Highly correlated calculations with a polynomial cost algorithm: A study of the density matrix renormalization group

Garnet Kin-Lic Chan and Martin Head-Gordon

Citation: **116**, (2002); doi: 10.1063/1.1449459
View online: http://dx.doi.org/10.1063/1.1449459
View Table of Contents: http://aip.scitation.org/toc/jcp/116/11
Published by the American Institute of Physics

Highly correlated calculations with a polynomial cost algorithm: A study of the density matrix renormalization group

Garnet Kin-Lic Chan[a] and Martin Head-Gordon[b]
Department of Chemistry, University of California, Berkeley, California 94720-1460

(Received 16 October 2001; accepted 18 December 2001)

We study the recently developed Density Matrix Renormalization Group (DMRG) algorithm in the context of quantum chemistry. In contrast to traditional approaches, this algorithm is believed to yield arbitrarily high accuracy in the energy with only polynomial computational effort. We describe in some detail how this is achieved. We begin by introducing the principles of the renormalization procedure, and how one formulates an algorithm for use in quantum chemistry. The renormalization group algorithm is then interpreted in terms of familiar quantum chemical concepts, and its numerical behavior, including its convergence and computational cost, are studied using both model and real systems. The asymptotic convergence of the algorithm is derived. Finally, we examine the performance of the DMRG on widely studied chemical problems, such as the water molecule, the twisting barrier of ethene, and the dissociation of nitrogen. In all cases, the results compare favorably with the best existing quantum chemical methods, and particularly so when the nondynamical correlation is strong. Some perspectives for future development are given. © *2002 American Institute of Physics.* [DOI: 10.1063/1.1449459]

I. INTRODUCTION

In recent years, the Renormalization Group (RG) has become an important concept in many areas of physics.[1,2] The goal of the approach, to treat correlations beyond mean-field theory, is essentially the same as the problem of capturing correlations in quantum chemistry beyond the Hartree–Fock approximation. With the development by White[3,4] of the Density Matrix Renormalization Group (DMRG), which is a RG method particularly suited to numerical computation, RG approaches are now the methods of choice when studying low-dimensional quantum lattice systems.

In a number of interesting recent papers, Fano *et al.*[5] and White *et al.*[6,7] showed how the DMRG may be applied to quantum chemical Hamiltonians. (Since this article was submitted, another study of the DMRG as applied to quantum chemistry, with a different emphasis from our own, has also appeared.[8]) In principle, the approach is extremely powerful, as it allows one to approximate the full configuration interaction solution as accurately as one desires, with a reasonable cost. In particular, it appears that arbitrarily high accuracy can be obtained with only *polynomial* computational effort. This contrasts strongly with the conventional quantum chemical hierarchy, where successive excitations in a method such as coupled cluster or configuration interaction (CI) theory, lead to exponentially greater cost in both computation and memory. It also differs greatly from selected configuration interaction,[9] as it does not involve expansions in Slater determinants, but rather in an *adaptive* many-body basis. However, despite the apparent potential of the method, it is still relatively untested in chemical problems. The systems studied in Refs. 6, 7 are not sufficient to understand the power or limitations of this approach, especially in relation to existing quantum chemical methods.

In this work, we attempt to understand in detail, the DMRG as applied to quantum chemistry. Our goals are to understand how the algorithm may be efficiently implemented, how the algorithm works from the viewpoint of quantum chemistry, its performance (in terms of cost and accuracy) over a range of dynamically and nondynamically correlated systems, and its general strengths and weaknesses.

In Sec. II, after a brief introduction to renormalization group ideas, we describe our version of the DMRG algorithm. In particular, we have taken care to describe the algorithm in greater detail than is usual, because we encourage others to implement these ideas, and because we have found that presentations elsewhere are often very brief. We tackle concerns of efficiency, and algorithmic details, such as an optimal orbital ordering, which we have not found addressed elsewhere.

In the first part of Sec. III, we analyze the DMRG in terms of concepts from quantum chemistry. We examine how the DMRG incorporates N-fold excitations into the wave function, thus approximating the full CI. We discuss in a general way the relative merits of RG approaches, mean-field theory, and perturbation theory. We also show why we can expect to capture a good energy (but not a good wave function) with a polynomial cost algorithm such as the DMRG, by considering the reduced density matrix.

In the second part of Sec. III, we analyze the numerical behavior of the DMRG. We begin with a study of the accuracy of the DMRG. Crucial to the success of the method is the rate of the convergence of the energy to the full CI energy, as the number of DMRG states M is increased. It has been previously believed[6,10] that the error decreases exponentially fast with M, as $\exp-\kappa M$. By contrast, here we

[a]Electronic mail: garnet@bastille.cchem.berkeley.edu
[b]Electronic mail: mhg@bastille.cchem.berkeley.edu

show that the rate of convergence is actually described by $\exp-\kappa(\ln M)^2$, which is still faster than algebraic. We also briefly study the dependence of the DMRG on the correlation length of the system in a Hückel model, and hydrogen chains, with its implications for large quantum chemical calculations. We verify the polynomial cost of the algorithm, which is $O(M^2k^4+M^3k^3)$, where k is the number of one-particle orbitals.

In Sec. IV, we present calculations on a number of benchmark chemical problems, namely water, twisted ethene, and nitrogen dissociation, and compare the performance of the DMRG to the best existing quantum chemical methods. In all cases, we obtain results that compare very favorably with CCSD(T) calculations. Nonetheless, we observe that the DMRG in its present formulation is not ideally suited to the treatment of systems with large amounts of dynamical correlation (i.e., with very large basis sets), but that the DMRG handles nondynamical correlation quite successfully.

We present conclusions in Sec. V. Our findings are presented in the context of future developments in quantum chemistry, and we discuss how the DMRG algorithm may itself be further improved.

II. THE DENSITY MATRIX RENORMALIZATION GROUP ALGORITHM

Systems for which there are two distinct length or time scales, are well treated through mean field theory. The classic example of this in chemistry is the Born–Oppenheimer approximation, where the nuclei can be considered to move in the effective field (the potential energy surface) generated by ''integrating out'' electronic motion.

Renormalization techniques are applicable when such a length scale separation does not exist. This can happen even when the microscopic mechanism of correlation is relatively short ranged, through a ''domino effect.'' When hierarchical fluctuations lead to a correlation length that diverges in this way, the system is called ''critical.'' Although most quantum chemical systems of interest are not critical systems (in a sense that is made more precise in Sec. III A), electron correlation is hierarchical in nature; correlation is only pairwise, but it gives rise to non-negligible three and higher-particle correlations via a domino effect.

When there is no decoupling between the different scale lengths in the problem, all the microscopic degrees of freedom of the system are coupled. In the context of electron correlation, this corresponds to the very large number of non-negligible Slater determinants arising from high-order excitations, which are necessary for the accurate description of strongly correlated systems.

Renormalization is a procedure which allows us to describe the thermodynamic properties of a system on a large scale, without an explicit consideration of all the coupled degrees of freedom. This is through the *renormalization transform*, which relates the description of a system at a length scale λ to that at a longer length $\lambda+\delta\lambda$ by integrating out the degrees of freedom in between. At each length scale, the variables used to describe the system depend in some complicated way on the variables at some shorter length scale. By progressively integrating to longer and longer length scales, one can recover a description of the system even when the number of microscopic variables is prohibitively large. The procedure of changing the length scale is known as *blocking*, and that of integrating out degrees of freedom is known as *decimation*. Each renormalization transform must involve a modification of the ''structure rule'' of the system (such as the Hamiltonian or partition function), so as to preserve the physical invariant of interest, e.g., the electronic energy.

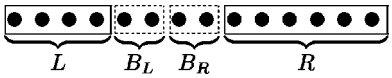

FIG. 1. A standard block configuration. B_L is to be blocked with L, and B_R with R.

A. Real-space renormalization and the DMRG

Renormalization algorithms fall broadly into two classes: real space and conjugate space. We shall not discuss renormalization in conjugate space further, despite its importance in other areas of physics.[1]

The real space renormalization transform, of which the DMRG is an example, is an algebraic transformation. In the following description, we shall use some terminology from the study of lattice systems. In quantum chemistry, the ''lattice'' is simply an ordering of orbitals on a line, with each orbital occupying a *site* on the lattice. We will use orbitals and sites more or less interchangeably, though in principle, a site can carry much more general states. As a caveat, the following description is *not* intended as a detailed renormalization algorithm as applied to a quantum chemical problem (which is covered in Sec. II B), but is intended rather to illustrate the basic principles of blocking and decimation.

Consider some such lattice of orbitals: each spin–orbital carries a set of states in Fock space, $|\mathrm{vac}\rangle$, $|\phi\rangle$. This can be viewed diagrammatically in Fig. 1, where we have denoted single sites by ●, and groups of many orbitals (or *blocks*) by boxes. We begin with the lattice partitioned into four blocks L, B_L, B_R, R, with corresponding Fock spaces F_L, F_{B_L}, F_{B_R}, F_R.

We work with the second-quantized Hamiltonian, which for quantum chemical systems is of the form,

$$H=\sum_{ij} t_{ij}a_i^\dagger a_j+\sum_{ijkl} v_{ijkl}a_i^\dagger a_j^\dagger a_k a_l. \tag{1}$$

For each block, we may write down Hamiltonians H_L, H_{B_L}, H_{B_R}, H_R in this form, where the orbital indices are restricted to the orbitals within the block. The Hamiltonian for a combined block, such as block $L'=LB_L$ may be written as

$$H_{L'}=H_L+H_{B_L}+\sum_{IJ} v_{IJ}I_{B_L}J_L, \tag{2}$$

where $\Sigma_{IJ}v_{IJ}I_LJ_{B_L}$ describes the interactions between blocks L and B_L. More explicitly, we find

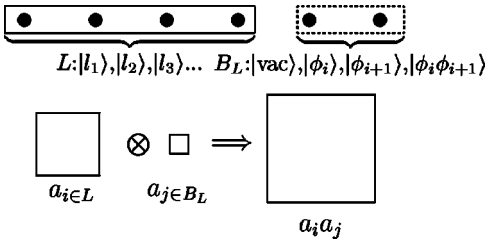

FIG. 2. Blocking. Here L, spanned by states $\{l\}$, is blocked with B_L, giving a new block spanned by states $\{l\}\otimes\{b_L\}$. Operator matrices (such as $a_i a_j$) are formed through direct products of operators on each of the blocks, viz., Eqs. (6) and (7).

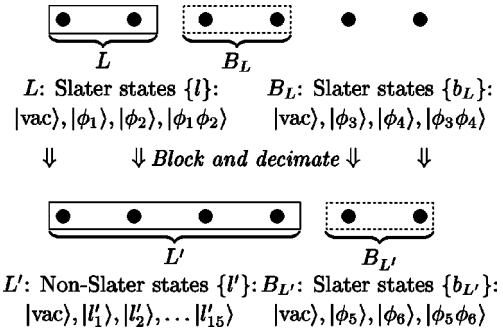

FIG. 3. The first renormalization transform. Initially L covers only two orbitals, and is spanned by only a few Slater states, which are known explicitly. The operators of L may thus be initialized. After the first renormalization transform, the states of L' are no longer determinants, and only the operators are known explicitly.

$$H_{L'}=H_{B_L}+H_L+\sum_{i\in B_L j\in L} t_{ij}(a_i^\dagger a_j+a_j^\dagger a_i)\cdots + \sum_{i\in B_L,jkl\in L} w_{ijkl}a_i^\dagger a_j^\dagger a_k a_l\cdots + \sum_{i\in L,jkl\in B_L} w_{ijkl}a_i^\dagger a_j^\dagger a_k a_l\cdots + \sum_{ij\in B_L,kl\in L} v_{ijkl}(a_i^\dagger a_j^\dagger a_k a_l+a_k^\dagger a_l^\dagger a_i a_j)\cdots + \sum_{ik\in B_L,jl\in L} x_{ijkl}a_i^\dagger a_j^\dagger a_k a_l, \quad (3)$$

$$w_{ijkl}=v_{ijkl}-v_{jikl}, \quad (4)$$

$$x_{ijkl}=v_{ijkl}-v_{jikl}-v_{jikl}+v_{jilk}. \quad (5)$$

Note that the structure of the interaction terms $I_{B_L}J_L$ is such that one operator is on B_L and three on L, or two on B_L and two on L, etc.

Let us now assume that blocks B_L and L are each described by a set of many-body states $\{b\}\in F_{B_L}$, $\{l\}\in F_L$, and that we have matrix representations of the operators H_{B_L}, H_L, I_{B_L}, J_L in this basis. Then the real space blocking procedure consists of constructing the representations of the operators $H_{L'}$, $I_{L'}$ of the combined block L', in the tensor-product space $\{l'\}\in F_{L'}=F_L\otimes F_{B_L}$. This is through simple direct products which for operators O_L, O_{B_L}, are (see also Fig. 2),

$$\langle l_1 b_1|O_L|l_2 b_2\rangle=\langle l_1|\langle b_1|O_L|b_2\rangle|l_2\rangle = \delta_{b_1 b_2}[O_L]_{l_1 l_2}p(b_1,O_L), \quad (6)$$

$$\langle l_1 b_1|O_L O_{B_L}|l_2 b_2\rangle=\langle l_1|\langle b_1|O_L O_{B_L}|b_2\rangle|l_2\rangle =[O_L]_{l_1 l_2}[O_{B_L}]_{b_1 b_2}p(b_1,O_L). \quad (7)$$

Here p a parity operator which gives the factor of 1 or -1 associated with the operation $\langle b_1|O_L\rightarrow O_L\langle b_1|$. It depends only on the number of particles in the state $\langle b_1|$, and the number of second-quantized operators in O_L; for example, if state $\langle b_1|$ contains an odd number of particles, then $\langle b_1|a_i=-a_i\langle b_1|$. The second relation (7) allows us to calculate the coupling matrix elements between blocks L and B_L, since the interaction is a sum of terms of the form $O_L O_{B_L}$, as in Eq. (2).

Note that to construct the operators of the combined block L', we do not need any details of the structure of the states of L and B_L, other than their particle number (which is necessary to evaluate p). Instead, each block is simply described by the matrix representations of its operators, and the quantum numbers of the states.

It may seem strange that we do not know the structure of the states (for example, their expansion in terms of Slater determinants). However, if one only asks the question: what information does one need to construct the operators of a larger block from those of the smaller block, one sees that no explicit wave function knowledge is needed, and rather, only the relevant operator matrices to use in Eqs. (6) and (7). For example, to construct $a_i a_j$, where $i\in L$, $j\in B_L$, one takes the direct product of the a_i and a_j matrices *assuming that we have been able to obtain these in some way*. Without going into further detail here, we note that to successively block $LB_L\rightarrow L'$, we need only to *start* with some L, for which all the operator matrices can be explicitly evaluated, and also to explicitly know the operator matrices of B_L at each stage. This is achieved, as shown in Fig. 3, by performing the first renormalization transform with L sufficiently small (i.e., only a few sites) so one can construct all the operators carried by L explicitly in the Fock space of all determinants spanned by L. Also, B_L is always kept small, carrying only a few states (for example, in Fig. 2, B_L carries states $|\text{vac}\rangle$, $|\phi_i\rangle|\phi_{i+1}\rangle$, and $|\phi_i\phi_{i+1}\rangle$), and thus we can always explicitly construct all the operators carried by B_L.

Blocking is followed by the decimation procedure which reduces the degrees of freedom of the combined block L', while approximately preserving physical invariants, such as the eigenvalue spectrum of the Hamiltonian. Clearly, the degrees of freedom that we wish to retain will depend on the phenomenon we wish to study at the end of the renormalization procedure. In the original real-space renormalization group which was used to study low energy phenomena,[2] Wilson proposed to diagonalize $H_{L'}$ in the space $F_{L'}$, and to select the *lowest M* eigenstates of the Hamiltonian of block L', namely, $\{|\mu\rangle=C^\mu_{l'}|l'\rangle:H_{L'}C^\mu_{l'}=\epsilon_\mu C^\mu_{l'},\mu=1,...,M\}$, as

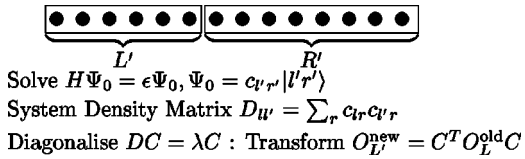

FIG. 4. The DMRG decimation. Here L' is the system block and R' is the environment.

the new representation of block L'. Consequently, the various operators of block L' are transformed via

$$O_{L'}^{\text{new}}=C^T O_{L'}^{\text{old}} C, \tag{8}$$

where C is the $M^2\times M$ matrix of coefficients of the lowest M eigenvectors of $H_{L'}$. The above multiplication reduces the $M^2\times M^2$ matrix $O_{L'}^{\text{old}}$ to an $M\times M$ matrix. This decimation reduces the number of states used to describe the superblock L' from M^2 tensor product states to M coarse-grained states relevant to low energy phenomena. The renormalization procedure may then be iterated down the chain to generate a coarse grained representation of the entire lattice (from which, for example, we may deduce an approximation to the lattice ground-state energy). With every decimation, the degrees of freedom are chosen to be relevant to the scale of the system, in contrast to configuration interaction, where one works with Slater determinants, which are parametrized by the one-particle degrees of freedom.

However, despite the early promise of the above real-space renormalization for quantum lattice problems, a naive application of the procedure does not usually work. As discussed in detail in the work of many authors,[4] the problem is with the decimation procedure, as the lowest energy eigenstates of the system block L' are not generally the best states to keep.[31] In particular, there may be significant interactions between the discarded states, and the rest of the lattice, i.e., the environment. Thus one should choose to retain states that in some way describe well the significant interactions between the system and the environment.

The question is, what is the best representation of L' to approximate the ground-state wave function Ψ_0 of the entire lattice $L'R'$? (see Fig. 4). $|\Psi_0\rangle$ may be approximated in the Hilbert space of the lattice, in the form $|\Psi\rangle=\Sigma c_{l'r'}|l'\rangle|r'\rangle$, where $|l'\rangle\in F_{L'}$ and $|r'\rangle\in F_R$. If we minimize the distance between the approximate and true wave functions, $\langle\Psi-\Psi_0|\Psi-\Psi_0\rangle$, it is simple to show that the "optimal" states of L' are the eigenfunctions θ_i of the projected density matrix $D_{L'}$ given by

$$D_{L'}=\mathrm{Tr}_{F_{B_R}}|\Psi_0\rangle\langle\Psi_0| \tag{9}$$

$$=\sum_{r'l_1'l_2'} c_{l_1'r'}c_{l_2'r'}|l_1'\rangle\langle l_2'| \tag{10}$$

$$=\sum_i w_i|\theta_i\rangle\langle\theta_i|. \tag{11}$$

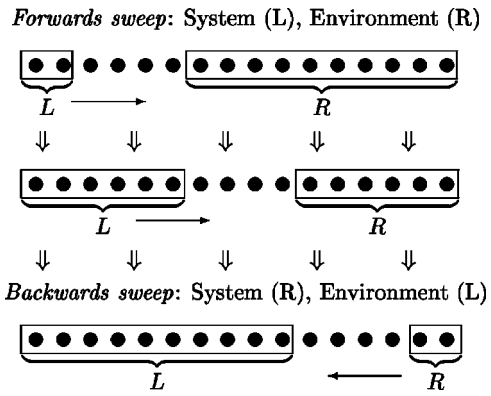

FIG. 5. The DMRG sweep algorithm. In the forwards sweep, the system block L is grown two sites at a time. In the backwards sweep, R becomes the system block.

The eigenfunctions θ_i form a set of expansion functions that yield most rapid convergence in the above metric sense, in the same way that the natural orbitals form a set of optimal expansion functions for full CI.

With this in mind, White and Noack proposed the following modified decimation procedure.[4] After one blocking operation $LB_L\rightarrow L'$, one performs an additional blocking, $L'R'$, where we assume we have some approximation representation of the environment, R', i.e., we have approximate operator matrices $H_{R'}, I_{R'}$. We may then form the Hamiltonian of the superblock $H_{L'R'}$, from which we obtain an approximate lattice ground-state wave function Ψ, subject to any necessary quantum constraints (such as total particle number and spin number constraints). This is projected onto the density matrix for block L' via Eq. (9), and the density matrix is diagonalized. The density matrix eigenfunctions with the M largest eigenvalues w_i are then selected to form the C matrix, and the operators of L' are transformed through Eq. (8). Because the M retained eigenfunctions are optimal, in the sense that they lead to the best approximate expansion of Ψ, this is an optimal choice of decimation. It is this form of decimation, which distinguishes the DMRG from other RG algorithms.

A further improvement to the renormalization algorithm was also introduced by White.[10] This consists of arranging successive renormalization transforms into a sweep algorithm, as shown in Fig. 5. After iteratively performing renormalization transforms down the lattice, one eventually reaches the end of the lattice, with the system block spanning the entire lattice and an environment block of negligible size. We call such a group of transformations, a sweep. Note that, as previously mentioned, we start each sweep with a system block that spans sufficiently few sites such that its Fock space (i.e., all possible determinants, for all particle numbers) can be *exactly represented* by M states. At the end of a sweep, the direction of the sweep is reversed, so that the system blocks of the *preceding sweep* become environment blocks and vice versa. In this way, the environment and system states may both be improved, until a self-consistent state is reached. Such a self-consistent state, which is the best representation of Ψ in the product form $\Psi=\Sigma_{l'r'}c_{l'r'}|l'r'\rangle$ for the M states (and the given orbital or-

dering of the lattice), may be considered the *fixed point* of the renormalization algorithm. Of course, a fixed point does not always exist, and in practice the algorithm sometimes enters a limit cycle, although we have found that the fluctuations in the energy are very small (below our convergence threshold), and decrease with increasing M. Note that the accuracy of the fixed point energy (i.e., the energy of the self-consistently determined Ψ) is determined by the number of states M kept at each transformation, and consequently increases as M increases.

The procedure of real-space renormalization using the eigenfunctions of the density matrix, and the combination with a self-consistent sweep procedure, constitutes the density matrix renormalization group (DMRG) algorithm.

B. A density matrix algorithm for electronic structure theory

The first application of the DMRG to a quantum chemical Hamiltonian was in the work of Fano *et al.*,[5] who studied Pariser–Parr–Pople[11] Hamiltonians in the context of cyclic polyenes. This was followed by the interesting papers of White *et al.*,[6,7] which showed how the DMRG may be efficiently formulated for *ab initio* Hamiltonians. The algorithm we have implemented is not different in the essentials from that described in Ref. 6; however, we will describe our algorithm in considerable detail, as such a presentation is lacking elsewhere.

As discussed in the previous section, the renormalization algorithm may be separated into two stages, blocking and decimation. The algorithm as a whole is summarized in pseudocode form, in Appendix A. We now proceed to give a systematic account of these steps.

1. The renormalization transform

The renormalization transform is applied to the lattice configuration $L \bullet \bullet \bullet \bullet R$ as shown in Fig. 1. As before, L and R are blocks spanned by M states, and $\bullet$ represents a single spin–orbital or site, spanning states $|\mathrm{vac}\rangle$, $|\phi\rangle$. We emphasize again, that the operator matrices carried by each $\bullet$ at each configuration, are immediately known because of their simple structure. The first two sites are blocked with the system $L \rightarrow L \bullet \bullet$ while the second two are blocked with the environment $R \rightarrow \bullet \bullet R$. The idea is to then use the enlarged environment $R' = \bullet \bullet R$ during the decimation phase. Some comment should be made regarding this choice of block configuration. When blocking L with a new block, it is advantageous for the new block to span only a small number of sites (and consequently a small number of states), to reduce the cost of the operations Eqs. (6) and (7). It is also advantageous to augment the environment states of R by a set of exactly described sites $\bullet \bullet$, that "strongly interact" with the system block $L\bullet$ (from the assumed short range nature of the Hamiltonian), since at the beginning of our iterative sweeps, our environment block is only very approximate. Consequently this leads to faster convergence of the iterative sweeps.

The main difficulty in applying the DMRG methodology to quantum chemical systems lies in the large number of terms in the electronic Hamiltonian (1), which is k^4, where k is the number of orbitals. With each blocking operation, we need $M \times M$ matrix representations of all the operators of the blocks, and these must be combined [as in Eqs. (6) and (7)] in many ways, and consequently both memory and computational costs are high. Several optimizations are therefore necessary. As first described by Xiang[12] in the momentum space formulation of the DMRG, an explicit consideration of all k^4 terms is not required if we restrict our interest to expectation values, such as the energy. In this case, many operators can be contracted leading to so-called *complementary operators*. For an arbitrary block combination AB, the combined Hamiltonian may be rewritten in terms of these complementary operators, as (see also White and Martin[6])

$$H_{AB} = H_A + H_B + \Delta_{AB}, \tag{12}$$

$$\Delta_{AB} = \left[\sum_{i \in A} a_i^\dagger S_i^B + \sum_{ij \in A} (a_i^\dagger a_j^\dagger P_{ij}^B + a_i^\dagger a_j Q_{ij}^B) \cdots + \sum_{i \in A} a_i^\dagger R_i^B + \sum_{i \in B} a_i^\dagger R_i^A \right] + \text{adjoint}, \tag{13}$$

$$P_{ij}^B = \sum_{kl \in B} v_{ijkl} a_k a_l, \tag{14}$$

$$Q_{ij}^B = \sum_{kl \in B} x_{ijkl} a_k^\dagger a_l, \tag{15}$$

$$R_i^B = \sum_{jkl \in B} w_{ijkl} a_j^\dagger a_k a_l, \tag{16}$$

$$S_i^B = \sum_{j \in B} t_{ij} a_j. \tag{17}$$

Note that the decomposition is not symmetrical with respect to A and B; one block carries uncontracted operators such as $a_i a_j$ (normal operators) while the other carries only the complementary operators, such as P_{ij} (though R_i is carried by both blocks). In the block configuration shown in Fig. 1, each block L and R then carries $O(k^2)$ operators, with storage of $O(M^2k^2)$. The most expensive part of blocking $L \bullet \bullet$ is consequently the formation of the complementary operators P_{ij}, Q_{ij}, which is $O(M^2k^3)$.

Further efficiency is gained by utilizing the symmetry properties of the two-index operators, such as $a_i a_j = -a_j a_i$, $Q_{ij} = Q_{ji}^T$, and by storing the operators in sparse form. We keep track of the m_s and N values in each state. When states are thus grouped by particle number and spin, the operators exhibit block-sparse structure; for example, a_i will only connect states that differ by one particle, and by the spin associated with site i. This yields the savings of sparse-storage while still allowing the benefits of machine-optimized BLAS matrix operations per matrix block, which we find leads to one to two orders of magnitude improvement in compute and memory costs. Finally, to further reduce storage one never stores all the operators of the combined block $L\bullet \bullet$. Instead, when such operators are needed, for example in solving for the ground-state wavefunction of the entire lattice, they are computed one-by-one in a "direct fashion."

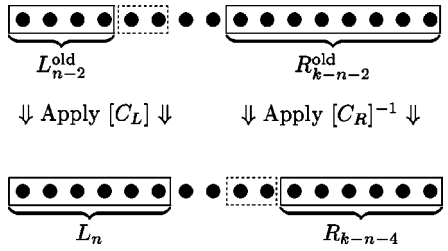

FIG. 6. Wave function transformation. The old wave function (top) is expressed in the product space of the states of L^{old} and R^{old}, and the two sites in the middle. The coefficients in the new block configuration are related by transforming the L^{old} states by C_L, and the R^{old} states by C_R^{-1}.

After we have performed the blockings $L'=L\bullet\bullet$ and $R'=\bullet\bullet R$, the next stage is to decimate L'. For this, we need the ground-state wavefunction of the superblock $L'R'$. As we only need a few states of the superblock, we use the iterative Davidson procedure,[13] where the key operation is the construction of $\mathbf{v}_{L'R'}=\mathbf{H}_{L'R'}\mathbf{c}_{L'R'}$. The dimension of $\mathbf{c}$ is the number of states in the superblock $L'R'$, which is $O(16M^2)$ (In practice this is an overestimate, because at this stage, we only consider those states of $L'R'$ which satisfy the necessary quantum constraints on the total particle number and m_s number). Thus the matrix-vector product might appear to be an $O(M^4)$ operation. However, as described by White *et al.* (Refs. 6 and 7), we can take advantage of the fact that H is composed of products of operators on block L' and block R', i.e., $H=\Sigma O_{L'}O_{R'}$. Consequently, we can define intermediates U for each $O_{L'}O_{R'}$ pair, through the structure (dropping primes for simplicity),

$$v_{l_1r_1}=[O_L]_{l_1l_2}[O_R]_{r_1r_2}c_{l_2r_2}, \tag{18}$$

$$[U]_{l_1r_2}=\sum_{r_2} c_{l_1r_1}[O_R]_{r_1r_2}, \tag{19}$$

$$v_{l_1r_1}=p\sum_{m} [U]_{l_1r_2}[O_R]_{r_1r_2}, \tag{20}$$

where p is the appropriate parity operator in the sense of Eqs. (6) and (7). Since each operator $O_{L'}$, $O_{R'}$ is an $M\times M$ matrix, the cost for the operations described above is only $O(M^3)$. Thus the cost for each matrix-vector multiply $H_{L'R'}c_{L'R'}$ is $O(M^3)$ times the number of operator pairs $O_{L'}O_{R'}$, which yields a $O(M^3k^2)$ cost per multiply.

Further improvement can be made by supplying a good initial guess to the Davidson algorithm. As suggested in Ref. 14, a suitable candidate is the wave function from the previous block configuration in the sweep. We have the two successive block configurations $L_{n-2}^{\text{old}}\bullet\bullet\bullet\bullet R_{k-n-2}^{\text{old}}$ and $L_n\bullet\bullet\bullet\bullet R_{k-n-4}$, where L_n denotes a block covering n sites, as shown in Fig. 6. The states of L are obtained from those of $L^{\text{old}'}$ through decimation, and those of R^{old} are obtained from those of R' by decimation. Also, the states of B_L are the same as those of B_R^{old} in the preceding configuration, and we shall denote them by $\{b\}$. The old wavefunction is expanded in the product space $\{l^{\text{old}'}\}\otimes\{b\}\otimes\{r^{\text{old}}\}$, as

$$\Psi^{\text{old}}=c_{l^{\text{old}'}br^{\text{old}}}\langle l^{\text{old}'}br^{\text{old}}|, \tag{21}$$

while the new wavefunction is expanded in the space $\{l\}\otimes\{b\}\otimes\{r'\}$. To transform the old wave function Ψ^{old} into the guess wave function in the new space, we can use the relevant transformation matrices $C_L(L^{\text{old}'}\rightarrow L)$ and $C_{R^{\text{old}}}(R'\rightarrow R^{\text{old}})$. Such a transformation is not exact, as it requires inverting the transformation matrix, which cannot be done (since it is not square). Instead, we use the generalized inverse, computed through singular value decomposition, and using the inverse in this sense, we obtain

$$\Psi^{\text{guess}}=\sum_{lbr'} c_{lbr'}^{\text{guess}}\langle lbr'|, \tag{22}$$

$$c_{lbr'}^{\text{guess}}=\sum_{l^{\text{old}'}r^{\text{old}}} [C_L]_{ll^{\text{old}'}}[C_{R^{\text{old}}}]^{-1}_{r'r^{\text{old}}}c_{l^{\text{old}'}br^{\text{old}}}. \tag{23}$$

Using such a guess vector, usually only three or four Davidson iterations are necessary to achieve our target tolerance of 10^{-7} in the residual wavefunction norm.

After obtaining Ψ_0, one constructs and diagonalizes the projected density matrix D, through Eq. (9). For reasons described in the next section, it is sometimes advantageous to add a small amount of noise to the density matrix. This is done by constructing the density matrix from $(1-\delta)\Psi+\delta\chi$, where χ is a random wavefunction, and δ is a small number (for example, $O(10^{-9})$). Using the lowest M eigenvectors of the density matrix D, the operators of L' are then rotated through (8), in $O(M^3k^2)$ time. These operators (together with the wave function and transformation matrix) are then saved to disk for use in the next sweep iteration or renormalization transform.

Overall, the most expensive parts of the renormalization transform are the blocking, Davidson diagonalization, and decimation steps, which cost $O(M^2k^3)$, $O(M^3k^2)$, and $O(M^3k^2)$ time, respectively. All these operations may be easily parallelized on a shared memory architecture, by associating the manipulations of the operators of a set of sites with each processor. For example, if we consider an operator P_{ij}, a given processor can handle its formation during blocking, then the multiplication with $a_i^\dagger a_j^\dagger$ during the Davidson step, and finally the decimation. Our preliminary tests have shown good speedups on a 4 processor SMP node of an IBM SP2, when our code is thus arranged. Of course, massive parallelization on a distributed architecture requires more careful consideration of data communication, particularly in the blocking step, and such an algorithm will be presented elsewhere.

2. Sweep algorithms

Successive renormalization transforms are arranged into a self-consistent sweep algorithm as described in Sec. II A. The nature of the DMRG, as explained in more detail later, is such that the accuracy of the calculation depends both on the choice and ordering of the orbitals on the lattice. In our work, we have used Hartree–Fock orbitals, which provide a good low-energy starting point, since the DMRG calculation with $M=1$ will then reproduce the Hartree–Fock energy. We now specify the order in which the orbitals are traversed. For better accuracy, we should try to minimize the range of in-

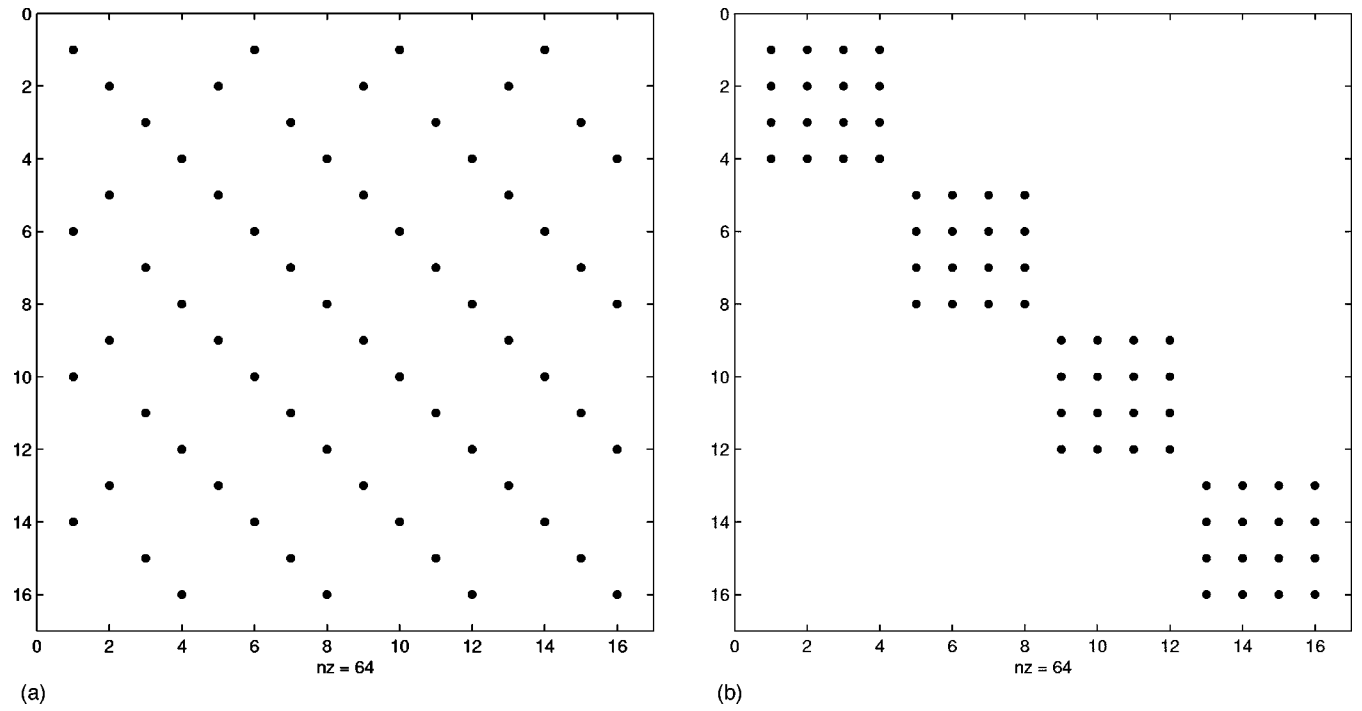

FIG. 7. One-electron integral matrices for a hydrogen chain. (a) Integral matrix, when orbitals are ordered in the Hartree–Fock energy ordering. (b) Integral matrix, after RCM reordering.

teractions of the Hamiltonian, so as to minimize the correlation length of the system. We have tried various schemes, and the simplest appears to be to minimize the bandwidth of the integral matrix t_{ij}, by reordering columns and rows. We have found the symmetric reverse Cuthill–McKee (RCM) reordering,[15,16] which swaps columns and rows so as to make a sparse matrix more closely band diagonal, to be generally satisfactory. An example of a one-electron integral reordering is given in Fig. 7.

We have previously mentioned that one starts the sweep with a small system block, for example, in the configuration ○ ○ ● ● ● ● R, where ○ ○ represents the system, so that one can construct the initial representations of the operators of the system. However, during the initial sweep, the environment states are also not specified, and consequently some sort of guess states are needed. We choose low-energy Slater determinants which are ''complementary'' to the existing system states, in that they combine to give the correct total particle number and m_s. This is because, if the environment states do not couple all the quantum numbers of the system, some quantum numbers of the system will appear with zero weight in the density matrix (9) and will consequently be lost during decimation. Due to the self-consistent nature of the sweeps, they will also not reappear in later sweeps.

As a further measure against losing states in this way, we add some noise to the density matrix during decimation, which ensures that each system state is represented with nonzero weight in the density matrix. It should be stressed that other than the requirement that the guess states should couple to the quantum numbers of the system states, their detailed choice does not affect the final converged answer of the DMRG algorithm.

The lowest energy of the superblock L ● ● ● ● R occurs near the middle of the sweep, although the energies of the other block configurations are very similar. When we refer to the sweep energy, we mean the lowest energy out of all the block configurations in the sweep. While the renormalization transform does not strictly lead to lower energies each sweep, in practice the sweep energy converges smoothly. Slow convergence can be a problem, particularly when the number of orbitals is large, and M is small. In some cases, we find that tens of iterations are necessary for convergence to a few tens of microHartrees. Such convergence issues may be traced to long wavelength fluctuations over the lattice, and are related to the slow convergence problems associated with solving differential equations on a lattice with iterative methods.[17] It has been suggested elsewhere (for example, Ref. 14) that starting sweeps at small M and then gradually increasing M with each sweep shortens calculation time. As our calculation time is usually dominated by the many sweeps conducted at the largest M value, we have not observed any great advantage in such a scheme.

Finally, we note that each sweep consists of $O(k)$ renormalization transforms. Consequently the total computational cost per sweep is $O(M^2k^4+M^3k^3)$, with memory usage $O(M^2k^2)$ and $O(M^2k^3)$ disk usage.

3. Remarks on properties and symmetries

How does one compute properties other than the total energy? Because one does not have the detailed structure of the states, we can only compute such expectation values as can be constructed from the operators we store on each block. Thus if we store up to two-body type operators of the form $a_i^\dagger a_j^\dagger a_k a_l$, then one can evaluate the expectation value of any two body operator in the form $\Sigma o_{ijkl} a_i^\dagger a_j^\dagger a_k a_l$. In the

algorithm above, we never store four index quantities for reasons of efficiency. Instead we only keep the contracted operators P_{ij}, Q_{ij}, R_i, S_i. In the same way, we may construct contracted operators for any two particle quantity, in the form $\Sigma_{kl} o_{ijkl} a_i^\dagger a_j^\dagger$, etc. We can keep track of such quantities and block and decimate them in the same way as for the Hamiltonian operator. Thus we may compute up to any two-particle property without much additional effort (i.e., we do not effect the order of the polynomial scaling). Nuclear gradients, which depend only on the two-particle density matrix, may be computed in this way.

For fermionic systems, we need to keep track of the particle numbers per state, to compute the parity operator p in Eqs. (6) and (7). However, it may be advantageous to keep track of more information for each state, related to spin or symmetry labels. For example, in the above algorithm, we keep track of the m_s quantum numbers. It is also possible, though more complicated, to utilize symmetries in S^2 and in the point group. We have not done so here. However, the more symmetry information that is used, the higher the sparsity of the operator matrices. Furthermore, the expansion of the DMRG wave function in the tensor product space is much shorter.

Because the Hamiltonian commutes with any symmetry operator of the molecule, then for sufficiently large M, our DMRG wave function will also be an eigenstate of all the symmetry operators. Thus even if one does not actually retain the symmetry information in the states, one need not worry about such problems as spin contamination in the lattice wave function, so long as one retains enough states per block.

III. ANALYSIS OF THE DENSITY MATRIX RENORMALIZATION GROUP

A. Excitation structure, basis functions, and correlation

In quantum chemistry, it is usual to analyze a correlation treatment in terms of its excitation structure. Let us consider a Hartree–Fock reference $|\text{ref}\rangle = |\phi_1 \phi_2 \phi_3 \phi_4 \cdots \phi_n\rangle$. Now start with a system block $L_1 = \bullet_1$ and consider the blocking and decimation of $\bullet_1 \rightarrow \bullet_1 \bullet_2 (L_2)$ (adding one site at a time for simplicity). In decimating, we select the most important states of $L_2 = \bullet_1 \bullet_2$. These represent the significant excitations, and combinations of excitations (singles and doubles) out of orbitals ϕ_1 and ϕ_2 in the Hartree–Fock reference. In the next step, we add site ϕ_3, and perform a further decimation. Then the important states of our system describe the most significant singles and doubles excitations out of L_2, the single excitations out of ϕ_3, and the doubles and triples excitations that are *combinations of products* of the significant excitations of L_2 and ϕ_3. In this way, the renormalization procedure is able to capture high-body excitations in an economical way, relative to explicit configuration interaction.

This procedure bears some similarity to various selected configuration interaction schemes.[9] In addition, the product structure of the excitations is reminiscent of coupled-cluster theory. However, the difference lies in transforming the N-particle basis used to represent the lattice to a more optimal N-particle basis at each step, rather than using a basis of Slater determinants. Each basis function is a complicated mixture of many Slater determinants, although this representation is never explicit. Thus, from this perspective, one can view the DMRG is a method of performing configuration interaction, without the need to explicitly store any long Slater determinant expansion.

One might ask the question, is it possible to explicitly construct the Slater determinant expansion of the DMRG basis functions at any given stage? The answer is yes, provided that we store all the transformation matrices C, that are used in the decimation procedure. Then, one can retrace the steps of the renormalization group algorithm, applying the matrix C at each stage, and keep track of the mixing of the basis functions. Thus implicitly, the Slater determinant expansions of the basis functions are contained completely within the set of transformation matrices.

It is useful at this point to make some link between the common quantum chemical concepts of dynamical and nondynamical correlation, and ideas in renormalization group theory. Roughly speaking, the usual quantum chemical approach is to take the mean-field (Hartree–Fock) solution as a starting point, and then to correct this using low-order perturbation or configuration interaction theory. Such corrections are usually associated with dynamical correlation, while qualitative errors that arise from the insufficiency of the Hartree–Fock reference due to near degeneracies (which cannot be corrected through perturbation theory), are associated with nondynamical correlation.

RG approaches were originally conceived to handle cases where mean-field theory breaks down completely. The hypothesis in mean field theory is that the statistical correlations between one particle and all the others are summable; then for sufficiently large numbers of particles, the law of large numbers ensures that the ''error'' in the mean field is distributed as a Gaussian. However, if correlations between constituents do not decay sufficiently quickly, the correlations may indeed *not* be summable, and mean-field theory is thus invalid.

Although ''criticality'' is really a thermodynamic concept, in quantum chemical systems, the lack of convergence of perturbation theory when the Hartree–Fock reference is poor (for example in bond-breaking), may be seen as a form of ''unsummability'' due to slow decay of correlations (by order of perturbation theory). In this sense, there is an infinite correlation length, as associated with criticality. Criticality is also associated with divergence of thermodynamic derivatives. This is also the case in quantum systems; when there is a near degeneracy, the energy can change only by a very small amount, but the occupation numbers of the states jump discontinuously. Thus, though the analogy is not complete, we can view a RG approach as particularly suited to addressing problems of *nondynamical correlation*, where there are important large high-body excitations. We will witness this in some of our calculations in the later sections.

B. Variational nature

Because the DMRG simply restricts the Hilbert space in which the Hamiltonian is solved, it is a variational method. However, the energy need not go down from one block configuration to the next, or even from sweep to sweep, because there is no simple relation between the Hilbert spaces spanned by successive block configurations in a sweep. One exception to this is when both the left and right hand blocks carry M states, for example in the block configuration LB_LR. By this we mean that we do not add any extra sites B_R to the environment during the decimation procedure, unlike in Fig. 1. Then, since there are only M environment states, the number of nonzero density matrix eigenvalues carried by LB_L is also M, and thus the renormalization transform incurs no loss of information in the *ground-state wave function* (although, of course, information is lost in the representation of the total Hilbert space). Consequently, the ground-state wave function in the configuration LB_LR can be expressed exactly in the succeeding configuration $L'B_{L'}R'$, and thus by the variational theorem, the ground-state energy cannot go up between iterations.

As previously mentioned, in practice we do observe the sweep energy to converge smoothly, although the lowest energy per sweep is generally in the middle of the sweep.

C. The computational complexity of the DMRG

It is generally believed that the problem of obtaining the exact wave function is a very difficult, exponentially difficult problem.[32] The DMRG, on the other hand, proposes a polynomial time algorithm to compute an arbitrarily accurate energy. Such statements are not contradictory. As has previously been emphasized informally by Kohn,[18] good energies should correspond only to good *reduced density matrices*, but may come from poor wave functions.

This can be made more explicit, by considering the overlap of the DMRG wave function with the true wave function. In general, if each orbital has an error of $O(\epsilon)$, the overlap with the wave function will vanish exponentially, but the error in the energy is only a polynomial in ϵ. In more detail, consider a lattice of equivalent sites each of which carries two orthogonal states. Now consider a renormalization algorithm where we keep only $M=1$, and add on one site at a time. Since $M=1$, then when performing the decimation $L\bullet_i \rightarrow L'$, the discarded weight ϵ in the density matrix may be associated entirely with the error in the space spanned by site $\bullet_i$. The approximate $M=1$ wave function, for the lattice of k sites, may be written in product form, as

$$\Psi = \prod_n^k \psi_n . \tag{24}$$

Since the wave function is in product form, so is the overlap integral with the true wave function Ψ_0, and thus we find

$$\langle \Psi | \Psi_0 \rangle \approx \prod^k (l-\epsilon)\Phi e^{-\lambda k}, \tag{25}$$

(where $e^{-\lambda} \sim 1-\epsilon$), i.e., the overlap of the DMRG wave function with the true wave function decreases *exponentially fast* with the number of decimations.

On the other hand, the two-body density matrix, is a product of four operators. Roughly speaking, each operator is associated with an error ϵ, consequently, each term $a_i^\dagger a_j^\dagger a_k a_l$ in the two particle density matrix is associated with an error of $O(1-(1-\epsilon)^4)$. Consequently, the error in the total energy is only

$$\delta E \approx n(k)(1-(1-\epsilon)^4), \tag{26}$$

where $n(k)$ denotes the number of significant interaction elements v_{ijkl}.

The polynomial bound on the error, independent of k apart from the factor $n(k)$, is characteristic of any operator that acts on a fixed number of particles at a time. Of course, all thermodynamic operators are of this form, and therein lies the success of the renormalization group approach.

D. Convergence of the energy

The detailed numerical behavior of the DMRG algorithm is still imperfectly understood. In an early study of the accuracy of the algorithm for the Ising model, Legeza and Fath[19] demonstrated that the error in the energy for given M, is roughly proportional to the sum of the weights discarded during decimation. More strictly,

$$|E(M)-E_0| \sim \text{const} \sum_{i>M} w_i + C. \tag{27}$$

The linear relationship follows simply from the boundedness of the Hamiltonian and because the energy is a linear functional of the density matrix.

From Eq. (27), we deduce that the accuracy of the DMRG is strongly dependent on the rate of decay of the eigenvalues of the density matrix w_i [see Eq. (11)] with increasing index. From the normalization of the (infinite rank) density matrix, we can see that w_i must decay at least faster than $1/i$, but in practice, the decay rate is quicker, and probably faster than any polynomial in i. This is not strange, as w_i are the expansion coefficients of Ψ_0 in the optimal basis $\{|\theta\rangle\} \otimes \{|r'\rangle\}$, and for sufficiently smooth functions, expansions in an orthogonal basis always converge faster than algebraically.

In the infinite system limit, away from criticality, one can consider the environment as a heat bath, and consequently the density matrix of a subsystem is simply the thermal density matrix, $\propto e^{-\beta(H-\mu N)}$. We have recently proved,[20] that under fairly general conditions (see also Okunishi *et al.* for a simpler discussion specific to the Ising model[21]), the thermal density matrix eigenvalues decay asymptotically like

$$w_i \sim \text{const}\, e^{-\kappa(\ln i)^\alpha}, \tag{28}$$

where $\alpha \sim 2$, and constant κ is model specific and proportional to the one-particle level density.

In Fig. 8, we plot density matrix eigenvalues obtained from a neon DZP calculation, and from a water DZP calculation (described later). Here we observe good agreement

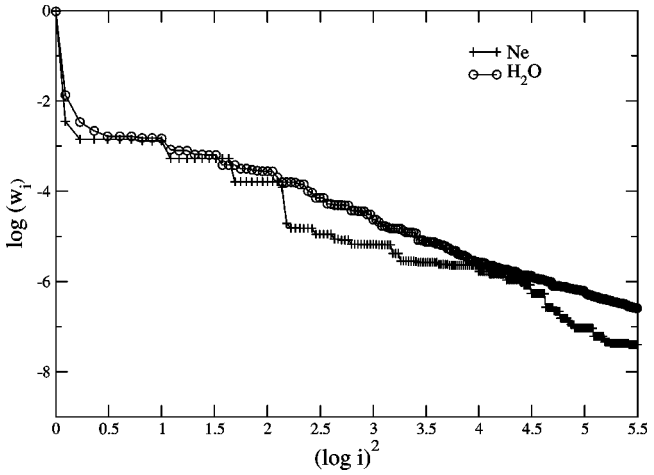

FIG. 8. Decay of the density matrix weights w_i with the eigenvalue index i for a neon DZP calculation and a water DZP calculation.

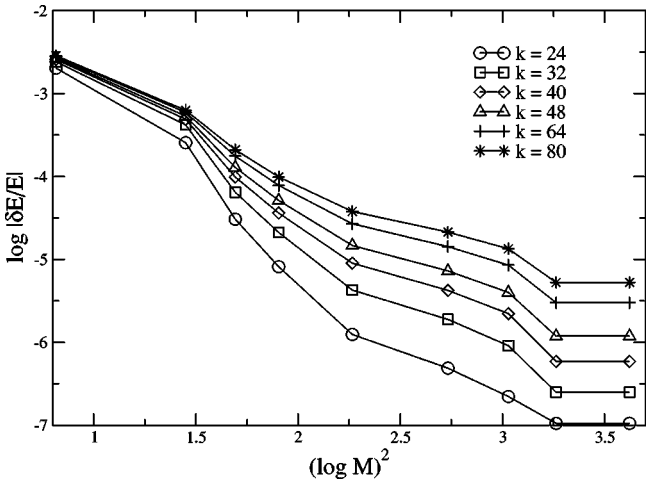

FIG. 9. Convergence of the energy error $|\Delta E/E|$ with the number of states M for Hückel chains of lengths k.

with the asymptotic form. This confirms the general validity of Eq. (28). Furthermore, from Eqs. (27) and (28), the error in the energy should asymptotically converge like

$$|\delta E| \sim \text{const} \int_M^\infty e^{-\kappa(\ln i)^2} di = \text{const} \left(\frac{e^{-\kappa(\ln M)^2 + \ln M}}{\ln M} \right). \tag{29}$$

Taking logarithms on both sides, we find that the leading term gives

$$\ln|\delta E| \sim -\kappa(\ln M)^2. \tag{30}$$

In Figs. 9, 11, and 14 we observe a fairly clear linear relationship between $\ln|\delta E|$ and $(\ln M)^2$ for calculations on Hückel systems, hydrogen chains (described in the next section), and the water molecule (described in Sec. IV A). We have noted a similar relationship in our other calculations. In particular, our energy data fits Eq. (30) far better than the form $\ln|\delta E| \sim \text{const}\, M$ proposed elsewhere by White *et al.*[6,10]

E. Accuracy and correlation lengths

As described above, the accuracy of conventional quantum chemical methods is strongly linked to the quality of the reference system, which may be one or a few Slater determinants. In the DMRG, if we choose our orbitals to be Hartree–Fock orbitals, then a calculation with $M=1$ will yield the Hartree–Fock energy. In this sense, if the Hartree–Fock reference is accurate, we may also expect an accurate DMRG calculation, and vice versa.

In addition, the accuracy of the DMRG depends strongly on the *correlation length* of the system. One can view this length as the characteristic length[33] of the exchange-correlation hole $h_{\text{xc}}(\mathbf{r}_1,\mathbf{r}_2)$, defined in the usual way from the diagonal of the two particle density matrix $\rho_2(\mathbf{r}_1,\mathbf{r}_2) = \frac{1}{2}\rho(\mathbf{r}_1)[\rho(\mathbf{r}_2) - h_{\text{xc}}(\mathbf{r}_1,\mathbf{r}_2)]$. In noninteracting systems, the exchange-correlation hole arises purely from exchange effects, and for insulators with a band gap Δ, is believed to have a length $\sim \Delta^{-1/2}$.

We use a simple Hückel like model to study the dependence of the accuracy of the DMRG on the correlation length of the system. For implementational purposes, each site is a spin–orbital, arranged $\alpha\beta\alpha\cdots$, and only electrons of like spin interact. Thus,

$$H = \sum_{ij} \beta_{ij} a_i^\dagger a_j, \tag{31}$$

where $\beta_{ij} = -8$, for $i=j$ or $\beta_{ij} = -1$ for nearest neighbors of the *same spin* (thus nearest interacting neighbors are in fact two sites apart). We studied this model for various chain lengths ranging from 24 to 80 spin–orbitals. In Fig. 9, we examine the rate of convergence of the fractional error in the DMRG energy as M is increased, for the various chain lengths. Again we observe an approximate linear relationship between $\log|\delta E|$ and $(\log M)^2$, with the rate of convergence of the energy dropping for increasing chain lengths. We can fit to find the slopes of the curves in Fig. 9, as an estimate of the constants κ in Eq. (30). In Fig. 10, these κ values are plotted against the "correlation length" $\Delta^{-1/2}$, and a suggestive near-linear relationship is seen. This is consistent with the proportionality of κ in Eq. (28) to the one-particle level spacing, as described in Ref. 20.

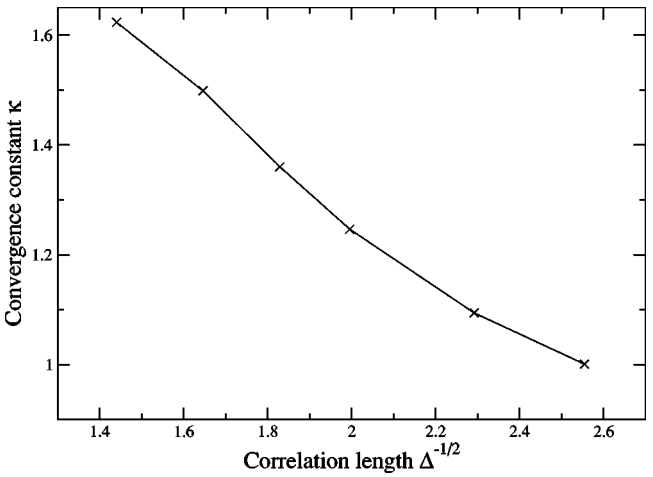

FIG. 10. Convergence constant κ in Eq. (30), from fitting above data, plotted against $|\epsilon_{\text{HOMO}} - \epsilon_{\text{LUMO}}|^{-1/2}$.

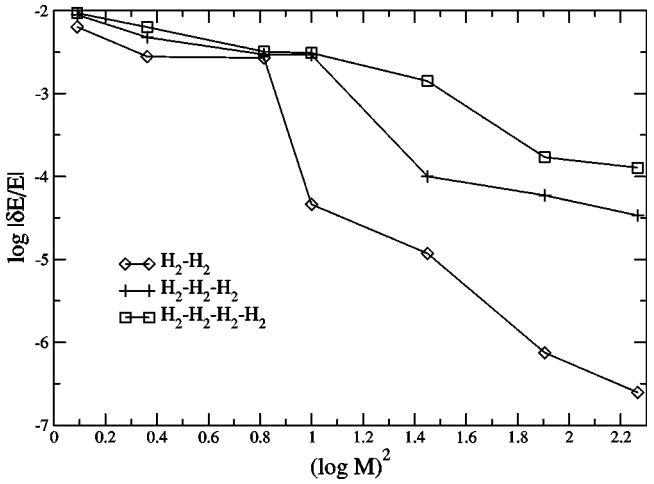

FIG. 11. Convergence of the energy error $|\Delta E/E|$ with the number of states M for a hydrogen chains of various lengths.

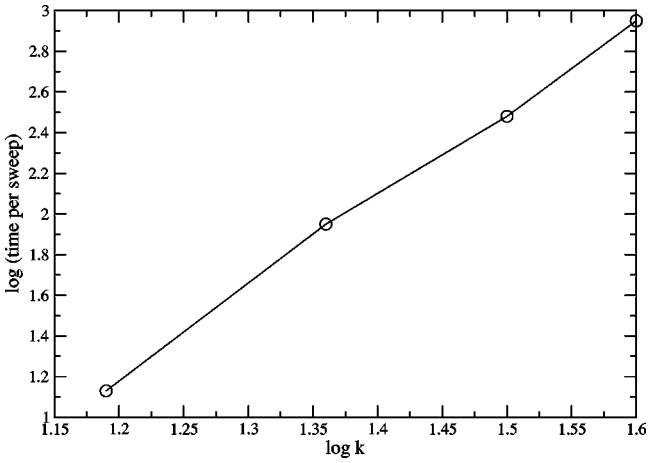

FIG. 12. Sweep time t as a function of the number of orbitals k for a set of hydrogen chains.

In interacting systems, the dependence of the correlation length on the effective range of the two-particle interaction is complicated. As we increase the range of interactions from 0, we expect the correlation length to increase. However, as the range of interaction$\rightarrow\infty$, mean field theory becomes exact, and the system becomes uncorrelated once more.

However, in many cases, the most important correlation length in the DMRG arises from the mapping of the Hamiltonian onto a one-dimensional lattice, which introduces an *artificial* lattice correlation length, Consequently, the DMRG depends in a more intricate fashion on the choice of orbitals than other correlation treatments, as certain choices or orderings lead to shorter ''lattice correlation lengths.'' As an example, consider the case of two systems A, B at infinite separation. In the case where the orbitals are arranged in two groups, where each group is localized on each system independently, the DMRG is exactly size consistent (i.e., for fixed M, the energy of the two systems is the sum of the energies of the systems calculated independently at the same M), as the wave function is in tensor product form $|A\rangle|B\rangle$. However, if we allow the orbitals to become delocalized across both systems (for example, as in a restricted Hartree–Fock description of bond-breaking), there exists long-range interactions between orbitals, and the exchange-correlation hole becomes delocalized. Artificial long-range correlations exist even for localized orbitals, *if we do not place them in the correct order on the lattice*, for example, if we arrange them as ϕ_1^A, ϕ_2^B, $\phi_3^A\cdots$. The DMRG calculation for the total system AB will be much less accurate using delocalized orbitals, or improperly ordered orbitals. Thus the DMRG is only size-consistent, with an appropriate choice of the one-particle space.

In a basis with approximately p states per unit length, for a system characterized by a correlation length λ, we expect the DMRG to become size consistent for $M\gg p^{\lambda}$. As we approach criticality (i.e., the metal–insulator transition), $\lambda\rightarrow\infty$, and we can no longer expect the DMRG to be size consistent, in a localized basis. Some studies of the DMRG near criticality are presented in Legeza and Fath.[19]

In Fig. 11, we plot the convergence of the DMRG energy (fractional error) as a function of $\log(M)^2$, for various polymers of H_2, i.e., hydrogen molecules arranged $H_2\cdots H_2\cdots$. Although the intermolecular spacing is very large (10 Å), and the molecular orbitals are themselves localized, we work with the molecular orbitals arranged in the usual energy ordering, which makes the Hamiltonian artificially long ranged, as shown by the one-electron integral matrix in Fig. 7. We observe that the quality of the DMRG calculation degrades as the hydrogen chain length and correlation length increases. By contrast, we find that the DMRG calculation is exact for $M=2$ for *all* chain lengths, if the molecular orbitals are properly reordered in the order of their respective hydrogen molecules (see Fig. 7), using the RCM scheme as described in Sec. II B 2.

Although in this case the RCM reordering is very successful, as previously mentioned, in complex chemical systems, interactions exist in three dimensions, and thus the usual lattice mapping of the Hamiltonian induces artificial long-range interactions far beyond the long-range nature of the Coulomb interaction. In such cases, a mapping onto a three-dimensional lattice is desirable.

F. Computational scaling

As described in Sec. II B 2, the DMRG algorithm formally costs $O(M^2k^4)+O(M^3k^3)$ time per sweep. Note that this is without any guaranteed bound on the error, however, we know that the error decays quite quickly with M as discussed earlier. In Fig. 12 we present calculation times per sweep on a series of hydrogen chains of increasing length. These calculations were performed with small M values, and hence the scaling is dominated by the M^2k^4 term. We observe quartic scaling in the logarithmic plot (the slope is 4.2). However, for most of the more accurate calculations presented in the following section, where M is an order of magnitude larger than k, the dominant cost is more like $O(M^3k^3)$. In Fig. 13, we plot the sweep times for a series of calculations on the water molecule, with increasing M. We see that for the larger M values, the sweep times begin to display a cubic dependence on M.

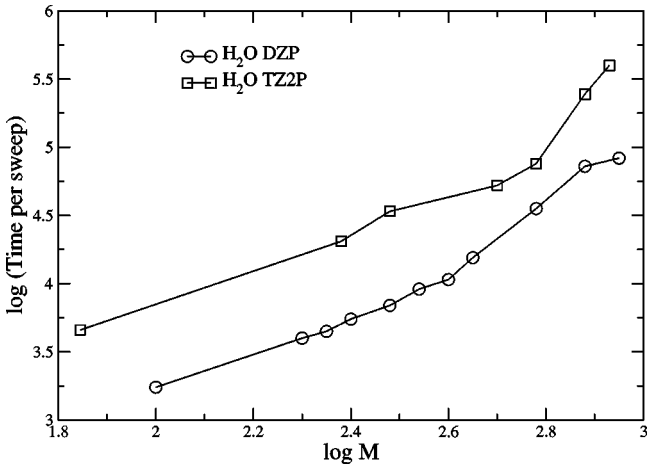

FIG. 13. Sweep time t as a function of the number of states M for a water calculation.

IV. ILLUSTRATIVE CALCULATIONS

In this section we present a number of illustrative DMRG calculations on H_2O, the N_2 curve, and C_2H_4 at the twisted and flat geometries. In the first half (''small systems''), full CI results are available for comparison, while in the second half (''larger systems''), we compare against coupled cluster calculations. Except where indicated, the coupled cluster results were obtained using the ACES package.[22] Some additional computational details, such as the orbital ordering, are given in Appendix B.

A. Small systems

The water molecule with a DZP basis is a commonly used benchmark system, as full CI energies have been computed by Bauschlicher and Taylor.[23] This system was also studied in the DMRG work of White and Martin.[6] The basis set used is the Dunning DZP basis,[24] with the exponents and geometry as specified in Ref. 23. To facilitate comparison with the full CI results, we freeze the O $(1s)$ core, yielding 8 active electrons and 25 active orbitals. It should be noted that the original Bauschlicher and Taylor energy is not in fact converged at the μH level, and thus we have recomputed a more accurate FCI energy for use in this work.

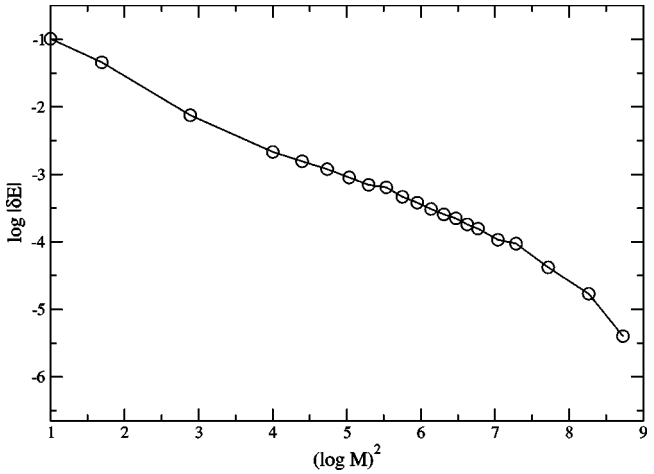

FIG. 14. Error of DMRG calculations on the water molecule at the DZP level, as a function of M.

TABLE I. DMRG results for water in a DZP basis, as we increase the number of states M. Our recomputed full CI energy is $-76.256\,634$ H [the original Bauschlicher and Taylor energy (Ref. 23) is $-76.256\,624$ H].

M	E/H	δE/mH
100	−76.2545	2.1
200	−76.2559	0.71
300	−76.25632	0.32
400	−76.256477	0.157
500	−76.256540	0.094
600	−76.256592	0.042
750	−76.256617	0.017
900	−76.256624	0.004
CCSD	−76.252503	4.131
CCSD(T)	−76.255907	0.727
CCSD(TQ)	−76.256846	0.202

Shown in Fig. 14 and Table I is the convergence in the DMRG energy as a function of $(\log M)^2$, to the full CI energy. The energies at lower M are given to fewer significant figures, as the sweep energies converge only slowly for such approximate calculations. Note that for the large M values, the error in the energy decreases more rapidly than expected from the form (30). This is an ''edge effect,'' which is due to the finite rank of the density matrices.

We compare our results against coupled cluster theory, with perturbative triples [denoted CCSD(T)] and perturbative triples and quadruples [denoted CCSD(TQ)].[25] White and Martin performed calculations up to $M=375$, with a highest accuracy of 0.24 mH. This result, however, was not fully converged, and with tighter thresholds, we obtained an error of 0.18 mH at the same M value. Using current technology, we have been able to further extend those results to $M=900$ yielding a final error of roughly 4 μH. The total computation time was roughly 5 days on an IBM RS/6000 workstation. Thus for systems of this size, we are able to achieve μH accuracy without too much effort. As can be seen from the table, already at $M\sim400$, the DMRG is more accurate than even CCSD(TQ).

B. Nitrogen dimer

Recently, Larsen, Olsen, and Jorgensen[26] have computed a benchmark N_2 curve at the full CI level using a Dunning cc-pVDZ basis,[27] and N $(1s)$ frozen (10 active electrons and 28 active orbitals).

As is well known, the dissociation of the nitrogen molecule within a restricted formalism is difficult within any conventional quantum chemical approach, due to the inadequacy of the restricted Hartree–Fock reference at long bond lengths. While we do not advocate the use of the restricted reference at long bond lengths for a DMRG calculation (since, as argued in Sec. III E, the DMRG will not be size-consistent with such a reference, as the spin-paired orbitals do not localize on their respective atoms), such calculations nonetheless provide a stringent test of the ability of the DMRG to recover nondynamic correlation.

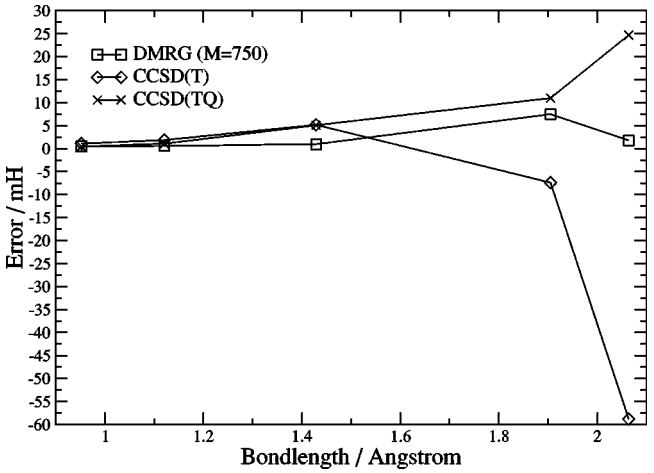

FIG. 15. Errors in various methods (from full CI) for the nitrogen molecule at various bond lengths.

In Fig. 15, we present the results of DMRG calculations, using restricted Hartree–Fock orbitals, at a number of N_2 geometries; the precise figures are given in Table II. The equilibrium geometry is at $r_e = 1.1208$ Å. As can be seen, our most accurate calculations, with $M = 750$, are considerably and consistently better than the CCSD(TQ) energies. As one stretches the bond, and the Hartree–Fock reference orbitals become a poorer starting point, the DMRG error also increases. However, the method does not fail as catastrophically as coupled-cluster theory, and even at the longest bond length of 2.0673 Å, the total energy error with $M = 750$ is only ~1.8 mH. The maximum error (~7 mH) occurs at the bond length 1.9050 Å, which is near where the CCSD(T) curve turns over. We do not fully understand the DMRG result at this geometry; the increased error may be due to the intrinsic complexity of the wave function, or due to noise from a poor choice of orbital ordering for this bond length. Nonetheless, it appears that the DMRG, by its optimal selection of states from each region in Fock space, overcomes to a large degree the problem of nondynamic correlation with single-reference methods. This is of course consistent with the ideas behind renormalization group theory, i.e., to correct when mean-field approaches fail.

We stress again, that one should use localized orbitals at the longer bond lengths for more accurate DMRG results. The unrestricted Hartree–Fock reference provides such a localized set. Note that with $M = 1$, in such as basis, the DMRG will reproduce the unrestricted HF result, and consequently will be considerably more accurate than the results with the restricted reference. The usual difficulty with the unrestricted Hartree–Fock reference wave function is that it is not a pure spin-state and consequently single reference methods based on this wave function suffer from spin contamination. Our view, as mentioned in Sec. III E, is that because the DMRG is an approximation to the exact solution, spin contamination is unimportant for sufficiently large M.

TABLE II. DMRG results for N_2 in a cc-pVDZ basis. Errors from the full CI energy, measured in mH.

Bond length/Å	DMRG	CCSD(T)	CCSD(TQ)
0.95250	0.47	1.11	0.44
1.12080 (r_e)	0.60	1.87	1.09
($M = 900$)	0.43		
1.42880	0.97	5.19	5.09
1.90500	7.5	−7.4	11.0
2.06376	1.8	−58.8	24.7

TABLE III. DMRG results for water in a TZ2P basis, as we increase the number of states M. Our estimated full CI energy is −76.3169 H.

M	E/H
70	−76.3019
240	−76.3147
300	−76.3154
500	−76.31637
600	−76.31656
750	−76.31669
850	−76.31676
CCSD	−76.31005
CCSD(T)	−76.31678
CCSD(TQ)	−76.31686

C. Larger systems

We now study some larger systems, by which we mean systems with a larger number of orbitals. From our analysis in Sec. III E, we expect such systems to pose challenges for the current formulation of the DMRG, due to the long correlation lengths across the lattice.

As a taxing test of our code, we examined the water molecule at the same geometry in a larger TZ2P basis set. We use the 6-311$G(2d,2p)$ basis,[28,29] with pure spherical functions, yielding 10 active electrons and 41 active restricted Hartree–Fock orbitals. We cannot yet conceive of full CI calculations (in C_1 symmetry, this requires roughly 5.6×10^{11} determinants) on such a large system, and so instead, we compare our results against other approximate quantum chemical methods. These results are shown in Table III.

As can be seen, with our largest calculation ($M = 850$), the energy we recover is essentially the same as that of CCSD(TQ), with an estimated error of 0.1–0.15 mH. This error is considerably larger than that obtained for a similar M value in the DZP basis. This is not surprising, from our arguments concerning the increased lattice correlation length. It should be noted that with our estimate, the error of the DMRG when going from the DZP to the TZ2P basis increases (31 μH $\rightarrow$ 0.1–0.15 mH), whereas that of CCSD(T) *decreases* (0.7 mH $\rightarrow$ 0.1–0.15 mH). We interpret this as due to the nonvariational nature of CCSD(T) theory. However, the improved performance of coupled cluster theory as compared to the DMRG in this system, dominated by dynamical correlation, is an indication of the weakness of the DMRG in this area.

As an indication of the effort expended for the largest DMRG calculation, each sweep at the $M = 850$ level took almost 2 days of CPU time, and we carried out 7 sweeps.

TABLE IV. Ethene (flat and twisted conformations) total energies in a DZP/DZ basis. The DMRG calculations are performed with $M=750$.

	Flat E/H	Twisted E/H
DMRG	−78.3495	−78.2304
CCSD(T)	−78.3516	−78.2274
CCSD(TQ)	−78.3524	−78.2313

D. Ethene

We now consider a more interesting system, namely the twisted and flat conformations of C_2H_4. As has been pointed out elsewhere,[30] rotating about the C_2H_4 bond changes the problem from one which is dominated by dynamical correlation to one which is dominated by nondynamical correlation. At the twisted configuration, for the wave function to transform as an irrep of the D_{2d} point group, both the π^2 and π^{*2} determinants must have equal weights in the full CI expansion. Consequently, all single reference methods have a difficult time correctly describing the wave function at this geometry. From the DMRG perspective, this system is also challenging because of the large number of orbitals and the long lattice correlation length.

We computed the DMRG energies using restricted Hartree–Fock orbitals, obtained with a Dunning DZP/DZ basis[24] ($4s2p1d$ on C, $2s$ on H, 38 basis functions in total), where the carbon d function was chosen with exponent 0.75, as in Ref. 30, and we correlate all electrons (16 electrons). We use $r_{CC}=1.33$ Å, $r_{CH}=1.076$ Å, $\theta_{CCH}=121.7°$. Our results, together with those of restricted coupled-cluster calculations, are presented in Table IV.

In the case of flat ethene, the CCSD(TQ) numbers lie more than 2 mH below our most accurate DMRG calculations. While the CCSD(TQ) result is not variational, it is probably more accurate than the DMRG answer in this case.

At the twisted conformation, the DMRG result is substantially more accurate than the CCSD(T) result, lying below it by roughly 2.5 mH, and comparable to the CCSD(TQ) result. The relative accuracies of the DMRG results at the two geometries, is again consistent with the view that the DMRG is better at picking up the nondynamical correlation than usual quantum chemical methods, while its performance for dynamically correlated systems is less impressive.

It should be noted that the DMRG wave function is exceedingly compact, as our most accurate result is obtained with a wave function expansion of only 600 000 states. By contrast the CISD space (without any symmetry constraints) is already 1.4×10^8 determinants in size. Thus the performance of the DMRG at the current M values, for dynamically correlated problems of this increased size, is not really ''poor,'' as the expansion length is so short. Of course, the DMRG may be systematically increased in accuracy by increasing M, but the current cost of the method, despite its favorable scaling, makes such calculations prohibitive.

V. SUMMARY AND CONCLUSIONS

In this work, we have undertaken an exploration of the density matrix renormalization group algorithm, as originally presented by White and co-workers. We have described, in some detail, how such an algorithm may be efficiently implemented for the study of quantum chemical problems. We then provided an analysis of the DMRG, including its error convergence behavior, its cost, and its links with existing chemical theory.

We applied our algorithm to calculations on water, ethene and the nitrogen dimer. At the highest level of accuracy used in this study, we find our results compare very favorably with coupled cluster theory. In particular, in situations where the nondynamic correlation is important, the DMRG performs much better than conventional single-reference methods. The renormalization group can thus be seen as providing a new reference, where conventional mean-field (i.e., Hartree–Fock) theory breaks down.

Although the cost of our most accurate calculations is considerably higher than that of CCSD(T) theory, the DMRG algorithm is not a perfected work, but instead is something which is amenable to substantial future improvement, and it is in this spirit that we have pursued this work. We feel that an essential component of the DMRG, the transformation away from the standard Slater determinant basis into an efficient, adaptive many-body basis, is a very important insight. Of course, the favorable scaling characteristics of the algorithm even in its present form lead one to be optimistic about its applicability to larger systems.

At present, the method is not without its difficulties. We have seen that it becomes increasingly difficult to capture dynamic correlation in larger basis sets, because of the increasing range of the lattice Hamiltonian. Whether this will be alleviated by a mapping to a three-dimensional lattice, or by a suitable clustering of strongly-interacting orbitals, or even through a perturbative approach, is a topic that requires further investigation. A related problem is the need to formulate the theory using localized orbitals, to preserve size-consistency in extended systems. Such questions will benefit from a formulation of the renormalization group theory in terms of a nonorthogonal basis, which we shall describe elsewhere.

ACKNOWLEDGMENTS

G.K.-L.C. acknowledges S. R. White for many interesting discussions, and the Miller Institute for Basic Research in Science for funding.

APPENDIX A: PSEUDOCODE FOR THE DMRG ALGORITHM

Forwards sweep: Set up initial system, e.g. $\bullet_1\bullet_2$,
  states $|\mathrm{vac}\rangle$, $|\phi_1\rangle, \phi_2\rangle$, $|\phi_1\phi_2\rangle$,
  operators a_1, a_2, a_1a_2, $a_1^\dagger a_2\cdots$.
Backwards sweep: Set up initial system, e.g., $\bullet_{n-1}\bullet_n$,
  states $|\mathrm{vac}\rangle$, $|\phi_{n-1}\rangle, \phi_n\rangle$, $|\phi_{n-1}\phi_n\rangle$,
  operators a_{n-1}, a_n, P_{12}, $R_{12}\cdots$.
Do until end of lattice (forwards sweep).
(1) $L\rightarrow L\bullet_i\bullet_{i+1}$:
  Form new states $|l\rangle$, $|l\phi_i\rangle$, $|l\phi_{i+1}\rangle$, $|l\phi_i\phi_{i+1}\rangle$,
  by combining quantum numbers of old states, e.g.,
  $\mathrm{N}[|l\phi_i\rangle]=\mathrm{N}[|l\rangle]+1,\ldots$.

Form operators of L in new states, Eq. (6).
Form operators of $\bullet_i\bullet_{i+1}$ in new states, Eq. (6).
Form coupling operators, e.g., $a_i a_{j\in L}$, Eq. (7).
(2) $R \rightarrow \bullet_{i+2}\bullet_{i+3}R$:
If initial sweep,
construct trial Slater Determinant states
and operators for R;
else
load R from previous sweep.
Form new states $|r\rangle$, $|\phi_{i+2}r\rangle$, $|\phi_{i+3}r\rangle$, $|\phi_{i+2}\phi_{i+3}r\rangle$,
by combining quantum numbers of old states.
Form operators of R in new states, Eq. (6).
Form operators of $\bullet_{i+2}\bullet_{i+3}$, Eq. (6).
Form coupling operators, e.g., $\Sigma t_{ij} a^\dagger_{i+2} a^\dagger_{j\in R}$, Eq. (7).
(3) Form superblock $L \bullet \bullet \bullet \bullet R$ states:
If not initial sweep,
Construct guess Ψ from
Ψ_0 of previous step, Eq. (23).
Find lowest Ψ for the superblock, Davidson method,
Hc is constructed using intermediates, Eq. (20).
Save wave function.
(4) Form density matrix of L', Eq. (9):
Add noise, diagonalize, and
transform L' operators, Eq. (8).
Save L' states and operators.
enddo
Repeat for backwards sweep,
interchange system, and environment,
until convergence in sweep energy.

The code was implemented in C++. The following summarizes the key data structures and their elements:
State: array of integers, holding the quantum numbers N and m_s.
Operator: Block-sparse matrix (array of matrices), identification data (e.g., a_i), etc.
Block: array of operators, array of states, identification data, etc.

APPENDIX B: COMPUTATIONAL DETAILS

The orbital orderings used were:
H_2O DZP: 1 17 11 25 9 2 16 18 13 7 4 22 23 10 5 12 8 20 24 3 6 21 14 19 15;
H_2O TZ2P: 41 39 38 34 32 30 25 24 22 1 18 9 10 4 13 6 17 2 40 35 31 3 29 16 19 8 23 14 26 7 15 36 20 33 5 27 12 28 37 11 21;
$N_2, r_e = 0.952\,50$ Å: 11 25 1 14 7 3 17 28 2 20 10 4 5 21 12 6 22 13 8 26 15 9 27 16 18 19 23 24;
$N_2, r_e = 1.120\,80$ Å: 11 25 1 14 5 3 17 28 2 18 10 4 6 21 12 7 22 13 8 26 15 9 27 16 19 20 23 24;
$N_2, r_e = 1.428\,80$ Å: 14 25 1 15 5 3 11 28 2 18 10 4 6 19 12 7 20 13 8 26 16 9 27 17 21 22 23 24;
$N_2, r_e = 1.905\,0$ Å: 14 19 1 17 5 3 13 28 2 18 10 4 6 20 11 7 21 12 8 26 15 9 27 16 22 23 24 25;
$N_2, r_e = 2.063\,76$ Å: 14 19 1 17 5 3 13 28 2 18 10 4 6 20 11 7 21 12 8 26 15 9 27 16 22 23 24 25;
C_2H_4, flat: 37 1 31 14 20 6 24 11 30 3 38 2 35 21 22 10 25 13 34 4 5 32 15 23 12 7 36 19 26 16 8 27 17 9 33 18 28 29;
C_2H_4, twisted: 38 37 35 33 32 31 25 23 22 21 20 1 14 6 10 3 11 4 12 2 26 36 28 17 7 15 18 9 8 19 5 13 29 16 24 34 30 27.
These numbers refer to the restricted Hartree–Fock orbitals arranged by energy, i.e., 1 is the lowest Hartree–Fock orbital. In the lattice, where calculations were performed with spin–orbitals, the pairs of equivalent α and β orbitals were placed in pairs $\alpha\beta\alpha\beta\cdots$.

[1] K. G. Wilson, Rev. Mod. Phys. **55**, 583 (1983).
[2] K. G. Wilson, Rev. Mod. Phys. **47**, 773 (1975).
[3] S. R. White, Phys. Rev. Lett. **69**, 2863 (1992).
[4] S. R. White and R. M. Noack, Phys. Rev. Lett. **68**, 3487 (1992), and references therein.
[5] G. Fano, F. Ortolani, and L. Ziosi, J. Chem. Phys. **108**, 9246 (1998).
[6] S. R. White and R. L. Martin, J. Chem. Phys. **110**, 4127 (1999).
[7] S. Daul, I. Ciofini, C. Daul, and S. R. White, Int. J. Quantum Chem. **79**, 331 (2000).
[8] A. O. Mitrushenkov, G. Fano, F. Ortolani, R. Linguerri, and P. Palmieri, J. Chem. Phys. **115**, 6815 (2001).
[9] N. Ben-Amor, S. Evangelisti, D. Maynau, and E. Rossi, Chem. Phys. Lett. **288**, 348 (1998), and references therein.
[10] S. R. White, Phys. Rev. B **48**, 10345 (1993).
[11] R. G. Parr, *Quantum Theory of Molecular Electronic Structure* (Benjamin, New York, 1963).
[12] T. Xiang, Phys. Rev. B **53**, 10445 (1996).
[13] E. R. Davidson, J. Comput. Phys. **17**, 87 (1975).
[14] S. R. White, Phys. Rev. Lett. **77**, 3633 (1996).
[15] E. Cuthill and J. McKee, in Proceedings of the 24th National Conference of the ACM, 1969.
[16] J. Liu and A. Sherman, SIAM (Soc. Ind. Appl. Math.) J. Numer. Anal. **13**, 198 (1975).
[17] W. L. Briggs, *A Multigrid Tutorial* (Soc. Ind. Appl. Math., Philadelphia, 1987).
[18] W. Kohn, 1999, remark made at the 8th International Conference on the Applications of Density Functional Theory to Chemistry and Physics, Rome.
[19] O. Legeza and G. Fath, Phys. Rev. B **53**, 14349 (1996).
[20] G. K.-L. Chan, P. W. Ayers, and E. S. Croot III, Phys. Rev. Lett. (submitted).
[21] K. Okunishi, Y. Hieida, and Y. Akutsu, Phys. Rev. E **59**, R6227 (1999).
[22] J. Stanton, J. Gauss, J. Watts, W. J. Lauderdale, and R. J. Bartlett, Int. J. Quantum Chem. **S26**, 879 (1992), ACES II.
[23] C. W. Bauschlicher and P. R. Taylor, J. Chem. Phys. **85**, 2779 (1986).
[24] T. H. Dunning, Jr., J. Chem. Phys. **53**, 2823 (1989).
[25] R. Bartlett, J. Watts, S. Kucharski, and J. Noga, Chem. Phys. Lett. **165**, 513 (1990).
[26] H. Larsen, J. Olsen, P. Jorgensen, and O. Christiansen, J. Chem. Phys. **113**, 6677 (2000).
[27] T. H. Dunning, Jr. and R. J. Harrison, J. Chem. Phys. **96**, 6796 (1992).
[28] R. Krishnan, J. Binkley, R. Seeger, and J. Pople, J. Chem. Phys. **72**, 650 (1980).
[29] M. Frisch, J. Pople, and J. Binkley, J. Chem. Phys. **80**, 3265 (1984).
[30] S. R. Gwaltney, C. D. Sherrill, M. Head-Gordon, and A. I. Krylov, J. Chem. Phys. **113**, 3548 (2000).

[31] Since the decimation procedure completely neglects the environment, it is only optimal in the limit of infinitely weak coupling between system and environment.
[32] By this we mean, exponentially difficult with the number of particles.
[33] In the cases we are considering, the characteristic length can be identified if the quantity has some decay like $e^{-\lambda r}$, in which case λ is the characteristic length. One can see this as meaning that one can roughly replace the exponential by a step function of length λ.