A Mixed-Precision Implementation of the Density Matrix Renormalization Group

Yingqi Tian, Zhaoxuan Xie, Zhen Luo,* and Haibo Ma*

School of Chemistry and Chemical Engineering, Nanjing University, Nanjing 210023, China

E-mail: luozhen@nju.edu.cn; haibo@nju.edu.cn

Abstract

Using the mixed precision strategy to optimize quantum chemistry codes has been proved promising in saving computational cost and maintaining chemical accuracy. Here, an efficient mixed-precision density matrix renormalization group (DMRG) scheme, containing a two-level mixed-precision hierarchy, is developed and demonstrated. At the coarse-grained level, based on the discovery that the single-precision orthogonalization may cause the DMRG generate a totally wrong answer, a feasible single-precision-sweep DMRG method with double-precision orthogonalization process is implemented. At the fine-grained level, a mixed-precision diagonalization algorithm is developed. This algorithm runs specific operations in the single-precision while preserving double-precision accuracy. Combining these two method, a hybrid mixed-precision scheme is presented. By applying this scheme, the DMRG single-point energy calculations are accelerated up to 131%. Mixed-precision DMRG yielded energies are accurate and deviate less than 0.01 kcal/mol compared with standard DMRG calculations.

1 Introduction

The density matrix renormalization group (DMRG) method was firstly developed by White to solve the strongly correlated quantum lattice models.[1,2] Later, this method was introduced into the quantum chemistry[3–9] to deal with the strongly correlated states in molecular electron structures, which helped DMRG method to stand out of other multi-reference methods by its attractive potential of handling large active space. A relatively large active space containing up to 108 orbitals[10] can be processed by the DMRG method, whereas the full configuration interaction (full-CI) method can only handle around 20 orbitals. This is because the computational cost of the full-CI method grows exponentially with the increase of the system size, resulting an unaffordable cost with this moderate size. On the other hand, the DMRG method holds a polynomial complexity of $O(N^4m^2) + O(N^3m^3)$, where the N indicating the number of the correlated orbitals and the m (usually in a few hundreds or thousands) representing the bond dimension of the wavefunction tensors, making DMRG a more conspicuous method than other methods. However, the computational cost of the DMRG method is still very expensive for practical chemistry applications.

Using single or mixed precision is a potential approach to increase the computational efficiency and reduce the computational cost. Because the single-precision (SP) floating point data form only takes 32 bits, as defined by the IEEE 754 standard,[11] to represent a certain number. However, 64 bits are needed for double-precision (DP) data form. This difference means that the memory usage of same data is saved by 50% when single precision is applied comparing with the double precision. Additionally, the memory access speed will be doubled. On the CPU platform, when the vectorization instruction set, such as streaming SIMD extensions (SSE) or advanced vector extensions (AVX) is used, the computational speed is also be doubled, resulting in a theoretical speed-up of 2. In modern architecture of clusters and supercomputers, the accelerators, including the graphic processing unit (GPU) and the field-programmable gate array (FPGA), play a more and more important role for obtaining high computing performance. And these special devices have been widely applied

in quantum chemistry methods, including the Hartree-Fock (HF) method[12–16] the integral evaluation,[17–20] and many post-HF methods.[21–24] Because the different design of the GPU hardware architecture, the speed-up for single precision is even larger on some GPUs.

Because of the potential advantages, using a relatively lower precision is a common approach for performance optimization in machine learning[25] and high-performance computing[26] and has also been applied in various scientific areas such as numerical linear algebra,[27–31] molecular dynamics,[32] and lattice quantum chromodynamics.[33] In quantum chemistry, some mixed-precision (MP) electron repulsion integral evaluation methods[34,35] have been implemented. The main idea is that the upper bound of the integrals can be previously identified. Thus, integrals with a rather small upper bound can be evaluated in single precision, while others in double precision. To make better use of this dynamic precision scheme, Parrish *et al.*[36] developed a "difference self-consistent field (dSCF)" approach, in which a difference density matrix instead of the actual density matrix is used. The difference density matrix is a numerically smaller matrix, resulting in a smaller integral upper bound. With the support of these mixed-precision integral methods, some mixed-precision Hartree-Fock and density functional theorey (DFT)[37] methods have been developed, as well as some single- or mixed-precision post-Hartree-Fock methods, including the MP2 method within resolution-of-the-identity (RI)[38] and Cholesky decomposition (CD)[39] schemes, coupled cluster with single and double excitation (CCSD)[40] and perturbative triples correction (CCSD(T)).[41] In 2018, Pokhilko *et al.*[42] presented an implementation of coupled cluster and equation-of-motion coupled-cluster with single and double excitations (EOM-CCSD) in single precision. They claim that the chemical accuracy can be maintained merely using single precision, while the full double precision accuracy can be recovered with a few cleanup iterations. Wang *et al.*[43] followed this idea and implemented the CCSD and CCSD(T) method with pure single precision on consumer GPUs and achieved 4–14 speed-up for CCSD calculation and 12–20 speed-up for (T) correction.

Since using single precision has several potential advantages, an intrinsic question arise,

can single precision be implemented for the DMRG numerical calculations? Actually, simply porting the double precision code into single precision is a wrong answer. The single-precision transformation should be carefully designed and examined. In this study, we investigated the DMRG procedure, and developed two types of mixed-precision transformation for DMRG method at two different levels. At a relatively coarse-grained level, we identified the Gram-Schmidt orthogonalization process as a fundamental step, that causes a tremendous loss in significant number and made the full single-precision DMRG strategy to fail. At the fine-grained level, we mainly focus on the mixed-precision transformation of the Davidson diagonalization method and developed an accuracy-preserved method. By combining these two methods, we developed a hybrid mixed-precision implementation for DMRG. In this article, we will first introduce the fundamental procedure of the DMRG method and these two mixed-precision methods in Section II. The benchmark result and some discussions are presented in Section III. In Section IV, some conclusions of this article are summarized.

2 Methods

2.1 DMRG method

To demonstrate the mixed-precision scheme, we first introduce the basis procedure of the DMRG calculation. Here we mainly focus on the algorithm and workflow, therefore, for detailed principles, we refer the interested readers to these nice review articles.[44–49]

The DMRG method is based on a wave function ansatz that the full-CI wave function can be represented by a production of a series of tensors, which is called the matrix product state (MPS) ansatz

$$
\begin{aligned}
|\Psi\rangle &= \sum_{\{n\}} \Psi^{n_1 n_2 \cdots n_k} |n_1 n_2 \cdots n_k\rangle \\
&= \sum_{\{n\}} \psi^{n_1} \psi^{n_2} \cdots \psi^{n_k} |n_1 n_2 \cdots n_k\rangle .
\end{aligned}
\tag{1}
$$

In quantum chemistry, the local state on a MO could be one of the four possible occupation states: the doubly-occupied configuration $|\uparrow\downarrow\rangle$, the spin-up singly-occupied configuration $|\uparrow\rangle$, the spin-down singly-occupied configuration $|\downarrow\rangle$ and the unoccupied configuration $|-\rangle$, *i.e.*

$$
\{n\} = \{|\uparrow\downarrow\rangle, |\uparrow\rangle, |\downarrow\rangle, |-\rangle\} . \tag{2}
$$

The singular value decomposition (SVD) method can be applied to generate these MPSs from the full-CI wavefunction. Then, these MPSs are truncated according to the significance of corresponding singular value. The truncated MPSs will have a dimension of $4 \times m \times m$, where the m is the truncation bond dimension of the wavefunction tensors. With this ansatz, the Hamiltonian operators are also represented by tensors

$$
\hat{H} = \sum_{\{n,n'\}} \hat{H}_1 \hat{H}_2 \cdots \hat{H}_k |n_1 n_2 \cdots n_k\rangle \langle n'_1 n'_2 \cdots n'_k| . \tag{3}
$$

And these operators are called the matrix product operators (MPOs). The MPOs and MPSs are combined together in a lattice tensor network form. Each site contains a MPO, a bra MPS and the corresponding ket MPS. The main purpose of the DMRG method in *ab-initio* quantum chemistry is to find the ground state energy E_0, which is the lowest energy of the system

$$
E_0 = \min_{|\Psi\rangle} \frac{\left\langle \Psi | \hat{H} | \Psi \right\rangle}{\langle \Psi | \Psi \rangle}. \tag{4}
$$

It is difficult to optimize all ψ^n tensors simultaneously, and accordingly an iterative

optimization scheme is adopted in DMRG through a few sweeps from left to right and from right to left. In each step of a forward or backward sweep, one MPS tensor is variationally optimized while others remain unchanged. This method is the single-site DMRG method. We can expand the variational space by optimize two site at a time to enhance the ability of escape the local minima. In this article, the mixed-precision scheme of this two-site DMRG method is discussed. The tensor lattice of the two-site DMRG method is presented in Fig. 1.

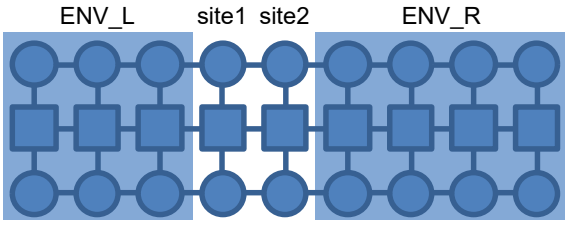


Figure 1: The tensor lattice of the two-site DMRG method. The dots represent the MPSs and the blocks represent the MPOs.

Conventionally, *ab-initio* quantum chemistry software apply the double precision as default standard. This double precision sweep can be represented with a simplified lattice form in Fig. 2. The detailed work flow of the double precision two-site DMRG algorithm is presented as Algorithm 1. The operations within each sweep step is presented in Algorithm 2. In Algorithm 2, the Lanczos algorithm is presented as an example of the diagonalization method, but other diagonalization methods (such as the Davidson algorithm or the Jacobi-Davidson algorithm) could also be used. The algorithm we presented can be transformed into other diagonalization methods by applying different form of the transformation function $T(x)$ specifically. Similarly, the following mixed-precision optimization is introduced with the Lanczos algorithm, but also can be applied for other diagonalization methods.

Figure 2: The simplified DMRG lattice for double precision sweep.

Algorithm 1 DMRG algorithm.

read in structure
generate MPO
generate initial guess MPS
pre-compute ENV
while DMRG unconverged **do**
    sweep left to right
    sweep right to left
end while
finish calculate and output energy

Algorithm 2 Sweep algorithm.

for each site **do**
    generate two-site MPS and two-site MPO
    procedure DIAGONALIZATION:
        set $v_0 = \left|\psi^{k,k+1}\right\rangle$
        contract tensor network to generate $w_0 = \hat{H}v_0$
        $v_1 = orthogonalize(v_0, w_0)$
        construct list $W = \{w_0\}$
        construct list $V = \{v_0, v_1\}$
        while diagonalization unconverged **do**
            $w_i = \hat{H}v_i$
            $H_k = V \cdot W$
            $\lambda =$ first eigen value of H_k
            $y =$ first eigen vector of H_k
            $x = V \cdot y$
            $r = \hat{H}x - \lambda x$
            $t = T(r)$
            $v_{i+1} = orthogonalize(V, t)$
            return x as new two-site MPS if converged
        end while
    end procedure
    decompose two-site MPS with SVD method
    update ENV
end for

2.2 Coarse-grained level method

The DMRG apply this sweep iterative method to approximate the ground state energy. We think that the first several steps can be evaluated in low precision while maintaining the double-precision convergent tendency. This leads to our coarse-grained level mixed-precision

DMRG method, which is introduced in Table 3.

Algorithm 3 Mixed-precision sweep algorithm.

```
read in structure
generate MPO
generate initial guess MPS
convert MPS and MPO to single precision
pre-compute ENV
while single-precision DMRG unconverged do
    sweep left to right
    sweep right to left
end while
convert MPS and MPO to double precision
while double-precision DMRG unconverged do
    sweep left to right
    sweep right to left
end while
finish calculate and output energy
```

This idea is simple and straightforward, however, it is infeasible to simply transform the double-precision codes into single-precision. Because the orthogonalization procedure fails in single-precision form. The numerical reason is: consider the Gram-Schmidt orthogonalization of two vector v_1 and v_2:

$$v_2' = v_2 - \frac{(v_2 \cdot v_1)}{(v_1 \cdot v_1)} v_1 \tag{5}$$

The most important step is the subtraction of two numbers that are extremely close to each other, hence resulting in a great loss of significant digits. Consequently, the Gram-Schmidt orthogonalization should be conducted under double-precision. In fact, using a higher precision in the orthogonalization process to improve convergence has already been applied by Alvermann, *et al.*[29] to achieve a robust orthogonalization in Jacobi-Davidson method. Even though they were applying a two-double projection in a double-precision diagonalization. This method was also been applied by Carson, *et al.*[31] in the mixed-precision s-step Lanczos method for the same reason and in a similar way to Alvermann. It was also found that if the evaluation of H_k remain in double-precision, the accuracy of the result will

be improved. In fact, in the diagonalization procedure, the $\hat{H}v$ calculation consumes more than 90% of the computational cost. Thus, other operations whether in double-precision or in single-precision, have little influence on the overall performance. Moreover, the additional conversion operations will have little effect of the entire computational cost.

Algorithm 4 Single-precision sweep algorithm. In the algorithm table, the upper mark "s" and "d"are used to identify the single- and double-precision data. At the right end of each line, we marked the single-precision operation as "S", double-precision as "D" and the conversion operation as "C".

for each site **do**
  generate two-site MPS and two-site MPO ▷ S
  procedure SP DIAGONALIZATION:
    set $v_0^s = \left|\psi^{k,k+1}\right\rangle$ ▷ S
    $w_0^s = \hat{H}^s v_0^s$ ▷ S
    convert v_0^s to v_0^d ▷ C
    convert w_0^s to w_0^d ▷ C
    construct list $W^d = \{w_0^d\}$ ▷ D
    $v_1^d = orthogonalize(v_0^d, w_0^d)$ ▷ D
    construct list $V^d = \{v_0^d, v_1^d\}$ ▷ D
    convert v_1^d to v_1^s ▷ C
    while diagonalization unconverged **do**
      $w_i^s = \hat{H}^s v_i^s$ ▷ S
      convert w_i^s to w_i^d ▷ C
      $H_k = V^d \cdot W^d$ ▷ D
      $\lambda =$ first eigen value of H_k ▷ D
      $y =$ first eigen vector of H_k ▷ D
      $x = V^d \cdot y$ ▷ D
      $r = W^d \cdot y - \lambda x$ ▷ D
      $t = T(r)$ ▷ D
      $v_{i+1}^d = orthogonalize(V^d, t)$ ▷ D
      convert v_{i+1}^d to v_{i+1}^s ▷ C
      return x as new two-site MPS if converged ▷ S
    end while
  end procedure
  decompose two-site MPS with SVD method ▷ S
  update ENV ▷ S
end for

After the single-precision sweep converged, the single-precision data (MPO, MPS and ENVs) are transformed into double precision. Subsequently, the double-precision sweep starts as in a standard DMRG calculation. The only remaining question is to determine

the switching point of these mixed-precision scheme, *i.e.* how to determine the convergence of the single-precision sweep. According to the IEEE 754 standard, the 32-bit floating point data form has a relative error of around 10^{-7}. Thus the convergence threshold of the single-precision sweep is around 10^{-4}–10^{-5}, based on the different systems. However a manually-appointed threshold number is unascertainable. Consequently, we choose another way to determine the convergence. In the first several single-precision sweep, the energy sustains to decrease sweep by sweep. But when the convergence point is reached, the energy starts to fluctuate. Therefore, if the energy of the current sweep is larger than the last sweep, one can determine that the convergence point has been reached.

2.3 Fine-grained level method

As mentioned in the last subsection, the convergence threshold of the single-precision sweep is around 10^{-4}–10^{-5} Hartree, which means that the overall accuracy of the single-precision DMRG method is around 10^{-4}–10^{-5} Hartree. Therefore, if higher accuracy is required (which is common in DMRG calculations), the following clean-up sweeps, *i.e.* the double-precision sweep, must be conducted, undermining the overall acceleration.

Fortunately, the clean-up sweeps can be optimized with the mixed-precision approach, which lead to our fine-grained level method. To maintain the double-precision accuracy, the MPS data, the MPO data, the ENV data and the related operations, including the SVD decomposition and the update of the ENV, should stay in double precision. But the diagonalization procedure can be transformed into the mixed-precision while still remain the double-precision accuracy. In fact, in the DMRG sweep, the diagonalization procedure consumed around 60%-90% of the total computational time, making it the most time-consuming step in the DMRG sweep. Therefore, this mixed-precision transformation of the diagonalization procedure can obtain a satisfying performance speed-up in double-precision sweeps, and maintain the double-precision accuracy.

At the fine-grained level, we developed a mixed-precision diagonalization method, which

can make use of the advantages of the single precision and achieve double-precision accuracy. First, let us reexamine the diagonalization procedure. The result $x = V \cdot y$, can be seen as the weighted summation of v_i, and should be not much different from the initial v_0 and v_1. Thus, the other v_i can be seen as the correction of these initial guesses, and should be not large. Then, if we make this initial guess (*i.e.* v_0 and v_1) evaluated in double-precision, and other corrections (*i.e.* $v_2, v_3, \dots$) evaluated in single precision, we can accelerate the overall calculation speed and maintain the double-precision accuracy. This mixed-precision diagonalization method is shown in Table 5. In this method, the orthogonalization procedure should also be calculated in double precision, as the coarse-grained level method, to guarantee the correctness.

Algorithm 5 Mixed-precision diagonalization algorithm.

procedure MP DIAGONALIZATION:
  set $v_0^d = \left|\psi^{k,k+1}\right\rangle$ ▷ D
  $w_0^d = \hat{H}^d v_0^d$ ▷ D
  contract list $W^d = \{w_0^d\}$ ▷ D
  $v_1^d = orthogonalize(v_0^d, w_0^d)$ ▷ D
  construct list $V^d = \{v_0^d, v_1^d\}$ ▷ D
  convert $\hat{H}^d$ to $\hat{H}^s$ ▷ C
  while diagonalization unconverged **do**
    convert v_i^d to v^s ▷ C
    $w^s = \hat{H}^s v^s$ ▷ S
    convert w^s to w_i^d ▷ C
    $H_k = V^d \cdot W^d$ ▷ D
    $\lambda =$ first eigen value of H_k ▷ D
    $y =$ first eigen vector of H_k ▷ D
    $x = V^d \cdot y$ ▷ D
    $r = W^d \cdot y - \lambda x$ ▷ D
    $t = T(r)$ ▷ D
    $v_{i+1}^d = orthogonalize(V, t)$ ▷ D
    return x as new two-site MPS if converged ▷ D
  end while
end procedure

With the combination of both the coarse-grained level method and the fine-grained level method, we can construct a hybrid mixed-precision scheme. For this hybrid scheme, the DMRG calculation starts with the single-precision sweep. After several sweeps, it reaches

the single-precision convergence point. Consequently, the data is transformed into double-precision. Subsequenly, the mixed-precision diagonalization method is applied until the required accuracy is reached. A simplified flow chart of this hybrid scheme is illustrated in Fig 3.

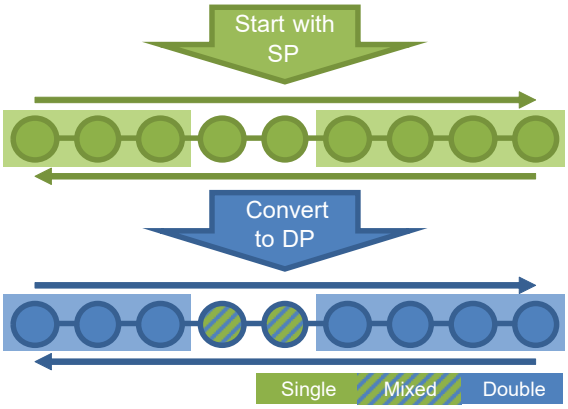


Figure 3: The hybrid mixed-precision DMRG scheme.

3 Result and discussion

3.1 Basic benchmark result

Several molecules are selected based on their intra- inter-molecular interaction representativity and atomic configuration complexity to benchmark the accuracy and performance of the mixed-precision DMRG scheme. Those are benzene molecule, two stretched water(H_2O) molecules (O-H bond at 1.2Å and 1.5Å), a carbon dimer(C_2) and a chromium dimer(Cr_2). The computing node is constructed with two Intel Xeon Gold 6126 12-core CPUs. The operation system is CentOS Linux under release 7.4.1708. The mixed-precision scheme is implemented in the Kylin quantum chemistry software[50] and the compiler is GCC (version 10.2.0). All of the tensor operations include the contraction, SVD decomposition and eigen solver is supported by Intel MKL library (version 2019.0.117). In the benchmark test, the 6-31G basis set was used for the benzene molecule and the Cr_2 molecule, the cc-pVTZ basis

set was applied for the water molecule and the cc-pVDZ basis set for the carbon dimer. The convergence threshold of these tests is 10^{-6}.

Table 1: System information and benchmark results for energy accuracy. The energy results in this table are in atomic units, and the differences of energy are in 10^{-3}kcal/mol. E_{dp} is the energy evaluated by double-precision DMRG. E_{sp} is evaluated by full single-precision DMRG. And E_{mp} is evaluated by the mixed-precision DMRG. Δ_{sp} is the difference between E_{dp} and E_{sp} (*i.e.* $|E_{dp} - E_{sp}|$). Similarly, $\Delta_m p = |E_{dp} - E_{mp}|$.

system	space	m	E_{dp}	E_{sp}	Δ_{sp}	E_{mp}	Δ_{mp}
benzene	(24e,24o)	5k	-230.7487820	-230.7487725	6.0	-230.7487813	0.5
benzene	(28e,28o)	3k	-230.7842899	-230.7842534	22.9	-230.7842875	1.5
H_2O(1.2Å)	(10e,40o)	2k	-76.1914944	-76.1917189	140.9	-76.1914951	0.5
H_2O(1.5Å)	(10e,58o)	1k	-76.1519778	-76.1521174	87.6	-76.1519929	9.5
C_2	(12e,28o)	2k	-75.7316808	-75.7318000	74.8	-75.7316818	0.6
C_2	(12e,28o)	3k	-75.7318504	-75.7319113	38.2	-75.7318505	0.0
Cr_2	(24e,24o)	2k	-2085.9825336	-2085.9828607	205.3	-2085.9825356	1.2
Cr_2	(24e,24o)	3k	-2085.9827180	-2085.9833185	377.0	-2085.9827184	0.6

The detailed information of the testing systems and the energy accuracy benchmark results of these systems are presented in Table 1. In this test, we benchmarked the mixed-precision scheme we presented in this article and compared with the original double-precision DMRG method which is implemented in the Kylin software. We also checked the energy result of the full single-precision DMRG method (*i.e.* all sweeps were calculated with the single-precision sweep method). The mixed-precision DMRG scheme shows good agreement with the double-precision result, thus the mixed-precision calculated energy deviates from double-precision for less than 0.01 kcal/mol. However, for most cases, the full single-precision DMRG may not deliver good result, especially for systems with complex orbital configuration like for the Cr_2 molecule. In fact, when transition metal elements are involved, the diagonalization and the DMRG sweep are harder to converge, which may affect the effectiveness and the performance of the mixed-precision scheme. This issue will be further investigated and discussed later. Generally speaking, the accuracy of the mixed-precision scheme will increase with the growth of m, which is the bond dimension of the MPS. As shown in the Fig. 4, when the bond dimension is larger than 1500, the error of the mixed-precision scheme

is under 0.001 kcal/mol.

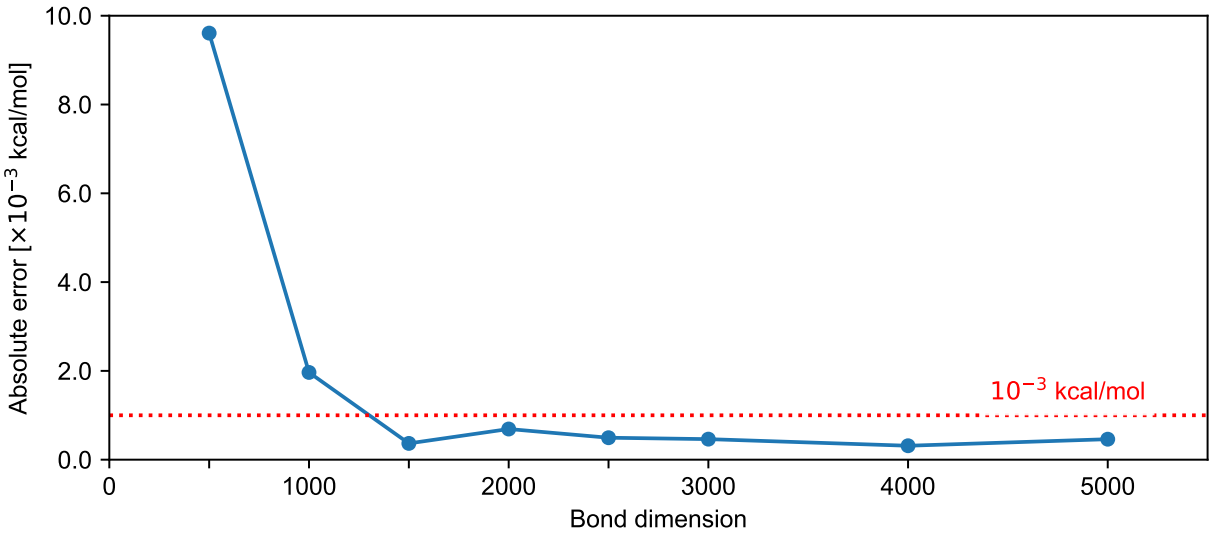


Figure 4: Energy error ($\times 10^{-3}$kcal/mol) of mixed-precision DMRG to double-precision DMRG for different bond dimensions. Benzene as benchmark system using the 6-31G basis set with a space of (24e,24o).

In Fig. 5, the benchmark result of the computational performance of the double-precision DMRG method and the mixed-precision DMRG scheme is presented, as well as the corresponding speed-up between these two approach. The testing systems are the same to those in Table 1. The performance of the mixed-precision scheme is, as expected, better than the double-precision method. Even a 2.31 speed-up is achieved for the benzene system with the active space of (24e,24o) and a bond dimension of 5000. This speed-up is larger than the theoretical upper bond of the speed-up for mixed-precision optimization. This is because the total number of sweeps is decreased for the mixed-precision scheme. Generally, the number of sweeps for the mixed-precision scheme is slightly different from the double-precision method, but will not be far from it.

3.2 Benchmarking different mixed-precision methods

In the performance test of the previous subsection, the mixed-precision approach for the chromium dimer system is different from other systems. This is because, for this special

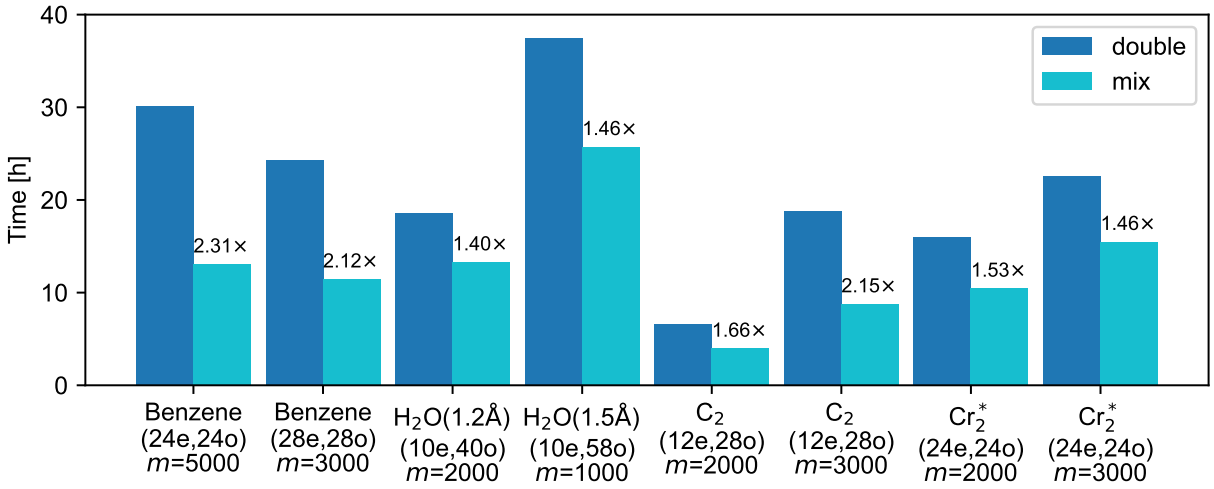


Figure 5: Benchmark result of the computational time of the double-precision method and the mixed-precision scheme. This is the time for the entire DMRG calculation, instead of the time for one sweep. The corresponding speed-up for each test is attached on top of the bar of the mixed-precision scheme. (*): Note that only mixed-diagonalization method was applied for chromium dimer system. This is because when transition metal elements are involved, the performance of the single-precision sweep will be slower than the mixed-diagonalization optimized double-precision sweep.

case, the fine-grained level method (mixed-precision diagonalization) surprisingly performed better than the coarse-grained level method (single-precision sweep). The deeper reason is that when dealing with the Cr_2 system, the diagonalization process is difficult to converge for the single-precision sweep method. To further investigate this phenomenon, we tested three different settings of the mixed-precision scheme, which are: (1) only the coarse-grained level method is applied; (2) only the fine-grained level method is applied; (3) both of these methods are applied. We also conducted a full double-precision test as a baseline and a full single-precision test as a contrast. Two different types of the system were chosen for this test. Besides the Cr_2 system, we also choose the benzene molecule as an example for those systems with no transition metal element. Both of these systems are using 6-31G basis set with the active space of (24e,24o). The bond dimension of Cr_2 is 2000 and benzene is 3000. The result of these tests are presented in Table 2 and 3.

From these two tables, we can infer that the mixed-precision diagonalization method performed better than other two methods for the Cr_2 system, while the hybrid mixed-

Table 2: Benchmark test for Cr_2, 6-31G, (24e,24o), $m = 2000$. Energy in atomic unit, energy difference in 10^{-3} kcal/mol, time in hour.

Method	E	Δ	T_{sweep}	speed-up$_{sweep}$	T_{total}	speed-up$_{total}$
Full double	-2085.9825336	0.0	2.06	1.00	15.94	1.00
Full single	-2085.9828607	205.3	1.77	1.17	3.87	4.12
Single sweep	-2085.9825371	2.2	1.77	1.17	18.51	0.86
Mixed diag.	-2085.9825356	1.2	1.18	1.74	10.43	1.53
Mixed	-2085.9825373	2.3	1.78/1.21	1.16/1.70	12.39	1.29

Table 3: Benchmark test for benzene, 6-31G, (24e,24o), $m = 3000$. Energy in atomic unit, absolute error in 10^{-3} kcal/mol, time in hour.

Method	E	Δ	T_{sweep}	speed-up$_{sweep}$	T_{total}	speed-up$_{total}$
Full double	-230.7484449	0.0	0.99	1.00	10.57	1.00
Full single	-230.7484264	11.6	0.45	2.21	2.58	4.11
Single sweep	-230.7484456	0.5	0.45	2.21	7.05	1.50
Mixed diag.	-230.7484474	1.6	0.55	1.80	8.53	1.24
Mixed	-230.7484456	0.5	0.45/0.55	2.22/1.81	5.85	1.81

precision scheme, combining both single-precision sweep and mixed-precision diagonalization, performed better than other two methods for the benzene system. For the Cr_2 system, the full single-precision DMRG may yield inaccurate energy, and have a poor performance improvement. The reason is, as said in the previous paragraph, that the single-precision diagonalization process has a big convergence issue. In each sweep, the total iteration number for diagonalization in single-precision is much larger than in double-precision. For example, on site 16, double-precision diagonalization took 15 iterations with 295.7 seconds, comparing to 31 iterations for the single-precision diagonalization, consuming 314.8 seconds. Hence, the speed-up for single-precision sweep method in one sweep is much lower. On the contrary, a 2.21 speed-up for the same method in one sweep is achieved when there is no convergence issue in the benzene system test. This convergence issue also effected the total number of sweeps. The double-precision DMRG method converged in 10 sweeps, while the hybrid mixed-precision scheme, and the single-precision sweep approach converged in 12 sweeps. On the other hand, the mixed-precision diagonalization method also converged in 10 sweeps, showing a good consistency with the double-precision method. In summary, the single-

precision diagnolization convergence issue become significant when a transition metal element is involved. While the mixed-precision diagonalization method does not have this issue. As results, the performance of single-precision sweep is slower than the mixed-precision diagonalization method for the Cr_2 system. The mixed-precision diagonalization method preserved the result in double-precision accuracy.

On the contrary, this convergence issue in diagonalization did not appear in the test of the benzene system, and other systems with no transition metal element. Thence, the performance of the single-precision sweep method for one sweep is much better than the double precision method. The speed-up of the single-precision sweep method is slightly larger than 2, the theoretical upper bond of the single-precision optimization. This is reasonable and can be regarded as common for single-precision optimization, when it is applied on a compute-intensive task. The reason is that in the single-precision sweep method, almost all of the data and operations are transformed into single-precision. Consequently, it is highly possible to reach the theoretical upper bond. As the single-precision data form saved memory space, more data can be stored in the CPU cache and the CPU cache hit rate is increased. Therefore, the overall computational performance can be slightly higher than the theoretical upper bond. However, we have to point out that although the performance of the single-precision sweep is very high for one sweep, the speed-up of total DMRG calculation is not as good as expected. This is because of the clean-up sweeps that are required to reach the double-precision accuracy. Therefore, the clean-up sweeps must run in double-precision mode, whether using full double-precision or using mixed-precision diagonalization method. Consequently, these clean-up sweeps will delay the overall DMRG calculations. On the other hand, if only the mixed-precision diagonalization method is applied, all the sweeps will run with fine-grained mixed-precision method, and there will be no clean-up sweeps. But the first several sweeps are set to run with a relatively small bond dimension, and will increase sweep by sweep. At this circumstance, the conversion process in the mixed-precision diagonalization method will consume much more time than other operations. resulting in a

deceleration of overall performance. This is why the overall speed-up of the mixed-precision method is not as good as the speed-up of a single sweep. Hence, the mixed-precision scheme, combining both of these two level mixed-precision methods, can achieve the best performance of these three mixed-precision settings, and also maintain the double-precision accuracy.

3.3 Performance for different space size and bond dimension

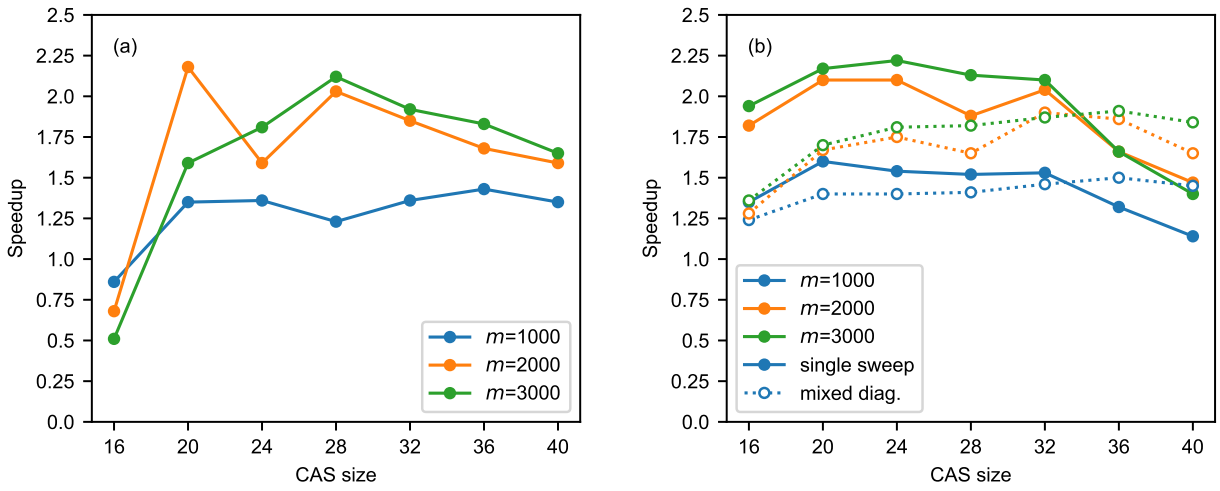


Figure 6: speed-ups of the mixed-precision scheme with different active space sizes and bond dimensions for the benzene molecule with 6-31G basis set. (a) The speed-up of total DMRG calculations. (b) The speed-up of one DMRG sweep. In this case, both the single-precision sweep method and the mixed-precision diagonalization method are presented in different type of the lines and the markers.

The performance of the mixed-precision scheme varies for different size of the active space and the bond dimension. Hence, further investigations on the performance and speed-up of the mixed-precision scheme is conducted on the benzene molecule and the water molecule with a bond length of 1.5 Å. In Fig. 6 and 7 the speed-ups of the mixed-precision scheme with different active space sizes and bond dimensions are presented. From these figures, it can be concluded that, the mixed-precision method performs well when the bond dimension is relatively large, especially when m is equal to and larger than 2000. For most cased in these two benchmark tests, the speed-up grows as the bond dimension grows. This is clearer on the speed-up of one single sweep, shown as the (b) sub-figure in each figure. Even though

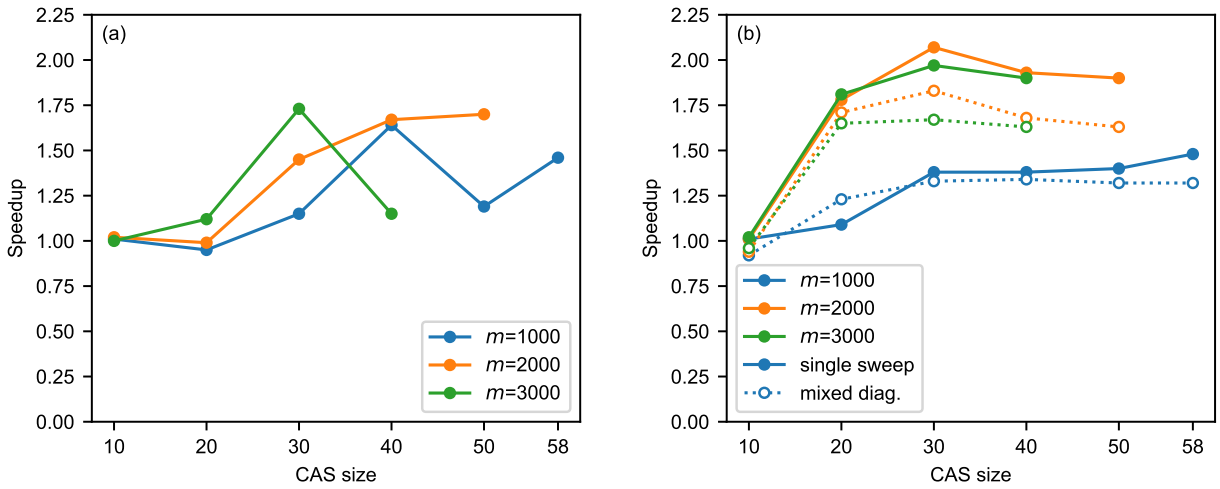


Figure 7: speed-ups of the mixed-precision scheme with different active space sizes and bond dimensions for the water molecule (1.5 Å) with cc-pVTZ basis set. (a) The speed-up of total DMRG calculations. (b) The speed-up of one DMRG sweep. In this case, both the single-precision sweep method and the mixed-precision diagonalization method are presented in different type of the lines and the markers.

the speed-up for $m = 3000$ is slightly smaller than $m = 2000$ for the water case, both of them are clearly much larger than $m = 1000$.

As for the the active space size aspect, the speed-up will firstly increase with the growth of the active space size, and then decrease. At the beginning, with a relatively small space, the total computational time is short. Therefore there is no much different of the total computational time between the double-precision method and the mixed-precision scheme, even though the performance of one single sweep is increased. Thus, with the growth of the system, the speed-up, both in total computational time and in a single sweep, will increase significantly. After reaching the maximum point, around 28–40 active orbitals for different system, the speed-up will slightly decrease. There are several reasons for this decrease of performance. The first and foremost reason is the performance of the matrix product operations, which depends on the implementation of the MKL library, that is slightly reduced with the growth of the matrix size when the size is large enough. The second reason is when the system is too big, the total amount of data would be larger than the memory of our computing node. Therefore, one must store the unused ENVs on the disk, when we doing

the sweep, and load the corresponding ENV back to memory when it is required. This memory store and load operations delays the overall performance, as the data is stored in text form instead of binary form. Another reason is that, with the growth of the system, the ratio of computational cost for the ENV update operation will grow. But this operation is not optimized in the fine-grained level method, causing a slight reduction of the performance. Because of these reasons, the speed-up has a small drop when the active space size is large enough.

Another interesting phenomenon observed in the figures is that, in general, the speed-up of single-precision sweep method is higher than the speed-up of the mixed-precision diagonalization method. This is reasonable, that for the coarse-grained level method, most operations including the diagonalization, the SVD decomposition and the update of the ENV are transformed into single-precision mode. However, in the fine-grained level method, only several iterations of the diagonalization procedure are transformed. Hence, the coarse-grained level method is naturally and rightfully achieved a higher speed-up than the fine-grained level method when there is no convergence issue in diagonalization.

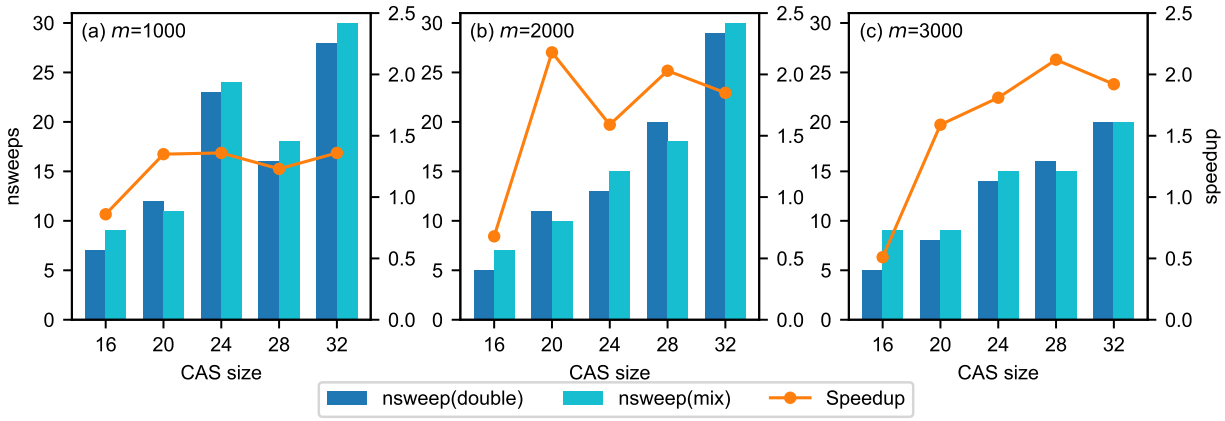


Figure 8: Total number of sweeps to convergence and the corresponding speed-up for benzene molecule with 6-31G basis set in different active space size and bond dimension.

To finish this section, we would like to investigate the convergence behavior in the DMRG sweep of the mixed-precision scheme at different circumstances. The benzene molecule was chosen as an example. The total number of sweeps to convergence for different active space

size and bond dimension is presented in Fig. 8 along with the corresponding performance speed-up. From the data in Fig. 8, we can see that the total number of sweeps for the mixed-precision scheme is tend to be a little bit larger than the double-precision method. In some special cases, the total number of sweeps is smaller. But most of the differences are no larger than 2 sweeps. Thus, it has no significant effect on the overall performance. After all, the computational speed of one sweep is increased considerably.

4 Conclusion

In this study, a feasible two-level mixed-precision DMRG scheme is implemented. At the coarse-grained level, a full single-precision sweep method, with the help of double-precision orthogonalization, was developed. The double-precision orthogonalization is the key point to make the full single-precision sweep successfully conducted and generate the right solution, however, it consumes little more computational resources. Thus the coarse-grained level method can achieve a rather high speed-up, even larger than the theoretical upper bond for some cases. However, it may face some convergence issues in diagonalization when dealing with systems containing transition metal elements. As for the fine-grained level method, a mixed-precision diagonalization method, that is free of the convergence issue in diagonalization, is provided. This method is mainly focus on the most time-consuming part of the DMRG sweep, which is the diagonalization procedure, specifically speaking, the $\hat{H}\psi$ calculation operation. With this mixed-precision diagonalization method, the double-precision accuracy is achieved. Nevertheless, the overall performance improvement is not as good as the coarse-grained level method. The only defect of the coarse-grained level method is that if one only apply the coarse-grained level method to run a full single-precision DMRG, the result may not be able to reach the required accuracy, as the numerical accuracy of the single-precision data form is only 10^{-7}, resulting a accuracy around 10^{-4}–10^{-5} Hartree. Thence, the clean-up sweeps are necessary to generate a higher accuracy, a common approach

in DMRG calculations. Therefore, the combination of these two methods is essential to construct a high accuracy and high performance mixed-precision DMRG scheme.

The benchmark results shows that the accuracy and the speed-up are well performed as expected. For all of these different systems, the accuracy stays within 0.01 kcal/mol. Specifically, when the bond dimension is larger than 1500, an error less than 0.002 kcal/mol can be achieved. The speed-up of the mixed-precision scheme is also ideal. Even a 2.31 speed-up was achieved for the benzene case with 6-31G basis set, active space (24e,24o), $m = 5000$. The chromium dimer is a special case, where the fine-grained level method performed better than the coarse-grained level method. This is because the coarse-grained level method may encounter a convergence issue in diagonalization, increasing the total number of diagonalization iterations, and delaying the overall performance. We think this convergence issue in diagonalization may be able to be fixed with other diagonalization methods such as the Jacobi-Davidson method, so that a better performance can be reached. We also discovered that the mixed-precision scheme prefers to be conducted on a relatively higher bond dimension, larger than 1000, and a reasonable active space size, from 24 orbitals to 40 orbitals.

As for the future work, some implementations and investigations are under consideration. The first quest is to find a possible way to fix the convergence issue in the diagonalization process. A potential approach is to implement some other diagonalization methods such as the Jacobi-Davidson method within single precision. Maybe the Jacobi-Davidson method has less issues than the Lanczos method. Moreover, maybe one can use the mixed-precision approach to solve the convergence problem. Although, in this work, the mixed-precision approach was only applied to achieve a better computational performance, the mix-precision approach can also be applied to achieve a better convergence. This approach has already been used by other numerical methods. Thence, we think a higher precision may be able to help dealing with the convergence issue for those systems with transition metal elements. Another work we plan to execute is porting this mixed-precision scheme onto the GPU

platform. We believe this mixed-precision scheme will perform even better on the GPU platform as the GPU architecture is more suitable for the single-precision calculations.

Acknowledgement

Financial support by the GHfund A (20220202,ghfund202202015504), the National Natural Science Foundation of China (No. 22073045), and the Fundamental Research Funds for the Central Universities are gratefully acknowledged. The authors thank Luis Vasquez, Yifan Cheng and Yinxuan Song for their helpful discussions.

References

(1) White, S. R. Density matrix formulation for quantum renormalization groups. *Phys. Rev. Lett.* **1992**, *69*, 2863–2866.

(2) White, S. R. Density-matrix algorithms for quantum renormalization groups. *Phys. Rev. B* **1993**, *48*, 10345–10356.

(3) Shuai, Z.; Brédas, J. L.; Pati, S. K.; Ramasesha, S. Quantum-confinement effects on the ordering of the lowest-lying excited states in conjugated chains. *Phys. Rev. B* **1997**, *56*, 9298–9301.

(4) White, S. R.; Martin, R. L. Ab initio quantum chemistry using the density matrix renormalization group. *J. Chem. Phys.* **1999**, *110*, 4127–4130.

(5) Chan, G. K.-L.; Head-Gordon, M. Highly correlated calculations with a polynomial cost algorithm: A study of the density matrix renormalization group. *J. Chem. Phys.* **2002**, *116*, 4462–4476.

(6) Moritz, G.; Hess, B. A.; Reiher, M. Convergence behavior of the density-matrix renor-

malization group algorithm for optimized orbital orderings. *J. Chem. Phys.* **2005**, *122*, 024107.

(7) Chan, G. K.-L.; Van Voorhis, T. Density-matrix renormalization-group algorithms with nonorthogonal orbitals and non-Hermitian operators, and applications to polyenes. *J. Chem. Phys.* **2005**, *122*, 204101.

(8) Kurashige, Y.; Yanai, T. High-performance ab initio density matrix renormalization group method: Applicability to large-scale multireference problems for metal compounds. *J. Chem. Phys.* **2009**, *130*, 234114.

(9) Guo, S.; Li, Z.; Chan, G. K.-L. A Perturbative Density Matrix Renormalization Group Algorithm for Large Active Spaces. *Journal of Chemical Theory and Computation* **2018**, *14*, 4063–4071, PMID: 29927592.

(10) Eriksen, J. J.; Anderson, T. A.; Deustua, J. E.; Ghanem, K.; Hait, D.; Hoffmann, M. R.; Lee, S.; Levine, D. S.; Magoulas, I.; Shen, J.; Tubman, N. M.; Whaley, K. B.; Xu, E.; Yao, Y.; Zhang, N.; Alavi, A.; Chan, G. K.-L.; Head-Gordon, M.; Liu, W.; Piecuch, P.; Sharma, S.; Ten-no, S. L.; Umrigar, C. J.; Gauss, J. The Ground State Electronic Energy of Benzene. *J. Phys. Chem. Lett.* **2020**, *11*, 8922–8929.

(11) IEEE Standard for Floating-Point Arithmetic. *IEEE Std 754-2019 (Revision of IEEE 754-2008)* **2019**, 1–84.

(12) Ufimtsev, I. S.; Martinez, T. J. Quantum chemistry on graphical processing units. 1. Strategies for two-electron integral evaluation. *J. Chem. Theory Comput.* **2008**, *4*, 222–231.

(13) Ufimtsev, I. S.; Martinez, T. J. Quantum Chemistry on Graphical Processing Units. 2. Direct Self-Consistent-Field Implementation. *J. Chem. Theory Comput.* **2009**, *5*, 1004–1015.

(14) Asadchev, A.; Gordon, M. S. New multithreaded hybrid CPU/GPU approach to Hartree–Fock. *J. Chem. Theory Comput.* **2012**, *8*, 4166–4176.

(15) Miao, Y.; Merz Jr, K. M. Acceleration of electron repulsion integral evaluation on graphics processing units via use of recurrence relations. *J. Chem. Theory Comput.* **2013**, *9*, 965–976.

(16) Barca, G. M.; Galvez-Vallejo, J. L.; Poole, D. L.; Rendell, A. P.; Gordon, M. S. High-Performance, Graphics Processing Unit-Accelerated Fock Build Algorithm. *J. Chem. Theory Comput.* **2020**, *16*, 7232–7238.

(17) Yasuda, K. Two-electron integral evaluation on the graphics processor unit. *J. Comput. Chem.* **2008**, *29*, 334–342.

(18) Asadchev, A.; Allada, V.; Felder, J.; Bode, B. M.; Gordon, M. S.; Windus, T. L. Uncontracted Rys quadrature implementation of up to g functions on graphical processing units. *J. Chem. Theory Comput.* **2010**, *6*, 696–704.

(19) Tornai, G. J.; Ladjánszki, I.; Rák, Á.; Kis, G.; Cserey, G. Calculation of quantum chemical two-electron integrals by applying compiler technology on GPU. *J. Chem. Theory Comput.* **2019**, *15*, 5319–5331.

(20) Tian, Y.; Suo, B.; Ma, Y.; Jin, Z. Optimizing two-electron repulsion integral calculations with McMurchie–Davidson method on graphic processing unit. *J. Chem. Phys.* **2021**, *155*, 034112.

(21) Song, C.; Martínez, T. J. Atomic orbital-based SOS-MP2 with tensor hypercontraction. I. GPU-based tensor construction and exploiting sparsity. *J. Chem. Phys.* **2016**, *144*, 174111.

(22) Isborn, C. M.; Luehr, N.; Ufimtsev, I. S.; Martínez, T. J. Excited-State Electronic Structure with Configuration Interaction Singles and Tamm–Dancoff Time-Dependent

Density Functional Theory on Graphical Processing Units. *J. Chem. Theory Comput.* **2011**, *7*, 1814–1823, PMID: 21687784.

(23) Ma, W.; Krishnamoorthy, S.; Villa, O.; Kowalski, K. GPU-Based Implementations of the Noniterative Regularized-CCSD(T) Corrections: Applications to Strongly Correlated Systems. *J. Chem. Theory Comput.* **2011**, *7*, 1316–1327, PMID: 26610126.

(24) Kim, J.; Panyala, A.; Peng, B.; Kowalski, K.; Sadayappan, P.; Krishnamoorthy, S. Scalable Heterogeneous Execution of a Coupled-Cluster Model with Perturbative Triples. Proceedings of the International Conference for High Performance Computing, Networking, Storage and Analysis. 2020.

(25) Jia, W.; Wang, H.; Chen, M.; Lu, D.; Lin, L.; Car, R.; Weinan, E.; Zhang, L. Pushing the Limit of Molecular Dynamics with Ab Initio Accuracy to 100 Million Atoms with Machine Learning. Proceedings of the International Conference for High Performance Computing, Networking, Storage and Analysis. 2020; pp 1–14.

(26) Shang, H.; Li, F.; Zhang, Y.; Zhang, L.; Fu, Y.; Gao, Y.; Wu, Y.; Duan, X.; Lin, R.; Liu, X.; Liu, Y.; Chen, D. Extreme-Scale *Ab Initio* Quantum Raman Spectra Simulations on the Leadership HPC System in China. Proceedings of the International Conference for High Performance Computing, Networking, Storage and Analysis. New York, NY, USA, 2021.

(27) Abdelfattah, A.; Anzt, H.; Boman, E. G.; Carson, E.; Cojean, T.; Dongarra, J.; Gates, M.; Grützmacher, T.; Higham, N. J.; Li, S., et al. A survey of numerical methods utilizing mixed precision arithmetic. *arXiv preprint arXiv:2007.06674* **2020**,

(28) Higham, N. J.; Mary, T. Mixed precision algorithms in numerical linear algebra. *preprint from http://eprints.maths.manchester.ac.uk/* **2021**,

(29) Alvermann, A.; Basermann, A.; Bungartz, H.-J.; Carbogno, C.; Ernst, D.; Fehske, H.; Futamura, Y.; Galgon, M.; Hager, G.; Huber, S., et al. Benefits from using mixed

precision computations in the ELPA-AEO and ESSEX-II eigensolver projects. *Japan J. Indust. Appl. Math.* **2019**, *36*, 699–717.

(30) Tsuchida, E.; Choe, Y.-K. Iterative diagonalization of symmetric matrices in mixed precision and its application to electronic structure calculations. *Comput. Phys. Commun.* **2012**, *183*, 980–985.

(31) Carson, E.; Gergelits, T.; Yamazaki, I. Mixed precision s-step Lanczos and conjugate gradient algorithms. *Numer. Linear Algebra Appl.* **2022**, *29*, e2425.

(32) Le Grand, S.; Götz, A. W.; Walker, R. C. SPFP: Speed without compromise—A mixed precision model for GPU accelerated molecular dynamics simulations. *Comput. Phys. Commun.* **2013**, *184*, 374–380.

(33) Clark, M.; Babich, R.; Barros, K.; Brower, R.; Rebbi, C. Solving lattice QCD systems of equations using mixed precision solvers on GPUs. *Computer Physics Communications* **2010**, *181*, 1517–1528.

(34) Luehr, N.; Ufimtsev, I. S.; Martínez, T. J. Dynamic Precision for Electron Repulsion Integral Evaluation on Graphical Processing Units (GPUs). *J. Chem. Theory Comput.* **2011**, *7*, 949–954, PMID: 26606344.

(35) Asadchev, A.; Gordon, M. S. Mixed-precision evaluation of two-electron integrals by Rys quadrature. *Computer Physics Communications* **2012**, *183*, 1563–1567.

(36) Parrish, R. M.; Liu, F.; Martínez, T. J. Communication: A difference density picture for the self-consistent field ansatz. *J. Chem. Phys.* **2016**, *144*, 131101.

(37) Das, S.; Motamarri, P.; Gavini, V.; Turcksin, B.; Li, Y. W.; Leback, B. Fast, Scalable and Accurate Finite-Element Based *Ab Initio* Calculations Using Mixed Precision Computing: 46 PFLOPS Simulation of a Metallic Dislocation System. Proceedings of

the International Conference for High Performance Computing, Networking, Storage and Analysis. New York, NY, USA, 2019.

(38) Vogt, L.; Olivares-Amaya, R.; Kermes, S.; Shao, Y.; Amador-Bedolla, C.; Aspuru-Guzik, A. Accelerating Resolution-of-the-Identity Second-Order Møller-Plesset Quantum Chemistry Calculations with Graphical Processing Units. *J. Phys. Chem. A* **2008**, *112*, 2049–2057, PMID: 18229900.

(39) Vysotskiy, V. P.; Cederbaum, L. S. Accurate Quantum Chemistry in Single Precision Arithmetic: Correlation Energy. *J. Chem. Theory Comput.* **2011**, *7*, 320–326, PMID: 26596154.

(40) DePrince, A. E.; Hammond, J. R. Coupled Cluster Theory on Graphics Processing Units I. The Coupled Cluster Doubles Method. *J. Chem. Theory Comput.* **2011**, *7*, 1287–1295, PMID: 26610123.

(41) Knizia, G.; Li, W.; Simon, S.; Werner, H.-J. Determining the Numerical Stability of Quantum Chemistry Algorithms. *J. Chem. Theory Comput.* **2011**, *7*, 2387–2398, PMID: 26606614.

(42) Pokhilko, P.; Epifanovsky, E.; Krylov, A. I. Double Precision Is Not Needed for Many-Body Calculations: Emergent Conventional Wisdom. *J. Chem. Theory Comput.* **2018**, *14*, 4088–4096, PMID: 29969560.

(43) Wang, Z.; Guo, M.; Wang, F. Single-precision open-shell CCSD and CCSD(T) calculations on graphics processing units. *Phys. Chem. Chem. Phys.* **2020**, *22*, 25103–25111.

(44) Chan, G. K.-L.; Sharma, S. The Density Matrix Renormalization Group in Quantum Chemistry. *Annu. Rev. Phys. Chem.* **2011**, *62*, 465–481, PMID: 21219144.

(45) Kurashige, Y. Multireference electron correlation methods with density matrix renormalisation group reference functions. *Mol. Phys.* **2014**, *112*, 1485–1494.

(46) Wouters, S.; Van Neck, D. The density matrix renormalization group for ab initio quantum chemistry. *Eur. Phys. J. D* **2014**, *68*, 1–20.

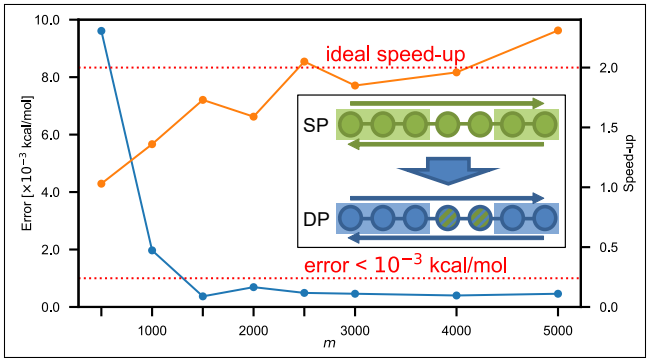
(47) Baiardi, A.; Reiher, M. The density matrix renormalization group in chemistry and molecular physics: Recent developments and new challenges. *J. Chem. Phys.* **2020**, *152*, 040903.

(48) Freitag, L.; Reiher, M. *Quantum Chemistry and Dynamics of Excited States*; John Wiley & Sons, Ltd, 2020; Chapter 7, pp 205–245.

(49) Cheng, Y.; Xie, Z.; Ma, H. Post-Density Matrix Renormalization Group Methods for Describing Dynamic Electron Correlation with Large Active Spaces. *J. Phys. Chem. Lett.* **2022**, *13*, 904–915, PMID: 35049302.

(50) Xie, Z.; Song, Y.; Li, J.; Chen, Y.; Tian, Y.; Luo, Z.; Ma, H. Kylin 1.0: An Ab-Initio Density Matrix Renormalization Group Quantum Chemistry Program. *in preparation*

TOC Graphic


ideal speed-up
SP
DP
error < 10^{-3} kcal/mol
Error [$\times 10^{-3}$ kcal/mol]
Speed-up
m
10.0
8.0
6.0
4.0
2.0
0.0
2.0
1.5
1.0
0.5
0.0
1000
2000
3000
4000
5000