RESEARCH ARTICLE | AUGUST 15 1988

# Determinant based configuration interaction algorithms for complete and restricted configuration interaction spaces

Jeppe Olsen; Björn O. Roos; Poul Jo/rgensen; Hans Jo/rgen Aa. Jensen



*J. Chem. Phys.* 89, 2185–2192 (1988)
https://doi.org/10.1063/1.455063

# Determinant based configuration interaction algorithms for complete and restricted configuration interaction spaces

Jeppe Olsen and Björn O. Roos
*Theoretical Chemistry, Chemical Center, University of Lund, S-22100 Lund, Sweden*

Poul Jørgensen and Hans Jørgen Aa. Jensen
*Department of Chemistry, Aarhus University, 8000 Aarhus C, Denmark*

(Received 8 January 1988; accepted 31 March 1988)

A restricted active space (RAS) wave function is introduced, which encompasses many commonly used restricted CI expansions. A highly vectorized algorithm is developed for full CI and other RAS calculations. The algorithm is based on Slater determinants expressed as products of alphastrings and betastrings and lends itself to a matrix indexing $C(I_\alpha, I_\beta)$ of the CI vector. The major features are: (1) The intermediate summation over determinants is replaced by two intermediate summations over strings, the number of which is only the square root of the number of determinants. (2) Intermediate summations over strings outside the RAS CI space is avoided and RAS calculations are therefore almost as efficient as full CI calculations with the same number of determinants. (3) An additional simplification is devised for $M_S = 0$ states, halving the number of operations. For a case with all single and double replacements out from 415 206 Slater determinants yielding 1 136 838 Slater determinants each CI iteration takes 161 s on an IBM 3090/150(VF).

## I. INTRODUCTION

The CASSCF (complete active space SCF) method has been very successful in describing the static part of the correlation energy (for a recent review see Ref. 1). In a CASSCF calculation a set of orbitals is chosen active, and all configurations with correct space and spin symmetry which can be constructed from these orbitals are included in the configuration space. Full configuration interaction (FCI) iterations are carried out a number of times to obtain a CASSCF wave function. Efficient algorithms for FCI calculations have recently been described by Siegbahn[2] and further developed by Knowles and Handy.[3] In FCI calculations the number of configurations increases very rapidly with the number of active orbitals and restrictions on the CI space from FCI are often necessary, e.g., to treat dynamical correlation, where a substantial number of orbitals is required. The CI space can be restricted by dividing the active orbital space into several subspaces and impose restrictions on the allowed orbital occupancies in each of the subspaces.[1,4] Here a restricted active subspace (RAS) is considered where the active orbital space is divided into three subsets I, II, and III, requiring a minimal number of occupied spin orbitals in space I and a maximal number in space III. The number of electrons in space II is unrestricted and is reminiscent of the active space in CAS calculations. A core of doubly occupied orbitals (inactive orbitals) can also be included in RAS. The RAS structure includes many usual CI spaces as special cases. The development of efficient CI algorithms for RAS calculations allows the use of CI spaces, that are neither complete in a small orbital basis, nor single and double excitations out from a small reference space.

The algorithm we describe is determinant based and uses the technique of Handy[5] of separating the determinants into alphastrings and betastrings. The two-electron contributions are divided into three terms—one involving excitations of two electrons with alpha spin, one involving excitations of two electrons with beta spin, and one term involving mixed excitations. The algorithm may be used to carry out RAS and FCI calculations very efficiently. To illustrate the efficiency we report RAS CI calculations with $10^3$–$10^5$ multireference Slater determinants and single, double, triple and quadruple excitations out of the multireference space. We also report FCI calculations with $\sim 10^6$ Slater determinants. A brief comparison is made to other recent results.

In Sec. II we describe briefly how alpha- and betastrings are used to describe a RAS wave function. In Sec. III we discuss the terms comprising the Hamiltonian matrix times a vector. Simplifications are introduced that may be used for $M_S = 0$ components of a state vector. In Sec. IV it is shown how the direct CI algorithm may be applied for a FCI space. The FCI algorithm is a special case of the RAS algorithm described in Sec. V and the FCI algorithm is only included for pedagogical reasons and in order to simplify the comparisons to other recent FCI algorithms. Sample calculations with FCI and RAS wave functions are reported in Sec. VI. Section VII contains some concluding remarks.

## II. SLATER DETERMINANTS AND STRINGS

A major step forward in determinant based CI was Handy's realization of the computational advances of splitting Slater determinants into alphastrings and betastrings:

$$|\alpha(I_\alpha)\beta(I_\beta)\rangle = \alpha(I_\alpha)\beta(I_\beta)|\text{vac}\rangle. \quad (1)$$

An alphastring $\alpha(I_\alpha)$ is an ordered product of $n_\alpha$ creation operators for molecular spin orbitals with alpha spin, and a betastring is an ordered product of $n_\beta$ creation operators for molecular spin orbitals with beta spin. The constant lengths of the alphastrings and betastrings are defined from the total number of electrons $n$ and the $z$ component of the spin $M_s$.



 0021-9606/88/162185-08$02.10  

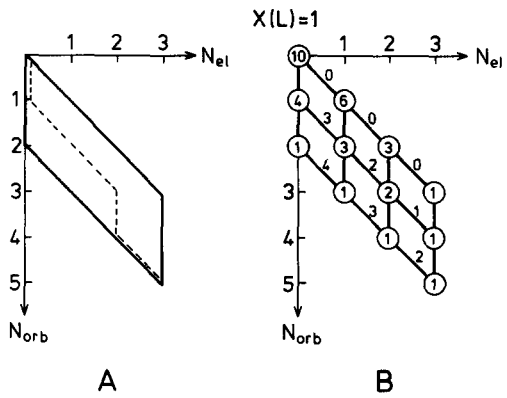


FIG. 1. Alphastring graph for $n_\alpha = 3$, $N = 5$. (A) The dashed line denotes the path for $a^\dagger_{2\alpha}a^\dagger_{3\alpha}a^\dagger_{5\alpha}$. (B) The numbers inside the circles are the vertex weights and the numbers on the sloped lines are the arc weights for lexical ordering.

Each alpha- and betastring is constructed from a set of $N$ orbitals.

The CI space of a given calculation is defined from the set of alphastrings and betastrings that are included, possibly with restrictions on the allowed combination of strings. In order to picturize and order the space of strings it is convenient to define the included strings from one or several graphs. We sketch below the graphical representations for strings and refer to the book by Duch for further details.[6] For the full CI case with $n_\alpha = 3$ and $N = 5$ the space of alpha strings is depicted in Fig. 1. To each string we associate a path in the graph. In Fig. 1(A) the dashed line represents the path for the string $a^\dagger_{2\alpha}a^\dagger_{3\alpha}a^\dagger_{5\alpha}$. A similar representation is obtained for the beta spin.

For restricted CI calculations the space of strings is described by a subset of the paths in the FCI graph, and we must in general use several graphs and introduce restrictions on the allowed combinations of graphs. To see this, consider the case with $n_\alpha = n_\beta = 3$ and $N = 5$. If at most two electrons are allowed in orbitals 4 and 5 we have the graphs for $\alpha$ electrons given in Fig. 2. The graphs have two, one, and zero alpha electrons in orbitals 4 and 5, respectively. For this simple example graph C corresponds to a single path. The space for beta electrons consists of the same three graphs. The alpha and beta graphs cannot be combined freely, since

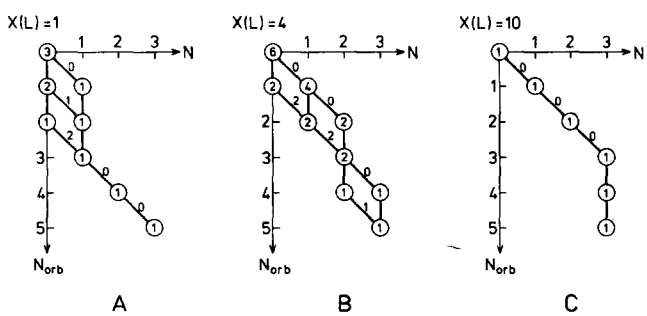


FIG. 2. Alphastring graphs for $n_\alpha = 3$, $N = 5$ with at most two electrons in orbitals 4 and 5. Vertex weights and arc weights are given for lexical ordering.

the total number of electrons in orbitals 4 and 5 must at most be two. Alphastrings of graph A can only be combined with betastrings of graph C; alphastrings of graph B can be combined with betastrings of types B and C; and alphastrings of type C can be combined with betastrings of type A, B, or C. The strings often can be divided into different graphs in several ways. As discussed later this can be used to reduce core storage.

The graphical representation of the alphastrings and betastrings can be used to order the paths, i.e., the strings, according to a sum of arc weights. From the path $L^\alpha$ of a given alphastring we obtain the unique index of that alphastring as

$$I_\alpha(L^\alpha) = X(L^\alpha) + \sum_{i=1}^{N} Y_{L_i}(N_i,i), \tag{2}$$

where $L_i$ is the occupation of the $i$th arc (0 or 1) and $(N_i,i)$ are the coordinates of vertices crossed by $L^\alpha$. We have followed the notation by Duch[6] for "lexical ordering," and we note that for FCI the element $Y_1(N_i,i)$ is identical to the addressing array element $Z(N_i,i)$ defined by Knowles and Handy.[3] We have extended the index definition with $X(L^\alpha)$ which denotes the offset to the alpha graph which path $L^\alpha$ belongs to. Figure 1(B) shows the weight of the vertices and arc weights for the $n_\alpha = 3$, $N = 5$ FCI example; using Eq. (2) on the path dashed in Fig. 1(A) we find that the index of this alpha string is $I_\alpha(L^\alpha) = 8$. Figure 2 illustrates how the offset $X(L^\alpha)$ is used to assure a sequential numbering of the strings when more than one graph is used to describe the desired CI space.

Following Knowles and Handy we write the CI coefficients as a rectangular matrix $C(I_\alpha, I_\beta)$. For restricted CI spaces, where not all combinations of alpha- and betastrings are allowed, only the allowed subblocks in the CI coefficient matrix are included.

The effect of single excitations, $E^\alpha_{ij}$ on alphastrings and $E^\beta_{ij}$ on betastrings, is conveniently calculated using the above ordering. For each alphastring $I_\alpha$ the lists of nonvanishing excitations $ij$, the excited strings $J_\alpha$, and a sign defined through

$$|\alpha(J_\alpha)\rangle = \mathrm{sgn}(ij)E^\alpha_{ij}|\alpha(I_\alpha)\rangle \tag{3}$$

can be calculated and stored in main memory together with the analogous information for betastrings.[3] (For $M_S = 0$ the same lists are used for alpha- and betastrings.) The effect of double, triple, and higher excitations on a given string is obtained by repeated table lookups. In the case of string spaces restricted from the full CI space, excitations out of the allowed string spaces can occur. In such cases we must add the additional graphs and the excitations to and from these to the tables. In Sec. IV we return to this problem and show that excitations out of the model space can be avoided for a RAS model space.

The above is straightforwardly modified to allow for Abelian spatial symmetry. The strings are then organized so that all strings of a given symmetry are stored consecutively. Within a given symmetry block the strings belonging to the same graph are stored together. The imposition of a given symmetry on the CI vector restricts the allowed combina-

tions of strings and the $C$ matrix of CI coefficients becomes a blocked matrix. The individual symmetry blocks of the matrix can again be split into subblocks, according to the separation of strings into different graphs.

## III. DIRECT CI

In a direct CI iteration we have a CI expansion,

$$|0\rangle = \sum_{I_\alpha I_\beta} C(I_\alpha,I_\beta)|\alpha(I_\alpha)\beta(I_\beta)\rangle \tag{4}$$

and wish to carry out a linear transformation on this vector[7]

$$\sigma(I_\alpha,I_\beta) = \sum_{J_\alpha J_\beta} \langle\beta(J_\beta)\alpha(J_\alpha)|\hat{H}|\alpha(I_\alpha)\beta(I_\beta)\rangle C(J_\alpha,J_\beta). \tag{5}$$

We will often refer to the CI vector as the $C$ matrix and the $\sigma$ vector as the $\sigma$ matrix because of the two-dimensional indexing of the determinants. $\hat{H}$ is the nonrelativistic electronic Hamiltonian in a finite basis

$$\hat{H} = \sum_{kl} h_{kl}\hat{E}_{kl} + \frac{1}{2}\sum_{ij,kl}(ij|kl)(\hat{E}_{ij}\hat{E}_{kl} - \delta_{jk}\hat{E}_{il}), \tag{6}$$

where $\hat{E}_{kl}$ is a generator of the linear group

$$\hat{E}_{kl} = \hat{E}^{\alpha}_{kl} + \hat{E}^{\beta}_{kl} \tag{7}$$

and $\hat{E}^{\alpha}_{kl}$ and $\hat{E}^{\beta}_{kl}$ contain the $\alpha$ and $\beta$ components of this generator, respectively. The $\sigma$ vector of Eq. (5) may be written as a sum of three terms:

$$\sigma(I_\alpha,I_\beta) = \sigma_1(I_\alpha,I_\beta) + \sigma_2(I_\alpha,I_\beta) + \sigma_3(I_\alpha,I_\beta), \tag{8}$$

with

$$\begin{aligned}\sigma_1(I_\alpha,I_\beta) &= \sum_{J_\beta}\sum_{kl}\langle\beta(J_\beta)|\hat{E}^{\beta}_{kl}|\beta(I_\beta)\rangle \\ &\times\left[h_{kl} - \frac{1}{2}\sum_j (kj|jl)\right]C(I_\alpha,J_\beta) \\ &+\frac{1}{2}\sum_{J_\beta}\sum_{ijkl}\langle\beta(J_\beta)|\hat{E}^{\beta}_{ij}\hat{E}^{\beta}_{kl}|\beta(I_\beta)\rangle \\ &\times(ij|kl)C(I_\alpha,J_\beta),\end{aligned} \tag{9a}$$

$$\begin{aligned}\sigma_2(I_\alpha,I_\beta) &= \sum_{J_\alpha}\sum_{kl}\langle\alpha(J_\alpha)|\hat{E}^{\alpha}_{kl}|\alpha(I_\alpha)\rangle \\ &\times\left[h_{kl} - \frac{1}{2}\sum_j (kj|jl)\right]C(J_\alpha,I_\beta) \\ &+\frac{1}{2}\sum_{J_\alpha}\sum_{ijkl}\langle\alpha(J_\alpha)|\hat{E}^{\alpha}_{ij}\hat{E}^{\alpha}_{kl}|\alpha(I_\alpha)\rangle \\ &\times(ij|kl)C(J_\alpha,I_\beta),\end{aligned} \tag{9b}$$

$$\begin{aligned}\sigma_3(I_\alpha,I_\beta) &= \sum_{J_\alpha J_\beta}\sum_{ijkl}\langle\beta(J_\beta)|\hat{E}^{\beta}_{ij}|\beta(I_\beta)\rangle \\ &\times\langle\alpha(J_\alpha)|\hat{E}^{\alpha}_{kl}|\alpha(I_\alpha)\rangle(ij|kl)C(J_\alpha,J_\beta).\end{aligned} \tag{9c}$$

In Secs. IV and V we describe how each of these terms may be calculated using vectorized algorithms for FCI as well as RAS model spaces. First some simplifications that may be used in the calculation of the $\sigma$ vector for the $M_S = 0$ component of a state vector will be introduced. In this case the coefficient matrix has the symmetry

$$C(I_\alpha,I_\beta) = (-1)^S C(I_\beta,I_\alpha). \tag{10}$$

To proceed we introduce the spin–flip operator[8]

$$\hat{U} = \exp[i\pi\hat{S}_x], \tag{11}$$

that interchanges alpha and beta spin,

$$\hat{U}\hat{a}^{\dagger}_{r\alpha}\hat{U}^{\dagger} = i\hat{a}^{\dagger}_{r\beta}, \tag{12}$$

$$\hat{U}\hat{a}^{\dagger}_{r\beta}\hat{U}^{\dagger} = i\hat{a}^{\dagger}_{r\alpha}. \tag{13}$$

Using Eqs. (11)–(13) and $\hat{U}|\text{vac}\rangle = |\text{vac}\rangle$ the $\sigma_2$ contribution may be expressed as

$$\begin{aligned}\sigma_2(I_\alpha,I_\beta) &= \sum_{J_\alpha}\sum_{kl}\langle\alpha(J_\alpha)|\hat{U}^{\dagger}\hat{U}\hat{E}^{\alpha}_{kl}\hat{U}^{\dagger}\hat{U}|\alpha(I_\alpha)\rangle \\ &\times\left[h_{kl} - \frac{1}{2}\sum_j (kj|jl)\right]C(J_\alpha,I_\beta) \\ &+\frac{1}{2}\sum_{J_\alpha}\sum_{ijkl}\langle\alpha(J_\alpha)|\hat{U}^{\dagger}\hat{U}\hat{E}^{\alpha}_{ij}\hat{U}^{\dagger}\hat{U}\hat{E}^{\alpha}_{kl}\hat{U}^{\dagger}\hat{U}| \\ &\times\alpha(I_\alpha)\rangle(ij|kl)C(J_\alpha,I_\beta) \\ &= \sum_{J_\alpha}\sum_{kl}\langle\beta(J_\alpha)|\hat{E}^{\beta}_{kl}|\beta(I_\alpha)\rangle \\ &\times\left[h_{kl} - \frac{1}{2}\sum_j (kj|jl)\right]C(J_\alpha,I_\beta) \\ &+\sum_{J_\alpha}\sum_{ijkl}\langle\beta(J_\alpha)|\hat{E}^{\beta}_{ij}\hat{E}^{\beta}_{kl}|\beta(I_\alpha)\rangle \\ &\times(ij|kl)C(J_\alpha,I_\beta).\end{aligned} \tag{14}$$

The $\sigma_2$ contribution to $\sigma$ can thus be expressed in terms of the $\sigma_1$ contribution, Eq. (9a), using Eq. (10):

$$\sigma_2(I_\alpha,I_\beta) = (-1)^S\sigma_1(I_\beta,I_\alpha). \tag{15}$$

In the same vein it may be shown that the $ijkl$ th component of the $\sigma_3$ satisfies

$$\sigma_3^{ijkl}(I_\alpha,I_\beta) = (-1)^S\sigma_3^{klij}(I_\beta,I_\alpha). \tag{16}$$

Equation (16) may be used to impose either the restriction $I_\alpha \geqslant I_\beta$ or $(ij)\geqslant(kl)$ when evaluating the $\sigma_3$ contribution. The optimal choice of these two alternatives depends on the problem in question. The choice $I_\alpha \geqslant I_\beta$ is optimal for eliminating certain subblocks of $\sigma$ from being calculated, e.g., in the case when $C$ is an expansion for a state which is not totally symmetric in the given pointgroup. For diagonal subblocks where the constraint $I_\alpha \geqslant I_\beta$ only can be used to eliminate the upper half of the subblock, the constraint $(ij)\geqslant(kl)$ is preferable, since it does not reduce the average vectorlength. To exploit the constraint $(ij)\geqslant(kl)$ we write

$$\begin{aligned}\sigma_3(I_\alpha,I_\beta) &= \sum_{ijkl}\sigma_3^{ijkl}(I_\alpha,I_\beta) \\ &= \sum_{(ij)\geqslant(kl)}\sigma_3^{ijkl}(I_\alpha,I_\beta) + \sum_{(ij)<(kl)}(-1)^S \\ &\times[\sigma_3^{klij}(I_\beta,I_\alpha)] \\ &= \sigma_3'(I_\alpha,I_\beta) + (-1)^S\sigma_3'(I_\beta,I_\alpha)\end{aligned} \tag{17}$$

with

$$\sigma_3'(I_\alpha,I_\beta) = \sum_{(ij)\geqslant(kl)}\sigma_3^{ijkl}(I_\alpha,I_\beta)(1+\delta_{(ij)(kl)})^{-1}. \tag{18}$$

In Eqs. (17) and (18) we have introduced an ordering $(ij)$ of orbital pair $(ij)$. Using Eqs. (15) and (17) the $\sigma$ vector becomes

$$\sigma(I_\alpha,I_\beta) = \sigma_1(I_\alpha,I_\beta) + \sigma_3'(I_\alpha,I_\beta) + (-1)^S[\sigma_1(I_\beta,I_\alpha) + \sigma_3'(I_\beta,I_\alpha)]. \quad (19)$$

The operation count is thus reduced roughly by a factor of 2 for the $M_S = 0$ component of a state vector at the expense of a matrix transposition.

## IV. FULL CI ALGORITHM

We now consider the calculation of the $\sigma$ vector for a FCI configuration space. A vectorizable algorithm for setting up the $\sigma_1$ vector in Eq. (9a) can be obtained using that the involved Hamiltonian matrix elements are independent of $I_\alpha$. The algorithm is sketched in Eq. (20).

Loop over $I_\beta$
  Initialize an array $F(J_\beta)$ to zero
  Loop over excitations $\hat{E}^\beta_{kl}$ from $|\beta(I_\beta)\rangle$

$$|\beta(K_\beta)\rangle = \mathrm{sgn}(kl)\hat{E}^\beta_{kl}|\beta(I_\beta)\rangle$$

$$F(K_\beta) = F(K_\beta) + \mathrm{sgn}(kl)\left[h_{kl} - \frac{1}{2}\sum_j (kj|jl)\right]$$

  Loop over excitations $\hat{E}^\beta_{ij}$ from $|\beta(K_\beta)\rangle$

$$|\beta(J_\beta)\rangle = \mathrm{sgn}(ij)\hat{E}^\beta_{ij}|\beta(K_\beta)\rangle$$

$$F(J_\beta) = F(J_\beta) + 1/2\ \mathrm{sgn}(kl)\mathrm{sgn}(ij)(ij)|kl)$$

  end of loop over $\hat{E}^\beta_{ij}$
end of loop over $\hat{E}^\beta_{kl}$

$$\sigma_1(I_\alpha,I_\beta) = \sum_{J_\beta} F(J_\beta)C(I_\alpha,J_\beta);\ \text{vectorized over } I_\alpha$$

end of loop over $I_\beta$. (20)

The construction of $F$ allows summation of all terms combining two strings. The time consuming step is the multiplication of the sparse vector $F$ on the matrix $C$ in the last step in Eq. (20). The number of multiplications involved in the last step is

$$N_1 = N_{\mathrm{det}}\{1/4n_\beta(n_\beta-1)(N-n_\beta)(N-n_\beta-1) + n_\beta(N-n_\beta)+1\} \simeq N_{\mathrm{det}}\frac{n_\beta^2(N-n_\beta)^2}{4}, \quad (21)$$

where we have used that the number of nonvanishing elements in $F$ is equal to the number of strings which differ from $\beta(I_\beta)$ in two, one, or zero spin orbitals. $N_{\mathrm{det}}$ is the number of determinants, $N$ the total number of orbitals, and $n_\beta$ is the number of beta electrons.

The second part of the $\sigma$ vector, $\sigma_2$ of Eq. (9b) can be constructed with an algorithm similar to the algorithm in Eq. (20). We have found it convenient to transpose the $C$ vector, as vectorization then can be performed on vectors with stride 1. The time-defining part in the construction of $\sigma_2$ has the operation count

$$N_2 = N_{\mathrm{det}}\{1/4n_\alpha(n_\alpha-1)(N-n_\alpha)(N-n_\alpha-1) + n_\alpha(N-n_\alpha)+1\} \simeq N_{\mathrm{det}}\frac{n_\alpha^2(N-n_\alpha)^2}{4}. \quad (22)$$

The third part of the $\sigma$ vector, $\sigma_3$ of Eq. (9c), is less trivial to vectorize. A gather operation and a scatter operation can be introduced to avoid use of indirect addressing in the time consuming part of the algorithm. An algorithm for the construction of $\sigma_3$ is given in Eq. (23).

Loop over $kl$
  set up $L(I)$, $R(I)$, and sgn(I) defined by
  $|\alpha[L(I)]\rangle = \hat{E}^\alpha_{kl}|\alpha[R(I)]\rangle\mathrm{sgn}(I)$
  $C'(I,J_\beta) = C[L(I),J_\beta]\mathrm{sgn}(I)$; vectorized gathering
  loop over $I_\beta$
    loop over excitations $\hat{E}^\beta_{ij}$ from $|\beta(I_\beta)\rangle$
      $|\beta(J_\beta)\rangle = \mathrm{sgn}(ij)\hat{E}^\beta_{ij}|\beta(I_\beta)\rangle$
      $F(J_\beta) = F(J_\beta) + \mathrm{sgn}(ij)(ij|kl)$
    end of loop over $\hat{E}^\beta_{ij}$
    $V(I) = \sum_{J_\beta} F(J_\beta)C'(I,J_\beta)$; vectorized over $I$
    $\sigma_3[R(I),I_\beta] = \sigma_3[R(I),I_\beta] + V(I)$; vectorized scattering
  end of loop over $I_\beta$
end of loop over $kl$. (23)

The construction of $V(I)$ from $F$ and $C'$ is usually the most operation intensive part of the construction of $\sigma_3$. The operation count for this part is

$$N_3 = N_{\mathrm{det}}n_\alpha(N-n_\alpha+1)[n_\beta(N-n_\beta)+1] \simeq N_{\mathrm{det}}n_\alpha n_\beta(N-n_\alpha)(N-n_\beta). \quad (24)$$

The gathering and scattering operations both involve $N_{\mathrm{det}}n_\alpha(N-n_\alpha)$ operations.

In the case that the integrals possess full eightfold permutational symmetry it is advantageous to restrict the loop over $kl$ in Eq. (23) to $k \geqslant l$. This is accomplished by using $(\hat{E}^\alpha_{kl} + \hat{E}^\alpha_{lk})(1+\delta_{kl})^{-1}$ instead of $\hat{E}^\alpha_{kl}$ to set up $|\alpha[L(I)]\rangle$, $|\alpha[R(I)]\rangle$ and sgn$(I)$. The operation count $N_3$ is not affected but vector efficiency is increased by doubling the lengths of $L(I)$ and $R(I)$.

For the case of $M_S = 0$ components the construction of $\sigma_2$ can be eliminated and the construction of $\sigma_3$ can be replaced with the construction of $\sigma_3'$ as discussed in Sec. III. In order to calculate $\sigma_3'$ the algorithm in Eq. (23) is slightly modified; only excitations $\hat{E}^\beta_{ij}$ from $|\beta(J_\beta)\rangle$ with $(ij) \geqslant (kl)$ are included, and integrals $(ij|ij)$ are divided by two.

The present algorithm is in many ways more in line with an earlier algorithm of Handy[5] than the later algorithms of Siegbahn[2] and Knowles and Handy.[3] The major progress, compared to Handy's earlier algorithm, is our efficient calculation of $\sigma_3$, Eq. (23).

The most computer intensive part of the Knowles–Handy algorithm is the multiplication of an $N_{\mathrm{det}} \times (N^2/2)$

matrix with the $(N^2/2)\times(N^2/2)$ integral list. The operation count $N^{KH}$ arising from this matrix multiplication is

$$N^{KH} = N_{det} N^4/4. \tag{25}$$

For the case $n_\alpha = n_\beta$ the described algorithm has an operation count

$$N = N_1 + N_2 + N_3 \simeq N_{det}\, 3/2 n_\alpha^2 (N - n_\alpha)^2. \tag{26}$$

When $n_\alpha = n_\beta = N/2$ one has a ratio $N^{KH}$:$N$ of about 5:2. When the total number of orbitals $N$ increases for a fixed number of electrons this ratio increases: for $n_\alpha = n_\beta = 4$, $N = 26$, as in the recent FCI calculations by Bauschlicher *et al.*,[11] a ratio of about 10:1 is obtained.

In order to transform the operation counts into indicators of relative speed on a vector computer the obtainable megaflop rates must be estimated . The matrix–matrix multiplication of the Knowles–Handy algorithm allows all vector machines to spin at maximal speed. The inner loops used in constructing $\sigma_1$ and $\sigma_2$ can be performed at similar megaflop rates, but the construction of $\sigma_3$ will often be less efficient, since only a few vectors enter in the multiplication in the rate determining step because $F(J_\beta)$ in Eq. (23) has fewer nonvanishing elements than in Eq. (20).

For singlet wave functions Knowles and Handy suggest the use of combinations $\alpha(I_\alpha)\beta(I_\beta) + \alpha(I_\beta)\beta(I_\alpha)$ to reduce the matrix–matrix multiplication step. They report that the other parts of the algorithm become more time consuming under such circumstances. With the present algorithm only a matrix transposition and the addition of two matrices are required to reduce the calculational effort in the inner loops by a factor of 2 in the case of $M_S = 0$ symmetry.

## V. RESTRICTED ACTIVE SPACE CI ALGORITHM

A CI space for a restricted active space (RAS) can be obtained by dividing the orbital set for the correlating electrons into three subsets: I, II, and III and requiring a minimal number of occupied spin orbitals $N_1$ in space I and a maximal number of occupied spin orbitals $N_3$ in space III. The number of electrons in space II is unrestricted, and this space is a reminiscent of the normal active space of CAS. The RAS space includes many usual CI spaces as special cases, e.g., singles, doubles, triples, etc., from an HF state; and all singles and doubles from a CAS space (SOCI). Another useful RAS structure consists of all single and double excitations in a large orbital space, augmented with triple and quadruple excitations in a smaller orbital space.

The strings of a RAS space are in general obtained from several restricted graphs. Since all strings of the same graph are stored consecutively, the restrictions on the allowed combinations of alpha- and betastrings compared to the full $C(I_\alpha, I_\beta)$ matrix correspond to restricting the $C$ matrix to certain subblocks. Usual methods for symmetry blocking of the matrices are easily generalized to handle this additional blocking.

A feature of the RAS CI space is that intermediate excitations to strings not included in the RAS space can be avoided in the calculation of the $\sigma$ vector [Eq. (9)]. In the $\sigma_3$ contribution excitations out of the restricted CI space can never contribute. For $\sigma_1$ (and analogous for $\sigma_2$) some of the intermediate strings $|\beta(K_\beta)\rangle$, which are constructed in the fourth step of Eq. (20), could be outside the restricted CI space. All contributions from out-of-space excitations can be removed by using the particle symmetry $(ij|kl) = (kl|ij)$ and the ordering

$$(ij) = (i-1)N + j \tag{27}$$

of the excitations $\hat{E}_{ij}$. The order of the orbitals is first orbitals of set I, then orbitals of set II, and finally orbitals of set III. The $\sigma_1$ contribution in Eq.(9a) can then be rewritten as

$$\begin{aligned}\sigma_1(I_\alpha, I_\beta) = &\sum_{J_\beta}\sum_{kl}\langle\beta(J_\beta)|\hat{E}^\beta_{kl}|\beta(I_\beta)\rangle g_{kl} C(I_\alpha, J_\beta) \\ &+ \sum_{J_\beta}\sum_{(ij)>(kl)}\langle\beta(J_\beta)|\hat{E}^\beta_{ij}\hat{E}^\beta_{kl}|\beta(I_\beta)\rangle \\ &\times (ij|kl)(1+\delta_{(ij)(kl)})^{-1} C(I_\alpha, J_\beta),\end{aligned} \tag{28}$$

where $g$ is an asymmetric matrix

$$\begin{aligned}g_{kl} &= h_{kl} - \sum_{j<k}(kj|jl) - (kk|kl)(1+\delta_{kl})^{-1} \quad \text{for } k \geqslant l \\ g_{kl} &= h_{kl} - \sum_{j<k}(kj|jl) \quad \text{for } k < l.\end{aligned} \tag{29}$$

Let us now consider an out-of-space excitation $\hat{E}^\beta_{kl}|\beta(I_\beta)\rangle$. From Eq. (28) we know that this excitation has a nonvanishing contribution if there exists an $\hat{E}^\beta_{ij}$, where $(ij) \geqslant (kl)$, so that $\hat{E}^\beta_{ij}\hat{E}^\beta_{kl}|\beta(I_\beta)\rangle$ is back in the allowed space. An out-of-space excitation $\hat{E}^\beta_{kl}|\beta(I_\beta)\rangle$ can arise if $\hat{E}^\beta_{kl}$ excites an electron out from RAS I or if $\hat{E}^\beta_{kl}$ excites one electron into RAS III. If the excitation $\hat{E}^\beta_{kl}$ annihilates an electron in RAS I any operator $\hat{E}^\beta_{ij}$ that adds an electron to RAS I will have $k > i$ so $(kl) > (ij)$. This term does not occur in Eq. (28) since the summation is restricted to $(ij) \geqslant (kl)$. Along the same line one eliminates the possibility of the other out-of-space excitations.

We now briefly discuss the construction of the $\sigma_1$, $\sigma_2$, and $\sigma_3$ components of the $\sigma$ vector in Eq. (8) for a RAS space. The construction of $\sigma_1$ can be performed with a slightly modified version of the algorithm in Eq. (20). The only major difference is that the construction of $\sigma_1$ from $F$ and $C$ in the last step of Eq. (20) is modified, so that the inner loop over alphastrings is split up into two loops; one over graphs of alphastrings and one over alphastrings of this graph. This allows identification of the types of betastrings to be included in the sum over elements in $F$. Since the strings of the same type are grouped together, the innermost loop is still a matrix-vector operation. The constructions of the $\sigma_2$ and $\sigma_3$ vectors must be modified in a similar fashion.

In the strict sense of obtained MFLOPs, a CAS iteration may in some cases be more efficient than a RAS iteration. This is caused by less efficient use of vector facilities in RAS, since some types of strings may have only a few members. For example, in a calculation with singles and doubles out of a HF reference state the type of space defining the reference state has just one string. For excitations out from a large reference space there will usually be enough strings of each type to support efficient vectorization. For $M_S = 0$ states most blocks that do not use vector facilities efficiently can be eliminated.

TABLE I. Timings for FCI iterations on Ne correlating ten electrons in a varying number of active orbitals. The length of the Slater-determinant expansions and the CPU seconds for the construction of a $\sigma$ vector are given as a function of the size of the active space.

| Number of active orbitals | Number of Slater determinants | Number of operations in inner loops | CPU time (s) | Energy |
|---|---|---|---|---|
| 8 | 400 | $3.3\times10^4$ | 0.14 | − 128.612 765 |
| 9 | 2 076 | $2.8\times10^5$ | 0.49 | − 128.654 459 |
| 10 | 8 072 | $1.4\times10^6$ | 1.2 | ··· |
| 11 | 26 804 | $5.8\times10^6$ | 3.0 | ··· |
| 12 | 78 432 | $2.1\times10^7$ | 7.1 | ··· |
| 13 | 207 189 | $6.3\times10^7$ | 17 | ··· |
| 14 | 501 992 | $2.3\times10^8$ | 36 | − 128.728 455 |
| 15 | 1 129 089 | $6.2\times10^8$ | 82 | ··· |
| 16 | 2 387 458 | $1.6\times10^9$ | 171 | ··· |

## VI. SAMPLE CALCULATIONS

The proposed algorithm has been implemented on an IBM 3090/150(VF) computer. The whole program has been written in FORTRAN 77, without any use of assembler routines. Separate routines have been written for the FCI case. We first report FCI calculations on Ne, using a $5s3p1d$ basis.[9] The ordering of the orbitals was $1s2s2p3p3s3d\,4p4s5s$, and the orbitals were chosen as the natural orbitals of an all singles and doubles CI out from the HF state with the $1s$ orbital kept doubly occupied. All 10 electrons were correlated, keeping 10 to 16 orbitals active. The number of determinants, operation counts, and the CPU seconds for the construction of the $\sigma$ vectors are given in Table I. For those calculations in Table I that gives a pure special representation of the $^1S$ ground state we also report energies. An important finding is that the timings are approximately proportional to the number of determinants. The number of operations, Eqs. (21) and (24), increases faster than the number of determinants, but this rise is almost entirely offset by increased vector length.

The largest calculation in Table I, 10 electrons in 16 orbitals, results in 2.39 million Slater determinants and took 171 s per CI iteration. The construction of $\sigma_1$ required around 23 s for about $361\times10^6$ operations in the innermost loop. On the average 75 elements contributed to each element in $\sigma_1$, and the average vector length was 546. For $\sigma_3$ around 74 s were used to perform $1199\times10^6$ operations in the innermost loop while another 74 s were used for the gathering, scattering, and the construction of the $F$ vector.

Siegbahn[2] and Knowles and Handy[3] have both reported FCI calculations with 10 electrons distributed in 11 orbitals, of which 8 orbitals are of symmetry $a_g$ and 3 are of symmetry $a_u$. The iteration time in the present algorithm for such a calculation is 7.9 s, while Siegbahn reported around 20 s and Knowles and Handy reported 8.4 s for one iteration (the latter calculations were carried out on CRAY-1S computers).

As a final FCI calculation we redid the 1983 benchmark calculation of Harrison and Handy[10] on $H_2O$ in a DZ basis yielding 1 002 708 determinants. The construction of one $\sigma$ vector required 75 s.

For testing and timing of the RAS procedure we report Ne calculations with the same basis as in the FCI calculations. The $1s$ orbital was frozen in all the calculations and the RAS I space was empty. In the calculations in Table II the size of RAS II and the number of electrons in RAS III are varied. In Table II we report the number of determinants and the iteration time for the different choices of RAS II and RAS III spaces. Each row represents a fixed number of orbitals in RAS II, and each column refers to a maximum on the number of electrons in RAS III. The cases with no electrons in RAS III, $n_{III} = 0$, give FCI calculations in the RAS II orbital space. The first row in Table II (four orbitals in RAS II) gives excitations through level $n_{III}$ out from the single reference determinant. Each graph was defined by a given occupation in RAS III. Although the program can recognize the $n_{III} = 8$ case as FCI, it was forced to work with five graphs. Many of the following timings are therefore not optimal, since the number of graphs often were longer than necessary.

The full CI calculation obtained by expanding RAS II to all 18 orbitals requires 140 s per iteration. A similar iteration

TABLE II. RAS calculations on Ne where the $1s$ orbital is doubly occupied and eight electrons are distributed in an RAS II space containing 4–18 orbitals and an RAS III space with a maximum of 0–8 electrons. The upper number in each table entrance gives the number of Slater determinants and the lower number the CPU second per iteration for the considered RAS calculation.

| | Maximum number of electrons in RAS III | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| Orbitals in RAS II | 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 |
| 4 | | 23 | 615 | 8 623 | 66 689 | 279 807 | 681 599 | 1 046 251 | 1 172 112 |
| | | 0.07 | 2.23 | 10.31 | 34.33 | 53.21 | 126.39 | 147.53 | 172.14 |
| 7 | 169 | 3 683 | 32 562 | 153 382 | 435 005 | 801 123 | 1 062 896 | 1 158 244 | |
| | 0.23 | 2.75 | 18.66 | 44.44 | 91.83 | 128.01 | 183.16 | 199.05 | ··· |
| 8 | 676 | 10 556 | 71 868 | 265 132 | 602 444 | 934 164 | 1 115 876 | 1 166 500 | |
| | 0.44 | 4.83 | 38.00 | 53.77 | 114.43 | 152.32 | 198.62 | 220.25 | ··· |
| 13 | 64 331 | 320 551 | 715 863 | 1 017 847 | 1 141 707 | 1 168 809 | 1 171 921 | 1 172 105 | |
| | 9.92 | 41.69 | 108.42 | 153.11 | 199.30 | 213.05 | 218.91 | 219.36 | ··· |
| 16 | 415 306 | 925 754 | 1 136 838 | 1 170 246 | | | | | |
| | 48.68 | 115.57 | 161.45 | 173.43 | ··· | ··· | ··· | ··· | ··· |
| 18 | 1 172 112 | | | | | | | | |
| | 139.85 | | | | | | | | |

TABLE III. The difference in total energy between the RAS calculations reported in Table II and a FCI calculation where the 1*s* orbital is doubly occupied and eight electrons are distributed in 18 orbitals.

| Orbitals in RAS II | Maximum number of electrons in RAS III | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| | 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 |
| 2*s*2*p* | 0.219 228 | 0.218 373 | 0.007 752 | 0.005 186 | 0.000 178 | 0.000 042 | 0.000 001 | $<10^{-6}$ | FCI |
| 2*s*2*p*3*p* | 0.130 154 | 0.089 336 | 0.001 103 | 0.000 490 | 0.000 003 | 0.000 001 | $<10^{-6}$ | $<10^{-6}$ | ... |
| 2*s*2*p*3*p*3*s* | 0.089 566 | 0.079 616 | 0.000 604 | 0.000 358 | 0.000 001 | $<10^{-6}$ | $<10^{-6}$ | $<10^{-6}$ | ... |
| 2*s*2*p*3*p*3*s* 3*d* | 0.017 038 | 0.008 437 | 0.000 013 | 0.000 007 | $<10^{-6}$ | $<10^{-6}$ | $<10^{-6}$ | $<10^{-6}$ | ... |
| 2*s*2*p*3*p*3*s* 3*d* 4*p* | 0.003 615 | 0.000 49 | $<10^{-6}$ | $<10^{-6}$ | ... | ... | ... | ... | ... |
| 2*s*2*p*3*p*3*s* 3*d* 4*p*4*s*5*s* | FCI | ... | ... | ... | ... | ... | ... | ... | ... |

with the FCI code takes 98 s. This difference in timings is caused by different ways of treating the symmetry for $M_S = 0$. In the FCI code the restriction $(ij) \geqslant (kl)$ is used in the construction of $\sigma_3$, while the RAS code uses the $M_S = 0$ symmetry to eliminate blocks above the diagonal of the $C$ vector. [See discussion following Eq. (16).] The FCI case is described by a single graph and, therefore, has no blocks above the diagonal. The number of operations in the inner loop for the construction of $\sigma_3$, is thereby roughly doubled compared to the FCI code where the restriction $(ij) \geqslant (kl)$ is active. The FCI result is also recovered for the case with zero through octuple excitations from a single determinant (four orbitals in RAS II, up to eight electrons in RAS III). A full CI iteration then required 172 s. The $M_S = 0$ symmetry is better exploited in this case, but this gain is more than offset by the separation of strings into five graphs (0,1,2,3,4 electrons in RAS III) due to the increased overhead and reduced vector lengths.

For a single reference state (first row in Table II) the timings are satisfactory for the cases where more than two electrons are distributed in RAS III. The use of the $M_S = 0$ symmetry to eliminate all blocks above the block diagonal in $\sigma_3$ often halved the iteration times compared to the approach where the restriction $(ij) \geqslant (kl)$ was used. This gain was due to the elimination of blocks with a small number of rows. These blocks were otherwise time consuming, due to large overhead and small vector lengths. The SDT and SDTQ cases with a single reference are too small to judge the efficiency of the program on a vector computer. For the calculations with at least five electrons in RAS III and a single reference state, the efficiency is comparable to the efficiency obtained in the FCI calculation.

The timings for the multireference calculations demonstrate the feasibility of carrying out this type of calculations. The first four rows of the columns which have 1, 2, 3, or 4 electrons in RAS III give large increases in the number of Slater determinants and more moderate increases in the iteration times. The slower rise in the iteration time is due to the slow rise of overhead and increased efficiency of vector operations. The calculations with three or four electrons in RAS III and the smaller reference spaces, 169 or 676 Slater determinants, are large enough to sustain efficient use of vector capabilities. With two electrons in RAS III, large reference spaces, 64 331 or 415 306 Slater determinants, are needed before the CI expansion has a size that allows efficient use of vector capabilities.

The lower right corner of Table II contains iteration times which are significantly larger than the two FCI iteration times also given in Table II. This is due to the overhead connected with having five graphs and the presence of large diagonal blocks, where the $M_S = 0$ symmetry is not fully exploited.

Finally, we compare the difference in total energy between the RAS calculations in Table II and the corresponding FCI in Table III. Comparing row 1, where RAS II consists of 2*s*2*p*, and row 3, where RAS II consists of 2*s*2*p*3*p*3*s*, we notice that the energy error for the calculations allowing 2 electrons in RAS III diminishes much more rapidly than for the calculations restricting all eight electrons to RAS II. Similar trends are also found at other places.

The importance of different excitation levels out from general reference spaces differs from the behavior for a single reference state. with a RAS II consisting of 2*s*2*p*3*p*, triple excitations into RAS III are slightly more important than the quadruple excitations into RAS III (0.000 693 a.u. vs 0.000 487 a.u.). For a RAS II space consisting of 2*s*2*p*3*p*3*s*3*d* the contribution from triple excitations is 0.000 246 a.u. while the contribution from quadruple excitations is 0.000 487 a.u. The importance of Slater determinants with one electron in RAS III is also noticable; in the calculation with 2*s*2*p*3*p*3*s*3*d* as RAS II, the Slater determinants with one electron in RAS III accounts for 50% of the energy that is not contained in the CAS calculation in RAS II.

## VII. DISCUSSION

We have introduced a restricted active space (RAS) wave function and demonstrated how direct CI iterations may efficiently be carried out with this wave function.

The RAS wave function is obtained by dividing the active orbital space into three subsets I, II, and III, requiring a minimal number of electrons occupying spin orbitals in space I and a maximal number of electrons occupying spin orbitals in space III. The number of electrons in space II is unrestricted and is a reminiscent of the normal active space of CAS.

The direct CI algorithm is obtained using the technique of Handy of splitting the determinants into alphastrings and

betastrings. The two electron contribution is separated into three terms resulting in a highly vectorizable algorithm which treats the sparsity of the Hamiltonian matrix in an efficient manner.

A full CI configuration space (FCI) is obtained as a special case of a RAS wave function. The performance of the present algorithm has been compared to FCI procedures recently described in the literature. Similar or improved performance has been demonstrated, depending on the ratio between the number of active orbitals and the number of electrons in the active space.

In the RAS algorithm the $C$ and $\sigma$ vectors are divided into subblocks, where each subblock only contains alphastrings and betastrings belonging to single graphs. The algorithm is easily modified to have only single subblocks of $C$ and $\sigma$ residing in main memory simultaneously. The expense for this is increased $I/O$ due to several reads of $C$ and $\sigma$. By defining an FCI space as a RAS with several graphs of strings, $C$ and $\sigma$ can be blocked and this makes it possible to treat very large CI expansions on computers with limited main memory.

The algorithm may straightforwardly be used to carry out direct CI iterations also when the integrals do not possess eightfold permutational symmetry. This is important for example in MCSCF linear response calculations and in MCSCF calculations with complex orbitals. In both cases linear transformations are required where the two-electron integrals only possess the exchange symmetry between particle one and particle two. The operators involved do not need to be total symmetric in the given point group. This is important in, e.g., calculations of the CI Hessian.

## ACKNOWLEDGMENT

The present study has been supported by a grant from the Swedish Natural Science Research (NFR).

[1] B. O. Roos, Adv. Chem. Phys. **LXIX**, 399 (1987), and references therein.
[2] P. E. M. Siegbahn, Chem. Phys. Lett. **109**, 417 (1984).
[3] P. J. Knowles and N. C. Handy. Chem. Phys. Lett. **111**, 315 (1984).
[4] I. Shavitt, in *Modern Theoretical Chemistry*, edited by H. F. Schaefer III (Plenum, New York, 1977), Vol. 3, p. 189.
[5] N. C. Handy, Chem. Phys. Lett. **74**, 280 (1980).
[6] W. Duch, *GRMS or Graphical Representation of Model Spaces* (Springer, Berlin, 1986).
[7] B. O. Roos, Chem. Phys. Lett. **15**, 153 (1972).
[8] J. Linderberg and Y. Öhrn, *Propagators in Quantum Chemistry* (Academic, London, 1973), p. 24.
[9] C. W. Bauschlicher, Jr., S. R. Langhoff, P. R. Taylor, and H. Partridge, Chem. Phys. Lett. **126**, 436 (1986).
[10] R. J. Harrison and N. C. Handy, Chem. Phys. Lett. **95**, 386 (1983).
[11] C. W. Bauschlicher, Jr. and S. R. Langhoff, J. Chem. Phys. **87**, 2919 (1987).