

## Computational Bottlenecks in Molecular Orbital Calculations

H. Bernhard Schlegel  
*Dept. of Chemistry*  
*Wayne State University*  
*Detroit, MI 48202*

Michael J. Frisch  
*Lorentzian, Inc.*  
*127 Washington Ave.*  
*North Haven, CT 06473*

**ABSTRACT** This Chapter examines some of the major steps in molecular orbital calculations, such as energies and gradients at SCF, MCSCF, MPn, CI, CC and QCI levels and SCF second derivatives. The emphasis is on the computational requirements rather than the details of the theory. Specifically cpu times, memory size and disk usage are considered for integral evaluation and transformation, solution of the SCF, CPHF and configuration interaction problems, and calculation of energy derivatives. Conventional, direct and semi-direct algorithms are compared and some guidelines for choosing between them are given.

### 1. Introduction

Over the past decade, there have been rapid developments in ab initio molecular orbital calculations that have allowed these methods to be applied to larger molecules. With currently available computers, SCF calculations can be carried out for molecules with over 500 basis functions and MP2 calculations for molecules with up to 400 basis functions. These advances come primarily from the use of direct methods, as well as the increase in speed of affordable computers. Direct methods, in turn, have benefitted from improved algorithms for calculating two electron integrals. There have also been advances in both the speed and accuracy of geometry optimizations because of the availability energy derivatives, especially for levels of theory that include electron correlation. In this Chapter, we will look at the computation of energies and energy derivatives for SCF, MCSCF, MPn, CI, CC and QCI calculations. The emphasis is on the computational requirements of the various steps in MO calculations and how these costs increase with the size of the molecule. The appropriate equations for the various levels of theory are outlined briefly to show where the computational bottlenecks arise. Details of the theory, however, are left to the text books<sup>1-6</sup> and the original literature<sup>7-44</sup>.

## 2. Background

In the non-relativistic approximation for the time independent case, the wavefunction for an  $n$  electron system satisfies the Schrödinger equation:

$$(1) \quad \hat{H}\Psi = E \Psi$$

where the Hamiltonian is given by:

$$(2) \quad \hat{H} = \sum_{i=1}^n \left[ -\frac{1}{2} \nabla_i^2 - \sum_A^{\text{nuclei}} \frac{Z_A}{r_{iA}} \right] + \sum_{i<j}^n \frac{1}{r_{ij}} + \sum_{A<B}^{\text{nuclei}} \frac{Z_A Z_B}{r_{AB}}$$

where  $Z_A$  is the charge on nucleus  $A$ . Except for very simple systems, the Schrödinger equation cannot be solved in closed form; hence various approximations must be made. For an approximate wavefunction, the variational principle gives an upper bound to the energy:

$$(3) \quad E_{\text{var}} = \langle \Psi | \hat{H} | \Psi \rangle / \langle \Psi | \Psi \rangle > E_{\text{exact}}$$

Adjustable parameters in the approximate wavefunction  $\Psi$  can be varied to minimize  $E_{\text{var}}$ , yielding an improved energy and wavefunction. The Born-Oppenheimer approximation<sup>7</sup> (clamped nuclei approximation) is used to separate the electronic and the nuclear motions. Further restrictions on the form of the electronic wavefunction and the manner in which the adjustable parameters are determined are governed by level of theory used.

In the Hartree-Fock (HF) model<sup>8</sup>, the  $n$  electron wavefunction is written as a single Slater determinant<sup>9</sup> of  $n$  orthonormal spin orbitals,  $\phi_i$ :

$$(4) \quad \Psi_0 = \frac{1}{\sqrt{n!}} | \phi_1 \phi_2 \cdots \phi_n |$$

As discussed below, in the Hartree-Fock model each electron sees only the average field of the other electrons. In reality, the electrons must explicitly avoid each other because of their mutual coulombic repulsion; hence their motions are correlated. The difference between the Hartree-Fock energy and the exact, non-relativistic energy is termed electron correlation energy. The Hartree-Fock wavefunction can be improved by taking a linear combination of Slater determinants, yielding a configuration interaction (CI) wavefunction:

$$(5) \quad \Psi = a_0 \Psi_0 + \sum_{ia} a_i^a \Psi_i^a + \sum_{i<j, a<b} a_{ij}^{ab} \Psi_{ij}^{ab} + \cdots$$

where  $\Psi_i^a$ ,  $\Psi_{ij}^{ab}$ , etc. are Slater determinants in which occupied spin orbital  $\phi_i$ ,  $\phi_j$  etc. in the reference determinant  $\Psi_0$  are replaced by unoccupied or virtual spin orbitals  $\phi_a$ ,  $\phi_b$  etc. (i.e. determinants that are singly excited, doubly excited, etc.). The amplitudes,  $a_i^a$ ,  $a_{ij}^{ab}$  etc. can be determined variationally, using eq (3) or by perturbation theory, as will be outlined

in subsequent sections. If all possible excited configurations are included (full CI), this approach converges (slowly) to the exact wavefunction.

The spin orbitals are constructed from a linear combination of atomic orbitals (LCAO):

$$(6) \quad \phi_p = \sum_{\mu} C_{\mu p} \chi_{\mu} \alpha \quad \text{or} \quad \phi_p = \sum_{\mu} C_{\mu p} \chi_{\mu} \beta$$

where  $\alpha$  and  $\beta$  are spin functions and  $\chi_{\mu}$  are basis functions for the spatial part of the spin orbitals. For simplicity, the formulae presented here will be in terms of spin orbitals; in practice integration over spin reduces the formulae to expressions for each different spin case and involving only spatial orbitals. The basis functions  $\chi_{\mu}$  are themselves linear combinations of gaussian functions<sup>10</sup> (Slater functions<sup>11</sup> can also be used but the necessary integrals are much more difficult to calculate). Because the computational times depend strongly on the number of basis functions (see Table 1), economic reasons usually place severe restrictions on the number and type of basis functions that can be used.

For any of the levels of theory considered in this Chapter, the total electronic energy corresponding to the Hamiltonian given in eq 2 and a wavefunction such as eq 4 or 5, can be written in terms of integrals over molecular orbitals or directly in terms of integrals over basis functions

$$(7) \quad E = \sum_{pq} P_{pq} h_{pq} + \sum_{pqrs} \Gamma_{pqrs} (pq|rs) + V_{\text{nuc}} = \sum_{\mu\nu} P_{\mu\nu} h_{\mu\nu} + \sum_{\mu\nu\lambda\sigma} \Gamma_{\mu\nu\lambda\sigma} (\mu\nu|\lambda\sigma) + V_{\text{nuc}}$$

where  $P$  and  $\Gamma$  are the one and two particle density matrices and  $V_{\text{nuc}}$  is the nuclear repulsion energy. The one and two electron integrals over basis functions are

$$(8) \quad h_{\mu\nu} = \int \chi_{\mu}^* \left[ -\frac{1}{2} \nabla^2 - \sum_A^{\text{nuclei}} \frac{Z_A}{r_A} \right] \chi_{\nu} d\tau; \quad S_{\mu\nu} = \int \chi_{\mu}^* \chi_{\nu} d\tau$$

$$(9) \quad (\mu\nu|\lambda\sigma) = \int \int \chi_{\mu}^*(1) \chi_{\nu}(1) \frac{1}{r_{12}} \chi_{\lambda}^*(2) \chi_{\sigma}(2) d\tau_1 d\tau_2$$

Since the molecular orbitals are expressed as linear combinations of basis functions, the integrals over molecular orbitals can be obtained by a linear transformation:

$$(10) \quad h_{pq} = \int \phi_p^* \left[ -\frac{1}{2} \nabla^2 - \sum_A^{\text{nuclei}} \frac{Z_A}{r_A} \right] \phi_q d\tau = \sum_{\mu\nu} C_{\mu p}^* C_{\nu q} h_{\mu\nu}$$

$$(11) \quad (pq|rs) = \int \int \phi_p^*(1) \phi_q(1) \frac{1}{r_{12}} \phi_r^*(2) \phi_s(2) d\tau_1 d\tau_2 = \sum_{\mu\nu\lambda\sigma} C_{\mu p}^* C_{\nu q} C_{\lambda r}^* C_{\sigma s} (\mu\nu|\lambda\sigma)$$

A number of standard conventions are used in this Chapter to simplify the notation. Greek subscripts refer to basis functions, e.g.  $\chi_\mu$ . Indices  $i, j, k \dots$  run over occupied molecular orbitals;  $a, b, c \dots$  run over unoccupied or virtual orbitals;  $p, q, r \dots$  run over both. Unless otherwise stated the sums run over these implied ranges. The computational resources required for the various calculations depend strongly on the number of functions used in the expansion of the wavefunctions. The number of basis functions is  $N$ ; the number of occupied orbitals is  $O$ , and the number of unoccupied or virtual orbitals is  $V$  (since some orbitals may be frozen,  $O + V \leq N$ ). The number of first derivatives with respect to the atom positions is  $N_a$  (equal to 3 times the number of atoms). The notation  $O(n)$  indicates a number of the order of  $n$ , e.g. there are formally  $O(N^4)$  two electron integrals. Table 1 outlines the formal and the actual size dependence of various electronic structure calculations on  $N, O, V$  and  $N_a$ . These will be discussed in greater detail as each level of theory is considered in turn.

### 3. Integrals

All molecular orbital calculations require the evaluation of integrals over the basis functions (for leading references on methods for evaluating these integrals, see ref. 12). The one electron integrals are not time-consuming since there are only  $O(N^2)$  of them. On the other hand, the two electron integrals are quite expensive to compute since there are  $O(N^4)$  of them. Permutational symmetry reduces the number of integrals by a factor of 8 ( $\mu \leftrightarrow \nu, \lambda \leftrightarrow \sigma$  and  $\mu\nu \leftrightarrow \lambda\sigma$  in  $(\mu\nu|\lambda\sigma)$ ); if available, spatial symmetry can reduce the number of integrals further. One of the major choices to be made in molecular orbitals calculations is how to calculate and store the numerous two electron integrals. There are four possible approaches:

**AQ** The two-electron integrals over the atomic orbitals are generated once and stored externally (on disk). This is the approach used by traditional SCF calculations. Formally this requires  $N^4/8$  cpu and disk; in practice the work and disk space grows as  $O(N^{3.5})$  because integrals that are near zero can be eliminated before calculation.

**MO** The atomic orbital integrals are generated once and stored externally, then transformed to the molecular orbital basis. The transformed (MO) integrals are also stored externally. The four index transformation, eq 11, requires  $N^5$  work and  $N^4$  disk storage (see the Integral Transformation section). This is the approach used until recently for all correlated energy methods.

**In-Core** The atomic orbital integrals are generated once and stored in canonical order in main memory (i.e., including zeroes). This requires *large* amounts of memory -  $N^4/8$

Table 1: Size-Dependence of Electronic Structure Methods

Method	Formal			Actual	
	CPU	Memory	Disk	CPU	Disk
Conventional SCF	$N^4$	$N^2$	$N^4$	$N^{3.5}$	$N^{3.5}$
In-core SCF	$N^4$	$N^4$	----	$N^4$	$N^2$
Direct SCF	$N^4$	$N^2$	----	$N^{2.7}$	$N^2$
Conventional MP2	$ON^4$	$N^2$	$N^4$	$ON^4$	$N^4$
Direct MP2 Energy	$ON^4$	$OVN$	----	$O^2N^3$	$N^2$
Semidirect MP2 Energy	$ON^4$	$N^2$	$VN^2$	$O^2N^3$	$VN^2$
Conventional MP2 Grad	$ON^4$	$N^2$	$N^4$	$ON^4$	$N^4$
Direct MP2 Grad	$ON^4$	$N^3$	----	$ON^4$	$N^2$
Semi-direct MP2 Grad	$ON^4$	$N^2$	$N^3$	$ON^4$	$N^3$
MP3, CISD, QCISD	$O^2N^4$	$N^2$	$N^4$	$O^2N^4$	$N^4$
MP4, QCISD(T)	$O^3V^4$	$N^2$	$N^4$	$O^3V^4$	$N^4$
Full CI	$((O+V)!/O!V!)^2$				

$N$  = number of basis functions,  $O$  = number of occupied orbitals correlated,  $V$  = number of virtual orbitals

words. Keeping the full set of two electron integrals in memory allows the integrals to be processed using matrix operations and eliminates all of the I/O. Consequently this approach is very fast; however, the  $N^4$  memory dependence limits it to small molecules and/or large memory machines.

**Direct** The atomic orbital integrals are recomputed as needed. This does not require  $O(N^4)$  internal or external storage but does involve additional computational effort. For large molecules, substantial savings are possible which compensate for this additional effort. However, direct methods are the only choice when memory and disk are exhausted and consequently are inevitably used for the largest calculations.

#### 4. Integral Transformation

Calculations that go beyond the Hartree-Fock approximation require the transformation of the one and two electron integrals over the atomic basis to integrals over the molecular orbitals. The one electron integrals are few in number, hence are easy to transform (eq 10,  $O(N^3)$  work and  $O(N^2)$  storage). The transformation of the two electron

integrals is more challenging<sup>13</sup>. A full transformation of the two electron integrals requires  $O(N^5)$  work, provided that the indices are transformed one at a time. However, many levels of theory beyond Hartree-Fock do not need explicitly all  $N^4$  two electron integrals in the MO basis; it is sufficient to form  $(ip|qr)$ , where one of the indices runs only over the occupied MO's. This reduces the work to  $O(ON^4)$  and the storage to  $O(ON^3)$ .

There are several choices for the manner in which the transformation is carried out, depending on how the AO integrals are handled and how much memory is available.

**Unsorted AO integrals** Because integrals are written to disk in the order that they are generated and only the non-zero integrals are written, the transformation of the first index cannot be vectorized. The transformation of the remaining indices is vectorizable.

**Sorted AO integrals** If the integrals can be sorted into canonical order (with zeros included), then the entire transformation is vectorizable. The cpu time for the sort is small but the I/O demands are heavy unless enough memory and disk are available. If the two electron integrals can be held in core, they can be generated in canonical order.

If there is  $O(N^3)$  memory, the partially transformed integrals,  $(iv|\lambda\sigma)$  for one or more  $i$ 's can be held in core and no I/O is necessary for the transformation of the remaining indices. If only  $O(N^2)$  memory is available, then the half transformed integrals,  $(ip|\lambda\sigma)$  are accumulated on disk before the last two indices are transformed.

## 5. Hartree-Fock and MCSCF Energies

The Hartree-Fock approximation obtained by using a single determinantal wavefunction, eq 4, constructed from molecular orbitals given by eq 6, in the expression for the variational energy, eq 3, and minimizing the energy with respect to the molecular orbital coefficients,  $C_{\mu i}$ .

$$(12) \quad E_{\text{HF}} = \langle \Psi_0 | \hat{H} | \Psi_0 \rangle / \langle \Psi_0 | \Psi_0 \rangle ; \quad \frac{\partial E_{\text{HF}}}{\partial C_{\mu i}} = 0$$

This leads to the matrix form of the Hartree-Fock equations, i.e. the Roothaan-Hall equations<sup>14</sup>:

$$(13) \quad \sum_{\nu} F_{\mu\nu} C_{\nu i} = \epsilon_i \sum_{\nu} S_{\mu\nu} C_{\nu i} ; \quad F_{\mu\nu} = h_{\mu\nu} + \sum_{\lambda\sigma} [(\mu\nu|\lambda\sigma) - (\mu\sigma|\lambda\nu)] P_{\lambda\sigma} ; \quad P_{\lambda\sigma} = \sum_i C_{\lambda i}^* C_{\sigma i}$$

$$(14) \quad E_{\text{HF}} = \sum_{\mu\nu} P_{\mu\nu} h_{\mu\nu} + \frac{1}{2} \sum_{\mu\nu\lambda\sigma} [P_{\mu\nu} P_{\lambda\sigma} - P_{\mu\sigma} P_{\lambda\nu}] (\mu\nu|\lambda\sigma) + V_{\text{nuc}}$$

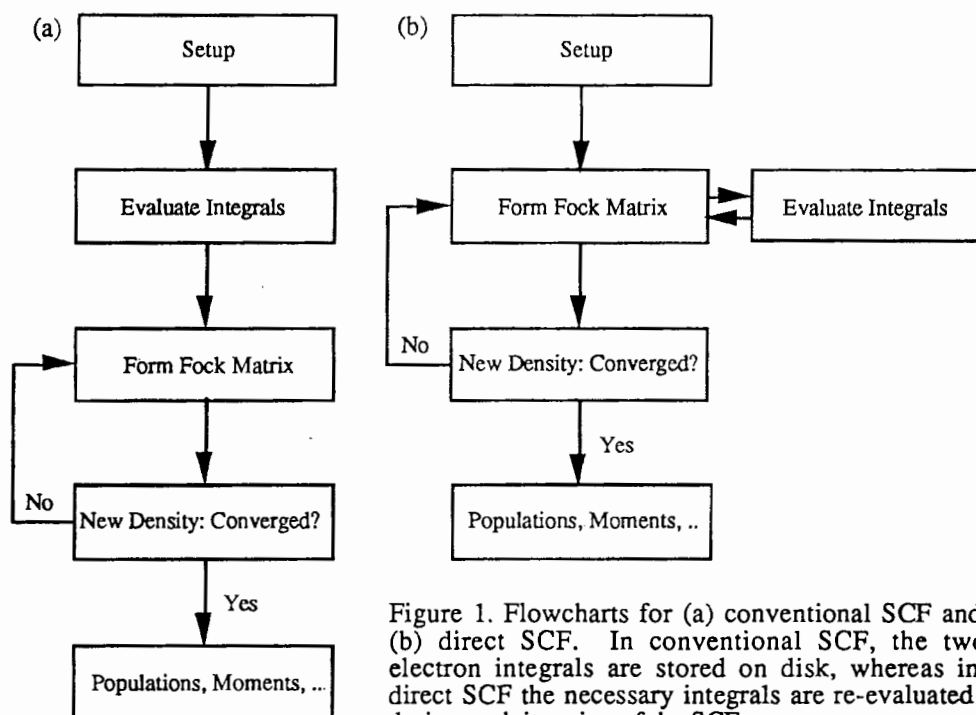
Given the Fock matrix,  $F_{\mu\nu}$ , the MO coefficients can be found from eq 13 by standard diagonalization methods. Since the Fock matrix, depends on the MO coefficients through the density matrix,  $P_{\lambda\sigma}$ , these equations must be solved iteratively until self consistency is

achieved<sup>15</sup> (hence, this is also known as the self consistent field (SCF) method). In effect, each spin orbital is optimized in the average field of all the other orbitals. Hartree-Fock calculations come in a number of flavors, depending on the constraints placed on the molecular orbitals. In restricted Hartree-Fock (RHF) calculations, the alpha (spin up) and the beta (spin down) MO coefficients are the same, whereas in the unrestricted Hartree-Fock (UHF) method, the alpha and beta MO coefficients are allowed to vary independently.

Some systems require a linear combination of determinants for a qualitatively correct description of the electronic wavefunction, e.g. homolytic bond breaking. The Hartree-Fock method can be extended by replacing the single Slater determinant by a linear combination of Slater determinants, eq 5, resulting in a multi-configurational self consistent field (MCSCF) method. The variational energy is minimized with respect to both the MO coefficients,  $C_{\mu i}$ , and the CI coefficients,  $a_i$ , appearing in eq 5. The CI coefficients can be determined in a manner similar to the conventional CI method (see below); the MO coefficients can be determined from a set of Fock-like matrices. As in the Hartree-Fock approach, the equations must be solved iteratively and the most time-consuming part of this process (for small CI expansions) is the construction of the Fock-like matrices. Details of MCSCF methods can be found elsewhere<sup>16</sup>.

The performance issues which arise for SCF calculations concern how and when the integrals are to be generated, and whether and in what format they are to be stored. A flow chart of the conventional SCF method is shown in Figure 1a. The two-electron integrals are computed once and stored externally. The wavefunction is then optimized iteratively, with each iteration consisting of forming a Fock matrix from the stored integrals and the density from the current wavefunction, and then diagonalizing the Fock matrix. Diagonalization requires only  $O(N^3)$  work and thus is not a rate determining step. However, both the generation of integrals and formation of the Fock matrix formally require  $O(N^4)$  arithmetic operations, external storage, and I/O since there are  $O(N^4)$  two electron integrals. For larger systems, many small integrals need not be computed, and empirically the calculations grow in cost as about  $N^{3.5}$ . On most systems, the amount of available external storage is as much of a limitation on the size of conventional SCF calculations as the availability of cpu cycles.

The primary alternative is direct SCF, pioneered by Almlöf and coworkers<sup>17</sup>, illustrated in Figure 1b. Here the two-electron integrals are recomputed as needed (once per iteration). This eliminates the need for external storage and the associated I/O. It might seem that direct SCF would be preferred only when disk space is insufficient, but this is not the case. Since the density matrix is available when the integrals are computed, the determination of which integrals can be neglected can take into account small density matrix elements, resulting in many more integrals being discarded. Furthermore, only the change



in Fock matrix need be computed each iteration, so as convergence is approached and the changes in the density matrix become small, even more integrals can be neglected. The result is that the CPU time required for direct SCF rises as about  $N^{2.7}$ . Consequently, a point is reached at which it actually consumes less CPU time to recompute the integrals as they are needed every time rather than compute all the integrals and store them externally. Where this crossover occurs depends on the speed of the integral evaluation in direct SCF and varies with the integral algorithms used as well as the type of computer involved. Thus there are several distinct cases where direct SCF is to be preferred over conventional SCF:

1. When storing the integrals exceeds available disk capacity.
2. When the calculation is being run on a workstation with a fast cpu but slow disks, so that I/O wait times are prohibitive.
3. When the calculation is big enough that the use of cutoffs and inherent efficiency in the direct SCF procedure more than compensates for the recomputation of integrals. This crossover point, above which all SCF calculations should be run direct even if sufficient disk space is available, is about 90 basis functions using *Gaussian 90*<sup>18</sup> on



vector machines such as the Cray or Convex, but about 180 basis functions on scalar machines such as the Sun or VAX.

Hence, with typical machine configurations one runs conventional SCF until out of disk space on scalar machines, runs all medium to large jobs direct on vector machines, and has a tradeoff for larger jobs on workstations.

The other alternative strategy for SCF calculations is to store the integrals in main memory rather than externally. This in-core SCF procedure eliminates the external storage and I/O while maximizing the speed of Fock matrix formation. The integrals can be generated and stored in memory in canonical order without any sorting. Formation of the Fock matrix is then a simple matrix operation, which is fast on vector machines. The principle limitation of in-core SCF is the requirement of  $N^4/8$  words of main memory for the integrals. This translates to 100 Megabytes of memory for 100 basis functions, which can be found on many machines, 1.6 Gigabytes of memory for 200 basis functions, which is only available on the Cray-2, and 8.1 Gigabytes of memory for 300 basis functions, which cannot currently be obtained on any machine. Since 200-300 basis function direct SCF calculations are quite feasible on departmental machines and super-workstations, this illustrates that memory availability controls the usefulness of in-core SCF and that larger systems will inevitably be done using direct SCF. It is also the case that the pure  $N^4$  size dependence of in-core SCF would cause it to become slower than direct SCF for sufficiently large systems; however, this asymptotic limit has not been reached in practice.

The same considerations apply to the closed-shell restricted, open-shell restricted and open-shell unrestricted SCF procedures. Likewise, both standard iteration procedures such as DIIS and quadratically convergent schemes can be utilized with any approach to handling the two-electron integrals. Only conventional procedures are currently available for MCSCF methods, although the same considerations apply in principle.

## 6. Hartree-Fock and MCSCF Gradients

Analytical gradients, or first derivatives of the energy with respect to atom positions<sup>19</sup> can improve the efficiency of optimizing equilibrium structures and searching for transition states is an order of magnitude more efficient<sup>20</sup>. The flow chart for a typical geometry optimization is shown in Figure 2. The energy and gradient are evaluated and a quasi-Newton algorithm is used to predict the next estimate of the optimized geometry; these steps are repeated until convergence on the gradient and the displacement are achieved. Second derivatives can be calculated by numerical differentiation of the analytical gradients, and proceeds by a flow chart similar to Figure 2 (see below for analytical second derivatives).

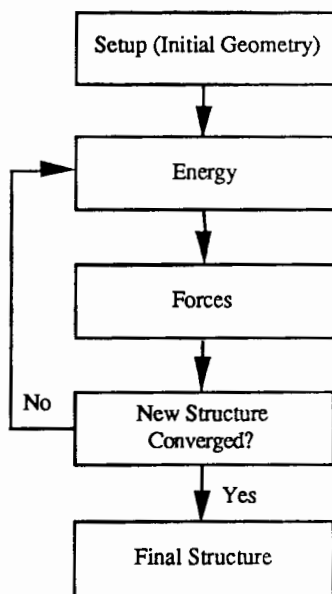


Figure 2. Flowchart for geometry optimization. Starting from an initial guess of the geometry, the energy and forces are calculated. The forces are used to predict a new structure. If the forces and the changes in the geometry are not below a preset threshold, the process is repeated.

The first derivative of the Hartree-Fock energy, eq 14, can be reduced to:

$$(15) \quad E_{\text{HF}}^x = \sum_{\mu\nu} P_{\mu\nu} h_{\mu\nu}^x + \frac{1}{2} \sum_{\mu\nu\lambda\sigma} P_{\mu\nu} P_{\lambda\sigma} (\mu\lambda||\nu\sigma)^x + \sum_{\mu\nu} W_{\mu\nu} S_{\mu\nu}^x + V_{\text{nuc}}^x$$

$$W_{\mu\nu} = -\sum_{\lambda\sigma} P_{\mu\lambda} F_{\lambda\sigma} P_{\sigma\nu}$$

where the superscript  $x$  denotes differentiation with respect to  $x$ . The term in  $W$  arises from the orthonormality constraint on the MO coefficients (or the idempotency condition for the density matrix,  $\mathbf{P} \mathbf{S} \mathbf{P} = \mathbf{P}$ ). Since the wavefunction has been optimized variationally, the derivative of the wavefunction is not needed. The MCSCF gradients can be cast into a similar form, except that the two particle density matrix is not fully separable into the product of one particle density matrices and the construction of  $W$  is a bit more involved<sup>21</sup>.

The derivatives of SCF and MCSCF energies with respect to nuclear coordinates present no new tradeoffs. In some ways, calculation of derivatives is simpler than the computation of the energy because the integral derivatives need not be stored since they can be contracted immediately with the density to form the contribution to the gradient. Since each two electron integral contributes to at most 12 nuclear coordinate derivatives, the cpu time grows only as the number of integrals  $O(N^4)$ , not  $O(N^4 N_a)$  where  $N_a$  is the number of derivatives with respect to nuclear coordinates. MCSCF derivatives are similar, except that

the active space two particle density must be back transformed to the AO basis as the integral derivatives are computed. This does not present storage problems because the number of orbitals in the active space is typically very small.

## 7. Hartree-Fock Second Derivatives

Analytic SCF second derivatives with respect to nuclear coordinates are useful for geometry optimizations and for calculating vibrational spectra. In addition to the force constants, which determine the fundamental frequencies, the dipole derivatives (second derivatives: once with respect to an electric field and once with respect to nuclear coordinates) determine the infra-red intensities and the polarizability derivatives (third derivatives: twice with respect to an electric field and once with respect to nuclear coordinates) determine the Raman intensities. Each of these requires the first derivatives of the MO coefficients along with first and second derivatives of the appropriate integrals.

The second derivatives of the Hartree-Fock energy with respect to the nuclear positions can be written as<sup>22</sup>:

$$(16) E_{\text{HF}}^{\text{xy}} = \sum_{\mu\nu} P_{\mu\nu} h_{\mu\nu}^{\text{xy}} + \frac{1}{2} \sum_{\mu\nu\lambda\sigma} P_{\mu\nu} P_{\lambda\sigma} (\mu\lambda||\nu\sigma)^{\text{xy}} + \sum_{\mu\nu} W_{\mu\nu} S_{\mu\nu}^{\text{xy}} + \sum_{\mu\nu} P_{\mu\nu}^{\text{x}} F_{\mu\nu}^{\text{y}} + \sum_{\mu\nu} W_{\mu\nu}^{\text{x}} S_{\mu\nu}^{\text{y}}$$

$$F_{\mu\nu}^{\text{x}} = h_{\mu\nu}^{\text{x}} + \sum_{\lambda\sigma} [(\mu\lambda||\nu\sigma)^{\text{x}} P_{\lambda\sigma} + (\mu\lambda||\nu\sigma) P_{\lambda\sigma}^{\text{x}}]; \quad P_{\mu\nu}^{\text{x}} = \sum_i [C_{\mu i}^{\text{x}*} C_{\nu i} + C_{\mu i}^* C_{\nu i}^{\text{x}}]$$

$$W_{\mu\nu}^{\text{x}} = - \sum_{\lambda\sigma} [P_{\mu\lambda}^{\text{x}} F_{\lambda\sigma} P_{\sigma\nu} + P_{\mu\lambda} F_{\lambda\sigma}^{\text{x}} P_{\sigma\nu} + P_{\mu\lambda} F_{\lambda\sigma} P_{\sigma\nu}^{\text{x}}]$$

The derivatives of the MO coefficients or the density matrix can be obtained by solving the the corresponding derivatives of the Hartree-Fock equations, i.e. the coupled perturbed Hartree-Fock (CPHF) equations<sup>23</sup>. These can be solved either in the MO basis<sup>23</sup> or in the AO basis<sup>24</sup>. After some manipulation, the CPHF equations in the MO basis can be reduced to:

$$(17) \sum_{bj} [(\epsilon_i - \epsilon_a) \delta_{ab} \delta_{ij} - (ab||ij) - (aj||ib)] P_{bj}^{\text{x}} = [h_{ai}^{\text{x}} - \epsilon_i S_{ai}^{\text{x}} + \sum_j (aj||ij)^{\text{x}} - \sum_{jk} (aj||ik) S_{jk}^{\text{x}}]$$

for the virtual-occupied block of the density. The occupied-occupied block is given by the orthonormality constraint,  $P_{ij}^{\text{x}} = -S_{ij}^{\text{x}}$ , and the virtual-virtual block is zero. The CPHF equations in the AO basis can be written as:

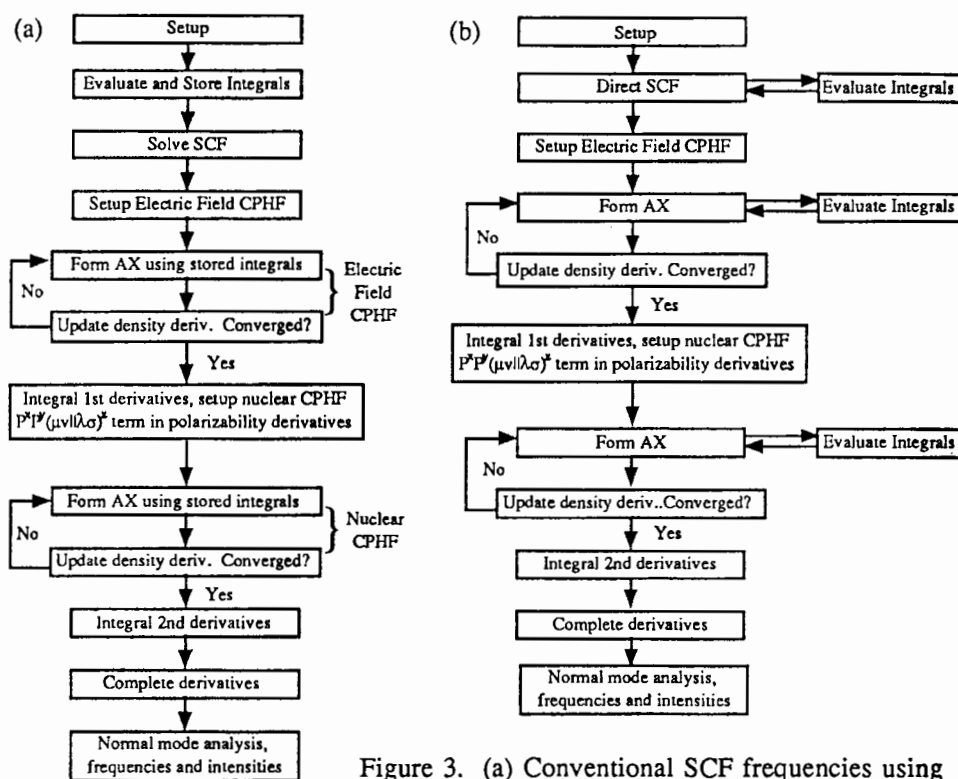


Figure 3. (a) Conventional SCF frequencies using integrals stored on disk. (b) Direct SCF frequencies, avoiding storage of the two electron integrals.

$$\begin{aligned}
 (18) \quad & \sum_{\lambda\sigma}^{ov} [F_{\mu\lambda} S_{\sigma\nu} - S_{\mu\lambda} F_{\sigma\nu} - (\mu\sigma||\nu\lambda)] {}^{ov}P_{\lambda\sigma}^x \\
 & = {}^{ov}h_{\mu\nu}^x + \sum_{\lambda\sigma}^{ov} [(\mu\sigma||\nu\lambda)^x P_{\lambda\sigma} - (\mu\sigma||\nu\lambda) {}^{oo}S_{\lambda\sigma}^x - F_{\mu\lambda} P_{\lambda\sigma} {}^{ov}S_{\sigma\nu}^x]
 \end{aligned}$$

where superscripts  $oo$  and  $ov$  denote projection on the occupied-occupied and occupied-virtual spaces; as in the MO basis  ${}^{oo}P_{\mu\nu}^x = -{}^{oo}S_{\mu\nu}^x$ .

The CPHF equations are linear in the density derivatives and have the form  $AX = B$  for each perturbation. In the MO basis  $A$  has  $O^2V^2$  elements and  $X$  and  $B$  have  $OV$  elements per perturbation. A straightforward solution of the linear equations by inversion of  $A$  would require  $O^2V^2$  memory and  $O^3V^3$  arithmetic operations. Instead, an iterative solution of these equations<sup>22</sup> permits  $A$  to be stored on disk and reduces the operation count to  $O(N_a O^2V^2)$ , where  $N_a$  is the number of derivatives (3 times the number of atoms

in the case of nuclear coordinates). Thus CPHF has quintic dependence on the size of the system. Similarly, the transformation of the integrals to the molecular orbital basis requires  $ON^4$  operations and  $O^2V^2$  storage. All other components of the second derivatives are no worse than quartic in cost.

Alternatively, the same iterative solution of the CPHF equations in the AO basis forms the products of  $AX$  from the integrals over atomic orbitals and the density derivatives without forming or storing  $A$  explicitly<sup>24</sup>. The two electron contributions to  $AX$  in the CPHF equations are the same as the two electron contributions to the Fock matrices in the HF equations but formed with density derivatives, reflecting the derivative nature of the CPHF equations. Neither integral transformation nor external storage for the transformed integrals is required. The CPHF solution requires  $O(N_aN^4)$  arithmetic operations, i.e. still quintic but with a larger constant factor than for solution in the MO basis. This increase in cost is compensated by the elimination of the transformation step and, in practice, solution in the AO basis is comparable in CPU cost to MO solution and hence preferred.

Figure 3a illustrates all the steps in an SCF frequency calculation using externally stored AO integrals. One term in the polarizability derivatives involves two density matrix derivatives with respect to electric field directions and two-electron integral derivatives with respect to nuclear coordinates. Since the integral derivatives must be computed before the nuclear coordinate CPHF (they contribute to the  $B$ 's in  $A X = B$ ), it is necessary to solve the electric field CPHF problem first. After the integral derivatives, the nuclear CPHF problem can be solved and the force constants completed with the integral second derivative terms. Finally, a normal mode analysis produces the frequencies and intensities from the Cartesian derivatives. If only the IR intensities are of interest, the nuclear and electric field CPHF problems can be solved at the same time. If only the force constants are required (i.e., if the second derivatives are to be used to assist a geometry optimization) then the electric field CPHF can be omitted altogether.

The SCF and CPHF equations may also be solved using integrals stored in main memory. This has the same advantages of speed and disadvantages of large memory requirements as for in-core SCF.

Finally, the integrals can be recomputed during the SCF and CPHF iterations<sup>25</sup>, as shown in Figure 3b. If  $N_aN^2$  memory is available, so that in a given CPHF iteration all perturbations can be handled at once, then the solution using recomputed integrals does not add to the quintic phase of the calculation. Because many Fock matrices are formed at a time, the integral cutoffs in direct CPHF are not as effective as in direct SCF, so that direct CPHF does not become faster than conventional for large systems, but the additional overhead of direct solution is modest (20-30%). Unlike the conventional solution, the extra overhead of separate solution of the electric field CPHF adds significantly (about 20%) to a direct frequency calculation, because the integrals are recomputed while solving for only 3 density derivatives.

## 8. MP2 Energy

One of the simplest means of including electron correlation energy is by second order perturbation theory:

$$(19) \quad (\hat{H}_0 + \lambda \hat{V}) (\Psi_0 + \lambda \sum_s a_s \Psi_s \dots) = (E^{(0)} + \lambda E^{(1)} + \lambda^2 E^{(2)} + \dots) (\Psi_0 + \lambda \sum_s a_s \Psi_s \dots)$$

$$\hat{H}_0 \Psi_s = E_s \Psi_s; \quad \hat{V} = \hat{H} - \hat{H}_0$$

$$(20) \quad E^{(0)} + E^{(1)} = \langle \Psi_0 | \hat{H}_0 + \hat{V} | \Psi_0 \rangle = E_{\text{HF}}$$

$$(21) \quad E^{(2)} = \sum_s a_s \langle \Psi_0 | \hat{V} | \Psi_s \rangle; \quad a_s = \langle \Psi_s | \hat{V} | \Psi_0 \rangle / (E_0 - E_s)$$

The most convenient choice for  $\hat{H}_0$  is the Fock operator since the Hartree-Fock wavefunction is an eigenfunction of the Fock operator, as are the excited configurations derived from it by replacing occupied orbitals by virtual orbitals. This choice of  $H_0$  yields Møller-Plesset (MP) perturbation theory or many-body perturbation theory (MBPT)<sup>26,27</sup>. Through second order, the energy is:

$$(22) \quad E_{\text{MP2}} = E_{\text{HF}} + E^{(2)} = E_{\text{HF}} + \frac{1}{4} \sum_{ijab} a_{ij}^{ab} (ij||ab); \quad a_{ij}^{ab} = - (ij||ab) / (\epsilon_a + \epsilon_b - \epsilon_i - \epsilon_j)$$

Note that single excitations do not contribute because of Brillouin's theorem<sup>28</sup> and excitations higher than double do not contribute because  $\hat{V}$  is only a two electron operator.

A conventional MP2 energy calculation, illustrated in Figure 4a, consists of an SCF calculation followed by an integral transformation and the formation of the antisymmetrized integrals which are summed into  $E^{(2)}$ . The only quintic step is the transformation; since only transformed integrals involving two occupied and two virtual orbitals are needed, this has  $O(N^4)$  cost.

The second-order energy can also be computed using direct methods<sup>29</sup>, Figure 4b. Following an SCF calculation (which might be either conventional or direct), a transformation is done in which the two-electron integrals are computed as needed. This can be done very efficiently by computing a double-length set of integrals (eliminating the  $\mu\nu \leftrightarrow \lambda\sigma$  permutation, yielding  $N^4/4$  rather than  $N^4/8$  integrals) which permits vectorization of both the integral evaluation and all steps in the transformation. Thus the direct method eliminates the use of external storage while at the same time offering maximal efficiency in the transformation step. Main memory requirements are kept bounded by restricting the number of occupied orbitals in the outer loop which are handled in each batch - that is, transformed integrals (ij||ab) are formed for a batch of i's and all occupied j, and virtual a,b at a time. A minimum of  $OVN$  memory is required and in this limit the integrals are computed  $O$  times (equivalent to 2  $O$  canonical integral evaluations).

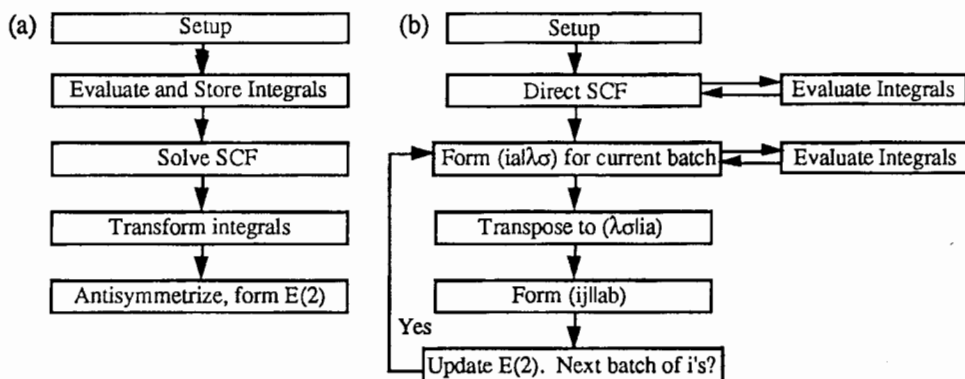


Figure 4. (a) Conventional MP2 energy and (b) direct MP2 energy.

The memory requirements for direct MP2 can be kept quadratic and the number of integral evaluations kept down by using a semi-direct algorithm<sup>30</sup>, similar to Figure 4b. This involves storing the half-transformed integrals ( $ia\lambda\sigma$ ) and transposing them using external storage. The amount of external storage used can be held at any fixed amount by restricting the number of occupied orbitals  $i$  in a batch. Memory requirements are reduced to  $O(N^2)$  and disk requirements are  $O(VN^2)$  per occupied orbital in the batch. In the limit of a single pass, the semi-direct MP2 method uses  $OVN^2$  disk, which is less than corresponding conventional algorithms and is considerably more efficient. Thus the combination of conventional SCF and one pass of semi-direct MP2 supersedes traditional MP2 algorithms entirely.

Finally, MP2 can be done with all integrals in memory. A double-length integral list is required, so that twice as much memory is required as for in-core SCF. This is again very CPU-efficient but has even more stringent memory requirements than in-core SCF.

## 9. MP2 Gradient

First derivatives of the MP2 energy are needed not only for geometry optimization but also for one electron properties such as the dipole moment. After some algebra, the gradient of the MP2 energy can be cast into a form similar to the Hartree-Fock gradient<sup>22,31</sup>:

$$(23) \quad E_{\text{MP2}}^x = \sum_{\mu\nu} P_{\mu\nu}^{\text{MP2}} h_{\mu\nu}^x + \frac{1}{2} \sum_{\mu\nu\lambda\sigma} \Gamma_{\mu\nu\lambda\sigma}^{\text{MP2}} (\mu\nu|\lambda\sigma)^x + \sum_{\mu\nu} W_{\mu\nu}^{\text{MP2}} S_{\mu\nu}^x + V_{\text{nuc}}^x$$

where  $P$ ,  $\Gamma$  and  $W$  are given by:

$$(24) \quad P^{\text{MP2}} = P^{\text{HF}} + P^{(2)}; \quad W^{\text{MP2}} = W^{\text{HF}} + W^{(2)}; \quad \Gamma_{\mu\nu\lambda\sigma}^{\text{MP2}} = \Gamma_{\mu\nu\lambda\sigma}^{\text{S}} + \Gamma_{\mu\nu\lambda\sigma}^{\text{NS}}$$

$$\Gamma_{\mu\nu\lambda\sigma}^{\text{S}} = (P_{\mu\nu}^{\text{HF}} + 2 P_{\mu\nu}^{(2)}) P_{\lambda\sigma}^{\text{HF}} - (P_{\mu\sigma}^{\text{HF}} + 2 P_{\mu\sigma}^{(2)}) P_{\lambda\nu}^{\text{HF}}; \quad \Gamma_{\mu\nu\lambda\sigma}^{\text{NS}} = 2 \sum_{ijab} a_{ij}^{ab} C_{\mu i}^* C_{\nu a} C_{\lambda j}^* C_{\sigma b}$$

The occupied-occupied and virtual-virtual blocks of  $P^{(2)}$  and  $W^{(2)}$  are given by:

$$(25) \quad P_{ij}^{(2)} = -\frac{1}{2} \sum_{kab} a_{ik}^{ab} a_{jk}^{ab}; \quad W_{ij}^{(2)} = \frac{1}{2} \sum_{kab} a_{ik}^{ab} (\text{killab}) - \epsilon_i P_{ij}^{(2)} - \sum_{pq} P_{pq}^{(2)} (\text{iplljq})$$

$$P_{ab}^{(2)} = \frac{1}{2} \sum_{ijc} a_{ij}^{ac} a_{ij}^{bc}; \quad W_{ab}^{(2)} = \frac{1}{2} \sum_{ijc} a_{ik}^{bc} (\text{ijllca}) - \epsilon_a P_{ab}^{(2)}$$

The occupied-virtual block requires the solution of a single CPHF-like equation:

$$(28) \quad \sum_{bj} [(\epsilon_a - \epsilon_i) \delta_{ab} \delta_{ij} + (\text{abllij}) + (\text{ajllib})] P_{ai}^{(2)} = L_{ai}$$

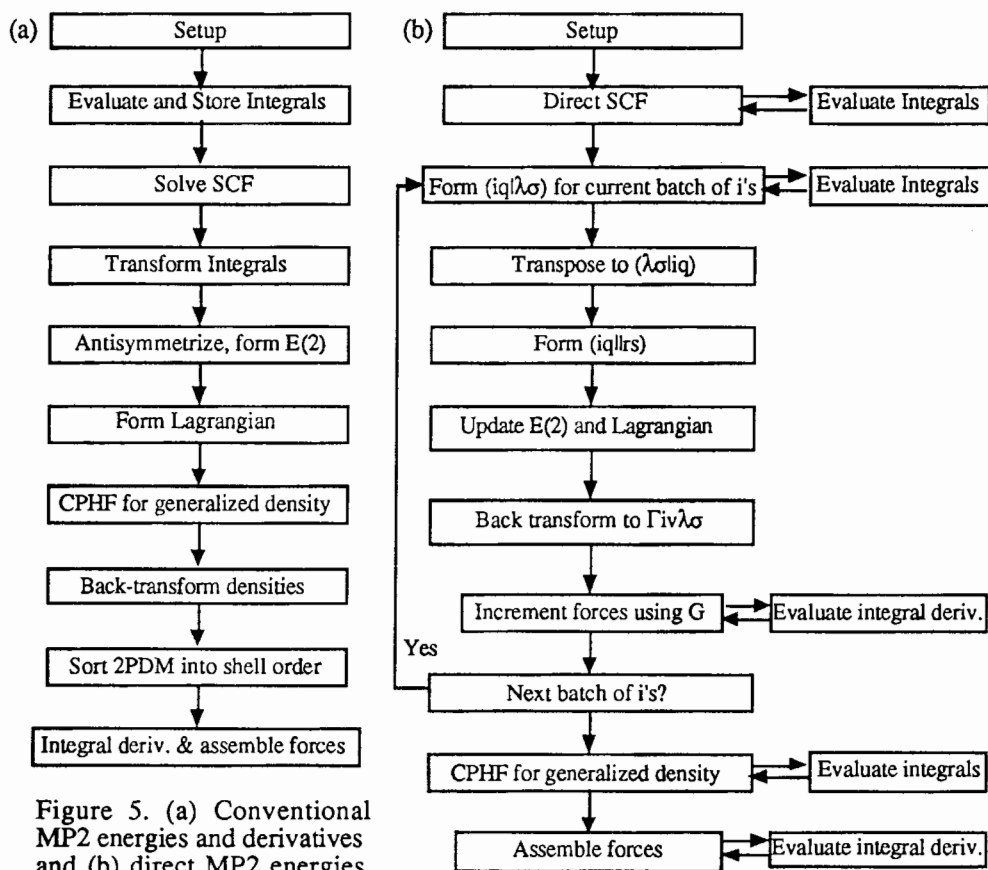
$$L_{ai} = \frac{1}{2} \sum_{jbc} a_{ij}^{ab} (\text{jallbc}) + \frac{1}{2} \sum_{jkb} a_{ij}^{ab} (\text{jkllib}) + \sum_{bc} P_{bc}^{(2)} (\text{icllab}) + \sum_{jk} P_{bc}^{(2)} (\text{killja})$$

$$W_{ai}^{(2)} = \frac{1}{2} \sum_{jkb} a_{ik}^{ba} (\text{jkllib}) - \epsilon_i P_{ai}^{(2)}$$

These equations can be derived by differentiating the expression for  $E_{\text{MP2}}$  and using the Z-vector method<sup>32</sup> to avoid solving the CPHF equations explicitly for each of the derivatives of the MO coefficients.

In a conventional MP2 gradient calculation, illustrated in Figure 5a, an SCF calculation is followed by an integral transformation producing all MO integrals involving at least one occupied orbital. The Lagrangian is then formed by multiplying integrals and the CPHF equation solved for the generalized density. The non-separable part of two-particle density matrix,  $\Gamma^{\text{NS}}$ , is transformed back to the AO basis and sorted into the order in which the integral derivatives will be evaluated. The integral derivatives can then be evaluated and immediately contracted with the sum of the non-separable part two-particle density (read from disk) and the separable part,  $\Gamma^{\text{S}}$ , which is a product of one-particle densities stored in memory. There are two quintic steps: the integral transformation and the two-particle density back transformation, both  $O(N^4)$ . External storage is required for the AO integrals, MO integrals, two-particle density and for sorting the partially and fully back transformed two-particle density, each  $O(N^4)$ .





The MP2 gradient can also be computed using direct and semi-direct techniques<sup>30</sup>, Figure 5b. Again the SCF is followed by a transformation involving batches of occupied orbitals,  $i$ . The transformed integrals are formed as for evaluation of just the MP2 energy, except that integrals  $(iq|lrs)$  are formed for all  $q$ ,  $r$ , and  $s$  and the current batch of  $i$ 's. These are used to form  $E^{(2)}$  and to add contributions to  $L_{ai}$ ,  $P^{(2)}$  and  $W^{(2)}$ . Then the two-particle density for the current batch is formed and back transformed. The integral derivatives are evaluated and contracted with the back-transformed two-particle density. Thus one double-length integral evaluation and one integral derivative evaluation are done for each batch. After all batches have been processed and the construction of Lagrangian is finished, the CPHF equation is solved and the computation of  $P^{(2)}$ ,  $W^{(2)}$  and  $\Gamma^S$  is completed. Finally, the contribution of the  $\Gamma^S$  and integral derivatives can be computed with an additional integral derivative evaluation. Thus if there are  $b$  batches, the complete MP2 gradient requires twice the time for a direct SCF plus  $(2b+1)$  integral evaluations and  $(b+1)$  integral derivative evaluations. Since the transformed integrals over all  $q,r,s$  are required, the fully

direct version requires  $N^3$  memory and the semi-direct algorithm requires  $N^2$  memory and  $N^3$  disk for each occupied in a batch. As for the energy, the semi-direct algorithm can run in any specified amount of disk by reducing the batch size and when run in a single pass with conventional SCF requires both less disk and less CPU time than the conventional algorithm. The additional overhead (as compared to the conventional algorithm) of doing two extra integral evaluations and one extra integral derivative evaluation is more than compensated for by the high efficiency of both the forward and back transformations.

## 10. MP3 and MP4 Energies

Although a sizeable fraction of the electron correlation energy is recovered by second order perturbation theory, it is often necessary to obtain a better estimate of the correlation energy. Among the non-iterative methods, third and fourth order perturbation theory are practical<sup>27</sup>:

$$(29) \quad E^{(3)} = \sum_{st}^D \frac{\langle \Psi_0 | \hat{V} | \Psi_s \rangle \langle \Psi_s | \bar{V} | \Psi_t \rangle \langle \Psi_t | \hat{V} | \Psi_0 \rangle}{(E_0 - E_s)(E_0 - E_t)}; \quad \bar{V} = \hat{V} - \langle \Psi_0 | \hat{V} | \Psi_0 \rangle$$

$$(30) \quad E^{(4)} = - \sum_{st}^D \frac{\langle \Psi_0 | \hat{V} | \Psi_s \rangle \langle \Psi_s | \hat{V} | \Psi_0 \rangle \langle \Psi_0 | \hat{V} | \Psi_t \rangle \langle \Psi_t | \hat{V} | \Psi_0 \rangle}{(E_0 - E_s)(E_0 - E_t)^2} \\ + \sum_{su}^D \sum_t^{\text{SDTQ}} \frac{\langle \Psi_0 | \hat{V} | \Psi_s \rangle \langle \Psi_s | \bar{V} | \Psi_t \rangle \langle \Psi_t | \hat{V} | \Psi_u \rangle \langle \Psi_u | \hat{V} | \Psi_0 \rangle}{(E_0 - E_s)(E_0 - E_t)(E_0 - E_u)}$$

To simplify the expressions for  $E^{(3)}$  and  $E^{(4)}$ , it is useful to define arrays  $u$ ,  $v$  and  $\Delta$ , and to expand them in terms of the integrals and amplitudes<sup>33</sup>:

$$(31) \quad u_s = \sum_t^D \frac{\langle \Psi_s | \bar{V} | \Psi_t \rangle \langle \Psi_t | \hat{V} | \Psi_0 \rangle}{(E_0 - E_t)}$$

$$(32) \quad u_i^a = - \sum_{jb} (j\text{all}ib) a_j^b - \frac{1}{2} \sum_{jbc} (j\text{all}bc) a_{ij}^{bc} - \frac{1}{2} \sum_{jkb} (jk\text{ll}ib) a_{jk}^{ab}$$

$$(33) \quad u_{ij}^{ab} = \sum_c [(ab\text{ll}cj) a_i^c - (i\leftrightarrow j)] - \sum_k [(k\text{bll}ij) a_k^a - (a\leftrightarrow b)] \\ + \frac{1}{2} \sum_{cd} (ab\text{ll}cd) a_{ij}^{cd} + \frac{1}{2} \sum_{kl} (k\text{ll}ij) a_{kl}^{ab} - \sum_{kc} [(k\text{bll}jc) a_{ik}^{ac} - (a\leftrightarrow b) - (i\leftrightarrow j) + (ia\leftrightarrow jb)]$$

$$\begin{aligned}
(34) \quad u_{ijk}^{abc} &= \\
&\sum_c [(bc|lek)a_{ij}^{ac} - (a \leftrightarrow b) - (a \leftrightarrow c) - (i \leftrightarrow k) - (j \leftrightarrow k) + (ia \leftrightarrow kb) + (ja \leftrightarrow kb) + (ia \leftrightarrow kc) + (ja \leftrightarrow kc)] \\
&+ \sum_m [(cm|ljk)a_{im}^{ab} - (a \leftrightarrow c) - (b \leftrightarrow c) - (i \leftrightarrow j) - (j \leftrightarrow k) + (ia \leftrightarrow jc) + (ja \leftrightarrow kc) + (ib \leftrightarrow jc) + (jb \leftrightarrow kc)] \\
(35) \quad v_s &= \sum_t \sum_u^Q \langle \Psi_s | \bar{V} | \Psi_t \rangle \langle \Psi_t | \bar{V} | \Psi_u \rangle \langle \Psi_u | \hat{V} | \Psi_0 \rangle / (E_0 - E_t) (E_0 - E_u) \\
(36) \quad v_i^a &= \frac{1}{2} \sum_{jkbc} (jkl|bc) [a_i^b a_{jk}^{ca} + a_j^a a_{ik}^{cb} + 2 a_j^b a_{ik}^{ac}] \\
(37) \quad v_{ij}^{ab} &= \frac{1}{4} \sum_{jkbc} (kl|lcd) [a_{ij}^{cd} a_{kl}^{ba} - 2(a_{ij}^{ac} a_{kl}^{bd} + a_{ij}^{bd} a_{kl}^{ac}) - 2(a_{ik}^{ab} a_{jl}^{cd} + a_{ik}^{cd} a_{jl}^{ab}) + 4(a_{ik}^{ac} a_{jl}^{bd} + a_{ik}^{bd} a_{jl}^{ac})] \\
(38) \quad \Delta_i^a &= (\epsilon_a - \epsilon_i); \quad \Delta_{ij}^{ab} = (\epsilon_a + \epsilon_b - \epsilon_i - \epsilon_j); \quad \Delta_{ijk}^{abc} = (\epsilon_a + \epsilon_b + \epsilon_c - \epsilon_i - \epsilon_j - \epsilon_k); \quad \text{etc.}
\end{aligned}$$

where  $(p \leftrightarrow q)$  denotes the term obtained by permuting indices  $p$  and  $q$ . The formation of  $u_i^a$  and  $v_i^a$  is only  $O(N^5)$  work. The computation of  $u(ab,ij)$  involves  $O(O^2N^4)$  cpu; the full transformation of the integrals can be avoided by constructing the  $(ab|cd)$  contribution directly from the AO integrals. The  $u_{ijk}^{abc}$  require  $O(O^3V^4)$  work and are the most costly terms. The  $v_{ij}^{ab}$  can be assembled in stages in which the largest fraction of the work is proportional to  $O(O^3V^3)$ .

With  $a_i^a = 0$  and  $a_{ij}^{ab} = -(ij|lab) / \Delta_{ij}^{ab}$  in the above definitions, the MP3 and MP4 energies become<sup>33</sup>

$$\begin{aligned}
(39) \quad E_{MP3} &= E_{MP2} + \frac{1}{4} \sum_{ijab} a_{ij}^{ab} u_{ij}^{ab} \\
(40) \quad E_{MP4} &= E_{MP3} - E^{(2)} \frac{1}{4} \sum_{ijab} |a_{ij}^{ab}|^2 - \sum_{ia} |u_i^a|^2 / \Delta_i^a - \frac{1}{4} \sum_{ijab} |u_{ij}^{ab}|^2 / \Delta_{ij}^{ab} \\
&\quad - \frac{1}{36} \sum_{ijkabc} |u_{ijk}^{abc}|^2 / \Delta_{ijk}^{abc} + \frac{1}{4} \sum_{ijab} a_{ij}^{ab} v_{ij}^{ab}
\end{aligned}$$

The five terms in the MP4 energy are the renormalization, singles, doubles, triples and quadruples contributions, respectively.

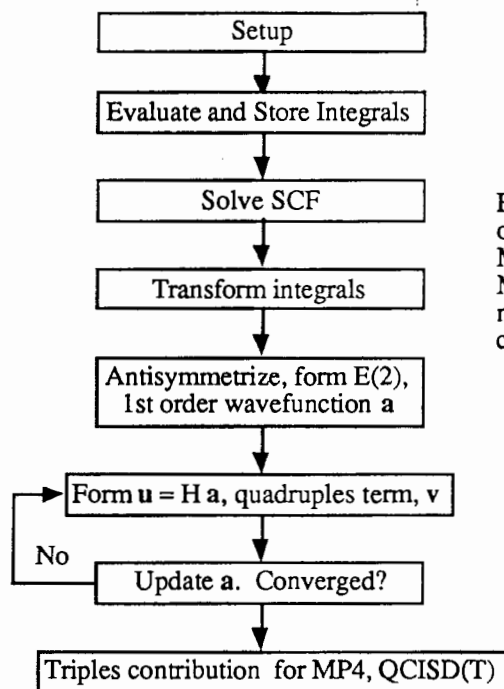


Figure 6. MP3, MP4, CI, CC or QCI energies. For MP3 and MP4, there is no iteration; for MP3, CID and CISD there is no triples or quadruples contribution.

### 11. CI, CC and QCI energies

The CI energy and wavefunction can be found by substituting the CI expansion of the wavefunction, eq 5, into the expression for the variational energy, eq 3, and solving the corresponding matrix eigenvalue problem.

$$(41) \quad \sum_t H_{st} a_t = E_{CI} a_s ; \quad \Psi_{CI} = \sum_s a_s \Psi_s ; \quad H_{st} = \langle \Psi_s | \hat{H} | \Psi_t \rangle$$

However, the Hamiltonian matrix  $H_{st}$  is too large to be stored explicitly, e.g.  $O^4 V^4$  elements if the CI wavefunction is limited to single and double excitations. Furthermore,  $H_{st}$  is sparse with only  $O(O^2 V^2 N^2)$  non-zero matrix elements. Hence the equations for CI and related methods are solved iteratively, with the sparsity of  $H_{st}$  explicitly taken into account in the formation of the product  $\sum_t H_{st} a_t$ .

To simplify the discussion of the CI, CC and QCI methods, it is desirable to define single, double, triple, etc. substitution operators:

$$(42) \quad \hat{T}_1 = \sum_{ia} a_i^\dagger a_a ; \quad \hat{T}_2 = \frac{1}{4} \sum_{ijab} a_{ij}^\dagger a_{ab} ; \quad \hat{T}_3 = \frac{1}{36} \sum_{ijkabc} a_{ijk}^\dagger a_{abc} \text{ etc.}$$

With these operators, the CI doubles and the CI singles and doubles wavefunctions are:

$$(43) \quad \Psi_{\text{CID}} = (1 + \hat{T}_2) \Psi_0; \quad \Psi_{\text{CISD}} = (1 + \hat{T}_1 + \hat{T}_2) \Psi_0$$

and the CISD equations can be written as:

$$(44) \quad \langle \Psi_0 | \hat{H} | \hat{T}_2 \Psi_0 \rangle = E_{\text{corr}}; \quad \langle \Psi_1^a | \bar{H} | (\hat{T}_1 + \hat{T}_2) \Psi_0 \rangle = a_1^a E_{\text{corr}}; \quad \langle \Psi_{ij}^{ab} | \bar{H} | (1 + \hat{T}_1 + \hat{T}_2) \Psi_0 \rangle = a_{ij}^{ab} E_{\text{corr}}$$

where  $\bar{H} = \hat{H} - E_{\text{HF}}$ . With the definition of the array  $u$  from the previous section, these equations can be rewritten<sup>34</sup>:

$$(45) \quad E_{\text{corr}} = \frac{1}{4} \sum_{ijab} a_{ij}^{ab} (ij||ab); \quad (\Delta_1^a - E_{\text{corr}}) a_1^a + u_1^a = 0; \quad (ab||ij) + (\Delta_{ij}^{ab} - E_{\text{corr}}) a_{ij}^{ab} + u_{ij}^{ab} = 0$$

These equations are solved iteratively for  $E_{\text{corr}}$  and the amplitudes. The work is dominated by the formation of  $u$  at each iteration,  $O(N^4)$  cpu.

Although the CI energy is variational (an upper bound to the exact energy), it is not size-consistent ( $E_{\text{CI}}$  for  $X$  and  $Y$  at large separation is not the sum of  $E_{\text{CI}}$  for  $X$  and  $E_{\text{CI}}$  for  $Y$  computed individually). This is obviously a severe disadvantage when computing reaction energetics. The MPn energies are size consistent; however, iterative methods can give better estimates of the correlation energy than perturbative methods, especially if the correlation energy is large. Hence, it is desirable to have a CI-like method that is size-consistent. This role is filled by the coupled-cluster (CC) method<sup>35</sup> (for a review see ref 27) and the quadratic configuration interaction (QCI) approach<sup>36</sup>. In the CC method, the wavefunction is written as an exponential of the excitation operators. In CCD and CCSD, the excitations involve  $\hat{T}_2$  and  $\hat{T}_1 + \hat{T}_2$ , respectively:

$$(46) \quad \Psi_{\text{CCD}} = \exp(\hat{T}_2) \Psi_0; \quad \Psi_{\text{CCSD}} = \exp(\hat{T}_1 + \hat{T}'_2) \Psi_0, \quad \text{where } \hat{T}'_2 = \hat{T}_2 - \frac{1}{2} \hat{T}_1^2$$

The equations for CCD and CCSD are:

$$(47) \quad \langle \Psi_0 | \hat{H} | \hat{T}_2 \Psi_0 \rangle = E_{\text{corr}}; \quad \langle \Psi_{ij}^{ab} | \bar{H} | (1 + \hat{T}_2 + \frac{1}{2} \hat{T}_2^2) \Psi_0 \rangle = a_{ij}^{ab} E_{\text{corr}}$$

$$(48) \quad \langle \Psi_0 | \hat{H} | \hat{T}_2 \Psi_0 \rangle = E_{\text{corr}}; \quad \langle \Psi_1^a | \bar{H} | (\hat{T}_1 + \hat{T}_2 + \hat{T}_1 \hat{T}_2 - \frac{1}{3} \hat{T}_1^3) \Psi_0 \rangle = a_1^a E_{\text{corr}}$$

$$\langle \Psi_{ij}^{ab} | \bar{H} | (1 + \hat{T}_1 + \hat{T}_2 + \hat{T}_1 \hat{T}_2 - \frac{1}{3} \hat{T}_1^3 + \frac{1}{2} \hat{T}_2^2 - \frac{1}{4} \hat{T}_1^4) \Psi_0 \rangle = a_{ij}^{ab} E_{\text{corr}}$$

With the above definitions of  $u$  and  $v$ , the CCD equations can be reduced to<sup>37</sup>:

$$(49) \quad E_{\text{corr}} = \frac{1}{4} \sum_{ijab} a_{ij}^{ab} (ij||ab); \quad a_i^a = 0; \quad (ab||ij) + \Delta_{ij}^{ab} a_{ij}^{ab} + u_{ij}^{ab} + v_{ij}^{ab} = 0$$

In the QCI approach, only the bare minimum number of excitation operators are added to the CI equations to make them size-consistent. For single and double excitations, this leads to:

$$(50) \quad \langle \Psi_0 | \hat{H} | \hat{T}_2 \Psi_0 \rangle = E_{\text{corr}}; \quad \langle \Psi_i^a | \hat{H} | (\hat{T}_1 + \hat{T}_2 + \hat{T}_1 \hat{T}_2) \Psi_0 \rangle = a_i^a E_{\text{corr}}$$

$$\langle \Psi_{ij}^{ab} | \hat{H} | (1 + \hat{T}_1 + \hat{T}_2 + \frac{1}{2} \hat{T}_2^2) \Psi_0 \rangle = a_{ij}^{ab} E_{\text{corr}}$$

which can be reduced to:

$$(51) \quad E_{\text{corr}} = \frac{1}{4} \sum_{ijab} a_{ij}^{ab} (ij||ab); \quad \Delta_i^a a_i^a + u_i^a + v_i^a = 0; \quad (ab||ij) + \Delta_{ij}^{ab} a_{ij}^{ab} + u_{ij}^{ab} + v_{ij}^{ab} = 0$$

For each iteration, the worst step is the formation of  $u$  and  $v$ , each are  $O(O^2N^4)$  if only a partial transformation of the integrals has been carried out or  $O(O^2V^4)$  for a full transformation (in this case, the  $O(O^3V^3)$  step in forming  $v$  is the most time consuming). The iterative solution of the CI, CC and QCI equations is illustrated in Figure 6.

For many cases, triple excitations are important as well. CCD, CCSD and QCISD omit the triples contributions to different extents. A simple remedy is to add in the missing contribution in a manner similar to perturbation theory<sup>38</sup>. This adds a single (non-iterative) step of  $O(O^3V^4)$  in cpu to an iterative process that is  $O(O^2N^4)$  for each step (Fig. 6). An alternative is to add triples in an iterative fashion<sup>39</sup> (at the cost of  $O(O^3V^4)$  work and storage per iteration).

It is possible to include all excitations within a fixed basis set size, i.e. full CI. Because the work increases roughly as  $((O+V)!/O!V!)^2$  or  $(O+V)^n/n!$ , where  $n$  is the number of electrons correlated, only small systems can be treated - few electrons correlated and/or small active space. However, full CI calculation provide valuable benchmarks for testing other methods for estimating electron correlation energy.

## 12. MP3, MP4, CI, CC and QCI gradients

The gradients for MP3, MP4 and CI methods<sup>40-42</sup> (Figure 7a) are very similar to MP2 gradients. Similar to the MP2 gradients, the full CPHF can be avoided by the  $Z$  vector method<sup>32</sup>. The assembly of  $L$  is, however, somewhat more complicated than the MP2 case. In the coupled cluster and QCI approaches, the amplitudes are not calculated in a variational manner. Hence, the derivatives of the amplitudes must be calculated by the

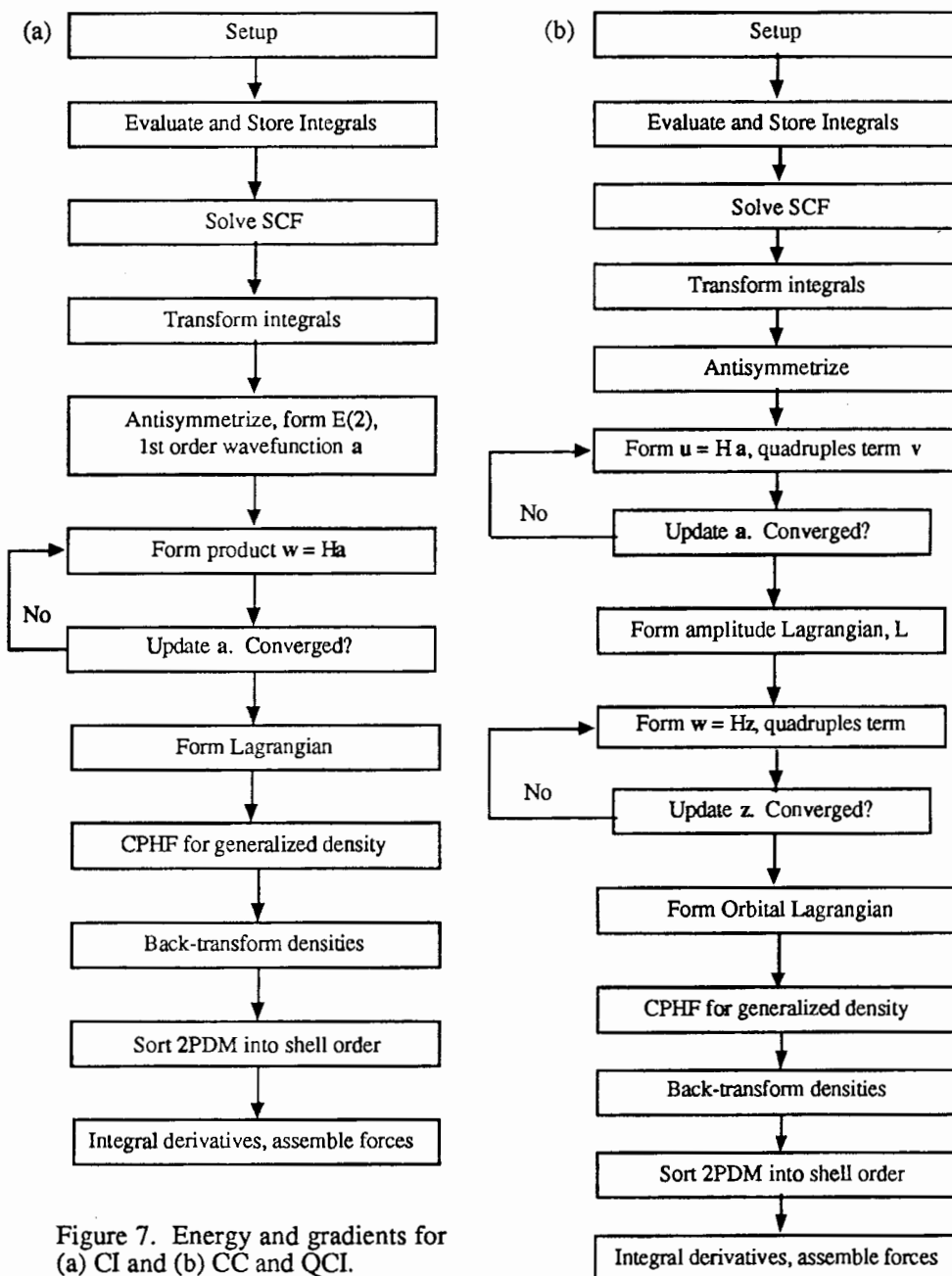


Figure 7. Energy and gradients for (a) CI and (b) CC and QCI.

appropriate set of coupled-perturbed equations<sup>43,44</sup> (Figure 7b). By using the Z vector approach, only one coupled-perturbed equation has to be solved, which requires about the same resources as the original unperturbed equations. The details can be found in the original literature.

### 13. Timing examples

Table 2 gives sample run times for a series of hydrocarbons using *Gaussian 90* on a Multiflow. In the neighborhood of 100 basis functions, the direct method becomes faster than the conventional SCF for these simple hydrocarbons. Figure 8 shows a log-log plot of  $n$  in  $C_nH_{2n+2}$  versus the cpu time. Beyond  $n=3$ , the curves are approximately linear. The crossover between the conventional and the direct SCF is clearly due to the difference in the slopes. The difference in the slopes, in turn, is due to the fact that in the direct SCF, the density matrix can be used in the integral cut-offs to discard more integrals than in the conventional integral evaluation. The position of the crossover depends on the relative speed of the integral evaluation (different algorithms are currently used in *Gaussian 90* for conventional and direct SCF), the spatial extent of the molecule (long molecules have more small integrals that can be eliminated more effectively in the direct procedure) and the number of SCF cycles (the direct method must re-evaluate a fraction of the integrals for every SCF cycle, but the number of SCF cycles is ca 10-15, independent of the size of the molecule).

The second example compares the cost of SCF and MP2 calculations for the normal hydrocarbons. Figure 9 shows the cpu times for MP2 energies (SCF + E2 times from Table 2) and direct SCF energies (also Table 2) using the 6-31G\* basis set. The MP2 calculations use the semi-direct algorithm with a maximum of 100 MW of disk. Below 100 basis functions, the MP2 calculations add only a small fraction to the SCF cpu time. In the 100 to 200 basis function range, the MP2 calculations are proportional to  $N^{3.5}$  -  $N^4$  rather than the formal  $ON^4$ . Even at 200 basis functions, the MP2 calculation costs only  $2^{1/2}$  times the SCF computation.

Figure 10 compares SCF algorithms on vector and scalar machines. For long instruction word machines like the Multiflow, the direct SCF becomes faster than the conventional SCF at ca. 90 basis functions (ratio of direct / conventional less than one). For scalar workstations such as the Sun, the cross-over may be beyond 130 basis functions, depending on the speed of the I/O. For vector machines like the Stardent, direct SCF becomes more efficient for more than 70 basis functions. The third curve in Figure 10 indicates that in-core SCF calculations are always more efficient than conventional SCF. With 48 MB of memory on the Stardent, in-core SCF can be used up to ca. 80 basis functions, beyond which direct SCF is more efficient than conventional SCF.



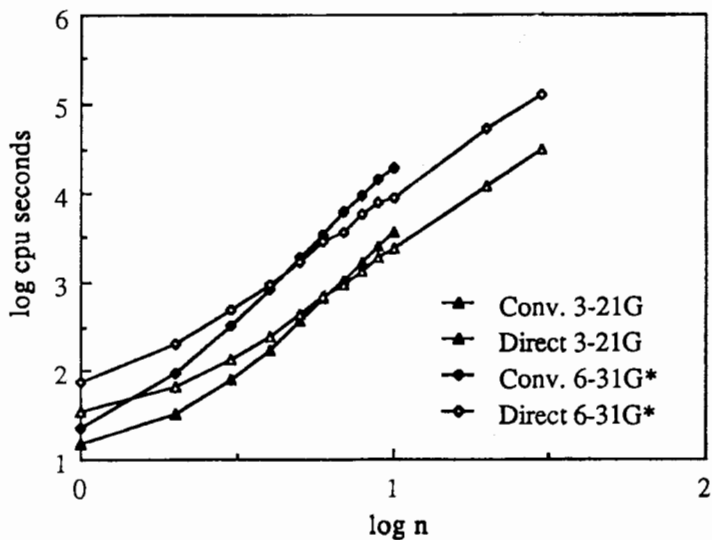


Figure 8. Log-log plot of cpu time vs chain length for normal hydrocarbons,  $C_nH_{2n+2}$  (all trans) for conventional and direct SCF using the 3-21G and 6-31G\* basis sets.

Table 2: Computational times for SCF calculations on normal hydrocarbons  $C_nH_{2n+2}$

n	3-21G			6-31G*			E2
	N	Conventional	Direct	N	Conventional	Direct	
1	17	16	35	23	23	76	10
2	30	32	65	42	97	208	43
3	43	77	132	61	332	501	103
4	56	174	246	80	860	940	275
5	69	360	431	99	1890	1679	606
6	82	652	685	118	3532	2868	1156
7	95	1048	966	137	6181	3759	2224
8	108	1729	1325	156	9229	5727	3503
9	121	2642	1891	175	13907	8041	7411
10	134	3699	2379	194	18771	9061	12452
20	264		12060	384		52906	
30	394		29910	574		127542	

<sup>1</sup> in seconds on a Multiflow Trace 14/300, E2 calculations semi-direct limited to 100MW disk

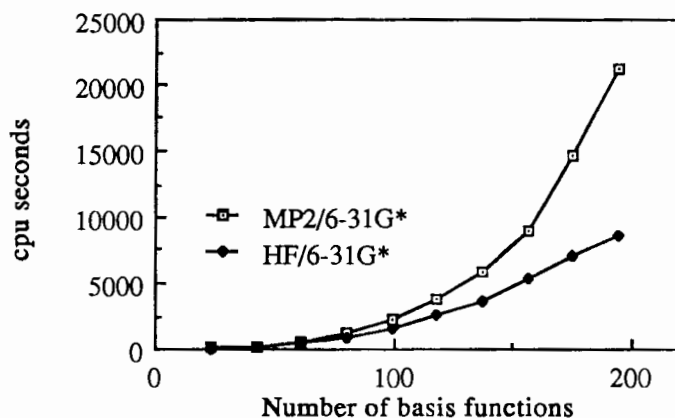


Figure 9. Comparison of cpu times for direct SCF and semi-direct MP2 calculations on normal hydrocarbons using the 6-31G\* basis set. The semi-direct MP2 calculation was limited to a maximum of 100 MW of disk.

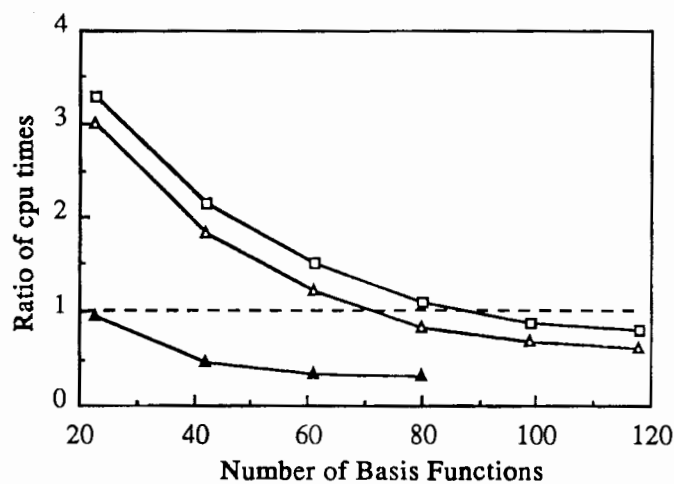


Figure 10. Ratio of cpu times for different SCF algorithms on vector and scalar machines: squares - direct / conventional on a Multiflow Trace 14/300, open triangles - direct / conventional on a Stardent 3020, filled triangles - in-core / conventional on a Stardent using up to 48 MB of memory.

#### 14. References

1. Hehre, W. J., Radom, L., Schleyer, P. vR., and Pople, J. A., *Ab Initio Molecular Orbital Theory*, Wiley-Interscience, New York, 1986.
2. Szabo, A. and Ostlund, N. S., *Modern Quantum Chemistry: Introduction to Advanced Electronic Structure Theory*, Macmillan, New York, 1982.
3. R. McWeeny and Craig, D. P., *Methods of Molecular Quantum Mechanics*, Academic Press, London, 1989.
4. Wilson, S., *Electron Correlation in Molecules*, Clarendon Press, Oxford, 1984.
5. Daudel, R., Leroy, G., Peeters, D. and Sana, M., *Quantum Chemistry*, Wiley-Interscience, Chichester, 1983.
6. Jørgensen, P. and Simons, J., *Second Quantization -Based Methods in Quantum Chemistry*, Academic Press, New York, 1981.
7. Born, M. and Oppenheimer, J. R., *Ann. Physik*, **84**, 457 (1927).
8. Hartree, D. R. *Proc. Cambridge Phil. Soc.*, **24**, 89, 11, 426 (1928); Fock, V. Z. *Phys.*, **61**, 126, (1930); **62**, 795, (1930); **75**, 622, (1932).
9. Slater, J. C., *Phys. Rev.*, **34**, 1293 (1929); **35**, 509 (1930).
10. Boys, S. F., *Proc. Roy. Soc. (London)*, **A200** 542 (1950); for a discussion of the utility of gaussians in quantum chemistry see Shavitt, I., *Methods in Computational Physics*, vol 2, Wiley, New York, 1962.
11. Slater, J. C., *Phys. Rev.*, **36**, 57 (19309)
12. Head-Gordon, M. and Pople, J. A., *J. Chem. Phys.*, **89**, 5777 (1988) and references therein.
13. Bauschlicher, C. W. *Theor. Chim. Acta*, **76**, 187 (1989).
14. Roothaan, C. C. J., *Rev. Mod. Phys.*, **23**, 69 (1951); Hall, G. G., *Proc. Roy. Soc. (London)*, **A205**, 541 (1951).
15. For a discussion of SCF stability and convergence see Schlegel, H. B.; McDouall, J. J. W. in *Computational Advances in Organic Reactions*, Ögretir, C. Csizmadia, I. G., eds., (Kluwer Academic, the Netherlands), 1990.
16. For a review of MCSCF methods see Werner, H. -J. *Adv. Chem. Phys.*, **69**, 1 (1987); Shepard, R. *Adv. Chem. Phys.*, **69**, 63 (1987); Roos, B. O., *Adv. Chem. Phys.*, **69**, 399 (1987).
17. Almlöf, J.; Korsell, K.; Faegri, K., Jr. *J. Comput. Chem.*, **3**, 385 (1982).
18. Frisch, M. J.; Head-Gordon, M.; Trucks, G.; Foresman, J. B.; Schlegel, H. B.; Raghavachari, K.; Robb, M. A.; Binkley, J. S.; Gonzalez, C.; DeFrees, D. J.; Fox, D. J.; Whiteside, R. A.; Seeger, R.; Melius, C. F.; Baker, J.; Martin, L. R.;

- Kahn, L. R.; Stewart, J. J. P.; Fluder, E. M.; Topiol, S.; Pople, J. A. *GAUSSIAN 90*, Gaussian, Inc., Pittsburgh PA, 1990.
19. For some reviews of analytical gradients see Pulay, P., *Adv. Chem. Phys.*, **69**, 241 (1987); Jørgensen, P., and Simons, J. (Eds.), *Geometrical Derivatives of Energy Surfaces and Molecular Properties*, Reidel, Dordrecht, 1986; Gaw, J. F., and Handy, N. C., *Annu. Rep. Prog. Chem. Sec. C*, **81**, 291 (1985); Fogarasi, G., and Pulay, P., *Annu. Rev. Phys. Chem.*, **35**, 191 (1984).
  20. Schlegel, H. B., *Adv. Chem. Phys.*, **67**, 249 (1987).
  21. Kato, S.; Morokuma, K. *Chem. Phys. Lett.*, **65**, 19 (1979); Dupuis, M. *J. Chem. Phys.*, **74**, 5758 (1981); Osamura, Y.; Yamagichi, Y.; Schaefer III, H. F. *J. Chem. Phys.*, **75**, 2919 (1981); **77**, 383 (1982); Schlegel, H. B.; Robb, M. A. *Chem. Phys. Lett.*, **93**, 43 (1982).
  22. Pople, J. A.; Krishnan, R.; Schlegel, H. B.; Binkley, J. S. *Int. J. Quantum Chem., Quantum Chem. Symp.*, **13**, 225 (1979).
  23. Gerratt, J.; Mills, I. M. *J. Chem. Phys.*, **49**, 1719, 1730 (1968).
  24. Osamura, Y.; Yamagichi, Y.; Schaefer III, H. F. *J. Chem. Phys.*, **77**, 383 (1982).
  25. Frisch, M. J.; Head-Gordon, M.; Pople, J. A. *Chem. Phys. Lett.*, **141**, 189 (1990).
  26. Møller, C.; Plesset, M. S. *Phys. Rev.*, **46**, 618 (1934).
  27. For a review of MBPT and coupled cluster methods see Bartlett, R. J. *Annu. Rev. Phys. Chem.*, **32** 359 (1981).
  28. Brillouin, L., *Actualities Sci, Ind.*, **71**, 159 (1934).
  29. Taylor, P. R. *Int. J. Quantum Chem.*, **31**, 521 (1987); Saebø, S.; Almlöf, J. *Chem. Phys. Lett.*, **154**, 83 (1989); Head-Gordon, M.; Pople, J. A.; Frisch, M. J. *Chem. Phys. Lett.*, **153**, 503 (1989).
  30. Frisch, M. J.; Head-Gordon, M.; Pople, J. A. *Chem. Phys. Lett.*, **166**, 275, 281 (1990).
  31. Rice, J. E.; Amos, R. D. *Chem. Phys. Lett.*, **122**, 585 (1985); Simandiras, E. D.; Amos, R. D.; Handy, N. C. *Chem. Phys.*, **114**, 9 (1987)
  32. Handy, N. C., and Schaefer, H. F., *J. Chem. Phys.*, **81**, 5031 (1984).
  33. Krishnan, R.; Pople, J. A. *Int. J. Quantum Chem.*, **14**, 91 (1978); Krishnan, R.; Frisch, M. J.; Pople, J. A. *J. Chem. Phys.*, **72**, 4244 (1980).
  34. Pople, J. A.; Seeger, R.; Krishnan, R. *Int. J. Quantum Chem., Quantum Chem. Symp.*, **11**, 149 (1977).
  35. Cizek, J. *Adv. Chem. Phys.*, **14**, 35 (1969).
  36. Pople, J. A.; Head-Gordon, M.; Raghavachari, K. *J. Chem. Phys.*, **87**, 5968 (1987).

37. Hurley, A. C. *Electron Correlation in Molecules*, Academic, New York, 1976; Pople, J. A.; Krishnan, R.; Schlegel, H. B.; Binkley, J. S. *Int. J. Quantum Chem.*, **14**, 545 (1978).
38. Raghavachari, K. *J. Chem. Phys.*, **82**, 4607 (1985); Raghavachari, K.; Trucks, G. W.; Pople, J. A.; Head-Gordon, M. *Chem. Phys. Lett.*, **157**, 479 (1989).
39. Urban, M.; Noga, J.; Cole, S. J.; Bartlett, R. J. *J. Chem. Phys.*, **83**, 4041 (1985); Noga, J.; Bartlett, R. J. *J. Chem. Phys.*, **86**, 7041 (1987) and references therein.
40. Fitzgerald, G.; Harrison, R.; Laidig, W. D.; Bartlett, R. J. *J. Chem. Phys.*, **82**, 4375 (1985).
41. Gauss, J.; Cremer, D. *Chem. Phys. Lett.*, **138**, 131 (1987); **153**, 303 (1988).
42. Krishnan, R.; Schlegel, H. B.; Pople, J. A. *J. Chem. Phys.*, **72**, 4654 (1980); Brooks, B. R.; Laidig, W. D.; Saxe, P.; Goddard, J. D.; Yamaguchi, Y.; Schaefer II, H. F. *J. Chem. Phys.*, **72**, 4652 (1980).
43. Scheiner, A. C.; Scuseria, G. E.; Rice, J. E.; Lee, T. J.; Schaefer II, H. F. *J. Chem. Phys.*, **87**, 5361 (1987); Salter, E. A.; Trucks, G. W.; Bartlett, R. J. *J. Chem. Phys.*, **90**, 1752 (1989).
44. Gauss, J.; Cremer, D. *Chem. Phys. Lett.*, **150**, 280 (1988); *ibid*, **163**, 549 (1989).