, whose components are given by:

$$I^{\alpha\beta} = \sum_{j=1}^{N_{sites}} m_j (d_j^2 \delta_{\alpha\beta} - d_j^\alpha r_j^\beta) \quad , \tag{3.181}$$

and COM stress and virial respectively written down as:

$$\sigma^{\alpha\beta} = \sum_{j=1}^{N_{sites}} d_j^\alpha f_j^\beta$$

$$\mathcal{W} = - \sum_{j=1}^{N_{sites}} \underline{d}_j \cdot \underline{f}_j \quad , \tag{3.182}$$

[1] An alternative approach is to define "basic" and "secondary" particles. The basic particles are the minimum number needed to define a local body axis system. The remaining particle positions are expressed in terms of the COM and the basic particles. Ordinary bond constraints can then be applied to the basic particles provided the forces and torques arising from the secondary particles are transferred to the basic particles in a physically meaningful way.

where $\underline{d}_j$ is the displacement vector of the atom j from the COM, and is given by:

$$\underline{d}_j = \underline{r}_j - \underline{R} \quad . \tag{3.183}$$

The rigid body angular velocity $\underline{\omega}$ is the scalar product of the moment of inertia, $\underline{\underline{\mathbf{I}}}$, inverse and the angular momentum, $\underline{J}$,:

$$\underline{\omega} = \underline{\underline{\mathbf{I}}}^{-1} \cdot \underline{J} \quad . \tag{3.184}$$

It is common practice in the treatment of rigid body motion to define the position $\underline{R}$ of the body in a universal frame of reference (the so called laboratory or inertial frame), but to describe the moment of inertia tensor in a frame of reference that is localised in the rigid body and changes as the rigid body rotates. Thus the local body frame is taken to be that in which the rotational inertia tensor $\underline{\underline{\hat{\mathbf{I}}}}$ is diagonal and the components satisfy $I_{xx} \geq I_{yy} \geq I_{zz}$. In this local frame (the so called *Principal Frame*) the inertia tensor is therefore constant.

The orientation of the local body frame with respect to the space fixed frame is described via a four dimensional unit vector, the quaternion:

$$\underline{q} = [q_0, q_1, q_2, q_3]^T \quad , \tag{3.185}$$

and the rotational matrix $\underline{\underline{\mathbf{R}}}$ to transform from the local body frame to the space fixed frame is the unitary matrix:

$$\underline{\underline{\mathbf{R}}} = \begin{pmatrix} q_0^2 + q_1^2 - q_2^2 - q_3^2 & 2\ (q_1\ q_2 - q_0\ q_3) & 2\ (q_1\ q_3 + q_0\ q_2) \\ 2\ (q_1\ q_2 + q_0\ q_3) & q_0^2 - q_1^2 + q_2^2 - q_3^2 & 2\ (q_2\ q_3 - q_0\ q_1) \\ 2\ (q_1\ q_3 - q_0\ q_2) & 2\ (q_2\ q_3 + q_0\ q_1) & q_0^2 - q_1^2 - q_2^2 + q_3^2 \end{pmatrix} \tag{3.186}$$

so that if $\hat{\underline{d}}_j$ is the position of an atom in the local body frame (with respect to its COM), its position in the universal frame (w.r.t. its COM) is given by:

$$\underline{d}_j = \underline{\underline{\mathbf{R}}} \cdot \hat{\underline{d}}_j \quad . \tag{3.187}$$

With these variables defined we can now consider the equations of motion for the rigid body unit.

3.6.2 Integration of the Rigid Body Equations of Motion

The equations of translational motion of a rigid body are the same as those describing the motion of a single atom, except that the force is the total force acting on the rigid body i.e. $\underline{F}$ in equation (3.179) and the mass is the total mass of the rigid body unit i.e. M in equation (3.174). These equations can be integrated by the standard Verlet (LFV or VV) algorithms described in the previous sections. Thus we need only consider the rotational motion here.

The rotational equation of motion for a rigid body relates the torque to the change in angular momentum:

$$\underline{\tau} = \frac{d}{dt}\underline{J} = \frac{d}{dt}\left(\underline{\underline{\mathbf{I}}} \cdot \underline{\omega}\right) \quad . \tag{3.188}$$

In a thermostat it can be written as:

$$\dot{\underline{J}}_i = \underline{\tau}_i - \underline{\omega}_i \times \underline{J}_i + \frac{\chi}{q_{mass}}\underline{J}_i \quad , \tag{3.189}$$

where i is the index of the rigid body, χ and q_{mass} are the thermostat friction coefficient and mass. In the local frame of the rigid body and without the thermostat term, these simplify to the Euler's equations

$$\begin{aligned} \dot{\hat{\omega}}_x &= \frac{\hat{\tau}_x}{\hat{I}_{xx}} + (\hat{I}_{yy} - \hat{I}_{zz})\ \hat{\omega}_y\ \hat{\omega}_z \\ \dot{\hat{\omega}}_y &= \frac{\hat{\tau}_y}{\hat{I}_{yy}} + (\hat{I}_{zz} - \hat{I}_{xx})\ \hat{\omega}_z\ \hat{\omega}_z \\ \dot{\hat{\omega}}_z &= \frac{\hat{\tau}_z}{\hat{I}_{zz}} + (\hat{I}_{xx} - \hat{I}_{yy})\ \hat{\omega}_x\ \hat{\omega}_y \quad . \end{aligned} \tag{3.190}$$

The vectors $\underline{\hat{\tau}}$ and $\underline{\hat{\omega}}$ are the torque and angular velocity acting on the body transformed to the local body frame. Integration of $\underline{\hat{\omega}}$ is complicated by the fact that as the rigid body rotates, so does the local reference frame. So it is necessary to integrate equations (3.190) simultaneously with an integration of the quaternions describing the orientation of the rigid body. The equation describing this is:

$$\begin{pmatrix} \dot{q}_0 \\ \dot{q}_1 \\ \dot{q}_2 \\ \dot{q}_3 \end{pmatrix} = \frac{1}{2} \begin{pmatrix} q_0 & -q_1 & -q_2 & -q_3 \\ q_1 & q_0 & -q_3 & q_2 \\ q_2 & q_3 & q_0 & -q_1 \\ q_3 & -q_2 & q_1 & q_0 \end{pmatrix} \begin{pmatrix} 0 \\ \hat{\omega}_x \\ \hat{\omega}_y \\ \hat{\omega}_z \end{pmatrix} \quad . \tag{3.191}$$

Rotational motion in DL_POLY_4 is handled by two different methods. For the LFV implementation, the Fincham Implicit Quaternion Algorithm (FIQA) is used [24]. The VV implementation uses the NOSQUISH algorithm of Miller *et al.* [25]. The implementation of FIQA is coded in Q_UPDATE and NOSQUSH in NO_SQUISH both contained within QUATERNION_CONTAINER.

The LFV implementation begins by integrating the angular velocity equation in the local frame:

$$\underline{\hat{\omega}}(t + \frac{\Delta t}{2}) = \underline{\hat{\omega}}(t - \frac{\Delta t}{2}) + \Delta t \, \underline{\underline{\hat{\mathbf{I}}}}^{-1} \cdot \underline{\dot{\hat{\omega}}}(t) \quad . \tag{3.192}$$

The new quaternions are found using the FIQA algorithm. In this algorithm the new quaternions are found by solving the implicit equation:

$$\underline{q}(t + \Delta t) = \underline{q}(t) + \frac{\Delta t}{2} \left(\underline{\underline{\mathbf{Q}}} \, [\underline{q}(t)] \cdot \underline{\hat{w}}(t) + \underline{\underline{\mathbf{Q}}} \, [\underline{q}(t + \Delta t)] \cdot \underline{\hat{w}}(t + \Delta t) \right) \quad , \tag{3.193}$$

where $\underline{\hat{w}} = [0, \underline{\hat{\omega}}]^T$ and $\underline{\underline{\mathbf{Q}}}[\underline{q}]$ is:

$$\underline{\underline{\mathbf{Q}}} = \frac{1}{2} \begin{pmatrix} q_0 & -q_1 & -q_2 & -q_3 \\ q_1 & q_0 & -q_3 & q_2 \\ q_2 & q_3 & q_0 & -q_1 \\ q_3 & -q_2 & q_1 & q_0 \end{pmatrix} \quad . \tag{3.194}$$

The above equation is solved iteratively with

$$\underline{q}(t + \Delta t) = \underline{q}(t) + \Delta t \, \underline{\underline{\mathbf{Q}}}[\underline{q}(t)] \cdot \underline{\hat{w}}(t) \tag{3.195}$$

as the first guess. Typically, no more than 3 or 4 iterations are needed for convergence. At each step the normalisation constraint:

$$\| \underline{q}(t + \Delta t) \| = 1 \tag{3.196}$$

is imposed.

While all the above is enough to build LFV implementations, the VV implementations, based on the NOSQUISH algorithm of Miller *et al.* [25], also require treatment of the quaternion momenta as defined by:

$$\begin{pmatrix} p_0 \\ p_1 \\ p_2 \\ p_3 \end{pmatrix} = 2 \begin{pmatrix} q_0 & -q_1 & -q_2 & -q_3 \\ q_1 & q_0 & -q_3 & q_2 \\ q_2 & q_3 & q_0 & -q_1 \\ q_3 & -q_2 & q_1 & q_0 \end{pmatrix} \begin{pmatrix} 0 \\ \hat{I}_{xx} \, \hat{\omega}_x \\ \hat{I}_{yy} \, \hat{\omega}_y \\ \hat{I}_{zz} \, \hat{\omega}_z \end{pmatrix} \quad , \tag{3.197}$$

and quaternion torques as defined by:

$$\begin{pmatrix} \Upsilon_0 \\ \Upsilon_1 \\ \Upsilon_2 \\ \Upsilon_3 \end{pmatrix} = 2 \begin{pmatrix} q_0 & -q_1 & -q_2 & -q_3 \\ q_1 & q_0 & -q_3 & q_2 \\ q_2 & q_3 & q_0 & -q_1 \\ q_3 & -q_2 & q_1 & q_0 \end{pmatrix} \begin{pmatrix} 0 \\ \hat{\tau}_x \\ \hat{\tau}_y \\ \hat{\tau}_z \end{pmatrix} \quad . \tag{3.198}$$

It should be noted that vectors $\underline{p}$ and $\underline{\Upsilon}$ are 4-component vectors. The quaternion momenta are first updated a half-step using the formula:

$$\underline{p}(t + \frac{\Delta t}{2}) \leftarrow \underline{p}(t) + \frac{\Delta t}{2} \underline{\Upsilon}(t) \quad . \tag{3.199}$$

Next a sequence of operations is applied to the quaternions and the quaternion momenta in the order:

$$e^{i\mathcal{L}_3(\delta t/2)}\ e^{i\mathcal{L}_2(\delta t/2)}\ e^{i\mathcal{L}_1(\delta t)}\ e^{i\mathcal{L}_2(\delta t/2)}\ e^{i\mathcal{L}_3(\delta t/2)}\quad , \tag{3.200}$$

which preserves the symplecticness of the operations (see reference [32]). Note that δt is some submultiple of Δt. (In DL_POLY_4 the default is $\Delta t = 10\delta t$.) The operators themselves are of the following kind:

$$\begin{aligned} e^{i\mathcal{L}(\delta t)}\ \underline{q} &= \cos(\zeta_k \delta t)\ \underline{q} + \sin(\zeta_k \delta t)\ P_k\ \underline{q} \\ e^{i\mathcal{L}(\delta t)}\ \underline{p} &= \cos(\zeta_k \delta t)\ \underline{p} + \sin(\zeta_k \delta t)\ P_k\ \underline{p}\quad , \end{aligned} \tag{3.201}$$

where P_k is a permutation operator with $k = 0, \ldots, 3$ with the following properties:

$$\begin{aligned} P_0\ \underline{q} &= \{\ q_0,\ q_1,\ q_2,\ q_3\} \\ P_1\ \underline{q} &= \{-q_1,\ q_0,\ q_3, -q_2\} \\ P_2\ \underline{q} &= \{-q_2, -q_3,\ q_0,\ q_1\} \\ P_3\ \underline{q} &= \{-q_3,\ q_2, -q_1,\ q_0\}\quad , \end{aligned} \tag{3.202}$$

and the angular velocity ζ_k is defined as:

$$\zeta_k = \frac{1}{4I_k}\underline{p}^T P_k\ \underline{q}\quad . \tag{3.203}$$

Equations (3.200-3.202) represent the heart of the NOSQUISH algorithm and are repeatedly applied (10 times in DL_POLY_4). The final result is the quaternion updated to the full timestep value i.e. $\underline{q}(t + \Delta t)$. These equations form part of the first stage of the VV algorithm (VV1).

In the second stage of the VV algorithm (VV2), new torques are used to update the quaternion momenta to a full timestep:

$$\underline{p}(t+\Delta t) \leftarrow \underline{p}(t+\frac{\Delta t}{2}) + \frac{\Delta t}{2}\underline{\Upsilon}(t+\Delta t)\quad . \tag{3.204}$$

3.6.3 Thermostats and Barostats coupling to the Rigid Body Equations of Motion

In the presence of rigid bodies in the atomic system the system's instantaneous pressure, equation (3.97):

$$\mathcal{P}(t) = \frac{[2E_{kin}(t) - \mathcal{W}_{\rm atomic}(t) - \mathcal{W}_{\rm COM}(t) - \mathcal{W}_{\rm constrain}(t-\Delta t) - \mathcal{W}_{\rm PMF}(t-\Delta t)]}{3V(t)} \tag{3.205}$$

and stress, equation (3.98):

$$\underline{\underline{\sigma}}(t) = \underline{\underline{\sigma}}_{kin}(t) + \underline{\underline{\sigma}}_{\rm atomic}(t) + \underline{\underline{\sigma}}_{\rm COM}(t) + \underline{\underline{\sigma}}_{\rm constrain}(t-\Delta t) + \underline{\underline{\sigma}}_{\rm PMF}(t-\Delta t) \tag{3.206}$$

are augmented to include the RBs' COM virial and stress contributions. **Note** that the kinetic energy and stress in the above also include the contributions of the RB's COM kinetic energy and stress!

It is straightforward to couple the rigid body equations of motion to a thermostat and/or barostat. The thermostat is coupled to both the translational and rotational degrees of freedom and so both the translational and rotational velocities are thermostated in the same manner as the purely atomic velocities. The barostat, however, is coupled only to the translational degrees of freedom and does not contribute to the rotational motion. Therefore, if we notion the change of the system's degrees of freedom as:

$$f \to F = f + f^{RB(\mathbf{tra})} + f^{RB(\mathbf{rot})} \tag{3.207}$$

then all equations of motion defining the ensembles as described in this chapter are subject to the following notional changes in order to include the RB contributions:

$$\begin{array}{rcccl} \sigma(f) & \to & \sigma(F) & = & \sigma(f + f^{RB(\mathbf{tra})} + f^{RB(\mathbf{rot})}) \\ \mathcal{H}(f) & \to & \mathcal{H}(F) & = & \mathcal{H}(f + f^{RB(\mathbf{tra})} + f^{RB(\mathbf{rot})}) \\ p_{mass}(f) & \to & p_{mass}(F - f^{RB(\mathbf{rot})}) & = & p_{mass}(f + f^{RB(\mathbf{tra})}) \\ \eta(f) & \to & \eta(F - f^{RB(\mathbf{rot})}) & = & \eta(f + f^{RB(\mathbf{tra})}) \\ \underline{\underline{\eta}}(f) & \to & \underline{\underline{\eta}}(F - f^{RB(\mathbf{rot})}) & = & \underline{\underline{\eta}}(f + f^{RB(\mathbf{tra})})\quad , \end{array} \tag{3.208}$$

where f refers to the degrees of freedom in the system (see equation (3.11)), σ is the system target energy (see equation (3.57)), $\mathcal{H}$ is the conserved quantity of the ensemble (if there is such defined), E_{kin} (!includes RB COM kinetic energy too) and E_{rot} are respectively the kinetic and rotational energies of the system, p_{mass} is the barostat mass, and η and $\underline{\underline{\eta}}$ are the barostat friction coefficient or matrix of coefficients respectively.

There are two slight technicalities with the Evans and Andersen ensembles that are worth mentioning.

Since both the translational and rotational velocities contribute towards temperature, equation (3.24), showing the derivation of the thermostat friction in the Evans ensemble by imposing a Gaussian constraint on the system's instantaneous temperature, changes to:

$$\begin{aligned}
\frac{d}{dt}\mathcal{T} = 0 \qquad &\propto \qquad \frac{d}{dt}\left(\frac{1}{2}\sum_i^{FP} m_i \underline{v}_i^2 + \frac{1}{2}\sum_j^{RB} M_j \underline{V}_j^2 + \frac{1}{2}\sum_j^{RB} \underline{\hat{\omega}}_j^T \cdot \underline{\underline{\hat{\mathbf{I}}}}_j \cdot \underline{\hat{\omega}}_j\right) = 0 \\
&\left\{\sum_i^{FP} \underline{v}_i(t)\cdot\underline{f}_i(t) + \sum_j^{RB} \underline{V}_j(t)\cdot\underline{F}_j(t) + \sum_j^{RB} \underline{\hat{\omega}}_j(t)\cdot\underline{\hat{\tau}}(t)\right\} - \\
&\qquad \chi(t)\left\{\sum_i^{FP} m_i \underline{v}_i^2(t) + \sum_j^{RB} M_j \underline{V}_j^2(t) + \sum_j^{RB} \underline{\hat{\omega}}_j^T(t)\cdot\underline{\underline{\hat{\mathbf{I}}}}_j\cdot\underline{\hat{\omega}}_j(t)\right\} = 0 \\
\chi(t) &= \frac{\left\{\sum_i^{FP} \underline{v}_i(t)\cdot\underline{f}_i(t) + \sum_j^{RB} \underline{V}_j(t)\cdot\underline{F}_j(t) + \sum_j^{RB} \underline{\hat{\omega}}_j(t)\cdot\underline{\hat{\tau}}(t)\right\}}{2\left[E_{kin}(t) + E_{rot}(t)\right]} \quad ,
\end{aligned} \tag{3.209}$$

where where $\mathcal{T}$ is the instantaneous temperature defined in equation (3.10) and E_{kin} in the final expression contains both the kinetic contribution form the free particles and the RBs' COMs.

In the case of the Andersen ensemble, if a Poisson selected particle constitutes a RB then the whole RB is Poisson selected. Poisson selected RBs' translational and angular velocities together with Poisson selected FPs' velocities sample the same Gaussian distribution isokinetically (Boltzmann distribution), where the isokineticity to target temperature is dependent upon the total of the Poisson selected FPs' and RBs' degrees of freedom.

Chapter 4

Coarse Graining Functionality

Scope of Chapter

This chapter describes the coarse-graining functionality available in DL_POLY_4.

4.1 User-Defined Coarse-Grain Models with Tabulated Force-Fields

One can use DL_POLY_4 for preparing and running simulations of (numerically) coarse-grained (CG) models by using tabulated effective force-fields (FF) derived from either **(i)** potentials of mean force (PMF); or **(ii)** iteratively optimised CG models.

In outline, systematic coarse-graining (SCG) of an atomistic system implies the application of a geometrical projection, or "mapping", of the original system onto a considerably reduced set of degrees of freedom (DoF), whence referred to as a "coarse-grained" model. The procedure is to recover the average configurational (often termed "physical") and topological (often termed "chemical") force-field related properties of the original full-atom (FA) model. Integrating out degrees of freedom ultimately leads to loss of information. However, if this is done selectively and consistently, the intrinsic information pertaining to the dominating interactions within the FA system (that govern its behaviour towards phenomena of interest) and the most important thermodynamic properties are retained. Then the greatly reduced phase-space of the CG model system allows for efficient access to much longer length and time scales for investigating microscopic phenomena using of the well-established classical MD machinery. The reduced DoF mapping leads to the generation of an effective CG FF in terms of the effective interaction potentials and forces between the CG particles, which are often tabulated numerically.

The initial coarse-grain mapping of the original FA trajectory can be done with the aid of DL_CGMAP tool – http://www.ccp5.ac.uk/software/. After the CG mapped trajectory has been obtained, the relevant distribution analysis and the Boltzmann Inversion procedure (producing tabulated PMF:s) can be performed by using DL_POLY_4 in "replay history" mode, with the corresponding directives in CONTROL file. For further iterative optimisation of the CG model the user is advised to use DL_POLY_4 as simulation engine within the framework of VOTCA package – http://www.votca.org/, which provides a handful of various systematic coarse-graining methodologies. For more details the user should refer to the manuals of these tools.

This section describes how to use DL_POLY_4 for two SCG tasks:

- Post-simulation analysis of the intramolecular (bonded and angular) mean-force interactions, based on the calculation of the corresponding intramolecular distributions.
- Preparation and use of the tabulated intramolecular potentials.

Note: Although these steps are also applicable to atomistic systems, which can be useful for benchmarking purposes, below we shall assume that the CG mapping has been done and the following data files have been generated for the CG mapped model system: CONTROL, FIELD, CONFIG, andi, alternatively, HISTORY.

4.2 Intramolecular Probability Distribution Function (PDF) Analysis

Albeit the distribution analysis can be performed on the fly, in the course of a simulation run, for CG purposes it has to be done on an existing CG mapped trajectory, by invoking the **replay history** and **analysis** directives, see Section 6.1.1. To trigger the PDF collection and subsequent output of PMF:s for any of the following intarmolecular interactions: bonds, angles, dihedrals and/or inversions, the CONTROL file must contain any combination of the options demonstrated in the example below.

```
TITLE: EXAMPLE OF DL_POLY_4 PDF ANALYSIS DIRECTIVES SNIPPET

# DIRECTIVES TO INVOKE INTRAMOLECULAR PDF ANALYSIS BY TYPE
analyse  bonds      sample every 100  nbins 250  rmax 5.0
analyse  angles     sample every 100  nbins 360  # [ 0 : pi]
analyse  dihedrals  sample every 100  nbins 720  # [-pi: pi]
analyse  inversions  sample every 100  nbins 360  # [ 0 : pi]

# DIRECTIVES TO INVOKE INTRAMOLECULAR PDF ANALYSIS FOR ALL TYPES
analyse  all         sample every 100  nbins 1000  rmax 5.0

# DIRECTIVES TO INVOKE PRINTING FOR ANY DEFINED
# INTER-(RDF->VDW) & INTRA-(bonded) MOLECULAR PDF ANALYSIS
print analysis
```

The **analyse** directive acts in a similar manner to that of the **rdf** directive outlining **(i)** the frequency of sampling (in steps/frames) and **(ii)** the number of bins over the cutoff interval of the specified interaction. It is only for bonded pairs that this cutoff needs specifying, whereas for angles the possible ranges are known *a priory* by definition. If no cutoff is supplied for bonds, then it defaults to 2 Å. It is worth noting that, for the sake of accuracy, the number of bins per PDF must be larger than or equal to $N_{\rm min} = \texttt{nint}\left(\frac{\rm PDF\ cutoff}{\Delta_{\rm max}}\right)$, where the max bin size, $\Delta_{\rm max}$, is defined internally for each type of PDF (see `setup_module.f90`). Otherwise, it will default to $\texttt{max}(N_{\rm min}, N_{\rm tab})$, where $N_{\rm tab}$ is the number of bins on the grid defined in the corresponding tabulated force-field data file (TAB*), if it is provided, otherwise $N_{\rm tab} = 0$. In particular, for bonds $\Delta_{\rm max}$ is 0.01 Å and for angles it is $0.2\pi/180$, i.e. 0.2 degree.

If **analyse all** option is used in conjunction with any specific directive, then it triggers the analysis on all PDFs while enforcing on the individually targeted PDF(s) the following parameters: **(i)** its sampling interval (frames) – only if it is smaller than, and **(ii)** its grid number – only if it is larger than those specified for the targeted PDFs. In the case of bonds it will also enforce the grid range **rmax** – only if it is larger than that specified in the **analyse bonds** directive. Hence, the **analyse all** directive allows to quickly override and/or unify the sampling frequencies and max grid numbers for all PDFs, provided its parameters improve on the accuracy of the collected data (compared to the specifications for the individually targeted PDF:s).

While the statistics are always collected and stored for the future use in the (binary) REVIVE file for all targeted PDF:s (see Section 6.2.8), using the **print analysis** directive will instruct DL_POLY_4 to additionally print the data in the OUTPUT and *DAT files, the latter containing each *type* of PDF:s separately (the asterisk stands for one of the following: BND, ANG, DIH, or INV). As a result, apart from OUTPUT, three data files will be created for each of the targeted distribution types: BNDDAT, BNDPMF & BNDTAB – for bonds, ANGDAT, ANGPMF & ANGTAB – for angles, DIHDAT, DIHPMF & DIHTAB – for dihedrals, and INVDAT, INVPMF & INVTAB – for inversions; the *PMF and *TAB files containing tabulated data for the respective potential of mean force and force/virial (*TAB files bearing the data from *PMF files but *resampled onto a finer grid*; see the last paragraph in this section).

Partial examples of the *DAT files for bonds and angles are given below.

```
[user@host]$ more BNDDAT
# TITLE: Hexane FA OPLSAA -> CG mapped with 3 beads (A-B-A)
# BONDS: Probability Density Functions (PDF) := histogram(bin)/hist_sum(bins)/dr_bin
# bins, cutoff, frames, types:        250    5.000       2285           1
#
# r(Angstroms)  PDF_norm(r)  PDF_norm(r)/dVol(r)   @   dr_bin =  0.02000
#

# type, index, instances: A        B                   1       2000
    0.01000  0.000000E+00  0.000000E+00
    0.03000  0.000000E+00  0.000000E+00
    0.05000  0.000000E+00  0.000000E+00
...
    4.95000  0.000000E+00  0.000000E+00
    4.97000  0.000000E+00  0.000000E+00
    4.99000  0.000000E+00  0.000000E+00

[user@host]$ more ANGDAT
# TITLE: Hexane FA OPLSAA -> CG mapped with 3 beads (A-B-A)
# ANGLES: Probability Density Functions (PDF) := histogram(bin)/hist_sum(bins)/dTheta_bin
# bins, cutoff, frames, types:        360        180       2285          1
#
# Theta(degrees)  PDF_norm(Theta)  PDF_norm(Theta)/sin(Theta)   @   dTheta_bin =  0.50000
#

# type, index, instances: A        B        A                  1       1000
    0.25000  0.000000E+00  0.000000E+00
    0.75000  0.000000E+00  0.000000E+00
    1.25000  0.000000E+00  0.000000E+00
...
  178.75000  1.380569E-02  6.328564E-01
  179.25000  8.368490E-03  6.393238E-01
  179.75000  2.901532E-03  6.649842E-01
```

One can see that all the header lines are commented out due to starting with the hash symbol, #. Nonetheless, the header contains some useful information. The title is, as usual, placed in the first line, which is followed by an explanatory line with the definition of a normalised PDF. The third line provides the four most important descriptors: the number of *bins* on the histogram grid, the *cutoff* interval (absolute value of the span) over which the distributions are sampled, the number of *frames* (samples) used, and the number of unique unit *types* analysed (where "unit" is one of the following: bonds, angles, dihedrals or inversions). The last explanatory line in the header, found in between two empty commented-out lines, defines the meaning of the columns and, at the end, the grid bin size.

The PDF histograms within a *DAT file are separated by uncommented empty lines, which makes it possible to directly import and plot all the data as separate lines in [Xm]Grace 2D plotter. For clarity, the data of each histogram are preceded by a commented-out line specifying the *type*, *index* and number of *instances* of the analysed interaction unit (the latter being counted over the entire system).

In all *DAT files the first column bears the *bin-centered* abscissa values (distance or angle), the second column is the distribution histogram normalized to unity, i.e. its *integral* equals 1, whereas the third column, if any, contains the PDF data corrected for the volumetric (entropic) degeneracy of the grid points. Thus, it is the data of the last column found that are used for calculating PMF $\sim -\ln(\text{PDF})$.

The OUTPUT file will contain copies of the first and second columns of all collected PDF:s, but normalised so that the figures in the second column *sum up* to 1, which can be checked by examining the third column as it bears the running sum of a PDF histogram.

In addition to the PDF:s, DL_POLY_4 also calculates the respective PMF:s and pairwise force dependencies (virial for bonds), which are stored in the *PMF and, upon resampling onto a *bin-edge* grid, *TAB files. **It is important to note** that, unlike the *PMF files containing the bare $-\ln(\text{PDF})$ data (converted to the requested energy units) *on the same grid as the PDF histograms*, the force-field data in *TAB files are *resampled* onto $\max(N_{\min}, N_{\text{tab}})$ grid points located at the bin edges (as expected by DL_POLY_4 when reading the potential and force tables), where N_{tab} is the grid number for the respective intramolecular unit type read-in from its TAB* file, if provided (otherwise $N_{\text{tab}} = 0$). Thus, the *TAB files obey the DL_POLY_4 format for numerically defined intramolecular force-field tables (TAB*, see below) and, hence, can be directly used as such upon renaming: BNDTAB → TABBND, ANGTAB → TABANG, DIHTAB → TABDIH and INVTAB → TABINV. **The user is, however, strongly advised to check the quality of the obtained tabulated force-fields before using those as input for a CG simulation.** Albeit DL_POLY_4 uses a simple smoothing algorithm for noise-reduction in PMF:s and implements capping of the forces in the regions of zero-valued PDF:s, in undersampled regions the PDF and PMF data are likely to suffer from inaccuracy and increased noise which, most often, require extra attention and re-fitting manually.

The general format of the above discussed files is shown in Section 6.2.12

4.3 Setting up Tabulated Intramolecular Force-Field Files

For a user-defined, e.g. coarse-grained, model system the effective potentials must be provided in a tabulated form. For non-bonded short-range (VdW) interactions the TABLE file must be prepared as described in Section 6.1.6. However, the tabulated data format for intramolecular interactions (bonds, bending angles, dihedral and inversion angles in a polymer) differs from that of the TABLE file and assumes three columns: abscissa (distance in Å or angle in degrees), and two ordinates: potential and force data (virial for distance dependent interactions – e.g. bonds, and force for angle dependent interactions – e.g. angles). Shown below are examples of TABBND and TABANG files, corresponding to the above PDF examples. **Note** that the PMF and force data have been resampled onto a finer grid with points located *at bin edges*.

```
[user@host]$ more TABBND
# TITLE: Hexane FA OPLSAA -> CG mapped with 3 beads (A-B-A)
# 5.0 500

# A B
 1.00000e-02  -9.0906600e+02   1.3954000e+00
 2.00000e-02  -9.1046200e+02   2.7908000e+00
 3.00000e-02  -9.1185700e+02   4.1862000e+00
...

[user@host]$ more TABANG
# TITLE: Hexane FA OPLSAA -> CG mapped with 3 beads (A-B-A)
# 1000

# A B A
 1.80000e-01   8.8720627e+01   6.9119576e-01
 3.60000e-01   8.8596227e+01   6.9119227e-01
 5.40000e-01   8.8471827e+01   6.9118704e-01
...
```

The input tables for bonds, angles, dihedrals and inversions are named TABBND, TABANG, TABDIH and TABINV, correspondingly. The format of these files is fixed in terms of the line, or *record*, order. In particular, the initial two header lines must contain a title and a record with the grid specification, and each of the following blocks of tabulated data must be preceded by an empty line and a one-line descriptor record containing the (white-space delimited) names of the atoms making up the given intermolecular interaction unit (same as a *unit type* in *DAT files). These descriptor lines can be commented out or not, i.e. having # as the first symbol would not affect the reading operation, but it would ease importing of the data for plotting and manipulating in [Xm]Grace software.

In all TAB* files the number of grid points (bins) must be specified on the second line (commented-out or not). For angles (TABANG, TABDIH, TABINV) no other information needs to be provided as their ranges are pre-determined: $0 < \Theta$ in TABANG & TABINV ≤ 180 and $-180 < \Theta$ in TABDIH ≤ 180. In the TABBND file, however, the "bond cutoff", $r_{\max}$ (Å), must be precede the grid number.

Note that all potential and force data are to be provided in the same energy units as specified by the user in the FIELD file (see Section 6.1.3) with distances in Å and angles in degrees. All the data related to angles are internally transformed and handled by DL_POLY_4 with angles measured in *radians*.

Finally, in order to instruct DL_POLY_4 to use tabulated intramolecular force-fields read from the TAB* files the user has to specify in the FIELD file the keyword **tab** or **-tab** for each intramolecular interaction thereby chosen for tabulation (similarly to how it is described in Section 6.1.3). The dash symbol (-) in front of the keyword **tab** is only valid for bonds and angles, and is interpreted in the same manner as in Table 6.8 and Table 6.9.

Note that VOTCA package is also capable of collecting both intra- and inter-molecular stats and producing correct TAB* files, provided the FIELD and HISTORY files exist (albeit VOTCA saves the distributions in a format different from *DAT files).

Below we summarise the sequence of operations the user has to follow in order to perform the CG distribution analysis and prepare the TAB* files for a newly coarse-grained system.

- Perform CG mapping of the original FA system with the aid of DL_CGMAP or VOTCA (in the case of using VOTCA follow its manual; the remainder of the list describes using DL_POLY_4 only);
- Move the data files for the newly CG-mapped system (FIELD_CG, CONFIG_CG, HISTORY_CG) into a separate directory under the standard names (FIELD, CONFIG, HISTORY) and create the corresponding CONTROL file containing the **analysis** and **replay history** directives.
- As no TAB* files are yet available for the CG system at this stage, one can either **(i)** create the initial TAB* files padded with zeros, or **(ii)** use a FIELD file with fictitious records for all the interactions to be tabulated, with the interaction keywords and parameters chosen arbitrarily in accord with Table 6.8, Table 6.9, Table 6.10 and Table 6.11. **Note** that DL_CGMAP (as well as VOTCA) creates FIELD_CG files that already contain interaction descriptors with the **tab** keyword(s) in place, so if the route *(ii)* is chosen, the user needs to replace those records with the fictitious ones (it is advisory to store the initial FIELD_CG for the future use, when the actual TAB* files are ready).
- Run DL_POLY_4 with the **replay history**, **rdf** and/or **analysis** options invoked in the CONTROL file, which will result in creation of the targeted inter- and intra-molecular PDF data files (RDFDAT, BNDDAT, ANGDAT, DIHDAT, INVDAT) and the respective PMF files as described above. **Note** that only and only when the **rdf** and **analysis** options are both active then the VDWPMF and VDWTAB files (derived from RDF:s) will be produced, along with RDFDAT. They are structured in the same manner and format as their intramolecular counterparts. The user can then convert the VDWTAB file into a correctly formatted TABLE file by using the utility called `pmf2tab.f` (subject to compilation; found in DL_POLY_4 directory `utility`) as follows.

  ```
  [user@host]$ pmf2tab.exe < VDWTAB
  ```

- Check the data for accuracy and amend the tabulated force-fields. Redo the analysis on coarser/finer grid(s), if necessary.
- When satisfied with the created TAB* files, run a DL_POLY_4 simulation for the prepared CG (or simply user-defined) model system.

Chapter 5

Construction and Execution

Scope of Chapter

This chapter describes how to compile a working version of DL_POLY_4 and run it.

5.1 Constructing DL_POLY_4: an Overview

5.1.1 Constructing the Standard Versions

DL_POLY_4 was designed as a package of useful subroutines rather than a single program, which means that users are to be able to construct a working simulation program of their own design from the subroutines available, which is capable of performing a specific simulation. However, we recognise that many, perhaps most, users will be content with creating a standard version that covers all of the possible applications and for this reason we have only provided the necessary tools to assemble such a version. The method of creating the standard version is described in detail in this chapter, however a brief step-by-step description follows.

1. DL_POLY_4 is supplied as a UNIX compressed file (tarred and gzipped). This must uncompressed and un-tared to create the DL_POLY_4 directory (Section 1.4).
2. In the *build* subdirectory you will find the required DL_POLY_4 makefiles (see Section 5.2.1 and Appendix **??**, where the main Makefiles are listed). This must be copied into the subdirectory containing the relevant source code. In most cases this will be the *source* subdirectory.
3. The chosen makefile is executed with an appropriate keyword (Section 5.2.1) which is selected for specific platforms. For DL_POLY_4 compilation in parallel mode (a FORTRAN90 compiler and an MPI implementation for the specific machine architecture are required) in many cases the user (sometimes with help from the administrator of their platform) will have to create their own keyword entry in the makefile due to the large variety of (i) software needed for the compilation of DL_POLY_4 and (ii) places where it could be installed (PATHS). To facilitate the user with the construction of their own keyword entry, examples are provided in the makefiles. In the case when users use a makefile for DL_POLY_4 compilation in serial mode they will have to provide a valid PATH to the FORTRAN90 compiler on their specific platform.
4. The makefile produces the executable version of the code, which as a default will be named DLPOLY.Z and located in the *execute* subdirectory.
5. DL_POLY_4 also has a Java GUI. The files for this are stored in the subdirectory *java*. Compilation of this is simple and requires running the javac compiler and the jar utility. Details for these procedures are provided in the GUI manual [21].
6. To run the executable for the first time you require the files CONTROL, FIELD and CONFIG (and possibly TABLE - if you have tabulated van der Walls potentials, TABEAM - if you have tabulated metal potentials and REFERENCE - if defect detection is opted for). These must be present in the directory from which the program is executed. (See Section 6.1 for the description of the input files.)
7. Executing the program will produce the files OUTPUT, STATIS, REVCON and REVIVE (and optionally HISTORY, RDFDAT, ZDNDAT, MSDTMP, REFERENCE, DEFECTS) in the executing directory. (See Section 6.2 for the description of the output files.)

This simple procedure is enough to create a standard version to run most simulations. There may however be some difficulty with array sizes. DL_POLY_4 contains features which allocate arrays after scanning the input files for a simulation. Sometimes these initial estimates are insufficient for a long simulation when, for example, the system volume changes markedly during the simulation or when a system is artificially constructed to have a non-uniform density. Usually, simply restarting the program will cure the problem, but sometimes, especially when the local atom density is somewhat higher than the global one or the system undergoes some form of clustering and the distribution of bonded-like interactions is far from uniform, it may be necessary to amend the array sizes in accordance with the error message obtained. To trigger lengthening of the density dependent global arrays the user may use the **densvar** option in the CONTROL (Section 6.1.1) file. However, lengthening these arrays will require a larger amount of memory from the execution machine for the simulation, which it may not be able to provide. See Section 7.2.2 for more insight on the DL_POLY_4 source code structure.

5.1.2 Constructing Non-standard Versions

In constructing a non-standard DL_POLY_4 simulation program, the first requirement is for the user to write a program to function as the root segment. The root segment /VV/DL_POLY is placed in the *source* directory and contains the set-up and close-down calls for a molecular dynamics simulation. It is the routine that first opens the OUTPUT file (Section 6.2), which provides the summary of the job. The root program calls the "molecular dynamics cycle" routines /LFV/MD_LFV or /LFV/MD_VV implementing the VV and LFV depending on which integrator has been specified for the simulation. These routines contain major routines required to perform the simulation, control the normal "molecular dynamics cycle" and monitor the *cpu* and *memory* usage. They also bring about a controlled termination of the program if the *cpu* usage approaches the allotted job time within a pre-set closure time and/or if the *memory* usage approaches the allocated limit for density dependent arrays. Users are recommended to study the aforementioned root directories as a model for other implementations of the package they may wish to construct. The dependencies and calling hierarchies of all the DL_POLY_4 subroutines can be found in Section 7.2.2.

Should additional functionality be added to DL_POLY_4 by the user, the SET_BOUNDS routine (and its support subroutines) may need modifying to allow specification of the dimensions of any new arrays.

Any molecular dynamics simulation performs five different kinds of operation: initialisation; forces calculation; integration of the equations of motion; calculation of system properties; and job termination. It is worth considering these operations in turn and to indicate which DL_POLY_4 routines are available to perform them. We do not give a detailed description, but provide only a guide. Readers are recommended to examine the different routines described in the DL_POLY_4 User Manual for further details (particularly regarding further dependencies i.e. additional routines that may be called).

The following outline assumes a system containing flexible molecules held together by rigid bonds.

Initialisation requires firstly that the program determine what platform resources are made available to the specific simulation job. This is done by the DL_POLY_4 routine MAP_DOMAINS in DOMAINS_MODULE that attempts to allocate and map the resources (nodes in parallel) in compliance with the DD strategy. MAP_DOMAINS is called within the routine SET_BOUNDS, which also sets the necessary limits for various simulation array sizes and all global variables as declared in SETUP_MODULE to convenient values based on a rough scan through the CONFIG, CONTROL, FIELD and optionally TABLE and TABEAM (Section 6.1) files. The routine also calls the READ_CONFIG routine to obtain atomic positions and optionally velocities and forces from the CONFIG file. After allocation of all necessary simulation arrays and variables (with compulsory initialisation to "zero" value), the job control information is required; this is obtained by the routine READ_CONTROL, which reads the CONTROL file. The description of the system to be simulated – the types of atoms and molecules present and the intermolecular forces – are obtained by the READ_FIELD routine, which reads the FIELD file. The SYSTEM_INIT routine is called next to initialise various simulation arrays and variables with the data available so far and detects if the job is a restart of previous simulation run. If so it reads the REVOLD (Section 6.1.5) to supply some arrays and variables with the necessary values as saved from the previous job. The domain halo is constructed immediately afterwards by the routine SET_HALO_PARTICLES. After gathering all these data, bookkeeping and exclusion arrays are created for the intramolecular and site related interactions (core-shell, constraint and tether units) by the BUILD_BOOK_INTRA and BUILD_EXCL_INTRA routines. Lastly, the thermodynamic properties of the system are checked and set by the SET_TEMPERATURE routine (which also generates the initial velocities if required to do so).

The calculation of the pair-like forces is carried out in the TWO_BODY_FORCES routine and represents the main part of any simulation. For calculation of the two-body contributions to the atomic forces, the Verlet neighbour list is constructed by the LINK_CELL_PAIRS routine using link-cell lists. Special measures are taken so that the list excludes: (i) pairs of atoms that are both in a *frozen state* as well as (ii) pairs in which one of the atoms has the other in its *exclusion list*. The last is built by BUILD_EXCL_INTRA where the specifications of bond-like interactions in the FIELD file are processed. Various other subroutines are then called to calculate specific contributions by different interactions. For example; VDW_FORCES for the short-range (van der Waals) forces (Section 2.3.1), METAL_LRC, METAL_LD_COMPUTE and METAL_FORCES for the metal interactions (Section 2.3.2), and EWALD_SPME_FORCES, EWALD_REAL_FORCES, EWALD_FRZN_FORCES and EWALD_EXCL_FORCES for the Coulombic forces (Section 2.4).

Higher order intermolecular, site-related and intramolecular forces require the routines
TERSOFF_FORCES, THREE_BODY_FORCES, FOUR_BODY_FORCES,
CORE_SHELL_FORCES or CORE_SHELL_RELAX, TETHERS_FORCES,
BONDS_FORCES, ANGLES_FORCES, DIHEDRALS_FORCES and INVERSIONS_FORCES.
The routines
EXTERNAL_FIELD_APPLY and EXTERNAL_FIELD_CORRECT are required if the simulated system has an external force field (e.g. electrostatic field) operating.

To help with equilibration simulations, routines such as CAP_FORCES, ZERO_K_OPTIMISE and MINIMISE_RELAX are sometimes required to reduce the magnitude of badly equilibrated forces and to steer the MD system towards an equilibrium state.

Integration of the equations of motion is handled by one of the routines listed and described in Chapter 3.

As mentioned elsewhere, DL_POLY_4 does not contain many routines for computing system properties during a simulation. Radial distributions may be calculated however, by using the routines RDF_COLLECT, RDF_EXCL_COLLECT, RDF_FRZN_COLLECT a and RDF_COMPUTE. Similarly, Z-density distributions may be calculated by using the routines Z_DENSITY_COLLECT and Z_DENSITY_COMPUTE, while velocity autocorrelation functions may be calculated using the routines VAF_COLLECT and VAF_COMPUTE. Ordinary thermodynamic quantities are calculated by the routine STATISTICS_COLLECT, which also writes the STATIS file (Section 6.2.13). Routine TRAJECTORY_WRITE writes the HISTORY (Section 6.2.1) file for later (postmortem) analysis. Routine DEFECTS_WRITE writes the DEFECTS (Section 6.2.3) file for later (postmortem) analysis. Routine MSD_WRITE writes the MSDTMP (Section 6.2.2) file for later (postmortem) analysis. Routine RSD_WRITE writes the RSDDAT (Section 6.2.4) file for later (postmortem) analysis.

Job termination is handled by the routine STATISTICS_RESULT which writes the final summaries in the OUTPUT file and dumps the restart files REVIVE and REVCON (Sections 6.2.8 and 6.2.7 respectively).

5.2 Compiling and Running DL_POLY_4

5.2.1 Compiling the Source Code

When you have obtained DL_POLY_4 from Daresbury Laboratory and unpacked it, your next task will be to compile it. To aid compilation three general makefiles have been provided in the sub-directory *build*. These are "Makefile_MPI" - for compiling a parallel version of DL_POLY_4, and "Makefile_SRL1" and "Makefile_SRL2" - for compiling a serial versions (see Appendix **??**). After choosing what the default compilation is to be, the appropriate makefile is to be copied as "Makefile" in the sub-directory *source*. The general DL_POLY_4 makefile will build an executable with the full range of functionality - sufficient for the test cases and for most users' requirements. In most cases, the user will have to modify few entries in the specification part of their makefile to match the location of certain software on their system architecture. **Note** that **only** FORTRAN90 compiler is required for successful build of DL_POLY_4 in serial mode, and **only** FORTRAN90 and MPI implementation - for DL_POLY_4 in parallel mode. Should the user add additional functionality to the code, major changes of the makefile may be required!

In UNIX environment the compilation of the program is initiated by typing the command:

make target

where *target* is the specification of the required machine. For many computer systems this is all that is required to compile a working version of DL_POLY_4. (To determine which targets are already defined in the makefile, examine it or type the command *make* without a nominated target - it will produce a list of known targets.)

The full specification of the *make* command is as follows

make <TARGET= ... > < EX=... > < BINROOT=... >

where some (or all) of the keywords may be omitted. The keywords and their uses are described below. **Note** that keywords may also be set in the UNIX environment (e.g. with the “setenv” command in a TCSH-shell, or “export” in BASH-shell).

5.2.1.1 Keywords in the Makefiles

1. **TARGET**
   The TARGET keyword indicates which kind of computer the code is to be compiled for. This **must** be specified - there is no default value. Valid targets can be listed by the makefile if the command *make* is typed, without arguments. The list frequently changes as more targets are added and redundant ones removed. Users are encouraged to extend the makefile for themselves, using existing targets as examples.

2. **EX**
   The EX keyword specifies the executable name. The default name for the executable is “DLPOLY.Z”.

3. **BINROOT**
   The BINROOT keyword specifies the directory in which the executable is to be stored. The default setting is “../execute”.

5.2.1.2 Modifying the Makefiles

1. **Changing the FORTRAN90 compiler and MPI implementation**
   To specify the FORTRAN90 compiler in a target platform, the user must type the full path to the executable in **FC=“...”** and all appropriate options (as defined in the relevant FORTRAN90 manual) and the path to the MPI implementation in **FCFLAGS=“...”**. The same must be done for the linker: the path to the executable in **LD=“...”** and the appropriate options and the path to the MPI implementation in **LDFLAGS=“...”**.

2. **Adding new functionality**
   To include a new subroutine in the code simply add *subroutine*.o to the list of object names in the makefile (‘OBJ_ALL”). **Note** that there is a hierarchal order of adding file names in the “OBJ_MOD” list whereas such order does not exist in the “OBJ_ALL” list. Therefore, should dependence exist between routines listed in the “OBJ_ALL” list, it **must** be explicitly declared in the makefile.

5.2.1.3 Note on the Interpolation Scheme

In DL_POLY_4 two-body-like contributions (van der Waals, metal and real space Ewald summation) to energy and force are evaluated by interpolation of tables constructed at the beginning of execution. The DL_POLY_4 interpolation scheme is based on a 3-point linear interpolation in r. **Note** that a 5-point linear interpolation in r is ised in DL_POLY_4 for interpolation of the EAM (metal) forces from EAM table data (TABEAM).

The number of grid points (mxgrid) required for interpolation in r to give good energy conservation in a simulation is:

$$\texttt{mxgrid} = \texttt{Max}(\texttt{mxgrid}, 1000, \texttt{Int}(r_{\texttt{cut}}/0.01 + 0.5) + 4) \quad ,$$

where $r_{\texttt{cut}}$ is the main cutoff beyond which the contributions from the short-range-like interactions are negligible.

5.2.2 Running

To run the DL_POLY_4 executable (DLPOLY.Z) you will initially require three to six input data files, which you must create in the *execute* sub-directory, (or whichever sub-directory you keep the executable program). The first of these is the CONTROL file (Section 6.1.1), which indicates to DL_POLY_4 what kind of simulation you want to run, how much data you want to gather and for how long you want the job to run. The second file you need is the CONFIG file (Section 6.1.2). This contains the atom positions and, depending on how the file was created (e.g. whether this is a configuration created from 'scratch' or the end point of another run), the velocities and forces also. The third file required is the FIELD file (Section 6.1.3), which specifies the nature of the intermolecular interactions, the molecular topology and the atomic properties, such as charge and mass. Sometimes you may require a fourth file: TABLE (Section 6.1.6), which contains short ranged potential and force arrays for functional forms not available within DL_POLY_4 (usually because they are too complex e.g. spline potentials) and/or a fifth file TABEAM (Section 6.1.7), which contains metal potential arrays for non-analytic or too complex functional forms and/or a sixth file: REFERENCE (Section 6.1.4), which is similar to the CONFIG file and contains the "perfect" crystalline structure of the system.

Examples of input files are found in the *data* sub-directory, which can be copied into the *execute* subdirectory using the *select* macro found in the *execute* sub-directory.

A successful run of DL_POLY_4 will generate several data files, which appear in the *execute* sub-directory. The most obvious one is the file OUTPUT (Section 6.2.6), which provides an effective summary of the job run: the input information; starting configuration; instantaneous and rolling-averaged thermodynamic data; minimisation information, final configurations; radial distribution functions (RDFs); Z-density profiles and job timing data. The OUTPUT file is human readable. Also present will be the restart files REVIVE (Section 6.2.8) and REVCON (Section 6.2.7). REVIVE contains the accumulated data for a number of thermodynamic quantities and RDFs, and is intended to be used as the input file for a following run. It is *not* human readable. The REVCON file contains the *restart configuration* i.e. the final positions, velocities and forces of the atoms when the run ended and is human readable. The STATIS file (Section 6.2.13) contains a catalogue of instantaneous values of thermodynamic and other variables, in a form suitable for temporal or statistical analysis. Finally, the HISTORY file (Section 6.2.1) provides a time ordered sequence of configurations to facilitate further analysis of the atomic motions. By default this file is formatted (human readable) but with little effort from the user it can be generated unformatted. You may move these output files back into the *data* sub-directory using the *store* macro found in the *execute* sub-directory.

Lastly, DL_POLY_4 may also create the files RDFDAT, ZDNDAT, MSDTMP, RSDDAT, and DEFECTS containing the RDF, Z-density, individual means square displacement and temperature, RSD and defects data respectively. They are all human readable files.

5.2.3 Parallel I/O

Many users that have suffered loss of data in the OUTPUT, especially running in parallel and when an error occurs on parallel architectures. In such circumstances the OUTPUT may be empty or incomplete, despite being clear that the actual simulation has progressed well beyond what has been printed in OUTPUT. Ultimately, this is due to OS's I/O buffers not being flushed as a default by the particular OS when certain kind of errors occurs, especially MPI related. The safest way to avoid loss of information in such circumstances is to write the OUTPUT data to the default output channel ("the screen"). There is an easy way to do this in DL_POLY_4, which is to use the **l_scr** keyword in the CONTROL file. The batch daemon will then place the output in the standard output file, which can then be of use to the user, or alternatively on many batch systems the output can be redirected into another file, allowing an easier following of the job progress over time. This latter technique is also useful on interactive systems where simply printing to the screen could lead to large amounts of output. However, such situations could be easily avoided by redirecting the output using the ">" symbol, for instance: "mpirun -n 4 DLPOLY.Z > OUTPUT".

It is also worth noting that the use of large batch and buffer numbers can speed up enormously the performance of the parallel I/O, for example putting in CONTROL (see Section 6.1.1):

```
io read mpiio         128 10000000 1000000
io write mpiio        512 10000000 1000000
```

at large processor count jobs (over 1000). However, this help comes at a price as larger batches and buffers also requires more memory. So at smaller processor counts the job will abort at the point of trying to use some of the allocated arrays responsible for these.

More information about DL_POLY_4 parallel I/O can be found in the following references [82, 83, 84].

5.2.4 Restarting

The best approach to running DL_POLY_4 is to define from the outset precisely the simulation you wish to perform and create the input files specific to this requirement. The program will then perform the requested simulation, but may terminate prematurely through error, inadequate time allocation or computer failure. Errors in input data are your responsibility, but DL_POLY_4 will usually give diagnostic messages to help you sort out the trouble. Running out of job time is common and provided you have correctly specified the job time variables (using the **close time** and **job time** directives - see Section 6.1.1) in the CONTROL file, DL_POLY_4 will stop in a controlled manner, allowing you to restart the job as if it had not been interrupted.

To restart a simulation after normal termination you will again require the original CONTROL file (*augment it to include the* **restart** *directive and/or extend the length and duration of the new targeted MD run*), the FIELD (and TABLE and/or TABEAM) file, and a CONFIG file, which is the exact copy of the REVCON file created by the previous job. You will also require a new file: REVOLD (Section 6.1.5), which is an exact copy of the previous REVIVE file. If you attempt to restart DL_POLY_4 without this additional file available, the job will most probably fail. **Note** that DL_POLY_4 will append new data to the existing STATIS and HISTORY files if the run is restarted, other output files will be **overwritten**.

In the event of machine failure, you should be able to restart the job in the same way from the surviving REVCON and REVIVE files, which are dumped at regular intervals to meet just such an emergency. In this case check carefully that the input files are intact and use the HISTORY and STATIS files with caution - there may be duplicated or missing records. The reprieve processing capabilities of DL_POLY_4 are not foolproof - the job may crash while these files are being written for example, but they can help a great deal. You are advised to keep backup copies of these files, noting the times they were written, to help you avoid going right back to the start of a simulation.

You can also extend a simulation beyond its initial allocation of timesteps, provided you still have the REVCON and REVIVE files. These should be copied to the CONFIG and REVOLD files respectively and the directive **timesteps** adjusted in the CONTROL file to the new total number of steps required for the simulation. For example if you wish to extend a 10000 step simulation by a further 5000 steps use the directive **timesteps 15000** in the CONTROL file and include the **restart** directive.

Further to the full restart option, there is an alternative **restart scale** directive that will reset the temperature at start or **restart noscale** that will keep the current kinetics intact. /bf Note that these two options are not correct **restart**s but rather modified **start**s as they make no use of REVOLD file and will reset internal accumulators to zero at start.

Note that all these options are mutually exclusive!

If none of the restart options is specified velocities are generated anew with Gaussian distribution of the target kinetic energy based on the provided temperature in the CONTROL file.

5.2.5 Optimising the Starting Structure

The preparation of the initial structure of a system for a molecular dynamics simulation can be difficult. It is quite likely that the structure created does not correspond to one typical of the equilibrium state for the required state point, for the given force field employed. This can make the simulation unstable in the initial stages and can even prevent it from proceeding.

For this reason DL_POLY_4 has available a selection of structure relaxation methods. Broadly speaking, these are energy minimisation algorithms, but their role in DL_POLY_4 is not to provide users with true structural optimisation procedures capable of finding the ground state structure. They are simply intended to help users improve the quality of the starting structure prior to a statistical dynamical simulation, which implies usage during the equilibration period only!

The available algorithms are:

1. 'Zero' temperature molecular dynamics. This is equivalent to a dynamical simulation at low temperature. At each time step the molecules move in the direction of the computed forces (and torques), but are not allowed to acquire a velocity larger than that corresponding to a temperature of 10 Kelvin. The subroutine that performs this procedure is ZERO_K_OPTIMISE.

2. Conjugate Gradients Method (CGM) minimisation. This is nominally a simple minimisation of the system configuration energy using the conjugate gradients method [69]. The algorithm coded into DL_POLY_4 is an adaptation that allows for rotation and translation of rigid bodies. Rigid (constraint) bonds however are treated as stiff harmonic springs - a strategy which we find does allow the bonds to converge within the accuracy required by SHAKE. The subroutine that performs this procedure is MINIMISE_RELAX which makes use of, MINIMISE_MODULE.

3. 'Programmed' energy minimisation, involving both MD and CGM. This method combines the two as minimisation is invoked by user-defined intervals of (usually low temperature) dynamics, in a cycle of minimisation - dynamics - minimisation etc., which is intended to help the structure relax from overstrained conditions (see Section 6.1.1). When using the programmed minimisation DL_POLY_4 writes (and rewrites) the file CFGMIN 6.2.5, which represents the lowest energy structure found during the programmed minimisation. CFGMIN is written in CONFIG file format (see section 6.1.2) and can be used in place of the original CONFIG file.

It should be noted that none of these algorithms permit the simulation cell to change shape. It is only the atomic structure that is relaxed. After which it is assumed that normal molecular dynamics will commence from the final structure.

Notes on the Minimisation Procedures

1. The zero temperature dynamics is really dynamics conducted at 10 Kelvin. However, the dynamics has been modified so that the velocities of the atoms are always directed along the force vectors. Thus the dynamics follows the steepest descent to the (local) minimum. From any given configuration, it will always descend to the same minimum.

2. The conjugate gradient procedure has been adapted to take account of the possibilities of constraint bonds and rigid bodies being present in the system. If neither of these is present, the conventional unadapted procedure is followed.

   (a) In the case of rigid bodies, atomic forces are resolved into molecular forces and torques. The torques are subsequently transformed into an equivalent set of atomic forces which are perpendicular both to the instantaneous axis of rotation (defined by the torque vector) and to the cylindrical radial displacement vector of the atom from the axis. These modified forces are then used in place of the original atomic forces in the conjugate gradient scheme. The atomic displacement induced in the conjugate gradient algorithm is corrected to maintain the magnitude of the radial position vector, as required for circular motion.
   (b) With regard to constraint bonds, these are replaced by stiff harmonic bonds to permit minimisation. This is not normally recommended as a means to incorporate constraints in minimisation procedures as it leads to ill conditioning. However, *if the constraints in the original structure are satisfied*, we find that provided only small atomic displacements are allowed during relaxation it is possible to converge to a minimum energy structure. Furthermore, provided the harmonic springs are stiff enough, it is possible afterwards to satisfy the constraints exactly by further optimising the structure using the stiff springs alone, without having a significant affect on the overall system energy.
   (c) Systems with independent constraint bonds and rigid bodies may also be minimised by these methods.

3. Of the three minimisation strategies available in DL_POLY_4, only the programmed minimiser is capable of finding more than one minimum without the user intervening.
4. Finally, we emphasise once again that the purpose of the minimisers in DL_POLY_4 is to help improve the quality of the starting structure and we believe they are adequate for that purpose. We do not recommend them as general molecular structure optimisers. They may however prove useful for relaxing crystal structures to 0 Kelvin for the purpose of identifying a true crystal structure.

5.2.6 Simulation Efficiency and Performance

Although the DL_POLY_4 underlining parallelisation strategy (DD and link-cells, see Section 7.1.1) is extremely efficient, it cannot always provide linear parallelisation speed gain with increasing processor count for a fixed size system. Nevertheless, it will always provide speedup of the simulation (i.e. there still is a sufficient speed gain in simulations when the number of nodes used in parallel is increased). The simplest explanation why this is is that increasing the processor count for a fixed size system decreases not only the work- and memory-load per processor but also the ratio size of domain to size of halo (both in counts of link cells). When this ratio falls down to values close to one and below, the time DL_POLY_4 spends on inevitable communication (MPI messages across neighbouring domains to refresh the halo data) increases with respect to and eventually becomes prevalent to the time DL_POLY_4 spends on numeric calculations (integration and forces). In such regimes, the **overall** DL_POLY_4 efficiency falls down since processors spend more time on staying idle while communicating than on computing.

It is important that the user recognises when DL_POLY_4 becomes vulnerable to decreased efficiency and what possible measures could be taken to avoid this. DL_POLY_4 calculates and reports the major and secondary link-cell algorithms ($M_x \cdot M_y \cdot M_z$) employed in the simulations immediately after execution. M_x (analogously for M_y and M_z) is the integer number of the ratio of the width of the system domains in x-direction (i.e. perpendicular to the (y,z) plane) to the major and secondary (coming from three- and/or four-body and/or Tersoff interactions) short-range cutoffs specified for the system:

$$\begin{aligned} M_x &= \texttt{Nint}\left[\frac{W_x/P_x}{\texttt{cutoff}}\right] \\ W_x &= \texttt{MD box width} \perp \texttt{plane}(y,z) \\ P_x &= \#(\texttt{nodes})_{x-\texttt{direction}} \quad , \end{aligned} \tag{5.1}$$

where x, y and z represent the directions along the MD cell lattice vectors. Every domain (node) of the MD cell is loaded with $(M_x+2)\cdot(M_y+2)\cdot(M_z+2)$ link-cells of which $M_x \cdot M_y \cdot M_z$ belong to that domain and the rest are a halo image of link-cells forming the surface of the immediate neighbouring domains. In this respect, if we define performance efficiency as minimising communications with respect to maximising computation (minimising the halo volume with respect to the node volume), best performance efficiency will require $M_x \approx M_y \approx M_z \approx M$ and $M \gg 1$. The former expression is a necessary condition and only guarantees good communication distribution balancing. Whereas the latter, is a sufficient condition and guarantees prevalence of computation over communications.

DL_POLY_4 issues a built-in warning when a link-cell algorithms has a dimension less than four (i.e. less than four link-cells per domain in given direction). A useful rule of thumb is that parallelisation speed-up inefficiency is expected when the ratio

$$R = \frac{M_x \cdot M_y \cdot M_z}{(M_x + 2) \cdot (M_y + 2) \cdot (M_z + 2) - M_x \cdot M_y \cdot M_z} \tag{5.2}$$

is close to or drops below one. In such cases there are three strategies for improving the situation that can be used singly or in combination. As obvious from equation (5.1) these are: **(i)** decrease the number of nodes used in parallel, **(ii)** decrease the cutoff and **(iii)** increase system size. It is crucial to note that increased parallelisation efficiency remains even when the link-cell algorithm is used inefficiently. However, DL_POLY_4 will issue an error message and cease execution if it detects it cannot fit a link-cell per domain as this is the minimum the DL_POLY_4 link-cell algorithm can work with - $(1 \cdot 1 \cdot 1)$ corresponding to ratio $R = 1/26$.

It is worth outlining in terms of the $\mathcal{O}($`computation` ; `communication`$)$ function what the rough scaling performance is like of the most computation and communication intensive parts of DL_POLY_4 in an MD timestep.

(a) Domain hallo re-construction in SET_HALO_PARTICLES, METAL_LD_SET_HALO and DEFECTS_REFERENCE_SET_HALO - $\mathcal{O}\left(\mathcal{N}/P\ ;\ \mathcal{N}/R\right)$

(b) Verlet neighbourlist construction by link-cells in LINK_CELL_PAIRS - $\mathcal{O}\left(\mathcal{N}/P\ ;\ 0\right)$, may take up to 40% of the time per timestep

(c) Calculation of k-space contributions to energy and forces from SMPE by EWALD_SPME_FORCES (depends on PARALLEL_FFT which depends on GPFA_MODULE) - $\mathcal{O}\left(\mathcal{N}\ log\ \mathcal{N}\ ;\ (\mathcal{N}\ log\ P)/P\right)$, may take up to 40% of the time per timestep

(d) Particle exchange between domains, involving construction and connection of new out of domain topology when bonded-like interactions exist, by RELOCATE_PARTICLES - $\mathcal{O}\left(\mathcal{N}\ ;\ (P/\mathcal{N})^{1/3}\right)$

(e) Iterative bond and PMF constraint solvers:
CONSTRAINTS_SHAKE_VV, CONSTRAINTS_RATTLE_VV, CONSTRAINTS_SHAKE_LFV
and PMF_SHAKE_VV, PMF_RATTLE_VV, PMF_SHAKE_LFV - $\mathcal{O}\left(\mathcal{N}\ ;\ (P/\mathcal{N})^{1/3}\right)$

where $\mathcal{N}$ is the number of particles, $P = P_x \cdot P_y \cdot P_z$ the total number of domains in the MD cell and the rest of the quantities are as defined in equations (5.1-5.2).

Performance may also affected by the fluctuations in the inter-node communication, due to unavoidable communication traffic when a simulation job does not have exclusive use of all machine resources. Such effects may worsen the performance much, especially when the average calculation time is of the same magnitude as or less than the average communication time (i.e. nodes spend more time communicating rather than computing).

5.3 A Guide to Preparing Input Files

The CONFIG file and the FIELD file can be quite large and unwieldy particularly if a polymer or biological molecule is involved in the simulation. This section outlines the paths to follow when trying to construct files for such systems. The description of the DL_POLY_4 force field in Chapter 2 is essential reading. The various utility routines mentioned in this section are described in greater detail in the DL_POLY_Classic User Manual. Many of these have been incorporated into the DL_POLY GUI [21] and may be conveniently used from there.

5.3.1 Inorganic Materials

The utility GENLAT can be used to construct the CONFIG file for relatively simple lattice structures. Input is interactive. The FIELD file for such systems are normally small and can be constructed by hand. Otherwise, the input of force field data for crystalline systems is particularly simple, if no angular forces are required (notable exceptions to this are zeolites and silicate glasses - see below). Such systems require only the specification of the atomic types and the necessary pair forces. The reader is referred to the description of the DL_POLY_4 FIELD file for further details (Section 6.1.3).

DL_POLY_4 can simulate zeolites and silicate (or other) glasses. Both these materials require the use of angular forces to describe the local structure correctly. In both cases the angular terms are included as *three-body terms*, the forms of which are described in Chapter 2. These terms are entered into the FIELD file with the pair potentials.

An alternative way of handling zeolites is to treat the zeolite framework as a kind of macromolecule (see below). Specifying all this is tedious and is best done computationally: what is required is to determine the nearest image neighbours of all atoms and assign appropriate bond and valence angle potentials. What must be avoided at all costs is specifying the angle potentials *without* specifying bond potentials. In this case DL_POLY_4 will automatically cancel the non-bonded forces between atoms linked via valence angles and the system will collapse. The advantage of this method is that the calculation is likely to be faster than using three-body forces. This method is not recommended for amorphous systems.

5.3.2 Macromolecules

To set up force fields for macromolecules, or indeed any covalent molecules, it is best to use DL_FIELD - http://www.ccp5.ac.uk/DL_FIELD/ . It is a program application tool developed to facilitate the construction of force field models for biological molecules and other molecules with complex geometries. For instance proteins, carbohydrates, polymers and networked molecules such as graphenes and organic cages. **Although created to assist DL_POLY_4, DL_FIELD is a separate program suite that requires separate registration!**

The primary functions of DL_FIELD are as follows:

1. **Force field model converter:** DL_FIELD converts the users atom models, supplied in PDB file format, into input files that are recognisable and ready to run with DL_POLY_Classic and DL_POLY_4 programs with minimum users intervention. This basically involves the conversion of the users atomic configuration in simple xyz coordinates into identifiable atom types base on a particular user-selectable potential schemes and then automatically generate the DL_POLY configuration file (CONFIG), the force field file (FIELD) and a generic control file (CONTROL).

2. **Force field editor:** DL_FIELD allows the user to edit or modify parameters of a particular force field scheme in order to produce a customised scheme that is specific to a particular simulation model. In addition, the standard force field model framework can also be easily modified. For instance, introduction of pseudo points and rigid body implementation to an otherwise standard potential scheme such as CHARMM or AMBER, etc.

3. **Force field library repertoire:** DL_FIELD contains a range of popular potential schemes (see below), all described in a single DL_FIELD format that are also easily recognisable by the user for maintenance purposes. Users can easily expand the existing library to include other new molecules.

Force Field Schemes

The available force field schemes are as follows:

CHARMM - proteins, ethers, some lipids and carbohydrates.

AMBER - proteins and Glycam for carbohydrates.

OPLSAA - proteins

DREIDING - General force field for covalent molecules.

PCFF - Polyorganics and other covalent molecules.

Model Construction

DL_FIELD does not have feature to construct molecular models. This can be achieved by either using DL_POLY GUI [21] or any other standard molecular building packages. The output files must be converted into the PDB format. In the case of proteins, these structures are usually obtained from data banks such as PDB. These *raw* PBD files must first be preprocessed by the user before they are in a *readable* format for DL_FIELD. To ensure this, it is advisable that users take into consideration the following steps:

1. Decide on inclusion/exclusion of and if necessary manually delete molecular residues that involve multiple occupancies in crystalline structures.

2. Usually, hydrogen atoms are not assigned in the raw PDB file. The molecules must therefore be pre-filled with hydrogen atoms (protonated) by using any standard packages available. The user must ensure that proper care is taken of terminal residues which must also be appropriately terminated.

3. Decide on the various charge states of some amino acids, such as histidine (HIS), lysine (LYS), glutamic acid (GLU), etc., by adding or deleting the appropriate hydrogen atoms. Force field schemes such as CHARMM will have different three-letter notations for amino acids of different charge states, DL_FIELD will automatically identify these differences and assign the appropriate potential parameters accordingly.

4. For cysteine (CYS) molecules with disulphide bonds, thiolate hydrogen atoms must be removed. DL_FIELD will automatically define disulphide bonds between the molecules, provided the S-S distance is within a sensible value.

5. DL_FIELD does not solvate the molecules and it is the user's responsibility to add water by using any standard package available (for example the DL_POLY GUI [21]).

Fore more details or further information, please consult the DL_FIELD manual and website http://www.ccp5.ac.uk/DL_FIELD/ .

5.3.3 Adding Solvent to a Structure

The utility WATERADD adds water from an equilibrated configuration of 256 SPC water molecules at 300 K to fill out the MD cell. The utility SOLVADD fills out the MD box with single-site solvent molecules from a fcc lattice. The FIELD files will then need to be edited to account for the solvent molecules added to the file.

Hint: to save yourself some work in entering the non-bonded interactions variables involving solvent sites to the FIELD file put two bogus atoms of each solvent type at the end of the CONNECT_DAT file (for AMBER force-fields) the utility AMBFORCE will then evaluate all the non-bonded variables required by DL_POLY_4. Remember to delete the bogus entries from the CONFIG file before running DL_POLY_4.

5.3.4 Analysing Results

DL_POLY_4 is not designed to calculate every conceivable property you might wish from a simulation. Apart from some obvious thermodynamic quantities and radial distribution functions, it does not calculate anything beyond the atomic trajectories. You must therefore be prepared to post-process the HISTORY file if you want other information. There are some utilities in the DL_POLY_4 package to help with this, but the list is far from exhaustive. In time, we hope to have many more. Our users are invited to submit code to the DL_POLY_4*public* library to help with this.

The utilities available are described in the DL_POLY_Classic User Manual. Users should also be aware that many of these utilities are incorporated into the DL_POLY GUI [21].

5.3.5 Choosing Ewald Sum Variables

5.3.5.1 Ewald sum and SPME

This section outlines how to optimise the accuracy of the Smoothed Particle Mesh Ewald sum parameters for a given simulation..

As a guide to beginners DL_POLY_4 will calculate reasonable parameters if the **ewald precision** directive is used in the CONTROL file (see Section 6.1.1). A relative error (see below) of 10^{-6} is normally sufficient so the directive

ewald precision 1d-6

will make DL_POLY_4 evaluate its best guess at the Ewald parameters α, `kmaxa`, `kmaxb` and `kmaxc`, or their doubles if **ewald** rather than **spme** is specified. (The user should note that this represents an *estimate*, and there are sometimes circumstances where the estimate can be improved upon. This is especially the case when the system contains a strong directional anisotropy, such as a surface.) These four parameters may also be set explicitly by the **ewald sum** directive in the CONTROL file. For example the directive

ewald sum 0.35 6 6 8

which is equvalent to

spme sum 0.35 12 12 16

would set $\alpha = 0.35$ Å^{-1}, `kmaxa` $= 12$, `kmaxb` $= 12$ and `kmaxc` $= 16$[1]. The quickest check on the accuracy of the Ewald sum is to compare the coulombic energy (U) and virial ($\mathcal{W}$) in a short simulation. Adherence to the relationship $U = -\mathcal{W}$, shows the extent to which the Ewald sum is correctly converged. These variables can be found under the columns headed `eng_cou` and `vir_cou` in the OUTPUT file (see Section 6.2.6).

The remainder of this section explains the meanings of these parameters and how they can be chosen. The Ewald sum can only be used in a three dimensional periodic system. There are five variables that control the accuracy: α, the Ewald convergence parameter; $r_{\rm cut}$ the real space force cutoff; and the `kmaxa`, `kmaxb` and `kmaxc` integers that specify the dimensions of the SPME charge array (as well as FFT arrays). The three integers effectively define the range of the reciprocal space sum (one integer for each of the three axis directions). These variables are not independent, and it is usual to regard one of them as pre-determined and adjust the others accordingly. In this treatment we assume that $r_{\rm cut}$ (defined by the **cutoff** directive in the CONTROL file) is fixed for the given system.

The Ewald sum splits the (electrostatic) sum for the infinite, periodic, system into a damped real space sum and a reciprocal space sum. The rate of convergence of both sums is governed by α. Evaluation of the real space sum is truncated at $r = r_{\rm cut}$ so it is important that α be chosen so that contributions to the real space

[1]**Important note**: As the SPME method substitues the standard Ewald the values of `kmaxa`, `kmaxb` and `kmaxc` are the double of those in the prescription of the standard Ewald since they specify the sides of a cube, not a radius of convergence.

sum are negligible for terms with $r > r_{\rm cut}$. The relative error (ϵ) in the real space sum truncated at $r_{\rm cut}$ is given approximately by

$$\epsilon \approx {\rm erfc}(\alpha\ r_{\rm cut})/r_{\rm cut} \approx \exp[-(\alpha\ r_{\rm cut})^2]/r_{\rm cut} \quad , \tag{5.3}$$

which reciprocally gives an estimate for α for a given ϵ:

$$\alpha \approx \frac{\sqrt{|\ln(\epsilon\ r_{\rm cut})|}}{r_{\rm cut}} \quad . \tag{5.4}$$

The recommended value for α is $3.2/r_{\rm cut}$ or greater (too large a value will make the reciprocal space sum very slowly convergent). This gives a relative error in the energy of no greater than $\epsilon = 4 \times 10^{-5}$ in the real space sum. When using the directive **ewald precision** DL_POLY_4 makes use of a more sophisticated approximation:

$${\rm erfc}(x) \approx 0.56\ \exp(-x^2)/x \tag{5.5}$$

to solve recursively for α, using equation 5.3 to give the first guess.

The relative error in the reciprocal space term is approximately

$$\epsilon \approx \exp(-k_{max}^2/4\alpha^2)/k_{max}^2 \tag{5.6}$$

where

$$k_{max} = \frac{2\pi}{L}\ \frac{\texttt{kmax}}{2} \tag{5.7}$$

is largest k-vector considered in reciprocal space, L is the width of the cell in the specified direction and `kmax` is an integer.

For a relative error of 4×10^{-5} this means using $k_{max} \approx 6.2\ \alpha$. `kmax` is then

$$\texttt{kmax} > 6.4\ L/r_{\rm cut}. \tag{5.8}$$

In a cubic system, $r_{\rm cut} = L/2$ implies `kmax` $= 14$. In practice the above equation slightly over estimates the value of `kmax` required, so optimal values need to be found experimentally. In the above example `kmax` $= 10$ or 12 would be adequate.

If you wish to set the Ewald parameters manually (via the **ewald sum** or **spme sum** directives) the recommended approach is as follows. Preselect the value of $r_{\rm cut}$, choose a working a value of α of about $3.2/r_{\rm cut}$ and a large value for the `kmax` (say 20 20 20 or more). Then do a series of ten or so *single* step simulations with your initial configuration and with α ranging over the value you have chosen plus and minus 20%. Plot the Coulombic energy (-$\mathcal{W}$) versus α. If the Ewald sum is correctly converged you will see a plateau in the plot. Divergence from the plateau at small α is due to non-convergence in the real space sum. Divergence from the plateau at large α is due to non-convergence of the reciprocal space sum. Redo the series of calculations using smaller `kmax` values. The optimum values for `kmax` are the smallest values that reproduce the correct Coulombic energy (the plateau value) and virial at the value of α to be used in the simulation. Note that one needs to specify the three integers (`kmaxa`, `kmaxb`, `kmaxc`) referring to the three spatial directions, to ensure the reciprocal space sum is equally accurate in all directions. The values of `kmaxa`, `kmaxb` and `kmaxc` must be commensurate with the cell geometry to ensure the same minimum wavelength is used in all directions. For a cubic cell set `kmaxa` = `kmaxb` = `kmaxc`. However, for example, in a cell with dimensions 2A = 2B = C, (ie. a tetragonal cell, longer in the c direction than the a and b directions) use 2`kmaxa` = 2`kmaxb` = `kmaxc`.

If the values for the kmax used are too small, the Ewald sum will produce spurious results. If values that are too large are used, the results will be correct but the calculation will consume unnecessary amounts of cpu time. The amount of cpu time increases proportionally to `kmaxa` $\times$ `kmaxb` $\times$ `kmaxc`.

It is worth noting that the working values of the k-vectors may be larger than their original values depending on the actual processor decomposition. This is to satisfy the requirement that the k-vector/FFT transform down each direction per domain is a multiple of 2, 3 and 5 only, which is due to the GPFA code (single 1D FFT) which the DaFT implementation relies on. This allowes for greater flexiblity than the power of 2 multiple restriction in DL_POLY_4 predicessor, DL_POLY_3. As a consequence, however, execution on different processor decompositions may lead to different working lengths of the k-vectors/FFT transforms and therefore slightly different SPME forces/energies whithin the same level of SPME/Ewald precision/accuracy specified. **Note** that although the number of processors along a dimension of the DD grid may be any number, numbers that have a large prime as a factor will lead to inefficient performance!

5.4 Warning and Error Processing

5.4.1 The DL_POLY_4 Internal Warning Facility

DL_POLY_4 contains a number of various in-built checks scattered throughout the package which detect a range of possible inconsistencies or errors. In all cases, such a check fails the subroutine WARNING is called, resulting in an appropriate message that identifies the inconsistency. In some cases an inconsistency is resolved by DL_POLY_4 supplying a default value or DL_POLY_4 assuming a priority of one directive over the another (in clash of mutually exclusive directives). However, in other cases this cannot be done and controlled termination of the program execution is called by the subroutine ERROR. In any case appropriate diagnostic message is displayed notifying the user of the nature of the problem.

5.4.2 The DL_POLY_4 Internal Error Facility

DL_POLY_4 contains a number of in-built error checks scattered throughout the package which detect a wide range of possible errors. In all cases, when an error is detected the subroutine ERROR is called, resulting in an appropriate message and termination of the program execution (either immediately, or after some additional processing). In some case, if the cause for error is considered to be mendable it is corrected and the subroutine WARNING results in an appropriate message.

Users intending to insert new error checks should ensure that all error checks are performed *concurrently* on *all* nodes, and that in circumstances where a different result may obtain on different nodes, a call to the global status routine GCHECK is made to set the appropriate global error flag on all nodes. Only after this is done, a call to subroutine ERROR may be made. An example of such a procedure might be:

```
Logical ::  safe
safe = (test_condition)
Call gcheck(safe)
If (.not.safe) Call error(message_number)
```

In this example it is assumed that the logical operation *test_condition* will result in the answer *.true.* if it is safe for the program to proceed, and *.false.* otherwise. The call to ERROR requires the user to state the `message_number` is an integer which used to identify the appropriate message to be printed.

A full list of the DL_POLY_4 error messages and the appropriate user action can be found in Appendix D of this document.

Chapter 6

Data Files

Scope of Chapter

This chapter describes all the input and output files for DL_POLY_4, examples of which are to be found in the *data* sub-directory.

6.1 The INPUT Files

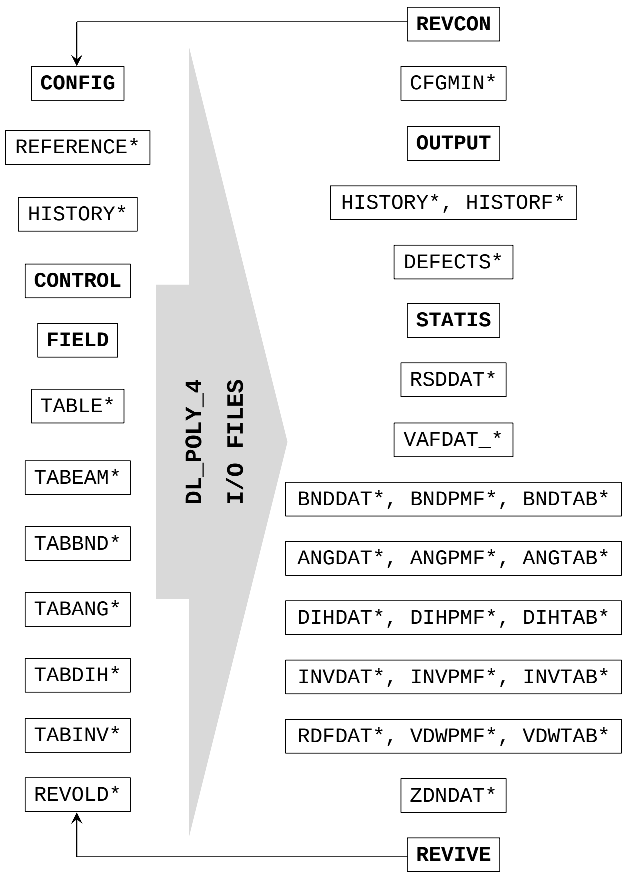

Figure 6.1: DL_POLY_4 input (left) and output (right) files. **Note**: files marked with an asterisk are non-mandatory.

DL_POLY_4 requires seven input files named CONTROL, CONFIG, FIELD, TABLE, TABEAM, REFERENCE and REVOLD. The first three files are mandatory, whereas TABLE and TABEAM are only used to input certain kinds of pair or metal potentials, and may not always be required. REFERENCE is required only if defect detection is switched on in CONTROL. REVOLD is required only if the job represents a continuation of a previous job. In the following sections we describe the form and content of these files.

6.1.1 The CONTROL File

The CONTROL file is read by the subroutine READ_CONTROL and defines the control variables for running a DL_POLY_4 job. (It is also read by the subroutine SCAN_CONTROL in the SET_BOUNDS routine.) It makes extensive use of **directives** and **keywords**. Directives are character strings that appear as the first entry on a data record (or line) and which invoke a particular operation or provide numerical parameters. Also associated with each directive may be one or more keywords, which may qualify a particular directive by, for example, adding extra options. Directives can appear in any order in the CONTROL file, except for the **finish** directive which marks the end of the file. Some of the directives are mandatory (for example the **timestep** directive that defines the timestep), others are optional.

This way of constructing the file is very convenient, but it has inherent dangers. It is, for example, quite easy to specify contradictory directives, or invoke algorithms that do not work together. By large DL_POLY_4 tries to sort out these difficulties and print helpful error messages, but it does not claim to be fully foolproof. Another common mistake is to specify more than once a directive that has no contradictory, disabling, altering or antagonistic directives - then the one specified last will be used as a control directive (for example **densvar, equil, steps, press, mxshak, shake, ...**). Fortunately, in most cases the CONTROL file will be small and easy to check visually. It is important to think carefully about a simulation beforehand and ensure that DL_POLY_4 is being asked to do something that is physically reasonable. It should also be remembered that the present capabilities the package may not allow the simulation required and it may be necessary for you yourself to add new features.

An example CONTROL file appears below. The directives and keywords appearing are described in the following section. The example lists all possible and not mutually excluding directives in a particular order. Although this order is not mandatory, it is highly recommended.

```
TITLE RECORD: DL_POLY_4 SAFE ORDER OF CONTROL DIRECTIVES

# SYSTEM REPLICATION & IMPACT OPTION
nfold              10 10 10
impact 1 2000 7.5 1.0 2.0 3.0

# DENSITY VARIATION ARRAY BOOST
densvar                  10 %

# INDEX AND VERIFICATION BYPASS AND NO TOPOLOGY REPORTING
no index
no strict
no topology

# INTERACTIONS BYPASS
no electostatics
no vdw

# APPLY MIXING TO ALLOWED & AVAILABLE VDW CROSS INTERACTIONS
# LorentzBerthelot || Fender-Halsey || Hogervorst (good hope) || Halgren HHG || Tang-Toennies
vdw mixing Lorentz

# DIRECT CALCULATION OF VDW/METAL INTERACTIONS INSTEAD OF
# EVALUATION BY SPLINING OVER TABULATED VALUES IN MEMORY
vdw direct
metal direct

# FORCE-SHIFT VDW INTERACTIONS SO THAT ENERGY AND FORCE
# CONTRIBUTIONS FALL SMOOTHLY TO ZERO WHEN APPROACHING R_CUT
vdw shift

# RANDOM NUMBER GENERATOR SEEDING
seed 100 200 300

# I/O READ: METHOD, READER COUNT, BATCH & BUFFER SIZES
io read  mpiio        2 2000000 20000

# I/O WRITE: METHOD, TYPE, WRITER COUNT, BATCH & BUFFER SIZES
io write mpiio sorted 8 2000000 20000
```

```
# SLAB SIMULATION PARALLEL CONTROL
slab

# RESTART OPTIONS
restart noscale
dump                 1000     steps

# SYSTEM TARGET TEMPERATURE AND PRESSURE
temperature           300.0   Kelvin
pressure                0.001 k-atmospheres

# SYSTEM CUTOFFS AND ELECTROSTATICS
rcut                   10.0  Angstroms
rpad                    0.35 Angstroms
rvdw                    8.0  Angstroms
exclude
epsilon                 1.0
ewald precision         1.0e-5
ewald evaluate          4

# RELAXED SHELL MODEL TOLERANCE
rlxtol                  1.0 force

# CONSTRANTS ITERATION LENGTH and TOLERANCE
mxshak                250 cycles
shake                   1.0e-5

# INTEGRATION FLAVOUR, ENSEMBLE AND PSEUDO THERMOSTAT
integration velocity verlet
ensemble nst hoover 0.5 0.5
pseudo langevin   2.0 150.0

# INTEGRATION TIMESTEP
variable timestep       0.001 pico-seconds
mindis                  0.03  Angstroms
maxdis                  0.10  Angstroms
mxstep                  0.005 pico-seconds

# SUMULATION & EQUILIBRATION LENGTH
steps               10000 steps
equilibration        1000 steps

# EQUILIBRATION DIRECTIVES
zero
cap                  2000 kT/Angstrom
scale                   5 steps
regauss                 3 steps
minimise force      20  1.0
optimise energy         0.001

# STATISTICS
collect
```

```
stack                  50 deep
stats                  10 steps

# OUTPUT
print                   2 steps

# HISTORY
replay
trajectory        20 30 0

# DEFECTS TRAJECTORY - DEFECTS
defects           40 15 0.75

# DISPLACEMENTS TRAJECTORY - RSDDAT
displacements     70 10 0.25

# MSDTMP
msdtmp          1000 100

# INTRAMOLECULAR PDF ANALYSIS BY TYPE IF PRESENT
analyse  bonds       sample every 100  nbins 250  rmax 5.0
analyse  angles      sample every 100  nbins 360  # [ 0 : pi]
analyse  dihedrals   sample every 100  nbins 720  # [-pi: pi]
analyse  inversions  sample every 100  nbins 360  # [ 0 : pi]

# INTRAMOLECULAR PDF ANALYSIS FOR ALL TYPES PRESENT
analyse  all         sample every 100  nbins 1000  rmax 5.0

# PRINT ANY DEFINED INTER-(RDF->VDW) & INTRA-(bonded) MOLECULAR PDF ANALYSIS
print analysis

# RDF & Z-DENSITY
binsize                 0.05 Angstroms
rdf                     7    steps
print rdf
zden                    7    steps
print zden

# EXECUTION TIME
job time             1000 seconds
close time             10 seconds

# FINISH
finish
```

6.1.1.1 The CONTROL File Format

The file is free-formatted and not case-sensitive. Every line is treated as a command sentence (record). Commented records (beginning with a #) and blank lines are not processed and may be added to aid legibility (see example above). Records must be limited in length to 100 characters. Records are read in words (**directives** and additional **keywords** and **numbers**), as a word must not exceed 40 characters in length. Words are recognised as such by separation by one or more space characters. Additional annotation is not recommended but may be added onto a directive line after the last control word in it.

- The first record in the CONTROL file is a header (up to 100 characters long) to aid identification of the file.
- The last record is a **finish** directive, which marks the end of the input data.

Between the header and the **finish** directive, a wide choice of control directives may be inserted. These are described below.

6.1.1.2 The CONTROL File Directives

The directives available are as follows:

directive:	**meaning:**
analyse **all** (sampling) (every) f **nbins** n **rmax** r	calculate and collect all intramolecular PDFs every f timesteps (default $f = 1$), using a grid with n bins and bonds cutoff of r Å (default r = 2 Å).
analyse **bon**ds (sampling) (every) f **nbins** n **rmax** r	calculate and collect bonds PDFs every f timesteps (default $f = 1$), using a grid with n bins in the range $(0 : r]$ (default r = 2 Å).
analyse **ang**les (sampling) (every) f **nbins** n	calculate and collect angles PDFs every f timesteps (default $f = 1$), using a grid with n bins in the range $(0 : 180]$ (degrees).
analyse **dih**edrals (sampling) (every) f **nbins** n	calculate and collect dihedrals PDFs every f timesteps (default $f = 1$), using a grid with n bins in the range $(-180 : 180]$ (degrees).
analyse **inv**ersions (sampling) (every) f **nbins** n	calculate and collect inversions PDFs every f timesteps (default $f = 1$), using a grid with n bins in the range $(0 : 180]$ (degrees).
binsize f	set the bin size for radial and z-density distribution functions to f Å (if 10^{-5} Å $\leq f \leq r_{\rm cut}/4$ or undefined, f defaults to 0.05 Å)
cap (forces) f	cap forces during equilibration period, f is maximum cap in units of k_BT/Å (default $f = 1000$ k_BT/Å)
close time f	set job closure time to f seconds
collect	include equilibration data in overall statistics
coulomb	calculate electrostatic forces using direct Coulomb sum
cutoff f ($\equiv$ **rcut** f)	set required long-ranged interactions cutoff, $r_{\rm cut}$, to f Å
defects i j f	write defects trajectory file, DEFECTS, with controls: i = start timestep for dumping defects configurations

	(default $i = 0$) j = timestep interval between configurations (default $j = 1$) f = site-interstitial cutoff (default $f = \text{Min}\,[0.75, r_{\text{cut}}/3]$ Å, $\text{Min}\,[0.3, r_{\text{cut}}/3]$ Å $\leq f \leq \text{Min}\,[3.5, r_{\text{cut}}/2]$ Å)
delr f ($\equiv$ **rpad** $4f$)	DL_POLY_Classic Verlet shell strip cutoff option is iterpreted by DL_POLY_4 as the **pad**ding ($\equiv$ **rpad**) $f/4$ option, so that r_{pad} gets set to $\text{Max}(r_{\text{pad}}, f_{\text{delr}}/4)$
densvar f	allow for local variation of $\approx f$ % in the system density of (i) particles and (ii) any present bonded-like entities (very useful for extremely non-equilibrium simulations, default $f = 0$)
distance	calculate electrostatic forces using Coulomb sum with distance dependent dielectric
displacements i j f	write displacements trajectory file, RSDDAT, with controls: i = start timestep for dumping displacements configurations (default $i = 0$) j = timestep interval between configurations (default $j = 1$) f = displacement qualifying cutoff (default $f = 0.15$ $\AA$)
dump n	set restart data dump interval to n steps (default $n = 1000$)
ensemble nve	select NVE ensemble (default ensemble)
ensemble nvt evans	select NVE$_{kin}$ ensemble, type Evans with Gaussian constraints thermostat
ensemble nvt langevin f	select NVT ensemble, type Langevin with thermostat relaxation speed (friction) constant f in ps^{-1}
ensemble nvt andersen f_1 f_2	select NVT ensemble, type Andersen with f_1, f_2 as the thermostat relaxation time in ps and softness ($0 \leq f_2 \leq 1$)
ensemble nvt berendesen f	select NVT ensemble, type Berendsen with thermostat relaxation constant f in ps
ensemble nvt hoover f	select NVT ensemble, type Nose-Hoover with thermostat relaxation constant f in ps
ensemble nvt gst f_1 f_2	select NVT ensemble, type Gentle Stochastic with thermostat relaxation constant f_1 in ps and Langevin friction f_1 in ps^{-1}
ensemble nvt dpd*type* γ	select NVT ensemble, with a DPD thermostat of type ***s1*** or ***s2*** for Shardlow's splitting of first or second order respectively, with an *optional* system global drag coefficient γ in Dalton/ps
ensemble npt langevin f_1 f_2	select NPT ensemble, type Langevin, with f_1, f_2 as the thermostat and barostat relaxation speed (friction) constants in ps^{-1}
ensemble npt berendsen f_1 f_2	select NPT ensemble, type Berendsen with f_1, f_2 as the thermostat and barostat relaxation times in ps

ensemble npt hoover f_1 f_2	select NPT ensemble, type Nose-Hoover, with $f_1,\ f_2$ as the athermostat and barostat relaxation times in ps
ensemble npt mtk f_1 f_2	select NPT ensemble, type Martyna-Tuckerman-Klein with $f_1,\ f_2$ as the thermostat and barostat relaxation times in ps
ensemble nst langevin f_1 f_2	select N$\underline{\underline{\sigma}}$T ensemble, type Langevin with $f_1,\ f_2$ as the thermostat and barostat relaxation speed (friction) constants in ps^{-1}
ensemble nst berendsen f_1 f_2	select N$\underline{\underline{\sigma}}$T ensemble, type Berendsen with $f_1,\ f_2$ as the thermostat and barostat relaxation times in ps
ensemble nst hoover f_1 f_2	select N$\underline{\underline{\sigma}}$T ensemble, type Nose-Hoover with $f_1,\ f_2$ as the thermostat and barostat relaxation times in ps
ensemble nst mtk f_1 f_2	select NσT ensemble, type Martyna-Tuckerman-Klein with $f_1,\ f_2$ as the thermostat and barostat relaxation times in ps
ensemble nst Q $f_1 f_2$ **area**	select NP$_n$AT ensemble, type Q (i.e. *lang*, *ber*, *hoover* or *mtk*), with $f_1,\ f_2$ as the thermostat and barostat relaxation times in ps
ensemble nst Q $f_1 f_2$ **tens**ion γ	select NP$_n\gamma$T ensemble, type Q (i.e. *lang*, *ber*, *hoover* or *mtk*), with $f_1,\ f_2$ as the thermostat and barostat relaxation times in ps and set required simulation (target/external) surface tension to γ dyn/cm
ensemble nst Q $f_1 f_2$ **tens** γ **semi**	select the same NP$_n\gamma$T ensemble as above but with the "semi-anisotropic constraint" so that the MD cell changes isotropically in the (x, y) plane
ensemble nst Q $f_1 f_2$ **orth**orhombic	select the NPT anisotropic ensemble for the orthorhombic MD cell ("orthorhombic constraint" - equivalent to the NP$_n\gamma$T ensemble when $\gamma = 0$: NP$_n\gamma = 0$T)
ensemble nst Q $f_1 f_2$ **orth semi**	select the same NP$_n\gamma = 0$T ensemble as above but with the "semi-anisotropic constraint" so that the MD cell changes isotropically in the (x, y) plane ("semi-orthorhombic constraint" - equivalent to the NP$_n\gamma = 0$T semi ensemble)
epsilon (constant) f	set relative dielectric constant to f (default $f\ =\ 1.0$)
equilibration (steps) n	equilibrate system for the first n timesteps (default $n\ =\ 0$)
ewald evaluate (every) n	evaluate the k-space contributions to the Ewald sum once every n timesteps ($1\ \leq\ n\ \leq\ 10$, activated when $n\ \geq\ 2$, $n\ <\ 1$ or undefined defaults to $n\ =\ 1$, $n\ >\ 10$ defaults to $n\ =\ 4$)
ewald precision f	calculate electrostatic forces using Ewald sum with automatic parameter optimisation for precission f ($10^{-20}\ \leq\ f\ \leq\ 0.5$, default $f\ =\ 10^{-20}$)

ewald (sum) α k_1 k_2 k_3	calculate electrostatic forces using Ewald sum with α = Ewald convergence parameter in Å^{-1} $k1$ = is the maximum k-vector index in x-direction $k2$ = is the maximum k-vector index in y-direction $k3$ = is the maximum k-vector index in z-direction
exclude	switch on extended coulombic exclusion affecting intra-molecular interactions such as: chemical bonds and bond angles; as well as bond constraints between ions that have shells and cores
finish	close the CONTROL file (last data record)
impact i j E x y z	initiate impact on the particle with index i ($i \geq 1$) at timestep j ($i \geq 0$) with energy E ($E \geq 0$) in kilo-eV and direction vector x y z from the Cartesian origin (centre) of the MD box (defaults: $i = 1, j = 0, E = 0, x = 1, y = 1, z = 1$)
integrator *string*	set the type of Verlet integrator, where *string* can only be *leapfrog* or *velocity*, as the later is the default
io read *method j k l e*	set the general I/O read interface to: ***method*** :: *mpiio* for MPI-I/O, *direct* for parallel direct access FORTRAN I/O or *master* for traditional master I/O or *netcdf* for netCDF I/O provided DL_POLY_4 is compiled in a netCDF-enabled mode (default *mpiio*) ***j, reader count*** :: $1 \leq j \leq$ `job size` (default $j = 2^{Int[Log\{Min(\texttt{job size},2\sqrt{\texttt{job size}})\}/Log(2)]}$) is the designated number of processes to carry out I/O read operations simultaneously **NOTE that *k* is not applicable for the *master* method** ***k, batch size*** :: $1 \leq k \leq 10,000,000$ (default $2,000,000$) is the maximum number of particle entities in a batch, i.e. multiples of (*species*,*index*,$\underline{r}$,$\underline{v}$,$\underline{f}$,*etc.*), transmitted between I/O groups (= I/O readers) for domain distribution purposes ***l, buffer size*** :: $100 \leq l \leq 100,000$ (default $20,000$) is the maximum number of ASCII line records read in a batch **NOTE that *e* is not applicable for the *master* method** ***e, parallel error check*** :: *Y*es (default *N*)
io write *method rp type j k l e*	set the general I/O write interface to: ***method*** :: *mpiio* for MPI-I/O, *direct* for parallel direct access FORTRAN I/O or *master* for traditional master I/O or *netcdf* for netCDF I/O provided DL_POLY_4 is compiled in a netCDF-enabled mode (default *mpiio*) **WARRNING: *direct* is not a platfotm portable solution (as it fails on LUSTRE but works on GPFS)** **NOTE that *rp* is only applicable for the *netcdf* method** ***rp, real precision*** :: *32bit* or *amber* for 32-bit (float), otherwise 64-bit (double) is defaulted if unspecified ***type*** :: *sorted* or *unsort*ed (DD scrambled) by global index output (default *sorted*)

j, writer count :: $1 \leq j \leq$ `job size`
(default $j = 2^{Int[Log\{Min(\texttt{job size},8\sqrt{\texttt{job size}})\}/Log(2)]}$)
is the designated number of processes to
carry out I/O write operations simultaneously

NOTE that *k* is not applicable for the *master* method

k, batch size :: $1 \leq k \leq 10,000,000$ (default $2,000,000$)
is the maximum number of particle entities in a batch, i.e.
multiples of ($species$,$index$,$\underline{r}$,$\underline{v}$,$\underline{f}$,$etc.$), transmitted between
I/O groups (= I/O writers) for global sorting purposes

l, buffer size :: $100 \leq l \leq 100,000$ (default $20,000$)
is the maximum number of ASCII line records written in a batch

NOTE that *e* is not applicable for the *master* method

e, parallel error check :: Yes (default N)

job time f	set job time to f seconds
maxdis f	set maximum distance allowed in variable timestep (control) to f Å (default $f = 0.10$ Å)
metal direct	enforce the direct calculation of metal interactions defined by explicit potential forms, i.e. it will not work for metal alloy systems using the EAM, EEAM, 2BEAM or 2BEEAM (TABEAM)
metal sqrtrho	swich the TABEAM default of reading embedding functions, $F(\rho)$, as interpolated over densities, ρ, to as interpolated over $\sqrt{\rho}$, i.e. $F = F(\sqrt{\rho})$
mindis f	set minimum distance allowed in variable timestep (control) to f Å (default $f = 0.03$ Å)
minimise *string* n f	minimise the instantaneous system configuration every n steps during equilibration (with respect to the last equilibration step) using conjugate gradient method (CGM) with respect to the criterion, *string*, and tolerance, f, where this criterion can only be *force* ($1 \leq f \leq 1000$, default $f = 50$) or *energy* ($0 < f \leq 0.01$, default $f = 0.005$) or *distance* (maximum absolute displacement in Å, $10^{-6} \leq f \leq 0.1$, default $f = 0.005$); the lowest *string* CGM minimised configuration during equilibration is saved in a file, CFGMIN which has the same format as CONFIG
msdtmp i j	write MSDTMP file, containing particles' individual $\sqrt{MSD}$ (in Å) and T_{mean} (in Kelvin), with controls: i = start timestep for dumping configurations (default $i = 0$) j = timestep interval between configurations (default $j = 1$)
multiple (timestep) n	act exactly the same as **ewald evalu**ate (every) n option above
mxquat n	set FIQA iterations limit to n (default $n = 100$)
mxshak n	set shake/rattle iterations limit to n (default $n = 250$)

mxstep *f*	set maximum timestep value in variable timestep (control) to f ps (no default $f = 0.0$ ps but if not opted sets to Huge(1.0))
nfold *i j k*	option to create matching CONFIG_i_j_k and FIELD_i_j_k for a volumetrically expanded version of the current system (CONFIG and FIELD) by replicating CONFIG's contents (i, j, k) times along the MD cell lattice vectors while preserving FIELD's topology template intact
no elec	ignore electrostatics in simulation
no index	ignore particles' indices as read from the CONFIG file and set particles' indexing by order of reading, this option assumes that the FIELD topology description matches the crystallographic sites from the CONFIG file by their order of reading rather than by their actual indexing
no strict	**(i)** abort strict checks such as; on existence of well defined system cutoff, on contiguity of particles' indices when connecting CONFIG (crystallographic listing) to FIELD (topology), on IO when **io** *mpiio/direct sorted* is selected, etc., **(ii)** abort display of warnings, non-leading to error messages and of iteration cycles in minimisation/relaxation routines, **(iii)** assume safe defaults for the general simulation cutoff and its padding, temperature, pressure and job times
no topology	skip detailed topology reporting during read of FIELD in OUTPUT (no FIELD replication), useful for large bio-chemcal simulations
no vafaveraging	ignore time-averaging of velocity autocorrelation functions (VAFs), report all calculated VAF profiles for individual species to VAFDAT files and final profile (for all species) in OUTPUT
no vdw	ignore short range (non-bonded) interactions in simulation
no vom	ignore Centre of Mass momentum removal during the simulation
optimise *string f*	minimise the system configuration at start during equilibration using conjugate gradient method (CGM) with respect to the criterion, *string*, and tolerance, f, where the criterion can only be *force* ($1 \leq f \leq 1000$, default $f = 50$) or *energy* ($0 < f \leq 0.01$, default $f = 0.005$) or *distance* (maximum absolute displacement in Å) ($10^{-6} \leq f \leq 0.1$, default $f = 0.005$); the CGM minimised configuration is saved in a file, CFGMIN which has the same format as CONFIG
padding f ($\equiv$ **rpad** f)	set optional padding to the major cutoff, $r_{\rm cut}$, to f Å ($f \geq {\rm Min}[0.05, 0.5\%.r_{\rm cut}]$, default $f = 0$)
pressure f	set required system pressure to f katms (target pressure for constant pressure ensembles)

print (every) n	print system data every n timesteps
print analysis	print any opted for analysis inter- & and intra-molecular PDFs in OUTPUT as well as their corresponding files *DAT, *PMF, *TAB
print rdf	print radial distribution functions
print vaf	print velocity autocorrelation functions
print zden	print Z-density profile
pseudo *string* f_1 f_2	attach a pseudo thermal bath with a thermostat of type *string*, where string can only be *langevin*, *gauss* or *direct* (if none is specified then *langevin*→*direct* are applied successively), f_1 is the thickness of the thermostat layers, attached on the inside of the MD cell boundaries, in units of Å (default $f1 = 2$ Å), f_2 is the thermostat temperature in Kelvin ($f2 \geq 1$), which when unspecified defaults to the system target temperature
quaternion (tolerance) f	set quaternion tolerance to f (default 10^{-8})
rdf (sampling) (every) f	calculate and collect radial distribution functions every f timesteps (default $f = 1$)
reaction (field)	calculate electrostatic forces using reaction field electrostatics
reaction (field) **damp** α	calculate electrostatic forces using reaction field electrostatics with Fennell [63] damping (Ewald-like convergence) parameter α in Å^{-1}
reaction (field) **precision** f	calculate electrostatic forces using reaction field electrostatics with Fennell [63] damping (Ewald-like convergence) derived by automatic parameter optimisation for precision f as for Ewald summation ($10^{-20} \leq f \leq 0.5$, default $f = 10^{-20}$)
regauss (every) n	resample the instantaneous system momenta distribution every n steps during equilibration (with respect to the last equilibration step)
replay (history)	abort simulation and replay HISTORY to recalculate structural properties such as RDFs, z-density profiles, defects and displacements trajectories (execution halts if no property is specified)
replay (history) **force**	abort simulation and replay the HISTORF file (a HISTORY copy) with full force evaluation driven by FIELD (different from the one used for the HISTORF generation)
restart	restart job from end point of previous run (i.e. continue current simulation, REVOLD required)

restart noscale	restart job from previous run without scaling system temperature (i.e. begin a new simulation from older run without temperature reset, REVOLD is not used)
restart scale	restart job from previous run with scaling system temperature (i.e. begin a new simulation from older run with temperature reset, REVOLD is not used)
rcut f ($\equiv$ **cut**off f)	act exactly the same as the **cut**off f option above
rlxtol f	set tolerance for relaxed shell model to f (default $f = 1$ in D Å ps^{-2})
rpad f ($\equiv$ **pad**ding f)	act exactly the same as **pad**ding f option above
rvdw (cutoff) f	set required short-ranged interactions cutoff to f Å
scale (temperature) (every) n	rescale system temperature every n steps during equilibration (with respect to the last equilibration step, (atomic velocities are scaled collectively)
seed n_1 n_2 n_3	seed control to the random number generator used in the generation of gaussian distributions and stochastic processes
shake (tolerance) f	set shake/rattle tolerance to f (default $f = 10^{-6}$)
shift	calculate electrostatic forces using force-shifted Coulomb sum
shift damp α	calculate electrostatic forces using force-shifted Coulomb sum with Fennell [63] damping (Ewald-like convergence) parameter α in Å^{-1}
shift precision f	calculate electrostatic forces using force-shifted Coulomb sum with Fennell [63] damping (Ewald-like convergence) derived by automatic parameter optimisation for precision f as for Ewald summation ($10^{-20} \leq f \leq 0.5$, default $f = 10^{-20}$)
slab	limits the number of processors in z-direction to 2 for slab simulations
spme evaluate (every) n	act exactly the same as **ewald evalu**ate (every) n
spme precision f	act exactly the same as **ewald precision** f
spme (sum) α k_1 k_2 k_3	calculate electrostatic forces using Ewald sum with α = Ewald convergence parameter in Å^{-1} $k1$ = is twice the maximum k-vector index in x-direction $k2$ = is twice the maximum k-vector index in y-direction $k3$ = is twice the maximum k-vector index in z-direction
stack (size) n	set rolling average stack to n timesteps

stats (every) n	accumulate statistics data every n timesteps
steps n	run simulation for n timesteps (default $n = 0$, corresponding to a ”dry” run)
temperature f	set required simulation temperature to f Kelvin (target temperature for constant temperature ensembles)
trajectory i j k	write HISTORY file with controls: i = start timestep for dumping configurations (default $i = 0$) j = timestep interval between configurations (default $j = 1$) k = data level (default $k = 0$, see Table 6.1)
vaf (sampling) (every) i (bin) (size) n	calculate and collect velocity autocorrelation function profiles every i timesteps (default $i = 50$) with a bin size set to n timesteps (default $n = 2i$, if i /ge 100, or $n = 100$ otherwise)
timestep f	set timestep to f ps
variable timestep f	variable timestep, start with timestep of f ps
vdw direct	enforces the direct calculation of van der Waals interactions defined by explicit potential forms, i.e. it will not work for systems using tabulated potentials (TABLE)
vdw mix *rule*	apply the mixing *rule* to all specified analytical, single-species van der Waals potential interactions to generate cross-species interactions when: (i) the latter has not already been specified; (ii) single species potentials are available (specified) and of the same type; and (iii) their type is allowed to mix, i.e. of 12-6 or LJ or DPD or AMOE or WCA type. The available mixing ***rule***s are as follows: *lore*ntz for the Lorentz-Berthelot type of mixing *fend*er for the Fender-Halsey type of mixing *hoge*rvorst for the Hogervorst (good hope) type of mixing *halg*ren for the Halgren HHG type of mixing *tang* for the Tang-Toennies type of mixing *func*tional for the Functional type of mixing The mixing formulae can be found in Section 2.3.1
vdw shift	apply a force-shifting procedure to all van der Waals potentials (except the shifted-force n-m potential) so that the VDW interactions’ energy and force contributions fall to zero smoothly for distances approaching $r_{\rm cut}$
zden (sampling) (every) f	calculate and collect the Z-density profile every f timesteps (default $f = 1$)
zero	perform zero temperature MD run (reset target system

	temperature 10 Kelvin)
*** SPECIAL OPTIONS ***	*** SPECIAL OPTIONS ***
l_dis	check and report on the minimum separation distance
l_fast	abandon global safety checks, which gives more speed to simulations run at too few link cells per domain regimes by abandoning global safety checks, however, in case of parallel failures no controlled manner termination will happen (OS's feedback dependence!!!)
l_his	generate a one frame HISTORY from CONFIG and terminate straight after
l_org x y z n	translate CONFIG along a vector (x, y, z), default (0,0,0), into CFGORG with configuration level n, default 0, after reading input & terminate
l_rin	read REVOLD in ASCII (default is binary)
l_rout	write REVIVE in ASCII (default is binary)
l_scr	redirect OUTPUT file's contents to the OS's default output channel (screen, interactive)
l_tor	abandon the production of REVIVE and REVCON between all Verlet neighbour list pairs at re/start

Note that in some cases additional keywords, shown in brackets "(...)", may also be supplied in the directives, or directives may be used in a long form. However, it is strongly recommended that the user uses only the **bold** part of these directives.

Table 6.1: Internal Trajectory/Defects File Key

`keytrj`	meaning
0	coordinates only in file
1	coordinates and velocities in file
2	coordinates, velocities and forces in file

6.1.1.3 Further Comments on the CONTROL File

1. A number of the directives (or their **mutually exclusive** alternatives) are **mandatory**:
   (a) **rcut** (≡ **cut**): specifying the short range forces cutoff. **It is compulsory in all circumstances as all DL_POLY_4 algorithms are directly or indirectly dependent on it.**
   (b) **temp** or **zero**: specifying the system temperature (not mutually exclusive but if **temp** has to proceed **zero** in CONTROL if **zero** is needed. **Use only one instance of these in CONTROL!** If a "dry run" is performed (see below) these can be omitted.
   (c) **timestep** or **variable timestep**: specifying the simulation timestep. **Use only one instance of these in CONTROL!** If a "dry run" is performed (see below) and a timestep length is not supplied a default one of 0.001 ps is provided.

(d) **ewald/spme sum/precision** or **coul** or **shift** or **distan** or **reaction** or **no elec**: specifying the required coulombic forces option. Apart from **no elec** the rest of the directives are mutually exclusive from one another. **If none is specified then none is applied!**

2. Some directives are optional. If not specified DL_POLY_4 may give default values if necessary. (Some but not all defaults are specified above in the list of directives.) However fail-safe DL_POLY_4 is, not always will it assume a default value for certain parameters. To enable DL_POLY_4 to be even more liberal in the fail-safe features, users are recommended to use **no strict** option.

3. The **steps** and **equilibration** directives have a default of zero. If not used or used with their default values a "dry run" is performed. This includes force generation and system dump (REVCON and REVIVE) and, depending on the rest of the options, may include; velocity generation, force capping, application of the CGM minimiser, application of the pseudo thermostat, and dumps of HISTORY, DEFECTS, RDFDAT, ZDNDAT and MSDTMP. **Note** that, since no actual dynamics is to be performed, the **temperature** and **pressure** directives do not play any role and are therefore not necessary.

4. If the CGM minimiser, **minim**ise, is specified with zero frequency, it is only applied at timestep zero if **equilibration** $\geq$ **steps** (i.e. optimise structure at start only!). This is equvalent to using the **optim**ise directive. In this way it can be used as a configuration optimiser at the beginning of the equilibration period or when a "dry run" (**steps** = 0) is performed (i.e. equilibrate without any actual dynamics!). **Note** that the CGM search algorithm stepping uses a step that is proportional to the instantaneous value of the **timestep** and thus its usage in may lead to algorithm's instability and failure in the cases when optimisation is applied before any dynamics occurs in a model system in an unphysical state (i.e. much away from equilibrium) and/or with an ill/badly defined forcefield for the state. Hence, special care should be taken when the option is used as **optimise** or/and in a "dry run" mode with a timestep value too large for the model system state.

5. The **variable timestep** (or also **timestep variable**) option requires the user to specify an initial guess for a reasonable timestep for the system (in picoseconds). The simulation is unlikely to retain this as the operational timestep however, as the latter may change in response to the dynamics of the system. The option is used in conjunction with the default values of **maxdis** (0.10 Å) and **mindis** (0.03 Å), which can also be optionally altered if used as directives (note the rule that **maxdis** $>$ 2.5 **mindis** applies). Also, an additional **mxstep** (in ps) control can be applied. These serve as control values in the variable timestep algorithm, which calculates the greatest distance a particle has travelled in any timestep during the simulation. If the maximum distance is exceeded, the timestep variable is halved and the step repeated. If the greatest move is less than the minimum allowed, the timestep variable is doubled and the step repeated provided it does not exceed the user specified **mxstep**. If it does then it scales to **mxstep** and the step is repeated. In this way the integration timestep self-adjusts in response to the dynamics of the system. **Note** that this option is abandoned when used in conjunction with DPD thermostats (**ensemble nvt dpd*type*** γ), since it cannot be applied efficiently and furthermore it is not well defined in a DPD sense either.

6. The **job time** and **close time** directives are required to ensure a controlled close down procedure when a job runs out of time. The time specified by the **job time** directive indicates the total time allowed for the job. (This must obviously be set equal to the time specified to the operating system when the job is submitted.) The **close time** directive represents the time DL_POLY_4 will require to write and close all the data files at the end of processing. This means the *effective* processing time limit is equal to the job time minus the close time. Thus when DL_POLY_4 reaches the effective job time limit it begins the close down procedure with enough time in hand to ensure the files are correctly written. In this way you may be sure the restart files etc. are complete when the job terminates. **Note** that setting the close time too small will mean the job will crash before the files have been finished. If it is set too large DL_POLY_4 will begin closing down too early. How large the close time needs to

be to ensure safe close down is system dependent and a matter of experience. It generally increases with increasing simulation system size.

7. The starting options for a simulation are governed by the keyword **restart**. If this is **not** specified in the control file, the simulation will start as new. When specified, it will continue a previous simulation (**restart**) provided all needed restart files are in place and not corrupted. If they are not in place or are found corrupted, it will start a new simulation without initial temperature scaling of the previous configuration (**restart noscale**). Internally these options are handled by the integer variable `keyres`, which is explained in Table 6.2.

Table 6.2: Internal Restart Key

`keyres`	meaning
0	start new simulation from CONFIG file and assign velocities from Gaussian distribution
1	continue current simulation
2	start new simulation from CONFIG file and rescale velocities to desired temperature
3	start new simulation from CONFIG file and **do not** rescale velocities

8. The various **ensemble** options (i.e. **nve**, **nvt evans**, **nvt ander**sen, **nvt lang**evin, **nvt ber**endsen, **nvt hoover**, **npt lang**evin, **npt ber**endsen, **npt hoover**, **npt mtk**, **nst lang**evin, **nst ber**endsen, **nst hoover**, **nst mtk**) are mutually exclusive, though none is mandatory (the default is the NVE ensemble). These options are handled internally by the integer variable `keyens`. The meaning of this variable is explained in Table 6.3. The **nst** keyword is also used in the NσT ensembles extension to NP$_n$AT and NP$_n\gamma$T ones. **Note** that these semi-isotropic ensembles are only correct for infinite interfaces placed perpendicularly to the z axis! This means that the interface is homogenious (unbroken) and continuous in the (x,y) plane of the MD cell, which assumes that that two of the cell vectors have a cross product only in the z direction. (For example, if the MD box is defined by its lattice vectors $(\underline{a}, \underline{b}, \underline{c})$ then $\underline{a} \times \underline{b} = \pm(0, 0, 1)$.) It is the users' responsibility to ensure this holds for their model system.

Table 6.3: Internal Ensemble Key

`keyens`	meaning
0	Microcanonical ensemble (NVE)
1	Evans NVT ensemble (NVE$_{kin}$)
10	Langevin NVT ensemble
11	Berendsen NVT ensemble
12	Nosé-Hoover NVT ensemble
20	Langevin NPT ensemble
21	Berendsen NPT ensemble
22	Nosé-Hoover NPT ensemble
23	Martyna-Tuckerman-Klein NPT ensemble
30	Langevin N$\underline{\underline{\sigma}}$T ensemble
31	Berendsen N$\underline{\underline{\sigma}}$T ensemble
32	Nosé-Hoover N$\underline{\underline{\sigma}}$T ensemble
33	Martyna-Tuckerman-Klein N$\underline{\underline{\sigma}}$T ensemble

9. The **zero** directive, enables a "zero temperature" optimisation. The target temperature of the simu-

lation is reset to 10 Kelvin and a crude energy minimiser:

$$\underline{v}_i \leftarrow \begin{cases} 0 & : \quad \underline{v}_i \cdot \underline{f}_i < 0 \\ \underline{f}_i \, \dfrac{\underline{v}_i \cdot \underline{f}_i}{\underline{f}_i \cdot \underline{f}_i} & : \quad \underline{v}_i \cdot \underline{f}_i \geq 0 \end{cases} \tag{6.1}$$

is used to help the system relax before each integration of the equations of motion (measures are taken to conserve the MD cell momentum). This must not be thought of as a true energy minimization method. **Note** that this optimisation is applied irrespectively of whether the simulation runs in equilibration or statistical mode.

The algorithm is developed in the DL_POLY_4 routine ZERO_K_OPTIMISE.

10. The **impact** $i\ j\ E\ x\ y\ z$ directive will not be activated if the particle index is beyond the one of the last particle. The option will fail in a controlled manner at application time if the particle is found to be in a frozen state or the shell of an ion or part of a rigid body. During application the center of mass momentum is re-zeroed to prevent any drifts. The user must take care to have the impact initiated after any possible equilibration. Otherwise, the system will be thermostatted and the impact energy dissipated during the equilibration.

11. The **pseudo** option is intended to be used in highly non-equilibrium simulations when users are primarily interested in the structural changes in the core of the simulated system as the the MD cell boundaries of the system are coupled to a thermal bath.

    The thermal bath can be used with two types of temperature scaling algorithms - **(i)** Langevin (stochastic thermostat), **(ii)** Gauss and **(iii)** Direct (direct thermostat). If no type is specified then the Langevin temperature control algorithm is applied first followed the Direct one. The user is also required to specify the width of the pseudo thermostat, f_1 (in Å), which must be larger than 2 Å and less than or equal to a quarter of minimum width of the MD cell. The thermostat is an f_1 Å thick buffer layer attached on the inside at the MD cell boundaries.

    The temperature of the bath is specified by the user, $T = f_2$ (in Kelvin), which must be larger than 1 Kelvin. If none is supplied by the user, T defaults to the system target temperature.

    - **pseudo langevin**
      The stochasticity of the Langevin thermostat emulates an infinite environment around the MD cell, providing a means for "natural" heat exchange between the MD system and the heath bath thus aiding possible heat build up in the system. In this way the instantaneous temperature of the system is driven naturally towards the bath temperature. Every particle within the thermostat buffer layer is coupled to a viscous background and a stochastic heat bath, such that

$$\begin{aligned} \frac{d\underline{r}_i(t)}{dt} &= \underline{v}_i(t) \\ \frac{d\underline{v}_i(t)}{dt} &= \frac{\underline{f}_i(t) + \underline{R}_i(t)}{m_i} - \chi(t)\ \underline{v}_i(t) \quad , \end{aligned} \tag{6.2}$$

      where $\chi(t)$ is the friction parameter from the dynamics in the the MD cell and $R(t)$ is stochastic force with zero mean that satisfies the fluctuation-dissipation theorem:

$$\left\langle R_i^\alpha(t)\ R_j^\beta(t') \right\rangle = 2\ \chi(t)\ m_i\ k_B T\ \delta_{ij}\ \delta_{\alpha\beta}\ \delta(t-t') \quad , \tag{6.3}$$

      where superscripts denote Cartesian indices, subscripts particle indices, k_B is the Boltzmann constant, T the bath temperature and m_i the particle's mass. The algorithm is implemented in routine PSEUDO and has two stages:

      – Generate random forces on all particles within the thermostat. Here, care must be exercised to prevent introduction of non-zero net force when the random forces are added to the system force field.

  - Rescale the kinetic energy of the thermostat bath so that particles within have Gaussian distributed kinetic energy with respect to the target temperature and determine the (Gaussian constraint) friction within the thermostat:

$$\chi(t) = Max \left(0, \frac{\sum_i [\vec{f_i}(t) + \vec{R_i}(t)] \cdot \vec{v_i}(t)}{\sum_i m_i \ \vec{v}_i^2(t)} \right) \quad . \tag{6.4}$$

    Care must be exercised to prevent introduction of non-zero net momentum. (Users are reminded to use for target temperature the temperature at which the original system was equilibrated in order to avoid simulation instabilities.)

  The effect of this algorithm is to relax the buffer region of the system on a local scale and to effectively dissipate the incoming excess kinetic energy from the rest of the system, thus emulating an infinite-like environment surrounding the MD cell. The thermostat width matters as the more violent the events on the inside of the MD cell, the bigger width may be needed in order to ensure safe dissipation of the excess kinetic energy.

- **pseudo Gauss**
  - Rescale the kinetic energy of the thermostat bath so that particles within have Gaussian distributed kinetic energy with respect to the target temperature.
- **pseudo direct**

  The Direct thermostat is the simplest possible model allowing for heat exchange between the MD system and the heath bath. All (mass, non-frozen) particles within the bath have their kinetic energy scaled to 1.5 $k_B T$ at the end of each time step during the simulation. Care is exercised to prevent introduction of non-zero net momentum when scaling velocities. (Users are reminded to use for target temperature the temperature at which the original system was equilibrated in order to avoid simulation instabilities.) Due to the "unphysical" nature of this temperature control the thermostat width does not matter to the same extent as in the case of the Langevin thermostat.

Note that embedding a thermostat in the MD cell walls is bound to produce **wrong ensemble averages**, and instantaneous pressure and stress build-ups at the thermostat boundary. Therefore, ensembles lose their meaning as such and so does the conserved quantity for true ensembles. *If the* **pseudo** *thermostat option is specified without any type of temperature control in CONTROL then both types will be applied in the order Langevin→Direct at each time step during the simulation.*

The algorithms are developed in the DL_POLY_4 routines PSEUDO_VV and PSEUDO_LFV respectively.

12. The **defe**cts option will trigger reading of REFERENCE (see Section 6.1.4), which defines a reference MD cell with particles' positions defining the crystalline lattice sites. If REFERENCE is not found the simulation will either
(i) halt if the simulation has been restarted, i.e. is a continuation of an old one - the **restart** option is used in CONTROL and the REVOLD (see Section 6.1.5) file has been provided.
or
(ii) recover using CONFIG (see Section 6.1.2) if it is a new simulation run, i.e **restart** option is not used in CONTROL or REVOLD has not been provided.

    The actual defect detection is based on comparison of the simulated MD cell to the reference MD cell based on a user defined site-interstitial cutoff, R_{def},

$$\text{Min}\,[0.3, r_{\text{cut}}/3] \ \ \mathring{A} \ \leq \ R_{def} \ \leq \ \text{Min}\,[1.2, r_{\text{cut}}/2] \ \ \mathring{A} \tag{6.5}$$

    with a default value of Min$\,[0.75, r_{\text{cut}}/3]$ Å. (If the supplied value exceeds the limits the simulation execution will halt). If a particle, p, is located in the vicinity of a site, s, defined by a sphere with its centre at this site and a radius, R_{def}, then the particle is a *first hand* claimee of s, and the site is not vacant. Otherwise, the site is presumed vacant and the particle is presumed a general interstitial. If a site, s, is claimed and another particle, $p\prime$, is located within the sphere around it, then $p\prime$ becomes an interstitial associated with s. After all particles and all sites are considered, it is clear which sites are vacancies. Finally, for every claimed site, distances between the site and its *first hand* claimee and interstitials are compared and the particle with the shortest one becomes the *real* claimee. If a *first hand* claimee of s is not the *real* claimee it becomes an interstitial associated with s. At this stage it is clear which particles are interstitials. The sum of interstitials and vacancies gives the total number of defects in the simulated MD cell.

Frozen particles and particles detected to be shells of polarisable ions are not considered in the defect detection.

Note that the algorithm cannot be applied safely if R_{def} is larger than half the shortest interatomic distance within the reference MD cell since a particle may; (i) claim more than one site, (ii) be an interstitial associated with more than one site, or both (i) and (ii). On the other hand, low values of R_{def} are likely to lead to slight overestimation of defects.

If the simulation and reference MD cell have the same number of atoms then the total number of interstitials is always equal to the total number of defects.

13. The **disp**lacements option will trigger dump of atom displacements based on a qualifying cutoff in a trajectory like manner. Dsiplacemets of atoms from their original position at the end of equilibration (the start of statistics), $t\ = 0$, is carried out at each timestep.

14. The tolerance for relaxed shell model **rlxtol**, is a last resort option to aid shell relaxation of systems with very energetic and/or rough potential surface. Users are advised to use it with caution, should there really need be, as the use of high values may result in physically incorrect dynamics.

15. The difference between the directives **ewald** and **spme** is only in the **ewald/spme sum** directive, in which the **ewald sum** specifies the indices of the maximum k-vector, whereas the **spme sum** the dimensions of the 3D charge array (which are exactly twice the maximum k-vector indices). **Note** that in either case, DL_POLY_4 will carry out the SPME coulombic evaluation.

16. The force selection directives **ewald/spme sum/precision**, **reaction**, **coul**, **shift**, **dist**, **no elec** are handled internally by the integer variable `keyfce`. See Table 6.4 for an explanation of this variable. **Note** that all these options with the exception of the last, **no elec**, are mutually exclusive.

Table 6.4: Electrostatics Key

`keyfce`	meaning
	Electrostatics are evaluated as follows:
0	Ignore electrostatic interactions
2	SPM Ewald summation
4	Coulomb sum with distance dependent dielectric
6	Standard truncated Coulomb sum
8	Force-shifted Coulomb sum
10	Reaction field electrostatics

17. **ewald evaluate** or **multiple** are not mutually exclusive and it is the first instance of these in CONTROL that is read and applied in the following simulation.

18. The choice of reaction field electrostatics (directive **reaction**) relies on the specification of the relative dielectric constant external to the cavity. This is specified by the **eps** directive.

19. The directive **ewald/spme evaluate** is only triggered when **ewald/spme sum/precision** is present. It sets an infrequent evaluation of the k-space contributions to the Ewald summation. Although this option decreases the simulation cost it also inherently decreases the accuracy of the dynamics. **Note**

that the usage of this feature may lead to inacuarte or even wrong and unphysical dynamics as the less frequent the evaluation, the greater the inacuarcy.

20. DL_POLY_4 uses two different potential cutoffs. These are as follows:

    (a) r_{cut} - the universal cutoff set by **cutoff**. It applies to the real space part of the electrostatics calculations and to the van der Waals potentials if no other cutoff is applied.

    (b) r_{vdw} - the user-specified cutoff for the van der Waals potentials set by **rvdw**. If not specified its value defaults to r_{cut}.

21. Constraint algorithms in DL_POLY_4, SHAKE/RATTLE (see Section 3.2), use default iteration precision of 10^{-6} and limit of iteration cycles of 250. Users may experience that during optimisation of a new built system containing constraints simulation may fail prematurely since a constraint algorithm failed to converge. In such cases directives **mxshak** (to increase) and **shake** (to decrease) may be used to decrease the strain in the system and stablise the simulation numerics until equilibration is achieved.

22. DL_POLY_4's DD strategy assumes that the local (per domain/node or link cell) density of various system entities (i.e. atoms, bonds, angles, etc.) does not vary much during a simulation and some limits for these are assumed empirically. This may not the case in extremely non-equilibrium simulations, where the assumed limits are prone to be exceeded or in some specific systems where these do not hold from the start. A way to tackle such circumstances and avoid simulations crash (by controlled termination) is to use the **densvar** *f* option. In the SET_BOUNDS subroutine DL_POLY_4 makes assumptions at the beginning of the simulation and corrects the lengths of bonded-like interaction lists arrays (`mxshl, mxcons, mxrgd, mxteth, mxbond, mxangl, mxdihd, mxinv`) as well as the lengths of link-cell (`mxlist`) and domain (`mxatms, mxatdm`) lists arrays when the option is activated with $f > 0$. Greater values of f will correspond to allocation bigger global arrays and larger memory consumption by DL_POLY_4 during the simulation. **Note** that this option may demand more memory than available on the computer architecture. In such cases DL_POLY_4 will terminate with an array allocation failure message.

23. As a default, DL_POLY_4 does not store statistical data during the equilibration period. If the directive **collect** is used, equilibration data will be incorporated into the overall statistics.

24. The **vaf** directive switches on velocity autocorrelation function (VAF) calculations for individual atomic species in DL_POLY_4 after equilibration or immediately at start if the directive **collect** is used. It controls how often VAF profiles are started what the size of each profile (in timesteps). Overlapping profiles are possible and require more memory to store them (and initial velocities) while they are being calculated. By default DL_POLY_4 will report time-averaged VAF profiles. This can be overridden using the **no vafav**eraging directive, which will instead report individual 'instantaneous' VAF profiles.

25. **io *action*** [***options***] controls how I/O is performed by DL_POLY_4. The options can help the performance of I/O operations within DL_POLY_4 for potentially large files during the run. The form of the command depends on the value of **action**, which may take the value either **read** or **write**. In general, this command should only be used for tuning the I/O subsystem in DL_POLY_4 for large runs. For small to average sized systems the built-in defaults usually suffice.

    (a) **io read *method*** [***options***]
    With action set to **read** the **io** command controls how the reading of large files is performed. *method* controls how the disk is accessed. Possible values are *mpiio*, in which case MPI-I/O is used, *direct*, which uses parallel FORTRAN direct access files, and *master* which performs all I/O through a master processor, or *netcdf* for netCDF I/O provided DL_POLY_4 is compiled in a netCDF-enabled mode. *mpiio* is the recommended method, and for large systems *master* should be avoided. Available options depend on which method is to be used, and all are optional in each

case. Where numerical values are to be supplied specifying 0 or a negative numbers indicates that DL_POLY_4 will resort to the default value. The possible options are:

- **io read mpiio|direct|netcdf** [*j* [*k* [*l* [*e*]]]]
  j specifies the number of processors that shall access the disk. *k* specifies the maximum number of particles that the reading processors shall deal with at any one time. Large values give good performance, but may results in an unacceptable memory overhead. *l* specifies the maximum number of particles that the reading processors shall read from the disk in one I/O transaction. Large values give good performance, but may results in an unacceptable memory overhead. *e* accepts *Y*es only to switch global error checking performed by the I/O subsystem, the default is *N*o.
- **io read master** [*l*]
  l specifies the maximum number of particles that the reading process shall read from the disk in one I/O transaction. Large values give good performance, but may results in an unacceptable memory overhead.

(b) **io writ**e ***method*** [***rp***] ***type*** [***options***]

With action set to **writ**e the **io** command controls how the writing of large files is performed. *method* controls how the disk is accessed. Possible values are *mpiio*, in which case MPI-I/O is used, *direct*, which uses parallel FORTRAN direct access files, and *master* which performs all I/O through a master processor, or *netcdf* for netCDF I/O provided DL_POLY_4 is compiled in a netCDF-enabled mode. *mpiio* is the recommended method, and for large systems *master* should be avoided and also THE DIRECT OPTION IS NOT STRICTLY PORTABLE, and so may cause problems on some machines. *rp* is an optional specification, only applicable to *netcdf* method for opting the binary precision for real numbers. It only takes *32bit* or *amber* for 32-bit (float) precision, otherwise 64-bit (double) precision is defaulted. *type* controls the ordering of the particles on output. Possible values are *sort*ed and *unsort*ed. *sort*ed ensures that the ordering of the particles the default - sequential, ascending. Whereas *unsort*ed uses the natural internal ordering of DL_POLY_4 which changes during the simulation. The recommended and default value is *sort*ed. If none is specified DL_POLY_4 defaultes to the *sort*ed type of I/O. It should be noted that the overhead of the *sort*ed otion compared to the *unsort*ed is usually very small. Available options depend on which method is to be used, and all are optional in each case. Where numerical values are to be supplied specifying 0 or a negative numbers indicates that DL_POLY_4 will resort to the default value. The possible options are:

- **io writ**e **mpiio|direct|netcdf [rp] sort|unsort** [*j* [*k* [*l* [*e*]]]]
  j specifies the number of processors that shall access the disk. *k* specifies the maximum number of particles that the writing processors shall deal with at any one time. Large values give good performance, but may results in an unacceptable memory overhead. *l* specifies the maximum number of particles that the writing processors shall write to the disk in one I/O transaction. Large values give good performance, but may results in an unacceptable memory overhead. *e* accepts *Y*es only to switch global error checking performed by the I/O subsystem, the default is *N*o.
- **io write master sort|unsort** [*l*]
  l specifies the maximum number of particles that the writing process shall write to the disk in one I/O transaction. Large values give good performance, but may results in an unacceptable memory overhead.

26. The **no vom** option will trigger a default bypass of the GETVOM routine which will return zero and thus no COM removal will happen. **Note that this will lead to COM momentum accumulation for many though not all ensembles!**. Such accumulation will propagate to the generation of flow in the MD cell and ultimately suppress the thermal motion of the particles in the system, leading to the so called "frozen ice cube effect"! It is worth nothing that this option must be turned on for the correct application of stochastic dynamics via the langevin temperature control (NVT Langevin)! If

the option is not applied then the dynamics will lead to peculiar thermalisation of different atomic species to mass- and system size-dependent temperatures.

27. The **replay force** option will attempt to open a renamed HISTORY as HISTORF and instead following any integration scheme specified in CONTROL read the positions as integration outcomes and proceed with the usual MD cycle. There are a couple of restrictions that may lead to failures of this feature's application.
    - The CONFIG file at (re)start must be the same as the frame in HISTORF that CONTROL (REVOLD) attempts to (re)start from.
    - If any two HISTROF consecutive frames are too far apart in time from each other then domain information reallocation follow-up may break when the feature is used in parallel.
28. The **rpad** ($\equiv$ **pad**) option will add extra distance, $r_{\rm pad}$, if larger than $f \geq {\rm Min}[0.05, 0.5\%.r_{\rm cut}]$ Å, to the major cutoff, $r_{\rm cut}$, to construct a larger link-cell width, $r_{\rm lnk} = r_{\rm cut} + r_{\rm pad}$, which will trigger a construction of a larger Verlet neighbour list (VNL) while at the same time facilitate its conditional update, rather at every timestpe. The VNL conditaional update is check at the end of each tiemstep and triggered only when the most travelled particle has moved a distance larger than $r_{\rm pad}/2$. It is worth noting that padding is at expense of extra memory but if used wisely it could improve time to solution from 10% to 100% depending on force-field complexity. If it is too large or too small (that is why the $f \geq {\rm Min}[0.05, 0.5\%.r_{\rm cut}]$ Å limit) it will lead to performace degradation. It is recomended that $r_{\rm pad}$ is set up at a value of $\approx 1 \div 5\%$ of the cutoff, $r_{\rm cut}$, as long as the major link-cell algorithm uses a link-cell decomposition that not worse than $4 \otimes 4 \otimes 4$ per domain. For such setups, in practice, one may expect average compute speedups [1] of the order of $10 \div 30\%$ for force-fields involving the Ewald summation methodology and $60 \div 100\%$ for force-fields without electrostatics evaluations involving the Ewald summation methodology.

Users are advised to study the example CONTROL files appearing in the *data* sub-directory to see how different files are constructed.

6.1.2 The CONFIG File

The CONFIG file contains the dimensions of the unit cell, the key for periodic boundary conditions and the atomic labels, coordinates, velocities and forces. This file is read by the subroutine READ_CONFIG (optionally by SCAN_CONFIG) in the SET_BOUNDS routine. The first few records of a typical CONFIG file are shown below:

```
IceI structure 6x6x6 unit cells with proton disorder
         2         3                276
  26.988000000000000   0.000000000000000   0.000000000000000
 -13.494000000000000  23.372293600000000   0.000000000000000
   0.000000000000000   0.000000000000000  44.028000000000000
       OW         1
    -2.505228382        -1.484234330        -7.274585343
    0.5446573999        -1.872177437       -0.7702718106
     3515.939287         13070.74357         4432.030587
       HW         2
    -1.622622646        -1.972916834        -7.340573742
     1.507099154        -1.577400769         4.328786484
     7455.527553        -4806.880540        -1255.814536
```

[1]I.e. I/O effects are excluded from comparison with a default simualtion and comparisons are carried over a few hundreds of timesteps. This is usually accounting for over 90% of the time tosolution.

```
      HW         3
    -3.258494716        -2.125627191        -7.491549620
     2.413871957        -4.336956694         2.951142896
    -7896.278327        -8318.045939        -2379.766752
      OW         4
    0.9720599243E-01    -2.503798635        -3.732081894
     1.787340483        -1.021777575        0.5473436377
     9226.455153         9445.662860         5365.202509
```

etc.

6.1.2.1 The CONFIG File Format

The file is free-formatted and not case sensitive. Every line is treated as a command sentence (record). However, line records are limited to 72 characters in length. Records are read in words, as a word must not exceed 40 characters in length. Words are recognised as such by separation by one or more space characters. The first record in the CONFIG file is a header (up to 72 characters long) to aid identification of the file. Blank and commented lines are not allowed.

6.1.2.2 Definitions of Variables in the CONFIG File

record 1		
`header`	a72	title line
record 2		
`levcfg`	integer	CONFIG file key. See Table 6.5 for permitted values
`imcon`	integer	Periodic boundary key. See Table 6.6 for permitted values
`megatm`	integer	Optinal, total number of particles (crystalographic entities)
record 3	omitted if `imcon` = 0	
`cell(1)`	real	x component of the a cell vector in Å
`cell(2)`	real	y component of the a cell vector in Å
`cell(3)`	real	z component of the a cell vector in Å
record 4	omitted if `imcon` = 0	
`cell(4)`	real	x component of the b cell vector in Å
`cell(5)`	real	y component of the b cell vector in Å
`cell(6)`	real	z component of the b cell vector in Å
record 5	omitted if `imcon` = 0	
`cell(7)`	real	x component of the c cell vector in Å
`cell(8)`	real	y component of the c cell vector in Å
`cell(9)`	real	z component of the c cell vector in Å

Note that **record 2** may contain more information apart from the mandatory as listed above. If the file has been produced by DL_POLY_4 then it also contains other items intended to help possible parallel I/O reading. Also, it is worth mentioning that the periodic boundary conditions (PBC), as specified in Table 6.6 and described in detail in Appendix B, refer generally to a **triclinic type of super-cell**, for which there are **no symmetry assumptions! Records 3, 4 and 5** contain the Cartesian components of the super-cell's lattice vectors in Å. DL_POLY_4 can only tract triclinic type of super-cells as the only types of super-cell shapes that are commensurate with the domain decomposition (DD) parallelisation strategy of it. However, this is not a restriction for the replicated data (RD) parallelisation that DL_POLY_Classic adopts and thus it can also accept truncated octahedral and rhombic dodecahedral periodic boundaries.

Subsequent records consists of blocks of between 2 and 4 records depending on the value of the `levcfg` variable. Each block refers to one atom. The atoms **do not** need to be listed sequentially in order of increasing index. Within each block the data are as follows:

record i		
`atmnam`	a8	atom name
`index`	integer	atom index
record ii		
`xxx`	real	x coordinate in Å
`yyy`	real	y coordinate in Å
`zzz`	real	z coordinate in Å
record iii	included only if `levcfg` > 0	
`vxx`	real	x component of velocity
`vyy`	real	y component of velocity
`vzz`	real	x component of velocity
record iv	included only if `levcfg` > 1	
`fxx`	real	x component of force
`fyy`	real	y component of force
`fzz`	real	z component of force

Note that on **record i** only the atom name is strictly mandatory, any other items are not read by DL_POLY_Classic but may be added to aid alternative uses of the file, for example alike DL_POLY GUI [21], DL_POLY_Classic assume that the atoms' indices are in a sequentially ascending order starting form 1. However, DL_POLY_4 needs the index or the **no ind**ex option needs to be specified in the CONTROL file! It is worth mentioning that DL_POLY_4 (as well as DL_POLY_Classic) assumes that the origin of Cartesian system with respect to which the particle positions are specified is the middle of MD cell. Also, as both the cell vectors and the particles' positions are specified in Å, there is a **fine connection** between them! This would not be the case if the particles' positions were kept in reduced space with fractional coordinates. Last but not least, it is worth pointing out that composite entities, such as velocities and forces, have their units expressed as composites of the DL_POLY units as shown in Section 1.3.7.

Table 6.5: CONFIG File Key (record 2)

`levcfg`	meaning
0	coordinates included in file
1	coordinates and velocities included in file
2	coordinates, velocities and forces included in file

Table 6.6: Periodic Boundary Key (record 2)

`imcon`	meaning
0	no periodic boundaries
1	cubic boundary conditions
2	orthorhombic boundary conditions
3	parallelepiped boundary conditions
6	x-y parallelogram boundary conditions with no periodicity in the z direction

6.1.2.3 Further Comments on the CONFIG File

The CONFIG file has the same format as the output file REVCON (Section 6.2.7). When restarting from a previous run of DL_POLY_4 (i.e. using the **restart**, **restart noscale** or **restart scale** directives in the

CONTROL file - above), the CONFIG file must be replaced by the REVCON file, which is renamed as the CONFIG file. The *copy* macro in the *execute* sub-directory of DL_POLY_4 does this for you.

The CONFIG file has the same format as the optional output file CFGMIN, which is only produced when the **minimise** (**optimise**) option has been used during an equilibration simulation or a "dry run".

6.1.3 The FIELD File

The FIELD file contains the force field information defining the nature of the molecular forces. This information explicitly includes the (site) topology of the system which sequence **must** be matched (implicitly) in the crystallographic description of the system in the CONFIG file. The FIELD file is read by the subroutine READ_FIELD. (It is also read by the subroutine SCAN_FIELD in the SET_BOUNDS routine.) Excerpts from a force field file are shown below. The example is the antibiotic Valinomycin in a cluster of 146 water molecules.

```
Valinomycin Molecule with 146 SPC Waters
UNITS kcal

MOLECULES    2
Valinomycin
NUMMOLS 1
ATOMS 168
    O        16.0000     -0.4160         1
    OS       16.0000     -0.4550         1
    "          "           "        "
    "          "           "        "
    HC         1.0080      0.0580         1
    C        12.0100      0.4770         1
BONDS 78
harm   31   19 674.000     1.44900
harm   33   31 620.000     1.52600
    "    "    "    "         "
    "    "    "    "         "
harm  168   19 980.000     1.33500
harm  168  162 634.000     1.52200
CONSTRAINTS 90
   20   19    1.000017
   22   21    1.000032
    "    "     "
    "    "     "
  166  164    1.000087
  167  164    0.999968
ANGLES 312
harm   43    2   44  200.00      116.40
harm   69    5   70  200.00      116.40
  "    "    "    "     "            "
  "    "    "    "     "            "
harm   18  168  162  160.00      120.40
harm   19  168  162  140.00      116.60
DIHEDRALS 371
harm    1   43    2   44  2.3000      180.00
harm   31   43    2   44  2.3000      180.00
```

```
  "     "     "     "     "      "            "
  "     "     "     "     "      "            "
cos   149   17   161   16  10.500      180.00
cos   162   19   168   18  10.500      180.00
FINISH
SPC Water
NUMMOLS 146
ATOMS  3
    OW       16.0000     -0.8200
    HW        1.0080      0.4100
    HW        1.0080      0.4100
CONSTRAINTS  3
    1    2    1.0000
    1    3    1.0000
    2    3    1.63299
FINISH
VDW   45
C       C          lj   0.12000      3.2963
C       CT         lj   0.08485      3.2518
"       "          "     "             "
"       "          "     "             "
"       "          "     "             "
OW      OS         lj   0.15100      3.0451
OS      OS         lj   0.15000      2.9400
CLOSE
```

6.1.3.1 The FIELD File Format

The file is free-formatted and not case-sensitive. Every line is treated as a command sentence (record). Commented records (beginning with a #) and blank lines are not processed and may be added to aid legibility (see example above). Records must be limited in length to 100 characters. Records are read in words, as a word must not exceed 40 characters in length. Words are recognised as such by separation by one or more space characters. The contents of the file are variable and are defined by the use of **directives**. Additional information is associated with the directives.

6.1.3.2 Definitions of Variables in the FIELD File

The file divides into three sections: general information, molecular descriptions, and non-bonded interaction descriptions, appearing in that order in the file.

General information

The first viable record in the FIELD file is the title. The second is the **units** directive. Both of these are mandatory.

record 1		
header	a100	field file header
record 2		
units	a40	Unit of energy used for input and output

The energy units on the **units** directive are described by additional keywords:

a. eV, for electron-Volts

b. kcal/mol, for k-calories per mol

c. kJ/mol, for k-Joules per mol

d. Kelvin/Boltzmann, for Kelvin per Boltzmann

e. internal, for DL_POLY internal units (10 Joules per mol).

If no units keyword is entered, DL_POLY internal units are assumed for both input and output. The **units** directive only affects the input and output interfaces, all internal calculations are handled using DL_POLY units. System input and output energies are read in **units per MD cell**.

Note that all energy bearing potential parameters are read in terms of the specified energy units. If such a parameter depends on an angle then the dependence is read in terms of radians although the following angle in the parameter sequence is read in terms of degrees.

Molecular details

It is important for the user to understand that there is an organisational correspondence between the FIELD file and the CONFIG file described above. It is required that the order of specification of molecular types and their atomic constituents in the FIELD file follows the order of indices in which they appear in the CONFIG file. Failure to adhere to this common sequence will be detected by DL_POLY_4 and result in premature termination of the job. It is therefore essential to work from the CONFIG file when constructing the FIELD file. It is not as difficult as it sounds!

The entry of the molecular details begins with the mandatory directive:

molecules *n*

where n is an integer specifying the number of different *types* of molecule appearing in the FIELD file. Once this directive has been encountered, DL_POLY_4 enters the *molecular description* environment in which only molecular description keywords and data are valid.

Immediately following the **molecules** directive, are the records defining individual molecules:

1. *name-of-molecule*
   which can be any character string up to 100 characters in length. (**Note**: this is not a directive, just a simple character string.)

2. **nummols *n***
   where n is the number of times a molecule of this type appears in the simulated system. The molecular data then follow in subsequent records:

3. **atoms *n***
   where n indicates the number of atoms in this type of molecule. A number of records follow, each giving details of the atoms in the molecule i.e. site names, masses and charges. Each record carries the entries:

`sitnam`	a8	atomic site name
`weight`	real	atomic site mass
`chge`	real	atomic site charge
`nrept`	integer	repeat counter
`ifrz`	integer	'frozen' atom (if `ifrz` $>$ 0)

The integer `nrept` need not be specified if the atom/site is not frozen (in which case a value of 1 is assumed.) A number greater than 1 specified here indicates that the next (`nrept` - 1) entries in the CONFIG file are ascribed the atomic characteristics given in the current record. The sum of the repeat numbers for all atoms in a molecule should equal the number specified by the **atoms** directive.

4. **shell *n***
   where n is the number of core-shell units. Each of the subsequent n records contains:

index 1 (i)	integer	site index of core
index 2 (j)	integer	site index of shell
k_2	real	force constant of core-shell spring
k_4	real	quartic (anharmonic) force constant of spring

   The spring potential is
   $$U(r) = \frac{1}{2}k_2 r_{ij}^2 + \frac{1}{4}k_4 r_{ij}^4 \quad , \tag{6.6}$$
   with the force constant k_2 entered in units of `engunit`×Å^{-2} and k_4 in `engunit` Å^{-4}, where usually $k_2 >> k_4$. The `engunit` is the energy unit specified in the **units** directive.

   Note that the atomic site indices referred to above are indices arising from numbering each atom in the molecule from 1 to the number specified in the **atoms** directive for this molecule. This same numbering scheme should be used for all descriptions of this molecule, including the **constraints**, **pmf**, **rigid**, **teth**, **bonds**, **angles**, **dihedrals** and **inversions** entries described below. DL_POLY_4 will itself construct the global indices for all atoms in the systems.

   Note that DL_POLY_4 determines which shell model to use by scanning shells' weights provided the FIELD file (see Section 2.5). If all shells have zero weight the DL_POLY_4 will choose the relaxed shell model. If no shell has zero weight then DL_POLY_4 will choose the dynamical one. In case when some shells are massless and some are not DL_POLY_4 will terminate execution controllably and provide information about the error and possible possible choices of action in the OUTPUT file (see Section 6.2.6). Shell models' extensions are dealt according to the user specifications in the FIELD files.

   This directive (and associated data records) need not be specified if the molecule contains no core-shell units.

5. **constraints *n***
   where n is the number of constraint bonds in the molecule. Each of the following n records contains:

index 1	integer	first atomic site index
index 2	integer	second atomic site index
bondlength	real	constraint bond length

   This directive (and associated data records) need not be specified if the molecule contains no constraint bonds. See the note on the atomic indices appearing under the **shell** directive above.

6. **pmf *b***
   where b is the potential of mean force bondlength (Å). There follows the definitions of two PMF units:

   (a) **pmf unit *n1***
   where *n1* is the number of sites in the first unit. The subsequent *n1* records provide the site indices and weighting. Each record contains:

index	integer	atomic site index
weight	real	site weighting

(b) **pmf unit *n2***
where *n2* is the number of sites in the second unit. The subsequent *n2* records provide the site indices and weighting. Each record contains:

index	integer	atomic site index
weight	real	site weighting

This directive (and associated data records) need not be specified if no PMF constraints are present. See the note on the atomic indices appearing under the **shell** directive.

Note that if a site weighting is not supplied DL_POLY_4 will assume it is zero. However, DL_POLY_4 detects that all sites in a PMF unit have zero weighting then the PMF unit sites will be assigned the masses of the original atomic sites.

The PMF bondlength applies to the distance between the centres of the two PMF units. The centre, $\vec{R}_i$, of each unit is given by

$$\underline{R}_i = \frac{\sum_{j=1}^{n_i} w_j \; \vec{r}_j}{\sum_{j=1}^{n_j} w_j} \quad , \tag{6.7}$$

where r_j is a site position and w_j the site weighting.

Note that the PMF constraint is intramolecular. To define a constraint between two molecules, the molecules must be described as part of the same DL_POLY_4 "molecule". DL_POLY_4 allows only one type of PMF constraint per system. The value of nummols for this molecule determines the number of PMF constraint in the system.

Note that in DL_POLY_4 PMF constraints are handeled in every available ensemble.

7. **rigid *n***
where *n* is the number of basic rigid units in the molecule. It is followed by at least *n* records, each specifying the sites in a rigid unit:

m	integer	number of sites in rigid unit
site 1	integer	first site atomic index
site 2	integer	second site atomic index
site 3	integer	third site atomic index
..	..	*etc.*
site m	integer	m'th site atomic index

Up to 15 sites can be specified on the first record. Additional records can be used if necessary. Up to 16 sites are specified per record thereafter.

This directive (and associated data records) need not be specified if the molecule contains no rigid units. See the note on the atomic indices appearing under the **shell** directive above.

8. **teth *n***
where *n* is the number of tethered atoms in the molecule. It is followed *n* records specifying the tehered sites in the molecule:

tether key	a4	potential key, see Table 6.7
index 1 (i)	integer	atomic site index
variable 1	real	potential parameter, see Table 6.7
variable 2	real	potential parameter, see Table 6.7

The meaning of these variables is given in Table 6.7.

This directive (and associated data records) need not be specified if the molecule contains no flexible chemical bonds. See the note on the atomic indices appearing under the **shell** directive above.

Table 6.7: Tethering Potentials

key	potential type	Variables (1-3)			functional form
harm	Harmonic	k			$U(r) = \frac{1}{2}\ k\ (r_i - r_i^{t=0})^2$
rhrm	Restraint	k	r_c		$U(r) = \frac{1}{2}\ k\ (r_i - r_i^{t=0})^2 \ :\ \lvert r_i - r_i^{t=0}\rvert \le r_c$ $U(r) = \frac{1}{2}\ k\ r_c^2 + k\ r_c(\lvert r_i - r_i^{t=0}\rvert - r_c)\ :\ \lvert r_i - r_i^{t=0}\rvert > r_c$
quar	Quartic	k	k'	k''	$U(r) = \frac{k}{2}\ (r_i - r_i^{t=0})^2 + \frac{k'}{3}\ (r_i - r_i^{t=0})^3 + \frac{k''}{4}\ (r_i - r_i^{t=0})^4$

9. **bonds *n***
   where n is the number of flexible chemical bonds in the molecule. Each of the subsequent n records contains:

bond key	a4	potential key, see Table 6.8
index 1 (i)	integer	first atomic site index in bond
index 2 (j)	integer	second atomic site index in bond
variable 1	real	potential parameter, see Table 6.8
variable 2	real	potential parameter, see Table 6.8
variable 3	real	potential parameter, see Table 6.8
variable 4	real	potential parameter, see Table 6.8

   The meaning of these variables is given in Table 6.8.

   This directive (and associated data records) need not be specified if the molecule contains no flexible chemical bonds. See the note on the atomic indices appearing under the **shell** directive above.

10. **angles *n***
   where n is the number of valence angle bonds in the molecule. Each of the n records following contains:

angle key	a4	potential key, see Table 6.9
index 1 (i)	integer	first atomic site index
index 2 (j)	integer	second atomic site index (central site)
index 3 (k)	integer	third atomic site index
variable 1	real	potential parameter, see Table 6.9
variable 2	real	potential parameter, see Table 6.9
variable 3	real	potential parameter, see Table 6.9
variable 4	real	potential parameter, see Table 6.9

   The meaning of these variables is given in Table 6.9.

   This directive (and associated data records) need not be specified if the molecule contains no angular terms. See the note on the atomic indices appearing under the **shell** directive above.

11. **dihedrals *n***
   where n is the number of dihedral interactions present in the molecule. Each of the following n records contains:

dihedral key	a4	potential key, see Table 6.10
index 1 (i)	integer	first atomic site index
index 2 (j)	integer	second atomic site index (central site)

Table 6.8: Chemical Bond Potentials

key	potential type	Variables (1-4)				functional form
harm **-hrm**	Harmonic	k	r_0			$U(r) = \frac{1}{2}\ k\ (r_{ij} - r_0)^2$
mors **-mrs**	Morse	E_0	r_0	k		$U(r) = E_0\ [\{1 - \exp(-k\ (r_{ij} - r_0))\}^2 - 1]$
12-6 **-126**	12-6	A	B			$U(r) = \left(\frac{A}{r_{ij}^{12}}\right) - \left(\frac{B}{r_{ij}^{6}}\right)$
lj **-lj**	Lennard-Jones	ϵ	σ			$U(r) = 4\epsilon \left[\left(\frac{\sigma}{r_{ij}}\right)^{12} - \left(\frac{\sigma}{r_{ij}}\right)^{6}\right]$
rhrm **-rhm**	Restraint	k	r_0	r_c		$U(r) = \frac{1}{2}\ k\ (r_{ij} - r_0)^2 \quad : \ \lvert r_{ij} - r_0\rvert \leq r_c$ $U(r) = \frac{1}{2}\ k\ r_c^2 + k\ r_c(\lvert r_{ij} - r_0\rvert - r_c) \ : \ \lvert r_{ij} - r_0\rvert > r_c$
quar **-qur**	Quartic	k	r_0	k'	k''	$U(r) = \frac{k}{2}\ (r_{ij} - r_0)^2 + \frac{k'}{3}\ (r_{ij} - r_0)^3 + \frac{k''}{4}\ (r_{ij} - r_0)^4$
buck **-bck**	Buckingham	A	ρ	C		$U(r) = A\ \exp\left(-\frac{r_{ij}}{\rho}\right) - \frac{C}{r_{ij}^6}$
coul **-cul**	Coulomb	k				$U(r) = k \cdot U^{Electrostatics}(r_{ij})\ \left(= \frac{k}{4\pi\epsilon_0\epsilon} \frac{q_i q_j}{r_{ij}}\right)$
fene **-fne**	Shifted* FENE [33, 34, 35]	k	R_o	Δ		$U(r) = -0.5\ k\ R_o\ ln \left[1 - \left(\frac{r_{ij}-\Delta}{R_o^2}\right)^2\right] \ : \ r_{ij} < R_o + \Delta$ $U(r) = \infty \ : \ r_{ij} \geq R_o + \Delta$
amoe **-amo**	AMOEBA [36] FF bond	k	r_o			$U(r) = k\ \delta^2\ [1 - 2.55\ \delta + (7/12)\ 2.55\ \delta^2]\ ;\ \delta\ = r - r_o$

* Note: Δ defaults to zero if $\lvert\Delta\rvert > 0.5\ R_o$ or if it is not specified in the FIELD file.
Note: Bond potentials with a dash (-) as the first character of the keyword, do not contribute to the *excluded atoms list* (see Section 2). In this case DL_POLY_4 will also calculate the non-bonded pair potentials between the described atoms, unless these are deactivated by another potential specification.

Table 6.9: Valence Angle Potentials

key	potential type	Variables (1-4)				functional form†
harm **-hrm**	Harmonic	k	θ_0			$U(\theta)=\frac{k}{2}\ (\theta-\theta_0)^2$
quar **-qur**	Quartic	k	θ_0	k'	k''	$U(\theta)=\frac{k}{2}\ (\theta-\theta_0)^2+\frac{k'}{3}(\theta-\theta_0)^3+\frac{k''}{4}(\theta-\theta_0)^4$
thrm **-thm**	Truncated harmonic	k	θ_0	ρ		$U(\theta)=\frac{k}{2}\ (\theta-\theta_0)^2\exp[-(r_{ij}^8+r_{ik}^8)/\rho^8]$
shrm **-shm**	Screened harmonic	k	θ_0	ρ_1	ρ_2	$U(\theta)=\frac{k}{2}\ (\theta-\theta_0)^2\exp[-(r_{ij}/\rho_1+r_{ik}/\rho_2)]$
bvs1 **-bv1**	Screened Vessal [37]	k	θ_0	ρ_1	ρ_2	$U(\theta)=\frac{k}{8(\theta-\pi)^2}\left\{\left[(\theta_0-\pi)^2-(\theta-\pi)^2\right]^2\right\}\times$ $\exp[-(r_{ij}/\rho_1+r_{ik}/\rho_2)]$
bvs2 **-bv2**	Truncated Vessal [38]	k	θ_0	a	ρ	$U(\theta)=k\ (\theta-\theta_0)^2\ [\theta^a(\theta+\theta_0-2\pi)^2$ $+\frac{a}{2}\pi^{a-1}(\theta_0-\pi)^3]\exp[-(r_{ij}^8+r_{ik}^8)/\rho^8]$
hcos **-hcs**	Harmonic Cosine	k	θ_0			$U(\theta)=\frac{k}{2}\ (\cos(\theta)-\cos(\theta_0))^2$
cos **-cos**	Cosine	A	δ	m		$U(\theta)=A\ [1+\cos(m\ \theta-\delta)]$
mmsb **-msb**	MM3 stretch-bend [39]	A	θ_0	r_{ij}^o	r_{jk}^o	$U(\theta)=A\ (\theta-\theta_0)\ (r_{ij}-r_{ij}^o)\ (r_{ik}-r_{ik}^o)$
stst **-sts**	Compass [40] stretch-stretch	A	r_{ij}^o	r_{jk}^o		$U(\theta)=A\ (r_{ij}-r_{ij}^o)\ (r_{ik}-r_{ik}^o)$
stbe **-stb**	Compass [40] stretch-bend	A	θ_0	r_{ij}^o		$U(\theta)=A\ (\theta-\theta_0)\ (r_{ij}-r_{ij}^o)$
cmps **-cmp**	Compass [40] all terms	A r_{ij}^o	B r_{jk}^o	C	θ_0	$U(\theta)=A\ (r_{ij}-r_{ij}^o)\ (r_{ik}-r_{ik}^o)+(\theta-\theta_0)\times$ $[B\ (r_{ij}-r_{ij}^o)+C\ (r_{ik}-r_{ik}^o)]$
amoe **-amo**	AMOEBA [36] FF angle	k	θ_0			$U(\theta)=k\ \Delta^2[1-1.4\cdot10^{-2}\Delta+5.6\cdot10^{-5}\Delta^2$ $-7.0\cdot10^{-7}\Delta^3+2.2\cdot10^{-8}\Delta^4)]\ ;\ \Delta=\theta-\theta_0$
kky **-kky**	KKY [36]	f_k	θ_0	g_r	r_o	$U(\theta)=f_k\ \sin[2(\theta-\theta_0)]\cdot\sqrt{K_{ij}\cdot K_{ik}}\ ;$ $K_{ij}=1/\exp[g_r(r_{ij}-r_o)]$

†θ is the *i*-*j*-*k* angle.

Note: valence angle potentials with a dash (-) as the first character of the keyword, do not contribute to the *excluded atoms list* (see Section 2). In this case DL_POLY_4 will calculate the non-bonded pair potentials between the described atoms.

index 3 (k)	integer	third atomic site index
index 4 (l)	integer	fourth atomic site index
variable 1	real	first potential parameter, see Table 6.10
variable 2	real	second potential parameter, see Table 6.10
variable 3	real	third potential parameter, see Table 6.10
variable 4	real	1-4 electrostatic interaction scale factor
variable 5	real	1-4 van der Waals interaction scale factor
variable 6	real	fourth potential parameter, see Table 6.10
variable 7	real	fifth potential parameter, see Table 6.10

The meaning of the variables 1-3,6-7 is given in Table 6.10. The variables 4 and 5 specify the scaling factor for the 1-4 electrostatic and van der Waals non-bonded interactions respectively.

This directive (and associated data records) need not be specified if the molecule contains no dihedral angle terms. See the note on the atomic indices appearing under the **shell** directive above.

Table 6.10: Dihedral Angle Potentials

key	potential type	Variables (1-3,6-7)			functional form‡
cos	Cosine	A	δ	m	$U(\phi) = A\ [1 + \cos(m\phi - \delta)]$
harm	Harmonic	k	ϕ_0		$U(\phi) = \frac{k}{2}\ (\phi - \phi_0)^2$
hcos	Harmonic cosine	k	ϕ_0		$U(\phi) = \frac{k}{2}\ (\cos(\phi) - \cos(\phi_0))^2$
cos3	Triple cosine	A_1	A_2	A_3	$U(\phi) = \frac{1}{2}\ \{A_1\ (1 + \cos(\phi)) + A_2\ (1 - \cos(2\phi)) + A_3\ (1 + \cos(3\phi))\}$
ryck	Ryckaert-Bellemans [43]	A			$U(\phi) = A\ \{a + b\ \cos(\phi) + c\ \cos^2(\phi) + d\ \cos^3(\phi) + e\ \cos^4(\phi) + f\ \cos^5(\phi)\}$
rbf	Fluorinated Ryckaert-Bellemans [44]	A			$U(\phi) = A\ \{a + b\ \cos(\phi) + c\ \cos^2(\phi) + d\ \cos^3(\phi) + e\ \cos^4(\phi) + f\ \cos^5(\phi)) + g\ \exp(-h(\phi - \pi)^2))\}$
opls	OPLS torsion	A_0 A_3	A_1 ϕ_0	A_2	$U(\phi) = A_0 + \frac{1}{2}\ \{A_1\ (1 + \cos(\phi - \phi_0)) + A_2\ (1 - \cos(2(\phi - \phi_0))) + A_3\ (1 + \cos(3(\phi - \phi_0)))\}$

‡ϕ is the i-j-k-l dihedral angle.

12. **inversions *n***
where n is the number of inversion interactions present in the molecule. Each of the following n records contains:

inversion key	a4	potential key, see Table 6.11
index 1 (i)	integer	first atomic site index (central site)
index 2 (j)	integer	second atomic site index
index 3 (k)	integer	third atomic site index
index 4 (l)	integer	fourth atomic site index
variable 1	real	potential parameter, see Table 6.11

variable 2	real	potential parameter, see Table 6.11
variable 3	real	potential parameter, see Table 6.11

The meaning of the variables 1-2 is given in Table 6.11.

This directive (and associated data records) need not be specified if the molecule contains no inversion angle terms. See the note on the atomic indices appearing under the **shell** directive above.

Table 6.11: Inversion Angle Potentials

key	potential type	Variables (1-3)			functional form‡
harm	Harmonic	k	ϕ_0		$U(\phi) = \frac{k}{2}\ (\phi - \phi_0)^2$
hcos	Harmonic cosine	k	ϕ_0		$U(\phi) = \frac{k}{2}\ (\cos(\phi) - \cos(\phi_0))^2$
plan	Planar	A			$U(\phi) = A\ [1 - \cos(\phi)]$
xpln	Extended planar	k	m	ϕ_0	$U(\phi) = \frac{k}{2}\ [1 - \cos(m\ \phi - \phi_0)]$
calc	Calcite	A	B		$U(u) = Au^2 + Bu^4$

‡ϕ is the i-j-k-l inversion angle.

Note that the calcite potential is not dependent on an angle ϕ, but on a *displacement* u. See Section 2.2.9 for details.

13. **finish**
This directive is entered to signal to DL_POLY_4 that the entry of the details of a molecule has been completed.

The entries for a second molecule may now be entered, beginning with the *name-of-molecule* record and ending with the **finish** directive.

The cycle is repeated until all the types of molecules indicated by the **molecules** directive have been entered.

The user is recommended to look at the example FIELD files in the *data* directory to see how typical FIELD files are constructed.

Non-bonded Interactions

Non-bonded interactions are identified by atom types as opposed to specific atomic indices. The following different types of non-bonded potentials are available in DL_POLY_4; **vdw** - van der Waals pair, **metal** - metal, **tersoff** - Tersoff, **tbp** - three-body and **fbp** - four-body. Each of these types is specified by a specific **keyword** as described bellow.

When DL_POLY_4 is cross-compiled with an openKIM functionality (see Section 2.9) it is possible to specify a complete model of inter-molecular interactions, **kim**, by calling the openKIM model name provided it available in your local KIM library.

1. **kim *model_name***
where *model_name* is the openKIM model identifier.

Note that although a KIM model fully describes a model system it is still possible to specify further, complementing intra- and inter-molecular interactions in the FIELD! This is a viable option only when the model system is extended beyond what the specific KIM model is intended to describes.

2. **vdw *n***
   where n is the number of pair potentials to be entered. It is followed by n records, each specifying a particular pair potential in the following manner:

atmnam 1	a8	first atom type
atmnam 2	a8	second atom type
key	a4	potential key, see Table 6.12
variable 1	real	potential parameter, see Table 6.12
variable 2	real	potential parameter, see Table 6.12
variable 3	real	potential parameter, see Table 6.12
variable 4	real	potential parameter, see Table 6.12
variable 5	real	potential parameter, see Table 6.12

   The variables pertaining to each potential are described in Table 6.12.

   Note that any pair potential not specified in the FIELD file, will be assumed to be zero.

3. **metal *n***
   where n is the number of metal potentials to be entered. It is followed by n records, each specifying a particular metal potential in the following manner:

atmnam 1	a8	first atom type
atmnam 2	a8	second atom type
key	a4	potential key, see Table 6.13
variable 1	real	potential parameter, see Table 6.13
variable 2	real	potential parameter, see Table 6.13
variable 3	real	potential parameter, see Table 6.13
variable 4	real	potential parameter, see Table 6.13
variable 5	real	potential parameter, see Table 6.13
variable 6	real	potential parameter, see Table 6.13
variable 7	real	potential parameter, see Table 6.13
variable 8	real	potential parameter, see Table 6.13
variable 9	real	potential parameter, see Table 6.13

   The variables pertaining to each potential are described in Table 6.13.

4. **rdf *n***
   where n is the number of RDF pairs to be entered. It is followed by n records, each specifying a particular RDF pair in the following manner:

atmnam 1	a8	first atom type
atmnam 2	a8	second atom type

   By default in DL_POLY_Classic and DL_POLY_4 every **vdw** and **met** potential specifies an RDF pair. If the control option **rdf *f*** is specified in the CONTROL file then all pairs defined in **vdw** and/or **met** potentials sections will also have their RDF calculated. The user has two choices to enable the calculation of RDFs in systems with force fields that do not have **vdw** and/or **met** potentials: (i) to define fictitious potentials with zero contributions or (ii) to use **rdf *n*** option - which not only provides a neater way for specification of RDF pairs but also better memory efficiency since DL_POLY_4 will not allocate (additional) potential arrays for fictitious interactions that will not be used. (This option is not available in DL_POLY_Classic.)

Table 6.12: Pair Potentials

key	potential type	Variables (1-5)					functional form
tab	Tabulation						tabulated potential
12-6	12-6	A	B				$U(r) = \left(\frac{A}{r^{12}}\right) - \left(\frac{B}{r^{6}}\right)$
lj	Lennard-Jones	ϵ	σ				$U(r) = 4\epsilon\left[\left(\frac{\sigma}{r}\right)^{12} - \left(\frac{\sigma}{r}\right)^{6}\right]$
nm	n-m [47, 48]	E_o	n	m	r_0		$U(r) = \frac{E_o}{(n-m)}\left[m\left(\frac{r_o}{r}\right)^{n} - n\left(\frac{r_o}{r}\right)^{m}\right]$
buck	Buckingham	A	ρ	C			$U(r) = A\ \exp\left(-\frac{r}{\rho}\right) - \frac{C}{r^6}$
bhm	Born-Huggins -Meyer	A	B	σ	C	D	$U(r) = A\ \exp[B(\sigma - r)] - \frac{C}{r^6} - \frac{D}{r^8}$
hbnd	12-10 H-bond	A	B				$U(r) = \left(\frac{A}{r^{12}}\right) - \left(\frac{B}{r^{10}}\right)$
snm	Shifted force[†] n-m [47, 48]	E_o	n	m	r_0	r_c[‡]	$U(r) = \frac{\alpha E_o}{(n-m)} \times \left[m\beta^n\left\{\left(\frac{r_o}{r}\right)^n - \left(\frac{1}{\gamma}\right)^n\right\} - n\beta^m\left\{\left(\frac{r_o}{r}\right)^m - \left(\frac{1}{\gamma}\right)^m\right\}\right] + \frac{nm\alpha E_o}{(n-m)}\left(\frac{r-\gamma r_o}{\gamma r_o}\right)\left\{\left(\frac{\beta}{\gamma}\right)^n - \left(\frac{\beta}{\gamma}\right)^m\right\}$
mors	Morse	E_0	r_0	k			$U(r) = E_0[\{1 - \exp(-k(r - r_0))\}^2 - 1]$
wca	Shifted* [49] Weeks-Chandler-Anderson	ϵ	σ[‡]	Δ			$U(r) = 4\epsilon\left[\left(\frac{\sigma}{r-\Delta}\right)^{12} - \left(\frac{\sigma}{r-\Delta}\right)^{6}\right] + \epsilon : r_{ij} < 2^{\frac{1}{6}}\ \sigma + \Delta$ $U(r) = 0 \quad : r_{ij} \geq 2^{\frac{1}{6}}\ \sigma + \Delta$
dpd	Standard DPD [?] (Groot-Warren)	A	r_c[‡]				$U(r) = \frac{A}{2}\ r_c\ \left(1 - \frac{r}{r_c}\right)^2 \ :\ r < r_c$ $U(r) = 0 \ :\ r \geq r_c$
amoe	AMOEBA [36] FF 14-7 pair	ϵ	r_o				$U(r) = \epsilon\left(\frac{1.07}{(r_{ij}/r_o)+0.07}\right)^7\left(\frac{1.12}{(r_{ij}/r_o)^7+0.12} - 2\right)$

[†] Note: in this formula the terms α, β and γ are compound expressions involving the variables E_o, n, m, r_0 and r_c. See Section 2.3.1 for further details.
[‡] Note: All local potential cutoffs, r_c, default to the general van der Waals cutoff, `rvdw`, or the general domain decomposition cutoff, `rcut`, if unspecified or set to zero in the FIELD file! Similarly, if the specified value of `rvdw` (and/or `rcut`) in CONTROL is found shorter than any of r_c (including the WCA equivalent $2^{\frac{1}{6}}\ \sigma + \Delta$) values specified in FIELD then `rvdw` (and/or `rcut`) will be reset by DL_POLY_4 to the largest of all values!
* Note: Δ defaults to zero if $|\Delta| > 0.5\ \sigma$ or it is not specified in the FIELD file.

Table 6.13: Metal Potential

key	potential type	Variables (1-5,6-9)					functional form
eam	EAM						tabulated potential
eeam	EEAM						tabulated potential
2bea	2BEAM						tabulated potential
2bee	2BEEAM						tabulated potential
fnsc	Finnis-Sinclair	c_0	c_1	c_2	c	A	$U_i(r) = \frac{1}{2} \sum_{j \neq i} (r_{ij} - c)^2 (c_0 + c_1 r_{ij} + c_2 r_{ij}^2) - A\sqrt{\rho_i}$;
		d	β				$\rho_i = \sum_{j \neq i} \left[(r_{ij} - d)^2 + \beta \frac{(r_{ij}-d)^3}{d} \right]$
exfs	Extended	c_0	c_1	c_2	c_3	c_4	$U_i(r) = \frac{1}{2} \sum_{j \neq i} (r_{ij} - c)^2 (c_0 + c_1 r_{ij} + c_2 r_{ij}^2 + c_3 r_{ij}^3 + c_4 r_{ij}^4)$
	Finnis-Sinclair	c	A	d	B		$-A\sqrt{\rho_i}$; $\rho_i = \sum_{j \neq i} \left[(r_{ij} - d)^2 + B^2 (r_{ij} - d)^4 \right]$
stch	Sutton-Chen	ϵ	a	n	m	c	$U_i(r) = \epsilon \left[\frac{1}{2} \sum_{j \neq i} \left(\frac{a}{r_{ij}} \right)^n - c\sqrt{\rho_i} \right]$; $\rho_i = \sum_{j \neq i} \left(\frac{a}{r_{ij}} \right)^m$
gupt	Gupta	A	r_0	p	B	q_{ij}	$U_i(r) = \sum_{j \neq i} A \exp\left(-p \frac{r_{ij} - r_0}{r_0} \right) - B\sqrt{\rho_i}$;
							$\rho_i = \sum_{j \neq i} \exp\left(-2 q_{ij} \frac{r_{ij} - r_0}{r_0} \right)$
mbpc	MBPC†	ϵ	a	m	α	r_o	$U_i(r) = -\epsilon\sqrt{\rho_i}$; $\rho_i = \sum_{j \neq i} \left(\frac{a}{r_{ij}^m} \right) \frac{1}{2} \left[1 + \text{erf}\left(\alpha (r_{ij} - r_\text{o}) \right) \right]$

†**Note** that the parameters α and r_o must be the same for all defined potentials of this type. DL_POLY_4 will set $\alpha = \text{Max}(0, \alpha_{pq})$ and $r_\text{o} = \text{Max}(0, r_{\text{o}\_pq})$ for all defined interactions of this type between species p and q. If after this any is left undefined, i.e. zero, the undefined entities will be set to their defaults: $\alpha = 20$ and $r_\text{o} = \text{Min}(1.5, 0.2\ r_\text{cut})$.

Note that **rdf** and **vdw/met** are not complementary - i.e. if the former is used in FIELD none of the pairs defined by the latter will be considered for RDF calculations.

The selected RDFs are calculated in the RDF_COLLECT, RDF_EXCL_COLLECT, RDF_FRZN_COLLECT and RDF_COMPUTE by collecting distance information from all two-body pairs as encountered in the Verlet neighbour list created in the LINK_CELL_PAIRS routine within the TWO_BODY_FORCES routine. In the construction of the Verlet neighbour list, pairs of particles (part of the exclusion list) are excluded. The exclusion list contains particles that are part of:

- core-shell units
- bond constraints
- chemical bonds, that are NOT distance restraints
- valence angles, that are NOT distance restraints
- dihedrals
- inversions

- frozen particles

RDF pairs containing type(s) of particles that fall in this list will be polluted. However, there are many ways to overcome such effects.

5. **tersoff *n***
where n is the number of specified Tersoff potentials. There are two types of Tersoff potential forms that cannot be mixed (used simultaneously). They are shorthanded as **ters** and **kihs** atomkeys.

- **ters** atomkey expects $2n$ records specifying n particular Tersoff single atom type parameter sets and $n(n+1)/2$ records specifying cross atom type parameter sets in the following manner:

potential 1 : record 1		
atmnam	a8	atom type
key	a4	potential key, see Table 6.14
variable 1	real	potential parameter, see Table 6.14
variable 2	real	potential parameter, see Table 6.14
variable 3	real	potential parameter, see Table 6.14
variable 4	real	potential parameter, see Table 6.14
variable 5	real	cutoff range for this potential (Å)
potential 1 : record 2		
variable 6	real	potential parameter, see Table 6.14
variable 7	real	potential parameter, see Table 6.14
variable 8	real	potential parameter, see Table 6.14
variable 9	real	potential parameter, see Table 6.14
variable 10	real	potential parameter, see Table 6.14
variable 11	real	potential parameter, see Table 6.14
...	...	...
...	...	...
potential n **: record** $2n-1$		
...	...	...
potential n **: record** $2n$		
...	...	...
cross term 1 **: record** $2n+1$		
atmnam 1	a8	first atom type
atmnam 2	a8	second atom type
variable a	real	potential parameter, see Table 6.14
variable b	real	potential parameter, see Table 6.14
variable c	real	potential parameter, see Table 6.14
...	...	...
...	...	...
cross term $n(n+1)/2$ **: record** $2n+n(n+1)/2$		
...	...	...

- **kihs** atomkey expects $3n$ records specifying n particular Tersoff single atom type parameter sets in the following manner:

potential 1 : record 1		
atmnam	a8	atom type
key	a4	potential key, see Table 6.14
variable 1	real	potential parameter, see Table 6.14
variable 2	real	potential parameter, see Table 6.14
variable 3	real	potential parameter, see Table 6.14
variable 4	real	potential parameter, see Table 6.14
variable 5	real	cutoff range for this potential (Å)

potential 1 : record 2

variable 6	real	potential parameter, see Table 6.14
variable 7	real	potential parameter, see Table 6.14
variable 8	real	potential parameter, see Table 6.14
variable 9	real	potential parameter, see Table 6.14
variable 10	real	potential parameter, see Table 6.14
variable 11	real	potential parameter, see Table 6.14

potential 1 : record 3

variable 12	real	potential parameter, see Table 6.14
variable 13	real	potential parameter, see Table 6.14
variable 14	real	potential parameter, see Table 6.14
variable 15	real	potential parameter, see Table 6.14
variable 16	real	potential parameter, see Table 6.14
...	...	...
...	...	...

potential n **: record** $2n-1$

...	...	...

potential n **: record** $2n$

...	...	...

The variables pertaining to each potential are described in Table 6.14.

Note that the fifth variable is the range at which the particular tersoff potential is truncated. The distance is in Å.

Table 6.14: Tersoff Potential

key	potential type	Variables (1-5,6-11,a-c/12-16)						functional form
ters	Tersoff	A	a	B	b	R		Potential forms
	(single)	S	β	η	c	d	h	
	(cross)	χ	ω	δ				as shown in
kihs	KIHS	A	a	B	b	R		
		S	η	δ	c_1	c_2	c_3	Section 2.3.3
		c_4	c_5	h	α	β		

6. **tbp** $\boldsymbol{n}$
where n is the number of three-body potentials to be entered. It is followed by n records, each specifying a particular three-body potential in the following manner:

atmnam 1 (i)	a8	first atom type
atmnam 2 (j)	a8	second (central) atom type
atmnam 3 (k)	a8	third atom type
key	a4	potential key, see Table 6.15
variable 1	real	potential parameter, see Table 6.15
variable 2	real	potential parameter, see Table 6.15
variable 3	real	potential parameter, see Table 6.15
variable 4	real	potential parameter, see Table 6.15
variable 5	real	cutoff range for this potential (Å)

The variables pertaining to each potential are described in Table 6.15.

Note that the fifth variable is the range at which the three body potential is truncated. The distance is in Å, measured from the central atom.

Table 6.15: Three-body Potentials

key	potential type	Variables (1-4)				functional form†
harm	Harmonic	k	θ_0			$U(\theta) = \frac{k}{2}\ (\theta - \theta_0)^2$
thrm	Truncated harmonic	k	θ_0	ρ		$U(\theta) = \frac{k}{2}\ (\theta - \theta_0)^2\ \exp[-(r_{ij}^8 + r_{ik}^8)/\rho^8]$
shrm	Screened harmonic	k	θ_0	ρ_1	ρ_2	$U(\theta) = \frac{k}{2}\ (\theta - \theta_0)^2 \exp[-(r_{ij}/\rho_1 + r_{ik}/\rho_2)]$
bvs1	Screened Vessal [37]	k	θ_0	ρ_1	ρ_2	$U(\theta) = \frac{k}{8(\theta-\theta_0)^2} \left\{ \left[(\theta_0 - \pi)^2 - (\theta - \pi)^2 \right]^2 \right\} \times$ $\exp[-(r_{ij}/\rho_1 + r_{ik}/\rho_2)]$
bvs2	Truncated Vessal [38]	k	θ_0	a	ρ	$U(\theta) = k\ (\theta - \theta_0)^2\ [\theta^a (\theta - \theta_0)^2 (\theta + \theta_0 - 2\pi)^2$ $+ \frac{a}{2} \pi^{a-1} (\theta_0 - \pi)^3]\ \exp[-(r_{ij}^8 + r_{ik}^8)/\rho^8]$
hbnd	H-bond [19]	D_{hb}	R_{hb}			$U(\theta) = D_{hb}\ \cos^4(\theta) \times$ $[5(R_{hb}/r_{jk})^{12} - 6(R_{hb}/r_{jk})^{10}]$

†θ is the i-j-k angle.

7. **fbp *n***
   where n is the number of four-body potentials to be entered. It is followed by n records, each specifying a particular four-body potential in the following manner:

atmnam 1 (i)	a8	first (central) atom type
atmnam 2 (j)	a8	second atom type
atmnam 3 (k)	a8	third atom type
atmnam 4 (l)	a8	fourth atom type
key	a4	potential key, see Table 6.16
variable 1	real	potential parameter, see Table 6.16
variable 2	real	potential parameter, see Table 6.16
variable 3	real	cutoff range for this potential (Å)

The variables pertaining to each potential are described in Table 6.16.

Note that the third variable is the range at which the four-body potential is truncated. The distance is in Å, measured from the central atom.

6.1.3.3 External Field

The presence of an external field is flagged by the directive:

extern

The following line in the FIELD file must contain another directive indicating what type of field is to be applied, followed by the field parameters in the following manner:

field key	a4	external field key, see Table 6.17

Table 6.16: Four-body Potentials

key	potential type	Variables (1-2)		functional form‡
harm	Harmonic	k	ϕ_0	$U(\phi) = \frac{k}{2}\ (\phi - \phi_0)^2$
hcos	Harmonic cosine	k	ϕ_0	$U(\phi) = \frac{k}{2}\ (\cos(\phi) - \cos(\phi_0))^2$
plan	Planar	A		$U(\phi) = A\ [1 - \cos(\phi)]$

‡ϕ is the i-j-k-l four-body angle.

variable 1	real	potential parameter, see Table 6.17
variable 2	real	potential parameter, see Table 6.17
variable 3	real	potential parameter, see Table 6.17
variable 4	real	potential parameter, see Table 6.17
variable 5	real	potential parameter, see Table 6.17

The variables pertaining to each field potential are described in Table 6.17.

Note: only one type of field can be applied at a time.

Note that external force parameters are read in terms of the specified energy units and the general DL_POLY units so that the two sides of the equation defining the field are balanced. For example, the magnetic field units, $\underline{H} = (H_1, H_2, H_3)$, in the DL_POLY FIELD scope will follow from the interaction definition as seen in Table 6.17:

$$\begin{aligned}
\underline{F} &= q\ (\underline{v} \times \underline{H}) \qquad \texttt{therefore,} \\
[H] &= \frac{[F]}{[q]\ \ [v]} = \frac{[m]\ \ [a]}{[q]\ \ [v]} \\
[H] &= \frac{\texttt{Dalton Å/ps}^2}{\texttt{proton Å/ps}} = \frac{\texttt{Dalton}}{\texttt{proton ps}} \qquad (6.8) \\
[H] &= 1.037837512 \times 10^4\ \texttt{Tesla} \qquad (6.9) \\
H(DL\_POLY) &= H(MKS)\ 1.037837512 \times 10^4\ \ .
\end{aligned}$$

Thus to apply a magnetic field of 1 Tesla along the y axis, one could specify in FIELD the following:

```
UNITS internal
...
external
magnetic 0  1/1.037837512e04   0
...
```

when working in DL_POLY internal units. If we worked in *unit* units then

$$\begin{aligned}
H &\propto Energy \\
Energy(DL\_POLY) &= Energy(unit)\ k_{unit \to DL\_POLY} \\
H(DL\_POLY) &= H(MKS)\ 1.037837512 \times 10^4 \qquad (6.10) \\
H(unit) &= H(MKS)\ \frac{1.037837512 \times 10^4}{k_{unit \to DL\_POLY}}\ \ ,
\end{aligned}$$

with the following conversion factors values:

$$k_{eV \to DL\_POLY} = 9648.530821$$

Table 6.17: External Fields

key	potential type	Variables (1-5)					functional form
elec	Electric Field	E_x	E_y	E_z			$\underline{F} = q\ \underline{E}$
oshr	Oscillating Shear	A	n				$\underline{F}_x = A\ \cos(2n\pi \cdot z/L_z)$
shrx	Continuous Shear	A	z_0				$\underline{v}_x = \frac{A}{2}\frac{\lvert z\rvert}{z}\ :\ \lvert z\rvert\ >\ z_0$
grav	Gravitational Field	G_x	G_y	G_z			$\underline{F} = m\ \underline{G}$
magn	Magnetic Field	H_x	H_y	H_z			$\underline{F} = q\ (\underline{v} \times \underline{H})$
sphr	Containing Sphere	A	R_0	n	$R_{\rm cut}$		$\underline{F} = A\ (R_0 - r)^{-n}\ :\ r > R_{\rm cut}$
zbnd	Repulsive Wall	A	z_0	p			$\underline{F} = A\ (z_0 - z)\ :\ p \cdot z > p \cdot z_0$
xpis	X-Piston	i_{ind}^{glob}	j_{ind}^{glob}	P_{k-atm}^{in}			$\underline{F}_x = \frac{P \cdot \underline{\rm Area}(\perp {\rm X}\text{-}dir)}{\left[\sum_{k=i}^{j} m_k\right]/m_k}\ :\forall\ k = i,..,j$
zres	Molecule in HR Zone	i_{ind}^{glob}	j_{ind}^{glob}	k	z_{mn}	z_{mx}	$\underline{F}_z = \begin{cases} A(z_{cm} - z_{mx}) : z_{cm} > z_{mx} \\ A(z_{mn} - z_{cm}) : z_{cm} < z_{mn} \end{cases}$
zrs-	HR Zone (pull out)	i_{ind}^{glob}	j_{ind}^{glob}	k	z_{mn}	z_{mx}	$\underline{F}_z = \begin{cases} A(z - z_{mx}) : z \geq \frac{z_{mx}+z_{mn}}{2} \\ A(z_{mn} - z) : z < \frac{z_{mx}+z_{mn}}{2} \end{cases}$
zrs+	HR Zone (pull in)	i_{ind}^{glob}	j_{ind}^{glob}	k	z_{mn}	z_{mx}	$\underline{F}_z = \begin{cases} A(z - z_{mx}) : z > z_{mx} \\ A(z_{mn} - z) : z < z_{mn} \end{cases}$
osel	Osc. Electric Field	E_x	E_y	E_z	$\omega_{ps^{-1}}^{in}$		$\underline{F} = q\ \underline{E}\ \sin(2\pi\omega t)$

$$\begin{aligned} k_{kcal/mol \rightarrow DL\_POLY} &= 418.4 \\ k_{kJ/mol \rightarrow DL\_POLY} &= 100.0 \\ k_{K/Boltz \rightarrow DL\_POLY} &= 0.831451115 \\ k_{DL\_POLY \rightarrow DL\_POLY} &= 1.0\ . \end{aligned} \tag{6.11}$$

Obviously, for eV units

$$\begin{aligned} H(unit) &= H(MKS)\ \frac{1.037837512 \times 10^4}{k_{unit \rightarrow DL\_POLY}} \\ k_{eV \rightarrow DL\_POLY} &= 9648.530821 \\ H(eV) &= H(MKS)\ 1.07564305\ , \end{aligned} \tag{6.12}$$

the FIELD file should be amended to read:

```
UNITS eV
...
external
magnetic 0  1/1.07564305   0
...
```

6.1.3.4 Closing the FIELD File

The FIELD file must be closed with the directive:

close
which signals the end of the force field data. Without this directive DL_POLY_4 will abort.

6.1.4 The REFERENCE File

The REFERENCE has the same format and structure as CONFIG (see Section 6.1.2) file with the exception that `imcon` **MUST BE** $\neq 0$. REFERENCE may contain more or less particles than CONFIG does and may have particles with identities that are not defined in FIELD (see Section 6.1.3). The positions of these particles are used to define the crystalline lattice sites to whitch the particles in CONFIG compare during simulation when the defect detection option, **defe**cts, is used. REFERENCE is read by the subroutine DEFECTS_REFERENCE_READ.

6.1.5 The REVOLD File

This file contains statistics arrays from a previous job. It is not required if the current job is not a continuation of a previous run (i.e. if the **restart** directive is not present in the CONTROL file - see above). The file is unformatted and therefore not human readable. DL_POLY_4 normally produces the file REVIVE (see Section 6.2.8) at the end of a job which contains the statistics data. REVIVE should be copied to REVOLD before a continuation run commences. This may be done by the *copy* macro supplied in the *execute* sub-directory of DL_POLY_4.

6.1.5.1 Format

The REVOLD file is unformatted. All variables appearing are written in native working precision (see Section 5.3.5) real representation. Nominally, integer quantities (e.g. the timestep number `nstep`) are represented by the nearest real number. The contents are as follows (the dimensions of array variables are given in brackets, in terms of parameters from the SETUP_MODULE file - see Section 7.2.8).

record 1:	
`nstep`	timestep of final configuration
`numacc`	number of configurations used in averages
`numrdf`	number of configurations used in RDF averages
`numzdn`	number of configurations used in Z-density averages
`time`	elapsed simulation time
`tmst`	elapsed simulation before averages were switched on
`chit`	thermostat related quantity (first)
`chip`	barostat related quantity
`cint`	thermostat related quantity (second)
record 2:	
`eta`	scaling factors for simulation cell matrix elements (9)
record 3:	
`stpval`	instantaneous values of thermodynamic variables (`mxnstk`)
record 4:	
`sumval`	average values of thermodynamic variables (`mxnstk`)
record 5:	
`ssqval`	fluctuation (squared) of thermodynamic variables (`mxnstk`)
record 6:	

`zumval`	running totals of thermodynamic variables (`mxnstk`)
record 7:	
`ravval`	rolling averages of thermodynamic variables (`mxnstk`)
record 8:	
`stkval`	stacked values of thermodynamic variables (`mxstak`×`mxnstk`)
record 9:	
`strcon`	constraint bond stress (9)
record 10:	
`strpmf`	PMF constraint stress (9)
record 11:	
`stress`	atomic stress (9)
record 12: (Optional)	
`rdf`	RDF array (`mxgrdf`×`mxrdf`)
record 13: (Optional)	
`zdens`	Z-density array (`mxgrdf`×`mxatyp`)
record 14: (Optional)	
`vaf`	VAF arrays (sizes dependent on sampling frequency and VAF bin size)

6.1.5.2 Further Comments

Note that different versions of DL_POLY_4 may have a different order of the above parameters or include more or less such. Therefore different versions of DL_POLY_4 may render any existing REVOLD file unreadable by the code.

6.1.6 The TABLE File

The TABLE file provides an alternative way of reading in the short range potentials - in tabular form. This is particularly useful if an analytical form of the potential does not exist or is too complicated to specify in the VDW_GENERATE subroutine. The table file is read by the subroutine VDW_TABLE_READ (see Chapter 7).

The option of using tabulated potentials is specified in the FIELD file (see above). The specific potentials that are to be tabulated are indicated by the use of the **tab** keyword on the record defining the short range potential (see Table 6.12).

6.1.6.1 The TABLE File Format

The file is free-formatted but blank and commented lines are not allowed.

6.1.6.2 Definitions of Variables

record 1		
`header`	a100	file header
record 2		
`delpot`	real	mesh resolution in Å ($\texttt{delpot} = \frac{\texttt{cutpot}}{\texttt{ngrid}-4}$)
`cutpot`	real	cutoff used to define tables in Å
`ngrid`	integer	number of grid points in tables

The subsequent records define each tabulated potential in turn, in the order indicated by the specification in the FIELD file. Each potential is defined by a header record and a set of data records with the potential and force tables.

header record:

`atom 1`	a8	first atom type
`atom 2`	a8	second atom type

potential data records: (*number of data records* = `Int`((`ngrid`+3)/4))

`data 1`	real	data item 1
`data 2`	real	data item 2
`data 3`	real	data item 3
`data 4`	real	data item 4

force data records: (*number of data records* = `Int`((`ngrid`+3)/4))

`data 1`	real	data item 1
`data 2`	real	data item 2
`data 3`	real	data item 3
`data 4`	real	data item 4

6.1.6.3 Further Comments

It should be noted that the number of grid points in the TABLE file should not be less than the number of grid points DL_POLY_4 is expecting. (This number is given by the parameter `mxgvdw` calculated in the SETUP_MODULE file - see Section 5.2.1.3 and 7.2.8.) DL_POLY_4 will re-interpolate the tables if `delpot` = $\frac{\texttt{cutpot}}{\texttt{ngrid}-4} < \texttt{dlrvdw} = \frac{\texttt{rvdw}}{\texttt{mxgvdw}-4}$ (usually when `ngrid` > `mxgvdw`), but will abort if `delpot` > `dlrvdw`.

The potential and force tables are used to fill the internal arrays `vvdw` and `gvdw` respectively (see Section 2.3.1). The contents of force arrays are derived from the potential via the formula:

$$G(r) = -r\frac{\partial}{\partial r}U(r) \quad . \tag{6.13}$$

Note, this is *not* the same as the true force.

During simulation, interactions beyond distance `cutpot` are discarded.

6.1.7 The TABEAM File

The TABEAM file contains the tabulated potential functions (no explicit analytic form) describing the EAM or EEAM metal interactions in the MD system. This file is read by the subroutine METAL_TABLE_READ (see Chapter 7).

6.1.7.1 The TABEAM File Format

The file is free-formatted but blank and commented lines are not allowed.

6.1.7.2 Definitions of Variables

record 1

`header`	a100	file header

record 2

`numpot`	integer	number of potential functions in file

For an n component alloy, `numpot` is

- $n(n + 5)/2$ for the EAM potential *or*
- $3n(n + 1)/2$ for the EEAM potential *or*

- $n(n+4)$ for the 2BEAM potential *or*
- $5n(n+1)/2$ for the 2BEEAM potential.

The subsequent records for an n component alloy define $n(n+1)/2$ *cross* pair potential functions - **pair**s keyword and

- EAM: n embedding functions (one for each atom type) - **embe**dding keyword, n electron density functions (one for each atom type) - **dens**ity keyword;
- EEAM: n embedding functions (one for each atom type) and n^2 for the EEAM potential (one for each non-commuting pair of atoms types);
- 2BEAM: n s-band embedding functions (one for each atom type) - **semb**edding keyword, $n(n+1)/2$ s-band density functions (one for each *cross*-pair of atoms types) - **sden**sity keyword, n d-band embedding functions (one for each atom type) - **demb**edding or **embe**edding keyword, and n d-band density functions (one for each atom types) - **dden**sity or **dens**ity keyword;
- 2BEEAM: n s-band embedding functions (one for each atom type) - **semb**edding keyword, n^2 s-band density functions (one for each non-commuting pair of atoms types) - **sden**sity keyword, n d-band embedding functions (one for each atom type) - **demb**edding or **embe**edding keyword, and n^2 d-band density functions (one for each non-commuting pair of atoms types) - **dden**sity or **dens**ity keyword.

The functions may appear in any random order in TABEAM as their identification is based on their unique keyword, defined first in the function's header record. The header record is followed by predefined number of data records **as a maximum of four data per record are read in** - allowing for incompletion of the very last record.

header record:

keyword	a4	type of EAM function: **pair**, **embe**d or **dens**ity, with 2B extension alternatives for the s-band - [**semb**ed and **sden**sity] and d-band - **demb**ed = **embe**d and **dden**sity = **dens**ity
`atom 1`	a8	first atom type
`atom 2`	a8	second atom type - only specified for **pair** potential functions and for the (*i*) **dens**ity functions in the EEAM potential case *or* (*ii*) **sden**sity functions in the 2BEAM potential case *or* (*iii*) **sden** and **dden** functions in the 2BEEAM potential case
`ngrid`	integer	number of function data points to read in
`limit 1`	real	lower interpolation limit in Å for **dens/sden/dden** and **pair** or in density units for **embe/semb/demb**
`limit 2`	real	upper interpolation limit in Å for **dens/sden/dden** and **pair** or in density units for **embe/semb/demb**

function data records: (*number of data records* = `Int((ngrid+3)/4)`)

`data 1`	real	data item 1
`data 2`	real	data item 2
`data 3`	real	data item 3
`data 4`	real	data item 4

6.1.7.3 Further Comments

The tabled data are used to fill the internal arrays `vmet`, `dmet` and `fmet`, and optionally `dmes` and `fmes` for the 2B extensions of EAM and EEAM (see Section 2.3.2). The force arrays are generated from these (by the METAL_TABLE_DERIVATIVES routine) using a five point interpolation procedure. During simulation, interactions beyond distance $Min(r_{\rm cut},\texttt{limit 2})$ are discarded, whereas interactions at distances shorter than `limit 1` will cause the simulation to abort. For the purpose of extrapolating the embedding functions $F(\rho)$ beyond its `limit 2` specified in the tabulated array, it is assumed that

$$F(\rho > \texttt{limit 2}) = F(\rho = \texttt{limit 2}) \quad . \tag{6.14}$$

The simulation will however abort if any local density is less than the `limit 1` for its corresponding embedding function.

It is worth noting that in the 2BEAM and 2BEEAM the s-band contribution is usually only for the alloy component, so that local concentrations of a single element revert to the standard EAM or EEAM! In such case, the densities functions must be zeroed in the DL_POLY_4 TABEAM file. A convenient way to do this, for example, will be data record of the type:

```
SDEN Atom1 Atom1 1 0 1
0
```

6.1.8 The TABBND, TABANG, TABDIH & TABINV Files

DL_POLY_4 allows the specification of tabulated data for intramolecular interactions:

- TABBND - for chemical bonds potentials - distance dependent
- TABANG - for bond angles potentials - angle dependent
- TABDIH - for dihedrals (torsional) potentials - angle dependent
- TABINV - for inversions potentials - angle dependent.

The files have the same formatting rules with examples shown in Section 4.3. Refer to Section 4.1 for their derivation and usage in coarse grained model systems.

6.1.8.1 Definitions of Variables

record 1

`header`	a100	file header

record 2

`#`	a1	a hash (#) symbol
`cutpot`	real	cutoff in Å - **only expected in TABBND** as the cutoff ranges are known for TABANG, TABDIH & TABINV
`ngrid`	integer	number of grid points in table for all potentials

record 3

`#`	a1	a hash (#) symbol

The subsequent records define each tabulated potential in turn, in the order indicated by the specification in the FIELD file. Each potential is defined by a header record and a set of data records with the potential and force tables.

empty record:

id record:

`#`	a1	a hash (#) symbol
`atom 1`	a8	first atom type
`atom 2`	a8	second atom type

`atom 3`	a8	third atom type - only required for TABANG
`atom 4`	a8	forth atom type - only required for TABDIH & TABINV

interaction data records 0/1–ngrid:

`abscissa`	real	consecutive value over the full cutoff/range in Å for TABBND and degrees for TABANG, TABDIH & TABINV
`potential`	real	potential at the abscissa grid point in **units** as specified in FIELD
`force`	real	complementary force (virial for TABBND) value

6.1.8.2 Further Comments

It should be noted that the number of grid points in the table files should not be less than the number of grid points DL_POLY_4 is expecting. For more information the reader is advised to examine SETUP_MODULE and inspect the `mxg`***int*** variables, where ***int*** refers to `bnd` for bonds, `ang` for angles, `dih` for dihedrals and `inv` for inversions.

The potential and force tables are used to fill the internal arrays `v`***int*** and `g`***int*** for the respective ***int***rmolecular potential (see Chapter 2).

The contents of force arrays for TABBND are derived from the potential via the formula:

$$G(r) = -r\frac{\partial}{\partial r}U(r) \quad . \tag{6.15}$$

Note, this is *not* the same as the true force. During simulation, interactions beyond distance `cutpot` will bring the run to a controlled termination.

6.2 The OUTPUT Files

DL_POLY_4 produces up to 20 kinds of output files. Some of these are HISTORY, DEFECTS, MSDTMP, CFGMIN, OUTPUT, REVCON, REVIVE, RDFDAT, ZDNDAT, VDFDAT_* and STATIS. These respectively contain: an incremental dump file of all atomic coordinates, velocities and forces; an incremental dump file of atomic coordinates of defected particles (interstitials) and sites (vacancies); an incremental dump file of of individual atomic mean square displacement and temperature; a dump file of all atomic coordinates of a minimised structure; an incremental summary file of the simulation; a restart (final) configuration file; a restart (final) statistics accumulators file; a radial distribution function data file; Z-density distribution data file; velocity autocorrelation function data files (one file for each species) and a statistical history file.

6.2.1 The HISTORY File

The HISTORY file is the dump file of atomic coordinates, velocities and forces. Its principal use is for off-line analysis. The file is written by the subroutine TRAJECTORY_WRITE. The control variables for this file are `ltraj, nstraj, istraj` and `keytrj` which are created internally, based on information read from the **traj** directive in the CONTROL file (see Section 6.1.1). The HISTORY file will be created only if the directive **traj** appears in the CONTROL file.

The HISTORY file can become *very* large, especially if it is formatted. For serious simulation work it is recommended that the file be written to a scratch disk capable of accommodating a large data file. Alternatively, the file may be written in netCDF format instead of in ASCII (users must change ensure this functionality is available), which has the additional advantage of speed.

The HISTORY has the following structure:

record 1

`header`	a72	file header

record 2

`keytrj`	integer	trajectory key (see Table 6.1) in last frame
`imcon`	integer	periodic boundary key (see Table 6.6) in last frame
`megatm`	integer	number of atoms in simulation cell in last frame
`frame`	integer	number configuration frames in file
`records`	integer	number of records in file

For timesteps greater than `nstraj` the HISTORY file is appended at intervals specified by the **traj** directive in the CONTROL file, with the following information for each configuration:

record i

`timestep`	a8	the character string "timestep"
`nstep`	integer	the current time-step
`megatm`	integer	number of atoms in simulation cell (again)
`keytrj`	integer	trajectory key (again)
`imcon`	integer	periodic boundary key (again)
`tstep`	real	integration timestep (ps)
`time`	real	elapsed simulation time (ps)

record ii

`cell(1)`	real	x component of a cell vector
`cell(2)`	real	y component of a cell vector
`cell(3)`	real	z component of a cell vector

record iii

`cell(4)`	real	x component of b cell vector
`cell(5)`	real	y component of b cell vector
`cell(6)`	real	z component of b cell vector

record iv

`cell(7)`	real	x component of c cell vector
`cell(8)`	real	y component of c cell vector
`cell(9)`	real	z component of c cell vector

This is followed by the configuration for the current timestep. i.e. for each atom in the system the following data are included:

record a

`atmnam`	a8	atomic label
`iatm`	integer	atom index
`weight`	real	atomic mass (a.m.u.)
`charge`	real	atomic charge (e)
`rsd`	real	displacement from position at $t = 0$ (Å)

record b

`xxx`	real	x coordinate
`yyy`	real	y coordinate
`zzz`	real	z coordinate

record c only for `keytrj` > 0

`vxx`	real	x component of velocity
`vyy`	real	y component of velocity
`vzz`	real	z component of velocity

record d only for `keytrj` > 1

`fxx`	real	x component of force
`fyy`	real	y component of force
`fzz`	real	z component of force

Thus the data for each atom is a minimum of two records and a maximum of 4.

6.2.2 The MSDTMP File

The MSDTMP file is the dump file of individual atomic mean square displacements (square roots in Å) and mean square temperature (square roots in Kelvin). Its principal use is for off-line analysis. The file is written by the subroutine MSD_WRITE. The control variables for this file are `l_msd, nstmsd, istmsd` which are created internally, based on information read from the **msdtmp** directive in the CONTROL file (see Section 6.1.1). The MSDTMP file will be created only if the directive **msdtmp** appears in the CONTROL file.

The MSDTMP file can become *very* large, especially if it is formatted. For serious simulation work it is recommended that the file be written to a scratch disk capable of accommodating a large data file.

The MSDTMP has the following structure:

record 1

`header`	a52	file header

record 2

`megatm`	integer	number of atoms in simulation cell in last frame
`frame`	integer	number configuration frames in file
`records`	integer	number of records in file

For timesteps greater than `nstmsd` the MSDTMP file is appended at intervals specified by the **msdtmp** directive in the CONTROL file, with the following information for each configuration:

record i

`timestep`	a8	the character string "timestep"
`nstep`	integer	the current time-step
`megatm`	integer	number of atoms in simulation cell (again)
`tstep`	real	integration timestep (ps)
`time`	real	elapsed simulation time (ps)

This is followed by the configuration for the current timestep. i.e. for each atom in the system the following data are included:

record a

`atmnam`	a8	atomic label
`iatm`	integer	atom index
$\sqrt{\texttt{MSD}(t)}$	real	square root of the atomic mean square displacements (in Å)
T_{mean}	real	atomic mean temperature (in Kelvin)

6.2.3 The DEFECTS File

The DEFECTS file is the dump file of atomic coordinates of defects (see Section 6.1.4). Its principal use is for off-line analysis. The file is written by the subroutine DEFECTS_WRITE. The control variables for this file are `ldef, nsdef, isdef` and `rdef` which are created internally, based on information read from the **defects** directive in the CONTROL file (see Section 6.1.1). The DEFECTS file will be created only if the directive **defects** appears in the CONTROL file.

The DEFECTS file may become *very* large, especially if it is formatted. For serious simulation work it is recommended that the file be written to a scratch disk capable of accommodating a large data file.

The DEFECTS has the following structure:

record 1

`header`	a72	file header

record 2

`rdef`	real	site-interstitial cutoff (Å) in last frame
`frame`	integer	number configuration frames in file
`records`	integer	number of records in file

For timesteps greater than `nsdef` the DEFECTS file is appended at intervals specified by the **defects** directive in the CONTROL file, with the following information for each configuration:

record i		
`timestep`	a8	the character string "timestep"
`nstep`	integer	the current time-step
`tstep`	real	integration timestep (ps)
`time`	real	elapsed simulation time (ps)
`imcon`	integer	periodic boundary key (see Table 6.6)
`rdef`	real	site-interstitial cutoff (Å)
record ii		
`defects`	a7	the character string "defects"
`ndefs`	integer	the total number of defects
`interstitials`	a13	the character string "interstitials"
`ni`	integer	the total number of interstitials
`vacancies`	a9	the character string "vacancies"
`nv`	integer	the total number of vacancies
record iii		
`cell(1)`	real	x component of *a* cell vector
`cell(2)`	real	y component of *a* cell vector
`cell(3)`	real	z component of *a* cell vector
record iv		
`cell(4)`	real	x component of *b* cell vector
`cell(5)`	real	y component of *b* cell vector
`cell(6)`	real	z component of *b* cell vector
record v		
`cell(7)`	real	x component of *c* cell vector
`cell(8)`	real	y component of *c* cell vector
`cell(9)`	real	z component of *c* cell vector

This is followed by the `ni` interstitials for the current timestep, as each interstitial has the following data lines:

record a		
`atmnam`	a10	i_atomic label from CONFIG
`iatm`	integer	atom index from CONFIG
record b		
`xxx`	real	x coordinate
`yyy`	real	y coordinate
`zzz`	real	z coordinate

This is followed by the `nv` vacancies for the current timestep, as each vacancy has the following data lines:

record a		
`atmnam`	a10	v_atomic label from REFERENCE
`iatm`	integer	atom index from REFERENCE
record b		
`xxx`	real	x coordinate from REFERENCE

`yyy`	real	y coordinate from REFERENCE
`zzz`	real	z coordinate from REFERENCE

6.2.4 The RSDDAT File

The RSDDAT file is the dump file of atomic coordinates of atoms that are displaced from their original position at $t = 0$ farther than a preset cutoff. Its principal use is for off-line analysis. The file is written by the subroutine RSD_WRITE. The control variables for this file are `lrsd, nsrsd, isrsd` and `rrsd` which are created internally, based on information read from the **displacements** directive in the CONTROL file (see Section 6.1.1). The RSDDAT file will be created only if the directive **defects** appears in the CONTROL file.

The RSDDAT file may become *very* large, especially if it is formatted. For serious simulation work it is recommended that the file be written to a scratch disk capable of accommodating a large data file.

The RSDDAT has the following structure:

record 1		
`header`	a72	file header
record 2		
`rdef`	real	displacement qualifying cutoff (Å) in last frame
`frame`	integer	number configuration frames in file
`records`	integer	number of records in file

For timesteps greater than `nsrsd` the RSDDAT file is appended at intervals specified by the **displacements** directive in the CONTROL file, with the following information for each configuration:

record i		
`timestep`	a8	the character string "timestep"
`nstep`	integer	the current time-step
`tstep`	real	integration timestep (ps)
`time`	real	elapsed simulation time (ps)
`imcon`	integer	periodic boundary key (see Table 6.6)
`rrsd`	real	displacement qualifying cutoff (Å)
record ii		
`displacements`	a13	the character string "displacements"
`nrsd`	integer	the total number of displacements
record iii		
`cell(1)`	real	x component of *a* cell vector
`cell(2)`	real	y component of *a* cell vector
`cell(3)`	real	z component of *a* cell vector
record iv		
`cell(4)`	real	x component of *b* cell vector
`cell(5)`	real	y component of *b* cell vector
`cell(6)`	real	z component of *b* cell vector
record v		
`cell(7)`	real	x component of *c* cell vector
`cell(8)`	real	y component of *c* cell vector
`cell(9)`	real	z component of *c* cell vector

This is followed by the `nrsd` displacements for the current timestep, as each atom has the following data lines:

record a

`atmnam`	a10	atomic label from CONFIG
`iatm`	integer	atom index from CONFIG
`ratm`	real	atom displacement from its position at $t = 0$
record b		
`xxx`	real	x coordinate
`yyy`	real	y coordinate
`zzz`	real	z coordinate

6.2.5 The CFGMIN File

The CFGMIN file only appears if the user has selected the programmed minimisation option (directive **minim**ise (or **optim**ise) in the CONTROL file). Its contents have the same format as the CONFIG file (see Section 6.1.2), but contains only atomic position data and will never contain either velocity or force data (i.e. parameter `levcfg` is always zero). In addition, three extra numbers appear on the end of the second line of the file:

1. an integer indicating the number of minimisation cycles required to obtain the structure,
2. the configuration energy of the minimised configuration expressed in DL_POLY_4 units (Section 1.3.7), and
3. the configuration energy of the initial structure expressed in DL_POLY_4 units (Section 1.3.7).

6.2.6 The OUTPUT File

The job output consists of 7 sections: Header; Simulation control specifications; Force field specification; System specification; Summary of the initial configuration; Simulation progress; Sample of the final configuration; Summary of statistical data; and Radial distribution functions and Z-density profile. These sections are written by different subroutines at various stages of a job. Creation of the OUTPUT file *always* results from running DL_POLY_4. It is meant to be a human readable file, destined for hardcopy output.

6.2.6.1 Header

Gives the DL_POLY_4 version number, the number of processors in use, the link-cell algorithm in use and a title for the job as given in the header line of the input file CONTROL. This part of the file is written from the subroutines DL_POLY_, SET_BOUNDS and READ_CONTROL.

6.2.6.2 Simulation Control Specifications

Echoes the input from the CONTROL file. Some variables may be reset if illegal values were specified in the CONTROL file. This part of the file is written from the subroutine READ_CONTROL.

6.2.6.3 Force Field Specification

Echoes the FIELD file. A warning line will be printed if the system is not electrically neutral. This warning will appear immediately before the non-bonded short-range potential specifications. This part of the file is written from the subroutine READ_FIELD.

6.2.6.4 System Specification

Echoes system name, periodic boundary specification, the cell vectors and volume, some initial estimates of long-ranged corrections the energy and pressure (if appropriate), some concise information on topology and degrees of freedom break-down list. This part of the file is written from the subroutines SCAN_CONFIG, CHECK_CONFIG, SYSTEM_INIT, REPORT_TOPOLOGY and SET_TEMPERATURE.

6.2.6.5 Summary of the Initial Configuration

This part of the file is written from the main subroutine DL_POLY_. It states the initial configuration of (a maximum of) 20 atoms in the system. The configuration information given is based on the value of `levcfg` in the CONFIG file. If `levcfg` is 0 (or 1) positions (and velocities) of the 20 atoms are listed. If `levcfg` is 2 forces are also written out.

6.2.6.6 Simulation Progress

This part of the file is written by the DL_POLY_4 root segment DL_POLY_. The header line is printed at the top of each page as:

```
--------------------------------------------------------------------------------------------------

    step   eng_tot  temp_tot   eng_cfg   eng_src   eng_cou   eng_bnd   eng_ang   eng_dih   eng_tet
time(ps)    eng_pv  temp_rot   vir_cfg   vir_src   vir_cou   vir_bnd   vir_ang   vir_con   vir_tet
cpu  (s)    volume  temp_shl   eng_shl   vir_shl     alpha      beta     gamma   vir_pmf     press

--------------------------------------------------------------------------------------------------
```

The labels refer to :

line 1

`step`	MD step number
`eng_tot`	total internal energy of the system
`temp_tot`	system temperature (in Kelvin)
`eng_cfg`	configurational energy of the system
`eng_src`	configurational energy due to short-range potential contributions
`eng_cou`	configurational energy due to electrostatic potential
`eng_bnd`	configurational energy due to chemical bond potentials
`eng_ang`	configurational energy due to valence angle and three-body potentials
`eng_dih`	configurational energy due to dihedral inversion and four-body potentials
`eng_tet`	configurational energy due to tethering potentials

line 2

`time(ps)`	elapsed simulation time (in pico-seconds) since the beginning of the job
`eng_pv`	enthalpy of system
`temp_rot`	rotational temperature (in Kelvin)
`vir_cfg`	total configurational contribution to the virial
`vir_src`	short range potential contribution to the virial
`vir_cou`	electrostatic potential contribution to the virial
`vir_bnd`	chemical bond contribution to the virial
`vir_ang`	angular and three-body potentials contribution to the virial
`vir_con`	constraint bond contribution to the virial
`vir_tet`	tethering potential contribution to the virial

line 3

`cpu (s)`	elapsed cpu time (in seconds) since the beginning of the job

`volume`	system volume (in Å^3)
`temp_shl`	core-shell temperature (in Kelvin)
`eng_shl`	configurational energy due to core-shell potentials
`vir_shl`	core-shell potential contribution to the virial
`alpha`	angle between b and c cell vectors (in degrees)
`beta`	angle between c and a cell vectors (in degrees)
`gamma`	angle between a and b cell vectors (in degrees)
`vir_pmf`	PMF constraint contribution to the virial
`press`	pressure (in kilo-atmospheres)

Note: The total internal energy of the system (variable `tot_energy`) includes all contributions to the energy (including system extensions due to thermostats etc.). It is nominally the *conserved variable* of the system, and is not to be confused with conventional system energy, which is a sum of the kinetic and configuration energies.

The interval for printing out these data is determined by the directive **print** in the CONTROL file. At each time-step that printout is requested the instantaneous values of the above statistical variables are given in the appropriate columns. Immediately below these three lines of output the rolling averages of the same variables are also given. The maximum number of time-steps used to calculate the rolling averages is controlled by the directive **stack** in file CONTROL (see above) and listed as parameter `mxstak` in the SETUP_MODULE file (see Section 7.2.2). The default value is `mxstak` = 100.

Energy Units

The energy unit for the energy and virial data appearing in the OUTPUT is defined by the **units** directive appearing in the FIELD file. System energies are therefore read in **units** per MD cell.

Pressure Units

The unit of pressure is katms, irrespective of what energy unit is chosen.

6.2.6.7 Sample of Final Configuration

The positions, velocities and forces of the 20 atoms used for the sample of the initial configuration (see above) are given. This is written by the main subroutine DL_POLY.

6.2.6.8 Summary of Statistical Data

This portion of the OUTPUT file is written from the subroutine STATISTICS_RESULT. The number of time-steps used in the collection of statistics is given. Then the averages over the production portion of the run are given for the variables described in the previous section. The root mean square variation in these variables follow on the next two lines. The energy and pressure units are as for the preceding section.

Also provided in this section are estimates of the diffusion coefficient and the mean square displacement for the different atomic species in the simulation. These are determined from a *single time origin* and are therefore approximate. Accurate determinations of the diffusion coefficients can be obtained using the MSD utility program, which processes the HISTORY file (see DL_POLY_Classic User Manual).

If an NPT (N$\underline{\underline{\sigma}}$T) simulation is performed the OUTPUT file also provides the mean pressure (stress tensor) and mean simulation cell vectors. In case when extended N$\underline{\underline{\sigma}}$T ensembles are used then further mean (x, y) plain area and mean surface tension are also displayed in the OUTPUT file.

6.2.6.9 Radial Distribution Functions

If both calculation and printing of radial distribution functions have been requested (by selecting directives **rdf** and **print rdf** in the CONTROL file) radial distribution functions are printed out. This is written from the subroutine RDF_COMPUTE. First the number of time-steps used for the collection of the histograms is stated. Then each pre-requested function is given in turn. For each function a header line states the atom types ('a' and 'b') represented by the function. Then r, $g(r)$ and $n(r)$ are given in tabular form. Output is given from 2 entries before the first non-zero entry in the $g(r)$ histogram. $n(r)$ is the average number of atoms of type 'b' within a sphere of radius r around an atom of type 'a'. Note that a readable version of these data is provided by the RDFDAT file (below).

6.2.6.10 Z-density Profile

If both calculation and printing of Z-density profiles have been requested (by selecting directives **zden** and **print zden** in the CONTROL file Z-density profiles are printed out as the last part of the OUTPUT file. This is written by the subroutine Z_DENSITY_COMPUTE. First the number of time-steps used for the collection of the histograms is stated. Then each function is given in turn. For each function a header line states the atom type represented by the function. Then z, $\rho(z)$ and $n(z)$ are given in tabular form. Output is given from $Z = [-L/2, L/2]$ where L is the length of the MD cell in the Z direction and $\rho(z)$ is the mean number density. $n(z)$ is the running integral from $-L/2$ to z of (xy cell area)$\times \rho(s)\ ds$. Note that a readable version of these data is provided by the ZDNDAT file (below).

6.2.6.11 Velocity Autocorrelation Functions

If both calculation and printing of velocity autocorrelation functions have been requested (by selecting directives **vaf** and **print vaf** in the CONTROL file the velocity autocorrelation function for the system (either time-averaged or the last complete sample) is printed out as the last part of the OUTPUT file. This is written by the subroutine VAF_COMPUTE. First the details of the calculations are stated: either the number of samples used to give a time-averaged profile or the number of the last completed sample with its starting time. The absolute value of the velocity autocorrelation function for the system at $t = 0$, $C(0)$, is then stated. Then t and $Z(t)$ are given in tabular form. $Z(t) = C(t)/C(0)$ is the value of the velocity autocorrelation function, $C(t) = \langle \underline{v}_i(0) \cdot \underline{v}_i(t) \rangle$, scaled by $C(0) \equiv 3k_BT/m$. Note that a readable version of these data for individual species is provided by the VAFDAT files (below).

6.2.7 The REVCON File

This file is formatted and written by the subroutine REVIVE. REVCON is the restart configuration file. The file is written every `ndump` time steps in case of a system crash during execution and at the termination of the job. A successful run of DL_POLY_4 will always produce a REVCON file, but a failed job may not produce the file if an insufficient number of timesteps have elapsed. `ndump` is controlled by the directive **dump** in file CONTROL (see above) and listed as parameter `ndump` in the SETUP_MODULE file (see Section 7.2.2). The default value is `ndump` = 1000. REVCON is identical in format to the CONFIG input file (see Section 6.1.2). REVCON should be renamed CONFIG to continue a simulation from one job to the next. This is done for you by the *copy* macro supplied in the *execute* directory of DL_POLY_4.

6.2.8 The REVIVE File

This file is unformatted and written by the subroutine SYSTEM_REVIVE. It contains the accumulated statistical data. It is updated whenever the file REVCON is updated (see previous section). REVIVE should be renamed REVOLD to continue a simulation from one job to the next. This is done by the *copy* macro supplied in the *execute* directory of DL_POLY_4. In addition, to continue a simulation from a previous job the **restart** keyword must be included in the CONTROL file.

The format of the REVIVE file is identical to the REVOLD file described in Section 6.1.5.

6.2.9 The RDFDAT File

This is a formatted file containing *Radial Distribution Function* (RDF) data. Its contents are as follows:

record 1		
`cfgname`	a72	configuration name
record 2		
`ntprdf`	integer	number of different RDF pairs tabulated in file
`mxgrdf`	integer	number of grid points for each RDF pair

There follow the data for each individual RDF, i.e. `ntprdf` times. The data supplied are as follows:

first record		
`atname 1`	a8	first atom name
`atname 2`	a8	second atom name
following records (*mxgrdf* records)		
`radius`	real	interatomic distance (Å)
`g(r)`	real	RDF at given radius

Note 1. The RDFDAT file is optional and appears when the **print rdf** option is specified in the CONTROL file.

Note 2. Along with the RDFDAT file, two other files will be created whenever the **print ana**lysis directive is invoked: VDWPMF & VDWTAB, both containing the data for potentials of mean force and the corresponding virials calculated based on the obtained RDF:s, i.e. PMF $\sim -\ln(\text{RDF})$ (in the energy units specified in the FIELD file). These files have a simple three column format, the same as that used for *PMF files in the case of bonded units, see Section 4.2. The purpose of these files is to provide the user with means of setting up a PMF-based force-field, for example in the case of initial coarse-graining of an atomistic system. In particular, one can convert the VDWTAB file into a correctly formatted TABLE file (Section 6.1.6) by using the utility called `pmf2tab.f` (subject to compilation; found in DL_POLY_4 directory `utility`) as follows,

```
[user@host]$ pmf2tab.exe < VDWTAB
```

see Section 4.1 for completeness.

6.2.10 The ZDNDAT File

This is a formatted file containing the Z-density data. Its contents are as follows:

record 1		
`cfgname`	a72	configuration name
record 2		
`ntpatm`	integer	number of unique atom types profiled in file
`mxgrdf`	integer	number of grid points in the Z-density function

There follow the data for each individual Z-density function, i.e. `ntpatm` times. The data supplied are as follows:

first record

`atname`	a8	unique atom name

following records (*mxgrdf* records)

`z`	real	distance in z direction (Å)
$\rho(z)$	real	Z-density at given height `z`

Note the ZDNDAT file is optional and appears when the **print rdf** option is specified in the CONTROL file.

6.2.11 The VAFDAT Files

These are formatted files containing *Velocity Autocorrelation Function* (VAF) data. An individual file is created for each atomic species, i.e. `VAFDAT_`*atname*. Their contents are as follows:

record

`cfgname`	a72	configuration name

There follow the data for the VAF, either a single time-averaged profile or successive profiles separated by two blank lines. The data supplied are as follows:

first record

`atname`	a8	atom name
`binvaf`	integer	number of data points in VAF profile, *excluding* $t = 0$
`vaforigin`	real	absolute value of VAF at $t = 0$ ($C(0) \equiv 3k_BT/m$)
`vaftime0`	real	simulation time (ps) at beginning of (last) VAF profile ($t = 0$)

following records (*binvaf*+1 records)

`t`	real	time (ps)
`Z(t)`	real	scaled velocity autocorrelation function ($C(t)/C(0)$) at given time t

Note the VAFDAT files are optional and appear when the **print vaf** option is specified in the CONTROL file.

6.2.12 The *INT*DAT, *INT*PMF & *INT*TAB Files

These files, where ***INT*** is referring to ***INT***ra-molecular interactions and ***VDW***(RDF derived intermolecular), have very similar formatting rules with some examples shown in Section 4.2. Refer to Section 4.2 for their meaning and usage in coarse grained model systems.

record 1

`title`	a100	file header title

record 2

`header`	a100	file information header

record 3

`# info`	a30	information to follow string
`bins`	integer	number of bins for all PDFs
`cutoff`	real	cutoff in Å for bonds and RDFs or degrees for angular intramolecular interactions
`types`	integer	number of unique types of these interactions

record 4

`#`	a1	a hash (#) symbol

record 5

`# info 1`	a100	information to follow string
record 6		
`#`	a1	a hash (#) symbol

The subsequent records define each PDF potential in turn, in the order indicated by the specification in the FIELD file. Each potential is defined by a header record and a set of data records with the potential-like and force-like tables.

empty record:
id record:

`# info`	a25	information to follow string
`atom 1`	a8	first atom type
`atom 2`	a8	second atom type
`atom 3`	a8	third atom type - only available in ANG* files
`atom 4`	a8	forth atom type - only available in DIH* & INV* files
`index`	integer	unique index of PDF in file
`instances`	integer	instances of this unique type of PDF
interaction data records 1–bins:		
`abscissa`	real	consecutive value over the full cutoff/range in Å for BNDTAB & VDWTAB and degrees for ANGTAB, DIHTAB & INVTAB
`potential`	real	potential at the abscissa grid point in **units** as specified in FIELD
`force`	real	complementary force (virial for BNDTAB & VDWTAB) value

6.2.13 The STATIS File

The file is formatted, with integers as "i10" and reals as "e14.6". It is written by the subroutine STATISTICS_COLLECT. It consists of two header records followed by many data records of statistical data.

record 1		
`cfgname`	a72	configuration name
record 2		
`string`	a8	energy units

Data records
Subsequent lines contain the instantaneous values of statistical variables dumped from the array `stpval`. A specified number of entries of `stpval` are written in the format "(1p,5e14.6)". The number of array elements required (determined by the parameter `mxnstk` in the SETUP_MODULE file) is

$$\begin{aligned}\texttt{mxnstk} \geq\ & 27 + \texttt{ntpatm}\text{ (number of unique atomic sites) } + \\ & 9\text{ (stress tensor elements) } + \\ & 9\text{ (if constant pressure simulation requested) } + \quad 2 * mxatdm\text{ (if msdtmp option is used)}\end{aligned}$$

The STATIS file is appended at intervals determined by the **stats** directive in the CONTROL file. The energy unit is as specified in the FIELD file with the **units** directive, and are compatible with the data appearing in the OUTPUT file. The contents of the appended information is:

record i		
`nstep`	integer	current MD time-step
`time`	real	elapsed simulation time
`nument`	integer	number of array elements to follow
record ii `stpval`(1) – `stpval`(5)		
`engcns`	real	total extended system energy

		(i.e. the conserved quantity)
`temp`	real	system temperature
`engcfg`	real	configurational energy
`engsrc`	real	short range potential energy
`engcpe`	real	electrostatic energy
record iii `stpval`(6) – `stpval`(10)		
`engbnd`	real	chemical bond energy
`engang`	real	valence angle and 3-body potential energy
`engdih`	real	dihedral, inversion, and 4-body potential energy
`engtet`	real	tethering energy
`enthal`	real	enthalpy (total energy + PV)
record iv `stpval`(11) – `stpval`(15)		
`tmprot`	real	rotational temperature
`vir`	real	total virial
`virsrc`	real	short-range virial
`vircpe`	real	electrostatic virial
`virbnd`	real	bond virial
record v `stpval`(16) – `stpval`(20)		
`virang`	real	valence angle and 3-body virial
`vircon`	real	constraint bond virial
`virtet`	real	tethering virial
`volume`	real	volume
`tmpshl`	real	core-shell temperature
record vi `stpval`(21) – `stpval`(25)		
`engshl`	real	core-shell potential energy
`virshl`	real	core-shell virial
`alpha`	real	MD cell angle α
`beta`	real	MD cell angle β
`gamma`	real	MD cell angle γ
record vii `stpval`(26) – `stpval`(27)		
`virpmf`	real	PMF constraint virial
`press`	real	pressure
the next `ntpatm` **entries**		
`amsd(1)`	real	mean squared displacement of first atom types
`amsd(2)`	real	mean squared displacement of second atom types
`...`	...	...
`amsd(ntpatm)`	real	mean squared displacement of last atom types
the next 9 entries for the stress tensor		
`stress(1)`	real	xx component of stress tensor
`stress(2)`	real	xy component of stress tensor
`stress(3)`	real	xz component of stress tensor
`stress(4)`	real	yx component of stress tensor
`...`	real	...
`stress(9)`	real	zz component of stress tensor
the next 9 entries - *if* a NPT or N$\underline{\underline{\sigma}}$T simulation is undertaken		
`cell(1)`	real	x component of a cell vector
`cell(2)`	real	y component of a cell vector
`cell(3)`	real	z component of a cell vector
`cell(4)`	real	x component of b cell vector
`...`	real	...
`cell(9)`	real	z component of c cell vector

Chapter 7

The DL_POLY_4 Parallelisation and Source Code

Scope of Chapter

This chapter we discuss the DL_POLY_4 parallelisation strategy, describe the principles used in the DL_POLY_4 modularisation of the source code and list the file structure found in the *source* subdirectory.

7.1 Parallelisation

DL_POLY_4 is a distributed parallel molecular dynamics package based on the Domain Decomposition parallelisation strategy [2, 3, 9, 10, 5, 6]. In this section we briefly outline the basic methodology. Users wishing to add new features DL_POLY_4 will need to be familiar with the underlying techniques as they are described in the above references.

7.1.1 The Domain Decomposition Strategy

The Domain Decomposition (DD) strategy [2, 3, 5] is one of several ways to achieve parallelisation in MD. Its name derives from the division of the simulated system into equi-geometrical spatial blocks or domains, each of which is allocated to a specific processor of a parallel computer. I.e. the arrays defining the atomic coordinates $\underline{r}_i$, velocities $\underline{v}_i$ and forces $\underline{f}_i$, for all N atoms in the simulated system, are divided in to sub-arrays of approximate size N/P, where P is the number of processors, and allocated to specific processors. In DL_POLY_4 the domain allocation is handled by the routine DOMAINS_MODULE and the decision of approximate sizes of various bookkeeping arrays in SET_BOUNDS. The division of the configuration data in this way is based on the location of the atoms in the simulation cell, such a *geometric* allocation of system data is the hallmark of DD algorithms. Note that in order for this strategy to work efficiently, the simulated system must possess a reasonably uniform density, so that each processor is allocated almost an equal portion of atom data (as much as possible). Through this approach the forces computation and integration of the equations of motion are shared (reasonably) equally between processors and to a large extent can be computed independently on each processor. The method is conceptually simple though tricky to program and is particularly suited to large scale simulations, where efficiency is highest.

The DD strategy underpinning DL_POLY_4 is based on the link cell algorithm of Hockney and Eastwood [85] as implemented by various authors (e.g. Pinches *et al.* [9] and Rapaport [10]). This requires that the cutoff applied to the interatomic potentials is relatively short ranged. In DL_POLY_4 the link-cell list is build by the routine LINK_CELL_PAIRS. As with all DD algorithms, there is a need for the processors to exchange 'halo data', which in the context of link-cells means sending the contents of the link cells at the boundaries of each domain, to the neighbouring processors, so that each may have all necessary information to compute the pair forces acting on the atoms belonging to its allotted domain. This in DL_POLY_4 is handled by the SET_HALO_PARTICLES routine.

Systems containing complex molecules present several difficulties. They often contain ionic species, which usually require Ewald summation methods [22, 86], and *intra*-molecular interactions in addition to *inter*-molecular forces. Intramolecular interactions are handled in the same way as in DL_POLY_Classic, where each processor is allocated a subset of intramolecular bonds to deal with. The allocation in this case is based on the atoms present in the processor's domain. The SHAKE and RATTLE algorithms [74, 23] require significant modification. Each processor must deal with the constraint bonds present in its own domain, but it must also deal with bonds it effectively shares with its neighbouring processors. This requires each processor to inform its neighbours whenever it updates the position of a shared atom during every SHAKE (RATTLE_VV1) cycle (RATTLE_VV2 updates the velocities), so that all relevant processors may incorporate this update into its own iterations. In the case of the DD strategy the SHAKE (RATTLE) algorithm is simpler than for the Replicated Data method of DL_POLY_Classic, where global updates of the atom positions (merging and splicing) are required [87]. The absence of the merge requirement means that the DD tailored SHAKE and RATTLE are less communications dependent and thus more efficient, particularly with large processor counts.

The DD strategy is applied to complex molecular systems as follows:

1. Using the atomic coordinates $\underline{r}_i$, each processor calculates the forces acting between the atoms in its domain - this requires additional information in the form of the halo data, which must be passed from the neighbouring processors beforehand. The forces are usually comprised of:

   (a) All common forms of non-bonded atom-atom (van der Waals) forces
   (b) Atom-atom (and site-site) coulombic forces
   (c) Metal-metal (local density dependent) forces
   (d) Tersoff (local density dependent) forces (for hydro-carbons) [17]
   (e) Three-body valence angle and hydrogen bond forces
   (f) Four-body inversion forces
   (g) Ion core-shell polarasation
   (h) Tether forces
   (i) Chemical bond forces
   (j) Valence angle forces
   (k) Dihedral angle (and improper dihedral angle) forces
   (l) Inversion angle forces
   (m) External field forces.

2. The computed forces are accumulated in atomic force arrays $\underline{f}_i$ independently on each processor
3. The force arrays are used to update the atomic velocities and positions of all the atoms in the domain
4. Any atom which effectively moves from one domain to another, is relocated to the neighbouring processor responsible for that domain.

It is important to note that load balancing (i.e. equal and concurrent use of all processors) is an essential requirement of the overall algorithm. In DL_POLY_4 this is accomplished quite naturally through the DD partitioning of the simulated system. Note that this will only work efficiently if the density of the system is reasonably uniform. THERE ARE NO LOAD BALANCING ALGORITHMS IN DL_POLY_4 TO COMPENSATE FOR A BAD DENSITY DISTRIBUTION.

7.1.2 Distributing the Intramolecular Bonded Terms

The intramolecular terms in DL_POLY_4 are managed through bookkeeping arrays which list all atoms (sites) involved in a particular interaction and point to the appropriate arrays of parameters that define the potential. Distribution of the forces calculations is accomplished by the following scheme:

1. Every atom (site) in the simulated system is assigned a unique 'global' index number from 1 to N.
2. Every processor maintains a list of the local indices of the atoms in its domain. (This is the local atom list.)
3. Every processor also maintains a sorted (in ascending order) local list of global atom indices of the atoms in its domain. (This is the local sorted atom list.)
4. Every intramolecular bonded term U_{type} in the system has a unique index number i_{type}: from 1 to N_{type} where *type* represents a bond, angle, dihedral, or inversion. Also attached there with unique index numbers are core-shell units, bond constraint units, PMF constraint units, rigid body units and tethered atoms, their definition by site rather than by chemical type.
5. On each processor a pointer array $key_{type}(n_{type}, i_{type})$ carries the indices of the specific atoms involved in the potential term labelled i_{type} . The dimension n_{type} will be 1 if the term represents a tether, $1,\ 2$ for a core-shell unit or a bond constraint unit or a bond, $1,\ 2,\ 3$ for a valence angle and $1,\ 2,\ 3,\ 4$ for a dihedral or an inversion, $1, .., n_{\texttt{PMF unit}_{1 \texttt{ or } 2}} + 1$ for a PMF constraint unit, or $-1,\ 0,\ 1, .., n_{\texttt{RB unit}}$ for a rigid body unit.

6. Using the *key* array, each processor can identify the global indices of the atoms in the bond term and can use this in conjunction with the local sorted atoms list *and a binary search algorithm* to find the atoms in local atom list.

7. Using the local atom identity, the potential energy and force can be calculated.

It is worth mentioning that although rigid body units are not bearing any potential parameters, their definition requires that their topology is distributed in the same manner as the rest of the intra-molecular like interactions.

Note that, at the start of a simulation DL_POLY_4 allocates individual bonded interactions to specific processors, based on the domains of the relevant atoms (DL_POLY_4 routine BUILD_BOOK_INTRA). This means that each processor does not have to handle every possible bond term to find those relevant to its domain. Also this allocation is updated as atoms move from domain to domain i.e. during the *relocation* process that follows the integration of the equations of motion (DL_POLY_4 routine RELOCATE_PARTICLES). Thus the allocation of bonded terms is effectively dynamic, changing in response to local changes.

7.1.3 Distributing the Non-bonded Terms

DL_POLY_4 calculates the non-bonded pair interactions using the link cell algorithm due to Hockney and Eastwood [85]. In this algorithm a relatively short ranged potential cutoff ($r_{\rm cut}$) is assumed. The simulation cell is logically divided into so-called link cells, which have a width not less than (or equal to) the cutoff distance. It is easy to determine the identities of the atoms in each link cell. When the pair interactions are calculated it is already known that atom pairs can only interact if they are in the same link cell, or are in link cells that share a common face. Thus using the link cell 'address' of each atom, interacting pairs are located easily and efficiently via the 'link list' that identifies the atoms in each link cell. So efficient is this process that the link list can be recreated every time step at negligible cost.

For reasons, partly historical, the link list is used to construct a Verlet neighbour list [22]. The Verlet list records the indices of all atoms within the cutoff radius ($r_{\rm cut}$) of a given atom. The use of a neighbour list is not strictly necessary in the context of link-cells, but it has the advantage here of allowing a neat solution to the problem of 'excluded' pair interactions arising from the intramolecular terms and frozen atoms (see below).

In DL_POLY_4, the neighbour list is constructed *simultaneously* on each node, using the DD adaptation of the link cell algorithm to share the total burden of the work reasonably equally between nodes. Each node is thus responsible for a unique set of non-bonded interactions and the neighbour list is therefore different on each node.

A feature in the construction of the Verlet neighbour list for macromolecules is the concept of *excluded atoms*, which arises from the need to exclude certain atom pairs from the overall list. Which atom pairs need to be excluded is dependent on the precise nature of the force field model, but as a minimum atom pairs linked via extensible bonds or constraints and atoms (grouped in pairs) linked via valence angles are probable candidates. The assumption behind this requirement is that atoms that are formally bonded in a chemical sense, should not participate in non-bonded interactions. (However, this is not a universal requirement of all force fields.) The same considerations are needed in dealing with charged excluded atoms.

The modifications necessary to handle the excluded and frozen atoms are as follows. A distributed *excluded atoms list* is constructed by the DL_POLY_4 routine BUILD_EXCL_INTRA at the start of the simulation and is then used in conjunction with the Verlet neighbour list builder LINK_CELL_PAIRS to ensure that excluded interactions are left out of the pair force calculations. Note that, completely frozen pairs of atoms are excluded in the same manner. The excluded atoms list is updated during the atom relocation process described above (DL_POLY_4 routine EXCHANGE_PARTICLES).

Once the neighbour list has been constructed, each node of the parallel computer may proceed independently to calculate the pair force contributions to the atomic forces (see routine TWO_BODY_FORCES).

The potential energy and forces arising from the non-bonded interactions, as well as metal and Tersoff interactions are calculated using interpolation tables. These are generated in the following routines: VDW_GENERATE, METAL_GENERATE, METAL_TABLE_DERIVATIVES and TERSOFF_GENERATE.

7.1.4 Modifications for the Ewald Sum

For systems with periodic boundary conditions DL_POLY_4 employs the Ewald Sum to calculate the coulombic interactions (see Section 2.4.5). It should be noted that DL_POLY_4 uses only the Smoothed Particle Mesh (SPME) form of the Ewald sum.

Calculation of the real space component in DL_POLY_4 employs the algorithm for the calculation of the non-bonded interactions outlined above, since the real space interactions are now short ranged (implemented in EWALD_REAL_FORCES routine).

The reciprocal space component is calculated using Fast Fourier Transform (FFT) scheme of the SMPE method [66, 88] as discussed in Section 2.4.5. The parallelisation of this scheme is entirely handled within the DL_POLY_4 by the 3D FFT routine PARALLEL_FFT, (using GPFA_MODULE) which is known as the Daresbury advanced Fourier Transform, due to I.J. Bush [89]. This routine distributes the SPME charge array over the processors in a manner that is completely commensurate with the distribution of the configuration data under the DD strategy. As a consequence the FFT handles all the necessary communication implicit in a distributed SPME application. The DL_POLY_4 subroutine EWALD_SPME_FORCES perfoms the bulk of the FFT operations and charge array construction, while SPME_FORCES calculates the forces.

Other routines required to calculate the Ewald sum include EWALD_MODULE,
EWALD_EXCL_FORCES,
EWALD_FRZN_FORCES and SPME_CONTAINER.

7.1.5 Metal Potentials

The simulation of metals (Section 2.3.2) by DL_POLY_4 makes use of density dependent potentials. The dependence on the atomic density presents no difficulty however, as this class of potentials can be resolved into pair contributions. This permits the use of the distributed Verlet neighbour list as outlined above. DL_POLY_4 implements these potentials in various subroutines with names beginning with METAL_.

7.1.6 Tersoff, Three-Body and Four-Body Potentials

DL_POLY_4 can calculate Tersoff, three-body and four-body interactions. Although some of these interactions have similar terms to some intramolecular ones (three-body to the bond angle and four-body to inversion angle), these are not dealt with in the same way as the normal bonded interactions. They are generally very short ranged and are most effectively calculated using a link-cell scheme [85]. No reference is made to the Verlet neighbour list nor the excluded atoms list. It follows that atoms involved these interactions can interact via non-bonded (pair) forces and ionic forces also. Note that contributions from frozen pairs of atoms to these potentials are excluded. The calculation of the Tersoff three-body and four-body terms is distributed over processors on the basis of the domain of the central atom in them. DL_POLY_4 implements these potentials in the following routines TERSOFF_FORCES, TERSOFF_GENERATE, THREE_BODY_FORCES and FOUR_BODY_FORCES.

7.1.7 Globally Summed Properties

The final stage in the DD strategy, is the global summation of different (by terms of potentials) contributions to energy, virial and stress, which must be obtained as a global sum of the contributing terms calculated on all nodes.

The DD strategy does not require a global summation of the forces, unlike the Replicated Data method used in DL_POLY_Classic, which limits communication overheads and provides smooth parallelisation to large processor counts.

7.1.8 The Parallel (DD tailored) SHAKE and RATTLE Algorithms

The essentials of the DD tailored SHAKE and RATTLE algorithms (see Section 3.2) are as follows:

1. The bond constraints acting in the simulated system are allocated between the processors, based on the location (i.e. domain) of the atoms involved.
2. Each processor makes a list of the atoms bonded by constraints it must process. Entries are zero if the atom is not bonded.
3. Each processor passes a copy of the array to the neighbouring processors which manage the domains in contact with its own. The receiving processor compares the incoming list with its own and keeps a record of the shared atoms and the processors which share them.
4. In the first stage of the algorithms, the atoms are updated through the usual Verlet algorithm, without regard to the bond constraints.
5. In the second (iterative) stage of the algorithms, each processor calculates the incremental correction vectors for the bonded atoms in its own list of bond constraints. It then sends specific correction vectors to all neighbours that share the same atoms, using the information compiled in step 3.
6. When all necessary correction vectors have been received and added the positions of the constrained atoms are corrected.
7. Steps 5 and 6 are repeated until the bond constraints are converged.
8. Finally, the change in the atom positions from the previous time step is used to calculate the atomic velocities.

The compilation of the list of constrained atoms on each processor, and the circulation of the list (items 1 - 3 above) is done at the start of the simulation, but thereafter it needs only to be done every time a constraint bond atom is relocated from one processor to another. In this respect DD-SHAKE and DD-RATTLE resemble every other intramolecular term.

Since the allocation of constraints is based purely on geometric considerations, it is not practical to arrange for a strict load balancing of the DD-SHAKE and DD-RATTLE algorithms. For many systems, however, this deficiency has little practical impact on performance.

7.1.9 The Parallel Rigid Body Implementation

The essentials of the DD tailored RB algorithms (see Section 3.6) are as follows:

1. Every processor works out a list of all local and halo atoms that are qualified as free (zero entry) or as members of a RB (unit entry.
2. The rigid body units in the simulated system are allocated between the processors, based on the location (i.e. domain) of the atoms involved.
3. Each processor makes a list of the RB and their constituting atoms that are fully or partially owned by the processors domain.

4. Each processor passes a copy of the array to the neighbouring processors which manage the domains in contact with its own. The receiving processor compares the incoming list with its own and keeps a record of the shared RBs and RBs' constituent atoms, and the processors which share them. *Note that a RB can be shared between up to* **eight** *domains!*

5. The dynamics of each RB is calculated in full on each domain but domains only update $\{\underline{r}, \underline{v}, \underline{f}\}$ of RB atoms which they own. *Note that a site/atom belongs to* **one and only one** *domain at a time (no sharing) !*

6. Strict bookkeeping is necessary to avoid multiple counting of kinetic properties. $\{\underline{r}, \underline{v}, \underline{v}\}$ updates are necessary for halo parts (particles) of partially shared RBs. For all domains the kinetic contributions from each fully or partially present RB are evaluated in full and then waited with the ratio - number of RB's sites local to the domain to total RB's sites, and then globally summed.

The compilation of the lists in items 1 - 3 above and their circulation of the list is done at the start of the simulation, but thereafter these need updating on a local level every time a RB site/atom is relocated from one processor to another. In this respect RBs topology transfer resembles every other intramolecular term.

Since the allocation of RBs is based purely on geometric considerations, it is not practical to arrange for a strict load balancing. For many systems, however, this deficiency has little practical impact on performance.

7.2 Source Code

7.2.1 Modularisation Principles

Modules in DL_POLY_4 are constructed to define parameters and variables (scalars and arrays) and/or develop methods that share much in common. The division is far from arbitrary and module interdependence is reduced to minimum. However, some dependencies exist which leads to the following division by groups in hierarchical order:

- **precision module**:: KINDS_F90

  The precision module defines the working precision `wp` of all real variables and parameters in DL_POLY_4. By default it is set to 64-bit (double) precision. If the precision is changed, the user must check whether the specific platform supports it and make sure it is allowed for in the MPI implementation. If all is OK then the code must be recompiled.

- **MPI module**:: MPI_MODULE

  The MPI module implements all MPI functional calls used in DL_POLY_4. It is only used when DL_POLY_4 is to be compiled in serial mode.

- **communication module**:: COMMS_MODULE (MPI_MODULE)

  The communication module defines MPI related parameters and develops MPI related functions and subroutines such as: initialisation and exit; global synchronisation, sum, maximum and minimum; node ID and number of nodes; simulation time. It is dependent on KINDS_F90 and on MPI_MODULE if MPI is emulated for DL_POLY_4 compilation in serial mode. The MPI_MODULE implements all MPI functional calls used in DL_POLY_4.

- **global parameters module**:: SETUP_MODULE

  The global parameters module holds all important global variables and parameters (see above). It is dependent on KINDS_F90.

- **parse module**:: PARSE_MODULE

The parse module develops several methods used to deal with textual input: `get_line` `strip_blanks` `lower_case` `get_word` `word_2_real`. Depending on the method dependencies on KINDS_F90 COMMS_MODULE SETUP_MODULE DOMAINS_MODULE are found.

- **development module**:: DEVELOPMENT_MODULE

  The development module contains several methods used to help with testing and debugging DL_POLY_4. Depending on the method dependencies on KINDS_F90 COMMS_MODULE SETUP_MODULE DOMAINS_MODULE are found.

- **I/O module**:: IO_MODULE

  The I/O module contains all important global variables that define the I/O methods and types used in the package and contains basic routines essential for the I/O in DL_POLY_4. It is dependent on KINDS_F90.

- **domains module**:: DOMAINS_MODULE

  The domains module defines DD parameters and maps the available computer resources on a DD grid. The module does not depend on previous modules but its mapping subroutine is dependent on KINDS_F90 and COMMS_MODULE.

- **site module**:: SITE_MODULE

  The site module defines all site related arrays (FIELD) and is dependent on KINDS_F90 only. However, it also develops an allocation method that is dependent on SETUP_MODULE.

- **configuration module**:: CONFIG_MODULE

  The configuration module defines all configuration related arrays (CONFIG) and is dependent on KINDS_F90 only. However, it also develops an allocation method that is dependent on SETUP_MODULE.

- **vnl module**:: VNL_MODULE

  The Verlet neighbour list (VNL) module defines all VNL related control variables and arrays needed for the VNL conditional update functionality, and is dependent on KINDS_F90 only. However, it is assisted by a VNL_CHECK routine that is dependent on more modules.

- **defects module**:: DEFECTS_MODULE

  The defects module defines all defects and configuration related arrays (REFERENCE) and is dependent on KINDS_F90 only. However, it also develops an allocation method that is dependent on SETUP_MODULE.

- ***inter*-molecular interactions modules**:: VDW_MODULE METAL_MODULE TERSOFF_MODULE THREE_BODY_MODULE FOUR_BODY_MODULE

  The intermolecular modules define all variables and potential arrays needed for the calculation of the particular interaction in the DL_POLY_4 scope. They depend on KINDS_F90. Their allocation methods depend on SETUP_MODULE.

- ***intra*-molecular interactions and site-related modules**:: RDF_MODULE Z_DENSITY_MODULE CORE_SHELL_MODULE CONSTRAINTS_MODULE PMF_MODULE RIGID_BODIES_MODULE TETHERS_MODULE BONDS_MODULE ANGLES_MODULE DIHEDRALS_MODULE INVERSIONS_MODULE

  These modules define all variables and potential or statistical grid arrays needed for the calculation of the particular interaction or distribution function in the DL_POLY_4 scope. They all depend on KINDS_F90 with allocation methods depending on SETUP_MODULE.

- **external field module**:: EXTERNAL_FIELD_MODULE

  This module defines all variables and potential arrays needed for the application of an external field in the DL_POLY_4 scope. It depends on KINDS_F90 and its allocation method on SETUP_MODULE.

- **langevin module**:: LANGEVIN_MODULE

  This module defines all variables and arrays needed for the application of NPT and N$\underline{\underline{\sigma}}$T Langevin routines in the DL_POLY_4 scope. It depends on KINDS_F90 and its allocation method on SETUP_MODULE.

- **minimise module**:: MINIMISE_MODULE

  This module defines all variables and arrays needed for the application of a Conjugate Gradient Method minimisation routine in the DL_POLY_4 scope. It depends on KINDS_F90 and its allocation method on SETUP_MODULE.

- **ewald module**:: EWALD_MODULE

  This module defines all variables and arrays needed for the refreshment of SPME k-space driven properties in the DL_POLY_4 scope when an infrequent SPME option is opted for in CONTROL. It depends on KINDS_F90 and its allocation method on SETUP_MODULE.

- **msd module**:: MSD_MODULE

  This module globalises a CONTROL variable.

- **statistics module**:: STATISTICS_MODULE

  This module defines all variables and arrays needed for the statistical accountancy of a simulation in DL_POLY_4. It depends on KINDS_F90 and its allocation methods on SETUP_MODULE and COMMS_MODULE.

- **greenkubo module**:: GREENKUBO_MODULE

  This module defines all variables and arrays needed for calculation of Green-Kubo relations during a simulation in DL_POLY_4. It depends on KINDS_F90 and its allocation methods on SETUP_MODULE.

- **kinetic module**:: KINETIC_MODULE

  The kinetic module contains a collection of routines for the calculation of various kinetic properties. It is dependent on KINDS_F90.

7.2.2 File Structure

Generally, the DL_POLY_4 file structure can be divided into four groups as follows:

- **module files** in the *source* directory::

  KINDS_F90 COMMS_MODULE SETUP_MODULE
  PARSE_MODULE DEVELOPMENT_MODULE IO_MODULE
  DOMAINS_MODULE
  SITE_MODULE CONFIG_MODULE VNL_MODULE DEFECTS_MODULE DEFECTS1_MODULE
  VDW_MODULE METAL_MODULE TERSOFF_MODULE
  THREE_BODY_MODULE FOUR_BODY_MODULE
  RDF_MODULE Z_DENSITY_MODULE
  CORE_SHELL_MODULE
  CONSTRAINTS_MODULE PMF_MODULE
  RIGID_BODIES_MODULE
  TETHERS_MODULE
  BONDS_MODULE ANGLES_MODULE DIHEDRALS_MODULE INVERSIONS_MODULE

  EXTERNAL_FIELD_MODULE LANGEVIN_MODULE MINIMISE_MODULE
  EWALD_MODULE MSD_MODULE STATISTICS_MODULE GREENKUBO_MODULE

  KINETIC_MODULE GPFA_MODULE PARALLEL_FFT

- **general files** in the *source* directory::

  WARNING ERROR SCAN_CONTROL_IO
  NUMERIC_CONTAINER SPME_CONTAINER QUATERNIONS_CONTAINER
  SCAN_FIELD SCAN_CONTROL_PRE
  READ_CONFIG_PARALLEL SCAN_CONFIG SCAN_CONTROL READ_CONFIG
  SET_BOUNDS
  READ_CONTROL
  VDW_GENERATE VDW_TABLE_READ VDW_DIRECT_FS_GENERATE
  METAL_GENERATE_ERF METAL_GENERATE METAL_TABLE_READ METAL_TABLE_DERIVATIVES
  TERSOFF_GENERATE DIHEDRALS_14_CHECK READ_FIELD
  CHECK_CONFIG ORIGIN_CONFIG SCALE_CONFIG WRITE_CONFIG
  TRAJECTORY_WRITE SYSTEM_EXPAND
  RIGID_BODIES_TAGS RIGID_BODIES_COMS RIGID_BODIES_WIDTHS
  RIGID_BODIES_SETUP
  INIT_INTRA TAG_LEGEND REPORT_TOPOLOGY PASS_SHARED_UNITS
  BUILD_BOOK_INTRA BUILD_EXCL_INTRA
  SCALE_TEMPERATURE UPDATE_SHARED_UNITS
  CORE_SHELL_QUENCH CONSTRAINTS_TAGS CONSTRAINTS_QUENCH
  PMF_COMS PMF_TAGS PMF_VCOMS PMF_QUENCH
  RIGID_BODIES_QUENCH
  SET_TEMPERATURE
  VDW_LRC METAL_LRC SYSTEM_INIT VNL_CHECK
  EXPORT_ATOMIC_DATA SET_HALO_PARTICLES
  EXPORT_ATOMIC_POSITIONS REFRESH_HALO_POSITIONS
  RIGID_BODIES_STRESS
  READ_HISTORY
  DEFECTS_REFERENCE_READ DEFECTS_REFERENCE_READ_PARALLEL
  DEFECTS_REFERENCE_WRITE
  DEFECTS_REFERENCE_EXPORT DEFECTS_REFERENCE_SET_HALO
  DEFECTS_LINK_CELLS DEFECTS1_WRITE DEFECTS_WRITE
  MSD_WRITE RSD_WRITE VAF_WRITE
  IMPACT CORE_SHELL_ON_TOP
  DEPORT_ATOMIC_DATA PMF_UNITS_SET COMPRESS_BOOK_INTRA
  RELOCATE_PARTICLES
  LINK_CELL_PAIRS
  METAL_LD_COLLECT_EAM METAL_LD_COLLECT_FST
  METAL_LD_EXPORT METAL_LD_SET_HALO
  METAL_LD_COMPUTE
  EXCHANGE_GRID EWALD_SPME_FORCES
  METAL_FORCES VDW_FORCES EWALD_REAL_FORCES
  COUL_DDDP_FORCES COUL_CP_FORCES COUL_FSCP_FORCES
  COUL_RFP_FORCES RDF_COLLECT RDF_EXCL_COLLECT RDF_FRZN_COLLECT
  EWALD_EXCL_FORCES
  EWALD_FRZN_FORCES TWO_BODY_FORCES
  TERSOFF_FORCES THREE_BODY_FORCES FOUR_BODY_FORCES
  CORE_SHELL_FORCES TETHERS_FORCES
  INTRA_COUL BONDS_FORCES ANGLES_FORCES
  INVERSIONS_FORCES DIHEDRALS_14_VDW DIHEDRALS_FORCES
  EXTERNAL_FIELD_APPLY EXTERNAL_FIELD_CORRECT
  LANGEVIN_FORCES
  CONSTRAINTS_PSEUDO_BONDS PMF_PSEUDO_BONDS
  RIGID_BODIES_SPLIT_TORQUE RIGID_BODIES_MOVE MINIMISE_RELAX

CORE_SHELL_RELAX ZERO_K_OPTIMISE VAF_COLLECT
NVT_E0_SCL NVT_E1_SCL NVT_B0_SCL NVT_B1_SCL

XSCALE CORE_SHELL_KINETIC REGAUSS_TEMPERATURE

Z_DENSITY_COLLECT STATISTICS_COLLECT
STATISTICS_CONNECT_SET STATISTICS_CONNECT_SPREAD STATISTICS_CONNECT_FRAMES
SYSTEM_REVIVE
RDF_COMPUTE Z_DENSITY_COMPUTE VAF_COMPUTE
BONDS_COMPUTE ANGLES_COMPUTE DIHEDRALS_COMPUTE INVERSIONS_COMPUTE
STATISTICS_RESULT

W_IMPACT_OPTION W_WRITE_OPTIONS W_CALCULATE_FORCES W_REFRESH_MAPPINGS
W_KINETIC_OPTIONS W_STATISTICS_REPORT W_REFRESH_OUTPUT
W_REPLAY_HISTORY W_REPLAY_HISTORF

DL_POLY

- **VV specific** files in the *source/VV* directory::

  PSEUDO_VV
  CONSTRAINTS_SHAKE_VV PMF_SHAKE_VV
  CONSTRAINTS_RATTLE PMF_RATTLE
  NVT_H0_SCL NVT_G0_SCL NPT_H0_SCL NST_H0_SCL
  NVE_0_VV NVT_E0_VV
  NVT_L0_VV NVT_A0_VV NVT_B0_VV NVT_H0_VV NVT_G0_VV
  NPT_L0_VV NPT_B0_VV NPT_H0_VV NPT_M0_VV
  NST_L0_VV NST_B0_VV NST_H0_VV NST_M0_VV
  NVT_H1_SCL NVT_G1_SCL NPT_H1_SCL NST_H1_SCL
  NVE_1_VV NVT_E1_VV
  NVT_L1_VV NVT_A1_VV NVT_B1_VV NVT_H1_VV NVT_G1_VV
  NPT_L1_VV NPT_B1_VV NPT_H1_VV NPT_M1_VV
  NST_L1_VV NST_B1_VV NST_H1_VV NST_M1_VV
  W_AT_START_VV W_INTEGRATE_VV W_MD_VV

- **LFV specific** files in the *source/LFV* directory::

  PSEUDO_LFV
  CONSTRAINTS_SHAKE_LFV PMF_SHAKE_LFV
  NVE_0_LFV NVT_E0_LFV
  NVT_L0_LFV NVT_A0_LFV NVT_B0_LFV NVT_H0_LFV NVT_G0_LFV
  NPT_L0_LFV NPT_B0_LFV NPT_H0_LFV NPT_M0_LFV
  NST_L0_LFV NST_B0_LFV NST_H0_LFV NST_M0_LFV
  NVT_L1_LFV NVT_A1_LFV NVT_B1_LFV NVT_H1_LFV NVT_G1_LFV
  NPT_L1_LFV NPT_B1_LFV NPT_H1_LFV NPT_M1_LFV
  NST_L1_LFV NST_B1_LFV NST_H1_LFV NST_M1_LFV
  W_AT_START_LFV W_INTEGRATE_LFV W_MD_LFV

- **SERIAL specific** files in the *source/SERIAL* directory::

  MPIF.H MPI_MODULE EWALD_SPME_FORC~S

The files in each group are listed in hierarchal order as closely as possible. The further down the list the file, the more dependent it is on the files listed above it. The same hierarchal order is followed in the makefiles (see Appendix **??**).

It is worth noting that the all W_*.F90 files in *source* and *source/*V* are in fact inclusion files that are wrapped as routines within DL_POLY. This wrapping up of functional calls is done to shorten the length of the code by reusing general functional sequences where possible.

7.2.3 Module Files

The DL_POLY_4 module files contain all global variables (scalars and arrays) and parameters as well as some general methods and generic functions intrinsically related to the purpose or/and contents of the specific module. The file-names and the methods or/and functions developed in them have self-explanatory names. More information of their purpose can be found in their headers.

The rest of files in DL_POLY_4 are dependent on the module files in various ways. The dependency relation to a module file is explicitly stated in the declaration part of the code.

7.2.4 General Files

The DL_POLY_4 general files are common to both MPI and SERIAL version of the code. In most cases, they have self-explanatory names as their order is matched as closely as possible to that occurring in the main segment of the code - DL_POLY. Only the first five files are exception of that rule; WARNING and ERROR are important reporting subroutines that have call points at various places in the code, and NUMERIC_CONTAINER, and SPME_CONTAINER are containers of simple functions and subroutines related in some way to their purpose in the code.

7.2.5 VV and LFV Specific Files

These implement the specific integration scheme as file-names are finished with the flavour they develop if they have a counterpart implementing the same algorithm but in the alternative flavour. Names are self-explanatory.

7.2.6 SERIAL Specific Files

These implement an emulation of some general MPI calls used in DL_POLY_4 source code when compiling in serial mode as well as some modified counterparts of the general files changed to allow for faster and/or better memory optimised serial execution. Names are self-explanatory.

7.2.7 Comments on MPI Handling

Only a few files make explicit calls to MPI routines:
COMMS_MODULE IO_MODULE
READ_CONFIG_PARALLEL READ_CONFIG WRITE_CONFIG
VDW_TABLE_READ CHECK_CONFIG
SYSTEM_EXPAND SYSTEM_INIT
PASS_SHARED_UNITS UPDATE_SHARED_UNITS
EXPORT_ATOMIC_DATA READ_HISTORY DEPORT_ATOMIC_DATA
METAL_LD_EXPORT PARALLEL_FFT EXCHANGE_GRID
DEFECTS_REFERENCE_WRITE
DEFECTS_REFERENCE_READ_PARALLEL DEFECTS_REFERENCE_READ
DEFECTS_REFERENCE_EXPORT DEFECTS_WRITE DEFECTS1_WRITE
TRAJECTORY_WRITE MSD_WRITE RSD_WRITE SYSTEM_REVIVE.
The rest of the files that use MPI functionality in any way make implicit calls via generic functions developed in COMMS_MODULE.

7.2.8 Comments on SETUP_MODULE

The most important module, by far, is SETUP_MODULE, which holds the most important global parameters and variables (some of which serve as "parameters" for global array bounds, set in SET_BOUNDS). A brief account of these is given below:

parameter	**value**	**function**
`pi`	3.14159265358979312	π constant
`twopi`	6.28318530717958623	2π constant
`fourpi`	12.56637061435917246	4π constant
`sqrpi`	1.772453850905588	$\sqrt[2]{\pi}$ constant
`rtwopi`	0.15915494309189535	$\frac{1}{2\pi}$ constant
`rt2`	1.41421356237309515	$\sqrt[2]{2}$ constant
`rt3`	1.73205080756887719	$\sqrt[2]{3}$ constant
`r4pie0`	138935.4835	electrostatics conversion factor to internal units, i.e. $\frac{1}{4\pi\epsilon_o}$
`boltz`	0.831451115	Boltzmann constant in internal units
`prsunt`	0.163882576	conversion factor for pressure from internal units to katms
`nread`	5	main input channel
`nconf`	11	configuration file input channel
`nfield`	12	force field input channel
`ntable`	13	tabulated potentials file input channel
`nrefdt`	14	reference configuration input channel
`nrite`	6	main output channel
`nstats`	21	statistical data file output channel
`nrest`	22	output channel accumulators restart dump file
`nhist`	23	trajectory history file channel
`ndefdt`	24	output channel for defects data file
`nrdfdt`	25	output channel for RDF data
`nzdfdt`	26	output channel for Z-density data file
`nrsddt`	27	output channel for displacements data files
`npdfdt`	28	output channel for raw PDF files
`ngdfdt`	29	output channel for normalised RDF data files
`seed(1:3)`	*variable*	pair of seeds for the random number generator
`lseed`	*variable*	logical swich on/off indicator for seeding
`mxsite`	*variable*	max number of molecular sites
`mxatyp`	*variable*	max number of unique atomic types
`mxtmls`	*variable*	max number of unique molecule types
`mxexcl`	*variable*	max number of excluded interactions per atom
`mxspl`	*variable*	SPME FFT B-spline order
`mxspl1`	*variable*	SPME FFT B-spline possible extension when $r_{\text{pad}} > 0$
`kmaxa`	*variable*	SPME FFT amended array dimension (a direction)
`kmaxb`	*variable*	SPME FFT amended array dimension (b direction)
`kmaxc`	*variable*	SPME FFT amended array dimension (c direction)
`kmaxa1`	*variable*	SPME FFT original array dimension (a direction)
`kmaxb1`	*variable*	SPME FFT original array dimension (b direction)
`kmaxc1`	*variable*	SPME FFT original array dimension (c direction)
`mxtshl`	*variable*	max number of specified core-shell unit types in system
`mxshl`	*variable*	max number of core-shell units per node

`mxfshl`	*variable*	max number of related core-shell units (1+1)
`mxtcon`	*variable*	max number of specified bond constraints in system
`mxcons`	*variable*	max number of constraint bonds per a node
`mxfcon`	*variable*	max number of related constraint units (6+1)
`mxlshp`	*variable*	max number of shared particles per node $\mathtt{Max}(2\,\frac{\mathtt{mxshl}}{2}, 2\,\frac{\mathtt{mxcons}}{2}, \frac{\mathtt{mxlrgd} * \mathtt{mxrgd}}{2})$
`mxproc`	*variable*	number of neighbour nodes in DD hypercube (26)
`mxtpmf(1:2)`	*variable*	max number of specified particles in a PMF unit (1:2)
`mxpmf`	*variable*	max number of PMF constraints per a node
`mxfpmf`	*variable*	max number of related PMF units (1+1)
`mxtrgd`	*variable*	max number of types RB units
`mxrgd`	*variable*	max number of RB units per node
`mxlrgd`	*variable*	max number of constituent particles of an RB unit
`mxfrgd`	*variable*	max number of related RB units (1+1)
`mxtteth`	*variable*	max number of specified tethered potentials in system
`mxteth`	*variable*	max number of tethered atoms per node
`mxftet`	*variable*	max number of related tether units (1+1)
`mxpteth`	*variable*	max number of parameters for tethered potentials (3)
`mxtbnd`	*variable*	max number of specified chemical bond potentials in system
`mxbond`	*variable*	max number of chemical bonds per node
`mxfbnd`	*variable*	max number of related chemical bonds (1+(6*(6+1))/2)
`mxpbnd`	*variable*	max number of parameters for chemical bond potentials (4)
`mxgbnd`	*variable*	max number of grid points in chemical bond pot. arrays ($>$ 1004)
`mxtang`	*variable*	max number of specified bond angle potentials in system
`mxangl`	*variable*	max number of bond angles per node
`mxfang`	*variable*	max number of related bond angles (1+(6*(6+1))/2)
`mxpang`	*variable*	max number of parameters for bond angle potentials (6)
`mxgang`	*variable*	max number of grid points in bond angle pot. arrays ($>$ 1004)
`mxtdih`	*variable*	max number of specified dihedral angle potentials in system
`mxdihd`	*variable*	max number of dihedral angles per node
`mxfdih`	*variable*	max number of related dihedral angles (1+((6-2)6*(6+1))/2)
`mxpdih`	*variable*	max number of parameters for dihedral angle potentials (7)
`mxgdih`	*variable*	max number of grid points in dihedral angle pot. arrays ($>$ 1004)
`mxtinv`	*variable*	max number of specified inversion angle potentials in system
`mxinv`	*variable*	max number of inversion angles per node
`mxfinv`	*variable*	max number of related inversion angles (1+(6*(6+1))/4)
`mxpinv`	*variable*	max number of parameters for inversion angle potentials (3)
`mxginv`	*variable*	max number of grid points in inversion angle pot. arrays ($>$ 1004)
`mxrdf`	*variable*	max number of pairwise RDF in system
`mxgrdf`	*variable*	number of grid points for RDF and Z-density arrays ($>$ 1004)
`mxgele`	*variable*	max number of grid points for ewald exclusion potential arrays
`mxvdw`	*variable*	max number of van der Waals potentials in system
`mxpvdw`	*variable*	max number of van der Waals potential parameters (5)
`mxgvdw`	*variable*	max number of grid points in vdw potential arrays ($>$ 1004)
`mxmet`	*variable*	max number of metal potentials in system
`mxmed`	*variable*	max number of metal density potentials in system
`mxmds`	*variable*	max number of metal extra density potentials in system
`mxpmet`	*variable*	max number of metal potential parameters (9)
`mxgmet`	*variable*	max number of grid points in metal potential arrays ($>$ 1004)
`mxter`	*variable*	max number of Tersoff potentials in system
`mxpter`	*variable*	max number of Tersoff potential parameters (11)
`mxgter`	*variable*	max number of grid points in tersoff potential arrays ($>$ 1004)

`mxgrid`	*variable*	max number of grid points in potential arrays (> 1004)
`mxtana`	*variable*	max number of PDFs per type
`mxgana`	*variable*	max number of grid points for PDFs arrays
`mxgbnd`	*variable*	max number of grid points for chemical bonds PDFs
`mxgang`	*variable*	max number of grid points for bond angles PDFs
`mxgdih`	*variable*	max number of grid points for dihedral angles PDFs
`mxginv`	*variable*	max number of grid points for inversion angles PDFs
`mxtbp`	*variable*	max number of three-body potentials in system
`mx2tbp`	*variable*	array dimension of three-body potential parameters
`mxptbp`	*variable*	max number of three-body potential parameters (5)
`mxfbp`	*variable*	max number of four-body potentials in system
`mx2fbp`	*variable*	array dimension of four-body potential parameters
`mxpfbp`	*variable*	max number of four-body potential parameters (3)
`mxpfld`	*variable*	max number of external field parameters (5)
`mxstak`	*variable*	dimension of stack arrays for rolling averages
`mxnstk`	*variable*	max number of stacked variables
`mxlist`	*variable*	max number of atoms in the Verlet list on a node
`mxcell`	*variable*	max number of link cells per node
`mxatms`	*variable*	max number of local+halo atoms per node
`mxatdm`	*variable*	max number of local atoms per node
`mxbfdp`	*variable*	max dimension of the transfer buffer for deport functions
`mxbfss`	*variable*	max dimension of the transfer buffer for statistics functions
`mxbfxp`	*variable*	max dimension of the transfer buffer for export functions
`mxbfsh`	*variable*	max dimension of the transfer buffer for shared units functions
`mxbuff`	*variable*	max dimension of the principle transfer buffer
`zero_plus`	*variable*	the machine representation of $+0$ at working precision
`half_plus`	*variable*	the machine representation of $+0.5 \uparrow$ at working precision
`half_minus`	*variable*	the machine representation of $+0.5 \downarrow$ at working precision
`engunit`	*variable*	the system energy unit

Chapter 8

Examples

Scope of Chapter

This chapter describes the standard test cases for DL_POLY_4, the input and output files for which are in the *data* sub-directory.

8.1 Test Cases

Because of the size of the data files for the DL_POLY_4 standard test cases, they are not shipped in the standard download of the DL_POLY_4 source. Instead users are requested to download them from the CCP5 FTP server as follows:

```
FTP site : ftp.dl.ac.uk
Username : anonymous
Password : your email address
Directory: ccp5/DL_POLY/DL_POLY_4.0/DATA
Files    : test_X.tar.gz
```

where '_X' stands for the test case number.

Remember to use the BINARY data option when transferring these files.

Unpack the files in the 'data' subdirectory using firstly 'gunzip' to uncompress them and then 'tar -xf' to create the 'TEST_X' directory.

These are provided so that you may check that your version of DL_POLY_4 is working correctly. All the jobs are of a size suitable to test the code in parallel execution. They not not be suitable for a single processor computer. The files are stored in compressed format. The test cases can be run by typing

select n

from the *execute* directory, where n is the number of the test case. The *select* macro will copy the appropriate CONTROL, CONFIG, and FIELD files to the *execute* directory ready for execution. The output file OUTPUT may be compared with the file supplied in the *data* directory.

It should be noted that the potentials and the simulation conditions used in the following test cases are chosen to demonstrate a limited set of relevant functionality over a limited extent of molecular systems' complexity only. **They are not necessarily appropriate for serious simulation of the test systems.** In other words, the tests are not warranted to have well defined force field in terms of applicability, transferability and fullness as well as to have a well defined state point - thus initial configurations may be away from equilibrium, if physical at all.

8.1.1 Test Case 1 and 2: Sodium Chloride

These are a 27,000 and 216,000 ion systems respectively with unit electric charges on sodium and chlorine. Simulation at 500 K with a NVT Berendsen ensemble. The SPME method is used to calculate the Coulombic interactions.

8.1.2 Test Case 3 and 4: DPMC in Water

These systems consist of 200 and 1,600 DMPC molecules in 9379 and 75032 water molecules respectively. Simulation at 300 K using NVE ensemble with SPME and RATTLE algorithm for the constrained motion. Total system size is 51737 and 413896 atoms respectively.

8.1.3 Test Case 5 and 6: $KNaSi_2O_5$

Potassium Sodium disilicate glass ($NaKSi_2O_5$) using two and three-body potentials. Some of the two-body potentials are read from the TABLE file. Simulation at 1000 K using NVT Nosé-Hoover ensemble with SPME. Cubic periodic boundaries are in use. System size is 69120 and 552960 ions respectively.

8.1.4 Test Case 7 and 8: Gramicidin A molecules in Water

These systems consist of 8 and 16 gramicidin A molecules in aqueous solution (32,096 and 256,768 water molecules) with total number of atoms 99,120 and 792,960 respectively. Simulation at 300 K using NPT Berendsen ensemble with SPME and SHAKE/RATTLE algorithm for the constrained motion.

8.1.5 Test Case 9 and 10: SiC with Tersoff Potentials

These systems consist of 74,088 and 343,000 atoms respectively. Simulation at 300 K using NPT Nosé-Hoover ensemble with Tersoff forces and no electrostatics.

8.1.6 Test Case 11 and 12: Cu_3Au alloy with Sutton-Chen (metal) Potentials

These systems consist of 32,000 and 256,000 atoms respectively. Simulation at 300 K using NVT Nosé-Hoover ensemble with Sutton-Chen forces and no electrostatics.

8.1.7 Test Case 13 and 14: lipid bilayer in water

These systems consist of 12,428 and 111,852 atoms respectively. Simulation at 300 K using NVT Berendsen ensemble with SPME and SHAKE/RATTLE algorithm for the constrained motion.

8.1.8 Test Case 15 and 16: relaxed and adiabatic shell model MgO

These systems consist of 8,000 (4,000 shells) and 64,000 (32,000 shells) atoms respectively. Simulation at 3000 K using NPT Berendsen ensemble with SPME. FIELD and CONTROL files for each shell model are provided separately.

8.1.9 Test Case 17 and 18: Potential of mean force on K+ in water MgO

These systems consist of 13,500 (500 PMFs) and 53,248 (2,048 PMFs) atoms respectively. Simulation at 300 K using NPT Berendsen ensemble with SPME and SHAKE/RATTLE algorithm for the constrained motion.

8.1.10 Test Case 19 and 20: Cu_3Au alloy with Gupta (metal) Potentials

These systems consist of 32,000 and 256,000 atoms respectively. Simulation at 300 K using NVT Nosé-Hoover ensemble with Gupta forces and no electrostatics.

8.1.11 Test Case 21 and 22: Cu with EAM (metal) Potentials

These systems consist of 32,000 and 256,000 atoms respectively. Simulation at 300 K using NPT Berendsen ensemble with EAM tabulated forces and no electrostatics.

8.1.12 Test Case 23 and 24: Al with Sutton-Chen (metal) Potentials

These systems consist of 32,000 and 256,000 atoms respectively. Simulation at 300 K using NVT Evans ensemble with Sutton-Chen forces and no electrostatics.

8.1.13 Test Case 25 and 26: Al with EAM (metal) Potentials

These systems consist of 32,000 and 256,000 atoms respectively. Simulation at 300 K using NVT Evans ensemble with EAM tabulated forces and no electrostatics.

8.1.14 Test Case 27 and 28: NiAl alloy with EAM (metal) Potentials

These systems consist of 27,648 and 221,184 atoms respectively. Simulation at 300 K using NVT Evans ensemble with EAM tabulated forces and no electrostatics.

8.1.15 Test Case 29 and 30: Fe with Finnis-Sincair (metal) Potentials

These systems consist of 31,250 and 250,000 atoms respectively. Simulation at 300 K using NPT Berendsen ensemble with Finnis-Sinclair forces and no electrostatics.

8.1.16 Test Case 31 and 32: Ni with EAM (metal) Potentials

These systems consist of 32,000 and 256,000 atoms respectively. Simulation at 300 K using NPT Berendsen ensemble with EAM tabulated forces and no electrostatics.

8.1.17 Test Case 33 and 34: SPC IceVII water with constraints

These systems consist of 11,664 (34,992 atoms) and 93,312 (279,936 atoms) water molecules respectively. Simulation at 25 K using NVE ensemble with CGM force minimisation and SPME electrostatics. Both constraint bond and rigid body dynamics cases are available.

8.1.18 Test Case 35 and 36: NaCl molecules in SPC water represented as CBs+RBs

These systems consist of 64 (512) NaCl ion pairs with 4,480 (35,840) water molecules represented by constraint bonds and 4,416 (35,328) water molecules represented by ridig bodies. Totalling 26,816 (214,528) atoms. Simulation at 295 K using NPT Berendsen ensemble with CGM energy minimisation and SPME electrostatics.

8.1.19 Test Case 37 and 38: TIP4P water: RBs with a massless charged site

These systems consist of 7,263 and 58,104 TIP4P rigid body water molecules totaling 29,052 and 232,416 particles respectively. Simulation at 295 K using NPT Berendsen ensemble with CGM energy minimisation and SPME electrostatics.

8.1.20 Test Case 39 and 40: Ionic liquid dimethylimidazolium chloride

These systems consist of 44,352 and 354,816 ions respectively. Simulation at 400 K using NPT Berendsen ensemble, using both particle and rigid body dynamics with SPME electrostatics.

8.1.21 Test Case 41 and 42: Calcite nano-particles in TIP3P water

In this case 600 and 4,800 molecules of calcium carbonate in the calcite structure form 8 and 64 nano-particles which are suspended in 6,904 and 55,232 water molecules represented by a flexible 3-centre TIP3P model. Simulation with SPME electrostatics at 310 K and 1 atmosphere maintained in a Hoover NPT ensemble. These systems consist of 23,712 and 189,696 ions respectively.

8.1.22 Test Case 43 and 44: Iron/Carbon alloy with EEAM

In this case a steel alloy of iron and carbon in ratio 35132 to 1651 is modelled using an EEAM potential forcefield. Simulation at 1000 K and 0 atmosphere is maintained in a Berendsen NPT ensemble. These systems consist of 36,803 and 294,424 particles respectively.

8.1.23 Test Case 45 and 46: Iron/Cromium alloy with 2BEAM

In this case a steel alloy of iron and chromium in ratio 27635 to 4365 is modelled using an 2BEAM potential forcefield. Simulation at 300 K and 0 atmosphere is maintained in an Evans NVT isokinetic ensemble. These systems consist of 32,000 and 256,000 particles respectively.

8.1.24 Test Case 47 and 48: Hexane and methanol melts, full atomistic and coarse-grained simulations

TEST47 and TEST48 contain a Hexane and a Methanol melts respectively, (1000 molecules each) modelled by the OPLSAA force-field (FF). Each system is also supplied in a CG-mapped representation as converted by VOTCA, http://www.votca.org/, or DL_CGMAP http://www.ccp5.ac.uk/projects/ccp5_cg.shtml.

These test cases are to exemplify the Coarse-Graining (CG) procedure (see Chapter 4), including FA-to-CG mapping and obtaining the PMF data by means of Boltzmann Inversion [90]. As a result, DL_POLY_4 could be used for simulating a CG system with numerically defined, tabulated FFs, see TABBND, TABANG, TABDIH and TABINV files for intra-molecular potentials, and TABLE for inter-molecular (short-range, VDW) potentials.

Both tests are also available as parts of the tutorial cases from the VOTCA package [91]. Therefore, the CONFIG, CONTROL and FIELD input files are fully consistent with the corresponding setup files found in the VOTCA tutorial directories “csg-tutorials/hexane” and “csg-tutorials/methanol’.

8.2 Benchmark Cases

DL_POLY_4 benchmark test cases are available to download them from the CCP5 FTP server as follows:

```
FTP site : ftp.dl.ac.uk
Username : anonymous
Password : your email address
Directory: ccp5/DL_POLY/DL_POLY_4.0/BENCH
```

The DL_POLY_4 authors provide these on an ”AS IS” terms. For more information refer to the README.txt file within.

Appendix A

DL_POLY_4 Dissipative Particle Dynamics

A.1 Introduction

Although in a Molecular Dynamics sense Dissipative Particle Dynamics (DPD) is regarded as a type of thermostat, on its own it is an off-lattice, discrete particle method for modelling mesoscopic systems in a fluid state. It has little in common with Lattice Gas Automata and Lattice Boltzmann dynamics methods [**?**], except in its application to systems of similar length and time scales, and last but not least that it captures hydrodynamics behaviour.

The DPD method inherits its methodology from Brownian/Langevin Dynamics (BD). However, it differs from BD in an important way: it is *Galilean invariant* and for this reason conserves hydrodynamic behaviour, while the BD method does not (its microscopic behaviour is only diffusive. Many systems in their fluid state are crucially dependent on hydrodynamic interactions and it is essential to retain this feature in their models. DPD is particularly useful for simulating coarse-grained systems on the near-molecular scale, such as polymers, biopolymers, lipids, emulsions and surfactants – systems in which large scale structure evolves on a time scale that is too long to be modelled effectively by traditional MD. The particles in DPD [79] are not regarded as molecules in a fluid but as lumps of molecules grouped to form a *fluid particle* in much the same spirit as the renormalisation group has been applied in polymer physics where lumps of monomers are grouped to form a *bead*. Hence, the beads are regarded as carriers of momentum.

It is worth noting that DPD may also be used when such systems experience shear and flow gradients.

A.2 Outline of Method

Following [50] the DPD algorithm can be summarised by the following:

- A condensed phase system may be modelled as a system of 'free' particles interacting directly through *soft* forces. Note that DL_POLY_4 allows for the application of the DPD thermostat beyond systems of free particles only. Thus it will be valid on systems with any inratamolecular like interactions.
- The system is coupled to a heat bath via stochastic forces, which act on the particles in a pairwise manner.
- The particles also experience a damping or drag force, which also acts in a pairwise manner.
- Thermodynamic equilibrium is maintained through the balance of the stochastic and drag forces, i.e. the method satisfies the fluctuation-dissipation theorem.

- At equilibrium (or steady state) the properties of the system are calculated as averages over the individual particles, as in traditional Molecular Dynamics.

Therefore, the equation of motion are the same as these for the microcanonical ensemble (NVE) but force, f_i, on particle i is now a sum of pair forces:

$$\underline{f}_i = \sum_{j \neq i}^{N} \left(\underline{f}_{ij}^C + \underline{f}_{ij}^D + \underline{f}_{ij}^R \right) \quad , \tag{A.1}$$

in which $\underline{f}_{ij}^C$, $\underline{f}_{ij}^D$ and $\underline{f}_{ij}^R$ are the *conservative*, *drag* and *random* (or *stochastic*) pair forces respectively. Each represents the force exerted on particle i due to the presence of particle j.

The conservative interactions are usually *soft* (i.e. weakly interacting) so that the particles can pass by each other (or even through each other) relatively easily so that equilibrium is achieved quickly. A common form of interaction potential is an inverse parabola (part of VDW types of potentials, see Section 2.3.1)

$$V(r_{ij}) = \begin{cases} \frac{A_{ij}}{2}\ r_c\ \left(1 - \frac{r_{ij}}{r_c}\right)^2 & : \quad r_{ij} < r_c \\ 0 & : \quad r_{ij} \geq r_c \end{cases} \quad , \tag{A.2}$$

where $r_{ij} = |\underline{r}_j - \underline{r}_i|$, r_c is a cutoff radius and A_{ij} is the interaction strength (that may be the same for all particle pairs or may be different for different particle types).

Equation (A.1) gives rise to a repulsive force of the form:

$$\underline{f}_{ij}^C = A_{ij}\ w^C(r_{ij}) \frac{\underline{r}_{ij}}{r_{ij}} = A_{ij} \left(1 - \frac{r_{ij}}{r_c}\right) \frac{\underline{r}_{ij}}{r_{ij}} \quad . \tag{A.3}$$

This is the deterministic or *conservative* force $\underline{f}_{ij}^C$ exerted on particle i by particle j. Note the switching function:

$$w^C(r_{ij}) = \begin{cases} \left(1 - \frac{r_{ij}}{r_c}\right) & : \quad r_{ij} < r_c \\ 0 & : \quad r_{ij} \geq r_c \end{cases} \quad , \tag{A.4}$$

and the force are zero when $r_{ij} \geq r_c$ and thus the particles have an effective diameter of 1 in units of the cutoff radius r_c. In the DL_POLY_4 context all inter- and intra-molecular forces will fall into this category of force!

The stochastic forces experienced by the particles is again pairwise in nature and takes the form:

$$\underline{f}_{ij}^R = \sigma_{ij} w^R(r_{ij}) \zeta_{ij} \Delta t^{-\frac{1}{2}} \frac{\underline{r}_{ij}}{r_{ij}} \quad , \tag{A.5}$$

in which Δt is the time step and $w^R(r_{ij})$ is a switching function which imposes a finite limit on the range of the stochastic force. ζ_{ij} is a random number with zero mean and unit variance. The constant σ_{ij} is related to the temperature, as is understood from the role of the stochastic force in representing a heat bath.

Finally, the particles are subject to a drag force, which depends on the relative velocity between interacting pairs of particles:

$$\underline{f}_{ij}^D = -\gamma_{ij} w^D(r_{ij}) \left(\underline{r}_{ij} \cdot \underline{v}_{ij}\right) \frac{\underline{r}_{ij}}{r_{ij}^2} \quad , \tag{A.6}$$

where $w^D(r_{ij})$ is once again a switching function and $\underline{v}_{ij} = \underline{v}_j - \underline{v}_i$ is the inter-particle relative velocity. The constant γ_{ij} is the drag coefficient. It follows from the fluctuation-dissipation theorem that for thermodynamic equilibrium to result from this method the following relations must hold:

$$\sigma_{ij}^2 = 2\ \gamma_{ij} k_B T \tag{A.7}$$

$$w^D(r_{ij}) = \left[w^R(r_{ij})\right]^2 \quad . \tag{A.8}$$

In practice, the switching functions are defined through:

$$w^R(r_{ij}) = \left[w^C(r_{ij})\right]^2 \quad , \tag{A.9}$$

which ensures that all interactions are switched off at the range $r_{ij} = r_c$.

In many DPD simulations, the stochastic and drag coefficients are often constant for all interactions, i.e. $\sigma_{ij} \equiv \sigma$ and $\gamma_{ij} \equiv \gamma$, although this assumption does not have to apply. In DL_POLY_4 the γ_{ij} coefficients may be supplied at the end of each specified vdw interaction potential as a parameter further to the last one for the particular vdw potential form. For a DPD thermostat to work correctly all possible two body interactions must be defined and all $\gamma_{ij} \neq 0$. What DL_POLY_4 will attempt first, if a two body interaction is missing, is to derive it using mixing rules (default may be overridden by user specification). However, if any of $\gamma_{ij} = 0$ then DL_POLY_4 will check for the existence of a global γ that may be optionally supplied by the user on the **ensemble nvt dpd*INT*** line and if it is non-zero a global override will occur. Otherwise, when the requirements for a DPD thermostat are not satisfied, everything else will result in a controlled termination.

A.3 Equation of state and dynamic properties

The form of the conservative force determines the equation of state for a DPD fluid, which can be derived using the virial theorem to express system pressure as follows:

$$\begin{aligned} \mathcal{P} &= \rho k_B T + \frac{1}{3V}\left\langle \sum_{j>i} (\underline{r}_i - \underline{r}_j)\cdot \underline{f}_{ij}^C \right\rangle && \text{(A.10)} \\ &= \rho k_B T + \frac{2\pi}{3}\rho^2 \int_0^{r_c} A\left(1-\frac{r}{r_c}\right) r^3 g(r)\; dr \quad , && \text{(A.11)} \end{aligned}$$

where $g(r)$ is a radial distribution function for the soft sphere model [50] and ρ is the DPD particle density. For sufficiently large densities ($\rho > 2$), $g(r)$ takes the same form and the equation of state can be well-approximated by:

$$\mathcal{P} = \rho k_B T + \alpha A \rho^2 \quad , \tag{A.12}$$

where the parameter $\alpha \approx 0.101 \pm 0.001$ has units equivalent to r_c^4. This expression permits the use of fluid compressibilities to obtain conservative force parameters for bulk fluids, e.g. for water $A \approx 75 k_B T/\rho$. Alternative equations of state may be obtained by modifying the functional form of conservative interactions to include localized densities (i.e. many-body DPD) [92, 93].

Transport coefficients for a DPD fluid can be derived using the expressions for the drag and stochastic forces[50, 94, 95]. The kinematic viscosity can be found to be

$$\nu \approx \frac{45 k_B T}{4\pi\gamma\rho r_c^3} + \frac{2\pi\gamma\rho r_c^5}{1575} \quad , \tag{A.13}$$

while the self-diffusion coefficient is given as

$$D \approx \frac{45 k_B T}{2\pi\gamma\rho r_c^3}. \tag{A.14}$$

The ratio of these two properties, the Schmidt number ($\mathrm{Sc} = \nu/D$), is therefore:

$$\mathrm{Sc} \approx \frac{1}{2} + \frac{(2\pi\gamma\rho r_c^4)^2}{70875 k_B T} \tag{A.15}$$

and for values of the drag coefficient and density frequently used in DPD simulations, this value is of the order of unity, which is an appropriate magnitude for gases but three orders of magnitude too small for liquids.

This property of standard DPD does *not* rule it out for simulations of liquid phases except when hydrodynamics are important. It may also be argued that the self-diffusion of DPD particles might not correspond to that of individual molecules and thus a Schmidt number of the order 10^3 is unnecessary for modelling liquids [96]. Alternative thermostats are available in the DL_MESO [97] - http://www.ccp5.ac.uk/DL_MESO/ package, which can model systems with higher Schmidt numbers [98, 99].

A.4 Derivation of Equilibrium

The derivation of the DPD algorithm is based on the Fokker-Planck equation

$$\frac{\partial \rho}{\partial t} = \mathcal{L}\rho \tag{A.16}$$

where ρ is the equilibrium distribution function and $\mathcal{L}$ is the evolution operator, which may be split into *conservative* and *stochastic+dissipative* parts:

$$\mathcal{L} = \mathcal{L}^C + \mathcal{L}^{R+D} \tag{A.17}$$

with

$$\mathcal{L}^C = -\sum_{i=1}^{N} \frac{\underline{p}_i}{m_i}\frac{\partial}{\partial \underline{r}_i} - \sum_{i \neq j}^{N} \underline{f}_{ij}^C \frac{\partial}{\partial \underline{p}_i} \tag{A.18}$$

$$\mathcal{L}^{R+D} = \sum_{i=1}^{N} \hat{e}_{ij} \cdot \frac{\partial}{\partial \underline{p}_i}\left[\frac{\sigma^2}{2}\left\{w^R(r_{ij})\right\}^2 \hat{e}_{ij} \cdot \left\{\frac{\partial}{\partial \underline{p}_i} - \frac{\partial}{\partial \underline{p}_j}\right\} + \gamma w^D \left(\hat{e}_{ij} \cdot \underline{v}_{ij}\right)\right] \quad , \tag{A.19}$$

where $\hat{e}_{ij} = \frac{r_{ij}}{\underline{\underline{r_{ij}}}}$.

When $\sigma = \gamma = 0$ then equation (A.16) becomes

$$\frac{\partial \rho}{\partial t} = \mathcal{L}^C \rho \quad , \tag{A.20}$$

for which the equilibrium solution is evidently

$$\rho^{eq} = \frac{1}{Z}\exp\left(\frac{1}{k_B T}\left[\sum_{i=1}^{N}\frac{p_i^2}{2m_i} + \frac{1}{2}\sum_{j \neq i}^{N}\phi(r_{ij})\right]\right) \tag{A.21}$$

which is, of course, the Boltzmann distribution function for an equilibrium system. Thus it is apparent that for the simulation based on equation (A.16) to maintain the same distribution function, the terms in the operator $\mathcal{L}^{R+D}$ of equation (A.19) must sum to zero. It follows that the conditions given in equations (A.7) and (A.8) must apply.

A.5 Summary of Dissipative Particle Dynamics

DPD is a simple method that can be viewed as a novel thermostatting method for molecular dynamics. All that is required is a system of spherical particles enclosed in a periodic box undergoing time evolution as a result of the above forces. It should be noted that all computed interactions are pairwise, which means that the principle of the conservation of momentum in the system, or *Galilean invariance*, is preserved. The conservation of momentum is required for the preservation of hydrodynamic forces. Therefore, the DPD method is an NVT method that *preserves hydrodynamics*. The presence of hydrodynamics is important in annealing defects in ordered mesophases [100]. Thus DPD has an intrinsic advantage over other methods such as traditional molecular dynamics, dynamic density functional theory (which are purely *diffusive*!) or Monte Carlo methods, in trying to evolve a system towards an ordered thermodynamic equilibrium state.

Appendix B

DL_POLY_4 Periodic Boundary Conditions

B.1 Introduction

DL_POLY_4 is designed to accommodate a number of different periodic boundary conditions, which are defined by the shape and size of the simulation cell. Briefly, these are as follows (which also indicates the IMCON flag defining the simulation cell type in the CONFIG file - see Section 6.1.2):

1.	None e.g. isolated polymer in space	(`imcon` = 0)
2.	Cubic periodic boundaries	(`imcon` = 1)
3.	Orthorhombic periodic boundaries	(`imcon` = 2)
4.	Parallelepiped periodic boundaries	(`imcon` = 3)
5.	Slab (X,Y periodic; Z non-periodic)	(`imcon` = 6)

We shall now look at each of these in more detail. Note that in all cases the cell vectors and the positions of the atoms in the cell are to be specified in Angstroms (Å).

B.2 No periodic boundary (`imcon` = 0)

Simulations requiring no periodic boundaries are best suited to *in vacuuo* simulations, such as the conformational study of an isolated polymer molecule. This boundary condition is not recommended for studies in a solvent, since evaporation is likely to be a problem.

Note this boundary condition have to be used with caution. DL_POLY_4 is not naturally suited to carry out efficient calculations on systems with great fluctuation of the local density in space, as is the case for clusters in vacuum. The parallelisation and domain decomposition is therefore limited to eight domains (maximum of two in each direction in space).

This boundary condition should not used with the SPM Ewald summation method.

B.3 Cubic periodic boundaries (`imcon` = 1)

The cubic MD cell is perhaps the most commonly used in simulation and has the advantage of great simplicity. In DL_POLY_4 the cell is defined with the principle axes passing through the centres of the faces. Thus for a cube with sidelength D, the cell vectors appearing in the CONFIG file should be: (D,0,0); (0,D,0); (0,0,D). Note the origin of the atomic coordinates is the centre of the cell.

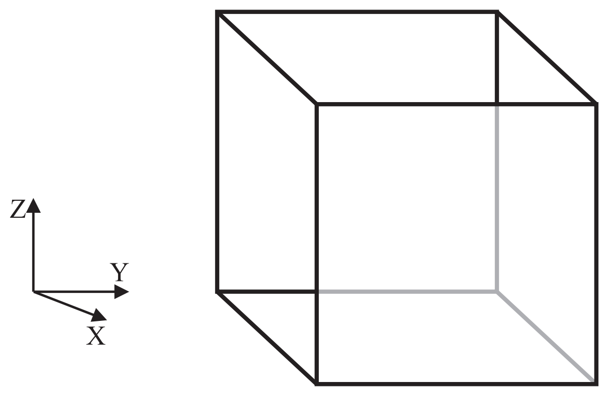

Figure B.1: The cubic MD cell

B.4 Orthorhombic periodic boundaries (`imcon` = 2)

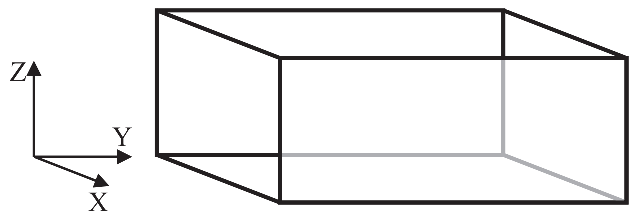

Figure B.2: The orthorhomic MD cell

The orthorhombic cell is also a common periodic boundary, which closely resembles the cubic cell in use. In DL_POLY_4 the cell is defined with principle axes passing through the centres of the faces. For an orthorhombic cell with sidelengths D (in X-direction), E (in Y-direction) and F (in Z-direction), the cell vectors appearing in the CONFIG file should be: (D,0,0); (0,E,0); (0,0,F). Note the origin of the atomic coordinates is the centre of the cell.

B.5 Parallelepiped periodic boundaries (`imcon` = 3)

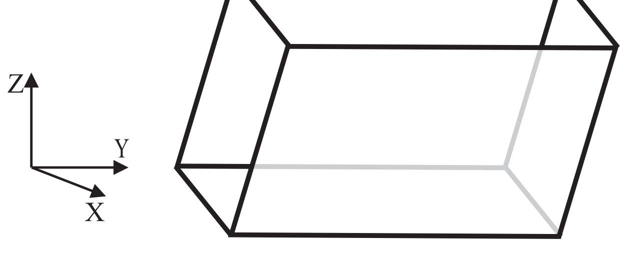

Figure B.3: The parallelepiped MD cell

The parallelepiped (e.g. monoclinic or triclinic) cell is generally used in simulations of crystalline materials, where its shape and dimension is commensurate with the unit cell of the crystal. Thus for a unit cell specified by three principal vectors $\underline{a}$, $\underline{b}$, $\underline{c}$, the MD cell is defined in the DL_POLY_4 CONFIG file by the vectors (La_1,La_2,La_3), (Mb_1,Mb_2,Mb_3), (Nc_1,Nc_2,Nc_3), in which L,M,N are integers, reflecting the multiplication of the unit cell in each principal direction. Note that the atomic coordinate origin is the centre of the MD cell.

B.6 Slab boundary conditions (`imcon` = 6)

Slab boundaries are periodic in the X- and Y-directions, but not in the Z-direction. They are particularly useful for simulating surfaces. The periodic cell in the XY plane can be any parallelogram. The origin of the X,Y atomic coordinates lies on an axis perpendicular to the centre of the parallelogram. The origin of the Z coordinate is where the user specifies it. However, it is recommended that it is in the middle of the slab. Domain decomposition division across Z axis is limited to 2.

If the XY parallelogram is defined by vectors $\underline{A}$ and $\underline{B}$, the vectors required in the CONFIG file are: (A_1,A_2,0), (B_1,B_2,0), (0,0,D), where D is any real number (including zero). If D is nonzero, it will be used by DL_POLY to help determine a 'working volume' for the system. This is needed to help calculate RDFs etc. (The working value of D is in fact taken as one of: 3×cutoff; or 2×max abs(Z coordinate)+cutoff; or the user specified D, whichever is the larger.)

The surface in a system with charges can also be modelled with DL_POLY_4 if periodicity is allowed in the Z-direction. In this case slabs of ions well-separated by vacuum zones in the Z-direction can be handled with `imcon` = 1, 2 or 3.

Appendix C

DL_POLY_4 Macros

Introduction

Macros are simple executable files containing standard UNIX commands. A number of the are supplied with DL_POLY_4 and are found in the *execute* sub-directory. These are not guaranteed to be immaculate but with little adaptation they can become a useful tool to a researcher. The available macros are as follows:

- *cleanup*
- *copy*
- *gopoly*
- *gui*
- *select*
- *store*

The function of each of these is described below. It is worth noting that most of these functions could be performed by the DL_POLY Java GUI [21].

cleanup

cleanup removes several standard data files from the *execute* sub-directory. It contains the UNIX commands:

```
rm OUTPUT STATIS REVCON REVOLD REVIVE RDFDAT ZDNDAT DEFECTS gopoly.*
```

and removes the files OUTPUT, REVCON, REVOLD, STATIS, REVIVE, DEFECTS and gopoly.* (all variants). It is useful for cleaning the sub-directory up after a run. (Useful data should be stored elsewhere however!)

copy

copy invokes the UNIX commands:

```
mv -v CONFIG CONFIG.OLD
mv -v REVCON CONFIG
mv -v REVIVE REVOLD
```

which collectively prepare the DL_POLY_4 files in the *execute* sub-directory for the continuation of a simulation. It is always a good idea to store these files elsewhere in addition to using this macro.

gopoly

gopoly is used to submit a DL_POLY_4 job to the HPCx, which operates a LOAD-LEVELER job queuing system. It invokes the following script:

```
#@ shell = /usr/bin/tcsh
#
#@ job_type = parallel
#@ job_name = gopoly
#
#@ cpus = 32
#
#@ node_usage = not_shared
#@ network.MPI = csss,shared,US
#
#@ wall_clock_limit = 00:30:00
#@ account_no = my_account
#
#@ output = $(job_name).$(schedd_host).$(jobid).out
#@ error  = $(job_name).$(schedd_host).$(jobid).err
#@ notification = never
#
#@ bulkxfer = yes
#@ data_limit = 850000000
#@ stack_limit = 10000000
#
#@ queue
#
# ENVIRONMENT SETTINGS
#
setenv MP_EAGER_LIMIT 65536
setenv MP_SHARED_MEMORY yes
setenv MEMORY_AFFINITY MCM
setenv MP_TASK_AFFINITY MCM
setenv MP_SINGLE_THREAD yes
#
poe  ./DLPOLY.Z
```

Using LOADLEVELLER, the job is submitted by the UNIX command:

llsubmit gopoly

where *llsubmit* is a local command for submission to the IBM SP4 cluster. The number of required nodes and the job time are indicated in the above script.

gui

gui is a macro that starts up the DL_POLY_4 Java GUI. It invokes the following UNIX commands:

```
java -jar ../java/GUI.jar $1 &
```

In other words the macro invokes the Java Virtual Machine which executes the instructions in the Java archive file GUI.jar, which is stored in the *java* subdirectory of DL_POLY_4. (Note: Java 1.3.0 or a higher version is required to run the GUI.)

select

select is a macro enabling easy selection of one of the test cases. It invokes the UNIX commands:

```
cp -vpLH ../data/TEST$1/CONTROL   .
cp -vpLH ../data/TEST$1/CONFIG    .
cp -vpLH ../data/TEST$1/HISTORY   .
cp -vpLH ../data/TEST$1/FIELD     .
cp -vpLH ../data/TEST$1/TAB*      .
cp -vpLH ../data/TEST$1/REFERENCE .
```

select requires one argument (an integer) to be specified:

select n

where *n* is test case number, which ranges from 1 to 18.

This macro sets up the required input files in the *execute* sub-directory to run the *n*-th test case. The last three copy commands may not be necessary in most cases.

store

The *store* macro provides a convenient way of moving data back from the *execute* sub-directory to the *data* sub-directory. It invokes the UNIX commands:

```
mkdir -pv          ../data/TEST$1
cp -vpLH CONTROL   ../data/TEST$1
cp -vpLH CONFIG    ../data/TEST$1
cp -vpLH FIELD     ../data/TEST$1
cp -vpLH TAB*      ../data/TEST$1
cp -vpLH REFERENCE ../data/TEST$1
cp -vpLH HISTORY   ../data/TEST$1
mv -v     REVCON   ../data/TEST$1
mv -v     CFGMIN   ../data/TEST$1
mv -v     OUTPUT   ../data/TEST$1
mv -v     HISTORF  ../data/TEST$1
mv -v     DEFECTS  ../data/TEST$1
mv -v     STATIS   ../data/TEST$1
mv -v     *DAT*    ../data/TEST$1
mv -v     *PMF     ../data/TEST$1
mv -v     *TAB     ../data/TEST$1
mv -v     REVIVE   ../data/TEST$1
chmod -R a-w       ../data/TEST$1
```

which first creates a new DL_POLY *data/TEST..* sub-directory and then moves the standard DL_POLY_4 output data files into it.

store requires one argument:

store n

where *n* is a unique string or number to label the output data in the *data/TESTn* sub-directory.

Note that *store* sets the file access to read-only. This is to prevent the *store* macro overwriting existing data without your knowledge.

Appendix D

DL_POLY_4 Error Messages and User Action

Introduction

In this appendix we document the error messages encoded in DL_POLY_4 and the recommended user action. The correct response is described as the **standard user response** in the appropriate sections below, to which the user should refer before acting on the error encountered.

The reader should also be aware that some of the error messages listed below may be either disabled in, or absent from, the public version of DL_POLY_4. Note that the wording of some of the messages may have changed over time, usually to provide more specific information. The most recent wording appears below.

The Standard User Response

DL_POLY_4 uses FORTRAN90 dynamic array allocation to set the array sizes at run time. This means that a single executable may be compiled to over all the likely uses of the code. It is not foolproof however. Sometimes an estimate of the required array sizes is difficult to obtain and the calculated value may be too small. For this reason DL_POLY_4 retains array dimension checks and will terminate when an array bound error occurs.

When a dimension error occurs, the **standard user response** is to edit the DL_POLY_4 subroutine SET_BOUNDS. Locate where the variable defining the array dimension is fixed and increase accordingly. To do this you should make use of the dimension information that DL_POLY_4 prints in the OUTPUT file prior to termination. If no information is supplied, simply doubling the size of the variable will usually do the trick. If the variable concerned is defined in one of the support subroutines SCAN_CONFIG, SCAN_FIELD, SCAN_CONTROL you will need to insert a new line in SET_BOUNDS to redefine it - after the relevant subroutine has been called! Finally the code must be recompiled, as in this case it will only be necessary to recompile SET_BOUNDS and not the whole code.

The DL_POLY_4 Error Messages

Message 1: error - word_2_real failure

The semantics in some of the INPUT files is wrong. DL_POLY_4 has tried to read a number but the has found a word in non-number format.

Action:

Look into your INPUT files and correct the semantics where appropriate and resubmit. DL_POLY_4 will have printed out in the OUTPUT file what the found non-uniform word is.

Message 2: error - too many atom types in FIELD (scan_field)

This error arises when DL_POLY_4 scans the FIELD file and discovers that there are too many different types of atoms in the system (i.e. the number of unique atom types exceeds the 1000).

Action:

Increase the number of allowed atom types (mmk) in SCAN_FIELD, recompile and resubmit.

Message 3: error - unknown directive found in CONTROL file

This error most likely arises when a directive is misspelt in the CONTROL file.

Action:

Locate the erroneous directive in the CONTROL file and correct error and resubmit.

Message 4: error - unknown directive found in FIELD file

This error most likely arises when a directive is misspelt or is encountered in an incorrect location in the FIELD file, which can happen if too few or too many data records are included.

Action:

Locate the erroneous directive in the FIELD file and correct error and resubmit.

Message 5: error - unknown energy unit requested

The DL_POLY_4 FIELD file permits a choice of units for input of energy parameters. These may be: electron-Volts (**eV**); k-calories per mol (**kcal**/mol); k-Joules per mol (**kJ**/mol); Kelvin per Boltzmann (**K**elvin/Boltzmann); or the DL_POLY_4 internal units, 10 Joules per mol (**internal**). There is no default value. Failure to specify any of these correctly, or reference to other energy units, will result in this error message. See documentation of the FIELD file.

Action:

Correct energy keyword on **units** directive in FIELD file and resubmit.

Message 6: error - energy unit not specified

A **units** directive is mandatory in the FIELD file. This error indicates that DL_POLY_4 has failed to find the required record.

Action:

Add **units** directive to FIELD file and resubmit.

Message 7: error - selected external field incompatible with selected ensemble (NVE only!!!)

Action:

Change the external field directive in FIELD file and or the type of ensemble in CONTROL and resubmit.

Message 8: error - ewald precision must be a POSITIVE real number

Ewald precision must be a positive non-zero real number. For example 10e-5 is accepted as a standard.

Action:

Put a correct number at the "ewald precision" directive in the CONTROL file and resubmit.

Message 10: error - too many molecular types specified

This should never happen! This indicates an erroneous FIELD file or corrupted DL_POLY_4 executable. Unlike DL_POLY_Classic, DL_POLY_4 does not have a set limit on the number of kinds of molecules it can handle in any simulation (this is not the same as the number of molecules).

Action:

Examine FIELD for erroneous directives, correct and resubmit.

Message 11: error - duplicate molecule directive in FIELD file

The number of different types of molecules in a simulation should only be specified once. If DL_POLY_4 encounters more than one **molecules** directive, it will terminate execution.

Action:

Locate the extra **molecule** directive in the FIELD file and remove and resubmit.

Message 12: error - unknown molecule directive in FIELD file

Once DL_POLY_4 encounters the **molecules** directive in the FIELD file, it assumes the following records will supply data describing the intra-molecular force field. It does not then expect to encounter directives not related to these data. This error message results if it encounters a unrelated directive. The most probable cause is incomplete specification of the data (e.g. when the **finish** directive has been omitted.)

Action:

Check the molecular data entries in the FIELD file, correct and resubmit.

Message 13: error - molecule species not specified

This error arises when DL_POLY_4 encounters non-bonded force data in the FIELD file, *before* the molecular species have been specified. Under these circumstances it cannot assign the data correctly, and therefore terminates.

Action:

Make sure the molecular data appears before the non-bonded forces data in the FIELD file and resubmit.

Message 14: error - too many unique atom types specified

This should never happen! This error most likely arises when the FIELD file or/and DL_POLY_4 executable are corrupted.

Action:

Recompile the program and/or recreate the FIELD file afresh. If no combination of these works, send the problem to us.

Message 15: error - duplicate vdw potential specified

In processing the FIELD file, DL_POLY_4 keeps a record of the specified short range pair potentials as they are read in. If it detects that a given pair potential has been specified before, no attempt at a resolution of the ambiguity is made and this error message results. See specification of FIELD file.

Action:

Locate the duplication in the FIELD file, rectify and resubmit.

Message 16: error - strange exit from FIELD file processing

This should never happen! It simply means that DL_POLY_4 has ceased processing the FIELD data, but has not reached the end of the file or encountered a **close** directive. Probable cause: corruption of the DL_POLY_4 executable or of the FIELD file. We would be interested to hear of other reasons!

Action:

See action notes on message 14 above.

Message 17: error - strange exit from CONTROL file processing

This should never happen! It simply means that DL_POLY_4 has ceased processing the CONTROL data, but has not reached the end of the file or encountered a **close** directive. Probable cause: corruption of the DL_POLY_4 executable or of the FIELD file. We would be interested to hear of other reasons!

Action:

Recompile the program and/or recreate the CONTROL file afresh. If no combination of these works, send the problem to us.

Message 18: error - duplicate three-body potential specified

DL_POLY_4 has encountered a repeat specification of a three-body potential in the FIELD file.

Action:

Locate the duplicate entry, remove and resubmit job.

Message 19: error - duplicate four-body potential specified

A 4-body potential has been duplicated in the FIELD file.

Action:

Locate the duplicated four-body potential, remove and resubmit job.

Message 20: error - too many molecule sites specified

This should never happen! This error most likely arises when the FIELD file or/and DL_POLY_4 executable are corrupted.

Action:

See action notes on message 14 above.

Message 21: error - molecule contains more atoms/sites than declared

The molecule contains more atom/site entries that it declares in the beginning.

Action:

Recreate or correct the erroneous entries in the FIELD file and try again.

Message 22: error - unsuitable radial increment in TABLE||TABBND||TABANG||TABDIH||TABIN file

This arises when the tabulated van der Waals potentials presented in the TABLE file have an increment that is greater than that used to define the other potentials in the simulation. Ideally, the increment should be $r_{\rm cut}/(\texttt{mxgrid}-4)$, where $r_{\rm cut}$ is the largest potential cutoff of all supplied ,for the short range potentials and the domain decomposition link cell size, and `mxgrid` is the parameter defining the length of the interpolation arrays. An increment less than this is permissible however. The same argument holds for the tabulated intra-molecular interactions that are possibly supplied via the TABBND, TABANG, TABDIH and TABINV files. All should have grids sized less than the generic `mxgrid`-4.

Action:

The tables must be recalculated with an appropriate increment.

Message 23: error - incompatible FIELD and TABLE file potentials

This error arises when the specification of the short range potentials is different in the FIELD and TABLE files. This usually means that the order of specification of the potentials is different. When DL_POLY_4 finds a change in the order of specification, it assumes that the user has forgotten to enter one.

Action:

Check the FIELD and TABLE files. Make sure that you correctly specify the pair potentials in the FIELD file, indicating which ones are to be presented in the TABLE file. Then check the TABLE file to make sure all the tabulated potentials are present in the order the FIELD file indicates.

Message 24: error - end of file encountered in TABLE||TABBND||TABANG||TABDIH||TABINV file

This means the TABLE||TABBND||TABANG||TABDIH||TABINV file is incomplete in some way: either by having too few potentials included, or the number of data points is incorrect.

Action:

Examine the TABLE file contents and regenerate it if it appears to be incomplete. If it look intact, check that the number of data points specified is what DL_POLY_4 is expecting.

Message 25: error - wrong atom type found in CONFIG file

On reading the input file CONFIG, DL_POLY_4 performs a check to ensure that the atoms specified in the configuration provided are compatible with the corresponding FIELD file. This message results if they are not *or the parallel reading wrongly assumed that CONFIG complies with the DL_POLY_3/4 style.*

Action:

The possibility exists that one or both of the CONFIG or FIELD files has incorrectly specified the atoms in the system. The user must locate the ambiguity, using the data printed in the OUTPUT file as a guide, and make the appropriate alteration. If the reason is in the parallel reading then produce a new CONFIG using a serial reading and continue working with it.

Message 26: error - neutral group option now redundant

DL_POLY_4 does not have the neutral group option.

Action:

Use the Ewald sum option. (It's better anyway.)

Message 27: error - unit's member indexed outside molecule's site range

An intra-molecular or intra-molecular alike interaction (topological) unit has member/site which is given a number outside the scope of the molecule it is part of.

Action:

Find the erroneous entry in FIELD, correct it and try running DL_POLY_4 again.

Message 28: error - wrongly indexed atom entries found in CONFIG file

DL_POLY_4 has detected that the atom indices in the CONFIG file do not form a contnual and/or non-repeating group of indices.

Action:

Make sure the CONFIG file is complies with the DL_POLY_4 standards. You may use the **no index** option in the CONTROL file to override the crystalographic sites' reading from the CONFIG file from reading by index to reading by order of the atom entries with consecutive incremental indexing. Using this option assumes that the FIELD topology description matches the crystalographic sites (atoms entries) in the CONFIG file by order (consecutively).

Message 30: error - too many chemical bonds specified

This should never happen! This error most likely arises when the FIELD file or/and DL_POLY_4 executable are corrupted.

Action:

See action notes on message 14 above.

Message 31: error - too many chemical bonds per domain

DL_POLY_4 limits the number of chemical bond units in the system to be simulated (actually, the number to be processed by each node) and checks for the violation of this. Termination will result if the condition is violated.

Action:

Use **densvar** option in CONTROL to increase `mxbond` (alternatively, increase it by hand in SET_BOUNDS and recompile) and resubmit.

Message 32: error - coincidence of particles in core-shell unit

DL_POLY_4 has found a fault in the definition of a core-shell unit in the FIELD file. The same particle has been assigned to the core and shell sites.

Action:

Correct the erroneous entry in FIELD and resubmit.

Message 33: error - coincidence of particles in constraint bond unit

DL_POLY_4 has found a fault in the definition of a constraint bond unit in the FIELD file. The same particle has been assigned to the both sites.

Action:

Correct the erroneous entry in FIELD and resubmit.

Message 34: error - length of constraint bond unit >= real space cutoff (rcut)

DL_POLY_4 has found a constraint bond unit length (FIELD) larger than the real space cutoff (`rcut`) (CONTROL).

Action:

Increase cutoff in CONTROL or decrease the constraint bondlength in FIELD and resubmit. For small system consider using DL_POLY_Classic.

Message 35: error - coincidence of particles in chemical bond unit

DL_POLY_4 has found a faulty chemical bond in FIELD (defined between the same particle).

Action:

Correct the erroneous entry in FIELD and resubmit.

Message 36: error - only one *bonds* directive per molecule is allowed

DL_POLY_4 has found more than one bonds entry per molecule in FIELD.

Action:

Correct the erroneous part in FIELD and resubmit.

Message 38: error - outgoing transfer buffer size exceeded in metal_ld_export

This should not usually happen!

Action:

Consider using `densvar` option in CONTROL for extremely non-equilibrium simulations. Alternatively, increase `mxbfxp` parameter in SET_BOUNDS recompile and resubmit. Send the problem to us if this is persistent.

Message 39: error - incoming data transfer size exceeds limit in metal_ld_export

See notes on message 38 above.

Action:

See action notes on message 38 above.

Message 40: error - too many bond constraints specified

This should never happen!

Action:

See action notes on message 14 above.

Message 41: error - too many bond constraints per domain

DL_POLY_4 limits the number of bond constraint units in the system to be simulated (actually, the number to be processed by each node) and checks for the violation of this. Termination will result if the condition is violated.

Action:

Use **densvar** option in CONTROL to increase `mxcons` (alternatively, increase it by hand in SET_BOUNDS and recompile) and resubmit.

Message 42: error - undefined direction passed to deport_atomic_data

This should never happen!

Action:

Send the problem to us.

Message 43: error - outgoing transfer buffer size exceeded in deport_atomic_data

This may happen in extremely non-equilibrium simulations or usually when the potentials in use do not hold the system stable.

Action:

Consider using `densvar` option in CONTROL for extremely non-equilibrium simulations. Alternatively, increase `mxbfdp` parameter in SET_BOUNDS recompile and resubmit.

Message 44: error - incoming data transfer size exceeds limit in deport_atomic_data

Action:

See action notes on message 43 above.

Message 45: error - too many atoms in CONFIG file or per domain

This can happen in circumstances when indeed the CONFIG file has more atoms listed than defined in FIELD, or when one of the domains (managed by an MPI process) has higher particle density than the system average and contains more particles than allowed by the default based on the system.

Action:

Check if CONFIG and FIELD numbers of particles match. Try executing on various number of processors. Try using the **densvar** option in CONTROL to increase `mxatms` (alternatively, increase it by hand in SET_BOUNDS and recompile) and resubmit. Send the problem to us if this is persistent.

Message 46: error - undefined direction passed to export_atomic_data

This should never happen!

Action:

Send the problem to us.

Message 47: error - undefined direction passed to metal_ld_export

This should never happen!

Action:

Send the problem to us.

Message 48: error - transfer buffer too small in *_table_read

Action:

Standard user response. Increase `mxgrid` parameter in SET_BOUNDS recompile and resubmit.

Message 49: error - frozen shell (core-shell) unit specified

The DL_POLY_4 option to freeze the location of an atom (i.e. hold it permanently in one position) is not permitted for the shells in core-shell units.

Action:

Remove the frozen atom option from the FIELD file. Consider using a non-polarisable atom instead.

Message 50: error - too many bond angles specified

This should never happen! This error most likely arises when the FIELD file or/and DL_POLY_4 executable are corrupted.

Action:

See action notes on message 14 above.

Message 51: error - too many bond angles per domain

DL_POLY_4 limits the number of valence angle units in the system to be simulated (actually, the number to be processed by each node) and checks for the violation of this. Termination will result if the condition is violated.

Action:

Use **densvar** option in CONTROL to increase `mxangl` (alternatively, increase it by hand in SET_BOUNDS and recompile) and resubmit.

Message 52: error - end of FIELD file encountered

This message results when DL_POLY_4 reaches the end of the FIELD file, without having read all the data it expects. Probable causes: missing data or incorrect specification of integers on the various directives.

Action:

Check FIELD file for missing or incorrect data, correct and resubmit.

Message 53: error - end of CONTROL file encountered

This message results when DL_POLY_4 reaches the end of the CONTROL file, without having read all the data it expects. Probable cause: missing **finish** directive.

Action:

Check CONTROL file, correct and resubmit.

Message 54: error - outgoing transfer buffer size exceeded in export_atomic_data

See notes on message 38 above.

Action:

See naction otes on message 38 above.

Message 55: error - end of CONFIG file encountered

This error arises when DL_POLY_4 attempts to read more data from the CONFIG file than is actually present. The probable cause is an incorrect or absent CONFIG file, but it may be due to the FIELD file being incompatible in some way with the CONFIG file.

Action:

Check contents of CONFIG file. If you are convinced it is correct, check the FIELD file for inconsistencies.

Message 56: error - incoming data transfer size exceeds limit in export_atomic_data

See notes on message 38 above.

Action:

See action notes on message 38 above.

Message 57: error - too many core-shell units specified

This should never happen!

Action:

See action notes on message 14 above.

Message 58: error - number of atoms in system not conserved

Either and an atom has been lost in transfer between nodes/domains or your FIELD is ill defined with respect to what is supplied in CONFIG/HISTORY.

Action:

If this error is issued at start before timestep zero in a simulation then it is either your FIELD file is ill defined or that your CONFIG file (or the first frame of your HISTRORY being replayed). Check out for mistyped number or identities of molecules, atoms, etc. in FIELD and for mangled/blank lines in CONFIG/HISTORY, or a blank line(s) at the end of CONFIG or missing FOF (End Of File) character in CONFIG. If this error is issued after timestep zero in a simulation that is not replaying HISTORY then it is big trouble and you should report that to the authors. If it is during replaying HISTORY then your HISTORY file has corrupted frames and you must correct it before trying again.

Message 59: error - too many core-shell units per domain

DL_POLY_4 limits the number of core-shell units in the system to be simulated (actually, the number to be processed by each node) and checks for the violation of this. Termination will result if the condition is violated.

Action:

Use **densvar** option in CONTROL to increase `mxshl` (alternatively, increase it by hand in SET_BOUNDS and recompile) and resubmit.

Message 60: error - too many dihedral angles specified

This should never happen!

Action:

See action notes on message 14 above.

Message 61: error - too many dihedral angles per domain

DL_POLY_4 limits the number of dihedral angle units in the system to be simulated (actually, the number to be processed by each node) and checks for the violation of this. Termination will result if the condition is violated.

Action:

Use **densvar** option in CONTROL to increase `mxdihd` (alternatively, increase it by hand in SET_BOUNDS and recompile) and resubmit.

Message 62: error - too many tethered atoms specified

This should never happen!

Action:

See action notes on message 14 above.

Message 63: error - too many tethered atoms per domain

DL_POLY_4 limits the number of tethered atoms in the system to be simulated (actually, the number to be processed by each node) and checks for the violation of this. Termination will result if the condition is violated.

Action:

Use **densvar** option in CONTROL to increase `mxteth` (alternatively, increase it by hand in SET_BOUNDS and recompile) and resubmit.

Message 64: error - incomplete core-shell unit found in build_book_intra

This should never happen!

Action:

Report problem to authors.

Message 65: error - too many excluded pairs specified

This should never happen! This error arises when DL_POLY_4 is identifying the atom pairs that cannot have a pair potential between them, by virtue of being chemically bonded for example (see subroutine BUILD_EXCL_INTRA). Some of the working arrays used in this operation may be exceeded, resulting in termination of the program.

Action:

Contact authors.

Message 66: error - coincidence of particles in bond angle unit

DL_POLY_4 has found a fault in the definition of a bond angle in the FIELD file.

Action:

Correct the erroneous entry in FIELD and resubmit.

Message 67: error - coincidence of particles in dihedral unit

DL_POLY_4 has found a fault in the definition of a dihedral unit in the FIELD file.

Action:

Correct the erroneous entry in FIELD and resubmit.

Message 68: error - coincidence of particles in inversion unit

DL_POLY_4 has found a fault in the definition of a inversion unit in the FIELD file.

Action:

Correct the erroneous entry in FIELD and resubmit.

Message 69: error - too many link cells required in three_body_forces

The number of link cells required for the build up of the Verlet neighbour list (as in link_cell_pairs) or the calculation of three- & four-body as well tersoff forces (as in three_body_forces, four_body_forces, tersoff_body_forces) in the given model exceeds the number allowed for by the DL_POLY_4 arrays. Probable cause: your system has expanded unacceptably much to DL_POLY_4. This may not be physically sensible!

Action:

Consider using `densvar` option in CONTROL for extremely non-equilibrium simulations.

Message 70: error - constraint_quench failure

When a simulation with bond constraints is started, DL_POLY_4 attempts to extract the kinetic energy of the constrained atom-atom bonds arising from the assignment of initial random velocities. If this procedure fails, the program will terminate. The likely cause is a badly generated initial configuration.

Action:

Some help may be gained from increasing the cycle limit, by using the directive **mxshak** in the CONTROL file. You may also consider reducing the tolerance of the SHAKE iteration using the directive **shake** in the CONTROL file. However it is probably better to take a good look at the starting conditions!

Message 71: error - too many metal potentials specified

This should never happen!

Action:

Report to authors.

Message 72: error - too many tersoff potentials specified

This should never happen!

Action:

Report to authors.

Message 73: error - too many inversion potentials specified

This should never happen!

Action:

Report to authors.

Message 74: error - unidentified atom in tersoff potential list

This shows that DL_POLY_4 has encountered and erroneous entry for Tersoff potentials in FIELD.

Action:

Correct FIELD and resubmit.

Message 76: error - duplicate tersoff potential specified

This shows that DL_POLY_4 has encountered and erroneous entry for Tersoff potentials in FIELD.

Action:

Correct FIELD and resubmit.

Message 77: error - too many inversion angles per domain

DL_POLY_4 limits the number of inversion units in the system to be simulated (actually, the number to be processed by each node) and checks for the violation of this. Termination will result if the condition is violated.

Action:

Use **densvar** option in CONTROL to increase `mxinv` (alternatively, increase it by hand in SET_BOUNDS and recompile) and resubmit.

Message 79: error - tersoff potential cutoff undefined

This shows that DL_POLY_4 has encountered and erroneous entry for Tersoff potentials in FIELD.

Action:

Correct FIELD and resubmit.

Message 80: error - too many pair potentials specified

This should never happen!

Action:

Report to authors.

Message 81: error - unidentified atom in pair potential list

This shows that DL_POLY_4 has encountered and erroneous entry for vdw or metal potentials in FIELD or cited TABle file.

Action:

Correct FIELD and/or cited TABle file.

Message 82: error - calculated pair potential index too large

This should never happen! In checking the vdw and metal potentials specified in the FIELD file DL_POLY_4 calculates a unique integer indices that henceforth identify every specific potential within the program. If this index becomes too large, termination of the program results.

Action:

Report to authors.

Message 83: error - too many three-body/angles potentials specified

This should never happen!

Action:

Report to authors.

Message 84: error - unidentified atom in three-body/angles potential list

This shows that DL_POLY_4 has encountered and erroneous entry at three-body or angles definitions in FIELD.

Action:

Correct FIELD and resubmit.

Message 85: error - required velocities not in CONFIG file

If the user attempts to start up a DL_POLY_4 simulation with any type of **restart** directive (see description of CONTROL file,) the program will expect the CONFIG file to contain atomic velocities as well as positions. Termination results if these are not present.

Action:

Either replace the CONFIG file with one containing the velocities, or if not available, remove the **restart ...** directive altogether and let DL_POLY_4 create the velocities for itself.

Message 86: error - calculated three-body potential index too large

This should never happen! DL_POLY_4 has a permitted maximum for the calculated index for any three-body potential in the system (i.e. as defined in the FIELD file). If there are m distinct types of atom in the system, the index can possibly range from 1 to $(m^2 * (m-1))/2$. If the internally calculated index exceeds this number, this error reports results.

Action:

Report to authors.

Message 88: error - legend array exceeded in build_book_intra

The second dimension of a legend array has been exceeded.

Action:

If you have an intra-molecular (like) interaction present in abundance in your model that you suspect is driving this out of bound error increase its legend bound value, `mxf`interaction, at the end of SCAN_FIELD, recompile and resubmit. If the error persists contact authors.

Message 89: error - too many four-body/dihedrals/inversions potentials specified

This should never happen!

Action:

Report to authors.

Message 90: error - specified tersoff potentials have different types'

This is not allowed! Only one general type of tersoff potential is allowed in FIELD as there are no mixing rules between different tersoff potentials!

Action:

Correct your model representation in FIELD and try again.

Message 91: error - unidentified atom in four-body/dihedrals/inversions potential list

The specification of a four-body or dihedrals or inversions potential in the FIELD file has referenced an atom type that is unknown.

Action:

Locate the errant atom type in the four-body/dihedrals/inversions potential definition in the FIELD file and correct. Make sure this atom type is specified by an `atoms` directive earlier in the file.

Message 92: error - specified metal potentials have different types

The specified metal interactions in the FIELD file are referencing more than one generic type of metal potentials. Only one such type is allowed in the system.

Action:

Locate the errant metal type in the metal potential definition in the FIELD file and correct. Make sure only one metal type is specified for all relevan atom interactions in the file.

Message 93: error - PMFs mixing with rigid bodies not allowed

Action:

Correct FIELD and resubmit.

Message 95: error - error - rcut or (rcut+rpad) > minimum of all half-cell widths

In order for the minimum image convention to work correctly within DL_POLY_4, it is necessary to ensure that the major cutoff, plus its possible padding distance, applied to the pair interactions does not exceed half the perpendicular width of the simulation cell. (The perpendicular width is the shortest distance between opposing cell faces.) Termination results if this is detected. In NVE and NVT simulations this can only happen at the start of a simulation, but in NPT and N$\underline{\underline{\sigma}}$T, it may occur at any time.

Action:

Supply a cutoff that is less than half the cell width. If running constant pressure calculations, use a cutoff that will accommodate the fluctuations in the simulation cell. Study the fluctuations in the OUTPUT file to help you with this.

Message 96: error - incorrect atom totals assignments in metal_ld_set_halo

This should never happen!

Action:

Big trouble. Report to authors.

Message 97: error - constraints mixing with rigid bodies not allowed

Action:

Correct FIELD and resubmit.

Message 99: error - cannot have shells as part of a constraint, rigid body or tether

Action:

Correct FIELD and resubmit.

Message 100: error - core-shell unit separation > rcut (the system cutoff)

This could only happen if FIELD and CONFIG do not match each other or CONFIG is damaged.

Action:

Regenerate CONFIG (and FIELD) and resubmit.

Message 101: error - calculated four-body potential index too large

This should never happen! DL_POLY_4 has a permitted maximum for the calculated index for any four-body potential in the system (i.e. as defined in the FIELD file). If there are m distinct types of atom in the system, the index can possibly range from 1 to $(m^2 * (m+1) * (m+2))/6$. If the internally calculated index exceeds this number, this error report results.

Action:

Report to authors.

Message 102: error - rcut < 2*rcter (maximum cutoff for tersoff potentials)

The nature of the Tersoff interaction requires they have at least twice shorter cutoff than the standard pair interctions (or the major system cutoff).

Action:

Decrease Tersoff cutoffs in FIELD or increase cutoff in CONTROL and resubmit.

Message 103: error - parameter mxlshp exceeded in pass_shared_units

Various algorithms (constraint and core-shell ones) require that information about 'shared' atoms be passed between nodes. If there are too many such atoms, the arrays holding the information will be exceeded and DL_POLY_4 will terminate execution.

Action:

Use **densvar** option in CONTROL to increase `mxlshp` (alternatively, increase it by hand in SET_BOUNDS and recompile) and resubmit.

Message 104: error - arrays listme and lstout exceeded in pass_shared_units

This should not happen! Dimensions of indicated arrays have been exceeded.

Action:

Consider using `densvar` option in CONTROL for extremely non-equilibrium simulations.

Message 105: error - shake algorithm (constraints_shake) failed to converge

The SHAKE algorithm for bond constraints is iterative. If the maximum number of permitted iterations is exceeded, the program terminates. Possible causes include: a bad starting configuration; too large a time step used; incorrect force field specification; too high a temperature; inconsistent constraints (over-constraint) etc..

Action:

You may try to increase the limit of iteration cycles in the constraint subroutines by using the directive **mxshak** and/or decrease the constraint precision by using the directive **shake** in CONTROL. But the trouble may be much more likely to be cured by careful consideration of the physical system being simulated. For example, is the system stressed in some way? Too far from equilibrium?

Message 106: error - neighbour list array too small in link_cell_pairs

Construction of the Verlet neighbour list in subroutine LINK_CELL_PAIRS non-bonded (pair) force has exceeded the neighbour list array dimensions.

Action:

Consider using `densvar` option in CONTROL for extremely non-equilibrium simulations or increase by hand `mxlist` in SET_BOUNDS.

Message 107: error - too many pairs for rdf look up specified

This should never happen! A possible reason is corruption in FIELD or/and DL_POLY_4 executable.

Action:

See action notes on message 14 above.

Message 108: error - unidentified atom in rdf look up list

During reading of RDF look up pairs in FIELD DL_POLY_4 has found an unlisted previously atom type.

Action:

Correct FIELD by either defining the new atom type or changing it to an already defined one in the erroneous line. Resubmit.

Message 109: error - calculated pair rdf index too large

This should never happen! In checking the RDF pairs specified in the FIELD file DL_POLY_4 calculates a unique integer index that henceforth identify every RDF pair within the program. If this index becomes too large, termination of the program results.

Action:

Report to authors.

Message 108: error - duplicate rdf look up pair specified

During reading of RDF look up pairs in FIELD DL_POLY_4 has found a duplicate entry in the list.

Action:

Delete the duplicate line and resubmit.

Message 111: error - bond constraint unit separation > rcut (the system cutoff)

This should never happen! DL_POLY_4 has not been able to find an atom in a processor domain or its bordering neighbours.

Action:

Probable cause: link cells too small. Use larger potential cutoff. Contact DL_POLY_4 authors.

Message 112: error - only one *constraints* directive per molecule is allowed

DL_POLY_4 has found more than one constraints entry per molecule in FIELD.

Action:

Correct the erroneous part in FIELD and resubmit.

Message 113: error - intra-molecular bookkeeping arrays exceeded in deport_atomic_data

One or more bookkeeping arrays for site-related interactions have been exceeded.

Action:

Consider using `densvar` option in CONTROL for extremely non-equilibrium simulations. Alternatively, you will need to print extra diagnostic data from the DEPORT_ATOMIC_DATA subroutine to find which boded-like contribution has exceeded its assumed limit and then correct for it in SET_BOUNDS, recompile and resubmit.

Message 114: error - legend array exceeded in deport_atomic_data

The array `legend` has been exceeded.

Action:

Try increasing parameter `mxfix` in SET_BOUNDS, recompile and resubmit. Contact DL_POLY_4 authors if the problem persists.

Message 115: error - transfer buffer exceeded in update_shared_units

The transfer buffer has been exceeded.

Action:

Consider increasing parameter `mxbfsh` in SET_BOUNDS, recompile and resubmit. Contact DL_POLY_4 authors if the problem persists.

Message 116: error - incorrect atom transfer in update_shared_units

An atom has become misplaced during transfer between nodes.

Action:

This happens when the simulation is very numerically unstable. Consider carefully the physical grounds of your simulation, i.e. are you using the adiabatic shell model for accounting polarisation with too big a timestep or too large control distances for the variable timestep, is the ensemble type NPT or N$\underline{\underline{\sigma}}$T and the system target temperature too close to the melting temperature?

Message 118: error - construction error in pass_shared_units

This should not happen.

Action:

Report to authors.

Message 120: error - invalid determinant in matrix inversion

DL_POLY_4 occasionally needs to calculate matrix inverses (usually the inverse of the matrix of cell vectors, which is of size 3×3). For safety's sake a check on the determinant is made, to prevent inadvertent use of a singular matrix.

Action:

Locate the incorrect matrix and fix it - e.g. are cell vectors correct?

Message 122: error - FIELD file not found

DL_POLY_4 failed to find a FIELD file in your directory.

Action:

Supply a valid FIELD file before you start a simulation

Message 124: error - CONFIG file not found

DL_POLY_4 failed to find a CONFIG file in your directory.

Action:

Supply a valid CONFIG file before you start a simulation

Message 126: error - CONTROL file not found

DL_POLY_4 failed to find a CONTROL file in your directory.

Action:

Supply a valid CONTROL file before you start a simulation

Message 128: error - chemical bond unit separation > rcut (the system cutoff)

This could only happen if FIELD and CONFIG do not match each other or if the instantaneous configuration is ill defined because of generation of large forces on bonded particles. This may be due to having a badly defined force-field and/or starting form a configuration which is too much away from equilibrium.

Action:

Regenerate CONFIG (and FIELD) and resubmit. Try topology verification by using `nfold 1 1 1` in CONTROL. Try using options as `scale`, `cap`, `zero` and `optimise`. Try using smaller SHAKE tolerance if constraints are present in the system. You may as well try using the `variable timestep` option.

Message 130: error - bond angle unit diameter > rcut (the system cutoff)

See action notes on message 128 above.

Action:

See action notes on message 128 above.

Message 132: error - dihedral angle unit diameter > rcut (the system cutoff)

See notes on message 128 above.

Action:

See action notes on message 128 above.

Message 134: error - inversion angle unit diameter > rcut (the system cutoff)

See notes on message 128 above.

Action:

See action notes on message 128 above.

Message 138: error - incorrect atom totals assignments in refresh_halo_positions

This should never happen although, sometimes, it could due to ill defined force field and/or and/or starting form a configuration which is too much away from equilibrium.

Action:

Try using the `variable timestep` option and/or running in serial to determine if particles gain too much speed and leave domains.

Message 141: error - duplicate metal potential specified

During reading of metal potentials (pairs of atom types) in FIELD DL_POLY_4 has found a duplicate pair of atoms in the list.

Action:

Delete one of the duplicate entries and resubmit.

Message 145: error - no two-body like interactions specified

This error arises when there are no two-body like interactions specified in FIELD and CONTROL. I.e. none of the following interactions exists or if does, it has been switched off; any coulombic, vdw, metal, tersoff. In DL_POLY_4 expects that particles will be kept apparat, stay separated and never go through each other due to one of the fore-specified interactions.

Action:

Users must alone take measures to prevent such outcome.

Message 150: error - unknown van der waals potential selected

DL_POLY_4 checks when constructing the interpolation tables for the short ranged potentials that the potential function requested is one which is of a form known to the program. If the requested potential form is unknown, termination of the program results. The most probable cause of this is the incorrect choice of the potential keyword in the FIELD file.

Action:

Read the DL_POLY_4 documentation and find the potential keyword for the potential desired.

Message 151: error - unknown EAM keyword in TABEAM

DL_POLY_4 checks when constructing the interpolation tables for the EAM metal potentials that the potential function requested is one which is of a form known to the program. If the requested potential form is unknown, termination of the program results. The most probable cause of this is the incorrect choice of the potential keyword in the FIELD file.

Action:

Read the DL_POLY_4 documentation and find the potential keyword for the potential desired.

Message 152: error - undefined direction passed to dpd_v_export

This should never happen!

Action:

Report to authors.

Message 154: error - outgoing transfer buffer size exceeded in dpd_v_export

See notes on message 38 above.

Action:

See action notes on message 38 above.

Message 156: error - incoming data transfer size exceeds limit in dpd_v_export

See notes on message 38 above.

Action:

See action notes on message 38 above.

Message 158: error - incorrect atom totals assignments in dpd_v_set_halo

This should never happen!

Action:

Big trouble. Report to authors.

Message 160: error - undefined direction passed to statistics_connect_spread

This should never happen!

Message 163: error - outgoing transfer buffer size exceeded in statistics_connect_spread

The transfer buffer has been exceeded.

Action:

Consider using `densvar` option in CONTROL for extremely non-equilibrium simulations. Alternatively, increase `mxbfss` parameters in SET_BOUNDS recompile and resubmit.

Message 164: error - incoming data transfer size exceeds limit in statistics_connect_spread

See notes on message 163 above.

Action:

See action notes on message 163 above.

Message 170: error - too many variables for statistics array

This error means the statistics arrays appearing in subroutine STATISTICS_COLLECT are too small. This should never happen!

Action:

Contact DL_POLY_4 authors.

Message 172: error - duplicate intra-molecular entries specified in TABBND||TABANG||TABDIH||TA

A duplicate entry has been encountered in the intra-molecular table file.

Action:

Contact DL_POLY_4 authors.

Message 200: error - rdf/z-density buffer array too small in system_revive

This error indicates that a global summation buffer array in subroutine SYSTEM_REVIVE is too small, i.e `mxbuff` < `mxgrdf`. This should never happen!

Action:

Contact DL_POLY_4 authors.

Message 210: error - only one *angles* directive per molecule is allowed

DL_POLY_4 has found more than one angles entry per molecule in FIELD.

Action:

Correct the erroneous part in FIELD and resubmit.

Message 220: error - only one *dihedrals* directive per molecule is allowed

DL_POLY_4 has found more than one dihedrals entry per molecule in FIELD.

Action:

Correct the erroneous part in FIELD and resubmit.

Message 230: error - only one *inversions* directive per molecule is allowed

DL_POLY_4 has found more than one inversions entry per molecule in FIELD.

Action:

Correct the erroneous part in FIELD and resubmit.

Message 240: error - only one *tethers* directive per molecule is allowed

DL_POLY_4 has found more than one tethers entry per molecule in FIELD.

Action:

Correct the erroneous part in FIELD and resubmit.

Message 300: error - incorrect boundary condition for link-cell algorithms

The use of link cells in DL_POLY_4 implies the use of appropriate boundary conditions. This error results if the user specifies octahedral or dodecahedral boundary conditions, which are only available in DL_POLY_Classic.

Action:

Correct your boundary condition or consider using DL_POLY_Classic.

Message 305: error - too few link cells per dimension for many-body and tersoff forces subroutines.

The link cells algorithms for many-body and tersoff forces in DL_POLY_4 cannot work with less than 3 (secondary) link cells per dimension. This depends on the cell size widths (as supplied in CONFIG) and the largest system cut-off (as specified in CONTROL although it may be drawn or overridden by cutoffs specified as part of some potentials' parameter sets in FIELD).

Action:

Decrease many-body and tersoff potentials cutoffs or/and number of nodes or/and increase system size.

Message 307: error - link cell algorithm violation

DL_POLY_4 does not like what you are asking it to do. Probable cause: the cutoff is too large to use link cells in this case.

Action:

Rethink the simulation model; reduce the cutoff or/and number of nodes or/and increase system size.

Message 308: error - link cell algorithm in contention with SPME sum precision

DL_POLY_4 does not like what you are asking it to do. Probable cause: you ask for SPME precision that is not achievable by the current settings of the link cell algorithm.

Action:

Rethink the simulation model; reduce number of nodes or/and SPME sum precision or/and increase cutoff.

Message 321: error - LFV quaternion integrator failed

This indicates unstable integration but may be due to many reasons.

Action:

Rethink the simulation model. Increase mxquat in CONTROL and resubmit or use VV integration to check system stability.

Message 340: error - invalid integration option requested

DL_POLY_4 has detected an incompatibility in the simulation instructions, namely that the requested integration algorithm is not compatible with the physical model. It *may* be possible to override this error trap, but it is up to the user to establish if this is sensible.

Action:

This is a non-recoverable error, unless the user chooses to override the restriction.

Message 350: error - too few degrees of freedom

This error can arise if a small system is being simulated and the number of constraints applied is too large.

Action:

Simulate a larger system or reduce the number of constraints.

Message 360: error - degrees of freedom distribution problem

This should never happen for a dynamically sensical system. This error arises if a model system contains one or more free, zero mass particles. Zero mass (mass-less) particles/sites are only allowed for shells in core-shell units and as part of rigid bodies (mass-less but charged RB sites).

Action:

Inspect your FIELD to find and correct the erroneous entries, and try again.

Message 380: error - simulation temperature not specified or < 1 K

DL_POLY_4 has failed to find a **temp** directive in the CONTROL file.

Action:

Place a **temp** directive in the CONTROL file, with the required temperature specified.

Message 381: error - simulation timestep not specified

DL_POLY_4 has failed to find a **timestep** directive in the CONTROL file.

Action:

Place a **timestep** directive in the CONTROL file, with the required timestep specified.

Message 382: error - simulation cutoff not specified

DL_POLY_4 has failed to find a **cutoff** directive in the CONTROL file.

Action:

Place a **cutoff** directive in the CONTROL file, with the required forces cutoff specified.

Message 387: error - system pressure not specified

The target system pressure has not been specified in the CONTROL file. Applies to NPT simulations only.

Action:

Insert a **press** directive in the CONTROL file specifying the required system pressure.

Message 390: error - npt/nst ensemble requested in non-periodic system

A non-periodic system has no defined volume, hence the NPT algorithm cannot be applied.

Action:

Either simulate the system with a periodic boundary, or use another ensemble.

Message 402: error - van der waals not specified

The user has not set any cutoff in CONTROL, (`rvdw`) - the van der Waals potentials cutoff is needed in order for DL_POLY_4 to proceed.

Action:

Supply a cutoff value for the van der Waals terms in the CONTROL file using the directive `rvdw`, and resubmit job.

Message 410: error - cell not consistent with image convention

The simulation cell vectors appearing in the CONFIG file are not consistent with the specified image convention.

Action:

Locate the variable `imcon` in the CONFIG file and correct to suit the cell vectors.

Message 414: error - conflicting ensemble options in CONTROL file

DL_POLY_4 has found more than one **ensemble** directive in the CONTROL file.

Action:

Locate extra **ensemble** directives in CONTROL file and remove.

Message 416: error - conflicting force options in CONTROL file

DL_POLY_4 has found incompatible directives in the CONTROL file specifying the electrostatic interactions options.

Action:

Locate the conflicting directives in the CONTROL file and correct.

Message 430: error - integration routine not available

A request for a non-existent ensemble has been made or a request with conflicting options that DL_POLY_4 cannot deal with.

Action:

Examine the CONTROL and FIELD files and remove inappropriate specifications.

Message 432: error - undefined tersoff potential

This shows that DL_POLY_4 has encountered an unfamiliar entry for Tersoff potentials in FIELD.

Action:

Correct FIELD and resubmit.

Message 433: error - rcut must be specified for the Ewald sum precision

When specifying the desired precision for the Ewald sum in the CONTROL file, it is also necessary to specify the real space cutoff `rcut`.

Action:

Place the **cut** directive *before* the **ewald precision** directive in the CONTROL file and rerun.

Message 436: error - unrecognised ensemble

An unknown ensemble option has been specified in the CONTROL file.

Action:

Locate **ensemble** directive in the CONTROL file and amend appropriately.

Message 440: error - undefined angular potential

A form of angular potential has been requested which DL_POLY_4 does not recognise.

Action:

Locate the offending potential in the FIELD file and remove. Replace with one acceptable to DL_POLY_4 if this is possible. Alternatively, you may consider defining the required potential in the code yourself. Amendments to subroutines READ_FIELD and ANGLES_FORCES will be required.

Message 442: error - undefined three-body potential

A form of three-body potential has been requested which DL_POLY_4 does not recognise.

Action:

Locate the offending potential in the FIELD file and remove. Replace with one acceptable to DL_POLY_4 if this is reasonable. Alternatively, you may consider defining the required potential in the code yourself. Amendments to subroutines READ_FIELD and THREE_BODY_FORCES will be required.

Message 443: error - undefined four-body potential

DL_POLY_4 has been requested to process a four-body potential it does not recognise.

Action:

Check the FIELD file and make sure the keyword is correctly defined. Make sure that subroutine THREE_BODY_FORCES contains the code necessary to deal with the requested potential. Add the code required if necessary, by amending subroutines READ_FIELD and THREE_BODY_FORCES.

Message 444: error - undefined bond potential

DL_POLY_4 has been requested to process a bond potential it does not recognise.

Action:

Check the FIELD file and make sure the keyword is correctly defined. Make sure that subroutine BONDS_FORCES contains the code necessary to deal with the requested potential. Add the code required if necessary, by amending subroutines READ_FIELD and BONDS_FORCES.

Message 445: error - r_14 > rcut in dihedrals_forces

The 1-4 coulombic scaling for a dihedral angle bonding cannot be performed since the 1-4 distance has exceeded the system short range interaction cutoff, `rcut`, in subroutine DIHEDRAL_FORCES.

Action:

To prevent this error occurring again increase `rcut`.

Message 446: error - undefined electrostatic key in dihedral_forces

The subroutine DIHEDRAL_FORCES has been requested to process a form of electrostatic potential it does not recognise.

Action:

The error arises because the integer key `keyfrc` has an inappropriate value (which should not happen in the standard version of DL_POLY_4). Check that the FIELD file correctly specifies the potential. Make sure the version of DIHEDRAL_FORCES does contain the potential you are specifying. Report the error to the authors if these checks are correct.

Action:

To prevent this error occurring again increase `rvdw`.

Message 447: error - only one *shells* directive per molecule is allowed

DL_POLY_4 has found more than one shells entry per molecule in FIELD.

Action:

Correct the erroneous part in FIELD and resubmit.

Message 448: error - undefined dihedral potential

A form of dihedral potential has been requested which DL_POLY_4 does not recognise.

Action:

Locate the offending potential in the FIELD file and remove. Replace with one acceptable to DL_POLY_4 if this is reasonable. Alternatively, you may consider defining the required potential in the code yourself. Amendments to subroutines READ_FIELD and DIHEDRAL_FORCES (and its variants) will be required.

Message 449: error - undefined inversion potential

A form of inversion potential has been encountered which DL_POLY_4 does not recognise.

Action:

Locate the offending potential in the FIELD file and remove. Replace with one acceptable to DL_POLY_4 if this is reasonable. Alternatively, you may consider defining the required potential in the code yourself. Amendments to subroutines READ_FIELD and INVERSIONS_FORCES will be required.

Message 450: error - undefined tethering potential

A form of tethering potential has been requested which DL_POLY_4 does not recognise.

Action:

Locate the offending potential in the FIELD file and remove. Replace with one acceptable to DL_POLY_4 if this is reasonable. Alternatively, you may consider defining the required potential in the code yourself. Amendments to subroutines READ_FIELD and TETHERS_FORCES will be required.

Message 451: error - three-body potential cutoff undefined

The cutoff radius for a three-body potential has not been defined in the FIELD file.

Action:

Locate the offending three-body force potential in the FIELD file and add the required cutoff. Resubmit the job.

Message 452: error - undefined vdw potential

A form of vdw potential has been requested which DL_POLY_4 does not recognise.

Action:

Locate the offending potential in the FIELD file and remove. Replace with one acceptable to DL_POLY_4 if this is reasonable. Alternatively, you may consider defining the required potential in the code yourself. Amendments to subroutines READ_FIELD, VDW_GENERATE* and DIHEDRALS_14_VDW will be required.

Message 453: error - four-body potential cutoff undefined

The cutoff radius for a four-body potential has not been defined in the FIELD file.

Action:

Locate the offending four-body force potential in the FIELD file and add the required cutoff. Resubmit the job.

Message 454: error - unknown external field

A form of external field potential has been requested which DL_POLY_4 does not recognise.

Action:

Locate the offending potential in the FIELD file and remove. Replace with one acceptable to DL_POLY_4 if this is reasonable. Alternatively, you may consider defining the required potential in the code yourself. Amendments to subroutines READ_FIELD and EXTERNAL_FIELD_APPLY will be required.

Message 456: error - external field xpis-ton is applied to a layer with at least one frozen particle

For a layer to emulate a piston no particle constituting it must be frozen.

Action:

Locate the offending site(s) in the FIELD file and unfreeze the particles.

Message 461: error - undefined metal potential

A form of metal potential has been requested which DL_POLY_4 does not recognise.

Action:

Locate erroneous entry in the FIELD file and correct the potental interaction to one of the allowed ones for metals in DL_POLY_4.

Message 462: error - thermostat friction constant must be > 0

A zero or negative value for the thermostat friction constant has been encountered in the CONTROL file.

Action:

Locate the **ensemble** directive in the CONTROL file and assign a positive value to the time constant.

Message 463: error - barostat friction constant must be > 0

A zero or negative value for the barostat friction constant has been encountered in the CONTROL file.

Action:

Locate the **ensemble** directive in the CONTROL file and assign a positive value to the time constant.

Message 464: error - thermostat relaxation time constant must be > 0

A zero or negative value for the thermostat relaxation time constant has been encountered in the CONTROL file.

Action:

Locate the **ensemble** directive in the CONTROL file and assign a positive value to the time constant.

Message 466: error - barostat relaxation time constant must be > 0

A zero or negative value for the barostat relaxation time constant has been encountered in the CONTROL file.

Action:

Locate the **ensemble** directive in the CONTROL file and assign a positive value to the time constant.

Message 467: error - rho must not be zero in valid buckingham potential

User specified vdw type buckingham potential has a non-zero force and zero rho constants. Only both zero or both non-zero are allowed.

Action:

Inspect the FIELD file and change the values in question appropriately.

Message 468: error - r0 too large for snm potential with current cutoff

The specified location (r0) of the potential minimum for a shifted n-m potential exceeds the specified potential cutoff. A potential with the desired minimum cannot be created.

Action:

To obtain a potential with the desired minimum it is necessary to increase the van der Waals cutoff. Locate the `rvdw` directive in the CONTROL file and reset to a magnitude greater than r0. Alternatively adjust the value of r0 in the FIELD file. Check that the FIELD file is correctly formatted.

Message 470: error - n < m in definition of n-m potential

The specification of a n-m potential in the FIELD file implies that the exponent m is larger than exponent n. (Not all versions of DL_POLY_4 are affected by this.)

Action:

Locate the n-m potential in the FIELD file and reverse the order of the exponents. Resubmit the job.

Message 471: error - rcut < 2*rctbp (maximum cutoff for three-body potentials)

The cutoff for the pair interactions is smaller than twice that for the three-body interactions. This is a bookkeeping requirement for DL_POLY_4.

Action:

Either use a smaller three-body cutoff, or a larger pair potential cutoff.

Message 472: error - rcut < 2*rcfbp (maximum cutoff for four-body potentials)

The cutoff for the pair interactions is smaller than twice that for the four-body interactions. This is a bookkeeping requirement for DL_POLY_4.

Action:

Either use a smaller four-body cutoff, or a larger pair potential cutoff.

Message 474: error - conjugate gradient mimimiser cycle limit exceeded

The conjugate gradient minimiser exceeded the iteration limit (100 for the relaxed shell model, 1000 for the configuration minimiser).

Action:

Decrease the respective convergence criterion. Alternatively, you may try to increase the limit by hand in CORE_SHELL_RELAX or in MINIMISE_RELAX respectively and recompile. However, it is unlikely that such measures will cure the problem as it is more likely to lay in the physical description of the system being simulated. For example, are the core-shell spring constants well defined? Is the system being too far from equilibrium?

Message 476: error - shells MUST all HAVE either zero or non-zero masses

The polarisation of ions is accounted via a core-shell model as the shell dynamics is either relaxed - shells have no mass, or adiabatic - all shells have non-zero mass.

Action:

Choose which model you would like to use in the simulated system and adapt the shell masses in FIELD to comply with your choice.

Message 478: error - shake algorithms (constraints & pmf) failed to converge

Your system has both bond and PMF constraints. SHAKE (RATTLE_VV1) is done by combined application of both bond and PMF constraints SHAKE (RATTLE_VV1) in an iterative manner until the PMF constraint virial converges to a constant. No such convergence is achieved.

Action:

See action notes on message 515 below.

Message 480: error - PMF constraint length > minimum of all half-cell widths

The specified PMF length has exceeded the minimum of all half-cell widths.

Action:

Specify shorter PMF length or increase MD cell dimensions.

Message 484: error - only one potential of mean force permitted

Only one potential of mean force is permitted in FIELD.

Action:

Correct the erroneous entries in FIELD.

Message 486: error - only one of the PMF units is permitted to have frozen atoms

Only one of the PMF units is permitted to have frozen atoms.

Action:

Correct the erroneous entries in FIELD.

Message 488: error - too many PMF constraints per domain

This should not happen.

Action:

Is the use of PMF constraints in your system physically sound?

Message 490: error - local PMF constraint not found locally

This should not happen.

Action:

Is your system physically sound, is your system equilibrated?

Message 492: error - a diameter of a PMF unit > minimum of all half cell widths

The diameter of a PMF unit has exceeded the minimum of all half-cell widths.

Action:

Consider the physical concept you are trying to imply in the simulation. Increase MD cell dimensions.

Message 494: error - overconstrained PMF units

PMF units are oveconstrained.

Action:

DL_POLY_4 algorithms cannot handle overconstrained PMF units. Decrease the number of constraints on the PMFs.

Message 497: error - pmf_quench failure

Action:

See notes on message 515 below.

Message 498: error - shake algorithm (pmf_shake) failed to converge

Action:

See action notes on message 515 below.

Message 499: error - rattle algorithm (pmf_rattle) failed to converge

See notes on message 515 below.

Action:

See action notes on message 515 below.

Message 500: error - PMF unit of zero length is not permitted

PMF unit of zero length is found in FIELD. PMF units are either a single atom or a group of atoms usually forming a chemical molecule.

Action:

Correct the erroneous entries in FIELD.

Message 501: error - coincidence of particles in PMF unit

A PMF unit must be constituted of non-repeating particles!

Action:

Correct the erroneous entries in FIELD.

Message 502: error - PMF unit member found to be present more than once

A PMF unit is a group of unique (distingushed) atoms/sites. No repetition of a site is allowed in a PMF unit.

Action:

Correct the erroneous entries in FIELD.

Message 504: error - cutoff too large for TABLE||TABBND file

The requested cutoff exceeds the information in the TABLE file or the TABBND cutoff is larger than half the system cutoff `rcut`.

Action:

In the case when this is received while reading TABLE, reduce the value of the vdw cutoff (`rvdw`) in the CONTROL file or reconstruct the TABLE file. In the case when this is received while reading TABBND then specify a larger `rcut` in CONTROL.

Message 505: error - EAM metal densities or pair crossfunctions out of range

The resulting densities or pair crossfunctions are not defined in the TABEAM file.

Action:

Recreate a TABEAM file with wider interval of defined densities and pair cross functions.

Message 506: error - EAM or MBPC metal densities out of range

The resulting densities are not defined in the TABEAM file if EAM is used or ill defined due to atoms nearly overlapping when MBPC metal potential is in use..

Action:

Recreate a TABEAM file with wider range of densities.

Message 507: error - metal density embedding out of range

In the case of EAM type of metal interactions this indicates that the electron density of a particle in the system has exceeded the limits for which the embedding function for this particle's type is defined (as supplied in TABEAM. In the case of Finnis-Sinclair type of metal interactions, this indicates that the density has become negative.

Action:

Reconsider the physical sanity and validity of the metal interactions in your system and this type of simulation. You MUST change the interactions' parameters and/or the way the physical base of your investigation is handled in MD terms.

Message 508: error - EAM metal interaction entry in TABEAM unspecified in FIELD

The specified EAM metal interaction entry found in TABEAM is not specified in FIELD.

Action:

For N metal atom types there are $(5N + N^2)/2$ EAM functions in the TABEAM file. One density (N) and one embedding (N) function for each atom type and $(N + N^2)/2$ cross-interaction functions. Fix the table entries and resubmit.

Message 509: error - duplicate entry for a pair interaction detected in TABEAM

A duplicate cross-interaction function entry is detected in the TABEAM file.

Action:

Remove all duplicate entries in the TABEAM file and resubmit.

Message 510: error - duplicate entry for a density function detected in TABEAM

A duplicate density function entry is detected in the TABEAM file.

Action:

Remove all duplicate entries in the TABEAM file and resubmit.

Message 511: error - duplicate entry for an embedding function detected in TABEAM

A duplicate embedding function entry is detected in the TABEAM file.

Action:

Remove all duplicate entries in the TABEAM file and resubmit.

Message 512: error - non-definable vdw/dpd interactions detected in FIELD

A VDW corss-interaction was uncpecified and recovering it by using a mixing rule proved impossible due to type difference of the single species potentials.

Action:

Rethink your FIELD file interactions before restarting the job with a new compatible FIELD and possibly CONTROL file.

Message 513: error - particle assigned to non-existent domain in read_config

This can only happen if particle coordinates do not match the cell parameters in CONFIG. Probably, due to negligence or numerical inaccuracy inaccuracy in generation of big supercell from a small one.

Action:

Make sure lattice parameters and particle coordinates marry each other. Increase accuracy when generating a supercell.

Message 514: error - allowed image conventions are: 0, 1, 2, 3 and 6

DL_POLY_4 has found unsupported boundary condition specified in CONFIG.

Action:

Correct your boundary condition or consider using DL_POLY_Classic.

Message 515: error - rattle algorithm (constraints_rattle) failed to converge

The RATTLE algorithm for bond constraints is iterative. If the maximum number of permitted iterations is exceeded, the program terminates. Possible causes include: incorrect force field specification; too high a temperature; inconsistent constraints (over-constraint) etc..

Action:

You may try to increase the limit of iteration cycles in the constraint subroutines by using the directive **mxshak** and/or decrease the constraint precision by using the directive **shake** in CONTROL. But the trouble may be much more likely to be cured by careful consideration of the physical system being simulated. For example, is the system stressed in some way? Too far from equilibrium?

Message 517: error - allowed configuration information levels are: 0, 1 and 2

DL_POLY_4 has found an erroneous configuration information level, $l : 0.le.l.le.2$, (i) for the trajectory option in CONTROL or (ii) in the header of CONFIG.

Action:

Correct the error in CONFIG and rerun.

Message 518: error - control distances for variable timestep not intact

DL_POLY_4 has found the control distances for the variable timestep algorithm to be in contention with each other.

Action:

`mxdis` MUST BE $>$ 2.5$\times$ `mndis`. Correct in CONTROL and rerun.

Message 519: error - REVOLD is incompatible or does not exist

Either REVOLD does not exist or its formatting is incompatible.

Action:

Change the `restart` option in CONTROL and rerun.

Message 520: error - domain decomposition failed

A DL_POLY_4 check during the domain decomposition mapping has been violated. The number of nodes allowed for imcon = 0 is only 1,2,4 and 8! The number of nodes allowed for imcon = 6 is restricted to 2 along the z direction! The number of nodes should not be a prime number since these are not factorisable/decomposable!

Action:

You must ensure DL_POLY_4 execution on a number of processors that complies with the advise above.

Message 530: error - pseudo thermostat thickness MUST comply with: 2 Angs <= thickness < a quarter of the minimum MD cell width

DL_POLY_4 has found a check violated while reading CONTROL.

Action:

Correct accordingly in CONTROL and resubmit.

Message 540: error - pseudo thermostat MUST only be used in bulk simulations, i.e. imcon MUST be 1, 2 or 3

DL_POLY_4 has found a check violated while reading CONTROL.

Action:

Correct accordingly in CONTROL nve or in CONFIG (`imcon`) and resubmit.

Message 551: error - REFERENCE not found !!!

The `defect` detection option is used in conjunction with `restart` but no REFERENCE file is found.

Action:

Supply a REFERENCE configuration.

Message 552: error - REFERENCE must contain cell parameters !!!

REFERENCE MUST contain cell parameters i.e. image convention MUST be **imcon** = 1, 2, 3 or 6.

Action:

Supply a properly formatted REFERENCE configuration.

Message 553: error - REFERENCE is inconsistent !!!

An atom has been lost in transfer between nodes. This should never happen!

Action:

Big trouble. Report problem to authors immediately.

Message 554: error - REFERENCE's format different from CONFIG's !!!

REFERENCE complies to the same rules as CONFIG with the exception that image convention MUST be **imcon** = 1, 2, 3 or 6.

Action:

Supply a properly formatted REFERENCE configuartion.

Message 555: error - particle assigned to non-existent domain in defects_read_reference

See notes on message 513 above.

Action:

See action notes on message 513 above.

Message 556: error - too many atoms in REFERENCE file

See notes on message 45 above.

Action:

See action notes on message 45 above.

Message 557: error - undefined direction passed to defects_reference_export

See notes on message 42 above.

Action:

See action notes on message 42 above.

Message 558: error - outgoing transfer buffer exceeded in defects_reference_export

See notes on message 38 above.

Action:

See action notes on message 38 above.

Message 559: error - incoming data transfer size exceeds limit in defects_reference_export

See notes on message 38 above.

Action:

See action notes on message 38 above.

Message 560: error - rdef found to be > half the shortest interatomic distance in REFERENCE

The defect detection option relies on a cutoff, rdef, to define the vicinity around a site (defined in REFERENCES) in which a particle can claim to occupy the site. Evidently, rdef MUST be < half the shortest interatomic distance in REFERENCE.

Action:

Decrease the value of rdef at directive **defect** in CONTROL.

Message 570: error - unsupported image convention (0) for system expansion option nfold

System expansion is possible only for system with periodicity on their boundaries.

Action:

Change the image convention in CONFIG to any other suitable periodic boundary condition.

Message 580: error - replay (HISTORY) option can only be used for structural property recalculation

No structural property has been specified for this option to activate itself.

Action:

In CONTROL specify properties for recalculation (RDFs,z-density profiles, defect detection) or alternatively remove the option.

Message 585: error - end of file encountered in HISTORY file

This means that the HISTORY file is incomplete in some way: Either should you abort the replay (HISTORY) option or provide a fresh HISTORY file before restart.

Action:

In CONTROL specify properties for recalculation (RDFs,z-density profiles, defect detection) or alternatively remove the option.

Message 590: error - uknown minimisation type, only ”force”, ”energy” and ”distance” are recognised

Configuration minimisation can take only these three criteria.

Action:

In CONTROL specify the criterion you like followed by the needed arguments.

Message 600: error - ”impact” option specified more than once in CONTROL

Only one instance of the ”impact” option is allowed in CONTROL.

Action:

Remove any extra instances of the "impact" option in CONTROL.

Message 610: error - "impact" applied on particle that is either frozen, or the shell of a core-shell unit or part of a RB

It is the user's responsibility to ensure that impact is initiated on a "valid" particle.

Action:

In CONTROL remove the "impact" directive or correct the particle identity in it so that it complies with the requirements.

Message 620: error - duplicate or mixed intra-molecular entries specified in FIELD

The FIELD parser has detected an inconsistency in the description of bonding interactions. It is the user's responsibility to ensure that no duplicate or mixed-up intra-molecular entries are specified in FIELD.

Action:

Look at the preceding warning message in OUTPUT and find out which entry of what intra-molecular-like interaction is at fault. Correct the bonding description and try running again.

Message 625: error - only one *rigid* directive per molecule is allowed

DL_POLY_4 has found more than one rigids entry per molecule in FIELD.

Action:

Correct the erroneous part in FIELD and resubmit.

Message 630: error - too many rigid body units specified

This should never happen! This indicates an erroneous FIELD file or corrupted DL_POLY_4 executable. Unlike DL_POLY_Classic, DL_POLY_4 does not have a set limit on the number of rigid body types it can handle in any simulation (this is not the same as the total number of RBs in the system or per domain).

Action:

Examine FIELD for erroneous directives, correct and resubmit.

Message 632: error - rigid body unit MUST have at least 2 sites

This is likely to be a corrupted FIELD file.

Action:

Examine FIELD for erroneous directives, correct and resubmit.

Message 634: error - rigid body unit MUST have at least one non-massless site

No RB dynamics is possible if all sites of a body are massless as no rotational inertia can be defined!

Action:

Examine FIELD for erroneous directives, correct and resubmit.

Message 638: error - coincidence of particles in rigid body unit

This indicates a corrupted FIELD file as all members of a RB unit must be destinguishable from one another.

Action:

Examine FIELD for erroneous directives, correct and resubmit.

Message 640: error - too many rigid body units per domain

DL_POLY_4 limits the number of rigid body units in the system to be simulated (actually, the number to be processed by each node) and checks for the violation of this. Termination will result if the condition is violated.

Action:

Use **densvar** option in CONTROL to increase `mxrgd` (alternatively, increase it by hand in SET_BOUNDS and recompile) and resubmit.

Message 642: error - rigid body unit diameter > rcut (the system cutoff)

DL_POLY_4 domain decomposition limits the size of a RB to a largest diagonal < system cutoff. I.e. the largest RB type is still within a linked cell volume.

Action:

Increase cutoff.

Message 644: error - overconstrained rigid body unit

This is a very unlikely message which usually indicates a corrupted FIELD file or unphysically overconstrained system.

Action:

Decrease constraint on the system. Examine FIELD for erroneous directives, if any, correct and resubmit.

Message 646: error - overconstrained constraint unit

This is a very unlikely message which usually indicates a corrupted FIELD file or unphysically overconstrained system.

Action:

Decrease constraint on the system. Examine FIELD for erroneous directives, if any, correct and resubmit.

Message 648:error - quaternion setup failed

This error indicates that the routine Q_SETUP has failed in reproducing all the atomic positions in rigid units from the centre of mass and quaternion vectors it has calculated.

Action:

Check the contents of the CONFIG file. DL_POLY_4 builds its local body description of a rigid unit type from the *first* occurrence of such a unit in the CONFIG file. The error most likely occurs because subsequent occurrences were not sufficiently similar to this reference structure. If the problem persists increase the value of `tol` in Q_SETUP and recompile. If problems still persist double the value of `dettest` in RIGID_BODIES_SETUP and recompile. If you still encounter problems contact the authors.

Message 650:error - failed to find principal axis system

This error indicates that the routine RIGID_BODIES_SETUP has failed to find the principal axis for a rigid unit.

Action:

This is an unlikely error. DL_POLY_4 should correctly handle linear, planar and 3-dimensional rigid units. There is the remote possibility that the unit has all of its mass-bearing particles frozen while some of the massless are not or the unit has just one mass-bearing particle. Another, more likely, possibility, in case of linear molecules is that the precision of the coordinates of these linear molecules' constituentsi, as produced by the user, is not good enough, which leads DL_POLY_4 to accepting it as non-linear while, in fact, it is and then failing at the current point. It is quite possible, despite considered as wrong practice, that the user defined system of linear RBs is, in fact, generated from a system of CBs (3 per RB) which has not been run in a high enough SHAKE/RATTLE tolerance accuracy (10ˆ-8 and higher may be needed). Check the definition of the rigid unit in the CONFIG file, if sensible report the error to the authors.

Message 655: error - FENE bond breaking failure

A FENE type bond was broken.

Action:

Examine FIELD for erroneous directives, if any, correct and resubmit.

Message 660: error - TABBND or PDF bond breaking failure

A bond with potential defined in TABBND or for which intra-molecular potential distribution is collected has exceeded its bondleghth limit.

Action:

If there is a TABBND present, reconstruct TABBND with potentials defined over larger cutoff and try again. If bonds PDF are collected, increase their cutoff value in CONTROL.

Message 1000: error - working precision mismatch between FORTRAN90 and MPI implementation

DL_POLY_4 has failed to match the available modes of MPI precision for real numbers to the defined in sc kinds_f90 FORTRAN90 working precision `wp` for real numbers. `wp` is a precompile parameter.

Action:

This simply mean that `wp` must have been changed from its original value to something else and the new value is not matched by the `mpi_wp` variable in COMMS_MODULE. It is the user's responsibility to ensure that `wp` and `mpi_wp` are compliant. Make the necessary corrections to sc kinds_f90 and/or COMMS_MODULE.

Message 1001: error - allocation failure in comms_module $->$ gcheck_vector

DL_POLY_4 has failed to find available memory to allocate an array or arrays, i.e. there is lack of sufficient memory (per node) on the execution machine.

Action:

This may simply mean that your simulation is too large for the machine you are running on. Consider this before wasting time trying a fix. Try using more processing nodes if they are available. If this is not an option investigate the possibility of increasing the heap size for your application. Talk to your systems support people for advice on how to do this.

Message 1002: error - deallocation failure in comms_module − > gcheck_vector

DL_POLY_4 has failed to deallocate an array or arrays, i.e. to free memory that is no longer in use.
Action:
Talk to your systems support people for advice on how to manage this.

Message 1003: error - allocation failure in comms_module − > gisum_vector

See notes on message 1001 above.
Action:
See action notes on message 1001 above.

Message 1004: error - deallocation failure in comms_module − > gisum_vector

See notes on message 1002 above.
Action:
See action notes on message 1002 above.

Message 1005: error - allocation failure in comms_module − > grsum_vector

See notes on message 1001 above.
Action:
See action notes on message 1001 above.

Message 1006: error - deallocation failure in comms_module − > grsum_vector

See notes on message 1002 above.
Action:
See action notes on message 1002 above.

Message 1007: error - allocation failure in comms_module − > gimax_vector

See notes on message 1001 above.
Action:
See action notes on message 1001 above.

Message 1008: error - deallocation failure in comms_module − > gimax_vector

See notes on message 1002 above.
Action:
See action notes on message 1002 above.

Message 1009: error - allocation failure in comms_module − > grmax_vector

See notes on message 1001 above.

Action:

See action notes on message 1001 above.

Message 1010: error - deallocation failure in comms_module − > grmax_vector

See notes on message 1002 above.

Action:

See action notes on message 1002 above.

Message 1011: error - allocation failure in parse_module − > get_record

See notes on message 1001 above.

Action:

See action notes on message 1001 above.

Message 1012: error - deallocation failure in parse_module − > get_record

See notes on message 1002 above.

Action:

See action notes on message 1002 above.

Message 1013: error - allocation failure in angles_module − > allocate_angles_arrays

See notes on message 1001 above.

Action:

See action notes on message 1001 above.

Message 1014: error - allocation failure in bonds_module − > allocate_bonds_arrays

See notes on message 1001 above.

Action:

See action notes on message 1001 above.

Message 1015: error - allocation failure in core_shell_module − > allocate_core_shell_arrays

See notes on message 1001 above.

Action:

See action notes on message 1001 above.

Message 1016: error - allocation failure in statistics_module − > allocate_statitics_arrays

See notes on message 1001 above.

Action:

See action notes on message 1001 above.

Message 1017: error - allocation failure in tethers_module − > allocate_tethers_arrays

See notes on message 1001 above.

Action:

See action notes on message 1001 above.

Message 1018: error - allocation failure in constraints_module − > allocate_constraints_arrays

See notes on message 1001 above.

Action:

See action notes on message 1001 above.

Message 1019: error - allocation failure in external_field_module − > allocate_external_field_arrays

See notes on message 1001 above.

Action:

See action notes on message 1001 above.

Message 1020: error - allocation failure in dihedrals_module − > allocate_dihedrals_arrays

See notes on message 1001 above.

Action:

See action notes on message 1001 above.

Message 1021: error - allocation failure in inversions_module − > allocate_inversion_arrays

See notes on message 1001 above.

Action:

See action notes on message 1001 above.

Message 1022: error - allocation failure in vdw_module − > allocate_vdw_arrays

See notes on message 1001 above.

Action:

See action notes on message 1001 above.

Message 1023: error - allocation failure in metal_module − > allocate_metal_arrays

See notes on message 1001 above.

Action:

See action notes on message 1001 above.

Message 1024: error - allocation failure in three_body_module − > allocate_three_body_arrays

See notes on message 1001 above.

Action:

See action notes on message 1001 above.

Message 1025: error - allocation failure in config_module − > allocate_config_arrays

See notes on message 1001 above.

Action:

See action notes on message 1001 above.

Message 1026: error - allocation failure in site_module − > allocate_site_arrays

See notes on message 1001 above.

Action:

See action notes on message 1001 above.

Message 1027: error - allocation failure in tersoff_module − > alocate_tersoff_arrays

See notes on message 1001 above.

Action:

See action notes on message 1001 above.

Message 1028: error - deallocation failure in angles_module − > deallocate_angles_arrays

See notes on message 1002 above.

Action:

See action notes on message 1002 above.

Message 1029: error - deallocation failure in bonds_module − > deallocate_bonds_arrays

See notes on message 1002 above.

Action:

See action notes on message 1002 above.

Message 1030: error - deallocation failure in core_shell_module − > deallocate_core_shell_arrays

See notes on message 1002 above.

Action:

See action notes on message 1002 above.

Message 1031: error - deallocation failure in tethers_module − > deallocate_tethers_arrays

See notes on message 1002 above.

Action:

See action notes on message 1002 above.

Message 1032: error - deallocation failure in constraints_module − > deallocate_constraints_arrays

See notes on message 1002 above.

Action:

See action notes on message 1002 above.

Message 1033: error - deallocation failure in dihedrals_module − > deallocate_dihedrals_arrays

See notes on message 1002 above.

Action:

See action notes on message 1002 above.

Message 1034: error - deallocation failure in inversions_module − > deallocate_inversions_arrays

See notes on message 1002 above.

Action:

See action notes on message 1002 above.

Message 1035: error - allocation failure in defects_module − > allocate_defects_arrays

See notes on message 1001 above.

Action:

See action notes on message 1001 above.

Message 1036: error - allocation failure in pmf_module − > allocate_pmf_arrays

See notes on message 1001 above.

Action:

See action notes on message 1001 above.

Message 1037: error - deallocation failure in pmf_module − > deallocate_pmf_arrays

See notes on message 1002 above.

Action:

See action notes on message 1002 above.

Message 1038: error - allocation failure in minimise_module − > allocate_minimise_arrays

See notes on message 1001 above.

Action:

See action notes on message 1001 above.

Message 1039: error - deallocation failure in minimise_module − > deallocate_minimise_arrays

See notes on message 1002 above.

Action:

See action notes on message 1002 above.

Message 1040: error - allocation failure in ewald_module − > ewald_allocate_kall_arrays

See notes on message 1001 above.

Action:

See action notes on message 1001 above.

Message 1041: error - allocation failure in langevin_module − > langevin_allocate_arrays

See notes on message 1001 above.

Action:

See action notes on message 1001 above.

Message 1042: error - allocation failure in rigid_bodies_module − > allocate_rigid_bodies_arrays

See notes on message 1001 above.

Action:

See action notes on message 1001 above.

Message 1043: error - deallocation failure in rigid_bodies_module − > deallocate_rigid_bodies_arrays

See notes on message 1002 above.

Action:

See action notes on message 1002 above.

Message 1044: error - allocation failure in comms_module − > gimin_vector

See notes on message 1001 above.

Action:

See action notes on message 1001 above.

Message 1045: error - deallocation failure in comms_module − > gimin_vector

See notes on message 1002 above.

Action:

See action notes on message 1002 above.

Message 1046: error - allocation failure in comms_module − > grmin_vector

See notes on message 1001 above.

Action:

See action notes on message 1001 above.

Message 1047: error - deallocation failure in comms_module − > grmin_vector

See notes on message 1002 above.

Action:

See action notes on message 1002 above.

Message 1048: error - error - allocation failure in comms_module − > grsum_matrix

See notes on message 1001 above.

Action:

See action notes on message 1001 above.

Message 1049: error - deallocation failure in comms_module − > grsum_matrix

See notes on message 1002 above.

Action:

See action notes on message 1002 above.

Message 1050: error - sorted I/O base communicator not set

Possible corruption if IO_MODULE. This should never happen!

Action:

Make sure you have a clean copy of DL_POLY_4, compiled without any suspicious warning messages. Contact authors if the problem persists.

Message 1053: error - sorted I/O allocation error

Your I/O buffer (and possibly batch) size is too big.

Action:

Decrease the value of the I/O buffer (and possibly batch) size in CONTROL and restart your job.

Message 1056: error - unkown write option given to sorted I/O

This should never happen!

Action:

Contact authors if the problem persists.

Message 1059: error - unknown write level given to sorted I/O

This should never happen!

Action:

Contact authors if the problem persists.

Message 1060: error - allocation failure in statistics_module -¿ allocate_statitics_connect

See notes on message 1001 above.

Action:

See action notes on message 1001 above.

Message 1061: error - allocation failure in statistics_module -¿ deallocate_statitics_connect

See notes on message 1001 above.

Action:

See action notes on message 1001 above.

Message 1063: error - allocation failure in vdw_module − > allocate_vdw_table_arrays

See notes on message 1001 above.

Action:

See action notes on message 1001 above.

Message 1066: error - allocation failure in vdw_module − > allocate_vdw_direct_fs_arrays

See notes on message 1001 above.

Action:

See action notes on message 1001 above.

Message 1069: error - allocation failure in metal_module − > allocate_metal_table_arrays

See notes on message 1001 above.

Action:

See action notes on message 1001 above.

Message 1070: error - allocation failure in ewald_module − > ewald_allocate_kfrz_arrays

See notes on message 1001 above.

Action:

See action notes on message 1001 above.

Message 1072: error - allocation failure in bonds_module -¿ allocate_bond_pot_arrays

See notes on message 1001 above.

Action:

See action notes on message 1001 above.

Message 1073: error - allocation failure in bonds_module -¿ allocate_bond_dst_arrays

See notes on message 1001 above.

Action:

See action notes on message 1001 above.

Message 1074: error - allocation failure in angles_module -¿ allocate_angl_pot_arrays

See notes on message 1001 above.

Action:

See action notes on message 1001 above.

Message 1075: error - allocation failure in angles_module -¿ allocate_angl_dst_arrays

See notes on message 1001 above.

Action:

See action notes on message 1001 above.

Message 1076: error - allocation failure in dihedrals_module -¿ allocate_dihd_pot_arrays

See notes on message 1001 above.

Action:

See action notes on message 1001 above.

Message 1077: error - allocation failure in dihedrals_module -¿ allocate_dihd_dst_arrays

See notes on message 1001 above.

Action:

See action notes on message 1001 above.

Message 1078: error - allocation failure in inversions_module -¿ allocate_invr_pot_arrays

See notes on message 1001 above.

Action:

See action notes on message 1001 above.

Message 1079: error - allocation failure in inversions_module -¿ allocate_invr_dst_arrays

See notes on message 1001 above.

Action:

See action notes on message 1001 above.

Message 1080: error - allocation failure in greenkubo_module -¿ allocate_greenkubo_arrays

See notes on message 1001 above.

Action:

See action notes on message 1001 above.

Message 1081: error - allocation failure in dpd_module -¿ allocate_dpd_arrays

See notes on message 1001 above.

Action:

See action notes on message 1001 above.

Appendix E

DL_POLY_4 README wisdom

E.1 README.txt

```
DL_POLY_4.07
============

The source is in fully self-contained free formatted FORTRAN90+MPI2
code (specifically FORTRAN90 + TR15581 + MPI1 + MPI-I/O only).  The
available NetCDF functionality makes the extended code dependent upon
it.  The non-extended code complies with the NAGWare and FORCHECK F90
standards with exception of the FORTRAN2003 feature TR15581, which is
very rarely unavailable in the nowadays FORTRAN90/95 compilers.

This version supports ALL features that are available in the
standard DL_POLY_Classic version with the exceptions of:
  (1) Rigid bodies (RBs) linked via (i) constraint bonds (CBs) or
      (ii) potential of mean field constraints (PMFs).
  (2) Truncated octahedral (imcon = 4), Rhombic Dodecahedral
      (imcon = 5) and Hexagonal Prism (imcon = 7) periodic boundary
      MD cell conventions (PBC).  Note that the last one is easily
      convertible to orthorhombic (imcon = 2), see utility/nfold.f90.
  (3) Classic Ewald and Hautman-Klein Ewald Coulomb evaluations.
  (4) Hyper-Dynamics: Temperature Accelerated Dynamics, Biased
      Potential Dynamics; Meta-Dynamics and solvation features.

No previous DL_POLY_3/4 feature is deprecated.  ALL NEW features are
documented in the "DL_POLY_4 User Manual".

Reference
---------
Thank you for using the DL_POLY_4 package in your work.  Please,
acknowledge our efforts by including the following reference when
publishing data obtained using DL_POLY_4: "I.T. Todorov, W. Smith,
K. Trachenko & M.T. Dove, J. Mater. Chem., 16 (20), 1911-1918 (2006)".

Warnings:
---------
  (1) DL_POLY_4 can produce index ordered REVCON, HISTORY and MSDTMP
```

```
      files which are restartable by DL_POLY_Classic.  Although such
      printed outputs look unscrambled, the actual printing process
      is not.  Unscrambled printing is slightly more expensive than
      natural (scrambled) printing.  The cost time-wise is little,
      < 1%, but HD space-wise is approximately 20%.  This is due to
      the necessary addition of blanks at the end of data record,
      included to align the (ASCII) lines of output files (human
      readable) to a constant length.  Printing scrambled outputs
      is optional.  Note that these too have blanks aligned records.
      The parallel I/O ensures (i) writing speeds of 10^5 to 10^6
      particle per second with optimal number of writers and (ii)
      reading speeds of 10^4 to 10^5 particles per second per reader.
      For more information on I/O options consult the user manual.
  (2) REVIVE files produced by different versions are not compatible.
      Furthermore, restarting runs across different sub-versions
      may not be possible.
  (3) The DL_POLY_4 parallel performance and efficiency are considered
      very-good-to-excellent as long as (i) all CPU cores are loaded
      with no less than 500 particles each and (ii) the major linked
      cells algorithm has no dimension less than 4.
  (4) Although DL_POLY_4 can be compiled in a serial mode, users are
      advised to consider DL_POLY_Classic as a suitable alternative to
      DL_POLY_4 when simulations are likely to be serial jobs for
      systems containing < 500 particles-per-processor.  In such
      circumstances, with both codes compiled in serial mode, the
      difference in performance, measured by the time-per-timestep
      ratio [DL_POLY_Classic(t)-DL_POLY_4(t)]/DL_POLY_Classic(t),
      varies in the range -5:+5%.  This variation depends strongly on
      the system force-field complexity and very weakly on the system
      size.
  (5) The ‘rpad’ & ‘no strict’ CONTROL options should be used with
      care especially in conjunction with the ‘variable timestep‘
      option when iterative algorithms are present in the simulation.
      Such may be driven by a combination of options such as:
      ‘minimise’, ‘ensemble npt’, ‘ensemble nst’ in the presence of
      constraints (associated with the ‘tolerance’ and ‘mxshake’
      options) and/or core shells units (dealt by the relaxed shell
      model and associated with the ‘rlxtol’ option) in the model
      system as defined in the FIELD file.

Integration Defaults:
---------------------
The default ensemble is NVE.

The default integration scheme is Trotter derived Velocity Verlet
(VV), although Leapfrog Verlet (LFV) is also available.  VV is
considered superior (to LFV) since:
  (1) Integration can be developed in symplectic manner for certain
      ensembles, such as: NVE, NVEk (NVT Evans) as well as all
      Nose-Hoover ensembles (NVT, & NPT & NsT when there is no
      external field applied on the system, otherwise they do not
      conserve the phase space volume) and MTK ensembles (NPT & NsT).
  (2) All ensemble variables are updated synchronously and
      thermodynamic quantities and estimators are exact at the every
      step, whereas in LFV particle velocities and thermostat and
      barostat friction velocities are half an integration time-step
      behind the rest of the ensemble variables and due to this
      certain estimators are approximated at full timestep.
  (3) It offers better numerical stability and faster convergence
      when (i) constraint solvers (CB/PMF: RATTLE/VV versus SHAKE/LFV)
      are involved and/or (ii) RB dynamics is integrated.

The LFV integration may take less cpu time than the VV one for the
certain ensembles - type of system (CB/PMF/RB) and type of ensemble
dependent.  Usually, LFV is slightly faster than VV when CB/PMF/RB
are present in the system.  The relative performance between the LVF
and VV integration (per timestep) is observed to vary in the limits
*** [LFV(t)-VV(t)]/VV(t) = -5:+5% ***.  However, the VV algorithms
treat CB/PMF/RB entities in a more precise (symplectic) manner than
the LFV ones and thus not only have better numerical stability but
also produce more accurate dynamics.

Makefiles & compilation
-----------------------
From within the ‘source’ directory the user may compile the code by
selecting the appropriate Makefile from the ‘build’ directory and
copying it across by typing at the command line:

   cp ../build/Makefile_MPI Makefile

- intended for parallel execution on multi-processor platforms (an MPI
  implementation is needed), or

   cp ../build/Makefile_SRLx Makefile

- intended for serial execution (no MPI required)

followed by <Enter>.

Note that in ‘comms_module.f90’ it is crucial that line 13 reads as:
‘Use mpi_module’    - for serial compilation or as
‘Use mpi’           - for parallel compilation (which is the default).

If the parallel OS environment, one is compiling on, is not fully F90
compatible then the ‘Use mpi’ entry in the ‘comms_module.f90’ file may
be interpreted as erroneous.  This is easily overcome by commenting out
the ‘Use mpi’ line and uncommenting the ‘Include ’mpif.h’’ one situated
immediately after the ‘Implicit None’ line.

If there is an ‘entry’ in the Makefile for the particular combination
of architecture, compiler & MPI implementation, then the user may
instantiate the compilation by issuing at the command line:

   make ‘entry’
```

```
and then pressing <Enter>.

Usually the one named ‘hpc’ is suitable for the majority of platforms.
To find out the keywords for all available entries within the Makefile
issue:

   make

press <Enter> and then examine the Makefile entries corresponding to
the keywords reported.  If there is not a suitable entry then you
should seek advice from a computer scientist or the support staff of
the particular machine (HPC service).

The necessary components for the source compilation are:
  (1) a FORTRAN90 + TR15581 compliant compiler (if the full PATH to it
      is not passed to the DEFAULT ENVIRONMENT PATH, then it MUST be
      explicitly supplied in the Makefile)
  (2) an MPI2 (or MPI1 + MPI-I/O) implementation, COMPILED for the
      architecture/OS and the targeted compiler (if the full PATH to
      these is not passed to the DEFAULT ENVIRONMENT PATH, then it MUST
      be explicitly supplied in the Makefile)
  (3) a MAKE command (Makefile interpreter in the system SHELL)
where (2) is not necessary for compilation in SERIAL mode!

Note that for the TR15581 compliance, gfortran requires a gcc version
4.2.0 or above!

By default, if the compilation process is successful then an executable
(build) will be placed in ‘../execute’ directory (at the same level as
the ‘source’ directory where the code is compiled).  Should the
‘../execute’ directory not exist then it will be created automatically
by the Makefile script.  The build may then be moved, renamed, etc. and
used as the user wishes.  However, when executed, the program will look
for input files in the directory of execution!

Compilation on Windows
----------------------
The best way to get around it is to install cygwin on the system
(http://www.cygwin.com/) to emulate a UNIX/Linux like environment
and then use the "make" command.  During the cygwin installation please
make sure that the "make" and "gfortran" components are specifically
opted for components (as they may not be included as default ones) in
the install.  A potential problem for Windows based FORTRAN compilers,
one may encounter, is that the compiler may not pick symbolic links.
This can be resolved by substituting the soft links with hard in the
Makefile.  For parallel compilation all "openMPI" components must also
be opted for during the cygwin install!!!

Compiling with NetCDF functionality
-----------------------------------
The targeted Makefile needs the following substitution within before
```

```
attempting compilation:

"netcdf_modul~.o -> netcdf_module.o"

Note that suitable entry may need to be created within the Makefile
so that it matches the particular combination of architecture/OS,
compiler, MPI library & netCDF library.

Contacts at STFC Daresbury Laboratory
-------------------------------------
Dr. I.T. Todorov :: ilian.todorov@stfc.ac.uk
```

E.2 README_KIM.txt

```
KIM API Installation Instructions:
----------------------------------

0) Presumptions

Linux OS (for wget downloads)
internet
starting in the DL_POLY_4.XX main directory

1) Download the kim-api-v1.6.X.tgz

% wget http://s3.openkim.org/kim-api/kim-api-v1.6.X.tgz

2) Unpack the tarball in the DL_POLY_4.ZZ source directory

% mv kim-api-v1.6.X.tgz DL_POLY_4.ZZ/source/
% cd DL_POLY_4.ZZ/source
% tar xzvf kim-api-v1.6.X

3) Download and unpack the testing Model and its Model Driver

% cd kim-api-v1.6.X/src/model_drivers
% wget http://www.aem.umn.edu/~elliott/Pair_Morse__MD.tgz
% tar xzvf Pair_Morse__MD.tgz
% cd ../models
% wget http://www.aem.umn.edu/~elliott/Pair_Morse_GirifalcoWeizer_LowCutoff_Cu__MO.tgz
% tar xzvf Pair_Morse_GirifalcoWeizer_LowCutoff_Cu__MO.tgz
% cd ../../

4) Configure the openkim-api package according to the instructions in INSTALL

% cp Makefile.KIM_Config.example Makefile.KIM_Config
% vi Makefile.KIM_Config  # Edit file as appropriate.
                          # In particular, change the
                          # KIM_DIR variable.
% make
```

```
5) Copy the DL_POLY_4.ZZ makefile and compile

% cd ..
% cp ../build/Makefile_SRL2_KIM Makefile
        % vi Makefile # Edit the KIM_API_DIR variable as appropriate.
% make gnu
% cd ../execute
% ls -haltr   # The last listed file should be DLPOLY.Z
              # and have the current date stamp

  Similarly, one may build DL_POLY_4 in parallel by copying
  ../build/Makefile_MPI_KIM and compiling accordingly.

Test & Verify KIM:
------------------

1) Obtain DL_POLY_4 TEST21 following the README.txt advice from the DL_POLY_4.ZZ/data directory

% cd ../data
% ftp ftp.dl.ac.uk
ftp username: anonymous
ftp password: email@google.com        # enter your email
ftp> cd ccp5/DL_POLY/DL_POLY_4.0/DATA
ftp> binary                            # transfer mode
ftp> get TEST21.tar.gz
ftp> quit
% tar xzvf TEST21.tar.gz

2) Move DL_POLY input files to the DL_POLY_4.ZZ/execute directory

% cd ../execute
% cp ../data/TEST21/CONFIG .
% cp ../data/TEST21/CONTROL .
% cp ../data/TEST21/FIELD .

3) Edit FIELD and change

metal 1 table
Cu      Cu      eam
close

   to

vdw 1
Cu      Cu      mors 3.42900e-01 2.86600e+00 1.35880e+00
close

4) Edit CONTROL and:

   remove

restart scale
```

```
   change

steps               6000     steps
equilibration       5000     steps

   to

steps                100     steps
equilibration         10     steps

   and

cutoff                 5.0   Angstroms

   to

cutoff                 6.76342e+00   Angstroms

5) Run and save the "native" simulation

% ./DLPOLY.Z
% mv OUTPUT OUTPUT.native

6) Edit FILED and change

vdw 1
Cu      Cu      mors 3.42900e-01 2.86600e+00 1.35880e+00
close

   to

kim  Pair_Morse_GirifalcoWeizer_LowCutoff_Cu__MO

7) Run and save the "kim" simulation

% ./DLPOLY.Z
% mv OUTPUT OUTPUT.kim

8) Compare simulation outputs form the native and KIM interaction models

% diff OUTPUT.native OUTPUT.kim

Good luck!

January 2015
Ryan Elliott, Ilian Todorov and Henry Boateng
```

Appendix F

DL_POLY_4 Academic Licence Agreement

F.1 The short story

By registering for any version of DL_POLY_4 you confirm that:

- the institution whose name appears as the Licensee in the Registration Form agrees to the terms of the Licence Agreement;
- you have authority to agree to the terms of the Licence Agreement and to enter into a contract with STFC on behalf of that institution;
- the DL_POLY_4 Software will be used only for Academic Purposes (as defined in the Licence Agreement);
- you will acknowledge the use of DL_POLY_4 when publishing any data obtained with its assistance by including the following reference: ”I.T. Todorov, W. Smith, K. Trachenko & M.T. Dove, *J. Mater. Chem.*, **16**, 1911-1918 (2006)”;
- the information given on the Registration Form is true; and
- you agree to your personal data being used for the purpose of managing your institution’s licence of the DL_POLY_4 Software, including contacting you to tell you about changes to and error corrections for that software and generally about the software,

which becomes a contractual agreement between the Licensee and STFC.

STFC is an executive Non-Departmental Public Body established as a Research Council by Royal Charter under the Science and Technology Act 1965. Its address for this agreement is the Daresbury Laboratory, Keckwick Lane, Warrington, WA4 4AD, Cheshire, United Kingdom and its VAT number is GB 618 3673 25.

If you wish to contact STFC about the Licence Agreement or the DL_POLY_4 Software, please contact me, Dr. I.T. Todorov, by emailing to ilian.todorov@stfc.ac.uk.

F.2 The full version

```
                        LICENCE AGREEMENT

PLEASE READ THE LICENCE BELOW, COMPLETE THE REGISTRATION INFORMATION,
```

AND ACCEPT THE TERMS OF THE LICENCE IN ORDER TO DOWNLOAD THE SOFTWARE. BY CLICKING ACCEPT AND DOWNLOADING THE SOFTWARE YOU ARE CONFIRMING YOUR INTENTION TO ENTER INTO A LEGALLY BINDING AGREEMENT GOVERNED BY THE TERMS AND CONDITIONS SET OUT BELOW (THE AGREEMENT). IN THE FOLLOWING TERMS AND CONDITIONS YOU ARE REFERRED TO AS THE LICENSEE.

1. DEFINITIONS AND INTERPRETATION

1.1 In this Licence Agreement the following expressions have the meanings set opposite:

Academic Purposes	fundamental or basic research or academic teaching, including any fundamental research that is funded by any public or charitable body, but not any purpose that generates revenue (as opposed to grant income) for the Licensee or any third party. Any research that is wholly or partially sponsored by any profit making organisation or that is carried out for the benefit of any profit-making organisation is not an Academic Purpose;
a Derived Work	any modification of, or enhancement or improvement to, any of the DL_POLY_4 Software and any software or other work developed or derived from any of the DL_POLY_4 Software;
the DL_POLY_4 Software	the release and version of the DL_POLY_4 Software downloaded by the Licensee from the DL_POLY_4 Website immediately after the Licensee agrees to the terms and conditions of this Licence Agreement;
the DL_POLY_4 Website	the website with the URL http://www.ccp5.ac.uk/DL_POLY/ , and any website that from time to time replaces that website;
a Harmful Element	any virus, worm, time bomb, time lock, drop dead device, trap and access code or anything else that might disrupt, disable, harm or impede the operation of any information system, or that might corrupt, damage, destroy or render inaccessible any software, data or file on, or that may allow any unauthorised person to gain access to, any information system or any software, data or file on it;
Intellectual Property	patents, trade marks, service marks, registered designs, copyrights, database rights, design rights, know-how, confidential information,

	applications for any of the above, and any similar right recognised from time to time in any jurisdiction, together with all rights of action in relation to the infringement of any of the above;
the Licence Period	the period beginning when the Licensee agrees to the terms and conditions of this Licence Agreement and downloads the DL_POLY_4 Software from the DL_POLY_4 Website and ending on the termination of this Licence Agreement under clause 5.2.

2. LICENCE

2.1 STFC grants the Licensee an indefinite, non-exclusive, non-transferable, royalty free licence to use, copy, modify, and enhance the DL_POLY_4 Software during the Licence Period on the terms and conditions of this Licence Agreement provided that:

2.1.1 the Licensee may not distribute any of the DL_POLY_4 Software or any Derived Work to any third party, or share their use with any third party (whether free of charge or otherwise), and the Licensee may not sub-license the use of any of the DL_POLY_4 Software to any third party unless, in each case, that third party has complied with clause 2.3 below;

2.1.2 the Licensee may not use the DL_POLY_4 Software on behalf of, or for the benefit of, any third party (including, without limitation, using it to provide bureau, outsourcing or application services or facilities management services) party unless that third party has complied with clause 2.3 below; and

2.1.3 the DL_POLY_4 Software and any Derived Work may be used by the Licensee and its employees and registered students for Academic Purposes only.

2.2 If the Licensee wishes to use the DL_POLY_4 Software or any Derived Work in any way except for Academic Purposes, or wishes to distribute or make the DL_POLY_4 Software or any Derived Work available to any third party for non-Academic Purposes, it must obtain a commercial licence from STFC. STFC may refuse to grant the Licensee a commercial licence. If STFC agrees to grant a commercial licence, that licence will be on such terms and conditions as STFC sees fit.

2.3 If the Licensee wishes to carry out any collaboration for Academic Purposes with any third party and that third party needs to use the DL_POLY_4 Software in connection with that collaboration, or if the Licensee wishes to make the DL_POLY_4 Software available on-line to any third party for Academic Purposes, the Licensee must direct that third party to the DL_POLY_4 Website. That third party may use the DL_POLY_4 Software and any Derived Work (whether obtained from STFC or from the Licensee) only if it has completed the registration process on the DL_POLY_4 Website and agreed to the terms and conditions of the Licence Agreement for the use of the DL_POLY_4 Software for Academic Purposes that then appear on the DL_POLY_4 Website.

2.4 This Licence Agreement allows the Licensee to use only the release or version of the DL_POLY_4 Software downloaded by the Licensee from the DL_POLY_4 Website immediately after the Licensee agrees to the terms and conditions of this Licence Agreement; the Licensee must acquire a new licence for any future release or version of the DL_POLY_4 Software that it wishes to use.

2.5 The Licensee will not tamper with, or remove, any copyright or other proprietary notice or any disclaimer that appears on or in any part of the DL_POLY_4 Software, and will reproduce the same in all copies of any of the DL_POLY_4 Software and in all Derived Works.

2.6 The Licensee may not itself distribute any Derived Work. If the Licensee would like any Derived Work made by or for the Licensee, or by any of its employees or students, based on any of the DL_POLY_4 Software to be considered for wider distribution, it should notify STFC of such Derived Work and provide STFC with a copy of that Derived Work (in source code) (a Contribution). In providing STFC with such a Contribution the Licensee grants STFC an irrevocable, indefinite licence to make that Contribution available to any third party on such terms and conditions as STFC may from time to time decide. The Licensee warrants that it owns (or has the necessary rights to) such Contribution to give effect to this clause.

3. WARRANTIES AND LIABILITY

3.1 The DL_ POLY Software is provided for Academic Purposes free of charge. Therefore, STFC gives no warranty and makes no representation in relation to the DL_POLY_4 Software or any assistance or advice that STFC may give in connection with the DL_POLY_4 Software. The Licensee will indemnify STFC against any and all claims arising as a result of the Licensee having made any of the DL_POLY_4 Software or any Derived Work available to any third party.

3.2 Before using any of the DL_POLY_4 Software, the Licensee will check that the DL_POLY_4 Software does not contain any Harmful Element. STFC does not warrant that the DL_POLY_4 Software will run without interruption or be error free, or be free from any Harmful Element. STFC is not obliged to provide any support or error correction service, assistance or advice in relation to the DL_POLY_4 Software, but the Licensee may access any error corrections and on-line assistance that STFC chooses to make available on the

DL_POLY_4 Website from time to time. If STFC does provide that sort of service, assistance or advice, subject to clause 3.7, STFC will not be liable for any loss or damage suffered by the Licensee as a result.

3.3 STFC will not be liable to the Licensee to the extent that any loss or damage is caused: by the Licensee's failure to implement, or the Licensee's delay in implementing, any correction or advice in in relation to the DL_POLY_4 Software that STFC has made available on the DL_POLY_4 Website; or by the Licensee's failure to acquire a licence of and to implement any new release or version of the DL_POLY_4 Software that would have remedied or mitigated the effects of any error, defect, bug or deficiency in the DL_POLY_4 Software.

3.4 The Licensee acknowledges that proper use of the DL_POLY_4 Software and any Derived Work is dependent on the Licensee, its employees and students exercising proper skill and care in inputting data and interpreting the output provided by the DL_POLY_4 Software or that Derived Work. STFC will not be liable for the consequences of decisions taken by the Licensee or any other person on the basis of that output. STFC does not accept any responsibility for any use which may be made by the Licensee of that output, nor for any reliance which may be placed on that output, nor for advice or information given in connection with that output.

3.5 Subject to clause 3.7, STFC's liability or any breach of this Licence Agreement, any negligence or arising in any other way out of the subject matter of this Licence Agreement or the use of the DL_POLY_4 Software, will not extend to any incidental or consequential damages or losses, or any loss of profits, loss of revenue, loss of data, loss of contracts or opportunity, whether direct or indirect, even if the Licensee has advised STFC of the possibility of those losses arising or if they were or are within STFC's contemplation.

3.6 Subject to clause 3.7, the aggregate liability of STFC for any and all breaches of this Licence Agreement, any negligence or arising in any other way out of the subject matter of this Licence Agreement or the use of the DL_POLY_4 Software will not exceed in total 5,000.

3.7 Nothing in this Licence Agreement limits or excludes STFC's liability for death or personal injury caused by its negligence or for any fraud, or for any sort of liability that, by law, cannot be limited or excluded.

3.8 The express undertakings and given by STFC in this Licence Agreement and the terms of this Licence Agreement are in lieu of all warranties, conditions, terms, undertakings and obligations on the part of STFC, whether express or implied by statute, common law, custom, trade usage, course of dealing or in any other way.

```
    All of these are excluded to the fullest extent permitted by law.

4. INTELLECTUAL PROPERTY RIGHTS AND ACKNOWLEDGEMENTS

4.1 Nothing in this Licence Agreement assigns or transfers any
    Intellectual Property Rights in any of the DL_POLY_4 Software.
    Those rights are reserved to STFC.

4.2 The Licensee will ensure that, if any of its employees or students
    publishes any article or other material resulting from, or relating
    to, a project or work undertaken with the assistance of any part of
    the DL_POLY_4 Software, that publication will contain the following
    acknowledgement:

    "DL_POLY_4 is a molecular dynamics simulation package written by
    I.T. Todorov and W. Smith, and has been obtained from STFC's
    Daresbury Laboratory via the website
    http://www.ccp5.ac.uk/DL_POLY"

    and cite the following reference:

    "I.T. Todorov, W. Smith, K. Trachenko& M.T. Dove,
    'J. Mater. Chem.', 16, 1911-1918 (2006)".

5.  TERMINATION

5.1 This Licence Agreement will take effect and the Licence Period will
    start when the Licensee has agreed to the terms and conditions of
    this Licence Agreement and downloaded the DL_POLY_4 Software from
    the DL_POLY_4 Website.

5.2 This Licence Agreement will terminate immediately and automatically
    if:

    5.2.1 the Licensee is in breach of this Licence Agreement; or

    5.2.2 the Licensee becomes insolvent, or if an order is made or a
          resolution is passed for its winding up (except voluntarily
          for the purpose of solvent amalgamation or reconstruction),
          or if an administrator, administrative receiver or receiver
          is appointed over the whole or any part of its assets, or if
          it makes any arrangement with its creditors.

5.3 The Licensee's right to use the DL_POLY_4 Software will cease
    immediately on the termination of this Licence Agreement, and the
    Licensee will destroy all copies of the DL_POLY_4 Software that it
or any of its employees or students then holds.

5.4 Clauses 1, 2.2, 3, 4, 5.3, 5.4, 5.5 and 6 will survive the expiry
    of the Licence Period and the termination of this Licence
    Agreement, and will continue indefinitely.
```

5.5 STFC may stop providing any assistance or advice in relation to, or any corrections, new releases or versions of the DL_POLY_4, and may stop updating or publishing the DL_POLY_4 Website at any time.

6. GENERAL

6.1 Headings: The headings in this Licence Agreement are for ease of reference only; they do not affect its construction or interpretation.

6.2 Assignment etc: The Licensee may not assign or transfer this Licence Agreement as a whole, or any of its rights or obligations under it, without first obtaining the written consent of STFC.

6.3 Illegal/unenforceable provisions: If the whole or any part of any provision of this Licence Agreement is void or unenforceable in any jurisdiction, the other provisions of this Licence Agreement, and the rest of the void or unenforceable provision, will continue in force in that jurisdiction, and the validity and enforceability of that provision in any other jurisdiction will not be affected.

6.4 Waiver of rights: If STFC fails to enforce, or delays in enforcing, an obligation of the Licensee, or fails to exercise, or delays in exercising, a right under this Licence Agreement, that failure or delay will not affect its right to enforce that obligation or constitute a waiver of that right. Any waiver by STFC of any provision of this Licence Agreement will not, unless expressly stated to the contrary, constitute a waiver of that provision on a future occasion.

6.5 Entire agreement: This Licence Agreement constitutes the entire agreement between the parties relating to its subject matter. The Licensee acknowledges that it has not entered into this Licence Agreement on the basis of any warranty, representation, statement, agreement or undertaking except those expressly set out in this Licence Agreement. The Licensee waives any claim for breach of, or any right to rescind this Licence Agreement in respect of, any representation which is not an express provision of this Licence Agreement. However, this clause does not exclude any liability which STFC may have to the Licensee (or any right which the Licensee may have to rescind this Licence Agreement) in respect of any fraudulent misrepresentation or fraudulent concealment before the signing of this Licence Agreement.

6.6 Amendments: No variation of, or amendment to, this Licence Agreement will be effective unless it is made in writing and signed by each party's representative.

6.7 Third parties: No one who is not a party to this Licence Agreement has any right to prevent the amendment of this Licence Agreement or its termination, and no one except a party to this Licence Agreement may enforce any benefit conferred by this Licence

```
    Agreement, unless this Licence Agreement expressly provides
    otherwise.

6.8 Governing law: This Licence Agreement is governed by, and is to be
    construed in accordance with, English law.  The English Courts will
    have exclusive jurisdiction to deal with any dispute which has
    arisen or may arise out of or in connection with this Licence
    Agreement, except that STFC may bring proceedings against the
    Licensee or for an injunction in any jurisdiction.
```

Bibliography

[1] Smith, W., and Forester, T. R., 1996, *J. Molec. Graphics*, **14**, 136. 3

[2] Todorov, I. T., and Smith, W., 2004, *Phil. Trans. R. Soc. Lond. A*, **362**, 1835. 3, 182

[3] Todorov, I. T., Smith, W., Trachenko, K., and Dove, M. T., 2006, *J. Mater. Chem.*, **16**, 1611–1618. 3, 182

[4] Smith, W., 1987, *Molecular Graphics*, **5**, 71. 3

[5] Smith, W., 1991, *Comput. Phys. Commun.*, **62**, 229. 3, 5, 182

[6] Smith, W., 1993, *Theoretica. Chim. Acta.*, **84**, 385. 3, 5, 182

[7] Smith, W., and Forester, T. R., 1994, *Comput. Phys. Commun.*, **79**, 52. 3

[8] Smith, W., and Forester, T. R., 1994, *Comput. Phys. Commun.*, **79**, 63. 3, 5, 64

[9] Pinches, M. R. S., Tildesley, D., and Smith, W., 1991, *Molecular Simulation*, **6**, 51. 3, 5, 182

[10] Rapaport, D. C., 1991, *Comput. Phys. Commun.*, **62**, 217. 3, 5, 182

[11] Daw, M. S., and Baskes, M. I., 1984, *Phys. Rev. B*, **29**, 6443. 4, 31

[12] Foiles, S. M., Baskes, M. I., and Daw, M. S., 1986, *Chem. Phys. Lett.*, **33**, 7983. 4, 31

[13] Finnis, M. W., and Sinclair, J. E., 1984, *Philos. Mag. A*, **50**, 45. 4, 31, 32

[14] Sutton, A. P., and Chen, J., 1990, *Philos. Mag. Lett.*, **61**, 139. 4, 32

[15] Rafii-Tabar, H., and Sutton, A. P., 1991, *Philos. Mag. Lett.*, **63**, 217. 4, 32, 40

[16] Todd, B. D., and Lynden-Bell, R. M., 1993, *Surf. Science*, **281**, 191. 4, 32

[17] Tersoff, J., 1989, *Phys. Rev. B*, **39**, 5566. 4, 41, 183

[18] van Gunsteren, W. F., and Berendsen, H. J. C. 1987, *Groningen Molecular Simulation (GROMOS) Library Manual.* BIOMOS, Nijenborgh, 9747 Ag Groningen, The Netherlands. Standard GROMOS reference. 4, 12

[19] Mayo, S. L., Olafson, B. D., and Goddard, W. A., 1990, *J. Phys. Chem.*, **94**, 8897. 4, 12, 45, 160

[20] Weiner, S. J., Kollman, P. A., Nguyen, D. T., and Case, D. A., 1986, *J. Comp. Chem.*, **7**, 230. 4, 12

[21] Smith, W., 2003, *Daresbury Laboratory.* 4, 10, 106, 115, 116, 117, 144, 208

[22] Allen, M. P., and Tildesley, D. J. 1989, *Computer Simulation of Liquids.* Oxford: Clarendon Press. 5, 50, 60, 62, 65, 182, 184

[23] Andersen, H. C., 1983, *J. Comput. Phys.*, **52**, 24. 5, 63, 182

[24] Fincham, D., 1992, *Molecular Simulation*, **8**, 165. 5, 96

[25] Miller, T. F., Eleftheriou, M., Pattnaik, P., Ndirango, A., Newns, D., and Martyna, G. M., 2002, *J. Chem. Phys.*, **116**, 8649. 5, 96

[26] Evans, D. J., and Morriss, G. P., 1984, *Computer Physics Reports*, **1**, 297. 5, 62, 65

[27] Adelman, S. A., and Doll, J. D., 1976, *J. Chem. Phys.*, **64**, 2375. 5, 62, 65

[28] Andersen, H. C., 1979, *J. Chem. Phys.*, **72**, 2384. 5, 62, 65

[29] Berendsen, H. J. C., Postma, J. P. M., van Gunsteren, W. F., DiNola, A., and Haak, J. R., 1984, *J. Chem. Phys.*, **81**, 3684. 5, 62, 65, 77

[30] Hoover, W. G., 1985, *Phys. Rev.*, **A31**, 1695. 5, 62, 65, 72, 74, 77

[31] Quigley, D., and Probert, M. I. J., 2004, *J. Chem Phys.*, **120**, 11432. 5, 62, 77

[32] Martyna, G. M., Tuckerman, M. E., Tobias, D. J., and Klein, M. L., 1996, *Molec. Phys.*, **87**, 1117. 5, 62, 77, 91, 97

[33] Warner, H. R. J., 1972, *ind. Eng. Chem. Fundam.*, **11**, 379. 14, 151

[34] Bird, R. B. e. a. 1977, *Dynamics of Polymeric Liquids*, volume 1 and 2. Wiley, New York. 14, 151

[35] Grest, G. S., and Kremer, K., 1986, *Phys. Rev. A*, **33**, 3628. 14, 151

[36] Ponder, J. W., Wu, C., Ren, P., Pande, V. S., Chodera, J. D., Schnieders, M. J., Haque, I., Mobley, D. L., Lambrecht, D. S., DiStasio Jr., R. A., Head-Gordon, M., Clark, G. N. I., Johnson, M. E., and Head-Gordon, T., 2010, *J. Phys. Chem. B*, **114**, 2549–2564. 14, 15, 17, 19, 29, 151, 152, 156

[37] Vessal, B., 1994, *J. Non-Cryst. Solids*, **177**, 103. 16, 18, 45, 152, 160

[38] Smith, W., Greaves, G. N., and Gillan, M. J., 1995, *J. Chem. Phys.*, **103**, 3091. 16, 18, 45, 152, 160

[39] Allinger, N. L., Yuh, Y. H., and Lii, J.-H., 1998, *J. Am. Chem. Soc.*, **111**, 8551. 16, 18, 152

[40] Sun, H., 1998, *J. Phys. Chem. B*, **102(38)**, 7338–7364. 17, 18, 19, 152

[41] Kumagai, N., Kawamura, K., and Yokokawa, T., 1994, *Mol. Simul.*, **12(3-9)**, 177. 17, 19

[42] Smith, W., 1993, *CCP5 Information Quarterly*, **39**, 14. 18, 21, 25

[43] Ryckaert, J. P., and Bellemans, A., 1975, *Chem. Phys. Lett.*, **30**, 123. 19, 22, 153

[44] Schmidt, M. E., Shin, S., and Rice, S. A., 1996, *J. Chem. Phys.*, **104**, 2101. 19, 22, 153

[45] Rohl, A. L., Wright, K., and Gale, J. D., 2003, *Amer. Mineralogist*, **88**, 921. 25

[46] Raiteri, P., and Gale, J. D., 2010, *J. Am. Chem. Soc.*, **132**, 17623–17634. 25

[47] Mie, G., 1903, *Annalen der Physik*, **11**, 657–697. 27, 28, 156

[48] Clarke, J. H. R., Smith, W., and Woodcock, L. V., 1986, *J. Chem. Phys.*, **84**, 2290. 27, 28, 156

[49] Weeks, J. D., Chandler, D., and Anderson, H. C., 1971, *J. Chem. Phys.*, **54**, 5237. 28, 156

[50] Groot, R. D., and Warren, P. B., 1997, *J. Chem. Phys.*, **107(11)**, 4423–4435. 28, 65, 201, 203

[51] Al-Matar, A. K., and Rockstraw, D. A., 2004, *J. Comput. Chem.*, **25**, 660–668. 30

[52] Hepburn, D. J., and Ackland, G. J., 2008, *Phys. Rev. B*, **78**(16), 165115. 31, 40

[53] Lau, T. T., Först, C. J., Lin, X., Gale, J. D., Yip, S., and Vliet, K. J. V., 2007, *Phys. Rev. Lett.*, **98**(21), 215501. 31, 40

[54] Cooper, M. W. D., Rushton, M. J. D., and Grimes, R. W., 2014, *J. Phys.: Condens. Matter*, **26**, 105401. 31, 33

[55] J., F., 1952, *Philos. Mag.*, **43**, 153. 31

[56] Ackland, G. J., and Reed, S. K., 2003, *Phys. Rev. B*, **67**, 1741081–1741089. 31

[57] Olsson, P., Wallenius, J., Domain, C., Nordlund, K., and Malerba, L., 2005, *Phys. Rev. B*, **72**, 2141191–2141196. 31

[58] Dai, X. D., Kong, Y., Li, J. H., and Liu, B. X., 2006, *J. Phys.: Condens. Matter*, **18**, 4527–4542. 32

[59] Cleri, F., and Rosato, F., 1993, *Phys. Rev. B*, **48**, 22. 32

[60] Johnson, R. A., 1989, *Phys. Rev. B*, **39**, 12556. 40

[61] Kumagai, T., Izumi, S., Hara, S., and Sakai, S., 2007, *Comput. Mat. Sci.*, **39**, 457. 41

[62] Eastwood, J. W., Hockney, R. W., and Lawrence, D. N., 1980, *Comput. Phys. Commun.*, **19**, 215. 44, 45, 46

[63] Fennell, C. J., and Gezelter, D. J., 2006, *J. Chem. Phys.*, **124**, 234104. 48, 50, 131, 132

[64] Neumann, M., 1985, *J. Chem. Phys.*, **82**, 5663. 49

[65] Fuchs, K., 1935, *Proc. R. Soc., A*, **151**, 585. 51, 53

[66] Essmann, U., Perera, L., Berkowitz, M. L., Darden, T., Lee, H., and Pedersen, L. G., 1995, *J. Chem. Phys.*, **103**, 8577. 52, 185

[67] Fincham, D., and Mitchell, P. J., 1993, *J. Phys. Condens. Matter*, **5**, 1031. 54

[68] Lindan, P. J. D., and Gillan, M. J., 1993, *J. Phys. Condens. Matter*, **5**, 1019. 54, 55

[69] Shewchuk, J. R. August 4, 1994, *An Introduction to the Conjugate Gradient Method Without the Agonizing Pain, Edition 1 1/4*. School of Computer Science, Carnegie Mellon University, Pittsburgh, PA 15213. 55, 112

[70] Schroder, U., 1966, *Solid State Commun.*, **4**, 347–349. 55

[71] Shardlow, T., 2003, *SIAM Journal on Scientific Computing*, **24**(4), 1267–1282. 62, 65, 76

[72] Leimkuhler, B., Noorizadeh, E., and Theil, F., 2009, *J. Stat. Phys.*, **135**, 261–277. 62, 65, 74

[73] Ikeguchi, M., 2004, *J. Comp. Chemi.*, **25**, 529–541. 62, 82, 85, 90, 92

[74] Ryckaert, J. P., Ciccotti, G., and Berendsen, H. J. C., 1977, *J. Comput. Phys.*, **23**, 327. 62, 182

[75] McCammon, J. A., and Harvey, S. C. 1987, *Dynamics of Proteins and Nucleic Acids*. Cambridge: University Press. 65

[76] Izaguirre, J. A. Langevin stabilisation of multiscale mollified dynamics. In Brandt A., Binder K., B. J., editor, *Multiscale Computational Methods in Chemistry and Physics*, volume 117 of *NATO Science Series: Series III - Computer and System Sciences*, pages 34–47. IOS Press, Amsterdam, 2001. 65, 68

[77] Samoletov, A., Chaplain, M. A. J., and Dettmann, C. P., 2007, *J. Stat. Phys.*, **128**, 1321–1336. 65, 74

[78] Hoogerbrugge, P. J., and Koelman, J. M. V. A., 1992, *EPL (Europhysics Letters)*, **19**(3), 155–160. 65

[79] Español, P., and Warren, P., 1995, *EPL (Europhysics Letters)*, **30**(4), 191–196. 65, 201

[80] Melchionna, S., Ciccotti, G., and Holian, B. L., 1993, *Molec. Phys.*, **78**, 533. 85

[81] Martyna, G. M., Tobias, D. J., and Klein, M. L., 1994, *J. Chem. Phys.*, **101**, 4177. 85, 93

[82] Todorov, I. T., Bush, I. J., and Porter, A. R., 2009, *Parallel Scientific Computing and Optimization. Springer Optimization and Its Applications, ISSN 1931-6828*, **27**. 111

[83] Todorov, I. T., Bush, I. J., and Smith, W., 2008, *Cray User Group 2008.* 111

[84] Bush, I. J., Todorov, I. T., and Smith, W., 2010, *Cray User Group 2010.* 111

[85] Hockney, R. W., and Eastwood, J. W. 1981, *Computer Simulation Using Particles.* McGraw-Hill International. 182, 184, 185

[86] Smith, W., 1992, *Comput. Phys. Commun.*, **67**, 392. 182

[87] Smith, W., and Fincham, D., 1993, *Molecular Simulation*, **10**, 67. 182

[88] Bush, I. J., Todorov, I. T., and Smith, W., 2006, *Computer Physics Communication*, **175**, 323. 185

[89] Bush, I. J., 2000, *Daresbury Laboratory.* 185

[90] Reith, D., Pütz, M., and Mü, ller-Plathe, F., 2003, *J. Comp. Chem.*, **24**, 1624. 200

[91] Rühle, V., Junghans, C., Lukyanov, A., Kremer, K., and Andrienko, D., 2009, *J. Chem. Theory Comput.*, **5**, 3211. 200

[92] Pagonabarraga, I., and Frenkel, D., 2001, *Journal of Chemical Physics*, **115**(11), 5015–5026. 203

[93] Trofimov, S. Y., Nies, E. L. F., and Michels, M. A. J., 2002, *Journal of Chemical Physics*, **117**(20), 9383–9394. 203

[94] Koelman, J. M. V. A., and Hoogerbrugge, P. J., 1993, *EPL (Europhysics Letters)*, **21**(3), 363–368. 203

[95] Mar, C. A., Backx, G., and Ernst, M. H., 1997, *Physical Review E*, **56**(2), 1676–1691. 203

[96] Peters, E. A. J. F., 2004, *EPL (Europhysics Letters)*, **66**(3), 311–317. 204

[97] Seaton, M., Anderson, R., Metz, S., and Smith, W., 2013, *Molecular Simulation*, **39**(10), 796–821. 204

[98] Lowe, C. P., 1999, *EPL (Europhysics Letters)*, **47**(2), 145–151. 204

[99] Stoyanov, S. D., and Groot, R. D., 2005, *Journal of Chemical Physics*, **122**(11), 114112. 204

[100] Gonella, G., orlandini, E., and Yeomans, J. M., 1997, *Phys. Rev. Lett.*, **78**, 1695. 204

Index