

The Guide 1.1 to CASTEP 3.9

Phil Lindan

Computational Materials Science Group,
Daresbury Laboratory, Warrington WA4 4AD, UK

`ukcp@dl.ac.uk`
<http://www.cse.clrc.ac.uk/Activity/UKCP>

April 26, 1999

About this Guide

This Guide covers the CASTEP 3.9 code and is part of the distribution to UK academia. The Guide is divided into three parts:

Part I provides a brief introduction to first-principles simulation and the history of CASTEP.

Part II tells you how to install and run CASTEP

Part III contains some useful reference material

In this, the first version, the scope is deliberately limited to explaining the basics of running the code. In later revisions more detail of the theory, methodology and the code itself will be included.

License terms and copyright

The entire academic CASTEP distribution, that is, the CASTEP code, associated library files and data, utility programs and this Guide, is strictly copyright and is distributed under license as part of an Agreement between CLRC Daresbury Laboratory, the United-Kingdom Car-Parrinello Consortium (UKCP) and Molecular Simulations Inc. (MSI). You must sign a License Agreement to obtain and use any part of the academic CASTEP distribution. Please note in particular that the code is provided for use *solely in UK academic institutions*, and only for non-commercial use.

Prerequisites and minimum requirements

CASTEP is a scientific research code under constant development. It is implicit therefore that its users know rather a lot about the science and methodology, enough about the computers they use, and that these computers are suitable for large-scale simulations. To be explicit, one cannot run CASTEP sensibly without at least:

- Some knowledge of solid-state theory
- A UNIX workstation with ≥ 1 GB free disk space and preferably at least 100 MB available memory
- A working knowledge of UNIX and the ability to edit text files

By far the most convenient way to use CASTEP is through the Cerius² User Interface (UI), available at modest cost to UK academics from MSI. If you

cannot obtain the Cerius², this Guide describes the essential features of the input files and some aspects of the code. Please bear in mind that CASTEP has been developed to run under the interface. Running the code without the interface requires considerable extra work and preparation of the files "by hand".

When all else fails

This Guide is by no means exhaustive, but it should help you to get started. You will also find that looking at the code is a good way (indeed, the only way) to understand fully what's going on. If you are at a complete loss, send a message to `ukcp@d1.ac.uk` describing your problem in detail. Whilst at this time we cannot provide full user support we will do our best to find a solution.

Credits and thanks

The CASTEP project was initiated by Mike Payne of Cambridge University. CASTEP is the product of the collaborative efforts of dozens of researchers, students, postocs and software professionals. Literally 100's of person-years' effort has gone into the code. Without naming names, which would inevitably result in omissions and unintentional offence, sincere thanks must go to all who have contributed over the years, and those who are actively developing the code right now.

Research and publications benefitting from the use of CASTEP should include the following citation:

CASTEP 3.9 academic version, licensed under the UKCP-MSI Agreement, 1999; *Rev. Mod. Phys.* **64** (1992) 1045

If you use the K290 code (see section 7.5) please cite the following reference: "Symmetry and Special Points Program K290", K. Kunc, R. J. Needs, O. H. Nielsen and R. M. Martin (to be published).

Finally, thanks to Victor Milman of MSI who kindly provided a lot of detailed information during the preparation of this Guide.

Contents

About this Guide	1
I Overview	5
1 Introduction	5
1.1 UKCP	5
1.2 First-principles simulation	5
2 CASTEP history	6
2.1 Beginnings	6
2.2 CETEP: Meeting the Grand Challenges	6
2.3 CASTEP Now	7
2.4 Parallel CASTEP	7
II Using CASTEP	8
3 The basics	8
3.1 The distribution	8
3.2 Installation	8
4 Running CASTEP	9
4.1 The file “seed”	9
4.2 Examining the output	9
4.3 Types of calculation and associated files	10
4.4 Input files	11
4.5 Output files	11
4.6 Restarting a calculation	12
4.7 Similarities between CASTEP and CETEP files	13
5 Running the test cases	13
5.1 Checking	14

<i>CONTENTS</i>	4
6 Compiling CASTEP	14
6.1 Software resources, libraries	14
6.2 The serial code	16
6.3 SGI shared-memory machines	16
6.4 Massively-parallel machines with MPI	17
6.5 MPI-Parallel Strategy	17
7 Detailed description of some CASTEP files	19
7.1 Geometry Information: <code>.geom</code>	19
7.2 Control parameters: <code>.rundat</code>	21
7.3 Runtime parameters: <code>.param</code>	32
7.4 The pseudopotentials: <code>.recpot</code> and <code>.realpot</code>	33
7.4.1 Preparing <code>.recpot</code> when using ultrasoft pseudopotentials	34
7.4.2 Preparing <code>.recpot</code> when using norm-conserving potentials	36
7.4.3 The real-space pseudopotential file <code>.realpot</code>	36
7.5 Symmetry information: <code>.symm</code>	37
III For Reference	38
8 Some UKCP publications	38

Part I

Overview

1 Introduction

Welcome to **CASTEP**. The code has been made available to UK academia through a unique agreement between the UK Car-Parrinello Consortium (UKCP), Molecular Simulations Inc. (MSI) and CLRC Daresbury Laboratory. As well as providing you with the code, the agreement embraces an aggressive and co-ordinated **CASTEP** development programme, encompassing work at MSI, Daresbury and within the UKCP community. The agreement rests on a long history of **CASTEP** development and groundbreaking applications. **CASTEP** is the result of much hard work and is a valuable resource, so please respect the terms of the license you signed.

1.1 UKCP

The United-Kingdom Car-Parrinello Consortium, or UKCP, was formed in 1990 with the aim of exploiting parallel supercomputers for first-principles simulation. This has been achieved with remarkable success, as witnessed by the large number of top-class publications from the UKCP research groups. UKCP has become much more than a computer consortium though. It is now a powerful collaborative network bringing together academic researchers, high-performance computing, MSI's software expertise and vast experience of Car-Parrinello code development and application. It is the focus for the UK's plane-wave DFT community of which you are now a part.

Currently there are nine groups in the Consortium, based in the Universities of Bath, Cambridge, Edinburgh, Oxford, Queen's Belfast, and University College London, and the Computational Materials science group at Daresbury Laboratory. EPSRC funds support for UKCP, which is provided by Daresbury.

1.2 First-principles simulation

CASTEP is a first-principles simulation code. First-principles calculations involve no adjustable parameters, are based on an explicit quantum treatment of the electrons in a model system, which means solving Schrödinger's equation to find the electronic ground state. There are still approximations in the solution, but these are of an extremely general kind. The calcula-

tions yield the total energy of assemblies of atoms, and forces on the atoms. This knowledge enables one to calculate a vast range of properties with near chemical accuracy.

Density Functional Theory (DFT) is a powerful, general theory relating the total energy of a system of interacting electrons in an external potential to the electron density. Many first-principles calculations employ DFT to describe the valence electrons. The expansion in the scope of such calculations has been breathtaking in the last fifteen years or so, and they are now routinely used to study fundamental aspects of the physics and chemistry of condensed matter in every conceivable guise, ranging from catalysts to proteins, minerals to liquid metals. The great importance of DFT and its application in simulation was recognised by the award of the 1998 Nobel Prize in chemistry to Walter Kohn and John Pople.

The DFT plane-wave pseudopotential method, was pioneered in the 80's by Roberto Car and Michele Parrinello. This approach has transformed the first-principles landscape, making possible calculations far beyond previous approaches. Throughout this time the UK has played a leading role in the development and application of the methods, an effort founded on Mike Payne's work with CASTEP.

The method relies on a plane-wave basis, pseudopotentials and the use of density-functional theory to describe the valence electrons in a model system. Other ingredients include fast-Fourier transforms and minimization of the total energy rather than matrix diagonalization.

2 CASTEP history

2.1 Beginnings

The CAMbridge Serial Total Energy Package or CASTEP was developed in the mid-'80s by Mike Payne, with numerous contributions from the students, postdocs and academic staff of many research groups. As mentioned, CASTEP is a DFT plane-wave pseudopotential code, and it calculates the total energy, forces and stresses in a 3D-periodic system. Having ionic forces allows for efficient structural optimization and for molecular dynamics calculations. The review article *Rev. Mod. Phys.* **64** (1992) 1045 sets out in detail the theory and technical implementation of CASTEP.

2.2 CASTEP: Meeting the Grand Challenges

It was recognised early that realistic and accurate simulations would inherently be on a grand scale, and that the use of parallel machines was essen-

tial. Out of this came **CETEP**, a highly efficient parallel version of **CASTEP** in which the grids spanning real- and reciprocal-space were decomposed and distributed among the nodes of the parallel machine. Once again, credit goes to many research groups and individuals for their contributions to **CETEP**. UKCP's ability to exploit parallel computers has consistently enabled it to tackle and solve the most difficult and demanding of problems of the day - real Grand Challenges. This in turn relies heavily on the UK's investment in supercomputing, from the Meiko Computing Surface in the early days through to the latest UK high-performance computer service, CSAR, at the Manchester Computing Centre.

2.3 CASTEP Now

Since 1993 MSI have developed and consolidated **CASTEP** as a commercial product, while **CETEP** remained the workhorse of UKCP academics. However, within the new UKCP-MSI agreement these strengths are combined. **CASTEP** is now a state-of-the-art plane-wave code, incorporating all the latest and best technology. Key Features include:

- Ultrasoft or norm-conserving pseudopotentials with non-linear core corrections
- Complete pseudopotential library
- Comprehensive range of minimisation methods: Density Mixing, RM-DIIS, Conjugate Gradients band-by-band and all-bands
- Full structural relaxation and molecular dynamics capabilities
- Local Density and Generalised Gradient approximations, spin-polarisation
- Supported by 100's of successful, published applications
- Compatible with MSI Cerius2 user interface giving automatic setup and full visualisation facilities

2.4 Parallel CASTEP

For much of the functionality available in this release, **CASTEP** runs in parallel under MPI. However, the most recently- added features will only work when the code is run serially. Calculations which access these features will generate an error message when started in parallel. Work is in progress (to complete mid-1999) to parallelise the latter.

Part II

Using CASTEP

3 The basics

3.1 The distribution

The CD contains the following items.

```
POTENTIALS/   castepexeMPI_v3.9/   guide/       utilities/
POTENTIALS_NC/ baselines.3.9/     quick_test/
README       castepexe_v3.9/    test_cases/
```

The contents of the files and directories are described in table 1

POTENTIALS	Library of ultrasoft pseudopotentials
POTENTIALS_NC	Library of norm-conserving pseudopotentials
README	Text file with basic instructions
baselines.3.9/	Reference output from the test cases
castepexe_v3.9/	Source code and object directories
castepexeMPI_v3.9/	MPI parallel source code and object directories
guide/	This document in several formats
quick_test/	Set of input files for a single quick test
test_cases/	Set of input files for the test cases
utilities/	Useful programs and scripts

Table 1: What's in the distribution

3.2 Installation

Copy all the files from the distribution CD to a suitable place on your workstation. For example, on an SGI machine:

```
cd
mkdir CASTEP
cd CASTEP
cp -r /CDROM/* .
```

For various purposes you will need to set the following environment variables:

```
setenv MACHINE <your_machine>
setenv PLATFORM x11
```

where <your_machine> is one of `clover`, `crayc90`, `crayj90`, `crayt90`,

crayt90ieee, crayymp, decalpha.x11, hp700, ibmrios or irix64 (compare with the subdirectories of `castepexe_v3.9/obj`).

4 Running CASTEP

If you are using a Silicon Graphics machine the code is ready to run. The executable is `castepexe_v3.9/obj/irir64.x11/castepexe` .

If you are using another type of machine you will have to compile the code yourself. Go to section 6 before continuing this section.

To place things in context, take a look at the files in the `quick_test` directory. This contains a copy of one of the test cases, which will be used for illustration.

4.1 The file “seed”

You will see that in `quick_test` there are several files with a common prefix, `BN`. For a given calculation, all CASTEP files are preceded by a “seed”, which serves to label the input and output from a job. Here the seed is `BN`. Note that the default seed is `castep`, and typing the executable name alone causes CASTEP to look for files with this prefix. Assuming you have a complete set of input files and an executable, the code is run with the command

```
<executable_name> <seed>
```

To run this test case from the `quick_test` directory, type

```
$castep_dir/castepexe_v3.9/obj/irir64.x11/castepexe BN
```

where `$castep_dir` is the path to the CASTEP distribution. After execution, the main CASTEP output is in the file `BN.cst`.

4.2 Examining the output

The contents of the `.cst` output file are largely self-explanatory. Firstly, the memory requirements for the job are stated¹. The next section of the file contains many of the control parameters and variables used in the calculation. In addition there is information about system geometry, pseudopotentials, k-points, the FFT grid and any advisory or warning messages.

During the calculation, information about convergence or progress is written to this file. For example, `BN.cst` will contain information about the

¹Note that at present CASTEP does not compute the memory requirement for MPI-parallel execution

convergence of the electronic minimisation, looking something like this:

SCF loop	Energy (eV)		Energy gain per atom	Timer (sec)	---	SCF
	Initial	Final			---	SCF
1	0.00000	-391.75302	195.876511	2.39	---	SCF
2	-391.75302	-399.35995	3.803465	3.47	---	SCF
3	-399.35995	-363.30127	-18.029340	4.63	---	SCF
4	-363.30127	-352.18237	-5.559450	5.72	---	SCF
5	-352.18237	-352.15629	-0.013041	6.82	---	SCF
6	-352.15629	-352.16147	0.002588	7.92	---	SCF
7	-352.16147	-352.16157	0.000054	9.08	---	SCF
8	-352.16157	-352.16157	-0.000003	9.86	---	SCF
TOTAL ENERGY IS		-352.1615687 (eV)				

It is obviously essential to check this information to see that the calculation has achieved the desired result. After the calculation is complete other results are written, such as the electronic eigenvalues and stress tensor, as well as the overall timings.

4.3 Types of calculation and associated files

There are three main types of calculation possible with CASTEP.

Single-point energy

Here the ground-state energy is calculated for fixed ionic positions. The BN calculation in `quick_test` is a single-point energy calculation. The calculation also yields the Khon-Sham eigenvalues at the set of k-points considered, the ionic forces and the stress tensor (if selected).

Geometry optimisation

Almost always one requires the total energy of a system in which the structural parameters have been optimized. This may involve finding the lattice vectors that minimize the total energy (cell optimization), or relaxing the structure by moving the ions to positions of zero force, or both.

Molecular dynamics

Molecular dynamics (MD) involves the stepwise integration of Newton's equations from a given starting point. It is the most natural method of performing equilibrium statistical-mechanical calculations via simulation.

Additionally, MD may be used to explore conformations, such as molecular adsorption sites on a surface.

4.4 Input files

The mandatory input files are:

`<seed>.ewdata` Tabulated data used for the Ewald summation. This file is identical for all calculations.

`<seed>.geom` Contains the geometry and k-point information.

`<seed>.param` Parameters for memory assignment.

`<seed>.recpot` The pseudopotentials in reciprocal space.

`<seed>.rundat` The switches and variables that control CASTEP execution.

Several files must be present only if certain options are selected in the `.rundat` file. These are:

`<seed>.optim` Contains the constraints and the initial Hessian. Needed if using BFGS geometry optimization (`IION=3` and/or `IBOX=2`).

`<seed>.realpot` Contains the real-space non-local pseudopotentials. Needed if evaluation is in real space.

`<seed>.symm` Contains the symmetry operations of the structure described in `<seed>.geom`. Needed if using symmetry information (`ISYMM = 1`).

4.5 Output files

The number of output files produced by CASTEP varies according to the type of calculation performed. All classes of calculation produce the following files:

`<seed>.cst` The output file containing details of the calculation, convergence information and any runtime analysis.

`<seed>.gm*` Geometry and eigenvalues.

`<seed>.charge*` The electronic charge.

`<seed>.wavfun*` The wavefunctions.

`<seed>.cst_esp` The local potential.

Additional files are written if certain options are chosen.

`<seed>.occnum*` Contains the occupation numbers, written when performing calculations on metals.

`<seed>.bands` The electron bands. Written if applying subspace rotation (`ISBROT` $\neq 0$) or performing a bandstructure calculation.

Files whose names are marked with an asterisk (*) are the binary files which must be present to re-start a calculation (see section 4.6).

Both MD and geometry optimization calculations generate additional output files. Both write these two files:

`<seed>.coord` Sequence of ionic co-ordinates.

`<seed>.force` Sequence of ionic forces.

Geometry optimisation calculations write these additional files:

`<seed>.optim_out` Like `<seed>.optim` but with the final Hessian.

`<seed>.cell` Evolution of the cell ($a, b, c, \alpha, \beta, \gamma$, internal and external stress tensors. Also written if the stress tensor is calculated (`ISTRESS = 1`))

MD calculations also produce these files

`<seed>.veloc` Sequence of ionic velocities ($\text{\AA}/\text{ps}$)

`<seed>.temper` Temperature (K)

`<seed>.hamilt` MD constant of motion (eV)

Several other intermediate and output files may be written depending on the settings in the `.rundat` file.

4.6 Restarting a calculation

Calculations may be restarted from the binary output files. The files required are:

`<seed>.gm` Geometry and eigenvalues.

`<seed>.charge` The electronic charge.

`<seed>.wavfun` The wavefunctions.

`<seed>.occnum` Contains the occupation numbers,

In addition, you must change the value of `ISTART` to 1 in the `.rundat` file. Note carefully that `CASTEP` will both read from and write to these files. Therefore, if you wish to keep intermediate results, such as charge densities, you must copy the relevant files before making a restart.

4.7 Similarities between CASTEP and CETEP files

If you are familiar with `CETEP` it is worth knowing that there are distinct relationships between `CASTEP` and `CETEP` input files. These are summarised in table 2. Examination of the subroutines in the `cst_inp_geom.f` file should enable you to cross-match precisely the two sets of files.

CETEP	CASTEP	Remarks
<code>fort.10</code>	<code><seed>.ewdata</code>	Identical.
<code>fort.11</code>	<code><seed>.recpot</code>	“Non-local” projector for the local potential not necessary in <code>CASTEP</code> file.
<code>fort.14</code>	<code><seed>.rundat</code>	Many parameters and variables in common, but formats differ.
<code>fort.18</code>	<code><seed>.realpot</code>	Identical.

Table 2: Relationships between `CASTEP` and `CETEP` files

5 Running the test cases

Several test cases are included in the `test_cases` directory. The major options selected and functionality used for each test case is summarised in table 3. It is strongly recommended that you run all the test cases to validate your copy of the code and its installation. In addition, they provide a range of examples of calculations that are possible with `CASTEP`. You may run any one of the cases as described in section 4.1.

Two simple scripts are provided with the test cases. If you wish to run all of them, `run_all` may be useful. This script will run all of the test cases and `diff` their output with that contained in the appropriate subdirectory of the `baselines_3.9/` directory. The output of the script is directed to a file of your choice. To use the script, type

```
run_all <output_file_name> <executable_name>
```

For example, this might be

```
run_all test.out $castep_obj_dir/castepexe
```

where `$castep_obj_dir` is the path to the CASTEP object directory.

The script `clean_all` removes all output files generated by running the test cases, except the `.cst` files.

The `silicon*` test cases read files containing the charge density (`.charge`) and geometry (`.gm`) from a previous calculation. Since these are binary files they are machine-specific. Files are provided in the `BINARY_CRAY` and `BINARY_NONCRAY` directories for Cray (vector) and other machines respectively.

5.1 Checking

Remember that there are bound to be small differences in the answers when the code is run on different machines. When checking the differences pay particular attention to total energies. These should agree to within the tolerance specified for the calculation. If they agree for all the test cases you may assume that everything is working properly.

The spin-polarised test case calculations need a special mention. Because these systems have multiple energy minima, it is possible that slight differences in initial conditions, caused by execution on different machines, will result in the location of different minima. This may arise, for example, if the sets of random numbers used to initialise the wavefunctions differ. This applies to test cases 3b, 3c, 4a-b and 6a-c. Comparison of the various baselines for these cases shows that the answers are somewhat machine-dependent. Therefore, if you compare your results for these cases with the baselines from a different machine, you may observe some discrepancies.

6 Compiling CASTEP

Academic CASTEP is distributed ready-compiled for SGI workstations and the Cray T3E. Here is some information to help with compilation on any platform, though please note that at present Daresbury can only offer support for CASTEP running on SGI and Cray T3E machines.

6.1 Software resources, libraries

Serial CASTEP requires few ancilliary software packages, and should work on an “average” workstation. Compilation requires both FORTRAN and C compilers. CASTEP adheres to standard F77 in most respects, but it does rely on non-standard pointers. There is scope for the use of the `blas` and

#	Sys	Sym	Fn	Mth	BSC	Ions	box	MD	PP	Space
1a	Si		LDA	0	•	SD			NL	k
1b	Si		GGA	0	•	SD			NL	k
1c	Si		GGA	0	•	BFGS			NL	k
1d	Si		GGA	1	•	BFGS			NL	k
1e	Si		GGA	1				•	NL	k
2a	Si	•	GGA	0	•				NL	k
2b	Si	•	GGA	0	•				NL	r
2c	MgCl		GGA	0		BFGS			NL	k
2d	Al	•	LDA	0	•		•		NL	k
3a	PuO2	•	GGA	0					NL	k
3b	PuO2	•	GGS	0					NL	k
3c	PuO2	•	GGS	0					NL	r
4a	Fe	•	LSDA	1					NL	k
4b	Fe	•	GGS	1					NL	k
5a	BN	•	GGA	2	•		•		US	k
5a BS	BN	•	GGA	2	•		•		US	k
5b	BN	•	GGA	1	•		•		US	k
5c	BN	•	GGA	2	•		•		NL	k
6a	BN	•	GGS	2	•		•		US	k
6a BS	BN	•	GGS	2	•		•		US	k
6b	BN	•	GGS	1	•		•		US	k
6c	BN	•	GGS	2	•		•		NL	k
7a	Pt		GGA	2					US	k
7b	Pt/Al		GGA	2					US/NL	k
DOS	Si		GGA	0					NL	k
BS	Si		GGA	0					NL	k
OPT	Si		GGA	0					NL	k

Table 3: The test cases. The headings are (with `.rundat` parameters in parenthesis): #, the test case number (with BS, DOS and OPT denoting band structure, densities-of-states and optics, respectively); Sys, the system; Sym, whether symmetry is exploited (ISYM); Fn, the exchange-correlation functional used (Theory level); Mth, the method used to minimise the electronic energy (METHOD); BSC, basis-set correction computed (BAS_CORR); Ions, ionic relaxation (IION); box, cell relaxation (IBOX); MD, molecular dynamics; PP, the pseudopotential type, either non-local or ultrasoft (IVPTYP); Space, how the pseudopotential is evaluated, in real (r) or reciprocal (k) space (NPOT).

lapack libraries in CASTEP, though these are not essential since software versions of the relevant routines are included in the distribution. However, the parallel version of the code does rely on the MPI library being present.

6.2 The serial code

The directory `castepexe_v3.9/obj` contains several sub-directories for different machines:

```

clover.x11/      crayt90.x11/    decalpha.x11/   irix64.x11/
crayc90.x11/    crayt90ieee.x11/ hp700.x11/
crayj90.x11/    crayymp.x11/   ibmrios.x11/

```

Each sub-directory contains a Makefile for the machines. The sub-directory `irix64.x11` for Silicon Graphics machines contains pre-compiled object and library files for CASTEP. The executable produced by linking these files is `$castep_dir/castepexe_v3.9/obj/irix64.x11/castepexe`. If you are using a machine other than an SGI, or if you make changes to the source code, you will need to compile the code.

CASTEP will compile on the machines for which there is a subdirectory to `obj/`. (In fact, it will compile on others too but you would have to do a little more work.) Assuming your machine is covered, set the following environment variables if you have not done so already.

```

setenv MACHINE <your_machine>
setenv PLATFORM x11

```

Then, to compile CASTEP, run the script `castepexe_v3.9/Make`. The script copies the source files to the appropriate directory, compiles and links them to produce `castepexe`. Note that while compilation may well generate many warnings, these should be benign.

6.3 SGI shared-memory machines

The serial code will also execute in parallel on SGI shared-memory multiple-processor machines such as the Origin series. On such machines, several costly parts of the calculation are shared between the processors, achieving some degree of parallelism. The environment variable `MP_SET_NUMTHREADS` must be set, for example:

```

setenv MP_SET_NUMTHREADS "min( 4,all)"

```

will select the smaller number from 4 and the total number of processors on the machine. Note that this shared-memory strategy is not a massively-parallel one, and useful speedup will only be achieved on a small number of

processors.

6.4 Massively-parallel machines with MPI

In this distribution the serial and MPI-parallel codes are distinct. The MPI-parallel code is on the `castepexeMPI_v3.9/` directory. Object files and an executable for the Cray T3E are included in the `castepexeMPI_v3.9/obj/t3e.x11` directory. Compilation is as for the serial code, except that the `MACHINE` environment variable should be set appropriately, e.g. to `t3e`. Running the parallel code is essentially the same as for the serial code, except that the number of processors must be specified either on the command line or in a batch script. A typical batch script for the T3E might look like this:

```
#QSUB -r castep_job
#QSUB -q normal
#QSUB -A <your_account>
#QSUB -l mpp_p=64
#QSUB -l mpp_t=2:00:00
cd CASTEP/bigsystem
mpprun -n64 /CASTEP/castepexeMPI_v3.9/obj/t3e.x11/castepexe si_surface
```

6.5 MPI-Parallel Strategy

The largest data structure that CASTEP deals with is the wavefunction and associated quantities, such as its derivative. The size of such a structure is proportional to

1. The number of k points
2. The granularity of the FFT grid (i.e. the number of plane-waves employed)
3. The number of bands

For example, a large calculation on an insulator may employ 2 k points, a $128 \times 128 \times 128$ FFT grid and 100 bands, a total of 6.4 Gigabytes, so it is imperative that on a distributed memory architecture data structures of this size are distributed among the processors. CASTEP can perform this distribution in three different ways:

1. **By k-point** Each processor stores the whole wavefunction for a subset of the required k points. For example if calculation that uses 8 k points is run on 8 processors, each processor would work on one k point. This distribution is best for very large numbers of k points, as might be required for a metal, and in general will require less communication between the processors than the other strategies.

2. **By G-Vector** Each processor stores a part of the wavefunction for all the k points. Thus, for a 2 k point calculation on 8 processors each processor will store 1/8 of the wavefunction at each k point (For users familiar with `CETEP` this is the strategy used there.) This approach works best for large unit cell calculations with few k points, but does require appreciably more inter-processor communication than the k-point strategy.
3. **By k-point and G-Vector** This is a hybrid of the previous two methods, best explained by example. Consider the same calculation as in the **G**-vector case. Here, initially, the k points are distributed amongst the processors, in this case the first four processors get k point 1 and the second four k point 2, and then at a given k point the wavefunction is distributed amongst the processors dealing with that k point as in the **G**-vector case.

There are also a number of other constraints on which distribution strategy may be used, for instance the schemes involving k-point distribution are only available when using the "All Bands" minimisation method (`METHOD=1` in `.rundat`). In general you need not concern yourself with which one to use since the code will automatically choose a suitable distribution strategy. However, should you wish to force the code to use a certain strategy, it is possible to set one of the flags `FORCE_K_POINT`, `FORCE_G_VECTOR` or `FORCE_K_POINT_AND_G_VECTOR` in the file `par_strategy.f`, and recompile.

It is possible to re-start a calculation on a different number of nodes and with a different distribution strategy than was used originally.

As already mentioned, the MPI code is currently under development, and while not a "beta" release, you may encounter some bugs. If you do, please let us know.

7 Detailed description of some CASTEP files

In this section the key CASTEP files are described with reference to the files in the `quick_test` directory. Those files relate to the unit cell of Boron Nitride, which contains one B and one N atom.

CASTEP relies on the information about different species appearing in the same order in the input files. For example, this means that the pseudopotentials for B and N appear in that order in the `BN.recpot` file, the positions of B atoms precede those for N atoms in `BN.geom`, and the specifications of core charge, numbers and masses in `BN.rundat` are in the same order.

7.1 Geometry Information: `.geom`

The file contains geometry information: cell parameters, atomic coordinates, k-points coordinates and weights. The whole file looks like this:

2.0871212231	0.0000000000	0.0000000000	
0.7379087850	2.2137263550	0.0000000000	
1.2780955070	1.2780955070	2.5561910140	
2.0871212231	0.0000000000	0.0000000000	
0.7379087850	2.2137263550	0.0000000000	
1.2780955070	1.2780955070	2.5561910140	
0.0000000000	0.0000000000	0.0000000000	0.0
0.0000000000	0.0000000000	0.0000000000	
0.2500000000	0.2500000000	0.2500000000	0.0
0.2500000000	0.2500000000	0.2500000000	
0.4000000000	0.4000000000	0.4000000000	
0.4000000000	0.4000000000	0.2000000000	
0.4000000000	0.4000000000	0.0000000000	
0.4000000000	0.4000000000	-0.2000000000	
0.4000000000	0.4000000000	0.6000000000	
0.4000000000	0.2000000000	0.2000000000	
0.4000000000	0.2000000000	-0.2000000000	
0.2000000000	0.2000000000	0.2000000000	
0.2000000000	0.2000000000	0.0000000000	
0.0000000000	0.0000000000	0.0000000000	
0.0640000000			
0.1920000000			
0.0480000000			
0.0960000000			
0.1920000000			
0.0960000000			
0.1920000000			
0.0640000000			
0.0480000000			
0.0080000000			

The first six lines are two sets of lattice vectors (in Å):

```
2.0871212231  0.0000000000  0.0000000000
```

```

0.7379087850  2.2137263550  0.0000000000
1.2780955070  1.2780955070  2.5561910140
2.0871212231  0.0000000000  0.0000000000
0.7379087850  2.2137263550  0.0000000000
1.2780955070  1.2780955070  2.5561910140

```

The format is columnar:

```

Ax          Bx          Cx
Ay          By          Cy
Az          Bz          Cz

```

Usually the two sets of lattice vectors are identical. For historical reasons it is possible to attempt using output files from one run to “restart” for a slightly different cell. In this case the first three lines give the new lattice vectors, and the next three give the old ones. This functionality is no longer supported.

The next section of the file contains the fractional positions of the ionic cores. In this example there are two species, B and N, and one atom of each species. The coordinates are sorted by species, and again for historical reasons, the set of positions for the ions in a given species appears twice.

```

0.0000000000  0.0000000000  0.0000000000  0.0
0.2500000000  0.2500000000  0.2500000000  0.0
0.2500000000  0.2500000000  0.2500000000

```

The line

```

0.0000000000  0.0000000000  0.0000000000  0.0

```

is the fractional coordinates of the atom of species 1, which is boron. The last number is the moveability switch: 0 - fixed, 1 - moveable atom. The number of lines for each species is determined by the `NIONSP` parameter in the `.rundat` file. The coordinates of this atom are repeated on the next line:

```

0.0000000000  0.0000000000  0.0000000000

```

This format is repeated for all species, so that in this example the fractional coordinates of the atom of species 2, nitrogen, appears next:

```

0.2500000000  0.2500000000  0.2500000000

```

After this coordinates section there might be a section with atomic velocities (only if the task is Molecular Dynamics). velocities would be written in a loop over species, one line of three numbers per atom. Note that velocities are in cartesian reference frame ($\text{\AA}/\text{ps}$).

The last part of the file contains the fractional coordinates of the k-points that are used for sampling the Brillouin zone. The number of lines is the

same as the NKPTS parameter in the `.param` file.

```

0.4000000000    0.4000000000    0.4000000000
0.4000000000    0.4000000000    0.2000000000
0.4000000000    0.4000000000    0.0000000000
0.4000000000    0.4000000000   -0.2000000000
0.4000000000    0.4000000000    0.6000000000
0.4000000000    0.2000000000    0.2000000000
0.4000000000    0.2000000000   -0.2000000000
0.2000000000    0.2000000000    0.2000000000
0.2000000000    0.2000000000    0.0000000000
0.0000000000    0.0000000000    0.0000000000

```

The weights of the k-points, in the same order as the coordinates, appear last.

```

0.0640000000
0.1920000000
0.0480000000
0.0960000000
0.1920000000
0.0960000000
0.1920000000
0.0640000000
0.0480000000
0.0080000000

```

7.2 Control parameters: `.rundat`

This file contains settings and switches for various operation modes, plus certain data specific for atomic species in the model. The whole file looks like this:

3	!	NITER	!	Number of iterations	
1	!	NPRINT	!	Forces are printed every NPRINTth iteration	
300.00000	!	ENMAX	!	Kinetic energy cut-off (eV)	
0	!	NPOT	!	Potential: 0-reciprocal, 1-real space	
5	!	NITMAX	!	Number of electronic CG steps	
1	!	NDELAY	!	Dummy starting from version 2.0	
0	!	ISTART	!	Mode: 0-new run, 1-restart	
1	!	ISBROT	!	Rotation: 0-NO,1-strict, 2-fast METALS only	
0	!	IOCCUP	!	Partial occupation: 0-no, 1-yes	
0.20000	!	DELMIN	!	Minimum smearing width (eV)	
4.00000	!	DELMAX	!	Starting smearing width (eV)	
6	!	NDEL	!	Smearing width is halved every NDEL SCF loops	
0	!	IION	!	Ionic relaxation: 0-no, 1-CG, 2-simple, 3-BFGS	
0	!	IBOX	!	Unit cell relaxation: 0-no, 1-uniform, 2-BFGS	
0	!	IPRINT	!	Amount of output	
1	!	NIONCG	!	Number of CG steps for ions	
1	!	NITFIX	!	Keep 1 (sort of dummy)	
1	!	INRAND	!	WFs initialization: 0-diamond 8at. 1-random	
1	!	ICLOCK	!	Timer: 0-off, 1-on	
1.00000	!	SITIM	!	Dummy (unit cell time step)	
0.20000	!	SIDAMP	!	Dummy (unit cell damping)	
100.00000	!	SIMASS	!	Dummy (unit cell mass)	
0.01000	!	SIDISP	!	Dummy (allowed unit cell change)	
0.00020	!	POTIM	!	Timestep for ionic dynamics (ps)	
0.05000	!	PODISP	!	Maximum displacement of ions	
3	!	ICHARG	!	The charge on atom species 1 (B)	
5	!	ICHARG	!	The charge on atom species 2 (N)	
10.81000	!	POMASS	!	The mass of atom species 1 (B)	
14.00670	!	POMASS	!	The mass of atom species 2 (N)	
1	!	NIONSP	!	The number of atoms of species 1 (B)	
1	!	NIONSP	!	The number of atoms of species 2 (N)	
2	!	IVPTYP	!	Type of potential(0-L,1-NL,2-USP)species 1 (B)	
2	!	IVPTYP	!	Type of potential(0-L,1-NL,2-USP)species 2 (N)	
2001	!	NPSPTN	!	Number of rec.space data points, species 1 (B)	
2001	!	NPSPTN	!	Number of rec.space data points, species 2 (N)	
1.15734	!	SCAL	!	Dummy	
1	!	ISYMM	!	Symmetrization switch: 0-off,1-on	
0	!	IBANS	!	Band structure mode: 0-no,1-yes	
0	!	ICNSTR	!	External constraints: 0-no,1-yes	
0	!	IDYNAM	!	MD mode: 0-off,1-on	
0.0000050	!	EPSIL	!	Convergence tolerance (eV/at)	
5	!	NEXTWR	!	Backup every NEXTWRth iteration	
0	!	ISTDYN	!	For MD: 0-new run, 1-restart	
100	!	NPREC	!	2 extra electronic steps every NPREC iterations	
0	!	INOSE	!	For MD: 0-microcanonical,1-Nose	
0.00007	!	QNOSE	!	Nose thermostat parameter	
273.00000	!	TEMPER	!	Temperature (K)	
1	!	NVEL	!	Velocity initialization (0-file,>0-random)	
0	!	NMDCG	!	Keep 0 (sort of dummy)	
CASTEP calculation from Cerius2				!	comment
1	!	ISTRESS	!	Stress output : 0-no, 1-yes	
GGA	!		!	Theory level	
2	!	METHOD	!	0-band_by_band, 1-all_bands, 2-density mixing	
20	!	ISEED	!	0-random, otherwise-pseudorandom wavefunctions	
0	!	IWRCHD	!	If nonzero, formatted charge density is created	

Continued...

```

30      !MAX_CYCL! Maximum number of SCF cycles at each step
1       !BAS_CORR! Finite sampling correction: 0-no, 1-yes
0       !MODE_COR! Mode of sampling correction: 0-auto, 1-manual
0.00000 !DE_DlogE! Derivative of Etotal w.r.t. log(Ecutoff)
0.00000 !PRESSURE! External pressure (GPa) and stress (next line)
0.000   0.000   0.000   0.000   0.000   0.000
3       !NUM_CUT ! Points for num. diff. for DE_DlogE
288.00000 ! ECUT  ! Cutoff # 1 for sampling correction (eV)
294.00000 ! ECUT  ! Cutoff # 2 for sampling correction (eV)
300.00000 ! ECUT  ! Cutoff # 3 for sampling correction (eV)
0.00100  !DISP_TOL! Tolerance for RMS displacement
0.05000  !RMSF_TOL! Tolerance for RMS force
0.10000  !RMS_STR ! Tolerance for RMS stress
0.0000200 !ETOT_TOL! Minimizer tolerance for total energy
0        ! OPTICS ! Calculate matrix elements: 0-no, 1-yes
4        !NEL_UP  ! Number of electrons with spin up
4        !NEL_DOWN! Number of electrons with spin down
0        !SPIN_FIX! Spin is fixed (0) or varied
0        !INIT_OCC! Init. occupancy: 0-lowest, 1-all,-1-file
0.20000  ! ROTMIX ! Mixing factor for WF rotation
0.50000  ! OCCMIX ! Mixing factor for occupation numbers
0.50000  !EL_STEP ! Step for all-bands minimizer
7        !MIX_HIST! Density mixing history (0:non-Pulay)
0.80000  ! CH_AMP ! Mixing amplitude for charge
1.50000  ! CH_Q   ! Cutoff wavevector for charge
2.00000  !SPIN_AMP! Mixing amplitude for spin density
1.50000  ! SPIN_Q ! Cutoff wavevector for spin density
0        !RM_DELAY! CG steps before RMM starts (0 is CG)

```

CASTEP does not read the comments part of the lines in this file, indicated by a !. They are included as a brief reminder of the purpose of the variables. Many of the names used in the comments are the same as the variable name in the code itself. CASTEP relies on the fixed order of lines in this file, and on the type of the number on a line, i.e. floating point or integer, being correct. This means that certain variables are purely historic and have no meaning in the current version (dummy variables). Dummy variables will not be discussed further.

The meaning and purpose of each variable is as follows.

```

3      ! NITER ! Number of iterations

```

Iteration has the meaning of a complete self-consistent calculation of the electronic wavefunctions and charge density for a fixed atomic geometry. NITER specifies the maximum number of iterations that CASTEP is allowed to make. Typically the single-point energy calculation just requires one iteration, and geometry optimization should converge in 10-20 iterations. CASTEP stops when the required tolerance is achieved, for both single-point or geometry optimization runs, so in these cases NITER is an upper limit. Molecular dynamics runs use NITER as the requested number of MD timesteps. The UI uses default values of NITER=3 for single-point energy, and 51 for geometry

optimization. CASTEP only allows a limited number of SCF loops within one iteration (see `MAX_CYCL`). This means that for some systems with poor electronic convergence CASTEP might need more than one iteration for the single-point energy run, which is why the default is 3, not 1.

```
1      ! NPRINT ! Forces are printed every NPRINTth iteration
```

`NPRINT` is really an historical variable. If the output mode is brief (i.e., `IPRINT=0`), then the `.cst` output file will only display forces on all atoms at the end of the run. The forces, however, will be stored in a separate file, `.forces`, which can be used to monitor forces as a function of number of iterations. If `NPRINT` set to 1, the `.forces` file will contain forces for each iteration, otherwise, for each `NPRINT`-th iteration. If the output mode is verbose (`IPRINT>0`), then the `.cst` file will also display forces for each `NPRINT`-th iteration.

```
300.00000      ! ENMAX ! Kinetic energy cut-off (eV)
```

`ENMAX` sets the energy cutoff for wavefunctions, E_c , in eV

```
0      ! NPOT  ! Potential: 0-reciprocal, 1-real space
```

Norm conserving pseudopotentials can be used either in reciprocal or real space representation (it's not possible to mix the two). In either case, the `.recpot` input file is always required. Real space potentials require an additional file, `.realpot`. The real-space option is not yet supported for ultrasoft pseudopotentials.

```
5      ! NITMAX ! Number of electronic CG steps
```

`NITMAX` has the meaning of the number of conjugate gradients steps in any CG-based electronic structure algorithm, or the history depth for the RMM-DIIS scheme. In either case, a value of 5 should be sufficient.

```
0      ! ISTART ! Mode: 0-new run, 1-restart
```

`ISTART` has to be set to 1 when the run is either a direct continuation of the previous one, or for a band structure run (`IBANS=1`).

```
1      ! ISBROT ! Rotation: 0-NO,1-strict, 2-fast METALS only
```

This switch is only relevant for norm-conserving pseudopotentials. It is recommended to set it to 1. Subspace rotation at the end of the conjugate gradients loop is necessary to obtain eigenvalues, and is thus required for metallic systems. There is no good reason not to use the subspace rotation (`ISBROT=0`), and the `ISBROT=2` option is not thoroughly tested.

```
0      ! IOCCUP ! Partial occupation: 0-no, 1-yes
```

IOCCUP determines whether the system is treated as a metal (IOCCUP=1) or as an insulator (IOCCUP=0). Metallic system should have NBANDS set to a number bigger than half the number of valence electrons. The occupation numbers for the bands are then determined by introducing artificial gaussian smearing on each band and finding the area of the smeared level which falls below the Fermi energy. IOCCUP=0 case means that each level has an occupancy of either 0 or 1.

```

0.20000      ! DELMIN ! Minimum smearing width (eV)
4.00000      ! DELMAX ! Starting smearing width (eV)
        6      ! NDEL   ! Smearing width is halved every NDEL SCF loops

```

The width of the smearing function which is used for metallic calculation is determined by these three numbers. The width is taken to be DELMAX for the first SCF loop, then it's halved every NDEL loops or when the total energy is converged at a given smearing, whichever happens first. The width is halved until it becomes equal to DELMIN. The DELMIN value should not be made too small, as this can lead to instabilities. A large DELMIN might help to converge metallic systems faster, but there will be an unaccounted-for error in atomic forces and stresses. Experience shows that a value of 0.2 provides a good intermediate solution. DELMAX should be set ≈ 4.0 unless the density mixing scheme is used for electronic minimization, when DELMAX should be ≈ 0.4 — in this case the algorithm is sufficiently stable not to require long quenching.

```

0      ! IION   ! Ionic relaxation: 0-no, 1-CG, 2-simple, 3-BFGS
0      ! IBOX   ! Unit cell relaxation: 0-no, 1-uniform, 2-BFGS

```

IION=0 means that the fractional coordinates of atoms are fixed. IION=3 is the properly supported mode, and 1 and 2 can only be used when the cell parameters are fixed. IION=3 implies that the `.optim` input file is required, containing an initial guess for the Hessian. The updated `.optim_out` file is written at the end of the run.

N.B: the use of ultrasoft pseudopotentials places restrictions on the choice of IION to 3, BFGS. This is because the CG and simple methods, after moving the ions, update only the data associated with norm-conserving pseudopotentials.

IBOX=0 means that the cell parameters are fixed. Full geometry optimization (cell + internal degrees of freedom) requires IBOX=2 and IION=3. IBOX=1 only makes sense for cubic crystals; in this case CASTEP changes the cell constants uniformly on the way to the energy minimum, and at the end it also produces the bulk modulus and its pressure derivative.

```

0      ! IPRINT ! Amount of output

```

IPRINT controls the amount of detail in the `.cst` file: produces brief output,

and 1 gives more information about the SCF process and prints components of stresses and forces. It rarely necessary to use higher settings, unless you need a very detailed description to help track down a problem with the code.

```
1      ! NIONCG ! Number of CG steps for ions
```

NIONCG is only relevant when performing conjugate gradients geometry optimisation (IIION=1). The conjugate gradients search is in the combined space of electronic and atomic degrees of freedom. NIONCG is the number of conjugate search directions examined before re-setting to the steepest-descent search vector. Therefore, 1 means that the steepest descents algorithm is used. Values of NIONCG between 3 and 10 are recommended. Higher values will probably not improve the serch efficiency.

```
1      ! INRAND ! WFs initialization: 0-diamond 8at. 1-random
```

An historical variable, but has to be set to 1.

```
1      ! ICLOCK ! Timer: 0-off, 1-on
```

ICLOCK=0 means that only the total timing will be included in the file. ICLOCK=1 provides timings after each SCF loop and in some other important places in the code. Note that the timings reported calculated from CPU time, not wall time.

```
0.00020      ! POTIM ! Timestep for ionic dynamics (ps)
```

MD calculations employ a fixed time step in picoseconds, defined by POTIM. The total length of the simulation is then POTIM×NITER

```
0.05000      ! PODISP ! Maximum displacement of ions
```

PODISP has different meanings within different geometry optimization modes. When IIION is 1 or 2, PODISP ;imits the step length in every cartesian component for every atom. When IIION is 3, i.e. when performing BFGS minimisation, PODISP limits the relative change in bond length between any two atoms. A value of 0.05 works quite well in either case.

```
3      ! ICHARG ! The charge on atom species 1 (B )
```

```
5      ! ICHARG ! The charge on atom species 2 (N )
```

ICHARG is number of valence electrons for each atomic species. This is not the nominal valence, but the number of electrons treated as valent during pseudopotential generation (these might include semicore states). The numbers can be extracted from the pseudopotential files: see section 7.4.

```
10.81000     ! POMASS ! The mass of atom species 1 (B )
```

```
14.00670     ! POMASS ! The mass of atom species 2 (N )
```

```
1           ! NIONSP ! The number of atoms of species 1 (B )
```

```

1      ! NIONSP ! The number of atoms of species 2 (N )
2      ! IVPTYP ! Type of potential(0-L,1-NL,2-USP)species 1 (B )
2      ! IVPTYP ! Type of potential(0-L,1-NL,2-USP)species 2 (N )

```

The parameters POMASS are the atomic masses for each species, and are only relevant for molecular dynamics. NIONSP is the number of atoms of each kind in the unit cell. The pseudopotential type IVPTYP is usually 1 (norm-conserving) or 2 (ultrasoft). In the POTENTIALS and POTENTIALS_NC sub-directories there are a few completely local (IVPTYP=0) potentials (e.g., the norm-conserving potential for Mg).

```

2001    ! NPSPTN ! Number of rec.space data points, species 1 (B )
2001    ! NPSPTN ! Number of rec.space data points, species 2 (N )

```

Defines the number of grid points used for storage of reciprocal space potentials.

```

1      ! ISYMM ! Symmetrization switch: 0-off,1-on

```

ISYMM=0 means that the structure is assumed to have no symmetry elements. If ISYMM=1, then the `.symm` input file is required, containing a list of symmetry operations. Note that the set of k-points in `.geom` should be consistent with the symmetry setting, i.e. if symmetry is exploited the k-point set should be symmetrised.

```

0      ! IBANS ! Band structure mode: 0-no,1-yes

```

A band structure calculation is performed when IBANS=1, determining the electronic energies in a fixed potential, i.e., it is the setting for a non self-consistent run. For single-point energy, geometry optimization or MD runs, IBANS is set to 0. If IBANS=1, ISTART should be set to 1 as well, since CASTEP will attempt to read the charge density from a previous calculation. The number of bands and the number/coordinates of k-points for the band structure run are unrelated to the settings of the original run.

```

0      ! ICNSTR ! External constraints: 0-no,1-yes

```

The setting ICNSTR=0 should be used. Any additional constraints may be added via the `.optim` file.

```

0      ! IDYNAM ! MD mode: 0-off,1-on

```

This is the MD switch.

```

0.0000050    ! EPSIL ! Convergence tolerance (eV/at)

```

This parameter specifies the accuracy of electronic structure optimization, in eV per atom. Atomic motion (either for MD or for geometry optimization) only starts when this tolerance is satisfied.

```
5      ! NEXTWR ! Backup every NEXTWRth iteration
```

CASTEP produces a full set of backup files at regular intervals. Unless the job is killed while writing one of those files, it should be possible to restart later by simply changing ISTART to 1.

```
0      ! ISTDYN ! For MD: 0-new run, 1-restart
```

Setting ISTDYN=0 allows you to continue a previous MD run using the final atomic positions but with re-generated thermal velocities. The value of ISTDYN is only relevant if ISTART=1 (restart).

```
100    ! NPREC ! 2 extra electronic steps every NPREC iterations
```

Not used.

```
0      ! INOSE ! For MD: 0-microcanonical,1-Nose
0.00007 ! QNOSE ! Nose thermostat parameter
```

The switch INOSE turns the Nosé thermostat on or off. INOSE=0 generates the NVE ensemble, INOSE=1 the NVT. The Nose thermostat effective capacity is QNOSE, and is only relevant for NVT MD runs.

```
273.00000 ! TEMPER ! Temperature (K)
```

TEMPER is used to initialize velocities during MD. The actual temperature fluctuates around TEMPER, even for NVT runs.

```
1      ! NVEL ! Velocity initialization (0-file,>0-random)
```

Setting NVEL=0 means that velocities are taken from the .geom file, while NVEL=1 will force CASTEP to generate initial velocities using the TEMPER value.

```
CASTEP calculation from Cerius2 ! comment
```

The title line: it is reproduced in the .cst file

```
1      ! ISTRESS! Stress output : 0-no, 1-yes
```

ISTRESS=1 forces CASTEP to calculate the stress tensor even when the cell parameters are fixed (IBOX=0).

```
GGA      ! Theory level
```

Allowed values for the exchange-correlation functional are: LDA, GGA, LSDA, GGS, LDA+Gradient Corr., LDA+GC The last two entries in the list are just old names for GGA.

```
2      ! METHOD ! 0-band_by_band, 1-all_bands, 2-density mixing
```

METHOD=0 selects the original Payne-Teter-Allan band-by-band conjugate

gradients scheme when using norm-conserving potentials, or employs a slightly modified CUSP code from David Vanderbilt when using ultrasoft potentials. METHOD=1 only applies to norm-conserving potentials, and it selects conjugate gradients minimisation applied to all bands simultaneously. METHOD=2 selects the recommended mode, where a self-consistent density mixing scheme is used. This operates with potentials of either class.

```
20      ! ISEED ! 0-random, otherwise-pseudorandom wavefunctions
```

ISEED is normally set to zero, and it results in different random initial wavefunctions each time the code is run. A non-zero value may be useful for quality control, since each nonzero ISEED generates a deterministic and reproducible pseudo-random sequence.

```
0      ! IWRCHD ! If nonzero, formatted charge density is created
```

CASTEP usually outputs the charge density as a binary file only. If you require a formatted file (e.g., for easy transfer from a Cray machine) then IWRCHD=1 should be used. The formatted file is only created at the very end of the run, while the binary file is backed-up regularly during the calculation.

```
30      !MAX_CYCL! Maximum number of SCF cycles at each step
```

Each CASTEP iteration can only make MAX_CYCL number of SCF loops. A value of 30 is usually sufficient, but if SCF convergence is not achieved in MAX_CYCL loops, then CASTEP still moves on to the next iteration (out of NITER). Note that because in general metals need more steps than insulators, CASTEP uses 5*MAX_CYCL as the number of SCF steps for metals (i.e. when partial band occupation are allowed).

```
1      !BAS_CORR! Finite sampling correction: 0-no, 1-yes
0      !MODE_COR! Mode of sampling correction: 0-auto, 1-manual
0.00000 !DE_DlogE! Derivative of Etotal w.r.t. log(Ecutoff)
```

BAS_CORR=1 means that the finite basis set correction will be used. This correction effectively allows one to get away with the k-points set and energy cutoff that are not high enough for absolute convergence, and still perform cell optimization.

MODE_COR=0 means that CASTEP will calculate the finite basis set correction during the run. The value that needs to be calculated is the derivative of the total energy with respect to $\log E_c$. MODE_COR=1 assumes that this value is already known (e.g., from the previous run performed for the same system with exactly the same run parameters). The value of $\partial E / \partial \log E_c$, DE_DlogE (in eV), is required when MODE_COR=1 is used.

```
0.00000 !PRESSURE! External pressure (GPa) and stress (next line)
0.000   0.000   0.000   0.000   0.000   0.000
```

This is the external applied stress (pressure is the trace of the tensor, and the tensor is given in a short notation as xx,yy,zz,yz,xz,xy).

```

      3      !NUM_CUT ! Points for num. diff. for DE_DlogE
288.00000  ! ECUT  ! Cutoff # 1 for sampling correction (eV)
294.00000  ! ECUT  ! Cutoff # 2 for sampling correction (eV)
300.00000  ! ECUT  ! Cutoff # 3 for sampling correction (eV)

```

When `MODE_COR=0`, $\partial E/\partial \log E_c$ is calculated automatically by finding the total energy at a number of reference cut-off energies. The value of `NUM_CUT`, specifies the number of energies which are given on the following `NUM_CUT` lines. The last `ECUT` should be the same as `ENMAX`. A default setting of `NUM_CUT=3` and a step between the reference energies of 2% of `ENMAX` should work in most cases. Note that if the difference between the reference energies is so small that the number of plane waves used does not change from one `ECUT` to the next, the calculation will not make sense.

```

      0.00100  !DISP_TOL! Tolerance for RMS displacement
      0.05000  !RMSF_TOL! Tolerance for RMS force
      0.10000  !RMS_STR ! Tolerance for RMS stress
      0.0000200 !ETOT_TOL! Minimizer tolerance for total energy

```

There are several parameters controlling geometry optimisation: It will continue until the ionic mean-square displacement is less than `DISP_TOL` (Å), or until the mean-square force on movable ions is less than `RMSF_TOL`. Geometry optimization continues until the mean-square stress is below `RMS_STR`. Only components that correspond to variable cell parameters are taken into account (e.g., only diagonal terms for a cubic cell). Finally, optimization continues until the change of the total energy between successive configurations is less than `ETOT_TOL`. NOTE: All the relevant criteria must be satisfied for the run to stop, i.e., there are four criteria to satisfy when both cell and internal degrees of freedom are allowed to relax.

```

      0      ! OPTICS ! Calculate matrix elements: 0-no, 1-yes

```

Running with `OPTICS=1` is only possible when `IBANS=1`. A calculation with `OPTICS=1` produces a `.cst_ome` file, containing matrix elements for optical transitions, after the band structure run. It is expected that the Monkhorst-Pack set of k-points is used when `OPTICS=1`.

```

      4      !NELUP  ! Number of electrons with spin up
      4      !NEL_DOWN! Number of electrons with spin down
      0      !SPIN_FIX! Spin is fixed (0) or varied

```

The use of a spin-polarized exchange-correlation functional is required when the numbers of spin=up and spin-down electrons differ. `SPIN_FIX` only affects runs using spin-polarized exchange-correlation functionals. If its value is zero, then the spin of the system is fixed. If its positive, then the spin is fixed

for the first SPIN_FIX SCF loops and is allowed to vary thereafter (which is required for magnetic metallic systems). A negative value of SPIN_FIX means that spin is allowed to vary from the very first loop.

```
0      !INIT_OCC! Init. occupancy: 0-lowest, 1-all,-1-file
```

The parameter INIT_OCC specifies how occupation numbers are initialized:

- 0 — the lowest bands are filled, and the higher ones are empty if their number is greater than half of the number of electrons.
- 1 — all bands are filled equally, even if this creates partial occupancies
- 1 — exotic setting, CASTEP reads starting occupation from the .ocnum file. This might be useful, e.g., for simple atoms where user knows the degeneracies.

```
0.20000      ! ROTMIX ! Mixing factor for WF rotation
0.50000      ! OCCMIX ! Mixing factor for occupation numbers
0.50000      !EL_STEP ! Step for all-bands minimizer
```

These parameters are only used by the all=bands minimiser. Recommended values are 0.2, 0.5 and 0.5 respectively, and these should not be altered.

```
7           !MIX_HIST! Density mixing history (0:non-Pulay)
0.80000     ! CH_AMP ! Mixing amplitude for charge
1.50000     ! CH_Q   ! Cutoff wavevector for charge
```

MIX_HIST specifies the depth of the history stack in the Pulay scheme of the density mixing method (DIIS history). A value of 0 means that the density mixing scheme is not using the Pulay algorithm. CH_AMP specifies the mixing amplitude for the charge density. It might be useful to decrease it slightly if convergence is not sufficiently stable. CH_Q specifies the cutoff wavevector: small- G components are mixed with smaller coefficients as prescribed by Kerker. A value of zero means that linear mixing is used, rather than Kerker mixing.

```
2.00000     !SPIN_AMP! Mixing amplitude for spin density
1.50000     ! SPIN_Q ! Cutoff wavevector for spin density
0           !RM_DELAY! CG steps before RMM starts (0 is CG)
```

SPIN_AMP and SPIN_Q have the same meanings as CH_AMP and CH_Q, but refer to the spin density (these variables are irrelevant for non spin-polarized calculations). RM_DELAY specifies whether to use the conjugate gradients scheme in the density mixing approach (RM_DELAY=0) or the DIIS-based residual minimization. Setting RM_DELAY>0 means that the RMM-DIIS algorithm is used, but the first RM_DELAY steps are still done by conjugate gradients.

7.3 Runtime parameters: .param

This file contains parameters used to define the dimensions of various CASTEP arrays.

```

12      * NGX      * FFT grid dimension
12      * NGY      * FFT grid dimension
12      * NGZ      * FFT grid dimension

```

These parameters define the FFT grid, whose size depends on the cutoff energy. The grid should be fine enough to prevent Fourier aliasing. This occurs when the period in reciprocal space is too small to prevent the sphere of plane waves overlapping with its periodic images. To completely avoid aliasing, the following prescription should be used:

$$NGX = 4L\sqrt{\frac{2m_e E_c}{\hbar^2}}, \quad (1)$$

and similarly for the other two dimensions. Here, L is the length of the real-space lattice vector, m_e is the electron mass, E_c the cutoff energy and \hbar is Planck's constant. With L in Å and E_c in eV this becomes

$$NGX = 0.32615L\sqrt{E_c}. \quad (2)$$

It is possible to run a plane-wave code with a grid smaller than the recommended one. For example, the Cerius2 interface typically sets the dimensions to 3/4 of those given by the above equation. The motivation for this is increased execution speed, but be warned, one must quantify errors that arise from using a smaller grid. Note also that ultrasoft pseudopotentials and gradient-corrected functionals (GGA) both require a finer FFT grid than is needed without them, assuming all other variables are the same.

```

2001      * NPSPTS * Max. dimension of reciprocal space data

```

This is the maximum size of the reciprocal-space dataset for the potentials being used, and is the largest of the NPSPTN values in the .rundat file. It is almost always 2001.

```

160      * NRPLWV * Max. number of plane waves accepted

```

This is the maximum number of plane waves needed to describe $|Psi_k\rangle$. The number of plane waves is different for different k , and it depends on the spacing between points in reciprocal space and the relative sizes of k and $G_{max} = \sqrt{2m_e E_c}/\hbar$. In the limit of fine reciprocal-space grid (large simulation box) and large E_c , $NRPLWV \approx NGX * NGY * NGZ / 16$. This assumes that a full-size, zero-overlap grid is used. If all grid dimensions are reduced by a factor f , then $NRPLWV \approx NGX * NGY * NGZ / (16 f^3)$. Note that the number of plane waves is constant even if the cell optimization is performed.

```

1      * NIONS * Max. number of atoms of any kind
2      * NSPEC * Number of different atomic species
2      * NIONST * Total number of atoms
4      * NBANDS * Number of electron bands
10     * NKPTS * Number of special k-points

```

NIONS is the largest of all NIONSP values in the `.rundat` file. NSPEC and NIONST are self-explanatory NBANDS should be at least a half of the number of electrons as given in the `.rundat` file (`NEL_UP + NEL_DOWN`). It is necessary to include some empty bands in calculations on metals. NKPTS should agree with the contents of `.geom` and `.symm` (where applicable): CASTEP will attempt to read NKPTS lines from those files.

```

1      * NRGRPT * Dimension of arrays for real-space potentials

```

This is usually calculated by the Cerius interface, but if set too small, CASTEP will increase it at runtime. This value can be also increased at runtime as a result of atomic motion during MD or geometry optimization.

```

1      * MXRLNL * Number of real-space nonlocal projectors
1      * MXRLSH * Number of real-space spherical harmonics
1      * NRLPTS * Max. dimension of real-space data

```

These parameters must be set if real-space pseudopotentials are used. In the BN example the potentials are evaluated in reciprocal space and these parameters are set to 1. However, if using the real-space option, their values might be: MXRLNL=2 if all the potentials have only two nonlocal components, e.g., s and p ; MXRLSH = 1 + 3 = 4 (1 for s , 3 for p ; NRLPTS= 2001, having the same meaning as NPSPTS, but for the real-space radial grid. It is usually 2001.

```

0      * N_COORD* Degrees of freedom for atomic coordinates
0      * N_CELL * Degrees of freedom for cell vectors

```

The parameter N_COORD=3×number of *moveable* atoms, i.e. those with a non-zero moveability switch in `.geom`/ N_COORD=0 if IION=0 (in `.rundat`). N_CELL is the number of independent degrees of freedom for cell variables. E.g., N_CELL=1 for cubic cells with cell optimization switched on, i.e. IBOX;0 in `.rundat`. Both N_COORD and N_CELL are only used for BFGS optimization, that is, when IION=3 and and/or IBOX=2.

7.4 The pseudopotentials: `.recpot` and `.realpot`

Each CASTEP calculation needs a single `.recpot` file, and if the pseudopotentials are to be evaluated in real space, a single `.realpot` file. These files contain data taken from the pseudopotential library files, or generated from them. The distribution contains two libraries, POTENTIALS and POTENTIALS_NC. In these libraries there are several kinds of files, whose suffixes

signify the type of pseudopotential they contain, as detailed in table 4. Note that the CASTEP input file containing the reciprocal space potentials, and the library files containing a *single* norm-conserving potential, share the same suffix, `.recpot`. You should note that the chosen ultrasoft pseudopotentials and similar technologies have been chosen as the preferred development route for CASTEP.

<code>.usp</code>	Vanderbilt Ultrasoft pseudopotentials
<code>.uspc</code>	As above but with core charge
<code>.recpot</code>	Norm-conserving reciprocal space pseudopotential in Kleinman-Bylander form
<code>.psp</code>	Teter pseudopotential
<code>.pspnc</code>	Troullier-Martins pseudopotential

Table 4: Pseudopotential files

A full description of the pseudopotentials will not be given here. However, you will need to know some details of the structure of the input files `.recpot` and `.realpot`, so that you can create them for your calculations.

7.4.1 Preparing `.recpot` when using ultrasoft pseudopotentials

All calculations require a `.recpot` file, though its contents depend on the types of potential you are using. In all cases the complete reciprocal-space pseudopotential for each species must be in this file, in the same order as the species appear in the `.geom` and `.rundat` files. You must also tell CASTEP which types of potentials are being used by setting the `IVPTYP` parameters in the `.rundat` file.

The overall structure of `.recpot` is as follows. The first line contains two numbers which denote the CASTEP version number for which the potentials were generated². Next comes the pseudopotential for the first species, with a line containing the number “1000” marking the end of the data. Then comes the pseudopotential for species 2, and so on, with each data set terminated by “1000”.

The library files contain ancilliary information which must be removed before they are assembled into a `.recpot` file. It is easier to understand this by working through the building of the `.recpot` file for the BN example. In this case, both pseudopotentials are ultrasoft ones. The boron potential is taken from the `POTENTIALS/B.00.usp` file. In this file you will see that there is a comments section, delimited by ”`START COMMENT` and `END COMMENT` lines.

²Versions earlier than 3.0 did not require this line in `.recpot`. If the version line is missing, CASTEP assumes that `.recpot` conforms to this older format.

All the `.usp` and `.recpot` files contain a comments section, in which the construction and testing of the pseudopotential is described. In addition there are recommended plane-wave cutoff energies for `COARSE`, `MEDIUM` and `FINE` grids. The entire comments section, including the delimiting lines, must be removed when inserting the pseudopotential data into `.recpot`.

The line after the `END COMMENT` line contains the version information for this potential. However, in `.recpot`, the first line contains version information for **all** the pseudopotentials. If `.recpot` contains a single potential there is no ambiguity. However, for systems with more than one species, the highest version among the potentials should be taken and inserted into `.recpot` as the first line.

In `BN.usp` and for all ultrasoft potentials, the next line gives the core charge for the potential. This is the value to which `ICHARG` should be set in the `.rundat` file. However, it is not required in the `.recpot` file. All data after this line is inserted into `.recpot`, and the termination line “1000” added immediately after the data.

Gathering this together, the structure of the ultrasoft pseudopotential files is as follows.

Comments Section delimited by `START COMMENT` and `END COMMENT` lines, containing generation and testing information

Version Line with the version of `CASTEP` for which the potential was made

Core Charge Line with the pseudopotential core charge

Pseudopotential Section containing all data for the pseudopotential

The stages in preparing `.recpot` are as follows.

1. Determine the highest version number of the potentials you are using from the Version Line and insert it as the first line of `.recpot`
2. Set the `ICHARG` parameters in `.rundat` to the core charges appearing on the Core Charge Lines.
3. Remove the Comments Section, Version Line and Core Charge Line from all the potentials, and append a line with the number “1000” to them.
4. Insert all the potentials into the `.rundat` file, in the same order as the species appear in the other input files.

7.4.2 Preparing .recpot when using norm-conserving potentials

The norm-conserving potentials are in the .recpot files in the POTENTIALS_NC/ directory. (Remember, both the CASTEP input file .recpot and the norm-conserving potential files share the same suffix.) Almost all the foregoing description relating to constructing the CASTEP input file .recpot when using ultrasoft pseudopotentials carries over to norm-conserving ones. There is, in fact, only one difference. The norm-conserving potentials do not contain the Core Charge line. The steps in assembling .recpot are exactly as described above, but you now need to find the core charges for the pseudopotentials by other means. Usually the Comments Section of the potential file contains enough information to deduce the core charge. The configurations used in generating the potential indicate which states were treated as valence states. Subtraction of the number of core states from the atomic number gives the core charge.

A second method, which is always reliable, is based on the nature of the local potential V . As the magnitude of the wavevector $q \rightarrow 0$, corresponding to large distances, V tends to that of a point charge, given by

$$V = Z/q^2. \quad (3)$$

For a norm-conserving pseudopotential we can find the smallest wavevector and the corresponding local potential. The line after the Version Line contains the radial grid maximum PSGMAX (in $1/\text{\AA}$), with the potential tabulated on a grid of (NPSPTN-1) points. NPSPTN is almost always 2001, though some very early potentials used a value of 501. The smallest wavevector is therefore (PSGMAX/(NPSPTN-1)). The value of the V at the smallest wavevector is second number in the following line. Including the necessary conversion factors, the pseudopotential core charge is given by the nearest integer to

$$\text{ICHARG} = \frac{|V|}{180.95263} \left(\frac{\text{PSGMAX}}{\text{NPSPTN} - 1} \right)^2 \quad (4)$$

For the boron norm-conserving pseudopotential POTENTIALS_NC/B_00.recpot this becomes:

$$\text{ICHARG} = \frac{217140.671875}{180.95263} \left(\frac{100}{2000} \right)^2, \quad (5)$$

yielding $\text{ICHARG} = 3$.

7.4.3 The real-space pseudopotential file .realpot

This file contains the non-local pseudopotential in real space. There is a considerable advantage to evaluating the pseudopotential in real space for large systems, since then the calculation scales more favourably with system

size. However, you will notice immediately that there are no `.realpot` files in the pseudopotential libraries. In fact, it is not possible to provide a single transformation of the reciprocal-space potential since the process is cutoff-dependent. Also, we are not able at present to provide a transformation code with the distribution. `CETEP` users should note that the `.realpot` file has the same structure as the `CETEP` input file `fort.18`. Note also that ultrasoft pseudopotentials may only be evaluated in reciprocal space in this version of `CASTEP`.

7.5 Symmetry information: `.symm`

This file contains information on the symmetry of the structure specified by `.geom`, as well as a symmetrised set of k-points. Symmetry is used in the calculation, and hence this file read, when `ISYMM=1` in the `.rundat` file. For high-symmetry crystals this can result in a substantial speedup, but the benefit is less for low-symmetry crystals and surfaces. Using symmetry during geometry and cell optimisation preserves the initial symmetry, which while useful, may result in metastable configurations at the end of the optimisation. Symmetry cannot be exploited during MD runs.

From a given `.geom` file, the utility program `K290` in the `utilities/K290` directory will generate the appropriate `.symm` file. This program was kindly provided by Karel Kunk, and if you make use of it please cite the reference given in section .

This utility is not supported as part of the distribution. However, you will find brief instructions in the file `sym.txt`. The control parameters are in the `fort.4` file. The input files `fort.14` and `fort.15` correspond to `<seed>.rundat` and `<seed>.geom` respectively. The output file `fort.3` is equivalent to the `CASTEP` input file `<seed>.symm`

Part III

For Reference

8 Some UKCP publications

1. I. Štich, M. C. Payne, R. D. King-Smith, J.-S. Lin and L. J. Clarke, “Ab initio total energy calculations for extremely large systems: application to the Takayanagi reconstruction of Si (111)”, *Phys. Rev. Lett.*, **68** 1351 (1992)
2. A. De Vita, M. J. Gillan, J.-S. Lin, M. C. Payne, I. Štich and L. J. Clarke, “Defect energetics in oxide materials from first principles”, *Phys. Rev. Lett.*, **68**, 3319 (1992)
3. L. J. Clarke, I. Štich and M. C. Payne, “Large-scale ab initio total energy calculations on parallel computers”, *Comput. Phys. Commun.*, **72**, 14 (1992)
4. M. C. Payne, I. Štich, R. D. King-Smith, J.-S. Lin and L. J. Clarke, “Ab initio total energy calculations on parallel computers: application to the Takayanagi reconstruction”, *Phys. Scripta*, **T 45**, 265 (1992)
5. A. De Vita, I. Manassidis, J.-S. Lin and M. J. Gillan, “The energetics of Frenkel defects in Li₂O from first principles”, *Europhys. Lett.*, **19**, 605 (1992)
6. A. De Vita, M. J. Gillan, J.-S. Lin, M. C. Payne, I. Štich and L. J. Clarke, “Defect energetics in MgO treated by first-principles methods”, *Phys. Rev. B*, **46**, 12964 (1992)
7. J. R. K. Bigger, D. A. McInnes, A. P. Sutton, M. C. Payne, I. Štich, R. D. King-Smith, D. M. Bird and L. J. Clarke, “The atomic and electronic structures of the 90° partial dislocation in silicon”, *Phys. Rev. Lett.*, **69**, 2224 (1992)
8. M. C. Payne, L. J. Clarke and I. Štich, “Role of parallel architectures in periodic boundary calculations”, *Phil Trans. Roy. Soc.*, A **341**, 211 (1992)
9. D. M. Bird, L. J. Clarke, R. D. King-Smith, M. C. Payne, I. Štich and A. P. Sutton, “First-principles calculation of the structure and energy of Si (113)”, *Phys. Rev. Lett.*, **69**, 3785 (1992)
10. I. Manassidis, A. De Vita and M. J. Gillan, “Structure and energetics of the (0001) surface of alpha-alumina from first-principles calculations”, *Surf. Sci. Lett.*, **285**, L517 (1993)

11. I. Manassidis, A. De Vita and M. J. Gillan, "Parallel computation and the structure of catalyst surfaces", *Proc Int. Conf. 'Physics Computing '92', Prague, August 1992* (Singapore, World Scientific 1993), p. 397
12. I. Štich, M. C. Payne, A. De Vita and M. J. Gillan, "First-principles dynamics used to study dissociative chemisorption", *Proc. Int. Conf. 'Physics Computing '92', Prague, August 1992* (Singapore, World Scientific, 1993), p. 477
13. M. C. Payne, I. Štich, R. D. King-Smith, J.-S. Lin, A. De Vita, M. J. Gillan and L. J. Clarke, "Ab initio simulations on parallel computers", *'Computer Aided Innovation of New Materials'*, ed. M. Doyama, J. Kihara, M. Tanaka and R. Yamamoto (Elsevier, 1993), p. 101
14. I. Štich, M. C. Payne, A. De Vita, M. J. Gillan and L. J. Clarke, "First-principles studies of semiconductor surfaces: reconstruction and dissociative chemisorption", *Proc. NATO ARW on 'Computations for the Nanoscale'*, NATO ASI Series E, Vol. 240 (1993), p. 63
15. A. De Vita, I. Štich, M. J. Gillan, M. C. Payne, R. D. King-Smith and L. J. Clarke, "Dynamics of dissociative chemisorption: $\text{Cl}_2/\text{Si}(111)\text{-}2\times 1$ ", *Phys. Rev. Lett.*, **71**, 1276 (1993)
16. J. Crain, G. J. Ackland, R. O. Piltz and P. D. Hatton, "Phase-transition-induced defect formation in III-V semiconductors", *Phys. Rev. Lett.*, **70**, 814 (1993)
17. M. J. Gillan, "Gulliver among the atoms", *New Scientist*, 3rd April 1993, p. 34
18. M. J. Gillan, "First applications of a new approach", *Europhys. News*, **23**, 176 (1992)
19. I. Štich, M. C. Payne, A. De Vita, M. J. Gillan and L. J. Clarke, "Chemically driven molecular decomposition at semiconductor surfaces", *Chem. Phys. Lett.*, **212**, 617 (1993)
20. I. Štich, A. De Vita, M. C. Payne, M. J. Gillan and L. J. Clarke, "Surface dissociation from first principles: dynamics and chemistry", *Phys. Rev. B*, accepted
21. I. Manassidis and M. J. Gillan, "The structure and energetics of alumina surfaces calculated from first principles", *J. Amer. Ceram. Soc.*, **77**, 335 (1994)
22. M. J. Gillan, I. Manassidis and A. De Vita, "The ab initio energetics of oxides", *Phil. Mag. B*, accepted

23. D. M. Bird, L. J. Clarke, M. C. Payne and I. Štich, "Dissociation of H₂ on Mg (0001)", *Chem. Phys. Lett.*, **212**, 518 (1993)
24. J. A. White and D. M. Bird, "Dissociative adsorption of H₂ on Cu (100)", *Chem. Phys. Lett.*, **213**, 422 (1993)
25. J. Crain, S. J. Clark, G. J. Ackland, M. C. Payne, V. Milman, P. D. Hatton and B. J. Reid, "Theoretical study of high-density phases of covalent semiconductors: I. Ab initio treatment", *Phys. Rev. B*, **49**, 5329 (1994)
26. E. S. Fois, J. N. Penman and P. A. Madden, "Self-interaction corrected density functionals and the structure of metal clusters", *J. Chem. Phys.*, **98**, 6352 (1993)
27. E. S. Fois, J. N. Penman and P. A. Madden, "Control of the adiabatic ground state in ab initio molecular dynamics", *J. Chem. Phys.*, **98**, 6361 (1993)
28. J. Crain, G. J. Ackland, P. D. Hatton, M. I. McMahon, R. J. Nelmes and R. O. Piltz, "Image plate studies of tetrahedral semiconductors under pressure: implications for experiment and theory", *Jap. J. Appl. Phys.*, **32**, 36 (1993)
29. H. Akbarzadeh, S. J. Clark and G. J. Ackland, "Theoretical study of selenium I under high pressure", *J. Phys.: Condens. Matter*, **5**, 8065 (1993)
30. J. Crain, G. J. Ackland, J. R. Maclean, R. O. Piltz, P. D. Hatton and G. S. Pawley, "Reversible pressure-induced structural transitions between metastable phases of silicon" *Phys. Rev. B* **50**, 13043 (1994).
31. F. Kirchhoff, J. Holender and M. J. Gillan, "The energetics and electronic structure of silver chloride", *Phys. Rev. B*, **49**, 17420 (1994)
32. I. Manassidis and M. J. Gillan, "First-principles study of SnO₂ (110)", *Proc. 10th Interdisciplinary Surface Science Conference, Liverpool, March 1994*, ed. S. Barrett, accepted
33. J. Crain, G. J. Ackland, S. J. Clark, Y. H. Nam, P. D. Hatton, J.-S. Lin, M. C. Payne and R. O. Piltz, "Tetrahedral structures and phase transitions in III-V semiconductors", *Phys. Rev. B*, **50**, 8389 (1994)
34. C. P. Ursenbach and P. A. Madden, "Cluster formation in alkali-doped zeolite-Y: an ab initio study", in preparation
35. J. A. White, D. M. Bird, M. C. Payne and I. Štich, "Surface corrugation in the dissociative adsorption of H₂ on Cu (100)", *Phys. Rev. Lett.* **73** 1404 (1994)

36. C.M. Goringe, L. J. Clark, M. S. Lee, M. C. Payne, I. Stich, J. A. White, M. J. Gillan and A. P. Sutton, "GaAs(001)-(2 × 4) Structure, Chemistry and Adsorbates" *J. Phys. Chem.* Invited feature article. In preparation.
37. S. Pugh and M. J. Gillan, "The energetics of NH₃ adsorption at the MgO (001) surface", *Surf. Science*, **320**, 331-43 (1994)
38. L. N. Kantorovich, J. M. Holender and M. J. Gillan, "The energetics and electronic structure of defective and irregular surfaces of MgO", *Surf. Sci.*, accepted.
39. S.J.Clark, G.J.Ackland, J.Crain and M.C.Payne, "Very low energy surface of silicon", *Phys. Rev. B* **50**, 5728 (1994)
40. G.J.Ackland, "Pressure induced symmetry breaking in tetrahedral networks" *Phys. Rev. B*, **50**, 7389 (1994).
41. S.J.Clark and G.J.Ackland, "Vibrational and elastic effects of point defects in silicon", *Phys. Rev. B* **48**, 10899 (1993).
42. J.Crain, G.J.Ackland, J.R.Maclean, R.O.Piltz, P.D.Hatton and G.S.Pawley, "Reversible pressure-induced structural transitions between metastable phases of silicon", *Phys. Rev. B* **50**, 13043 (1994).
43. S.J.Clark, G.J.Ackland and H.Akbarzadeh, "A theoretical study of pressure effects on selenium", *J.Phys.Chem. Solids* **56**, 329 (1995)
44. J.Crain, S.J.Clark and G.J.Ackland, "Dense tetrahedral structures of compound semiconductors", *J.Phys.Chem. Solids* **56**, 495 (1995)
45. R.O.Piltz, J.R.Maclean, S.J.Clark, G.J.Ackland, P.D.Hatton and J.Crain, "The Structure and Properties of Silicon XII: A Complex Tetrahedrally-bonded Phase", *Phys. Rev. B* **52**, 4072 (1995)
46. H-C Hsueh, J.R.Maclean, G.Y.Guo, M-H. Lee, S.J.Clark, G.J.Ackland and J.Crain, "Pressure-induced polymorphism in CuCl: An ab initio study", *Phys. Rev. B* **51** 12216 (1995)
47. H-C Hsueh, H.Vass, S.J.Clark, G.J.Ackland and J.Crain, "High-Pressure effects in the layered semiconductor Germanium Selenide", *Phys. Rev. B* **51**, 16750 (1995)
48. S.J.Clark, G.J.Ackland and J.Crain, "Theoretical stability limit of diamond at ultra-high pressure", *Phys. Rev. B* **52**, 15035 (1995)
49. S.J.Clark, G.J.Ackland and J.Crain, "A comparison of bonding in amorphous carbon and silicon", *Phys. Rev.B*, submitted.

50. J.A White, D.M. Bird and M. C. Payne, "Dissociation of H₂ on W(100)" *Phys. Rev. B*, **53** 1667 (1996)
51. J.A. White and D.M. Bird, "Implementation of Gradient-Corrected Exchange-Correlation Potentials in Car-Parrinello Total Energy Calculations", *Phys. Rev. B* **50** 4954 (1994)
52. Ruben Perez and Michael C. Payne, "First-Principles Simulations of Silicon Nanoindentation", *Phys. Rev. Lett.*, **75** 4748 (1995)
53. F. Kirchhoff, J. M. Holender and M. J. Gillan, "Complex ion formation in liquid Ag-Se alloys", submitted to *Phys. Rev. Lett.*
54. F. Kirchhoff, J. M. Holender and M. J. Gillan, "The liquid Ag-Se system studied by ab initio simulations", *Proc. 9th Int. Conf. on Liquid and Amorphous Metals, J. Non-Cryst. Solids*, **205-207**, 79 (1996)
55. C.M. Goringe and A. P. Sutton, "Tight binding and local density approximation studies of semiconductor surfaces and small organic adsorbates" *Mat. Sci. Eng. B* **37** 150 (1996)
56. J. M. Holender, M. J. Gillan, M. C. Payne and A. D. Simpson, "Static, dynamic and electronic properties of liquid gallium studied by first-principles simulation", *Phys. Rev. B* **52**, 967 (1995).
57. Philip Lindan and Nic Harrison, "And the rest is chemistry...", HPCProfile **1** 4 (March 1995)
58. E. Hernandez, M. J. Gillan and C. M. Goringe, "Towards a linear-scaling DFT technique: the density matrix approach", *Phys. Rev. B*, **53**, 7147 (1996)
59. R. Shah, A. De Vita and M. C. Payne, "Ab initio study of tritium defects in lithium oxide", *J. Phys.:Cond. Matter*, **6**, 6891 (1995)
60. R. Shah, J. D. Gale and M. C. Payne, "Acid-base catalysis in zeolites from first principles", *Int. J. Quantum Chem.*, **61**, 393 (1997)
61. M. Kay, G. R. Darling, S. Holloway, J. A. White and D. M. Bird, "Steering Effects in Non-Activated Adsorption", *Chem. Phys. Lett.*, **245** 311 (1995)
62. P. A. Gravil, J. A. White and D. M. Bird, "Chemisorption of O₂ on Ag(110)", *Surf. Sci.*, **352-354** 248 (1996)
63. R. Shah, A. De Vita, V. Heine and M. C. Payne, "The mechanism of tritium diffusion in lithium oxide", *Phys. Rev. B*, **53** 8257 (1996)

64. P. J. D. Lindan, N. M. Harrison, J. Holender, M. J. Gillan and M. C. Payne, "The TiO_2 (100) 1×3 reconstruction: Insights from ab initio calculations", *Surf. Sci.* **364** 431 (1996)
65. "First-principles molecular dynamics simulation of water dissociation on TiO_2 (110)", P. J. D. Lindan, N. M. Harrison, J. Holender and M. J. Gillan, *Chem. Phys. Lett.* **261** 246-252 (1996)
66. P. J. D. Lindan, N. M. Harrison, J. A. White and M. J. Gillan, "First-principles spin-polarised calculations on the reduced and reconstructed TiO_2 (110) surface", *Phys. Rev. B* (in press)
67. R. Shah, M. C. Payne, M. H. Lee and J. D. Gale, "Understanding the catalytic behaviour of zeolites: first principles study of adsorption of methanol", *Science*, **271**, 1395 (1996)
68. R. Shah, J. D. Gale and M. C. Payne, "Methanol adsorption in zeolites - a first principles study" *J. Phys. Chem.*, **100**, 11688 (1996)
69. J. Crain, G. J. Ackland and S. J. Clark, "Exotic Structures of tetrahedral semiconductors" (Invited review article), *Rep. Prog. Phys.*, **58**, 705 (1995)
70. M. C. Warren and G. J. Ackland, "Ab initio studies of structural instabilities in magnesium silicate perovskite", *Phys. Chem. Minerals* **23**, 107 (1996).
71. H. C. Hsueh, M. C. Warren, H. Vass, G. J. Ackland, S. J. Clark and J. Crain, "Vibrational properties of the layered semiconducting germanium sulfide under hydrostatic pressure", *Phys. Rev. B*, **53** 14806 (1996)
72. B. B. Karki, L. Stixrude, S. J. Clark, M. C. Warren, G. J. Ackland and J. Crain, "Structure and elasticity of MgO at high pressure", *American Mineralogist*, submitted
73. S. J. Clark, C. J. Adam, J. A. White, G. J. Ackland and J. Crain, "Properties of liquid crystal molecules by first-principles calculation", *Liquid Crystals*, in press.
74. G. J. Ackland, M. C. Warren and S. J. Clark, "Practical methods in ab initio lattice dynamics", *Phys. Rev. B*, submitted.
75. R. Shah, J. D. Gale and M. C. Payne, "Comparing the acidities of zeolites and SAPOs from first principles", *Chem. Comm. Chem. Commun.*, **1**, 131 (1997)
76. R. Shah, J. D. Gale and M. C. Payne, "The active sites of microporous solid acid catalysts" *Phase Transitions B*, submitted for publication, (1996)

77. I. Dawson, P. D. Bristowe, M-H. Lee, M. C. Payne, M. D. Segall and J. A. White, "A first principles study of a tilt grain boundary in rutile", *Phys. Rev. B* **54**, 13727 (1996)
78. M. Segall, R. Shah, C. J. Pickard, M. C. Payne, I. Dawson and P. D. Bristowe, "Population analysis of plane wave electronic structure calculations of bulk materials", *Phys. Rev. B* in press (1996)
79. P. A. Gravil, D. M. Bird and J. A. White, "Adsorption and Dissociation of O₂ on Ag(110)", *Phys. Rev. Lett.*, **77**, 3933 (1996)
80. D. M. Bird and P. A. Gravil, "First Principles Calculations of Molecular Dissociation at Surfaces", *Surf. Sci.*, submitted
81. R. Perez, M. C. Payne, I. Stich and K. Terakura, "First Principles Simulations of Nanoindentation and Atomic Force Microscopy on Silicon Surfaces", *Materials Theory, Simulations and Parallel Algorithms*, Editors: E. Kaxiras, J. Joannopoulos, P. Vashishta and R. K. Kalia, MRS Symposia Proceedings, Vol. **409** (MRS, Pittsburgh, 1996). (ISBN: 1-55899-311-8)
82. R. Perez, M. C. Payne, I. Stich and K. Terakura, "Role of Tip-Surface Covalent Interactions in Non-Contact Atomic Force Microscopy on Reactive Surfaces", *Phys. Rev. Lett.*, submitted
83. M. J. Gillan, L. N. Kantorovich and P. J. D. Lindan, "Modelling of oxide surfaces", *Current opinion in solid-state and materials science* **1** 820 (1996)
84. P. J. D. Lindan, N. M. Harrison and M. J. Gillan, "Paint by numbers: Oxide surface chemistry using first-principles molecular dynamics", *Psi-k Network newsletter* **17** (December 1996)
85. F. Kirchhoff, J. Holender and M. J. Gillan, "The structure, dynamics and electronic structure of liquid Ag-Se alloys investigated by ab initio simulation", *Phys. Rev. B*, **54**, 190 (1996)
86. L. N. Kantorovich, M. J. Gillan and J. A. White, "Adsorption of atomic oxygen on the MgO (100) surface", *J. Chem. Soc. Faraday Trans.*, **92**, 2075 (1996)
87. L. N. Kantorovich and M. J. Gillan, "Adsorption of atomic and molecular oxygen on the MgO and CaO (001) surfaces", *Proc. 13th Int. Conf. on Defects in Insulating Materials*, ed. R. T. Williams and G. E. Matthews, *Materials Science Forum*, **239-241**, 637 (1997)

88. C. M. Goringe, L. J. Clarke, M.-H. Lee, M. C. Payne, I. Štich, J. A. White, M. J. Gillan and A. P. Sutton, "GaAs (001)-(2 × 4) structure, chemistry and adsorbates", *J. Phys. Chem.*, submitted (invited feature article)
89. C. M. Goringe and A. P. Sutton, "Tight binding and local density approximation studies of semiconductor surfaces and small organic adsorbates", *Mat. Sci. Eng. B*, **37**, 150 (1996)
90. M. J. Gillan, "Quantum simulations of catalysis", *Physics World*, July 1996, p. 31
91. I. Dawson, P. D. Bristowe, M. C. Payne and M.-H. Lee, "An ab initio investigation of a grain boundary in a transition-metal oxide", *Mat. Res. Soc. Proc.*, **408**, 271 (1996)
92. I. Dawson, P. D. Bristowe, J. A. White and M. C. Payne, "First-principles simulation of the defect chemistry of rutile TiO₂" *Mat. Res. Soc. Proc.*, **409**, 453 (1997)
93. G. J. Ackland and M. C. Warren, "Soft-mode phase transitions from first principles", *Phase Transitions*, accepted.
94. B. B. Karki, S. J. Clark, M. C. Warren, G. J. Ackland and J. Crain, "Ab initio elasticity and lattice dynamics of AgGaSe₂", *J. Phys. Condens. Matter*, **9**, 375 (1997)
95. S.J.Clark, C.J.Adam, G.O.Bodammer, J.A.White, G.J.Ackland and J.Crain, "Structural and vibrational properties of large molecules from first principles: application to 5CB", *Phys. Rev. E* (in press)
96. C. M. Goringe, E. Hernández, C. M. Goringe and I. J. Bush, "Linear-scaling DFT-pseudopotential calculations on parallel computers", *Comput. Phys. Commun.*, submitted.
97. E. Hernández, C. M. Goringe and M. J. Gillan, "Basis functions for linear-scaling first-principles calculations", *Phys. Rev. B*, accepted
98. P. J. D. Lindan, N. M. Harrison and M. J. Gillan, "Mixed dissociative and molecular adsorption of water on the TiO₂ (110) surface", *Phys. Rev. Lett.* **80**, 762 (1998)
99. M. J. Gillan, P. J. D. Lindan, L. N. Kantorovich and S. P. Bates, "Molecular processes on oxide surfaces studied by first-principles calculations" *Mineral. Mag.* **62** (1998) 669