

Parallel implementation of the *ab initio* CRYSTAL program: Electronic structure calculations for periodic systems

I.J. Bush¹, S. Tomić², B.G. Searle², G. Mallia³, C.L. Bailey²,
B. Montanari⁴, L. Bernasconi⁴, J.M. Carr^{2,5} and N.M. Harrison^{2,3}

¹ *Numerical Algorithms Group Ltd.,*

Oxford OX2 8DR, United Kingdom

² *Computational Science and Engineering Department,*

STFC Daresbury Laboratory, Cheshire WA4 4AD, United Kingdom

³ *Department of Chemistry, Imperial College London,*

London SW7 2AZ, United Kingdom

⁴ *Computational Science and Engineering Department,*

STFC Rutherford Appleton Laboratory,

Oxfordshire OX11 0QX, United Kingdom

⁵ *The University Chemical Laboratory,*

Lensfield Road, Cambridge CB2 1EW, UK.

Abstract

CRYSTAL is an *ab initio* electronic structure program, based on the linear combination of atomic orbitals (LCAO), for periodic systems. This article concerns the ability of CRYSTAL to exploit massively parallel computer hardware. A brief review of the theory, numerical implementations, and parallel solutions will be given and some of the functionalities and capabilities highlighted. Some features which are unique to CRYSTAL will be described and development plans outlined.

I. INTRODUCTION

The development and application of new materials has played an important role in the technologies that impact on our daily lives. The global challenges of security, energy, climate change and health have further sharpened the focus on materials research and technology. The application of high performance computing to materials chemistry problems is currently contributing significantly to, for instance, the discovery and development of solar absorbers, new catalysts and hydrogen storage systems. The key relationship is that between composition, structure and desirable properties. In many cases it is an understanding of the electronic properties at an atomic and quantum mechanical level that is vital. As the need has grown the computer simulation of electronic structure has developed in accuracy, reliability and scale. There is now a rapidly growing number of cases where realistic models of important systems can be simulated and the value of experimental measurements significantly enhanced by direct comparison to *ab initio* theory. These developments can be attributed to advances in the underlying theory, algorithmic developments and the exploitation of high performance computers.

The CRYSTAL package [1, 2] was jointly developed by the Theoretical Chemistry Group at the University of Torino and the Computational Materials Science group in STFC's Daresbury and Rutherford Appleton Laboratories. The program performs *ab initio* calculations of the ground state energy, energy gradient, electronic wave function and properties of periodic systems. Hartree-Fock or Kohn-Sham Hamiltonians, which describe the effects of electronic exchange and correlation using a potential derived from density functional theory (DFT), can be used [3–5]. Systems periodic in 0 (molecules, 0D), 1 (polymers, 1D), 2 (slabs, 2D), and 3 dimensions (crystals, 3D) are treated on an equal footing. Symmetry is exploited and CRYSTAL automatically implements the 230 space groups, 80 layer groups, 99 rod groups and 45 point symmetry groups. In the case of polymers helical symmetries can also be applied and exploited.

This paper, which is aimed at providing information about developments of CRYSTAL that enable parallel computing of the energy and the wavefunction of a periodic system, is organised as follows: In Sec. II we review briefly the underlying theory; Hartree-Fock theory (II A), DFT (II B) and the numerical implementation of the theories (II C). In Sec. III we outline and compare parallelisation strategies adopted and implemented in the PCRYSTAL

and MPPCRYSTAL parallel versions of the program. In Sec. IV the parallel scalability of MPPCRYSTAL is demonstrated and a brief outline of recent and ongoing developments of CRYSTAL are discussed in Sec. V. Finally, conclusions are made in Sec. VI.

II. THEORETICAL BACKGROUND

A. Hartree-Fock Theory

The electronic Hamiltonian operator, $\hat{H}$, consists of a sum of three terms; the kinetic energy, the interaction with the external potential (V_{ext}) and the electron-electron interaction (V_{ee}). That is (in atomic unit);

$$\hat{H} = -\frac{1}{2} \sum_{i=1}^N \nabla_i^2 + \hat{V}_{ext} + \sum_{i=1}^N \sum_{j>i}^N \frac{1}{|\mathbf{r}_i - \mathbf{r}_j|} \quad (1)$$

In materials simulation the external potential of interest is generally the interaction of the electrons with the atomic nuclei;

$$\hat{V}_{ext} = - \sum_{i=1}^N \sum_{\alpha=1}^{N_{at}} \frac{Z_{\alpha}}{|\mathbf{r}_i - \mathbf{R}_{\alpha}|} \quad (2)$$

Here, $\mathbf{r}_i$ is the coordinate of electron i and the charge on the nucleus at $\mathbf{R}_{\alpha}$ is Z_{α} . Note that in order to simplify the notation and to focus the discussion on the main features of DFT the spin coordinate is omitted here and throughout this article.

In Hartree-Fock theory [6, 7] the ground state is found by minimising the total energy of the system with respect to a set of N normalised spin orbitals ψ_i . This leads to the Hartree-Fock (or SCF) equations;

$$\left[-\frac{1}{2} \nabla_1^2 - \sum_{\alpha=1}^{N_{at}} \frac{Z_{\alpha}}{|\mathbf{r}_1 - \mathbf{R}_{\alpha}|} + \int \frac{\rho(\mathbf{r}_2)}{|\mathbf{r}_1 - \mathbf{r}_2|} d\mathbf{r}_2 \right] \phi_i(\mathbf{r}_1) + \int \nu_x(\mathbf{r}_1, \mathbf{r}_2) \phi_i(\mathbf{r}_2) d\mathbf{r}_2 = \epsilon_i \phi_i(\mathbf{r}_1) \quad (3)$$

where $\nu_x(\mathbf{r}_1, \mathbf{r}_2)$ non-local exchange potential. Eq. 3 can be also written as $\hat{F}|\phi_i\rangle = \epsilon_i|\phi_i\rangle$, where $\hat{F}$ is the Fock operator.

B. Density Functional Theory

Hartree-Fock theory describes non-interacting electrons in a mean field potential consisting of a classical Coulomb potential and a non-local exchange potential. Electron correlation

is, however, neglected. DFT provides an improved approach for including the effect of electron correlation [5, 9];

$$\left[-\frac{1}{2}\nabla_1^2 - \sum_{\alpha=1}^{N_{at}} \frac{Z_{\alpha}}{|\mathbf{r}_1 - \mathbf{R}_{\alpha}|} + \int \frac{\rho(\mathbf{r}_2)}{|\mathbf{r}_1 - \mathbf{r}_2|} d\mathbf{r}_2 \right] \phi_i(\mathbf{r}_1) + \int \nu_{xc}(\mathbf{r}_1) \phi_i(\mathbf{r}_2) d\mathbf{r}_2 = \epsilon_i \phi_i(\mathbf{r}_1) \quad (4)$$

where $\nu_{xc}(\mathbf{r}) = \frac{\delta E_{xc}}{\delta \rho}$ is a local multiplicative potential which is the functional derivative of the exchange correlation energy with respect to the density.

The DFT energy can be written as the sum of the kinetic energy, the classical Coulomb interaction the electron-nucleus interaction and an exchange correlation term, E_{xc} . This later term, E_{xc} , is the sum of the errors in the approximations made in assuming a non-interacting kinetic energy term and in treating the electron-electron interaction classically.

Studies of the homogenous electron gas suggest that E_{xc} can be described in terms of the local electron density, several different functionals exist which exploit this property and they are known collectively as local density approximations (LDAs). The LDA can be improved upon by also considering the first derivatives of the density and functionals which use this are known as generalised gradient approximations (GGAs). The combination of non-local Fock exchange and density functionals was first proposed by Becke in the B3LYP hybrid functional which mixes 20% Fock exchange with a GGA exchange functional [10].

The energy functional and matrix elements of the exchange and correlation potentials are not analytic functions of the Gaussian basis set and are therefore integrated numerically on an atom centred grid of points. The integration over radial and angular coordinates is performed using Gauss-Legendre and Lebedev schemes, respectively.

In general bond enthalpies are significantly better described when using the GGA than when using either Hartree-Fock theory or the LDA. Hybrid exchange functionals partially correct for electronic self-interaction and therefore generally further improve the description of bond enthalpies and vibrational frequencies. The hybrid exchange approach also produces single particle band gaps for a variety of semiconductors, simple oxides and transition metal oxides more reliably than Hartree-Fock, LDA or GGA theories [11, 12].

C. Numerical Implementation in CRYSTAL

In CRYSTAL the crystalline orbitals, $\psi_i(\mathbf{r}; \mathbf{k})$, are expanded as a linear combination of Bloch functions, $\phi_{\mu}(\mathbf{r}; \mathbf{k})$, which are themselves expressed as linear combinations of atom

centred atomic orbitals, $\varphi_\mu(\mathbf{r})$.

$$\psi_i(\mathbf{r}; \mathbf{k}) = \sum_{\mu} a_{\mu,i}(\mathbf{k}) \phi_{\mu}(\mathbf{r}; \mathbf{k}) \quad (5)$$

$$\phi_{\mu}(\mathbf{r}; \mathbf{k}) = \sum_{\mathbf{g}} \varphi_{\mu}(\mathbf{r} - \mathbf{A}_{\mu} - \mathbf{g}) e^{i\mathbf{k} \cdot \mathbf{g}} \quad (6)$$

where $\mathbf{A}_{\mu}$ denotes the coordinate of the nucleus in the zero reference cell on which φ_{μ} is centred, and the $\sum_{\mathbf{g}}$ is extended to the set of all lattice vectors $\mathbf{g}$.

Each atomic orbital is expressed as a linear combination of individually normalised Gaussian type functions (GTFs), with fixed coefficients d_j and exponents, α_j :

$$\varphi_{\mu}(\mathbf{r} - \mathbf{A}_{\mu} - \mathbf{g}) = \sum_j^{n_G} d_j G(\alpha_j; \mathbf{r} - \mathbf{A}_{\mu} - \mathbf{g}) \quad (7)$$

The collection of all atomic orbitals being referred to as the basis set. The expansion coefficients of the Bloch functions, $a_{\mu,i}(\mathbf{k})$, are calculated by solving the matrix equation for each reciprocal lattice vector, $\mathbf{k}$:

$$\mathbf{F}(\mathbf{k})\mathbf{A}(\mathbf{k}) = \mathbf{S}(\mathbf{k})\mathbf{A}(\mathbf{k})\mathbf{E}(\mathbf{k}) \quad (8)$$

in which $\mathbf{S}(\mathbf{k})$ is the overlap matrix over the Bloch functions, $\mathbf{E}(\mathbf{k})$ is the diagonal energy matrix and $\mathbf{F}(\mathbf{k})$ is the Fock matrix in reciprocal space:

$$\mathbf{F}(\mathbf{k}) = \sum_{\mathbf{g}} \mathbf{F}^{\mathbf{g}} e^{i\mathbf{k} \cdot \mathbf{g}} \quad (9)$$

The matrix elements of $\mathbf{F}^{\mathbf{g}}$, the Fock matrix in direct space, can be written as a sum of one-electron and two-electron contributions in the basis set of the atomic orbitals:

$$F_{ij}^{\mathbf{g}} = H_{ij}^{\mathbf{g}} + B_{ij}^{\mathbf{g}} \quad (10)$$

The one-electron contribution is the sum of the kinetic and nuclear attraction terms:

$$H_{ij}^{\mathbf{g}} = T_{ij}^{\mathbf{g}} + Z_{ij}^{\mathbf{g}} = \langle \varphi_i^{\mathbf{0}} | \hat{T} | \varphi_j^{\mathbf{g}} \rangle + \langle \varphi_i^{\mathbf{0}} | \hat{Z} | \varphi_j^{\mathbf{g}} \rangle = \langle \varphi_i^{\mathbf{0}} | -\frac{1}{2} \nabla^2 | \varphi_j^{\mathbf{g}} \rangle + \langle \varphi_i^{\mathbf{0}} | \sum_{\alpha=1}^{N_{at}} \frac{-Z_{\alpha}}{|\mathbf{r}_i - \mathbf{R}_{\alpha}|} | \varphi_j^{\mathbf{g}} \rangle \quad (11)$$

In core pseudopotential calculations, $\hat{Z}$ includes the sum of the atomic pseudo-potentials.

The two-electron term is the sum of two contributions:

$$B_{ij}^{\mathbf{g}} = C_{ij}^{\mathbf{g}} + X_{ij}^{\mathbf{g}}$$

where $C_{ij}^{\mathbf{g}}$ is the Coulomb term given by

$$C_{ij}^{\mathbf{g}} = \sum_{k,l} \sum_{\mathbf{n}} P_{k,l}^{\mathbf{n}} \sum_{\mathbf{h}} (\varphi_i^{\mathbf{0}} \varphi_j^{\mathbf{g}} | \varphi_k^{\mathbf{h}} \varphi_l^{\mathbf{h}+\mathbf{n}}) \quad (12)$$

with

$$(\varphi_i^{\mathbf{0}} \varphi_j^{\mathbf{g}} | \varphi_k^{\mathbf{h}} \varphi_l^{\mathbf{h}+\mathbf{n}}) = \int \varphi_i^{\mathbf{0}}(\mathbf{r}_1) \varphi_j^{\mathbf{g}}(\mathbf{r}_1) \frac{1}{|\mathbf{r}_1 - \mathbf{r}_2|} \varphi_k^{\mathbf{h}}(\mathbf{r}_2) \varphi_l^{\mathbf{h}+\mathbf{n}}(\mathbf{r}_2) d\mathbf{r}_1 d\mathbf{r}_2 \quad (13)$$

The term $X_{ij}^{\mathbf{g}}$ is the exchange contribution in Hartree-Fock method, calculated as follows

$$X_{ij}^{\mathbf{g}} = -\frac{1}{2} \sum_{k,l} \sum_{\mathbf{n}} P_{k,l}^{\mathbf{n}} \sum_{\mathbf{h}} (\varphi_i^{\mathbf{0}} \varphi_k^{\mathbf{h}} | \varphi_j^{\mathbf{g}} \varphi_l^{\mathbf{h}+\mathbf{n}}) \quad (14)$$

while $X_{ij}^{\mathbf{g}}$ is the exchange and correlation contribution in DFT, obtained by integrating the exchange-correlation potential $\nu_{xc}(\mathbf{r})$, see Eq. 4.

The Coulomb interactions, that is, those of electron-nucleus, electron-electron and nucleus-nucleus, are individually divergent, due to the infinite size of the system. The grouping of corresponding terms is necessary in order to eliminate this divergence.

The $P^{\mathbf{n}}$ density matrix elements in the atomic orbitals basis set are computed by integration over the volume of the Brillouin zone,

$$P_{k,l}^{\mathbf{n}} = 2 \int_{\text{BZ}} d\mathbf{k} e^{i\mathbf{k} \cdot \mathbf{n}} \sum_m a_{k,m}^*(\mathbf{k}) a_{l,m}(\mathbf{k}) \theta[\epsilon_F - \epsilon_m(\mathbf{k})] \quad (15)$$

where $a_{i,m}$ denotes the i -th component of the m -th eigenvector, θ is the step function, ϵ_F , the Fermi energy and ϵ_n , the n -th eigenvalue. The total electronic energy per unit cell is given by:

$$E_{\text{tot}} = \frac{1}{2} \sum_{i,j} \sum_{\mathbf{g}} P_{ij}^{\mathbf{g}} (H_{ij}^{\mathbf{g}} + F_{ij}^{\mathbf{g}}) \quad (16)$$

III. PARALLEL IMPLEMENTATION IN CRYSTAL

The algorithm used to determine the ground state electron density and energy in CRYSTAL is similar to that used in other local orbital programs [13–16] and can be briefly summarised as follows.

Given an initial density matrix, $\mathbf{P}^{\mathbf{g}}$,

1. Calculate analytically the kinetic, Coulombic and, if necessary, the exact exchange contributions to the Fock matrix in the atomic orbital representation, $\mathbf{F}^{\mathbf{g}}$.

2. If required calculate by quadrature the DFT exchange and correlation contributions to $\mathbf{F}^g$.
3. Transform $\mathbf{F}^g$ into reciprocal space in its crystalline orbital representation $\mathbf{F}(\mathbf{k})$. This is done in two steps: first by a Fourier transform followed by a similarity transform.
4. At each $\mathbf{k}$ point diagonalise $\mathbf{F}(\mathbf{k})$.
5. Using the eigenvalues from step 4, calculate the Fermi level, and hence the occupation numbers of each orbital at each $\mathbf{k}$ point.
6. Sum over the occupied eigenvectors to construct a new density matrix, $\mathbf{P}^g$.
7. Repeat steps 1 to 6 until convergence of the total energy.

A parallel algorithm has been implemented for each of the steps (1-6) with the exception of step 5, which is not computationally demanding.

A replicated data approach is easily implemented. In step 1 the evaluation of each element of the Fock matrix is an independent task involving large numbers of analytic integrals. In step 2 the quadrature on a grid is a classic example of data parallelism as each cpu can evaluate the integral on different parts of the grid. Finally the transformation and diagonalisation of $\mathbf{F}(\mathbf{k})$ can be performed at each $\mathbf{k}$ point independently and after communication of the eigenvalues for the determination of the Fermi level the contribution of orbitals at each $\mathbf{k}$ point to $\mathbf{P}^g$ is also an independent calculation. A parallel version of CRYSTAL (PCRYSTAL) based on this approach was first released in 1996.

PCRYSTAL uses a replicated data paradigm; all cpus have access to a complete copy of all the objects required, but each cpu will be performing different calculations at any instant. The replication of data leads to a fairly straightforward parallelism. The terms evaluated analytically (step 1) may simply be farmed out across the cpus. There is a potential load balance problem when doing this as each term will take a different amount of time to compute. However in practice this is not usually a problem, as typically there are a very large number of terms to compute compared to the number of cpus, and a simple static load balancing procedure is sufficient to achieve good parallel scaling. This parallelisation method has also been implemented in several other quantum chemistry programs. [13, 15]

The evaluation of the quadrature (step 2) is a little more complex. In **CRYSTAL** the unit cell is divided into a number of parallelepipeds, and the quadrature within each of these parallelepipeds is evaluated independently and can be conducted in parallel. However unlike the analytic terms the number of parallelepipeds is often of a similar order of magnitude to the number of cpus. Hence load balance is a problem as each parallelepiped may contain a different number of grid points, and may contribute to different numbers of matrix elements. **PCRYSTAL** therefore dynamically load balances this part of the calculation by measuring the time taken to perform each of the quadratures and assigning the parallelepipeds to the cpus appropriately. For the first cycle it is reasonable to assume that the time required for a given parallelepiped is proportional to the number of grid points it contains.

For the calculations in reciprocal space, i.e. the Fourier and similarity transforms (step 3), the diagonalisation of $\mathbf{F}(\mathbf{k})$ (step 4) and the construction of $\mathbf{P}^g$ (step 6), **PCRYSTAL** exploits the independence of the $\mathbf{k}$ points. Each cpu is assigned a subset of $\mathbf{k}$ points, and performs the calculation on that subset constructing a partial $\mathbf{P}^g$. In unrestricted calculations the independence of the spin states is also exploited. The only synchronisation point is the evaluation of the Fermi level and the global summation of the partial $\mathbf{P}^g$.

The resulting code is simple and for many cases performs very well. However it does have a number of limitations:

1. The parallelism in reciprocal space is limited by the number of $\mathbf{k}$ points. As the system size increases the number of $\mathbf{k}$ points required decreases; very large systems can be accurately described using just one $\mathbf{k}$ point.
2. The maximum size of a calculation may be limited by the amount of memory available. As all the data is replicated the largest system that can be addressed by **PCRYSTAL** is no larger than that which serial **CRYSTAL** can address. In an era where the amount of memory per cpu is falling this is a particularly serious problem.
3. In principle the costs of re-replicating the data at each stage could become important as the cost of the procedure grows with the number of cpus. While this is important for other codes, in practice in **PCRYSTAL** each stage is sufficiently expensive that these costs are negligible.

Limitation 1 results in **PCRYSTAL** typically only scaling to a few 10s of cpus, with most

calculations becoming impractical due to runtime before limitation 2 is reached.

To address these problems a new massively parallel version of PCRYSTAL has been developed, MPPCRYSTAL. The main change from PCRYSTAL is that all large objects are distributed. In particular all the large reciprocal space matrices are distributed and operated on in parallel. For this the ScaLAPACK library is used, and thus a block cyclic distribution of the data is implemented. [17, 18]

Thus in MPPCRYSTAL for the reciprocal space part of the calculation a hierarchical parallelism is used; first the independence of the $\mathbf{k}$ points is exploited, and then for each $\mathbf{k}$ point a number of cpus perform the calculation in parallel by using the ScaLAPACK library. This addresses both the major limitations noted above, and MPPCRYSTAL can: (i) scale to 1,000s of cpus and (ii) address much larger problems than either CRYSTAL or PCRYSTAL. For instance it has been shown to be able to perform SCF iterations at over 40000 basis functions.

But what of the evaluation of $\mathbf{F}^{\mathbf{g}}$ in MPPCRYSTAL? With one major exception this is performed in almost the same manner as PCRYSTAL, i.e. $\mathbf{F}^{\mathbf{g}}$ and $\mathbf{P}^{\mathbf{g}}$ are replicated. This is not nearly as large an overhead as one might think as in CRYSTAL both these objects are stored in sparse format, and thus are very much smaller than the dense matrix representations used for the reciprocal space objects. However, ultimately, this is a potential problem, and we shall return to this later. The major exception is the grid used by DFT calculations which is distributed due to its memory requirement; as each cpu only performs quadratures across a small portion of the grid it only holds those parts of the grid which it needs.

The resulting code is, as stated above, much more scalable in terms of both time and memory than PCRYSTAL, and may exploit many more cpus to perform calculations on much larger systems. In fact benchmarks on a number of systems suggest that a very rough estimate of the number of cpus to which the code can scale reasonably efficiently is given by

$$(N_k \times N_s \times N_b)/30, \quad (17)$$

where N_k is the number of $\mathbf{k}$ points, N_s the number of spins and N_b the number of basis functions. Thus very roughly one might expect an MPPCRYSTAL calculation using 10 $\mathbf{k}$ points and 3000 basis functions to run efficiently on up to around 1000 cpus.

It is worth stressing the memory usage of the algorithm discussed above, as effective

and efficient memory usage is becoming much more important on modern high performance computers. In fact it is worth noting that the MPP code can optionally use algorithms which are less time efficient but more memory efficient than **PCRYSTAL** precisely because it is designed to handle large systems where memory considerations are paramount. If these low memory options (controlled by the new **LOWMEM** directive) are exploited, **MPPCRYSTAL** uses no replicated objects that scale with the square of the system size. The required total memory is reduced by: (i) storing only the symmetry irreducible form of the density matrix and computing elements of the reducible form as required, (ii) optimising the storage of tables which index the analytic Coulomb and exchange integrals and (iii) distributing the matrices associated with the symmetrised directions.

This optimisation has proved important for phase 2 of **HECToR** which consists of nodes with a total of 8 GB of memory and four cpus, that is just 2 GB memory per cpu; it was previously not possible to utilise all the cpus per node when running large systems on **HECToR**.

The current major limitations of **MPPCRYSTAL** are:

1. The time to solution of the ScaLAPACK diagonaliser does not scale well with cpu count. This has been somewhat helped by the introduction of the faster routines (**PDSYEVD** and **PZHEEVD**) in ScaLAPACK 1.7 (see Ref. [19, 20]), but it is still an issue, and this is the ultimate limitation in the scaling of the time to solution of **MPPCRYSTAL**.
2. As noted above $\mathbf{F}^g$ and $\mathbf{P}^g$ are still replicated. While these are much smaller than the equivalent reciprocal space objects, they are still a significant size, and it is these that limit the size of calculation that can be performed.

Work is in progress to address both of these issues.

IV. DEMONSTRATION OF THE SCALABILITY OF CRYSTAL

The massively parallel implementation of **CRYSTAL** allows systems with ~ 1000 atoms per unit cell to be studied routinely on parallel computers. The two key computational steps are the evaluation of the Fock matrix through the calculation of matrix element integrals and the solution of the Kohn-Sham equations through diagonalisation of the Fock matrix. The

parallel capability of **CRYSTAL** will be demonstrated with two examples of calculations on the UK’s national supercomputer, **HECToR**. It has to be noted that, as regards the analytic gradient calculation, the dominant component is also the evaluation of a set of bi-electronic integrals and integrals over the derived DFT potential in a manner highly analogous to the integrals for the potential and energy expression; the scaling is therefore very similar and is not reported.

The first system is an 864 atom supercell of zincblende GaSb with 6 Sb atoms randomly substituted by N atoms producing a $\text{Ga}_{432}\text{Sb}_{426}\text{N}_6$ cell with no spatial symmetry. Ga, Sb, and N are described using triple valence basis sets yielding 19,836 atomic orbitals per cell - this is the rank of the Fock matrix. The Fock matrix is diagonalised at two symmetry independent $\mathbf{k}$ points.

In Figure 1 the scalability of one SCF cycle is presented and separated into its two major components: Calculation of the two-electron integrals (**SHELLX**) and diagonalisation of the Fock matrix (**MPP_DIAG**). The overall speed up of an SCF cycle is 3.3 when comparing runs on 896 cpus and 3,584 cpus (the ideal speed-up is 4). The parallel scaling of the integral calculation is near ideal, 3.8^2 , while the diagonalisation step scales reasonably well up to 1,792 cpus but not to 3,584 cpus; this is as expected from the rough estimation of scalability given by Eq. 17. Although the diagonalisation is responsible for a small fraction of the total runtime it is clear that the scaling of the diagonalisation ultimately limits the parallel scaling of the calculation.

The diagonalisation is performed within the ScaLAPACK library using a divide and conquer algorithm (using the **DCDIAG** directive in **MPPCRYSTAL**) and is sensitive to the block size used to distribute the matrix over the cpus. This is illustrated in Figure 2; reducing the block size (using the **MPPBLOCK** directive) from the default value of 96 to 64 or 32 results in a speed-up of around 10% in the diagonalisation and about 20% speed up in the total time for similarity transform followed by diagonalisation and back transform. The optimal value of the block size is machine dependent.

The second example is a TiO_2 ($3 \times 3 \times 3$) supercell, consisting of 648 atoms with 13,608 atomic orbitals. There are 16 symmetry operators which are exploited in the calculation of the two-electron integrals. In Figure 3 the total runtime for a single SCF cycle and the contributions from **SHELLX** and **MPP_DIAG** are displayed for cases where a single $\mathbf{k}$ point and 8 $\mathbf{k}$ points are used. Comparing this system to the low symmetry $\text{Ga}_{432}\text{Sb}_{426}\text{N}_6$ system the inte-

gral calculation in **SHELLX** is almost two orders of magnitude faster while the diagonalisation is only marginally faster. The parallel scaling of the integral calculation is again near ideal while the diagonalisation at a single **k** point scales to only 256 cpus. Exploiting parallelism over 8 **k** points significantly increases the number of cpus that can be used effectively in **MPP_DIAG** to around 4,096. This observed scalability is similar to that estimated by Eq. 17. It is slightly lower than expected for the case where there is only one (real) **k** point. This is, in part, because the estimate given by Eq. 17 assumes a mixture of both real and complex **k** points. Complex **k** points require around three times more computation time than real **k** points; this characteristic is exploited in **MPPCRYSTAL**. Consequently, systems that have complex **k** points generally scale to larger numbers of cpus than those that only have real **k** points.

These two examples have been chosen as they are typical of the work undertaken on HECToR by members of the Materials Chemistry Consortium and confirm the expected parallel scalability of **MPPCRYSTAL** estimated in Eq. 17.

The third example is a TiO_2 ($5 \times 4 \times 4$) supercell with a single Fe dopant, consisting of 1920 atoms with 40320 atomic orbitals. The calculations are performed on HECToR phase 2b machine at the unrestricted Hartree-Fock level of theory, with one symmetry operator. In Figure 4 the total runtime for a single SCF cycle and the contributions from **SHELLX** and **MPP_DIAG** are displayed for cases where a single **k** point is used. Comparing this system to the second example with one **k** point, the integral calculation in **SHELLX** maintains perfect scalability over the whole range of cpus considered, while now scaling of the Fock matrix diagonalisation step **MPP_DIAG**, is dramatically improved. Although in this example we have considered only the real **k** point improved scalability in this part is basically due to the much larger rank of the Fock matrix (40320 vs. 13608), which increases the work load per cpu relative to the communication costs.

V. DEVELOPMENTS OF CRYSTAL RELEVANT TO PARALLEL SCALING

Currently, the major ongoing development of **CRYSTAL** is to include density functional perturbation theory for response properties and time dependent density functional theory for calculating excited states. Details of this development and its parallel implementation are given in Sec. V A.

In addition several other relevant developments have recently been made or are currently in progress. An interface to WaNT for computing electronic transport in nanostructures will soon be available and will exploit the ability of **MPPCRYSTAL** to describe systems of 1000+ atoms [21]. The nudged elastic band (NEB) method for transition state searches has recently been implemented [22]. This method exploits the task-farming parallelism in **MPPCRYSTAL** whereby multiple energy evaluations are performed simultaneously and independently. This division of labour is conceptually simple and efficient, requiring minimal communication. Furthermore, an individual energy evaluation can be parallelised as described in Sec. III.

For large scale parallel geometry optimisation an interface between the **DL_FIND** library [23] and **CRYSTAL** has recently been implemented. **DL_FIND** is open-source code written primarily at STFC Daresbury Laboratory and is available from CCPForge [24]. A wide range of algorithms for local and global minimisation, as well as transition state searching, are available via the library interface. The non-sequential optimisation methods, i.e. those involving multiple configurations at each iteration, such as the stochastic searches, exploit the task-farming parallelism in **CRYSTAL**. Consequently, this hierarchical approach will enable efficient calculations on 10,000s of cpus.

A. Dielectric Properties and Excited States

The calculation of the many-body dynamical polarisability and hyperpolarisability tensors is available in the latest release of **CRYSTAL** (**CRYSTAL09**). The calculation is based on the self-consistent solution of a set of coupled-perturbed equations, derived within the linear or quadratic response approximations of perturbation theory. This method is well known in molecular physics, [25] and can be implemented at the Hartree-Fock, DFT or hybrid-DFT levels of theory. Its extension to periodic crystalline systems in **CRYSTAL09** relies on an analytical representation of the $\mathbf{k}$ -dependent position operator for extended systems, which allows it to achieve consistent levels of accuracy across wide ranges of material structures and compositions. [26]

Solution of the coupled-perturbed equations at a frequency $\pm\omega$ yields a set of unitary matrices $U_{\pm\omega}^{\mathbf{k},a}$, where the superscript a indicates a Cartesian direction, which transform the unperturbed one-particle orbitals into their linear-response orbitals. Equivalently, a linear density-matrix response $D_{\pm\omega}^{\mathbf{k},a}$ can be defined in terms of the matrices $U_{\pm\omega}^{\mathbf{k},a}$. The many-body

dynamical polarisability tensor $\boldsymbol{\alpha}(\pm\omega)$ is then given by

$$\alpha_{ab}(\pm\omega_I) = -2 \sum_{\mathbf{k}}^{N_{\mathbf{k}}} w_{\mathbf{k}} \sum_i^N \sum_j^{N_{\text{occ}}} [U_{\pm\omega_I}^{\mathbf{k},a}]_{ji} \Omega_{ij}^{\mathbf{k},b} + \Omega_{ji}^{\mathbf{k},b} [U_{\pm\omega_I}^{\mathbf{k},a}]_{ij}, \quad (18)$$

where the subscript a and b indicate Cartesian components, and i and j are band indices. The polarisability is related to the dielectric tensor $\epsilon(\pm\omega)$ by

$$\epsilon_{ab}(\pm\omega) = \delta_{ab} + \frac{4\pi}{V} \alpha_{ab}(\pm\omega), \quad (19)$$

where V is the cell volume. Optically-allowed many-body electronic excitation energies can be computed from the poles of the mean dynamical polarisability

$$\bar{\alpha}(\pm\omega) = \frac{1}{3} \text{tr } \boldsymbol{\alpha}(\pm\omega) \quad (20)$$

by examining the behaviour of this quantity within a given range of frequencies. In the hybrid-DFT approximation (B3LYP), the quality of this method is sufficient to yield optical gaps within 0.1 eV of experimental estimates for several semiconductors and oxides, [27] including those exhibiting (bound) exciton transitions. In the **CRYSTAL09** implementation, this method scales linearly with the number of $\mathbf{k}$ points included in the sampling of the Brillouin zone. This approach is typically appropriate for studying the lowest-energy excitations of an extended quantum system.

The coupled-perturbed method can be extended to provide a general formalism for computing electronic excitation energies and transition probabilities which does not involve the self-consistent solution of a set of coupled equations. This is the so-called *RPA-matrix formalism*, [28] in which excitation energies correspond to the eigenvalues of a (in general non-Hermitian) coupling matrix. Transition probabilities are related to the eigenvectors of the coupling matrix. The coupling matrix is expressed in terms of super-operators $\mathcal{A}$ and $\mathcal{B}$, which couple single-particle excitations *via* a local (DFT) or non-local (Hartree-Fock and hybrid-DFT) two-electron response term. [28] In **CRYSTAL09**, $\mathcal{A}$ and $\mathcal{B}$ are represented in crystal-orbital basis, to yield the eigenvalue equation:

$$\begin{pmatrix} \mathbf{A} & \mathbf{B} \\ \mathbf{B} & \mathbf{A} \end{pmatrix} \begin{pmatrix} \mathbf{X} \\ \mathbf{Y} \end{pmatrix} = \omega \begin{pmatrix} \mathbf{1} & \mathbf{0} \\ \mathbf{0} & -\mathbf{1} \end{pmatrix} \begin{pmatrix} \mathbf{X} \\ \mathbf{Y} \end{pmatrix}, \quad (21)$$

where $\mathbf{A}$ and $\mathbf{B}$ are $(N_{\text{occ}} \times N_{\text{vir}})^2$ matrices ($N_{\text{occ}}/N_{\text{vir}}$ being the number of occupied/virtual one-particles states), and $\mathbf{X}/\mathbf{Y}$ are left/right eigenvectors. Depending on the form of the

matrices $\mathbf{A}$ and $\mathbf{B}$, Eq. 21 corresponds to either time-dependent DFT (TD-DFT), time-dependent Hartree-Fock, or hybrid TD-DFT. $\mathbf{A}$ and $\mathbf{B}$ can also be simplified to yield approximate forms of these theories, the Tamm-Dancoff approximation [29] to TD-DFT and the configuration interaction singles method. Eq. 21 can be solved either by direct diagonalisation, or *via* an iterative Davidson-like method. Calculation of matrix-vector products (exploiting possible sparsities in the matrices and/or Davidson guess vectors, which may arise in the crystal-orbital representation of $\mathcal{A}$ and $\mathcal{B}$) is carried out by means of a non-self consistent coupled-perturbed calculation, similar to the molecular approach of Bauernschmitt and Ahlrichs. [30]

These perturbation methods can explicitly exploit task farming in frequency scans and will also exploit parallelism for matrix operations and linear algebra using similar techniques to those implemented in the current version of MPPCRYSTAL and described in section III.

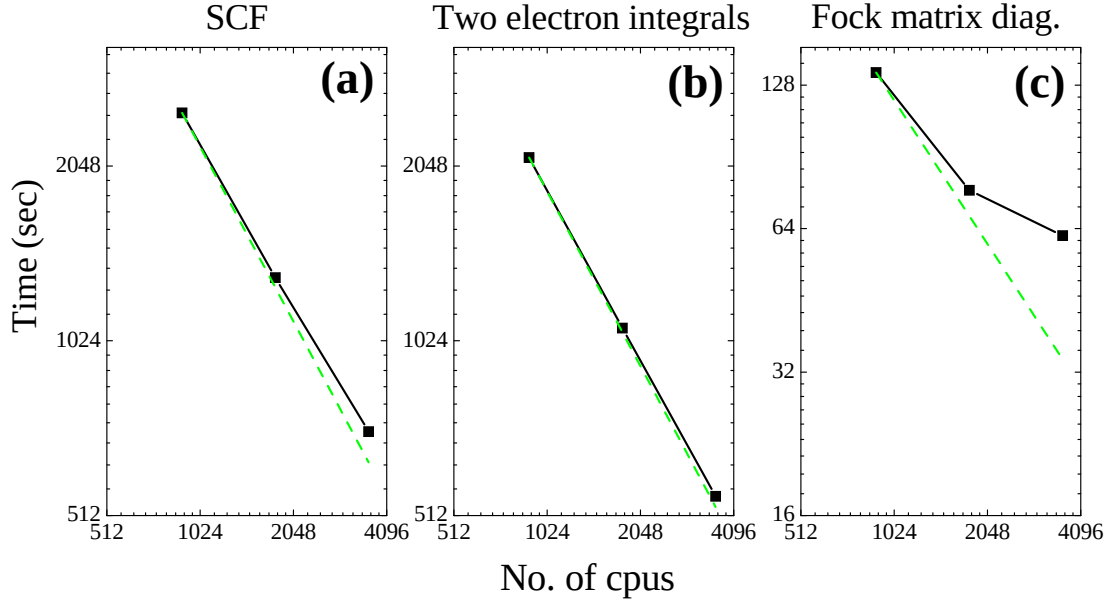
VI. CONCLUSIONS

An overview of the CRYSTAL code has been given, with particular emphasis on the adaptation of key algorithms for the exploitation of parallel computer hardware. It has been demonstrated that CRYSTAL can scale up to thousands of cpus for systems consisting of several hundred atoms and 10-20 thousand atomic orbitals. Recent developments have reduced the total memory required to run large CRYSTAL calculations. This has enabled CRYSTAL to run efficiently on the UK's national supercomputer, HECToR. Future developments include time dependent DFT for calculating excited states and an interface to WaNT [21] for computing electronic transport in nanostructures.

-
- [1] Dovesi, R., Saunders, V. R., Roetti, C., Orlando, R., Zicovich-Wilson, C. M., Pascale, F., Civalleri, B., Doll, K., Harrison, N. M., Bush, I. J., D'Arco, Ph. & Llunell, M., *CRYSTAL09 User's Manual*, (University of Torino, Torino, 2009)
 - [2] URL: <http://www.crystal.unito.it>.
 - [3] Kohn, W. & Sham, L. J. 1956 Self-Consistent Equations Including Exchange and Correlation Effects *Phys. Rev.* **140**, A1133 (10.1103/PhysRev.140.A1133)

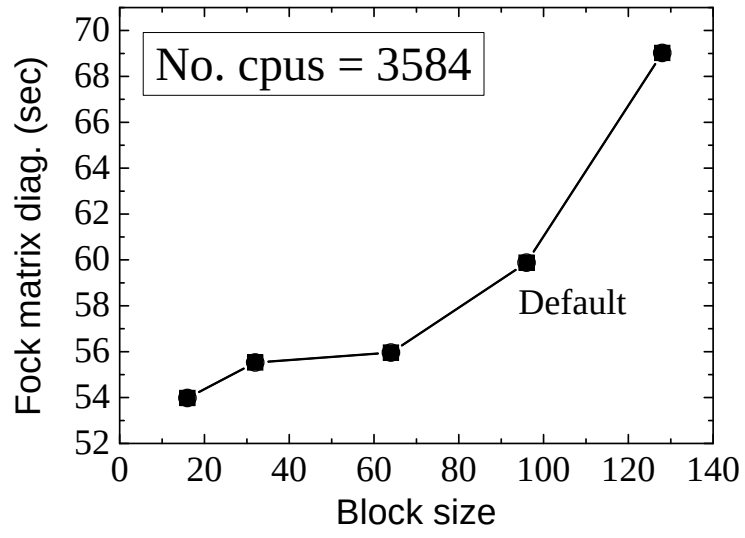
- [4] Hohenberg, P. & Kohn, W. 1964 Inhomogeneous Electron Gas *Phys. Rev.***136**, B864 (10.1103/PhysRev.136.B864)
- [5] Parr, R. G. & Yang, W. *Density-Functional Theory of Atoms and Molecules*, (Oxford University Press, Oxford, 1989).
- [6] Szabo, A. & Ostlund, N. S. *Modern Quantum Chemistry*, (Macmillan, New York, 1982).
- [7] Harrison, N. M. 2003 An introduction to density functional theory *Computational Materials Science NATO*, Volume: 187, 45-70, C. R. A. Catlow and E.A. Kotomin (Eds.), IOS Press
- [8] Stern, F. 1972 Self-Consistent Results for *n*-Type Si Inversion Layers *Phys. Rev. B***5**, 4891 (10.1103/PhysRevB.5.4891)
- [9] Martin, R. M. *Electronic Structure: Basic Theory and Practical Methods* (Cambridge University Press, UK, 2004)
- [10] Becke, A. D. 1993 A new mixing of Hartree-Fock and local density-functional theories *J. Chem. Phys.***98**, 1372 (10.1063/1.464304); Becke, A. D. 1993 Density-functional thermochemistry. III. The role of exact exchange *Ibid.* **98**, 5648 (10.1063/1.464913); Becke, A. D. 1988 Correlation energy of an inhomogeneous electron gas: A coordinate-space model *Ibid* **88**, 1053 (10.1063/1.454274)
- [11] Muscat, J., Wander, A. & Harrison, N. H. 2001 On the prediction of band gaps from hybrid functional theory *Chem. Phys. Lett.***342**, 397 (10.1016/S0009-2614(01)00616-9)
- [12] Tomić, S., Montanari, B. & Harrison, N. M. 2008 The group IIIV's semiconductor energy gaps predicted using the B3LYP hybrid functional *Physica E***40**, 2125 (10.1016/j.physe.2007.10.022)
- [13] Guest, M. F., Bush, I. J., Van Dam, H. J. J., Sherwood, P., Thomas, J. M. H., Van Lenthe, J. H., Havenith, R. W. A. & Kendrick, J. 2005 The GAMESS-UK structure package: algorithms, developments and applications *Molecular Physics***103** 719
- [14] Soler, J. M., Artacho, E., Gale, J. D., Garcia, A., Junquera, J., Ordejon, P. & Sanchez-Portal, D. 2002 The SIESTA method for ab initio order-N materials simulation *J. Phys. Condensed Matt.* **14**, 2745 (10.1088/0953-8984/14/11/302)
- [15] URL: <http://www.gaussian.com>.
- [16] URL: <http://www.molpro.net/>
- [17] URL: <http://www.netlib.org/scalapack>
- [18] URL: <http://www.mpi-forum.org>
- [19] URL: <http://www.cse.scitech.ac.uk/arc/diags.shtml>

- [20] URL:http://www.hector.ac.uk/cse/distributedcse/reports/castep/castep_performance_xt/4_Distributed_Diagonaliser_.html
- [21] Ferretti, A., Bonferroni, B., Calzolari A. & Buongiorno Nardelli, M. URL: <http://www.wannier-transport.org>.
- [22] Bailey, C. L., Wander, A., Searle B. G. & Harrison, N. H. 2005 Implementation of Nudged Elastic Band in CRYSTAL *DL Technical Report*, DL-TR-2005-003
- [23] Kastner, J., Carr, J. M., Keal, T.W., Thiel, W., Wander, A. & Sherwood, P. 2009 DL-FIND: An Open-Source Geometry Optimizer for Atomistic Simulations *J. Phys. Chem. A* **113** 11856
- [24] URL: <http://ccpforge.cse.rl.ac.uk>.
- [25] Hurst, G. J. B., Dupuis, M. & Clementi E. 1988 Ab initio analytic polarizability, first and second hyperpolarizabilities of large conjugated organic molecules: Applications to polyenes C₄H₆ to C₂₂H₂₄ *J. Chem. Phys.***89**, 385 (10.1063/1.455480)
- [26] Ferrero, M., Rérat, M., Orlando, R., & Dovesi R. 2008 Coupled perturbed Hartree-Fock for periodic systems: The role of symmetry and related computational aspects *J. Chem. Phys.***128**, 014110 (10.1063/1.2817596)
- [27] Bernasconi, L., Tomić, S., Ferrero, M., Rérat, M., Orlando, R., Dovesi, R. & Harrison, N. H. (submitted).
- [28] Hirata, S., Head-Gordon, M. & Bartlett, R. J. 1999 Configuration interaction singles, time-dependent HartreeFock, and time-dependent density functional theory for the electronic excited states of extended systems *J. Chem. Phys.***111**, 10774 (10.1063/1.480443)
- [29] Hirata, S., & Head-Gordon, M. 1999 Time-dependent density functional theory within the TammDancoff approximation *Chem. Phys. Lett.***314**, 291 (10.1016/S0009-2614(99)01149-5)
- [30] Bauernschmitt, R. and Ahlrichs, R. 1996 Treatment of electronic excitations within the adiabatic approximation of time dependent density functional theory *Chem. Phys. Lett.***256**, 454 (10.1016/0009-2614(96)00440-X)



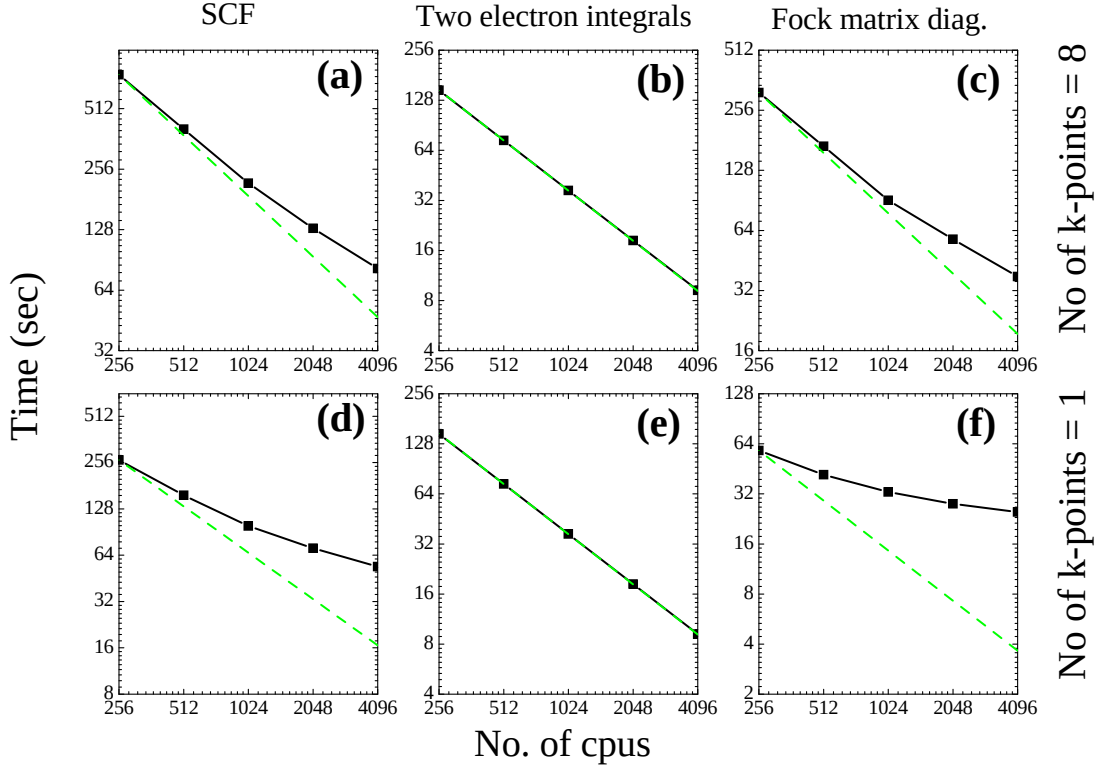
I.J. Bush et al, Fig 1 of 3

FIG. 1: Scalability of the major components in the total energy minimisation procedure using MPPCRYSTAL code on the $\text{Ga}_{432}\text{Sb}_{426}\text{N}_6$ system consisting of 19,836 atomic orbitals. The total time is given for the calculation of (a) one SCF cycle, (b) the two electron integrals (SHELLX), and (c) the diagonalisation of the Fock matrix (MPP_DIAG). The green dashed line is the ideal scaling.



I.J. Bush et al., Fig 2 of 3

FIG. 2: Dependence of the Fock matrix diagonalisation step on the block size (MPPBLOCK) of the matrix parallel distribution. The default value of MPPBLOCK is 96.



I.J. Bush et al, Fig 3 of 3

FIG. 3: Scalability of the major component in the total energy minimisation procedure using MPPCRYSTAL code on the TiO_2 ($3 \times 3 \times 3$) supercell consisting of 13,608 atomic orbitals. The total time is given for the calculation of (a) one SCF cycle, (b) the two electron integrals (SHELLX), and (c) the diagonalisation of the Fock matrix (MPP_DIAG). The green dashed line is the ideal scaling.

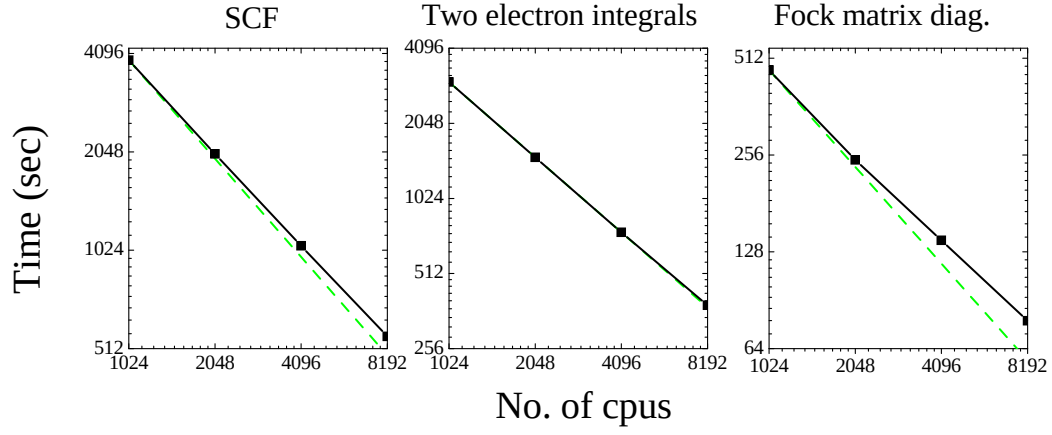


FIG. 4: Scalability of the major component in the total energy minimisation procedure using MPPCRYSTAL code on the TiO_2 ($5 \times 4 \times 4$) supercell consisting of 40,320 atomic orbitals. The total time is given for the calculation of (a) one SCF cycle, (b) the two electron integrals (SHELLX), and (c) the diagonalisation of the Fock matrix (MPP_DIAG). The green dashed line is the ideal scaling.