



Resolution of the Identity and Cholesky Representation of EOM-MP2 Approximation: Implementation, Accuracy and Efficiency

DINESH KUMAR^a, ACHINTYA KUMAR DUTTA^{b,*} and PRASHANT UDAY MANOHAR^{a,*} 

^aDepartment of Chemistry, Birla Institute of Technology and Science, Pilani, Rajasthan 333 031, India

^bPhysical Chemistry Division, CSIR-National Chemical Laboratory, Pune, Maharashtra 411 008, India

E-mail: achintyachemist@gmail.com; pumanohar@pilani.bits-pilani.ac.in

MS received 22 June 2017; revised 29 August 2017; accepted 7 September 2017; published online 23 October 2017

Abstract. We present a Resolution of Identity and Cholesky Decomposition Based Implementation of EOM-MP2 approximation. The RI and CD based EOM-MP2 shows significant speed-up and less storage requirement than the conventional canonical version and can be applied to very large systems. The new algorithm used for this implementation eliminates the most storage requiring four-index quantities resulting in the decrease of storage requirement, reduction in I/O penalties and improved parallel performance, at the expense of more floating point operations. Therefore, the speed-up compared to conventional EOM-MP2 method is more prominent in case of EA, EE and SF case where the storage bottleneck is significant than the EOM-IP-MP2 method, where the storage requirement is significantly less. However, the RI/CD based EOM-IP-MP2 can be coupled with frozen natural orbitals to gain further speed-up.

Keywords. Resolution-of-identity; RI; Cholesky decomposition; CD; equation-of-motion; EOM, CC; MP2; electron correlation; excited state; ionization; electron attachment; spinflip; IP; EA; EE; SF.

1. Introduction

Energy differences dictate chemistry. Be they thermodynamic properties like activation energy, free energy change, enthalpy change or spectroscopic properties like excitation energy (EE), ionization energy (IP) and electron affinity (EA). In recent times *ab-initio* computation has evolved as an efficient and robust tool to rationalize and interpret experimentally observed energy differences^{1,2} of both kinds. The spectroscopic properties are mainly computed using equation-of-motion (EOM)³ based approaches, propagator based approaches⁴ or extended Koopmans theorem (EKT)^{5,6} based approaches.

Among the various available theoretical methods, the EOM coupled cluster (EOM-CC)^{3,7–12} method has been one of the most attractive options for direct computation of EE, IP and EA, due to accuracy and scope for systematic improvements in the description of total and difference energies. The EOM-CC methods provide a

black-box way to various sectors of Fock-space (FS) multi-reference (MR) CC methods^{13–17} and for principal peaks in IP and EA problems are identical to the (0,1) and (1,0) sectors of (FS)MRCC.¹⁸ The EOM-CC is generally employed with single and double substitutions (EOM-CCSD) which provides an accuracy of 0.2 eV for low-lying excited states⁴⁰ and about 0.1–0.4 eV for principle ionized/electron-attached states. Inclusion of triples enables one to go for enhanced accuracy, particularly for studying strong near-degeneracies and high-lying excitations, obviously at the expense of added computational cost. The inclusion of partial, as well as, full triples has been tried within the framework of EOM-CC method.^{11,19–27} The linear response (LR) CC method arrives at the same expressions for the excited states as that of the EOM-CC method, although both the approaches start from very different theoretical viewpoints. In spite of having all other favourable features, EOM-CC method, like other coupled cluster methods, has the problem of steep computational scaling. For example, EOM-CCSD has prohibitively high scaling of N^6 and its variants such as EA, EE and SF

*For correspondence

have high storage requirement, which often becomes a limiting factor, for the available RAM and hard disks and slows down the computations by creating I/O bottlenecks. Due to this, EOM-CCSD computations with a reasonable basis beyond molecules containing more than ten second row atoms become computationally challenging. Significant efforts have been devoted to the development of lower scaling approximations to different flavors of EOM-CCSD method.^{28–31} All of them are based on the principle of approximating the CC and EOM equations by one way or the other.

The most obvious way of introducing an approximation is through perturbation order analysis.³² The CC and many body perturbation theory (MBPT) share an intriguing relationship³³ and suitable lower order iteration of coupled cluster amplitude equations recovers various orders of MBPT theory.³⁴ For example, the lowest order approximation to CCSD two-body amplitudes results in MBPT(2) method, more commonly referred to as MP2 method. Therefore, most of the strategies for generation of lower scaling approximation to EOM-CCSD has used perturbational truncation of the coupled cluster similarity transformed Hamiltonian for reduction of computational cost. Nooijen and Sniders³⁵ have first suggested the idea of using MBPT(2) amplitudes as an approximation to the CCSD amplitudes. However, the method would lead to loss of size-extensivity unless an appropriate modification of the Hamiltonian matrix elements would be performed. Stanton and Gauss,³⁶ later came up with a more generalized scheme and proposed a hierarchical approximation to full EOM-CCSD method, which guarantees size-extensivity of the energy for truncation of the similarity transformed Hamiltonian at each order of perturbation and can be applied to both energies and properties. They have denoted their approximation as EOM-CCSD(n), where n denotes that the effective Hamiltonian contains terms up to n^{th} order in perturbation. At large value of n, the truncated similarity transformed Hamiltonian converge to full CCSD similarity transformed Hamiltonian and the corresponding EOM-CCSD(n) energies and properties converge to their standard EOM-CCSD values. A second order approximation ($n = 2$) leads to EOM-MP2 method (or EOM-CCSD(2) method) and it gives significant savings both in terms of computational timing and storage requirements. Therefore, the EOM-MP2 approximation, in its original proposed form^{28,29,37} and with further modifications,^{30,31,38} has been extensively used to define lower scaling strategies for calculation of ionization potential, electron affinity and excitation energy. The excellent performance of the method for complex multi-reference situations,²⁹ potential energy surfaces,²⁹ geometry and IR frequency of large doublet

radicals²⁸ has also been reported. The EOM-MP2 relies on the perturbative truncation of the similarity transformed Hamiltonian with the framework of EOM-CC method. The CC2 method is, on the other hand, derived within the linear response framework, where the singles amplitudes in the CCSD Jacobian matrix are kept untruncated and the doubles amplitudes are truncated at the second order.³⁹ The method has an $O(N^5)$ scaling and gives an accuracy of comparable to the canonical EOM-CCSD method for valence excited state dominated by single excitations.⁴⁰ However, the method turns out to be less accurate for Rydberg excited states⁴¹ and ionization potentials.⁴² For Rydberg states⁴¹ and charge-transfer states,⁴² the accuracy of EOM-MP2 is comparable with that of EOM-CCSD. However, the results are much less accurate for valence excited states,⁴³ ionization potentials⁴² and electron affinities,³⁰ unless one makes special modifications.^{30,31,38} However, the real benefits of the EOM-MP2 method show up in the case of geometry and properties, where it gives accuracy comparable to that of the canonical EOM-CCSD method at a much lower computational cost.

Another alternative approach to reduce the scaling and storage requirements is by approximating the four-index electron repulsion integrals (ERI) by use of resolution-of-identity^{44–50} or by Cholesky decomposition techniques.^{51–58} Both the approaches have been widely used to reduce the cost of single-reference coupled cluster methods.^{59–69} Recently the RI/CD to ERI has been extended to EOM-CC methods by Krylov and co-workers.⁷⁰ They have shown that the use of RI/CD can give significant reduction of I/O requirements and at the same time, the speed-up in the EOM calculations. Both RI and CD result in lowering of the scaling of Coulomb-type integrals. However, the scaling of the exchange-type integrals is not improved by these approximations as a result of which the overall scaling is unaltered, though a significant speed-up is observed due to reduction in prefactor. These effects of RI and CD become more prominent for molecules of large computational size. Recently, Izs ac and co-workers have used pseudo-spectral chain of sphere approximation to accelerate the so called ‘exchange’ kind of terms in EOM-CC.⁷¹ Hattig and co-workers⁷² presented RI implementation on CC2, which is considered as the benchmark standard for doing large scale CC calculations on excited states.

The aim of this present study is to combine the perturbative truncation and RI/CD approach to further lower down the scaling and storage requirements of EOM-CC computations. The paper is organized as follows: the second section contains the theory and implementation

details of RI/CD approximation to EOM-MP2 methods. Numerical results and discussion on them are followed in section 3. Section 4 contains the concluding remarks.

2. Theory

The non-variational coupled cluster theory^{1,2} assumes that the exact wave-function can be generated by the action of an exponential operator on a single Slater determinant.

$$|\Psi\rangle = e^{\hat{T}}|\phi_0\rangle \quad (1)$$

The $|\phi_0\rangle$ is generally, but not restricted to, a Hartree-Fock determinant and $\hat{T} = \hat{T}_1 + \hat{T}_2 + \hat{T}_3 + \dots + \hat{T}_n$; where,

$$\begin{aligned} \hat{T}_1 &= \sum_{ia} t_i^a \{a_a^\dagger a_i\} \\ \hat{T}_2 &= \frac{1}{4} \sum_{ijab} t_{ij}^{ab} \{a_a^\dagger a_b^\dagger a_j a_i\} \\ \hat{T}_3 &= \frac{1}{6} \sum_{ijkabc} t_{ijk}^{abc} \{a_a^\dagger a_b^\dagger a_c^\dagger a_k a_j a_i\} \\ &\dots\dots\dots \end{aligned} \quad (2)$$

One can obtain these \hat{T} amplitudes by solving a system of coupled non-linear equations in an iterative manner. Inclusion of \hat{T} amplitudes up to n^{th} orders, where n equals to the number of electrons in the system, leads to full CC (FCC), which is equivalent to full configuration interaction (FCI). However, for practical applications, one need to truncate CC at finite order. The exponential nature of the correlation operator guarantees the size-extensivity of energy even in its truncated form. The truncation at singles and doubles level leads to the famous CCSD approximation. The inclusion of triples and more higher orders of excitation results in a systematic increase in the accuracy of calculations, accompanied by the consequent increase in the computational cost. For example, the CCSD scales as N^6 . The inclusion of full triples(CCSDT) elevates the scaling to N^8 , whereas, the inclusion of quadruples shoots up the scaling to N^{10} , which make them unsuitable for any practical application, beyond the purpose of benchmarking for very small molecules.

In EOM-CC method, the desired target electronic states are obtained by diagonalizing the similarity transformed Hamiltonian matrix formed over the suitable configurations.

$$\bar{H} = e^{-\hat{T}} \hat{H} e^{\hat{T}} = (\hat{H} e^{\hat{T}})_c \quad (3)$$

The subscript c in above equation denotes the connectedness of \hat{T} and \hat{H} . Since, T is not unitary, \hat{H} is not Hermitian. Consequently, there exists different right and left eigenvectors.

$$\langle \tilde{\Psi} | = \langle \phi_0 | \hat{L} e^{-\hat{T}} \quad (4)$$

$$|\Psi\rangle = e^{\hat{T}} \hat{R} |\phi_0\rangle \quad (5)$$

L and R are biorthogonal to each other and can be normalized to satisfy

$$\langle \phi_0 | \hat{L}_K \hat{R}_L | \phi_0 \rangle = \delta_{kl} \quad (6)$$

The form of the excitation operator R_K for IP, EA, EE and spin flip (SF) variants of EOM-CC is as follows

$$\begin{aligned} R_K^{\hat{E}E} &= R_0(K) + \sum_{i,a} R_i^a(K) a^\dagger i + \sum_{i,j,a,b} R_{ij}^{ab}(K) a^\dagger b^\dagger ij + \dots \\ R_K^{\hat{I}E} &= \sum_i R_i(K) i + \sum_{i,j,a} R_{ij}^a(K) a^\dagger ij + \dots \\ R_K^{\hat{E}A} &= \sum_a R^a(K) a^\dagger + \sum_{i,a,b} R_i^{ab}(K) a^\dagger b^\dagger i + \dots \\ R_K^{\hat{S}F} &= \sum_{i\uparrow,a\downarrow} R_{i\uparrow}^{a\downarrow}(K) a_{\downarrow}^\dagger i_{\uparrow} + \dots \end{aligned} \quad (7)$$

In the above equation, i, j and a, b denotes occupied and virtual orbitals, respectively. Following Gauss and Stanton's³⁶ suggestion equation (3) can be expanded in perturbation orders

$$\bar{H} = (H e^T)_c = \bar{H}^{[1]} + \bar{H}^{[2]} + \bar{H}^{[3]} + \dots + \bar{H}^{[n]} \quad (8)$$

The bracketed subscript in the above equation represents orders in perturbation. At large value of n , $\bar{H}^{[n]}$ converges to full \bar{H} and, consequently, the perturbative EOM variant converges to full EOM-CCSD. It can be shown that n^{th} order truncated form of the similarity transformed Hamiltonian contains terms only up to n^{th} order in perturbation, which ensures core-extensivity of the final energy for all values n . Truncation at $n = 2$, leads to EOM-MP2 method (sometimes referred to as EOM-CCSD(2)) and can be expressed as

$$\begin{aligned} \bar{H} &= (H e^T)_c \\ &\approx (H e^{T'})_c \end{aligned} \quad (9)$$

where the T' second order perturbative approximation to coupled cluster T amplitudes and can be expressed as

$$\begin{aligned} T_i^a &= \frac{f_{ia}}{\epsilon_i - \epsilon_a} \\ T_{ij}^{ab} &= \frac{\langle ab || ij \rangle}{\epsilon_i + \epsilon_j - \epsilon_a - \epsilon_b} \end{aligned} \quad (10)$$

T_1 is zero for RHF and UHF MP2 reference. Using these T' amplitudes one can define a modified effective \bar{H}' , which can be used as the reference for subsequent EOM calculation.

The diagonalization of \bar{H}' , within the $(N - 1)$ and $(N + 1)$ reference space lead to EOM-IP-MP2 and EOM-EA-MP2 methods, respectively, which offer significant computational simplifications over the standard EOM-IP-CCSD and EOM-EA-CCSD method. The CPU scaling EOM-IP and EOM-EA-CCSD is dominated by reference state CCSD calculation.

In EOM-IP/EA-CCSD, the diagonalization of the \bar{H} (using modified Davidsons iterative method⁷³) scales as N^5 . Therefore, truncating the reference state at MP2, would eliminate the N^6 -scaled CCSD computation and thereby reduces the overall EOM-MP2 computational scaling to N^5 . EOM-IP-MP2 has an additional advantage that the computation of costly $\langle AB || CD \rangle$ integrals is eliminated which results in

significant saving in terms of storage. Unfortunately, EOM-EA-MP2 cannot have this advantage in storage requirements. Recently, Pal and co-workers³⁰ have shown that a further truncation of the EOM matrix based on Lowdin's partitioning technique⁷⁴ can avoid the four-particle integrals in EOM-EA-MP2 method. The matrix diagonalization part in both EOM-EE-CCSD and EOM-SF-CCSD scales as iterative N^6 and both require the computation of the costly four-particle integrals. Therefore, the cost of EOM-EE/SF-MP2 is reduced only in the integral computation (resulting in a lower prefactor) but not in formal scaling and storage requirements compared to its EOM-CCSD analogue. However, for smaller number of roots (limited number of desired target states), the cost of EOM-CCSD is dominated by the CCSD (reference state calculations). Thus, speed-ups due to the lower prefactors in EOM-EE-MP2 and EOM-SF-MP2 are significant in comparison to their EOM-CCSD analogues.²⁹ It is also possible to use truncation of the EOM matrix based on Lowdin's partitioning technique⁷⁴ to reduce the formal scaling and storage requirements in both EOM-EE-MP2 and EOM-SF-MP2. However, we have not considered that in the present study.

2.1 RI and CD schemes

Recognizing that computation of the electron-repulsion integrals (ERI's) $(\mu\nu|\lambda\sigma)$, significantly affects the computational cost of an electronic structure method (and consequently, its applicability to large molecules), two cost-effective approximations to the ERI computation have now been commonly used, namely, the CD and RI approaches, respectively. Both the approaches⁷⁰ express the ERI's as

$$(\mu\nu|\lambda\sigma) = \sum_{P=1}^M B_{\mu\nu}^P B_{\lambda\sigma}^P \quad (11)$$

where, $B_{\mu\nu}^P$ is obtained as linear expansion of products of particular one-electron functions and M is the rank of decomposition and depends on the accuracy desired. The RI and CD approaches differ in the way, these densities are computed. In RI, an auxiliary basis is used to obtain the densities:

$$\begin{aligned} (\mu\nu|\lambda\sigma) &= \sum_{PQ} C_{\mu\nu}^P(P|Q) C_{\lambda\sigma}^Q \\ &= \sum_{PQ} (\mu\nu|P) (P|Q)^{-1} (Q|\lambda\sigma) \end{aligned} \quad (12)$$

$$(P|Q) = \int \frac{P(r_1)Q(r_2)}{|r_1 - r_2|} \quad (13)$$

The coefficients, $C_{\mu\nu}^P$ are obtained so as to minimize the difference between the actual and the fitted product densities:

$$\sum_N (M|N) C_{\mu\nu}^N = (M|\mu\nu) \quad (14)$$

$$\begin{aligned} B_{\mu\nu}^M &= \sum_N C_{\mu\nu}^N (N|M)^{1/2} \\ &= \sum_N (M|\mu\nu) (M|N)^{-1/2} (N|M)^{1/2} \end{aligned} \quad (15)$$

The accuracy of RI is strongly dependent on the quality of the auxiliary basis used. Preoptimized basis sets⁴⁴ corresponding to primary atomic basis and method (HF, MP2, etc) are usually available with many of the commonly used quantum chemistry softwares.⁷⁵⁻⁷⁸

CD approach⁵¹⁻⁵⁷ provides a more robust tool as it eliminates the requirement of auxiliary basis-functions. In this approach, the Cholesky decomposition of actual ERI matrix is iteratively performed to obtain and improve the Cholesky vectors. In this sense, CD is a more general approach and allows one to approximate original ERI matrix to an arbitrary accuracy. In this method, one computes all the diagonal elements $D_{\lambda\sigma}^0 = (\lambda\sigma|\lambda\sigma)$ of ERI matrix. The largest diagonal element, $(\lambda\sigma|\lambda\sigma)_{max}$ is identified and the first Cholesky vectors are obtained as follows:

$$B_{\mu\nu}^1 = (\lambda\sigma|\lambda\sigma)_{max}^{-1/2} (\mu\nu|\lambda\sigma)_{max} \quad (16)$$

The Cholesky vectors are optimized iteratively. For example, the Diagonal elements in j -th iteration are obtained from the $(j-1)$ -th iteration as

$$D_{\lambda\sigma}^j = D_{\lambda\sigma}^{j-1} - B_{\lambda\sigma}^j B_{\lambda\sigma}^j \quad (17)$$

After updating the diagonal elements, the largest element from the new set is identified and the Cholesky vectors for the next iterations are obtained using similar way as described in Eq (16). The procedure is repeated till the largest diagonal element of given iteration becomes smaller than the cut-off, which can be arbitrarily chosen by user depending on the desired accuracy. This is referred to as decomposition threshold, δ . In Q-CHEM, the threshold can be set to $\delta = 10^{-tol}$, where tol is a user-defined positive integer. The $tol = 3$ (that is, $\delta = 10^{-3}$) provides good balance between the accuracy and compactness of rank. The $tol = 2$ may provide only semi-qualitative results, whereas $tol = 4$ is computationally involved and may be used if very high accuracy is desired. In this article, we will present the CD results with $tol = 2, 3$ and 4.

2.2 RI and CD EOM-MP2

The equations for reference state: The reference state energy in RI and CD EOM-MP2 method is represented in the spin orbital notation as

$$E_{MP2} = \langle \Phi_0 | H | \Phi_0 \rangle + \frac{1}{2} \sum_{ijab} \left(\sum_P B_{ia}^P B_{jb}^P \right) t_{ij}^{ab} \quad (18)$$

where

$$t_{ij}^{ab} = \frac{\wp_{ab}^- \sum_P B_{ia}^P B_{jb}^P}{\Delta_{ij}^{ab}} \quad (19)$$

\wp_{ij}^- and \wp_{ab}^- implies the sum of two components differing by permutation of i,j and a,b indices, respectively.

The equations for EOM-EE-MP2 and EOM-SF-MP2: The sigma equations for k^{th} excited state EOM-EE-MP2 and

EOM-SF-MP2 using spin orbitals are as follows. EE and SF differ only in the spin-multiplicity of the EOM vectors. EOM-EE vector retain the electron spins, whereas in case of EOM-SF vectors, the spin of one electron is flipped.

$$r_0 = \frac{1}{2\omega_k} \sum_{ijab} \left(\sum_P B_{ia}^P B_{jb}^P \right) r_{ij}^{ab} \quad (20)$$

$$\begin{aligned} \sigma_i^a &= \sum_b r_i^b F_{ab} - \sum_a r_j^b F_{ij} \\ &\quad - \sum_{jP} B_{ji}^P (M_{jaP}^{2R} + M_{jaP}^{3R}) + \sum_P M_{iP}^{2R} M_{ia}^P \\ &\quad - \sum_{kc} \left(\sum_{jP} B_{jc}^P M_{kjP}^{2R} \right) t_{ik}^{ac} + \sum_{cP} M_{icP}^{2R} B_{ca}^P \end{aligned} \quad (21)$$

$$\begin{aligned} \sigma_{ij}^{ab} &= I_{ijab}^{(1i)} + \wp_{ab}^- \left\{ \sum_c (T_{ac}^{(2)} + T_{ac}^{(4)}) t_{ij}^{bc} + \sum_c r_{ij}^{ac} F_{bc} \right. \\ &\quad \left. - \sum_k I_{ijkb}^{(2b)} r_k^{ab} \right\} + \frac{1}{2} \sum_{kl} I_{ijkl}^{(4b)} r_{kl}^{ab} \\ &\quad + \wp_{ij}^- \wp_{ab}^- \left\{ \sum_{kd} t_{jk}^{ad} \left(\sum_P B_{db}^P M_{kiP}^{2R} + I_{kidx}^{(2i)} \right) \right. \\ &\quad \left. - \sum_P M_{ja}^P \left(\sum_c B_{db}^P r_i^c + M_{ibP}^{2R} \right) + I_{ijab}^{(3i)} \right\} \\ &\quad + \wp_{ij}^- \left\{ \sum_k r_{jk}^{ab} F_{ik} + \sum_k t_{jk}^{ab} \left(\sum_c T_{ik}^{(1)} + T_{ik}^{(3)} \right) \right\} \\ &\quad - \frac{1}{2} \sum_{kl} t_{kl}^{ab} \left(\wp_{ij}^- \wp_{kl}^- \sum_P M_{liP}^{2R} B_{kj}^P - \frac{1}{2} I_{ijkl}^{(4i)} \right) \end{aligned} \quad (22)$$

where ω_k is the corresponding excitation energy.

$$I_{ijab}^{(1i)} = \frac{1}{2} \sum_{cd} (\wp_{ab}^- \sum_P B_{ca}^P B_{db}^P) r_{ij}^{cd} \quad (23)$$

$$I_{kidx}^{(2i)} = \sum_{lc} \left(\sum_P B_{ld}^P B_{kc}^P \right) r_{il}^{bc} \quad (24)$$

$$I_{ijab}^{(3i)} = \sum_{kc} \left(\sum_P B_{ki}^P B_{cb}^P \right) r_{jk}^{ac} \quad (25)$$

$$I_{ijkl}^{(4i)} = \sum_{cd} (\wp_{cd}^- \sum_P B_{kc}^P B_{ld}^P) r_{ij}^{cd} \quad (26)$$

The intermediates described in Eq (22–25) are calculated using the three-tensor contraction procedure i.e. sequential multiplication of three tensor-like quantities and the intermediates are immediately added to the results.

M , F , T , $I^{(2b)}$, $I^{(4b)}$ are defined in Tables 1 and 2. The R -independent intermediates in Table 1 are computed only once after the computation of the MP2 approximated CC vectors while the R -dependent intermediates defined in Table 2 are recomputed in each Davidson iteration. As a result, there is a significant reduction in the computational cost in terms of storage requirements, though the larger no of O^3V^3 contractions (than in the canonical version) slightly increase the

Table 1. R -independent intermediates used in RI/CD EOM-MP2 energy expressions.

$$\begin{aligned} M_{iaP}^{2T} &= \sum_{jb} B_{jb}^P t_{ij}^{ab} \\ M_{ia}^P &= B_{ia}^P + M_{iaP}^{2T} \\ F_{ij} &= f_{ij} + \sum_{cP} B_{kc}^P M_{jcP}^{2T} F_{ab} = f_{ab} - \sum_{kP} B_{kc}^P M_{kbP}^{2T} \\ I_{ijkb}^{(2b)} &= \wp_{ij}^- \left\{ \sum_P B_{kj}^P M_{ia}^P + \right\} \\ I_{ijkl}^{(4b)} &= \wp_{ij}^- \sum_P B_{ki}^P B_{lj}^P + \sum_{ab} \left(\sum_P B_{ka}^P B_{lb}^P \right) t_{ij}^{ab} \end{aligned}$$

Table 2. R -dependent intermediates used in RI/CD EOM-MP2 energy expressions.

$$\begin{aligned} M_{iaP}^{2R} &= \sum_{ld} B_{ld}^P r_{il}^{ad} \\ M_{iaP}^{3R} &= \sum_c B_{ac}^P r_i^c \\ M_{iP}^{2R} &= \sum_{ld} B_{ld}^P r_l^d \\ M_{ijP}^{2R} &= \sum_c B_{ic}^P r_j^c \\ T_{ij}^{(1)} &= \sum_P B_{ji}^P M_{iP}^{2R} - \sum_{kP} M_{jkP}^{2R} B_{ki}^P \\ T_{ab}^{(2)} &= \sum_{kP} B_{kb}^P M_{kaP}^{3R} - \sum_P B_{ba}^P M_P^{2R} \\ T_{ab}^{(3)} &= \sum_{aP} B_{ja}^P M_{iP}^{2R} \\ T_{ab}^{(4)} &= \sum_{iP} B_{ib}^P M_{iP}^{2R} \end{aligned}$$

prefactor. Although it has pushed the domain of applicability to EOM-MP2 a little bit, the method still scales as N^6 , similar to the the original EOM-EE-CCSD method.

The equations for EOM-IP-MP2: The sigma equations for EOM-IP-MP2 using spin orbital notation are as follows

$$\sigma_i = \sum_j F_{ij} r_j - \sum_{jP} B_{ji}^P \left(\sum_{kb} B_{kb}^P r_{jk}^b \right) \quad (27)$$

$$\begin{aligned} \sigma_{ij}^a &= - \sum_k r_k I_{ijk}^{(2b)} + \wp_{ij}^- \left[\sum_k F_{ik} r_{jk}^a - I_{jia}^{(1i)} \right. \\ &\quad \left. + \sum_P M_{ja}^P \left(\sum_{kb} B_{kb}^P r_{ik}^b \right) - \sum_{lc} t_{jl}^{ac} I_{ilc}^{(2i)} \right] \\ &\quad + \sum_b r_{ij}^b F_{ab} + \sum_{kl} I_{ijkl}^{(4b)} r_{kl}^a + \sum_b t_{ij}^{ab} T_b^4 \end{aligned} \quad (28)$$

$$I_{ija}^{(1i)} = \sum_{kb} \left(\sum_P B_{kj}^P B_{ba}^P \right) r_{ik}^b \quad (29)$$

$$I_{ilc}^{(2i)} = \sum_{kb} \left(\sum_P B_{kc}^P B_{lb}^P \right) r_{ik}^b \quad (30)$$

$$T_b^4 = \sum_{kP} B_{kb}^P \left(\sum_{lc} B_{lc}^P r_{kl}^c \right) \quad (31)$$

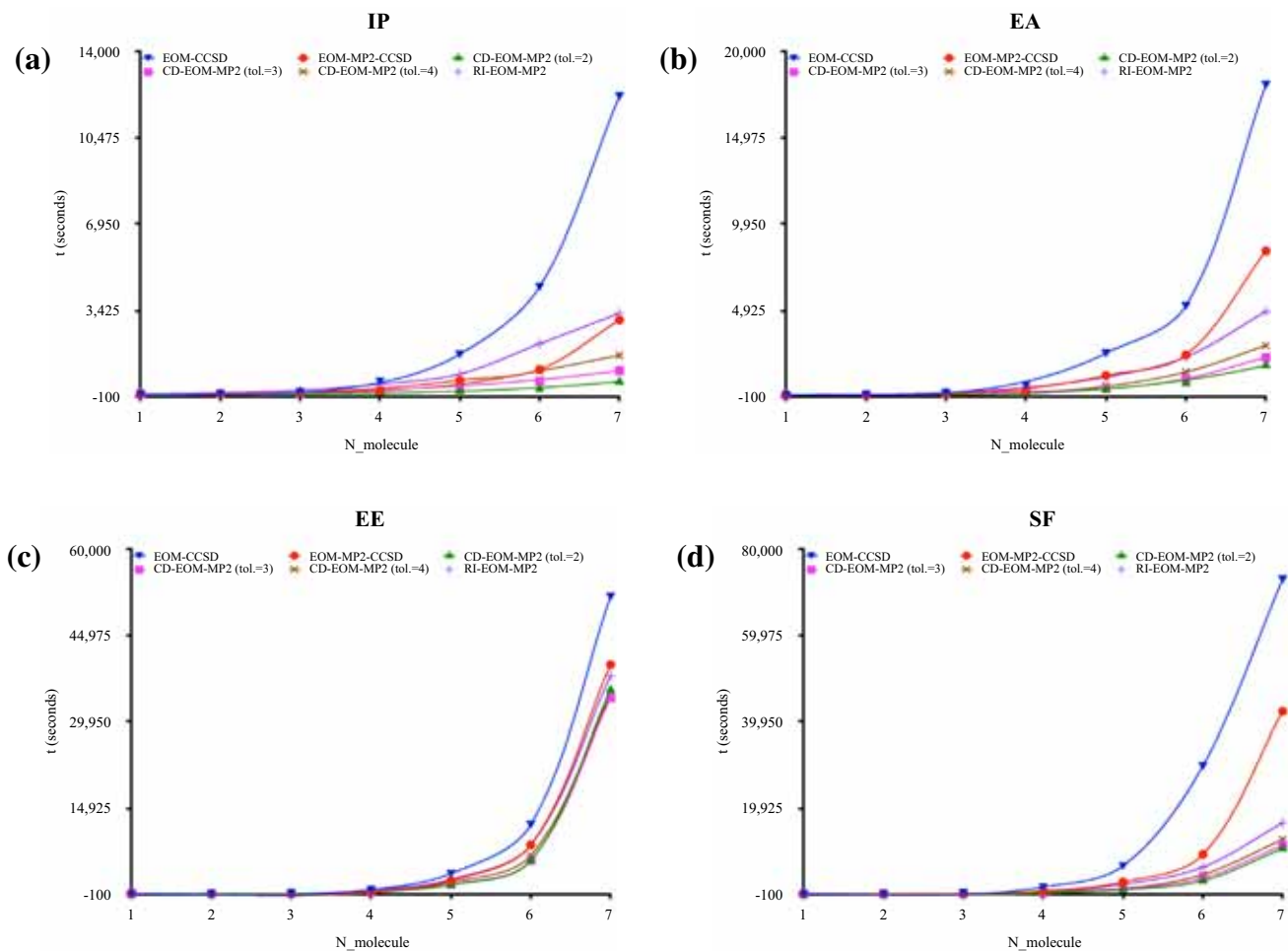


Figure 1. Plot of Total time vs no water clusters for different flavours of EOM-MP2.

The $I^{(1i)}$, $I^{(2i)}$ and $T^{(4)}$ intermediates are calculated using the three-tensor contraction procedure and M intermediates are described in Table 1. The sigma vector update procedure in both canonical and CD/RI-EOM-IP-MP2 method scales as N^5 . The CD/RI scheme have six N^5 , the dominant one being O^2V^2M . The canonical variant, on the other hand, has two N^5 contractions the dominant one being O^3V^2 . Therefore, the sigma vector update procedure should be three times faster in the canonical version than the RI/CD implementation. The storage requirement of both the methods is similar to the sigma update procedure in EOM-IP-MP2 does not involve the four external and three external intermediates. The computational gain achieved in CD/RI implementation of Krylov and co-workers⁷⁰ came from the reference state CCSD calculation, which is not present in MP2 approximation. Therefore, the CD/RI implementation of EOM-IP-MP2 is not expected to give any advantage over the canonical implementation.

The equations for EOM-EA-MP2: The sigma equations for EOM-EA-MP2 using spin orbitals are as follows

$$\sigma_a = \sum_c F_{ac} r^c - \sum_{dP} \left(\sum_{kc} B_{kc}^P r_k^{cd} \right) B_{da}^P \quad (32)$$

$$\begin{aligned} \sigma_i^{ab} = & \wp_{ab}^- \left[\sum_c F_{ac} r_i^{cb} - I_{iab}^{(1i)} \right. \\ & \left. + \sum_P M_{ia}^P \left(\sum_c B_{cb}^P r^c \right) - \sum_{kc} B_{kc}^P r_k^{bc} \right] \\ & \sum_{kl} t_{kl}^{ab} \left[\sum_P \left(\sum_c B_{kc}^P r^c \right) M_{li}^P - I_{kli}^{(2i)} \right] \\ & + \wp_{ab}^- \left\{ \sum_{ld} [I_{idb}^{(3i)} - \sum_P \left(\sum_c B_{lc}^P r^c \right) B_{db}^P] t_{li}^{ad} \right\} \\ & + \sum_l \sum_{dP} B_{ld}^P \left(\sum_{kc} B_{kc}^P r_k^{cd} \right) t_{il}^{ab} + I_{abcd}^{(4i)} \quad (33) \end{aligned}$$

$$I_{iab}^{(1i)} = \sum_{kc} \left(\sum_P B_{ki}^P B_{cb}^P \right) r_k^{ac} \quad (34)$$

$$I_{kli}^{(2i)} = \frac{1}{2} \sum_{cd} \left(\sum_P B_{kc}^P B_{ld}^P \right) r_i^{cd} \quad (35)$$

$$I_{idb}^{(3i)} = \sum_{kc} \left(\sum_P B_{kd}^P B_{lc}^P \right) r_k^{bc} \quad (36)$$

Table 3. Computational timings for EOM-CCSD, EOM-MP2, RI-EOM-MP2 and CD-EOM-MP2 (in seconds) for IP, EA, EE and SF energy calculations of water clusters using cc-pVDZ basis.

Cluster	EOM-CCSD	EOM-MP2	RI-MP2	CD-MP2			
				tol = 2	tol = 3	tol = 4	
IP	Monomer	13.98	11.95	11.84	13.08	14.10	16.94
	Dimer	30.69	17.99	14.39	23.94	35.49	57.79
	Trimer	115.43	47.26	23.80	48.80	78.23	164.65
	Tetramer	533.32	141.15	51.80	119.98	197.40	416.00
	Pentamer	1641.56	580.77	108.59	215.65	415.43	831.27
	Hexamer	4402.71	1016.19	219.60	539.77	956.55	2071.59
	Heptamer	12169.52	3046.57	517.67	967.92	1597.25	3318.21
EA	Monomer	17.35	15.89	12.92	13.51	19.67	17.32
	Dimer	33.75	21.72	17.88	26.70	37.26	59.94
	Trimer	138.22	70.68	33.90	64.59	97.02	178.36
	Tetramer	533.97	141.15	51.80	119.06	266.45	545.87
	Pentamer	2443.44	1141.59	319.34	375.64	544.33	1052.83
	Hexamer	5214.61	2347.02	683.92	949.09	1344.40	2424.51
	Heptamer	18053.36	8404.21	1725.44	2202.48	2887.38	4877.56
EE	Monomer	14.81	13.19	14.05	14.02	17.90	20.99
	Dimer	38.57	27.53	28.35	35.84	47.27	71.99
	Trimer	179.89	108.34	97.82	112.76	147.42	230.91
	Tetramer	732.30	411.22	379.61	311.46	428.53	674.45
	Pentamer	3543.40	2321.80	1575.87	1691.54	2046.66	2522.38
	Hexamer	12081.72	8542.70	5833.33	5867.38	6731.34	8545.06
	Heptamer	51694.74	39814.47	35411.09	34005.05	34536.32	37880.31
SF	Monomer	15.56	13.56	13.06	14.75	15.31	18.24
	Dimer	63.96	41.28	33.82	38.90	51.49	74.60
	Trimer	344.84	115.50	92.75	101.14	138.85	227.28
	Tetramer	1629.97	731.42	464.69	343.62	445.83	732.57
	Pentamer	6550.10	2787.66	1008.10	1203.63	1303.18	2266.16
	Hexamer	29757.69	9208.69	3083.73	3679.30	4517.95	6251.74
	Heptamer	72926.37	42278.97	10588.53	11383.11	12663.76	16541.78

$$I_{abcd}^{(4i)} = \sum_{cd} \left(\sum_P B_{ca}^P B_{db}^P \right) r_i^{cd} \quad (37)$$

The $I^{(1i)}$, $I^{(2i)}$, $I^{(3i)}$ and intermediates are computed using the three-tensor contraction procedure and M intermediates are described in Table 1. The floating point operations for both canonical and CD/RI implementation of EOM-EA-MP2 scales as N^5 , similar to that in EOM-IP-MP2. The canonical version has two N^5 scaling operations, the dominant one scales as OV^4 , whereas the number of N^5 scaling operation in the sigma update of procedure is ten, the dominant one being V^4M scaling. The storage requirement is significantly less for CD/RI-EOM-EA-MP2 due to which these are attractive options over the canonical version. The canonical EOM-EA-MP2 method also scales as N^5 , however, its huge storage restricts its domain of applicability beyond small molecules. The CD/RI-EOM-EA-MP2 method due to its lower scaling and storage requirements can be applied to much larger systems which cannot be achieved by CD/RI-EOM-EA-CCSD or canonical EOM-EA-MP2 alone.

3. Results and Discussion

3.1 Computational details

The RI-EOM-MP2 and CD-EOM-MP2 methods for IP, EA, EE and SF variants are implemented in Q-CHEM⁷⁸ developmental version. The geometries of water clusters was optimized at B3LYP/6-311++G** level of theory and aug-cc-pVDZ basis set were used for the subsequent EOM-CC calculations. The geometric parameters, nuclear repulsion energies, Hartree-Fock energies and MP2 and CCSD energies of the water clusters have been provided in the supplementary material. The optimized geometries of DNA nucleobases were taken from the work of Pal and co-workers.⁷⁹ All the computations were done using Q-CHEM⁷⁸ on an IBM x3530M4 Special Build Server (with 4×6 core processors, 10×16 GB RAM, 4×600 GB 2.4in HDD, etc) and Ubuntu 12.04 OS and the corresponding RI-MP2 basis sets have been used as the auxiliary fitting basis set for all the orbital basis set used.

Table 4. EOM-MP2 total target state energies^a of water clusters and deviations of RI-EOM-MP2 and CD-EOM-MP2 energies relative to EOM – MP2^b.

	Cluster	EOM-MP2	RI-MP2	CD-MP2		
				tol = 2	tol = 3	tol = 4
IP	Monomer	-75.81264	0.000 (14.34)	0.003 (112.05)	0.002 (74.61)	0.000 (7.53)
	Dimer	-152.11678	0.000 (18.20)	0.015 (545.02)	0.005 (188.02)	0.000 (17.27)
	Trimer	-228.38223	0.000 (1.35)	0.008 (307.67)	0.039 (1450.89)	0.001 (21.08)
	Tetramer	-304.66502	0.001 (24.56)	0.022 (795.54)	0.012 (452.21)	0.001 (34.65)
	Pentamer	-380.86701	0.000 (10.42)	0.045 (1641.36)	0.017 (613.74)	0.001 (44.70)
	Hexamer	-457.24425	0.001 (23.23)	0.078 (2860.95)	0.021 (757.46)	0.001 (53.56)
	Heptamer	-533.52662	0.001 (29.41)	0.052 (1918.54)	0.025 (933.63)	0.002 (68.37)
EA	Monomer	-76.23479	0.000 (3.01)	0.004 (160.89)	0.002 (69.76)	0.000 (7.16)
	Dimer	-152.51585	0.000 (5.47)	0.006 (225.48)	0.005 (191.46)	0.000 (17.63)
	Trimer	-228.79288	0.000 (9.87)	0.033 (1224.59)	0.009 (346.57)	0.001 (20.17)
	Tetramer	-305.07643	0.000 (16.51)	0.031 (1130.93)	0.012 (447.11)	0.001 (31.58)
	Pentamer	-381.28175	0.001 (43.88)	0.016 (597.80)	0.038 (1403.07)	0.001 (44.01)
	Hexamer	-457.64114	0.000 (15.15)	0.066 (2436.49)	0.021 (758.60)	0.001 (50.44)
	Heptamer	-533.92405	0.001 (21.52)	0.041 (1505.15)	0.025 (930.76)	0.002 (65.87)
EE	Monomer	-75.99408	0.000 (3.84)	0.009 (317.77)	0.001 (40.02)	0.000 (3.99)
	Dimer	-152.26798	0.000 (4.99)	0.003 (114.45)	0.004 (150.27)	0.000 (12.30)
	Trimer	-228.53729	0.000 (6.99)	0.019 (710.07)	0.007 (252.93)	0.000 (16.11)
	Tetramer	-304.81389	0.000 (12.71)	0.034 (1246.32)	0.010 (364.88)	0.001 (26.71)
	Pentamer	-381.01942	0.000 (1.93)	0.025 (933.14)	0.015 (534.57)	0.001 (39.82)
	Hexamer	-457.37036	0.000 (13.48)	0.058 (2137.74)	0.019 (679.99)	0.001 (44.82)
	Heptamer	-533.66119	0.000 (17.70)	0.031 (1152.41)	0.024 (891.42)	0.002 (63.19)
SF	Monomer	-76.25395	0.001 (30.83)	0.002 (91.24)	0.001 (52.73)	0.000 (4.62)
	Dimer	-152.52544	0.001 (28.88)	0.011 (396.58)	0.005 (182.48)	0.000 (13.95)
	Trimer	-228.80547	0.001 (25.47)	0.029 (1061.96)	0.009 (339.44)	0.000 (17.68)
	Tetramer	-305.08834	0.001 (19.23)	0.030 (1099.38)	0.012 (429.81)	0.001 (30.17)
	Pentamer	-381.28542	0.001 (34.64)	0.042 (1558.59)	0.017 (614.64)	0.001 (41.52)
	Hexamer	-457.64597	0.001 (19.92)	0.076 (2797.63)	0.020 (752.46)	0.001 (49.40)
	Heptamer	-533.93089	0.000 (15.98)	0.043 (1585.46)	0.025 (930.34)	0.002 (65.60)

^a EOM-MP2 energies of target states are expressed in hartrees.

^b The relative deviations of RI-EOM-MP2 and CD-EOM-MP2 are expressed in eV and (microhartrees).

3.2 Computational Timings

The speed-up achieved in RI/CD based methods is due to reduction in the formal scaling of the so-called ‘Coulomb’ type of terms in the EOM-CC expression, as well due to the reduction in I/O. For example, in EOM-CCSD, the $\sigma_{ij}^{ab} = (ia|kc)r_{ab}^{kj}$ scales as O^3V^3 . However, in RI, it can be written as two partial sums $I_{jb}^M = (M|kc)r_{cb}^{kj}$ and $\sigma_{ab}^{ij} = (ia|M)I_{jb}^m$. Both the steps scales as MO^2V^2 , where, M is the dimension of the aux basis and is larger than dimension of orbital basis set by an order of magnitude. This essentially leads to a reduction of scaling by a factor of one. Though the density fitting cannot reduce the formal scaling of the so-called exchange kind of terms, reduction in I/O does lead to some speed-up.

To investigate the performance RI/CD approximation of EOM-MP2 methods, we have computed lowest IP, EA, EE and SF excited states of series of water clusters.

(For convenience, we call them excited states or target states. For IP and EA, these actually correspond to the ionized and electron-attached states, respectively). The geometries of the cluster are provided in the supporting information. Figure 1 provides a plot of time required for the computation of a single root (for IP, EA, EE and SF variants of EOM-MP2 and RI and CD approximations) versus the no of water molecules. The data used for plotting can be found in Table 3.

It can be seen that use of RI/CD approximation leads to significant speed-up for IP and EA. However, the speed-up is much less for EE and SF variant of EOM-CCSD. Total computational timing depends mainly upon two factors. One of them is formal scaling and other is the I/O. In the case of IP and EA the formal scaling is $O(N^5)$ and the I/O is a big contributor to the total timing. Therefore, RI/CD approximation, which reduces the I/O bottleneck, leads to significant speed-up. The reduction in computational timing is much more

Table 5. First ionization energies, electron affinities, excitation energies and spin-flipping excitation energies ($E_{target} - E_{reference}$) (eV) of water clusters computed using EOM-MP2, RI-EOM-MP2 and CD-EOM-MP2 methods and cc-pVDZ basis.

	Cluster	MP2	RI-MP2	CD-MP2		
				tol = 2	tol = 3	tol = 4
IP	Monomer	12.262	12.263	12.264	12.262	12.262
	Dimer	11.387	11.387	11.387	11.386	11.387
	Trimer	11.818	11.818	11.818	11.816	11.818
	Tetramer	11.833	11.833	11.837	11.831	11.833
	Pentamer	11.712	11.713	11.710	11.710	11.712
	Hexamer	11.231	11.232	11.236	11.230	11.231
	Heptamer	11.306	11.306	11.308	11.304	11.306
EA	Monomer	0.775	0.775	0.769	0.774	0.775
	Dimer	0.527	0.527	0.519	0.527	0.527
	Trimer	0.644	0.644	0.638	0.643	0.644
	Tetramer	0.638	0.638	0.633	0.636	0.638
	Pentamer	0.426	0.426	0.418	0.424	0.425
	Hexamer	0.432	0.432	0.424	0.430	0.431
	Heptamer	0.491	0.491	0.482	0.490	0.491
EE	Monomer	7.325	7.325	7.315	7.323	7.325
	Dimer	7.272	7.272	7.261	7.270	7.272
	Trimer	7.599	7.599	7.579	7.595	7.599
	Tetramer	7.776	7.775	7.766	7.771	7.775
	Pentamer	7.565	7.565	7.543	7.561	7.565
	Hexamer	7.800	7.800	7.784	7.796	7.799
	Heptamer	7.644	7.644	7.625	7.642	7.644
SF	Monomer	-7.131	-7.132	-7.137	-7.131	-7.131
	Dimer	-7.495	-7.496	-7.496	-7.495	-7.495
	Trimer	-7.503	-7.504	-7.506	-7.503	-7.503
	Tetramer	-7.597	-7.598	-7.599	-7.597	-7.597
	Pentamer	-7.548	-7.549	-7.544	-7.548	-7.548
	Hexamer	-7.754	-7.755	-7.748	-7.754	-7.754
	Heptamer	-7.704	-7.704	-7.708	-7.703	-7.704

prominent in EA, where the 4e-integrals leads to significant I/O bottleneck. The absence of 4e-integrals makes RI/CD based methods less advantageous in IP than that in EA case. However, in case of EE and SF, the computational time is dominated by formal scaling ($O(N^6)$). Therefore, RI/CD gives very little advantage as compared to canonical EOM-MP2. Among the RI and CD based method, CD(tol = 2) is the fastest, followed by CD(tol = 3) and CD(tol = 4). The performance of RI is in between CD(tol = 2) and CD(tol = 3). The trend is uniform for IP, EA, EE, and SF.

3.3 Energies of water clusters

For convenience in the discussion, we now define some conventions to be used in this subsection. We will refer to the total energy of target state as TE and the total energy of the reference state as RE. We will use DE for target-reference energy difference. For EOM-IP, EOM-EA, EOM-EE and EOM-SF variants, the DE will correspond

to ionization energy, electron affinity, spin-conserving excitation energy and spin-flipping excitation energy, respectively. The energy-gap between two target states (or levels in case of degeneracy in the target states) will be denoted as DTE. Table 4 gives the canonical EOM-MP2 total energies for IP, EA, EE and SF target states and the relative energies of RI-EOM-MP2 and CD-EOM-MP2 methods for these variants. The corresponding excitation energies are summarized in Table 5. The standard deviations of TEs in canonical EOM-MP2 method relative to EOM-CCSD for IP, EA, EE and SF variants are 0.404, 0.431, 0.408 and 0.495 eV, respectively (see the supplementary information). This gives an idea about the error bar of EOM-MP2 methods. However, in this study, we present our RI/CD-EOM-MP2 results relative to canonical EOM-MP2 results, because that is the maximum accuracy possible to achieve in our RI/CD EOM-MP2 method. The TE values in RI-EOM-MP2 method are generally higher than the canonical EOM-MP2 values for all the variants, except the SF one.

Table 6. Energy difference (eV) between lowest two IP, EA, EE and SF target energy levels for water clusters computed using EOM-MP2, RI-EOM-MP2 and CD-EOM-MP2 methods and cc-pVDZ basis.

	Cluster	MP2	RI-MP2	CD-MP2		
				tol = 2	tol = 3	tol = 4
IP	Monomer	2.259	2.259	2.261	2.259	2.259
	Dimer	1.355	1.355	1.357	1.354	1.355
	Trimer	0.140	0.140	0.134	0.140	0.140
	Tetramer	1.322	1.322	1.324	1.322	1.322
EA	Monomer	0.728	0.729	0.730	0.729	0.728
	Dimer	0.847	0.847	0.850	0.847	0.847
	Trimer	0.575	0.575	0.575	0.575	0.575
	Tetramer	0.563	0.563	0.564	0.563	0.563
EE	Monomer	1.772	1.772	1.778	1.773	1.772
	Dimer	0.444	0.444	0.446	0.444	0.444
	Trimer	0.141	0.141	0.135	0.141	0.141
	Tetramer	1.311	1.311	1.319	1.311	1.311
SF	Monomer	7.038	7.041	7.078	7.041	7.038
	Dimer	7.399	7.402	7.462	7.401	7.399
	Trimer	7.411	7.414	7.491	7.413	7.411
	Tetramer	7.506	7.510	7.598	7.508	7.506

The maximum deviations of TEs in RI-EOM-MP2 relative to EOM-MP2 for all the EOM methods are 0.001 eV, only. However, the errors in CD variants depends strongly on the tolerance used. The standard deviation in CD-EOM-MP2 ($tol = 2$) TEs relative to EOM-MP2 for IP, EA, EE and SF variants are 0.028, 0.021, 0.018 and 0.024 eV, respectively. They are an order of magnitude less than the typical error bar of 0.2 eV -0.3 eV reported for EOM methods.⁴⁰ The error decreases on tightening the tolerance criteria. The standard deviation in errors with $tol = 3$ for IP, EA, EE and SF variants are 0.013, 0.13, 0.008 and 0.008 eV, respectively. The maximum deviations come down to 0.002 eV with $tol = 4$. It should be noted that the TE in CD-EOM-MP2 is always higher than their canonical version for all the flavors of EOM, including the SF one.

The DEs are much well reproduced in both RI and CD EOM-MP2 than that observed for TE. Table 5 presents the DE values for all flavors of EOM-MP2. The maximum deviations of DE in RI-EOM-MP2 is 0.001 eV for all flavors of EOM. The accuracy of CD-EOM-MP2 is strongly dependent on the decomposition threshold. The maximum deviation of DE with $tol = 2$ for IP, EA, EE and SF variants are 0.005, 0.009, 0.022 and 0.006 eV, respectively. The maximum error gets reduced to 0.002, 0.002, 0.005 and 0.001 eV, for IP, EA, EE and SF variants, respectively with Cholesky decomposition threshold $tol = 3$. The maximum errors come down within 0.001 eV with $tol = 4$. Table 6 summarises DTEs (ΔE_{12}) of water clusters computed for IP, EA,

EE and SF variants of RI-EOM-MP2 and CD-EOM-MP2. Both RI-EOM-MP2 and CD-EOM-MP2 predict the DTEs with good accuracy. For IP, the CD-EOM-MP2 with $tol = 4$ and RI-EOM-MP2 gives identical value as that of canonical EOM-MP2, whereas for CD-EOM-MP2 with $tol = 2, 3$, the maximum deviation increases to 0.006 eV and 0.001 eV, respectively. For EA, the maximum deviation for CD-EOM-MP2 with $tol = 2$ is 0.003 eV, whereas, CD-EOM-MP2 with $tol = 4$ gives identical value as that of canonical one. The maximum deviation for EA in the CD-EOM-MP2 with $tol = 3$ and RI-EOM-MP2 is 0.001 eV. For EE and SF, the CD-EOM-MP2 with $tol = 2$ is the least accurate with the maximum deviations of 0.008 and 0.092 eV, respectively. The maximum deviations for RI-EOM-MP2 in EE and SF is 0.001 eV and 0.004 eV, respectively. On the other hand, in the CD-EOM-MP2 with $tol = 3$, the EE and SF maximum deviations are 0.001 and 0.003 eV, respectively. The CD-EOM-MP2 with $tol = 4$ gives identical results as that of canonical value for both EE and SF.

3.4 Electron Affinity of DNA bases

The weakly bound electrons in DNA nucleic acid bases (NAB) lead to a diverse range of interesting properties,⁸⁰⁻⁸³ which has fascinated generations of theoreticians and experimentalists. A wide range of theories, starting from DFT⁸⁴⁻⁸⁷ with different functionals and basis set, to state of the art coupled cluster methods^{88,89}

were used in the simulation of electron attachment to NAB. However, the addition of an extra electron to NAB leads to multiple quasi-degenerate configurations, the simulation of which calls for a method which can go beyond the single determinant picture. The EOM-CC method is quite capable of treating such quasi-degenerate situations and can calculate electron attachment corresponding to multiple states in a single run.¹⁰ The EOM-IP-CC has been extensively used by Krylov and co-workers⁹⁰⁻⁹³ to study the ionization-induced changes in NAB.

However, the EOM-CC studies on the electron affinities of NAB are rather limited. To the best of our knowledge, the one available comes from the work of Pal and co-workers⁷⁹ and that also restricts its attention to isolated bases only. The large basis set required to model the NAB bound anions often makes conventional EOM-CC calculations difficult to perform beyond isolated DNA bases. However, our RI/CD based EOM-MP2 method because of its lower storage and memory requirements can be an efficient alternative. The electron affinity values of NAB in different theoretical methods are presented in Table 7. It can be seen that the electron affinities in all the theoretical methods are negative, which is consistent with the experimental results. It indicates the electron attached states of NAB are not bound states, rather quasi-bound or resonance states. It can be seen that the DFT method gives most scattered results ranging from highly negative values to values close to zero. The MP2 method leads to negative values. The CCSD and CCSD(T) method also leads to highly negative values and the values are almost identical to each other. However, the used basis set is too small to perform any meaningful comparison with experiments. The CASPT2 method, on the other hand, gives much less negative values. It can be seen that EOM based methods give the best agreement with experiments. The RI-EOM-MP2/aug-cc-pVDZ computed EA values are in good agreement with the EOM-CCSD/aug-cc-pVDZ results of Pal and co-workers.⁷⁹ On using the aug-cc-pVTZ basis, the absolute values of RI-EOM-MP2 computed EAs decrease by 0.04-0.08 eV from the corresponding EOM-CCSD results. However, it is justifiable, as Pal and co-workers⁷⁹ removed one of the f functions from their basis set and used aug-cc-pVDZ basis set for hydrogen. Whereas, in the present study the standard aug-cc-pVTZ basis set is used for all elements and it is well known that EA of NAB increases (i.e becomes less negative) with increase in basis set. the deviation is within the accuracy range of EOM-CCSD method. Among all the nucleobases, the uracil shows highest EA values of -0.12 eV in RI-EOM-MP2/aug-cc-pVTZ level of theory, which is in excellent agreement

Table 7. Vertical electron affinities (eV) of DNA and RNA nucleobases obtained by different theoretical methods.

Method ^a	Uracil	Thymine	Cytosine	Adenine	Guanine
experimental range ^b	-0.30~-0.22	-0.53~-0.29	-0.55~-0.32	-0.56~-0.45	-2.07~-0.08
DFT range ^c	-1.09~-0.11	-1.05~-0.28	-1.42~-0.31	-1.57~-0.34	-1.30
PMP2//MP2/aug-cc-pVDZ ^d	-0.56	-0.58	-0.73	-0.99	
CCSD//CCSD/aug-cc-pVDZ ^d	-0.63	-0.65	-0.77		
CCSD(T)//CCSD/aug-cc-pVDZ ^d	-0.64	-0.65	-0.79		
CASPT2/ANO-L 4321/321//CASSCF/ANO-L 431/21 ^d	-0.49	-0.45	-0.59	-0.74	-0.94
EOM-EA-CCSD/aug-cc-pVDZ/B3LYP/6-311++G** ^e	-0.23	-0.28	-0.27	-0.44	-0.23
EOM-EA-CCSD/aug-cc-pVTZ(-f)/B3LYP/6-311++G** ^e	-0.20	-0.24	-0.24	-0.40	-0.19
RI-EOM-EA-MP2/aug-cc-pVDZ ^f	-0.20	-0.26	-0.27	-0.46	-0.28
RI-EOM-EA-MP2/aug-cc-pVTZ ^f	-0.12	-0.17	-0.17	-0.35	-0.19

^a W/X/Y/Z denotes the single point calculations are done using W method using X basis set. The geometry was optimized using Y method and Z basis set.

^b Values taken from ref. ⁹⁴⁻⁹⁶

^c Values taken from ref. ⁹⁷

^d Values taken from ref. ⁸⁸

^e Values taken from ref. ⁷⁹

^f Present work.

with the optimized virtual orbital space (OVOS) based CCSD(T)/aug-cc-pVTZ value of -0.15 eV reported by Urban and co-workers.⁸⁹ The lowest electron affinity of -0.35 eV is displayed by adenine, which is much lower than the other theoretical methods, though it agrees well with the experimental value. The small storage requirements and favorable computational scaling of RI/CD of EOM-MP2 allows to calculate EA values beyond single DNA bases, but that is outside the scope of the present study and will be followed in some future work.

4. Conclusions

We have presented the formulation of cost-effective RI and CD based EOM-MP2 approximations to EOM-CCSD. The MP2 approximation reduces the timing and cost by eliminating iterative CCSD procedure. Moreover, use of RI (or CD) reduces the cost further with respect to the computational size of the system, thereby widening the applicability of EOM-MP2 to molecules of larger size such as DNA nucleobases. Although the RI/CD approximating doesn't change overall scaling of the methods, it greatly reduces the storage requirements and I/O bottlenecks and thereby resulting in a major speed-up. The iterative decomposition procedure in CD-EOM-MP2 is the rate-determining step and can alter the overall timing of the job depending on the CD threshold chosen. In terms of accuracy, RI-EOM-MP2 and CD-EOM-MP2 with $tol = 3, 4$ are comparable and provide fairly good results for DEs with maximum absolute error (MAE) with respect to the conventional EOM-MP2 to be of 0.002 eV for IPs, EAs and SFs, whereas of 0.005 eV for EEs. For DTEs, these methods perform even better with the MAE of 0.001 eV with respect to the EOM-MP2 values for IP, EA and EE, whereas of 0.004 eV for SF. In terms of computational timings, RI seems to work better than CD with $tol = 3, 4$ for water clusters. However, unlike RI, CD does not require use of a pre-optimized auxiliary basis and is thus, more advantageous when one needs to use non-standard basis for computations. Moreover, parallelization may significantly improve the timings for CD iterations. However, results for a few states in a series of water clusters is far from being conclusive enough and one can comment on the relative efficiency and accuracy of RI vs CD implementation of EOM-MP2 methods only after a more elaborate benchmarking with systems of big computational size, which is unfortunately outside the scope of the present study. For molecules of small and moderate computational size as discussed in this work, RI-EOM-MP2 is a preferred method over CD-EOM-MP2 in case of availability of a pre-optimized auxiliary

basis as it saves the time required for iterative CD procedure. However, CD-EOM-MP2 can be used for a more general choice of basis and suitable threshold can be chosen to tune the desired accuracy.

The small storage requirements and favorable computational scaling of RI/CD-EOM-MP2 allows one to use these methods for computing IPs, EAs, EEs and SFs of large molecules and clusters, which is beyond the reach of both the canonical EOM-MP2 as well as RI/CD-EOM-CCSD.

Supplementary Information

The geometric parameters (GP), nuclear repulsion energies (NRE), Hartree-Fock energies (EHF) and MP2 energies (EMP2) of water clusters and DNA bases are provided in the Supporting information (SI). For water clusters, the CCSD total energies (ECCSD) are also provided in the supplementary information. Following is the order in which data is provided in SI.

- GP, NRE, EHF, EMP2 and ECCSD of Water clusters (monomers to heptamer)
- GP, NRE, EHF and EMP2 of DNA bases (adenine, cytosine, guanine, tyamine and uracil)

The Supplementary Information is available at www.ias.ac.in/chemsci.

Acknowledgements

The authors are thankful to Prof. Anna I. Krylov for discussions and her valuable suggestions. PUM and DK acknowledge the Department of Science and Technology, India for financial support under the grant no. SR/FT/CS-18/2011.

References

1. Shavitt I and Bartlett R J 2009 *Many-Body Methods in Chemistry and Physics* (Cambridge University Press: Cambridge)
2. Helgaker T, Jørgensen P and Olsen J 2014 *Molecular Electronic-Structure Theory* (Wiley: Hoboken)
3. Rowe D J 1968 Equations-of-motion method and the extended shell model *Rev. Mod. Phys.* **40** 153
4. Trofimov A B, Krivdina I L, Weller J and Schirmer J 2006 Algebraic-diagrammatic construction propagator approach to molecular response properties *Chem. Phys.* **329** 1
5. Bozkaya U 2013 The extended Koopmans theorem for orbital-optimized methods: accurate computation of ionization potentials *J. Chem. Phys.* **139** 154105
6. Bozkaya U 2014 Accurate electron affinities from the extended Koopmans theorem based on orbital-optimized methods *J. Chem. Theory Comput.* **10** 2041
7. Geertsen J, Rittby M and Bartlett R J 1989 The equation-of-motion coupled-cluster method: excitation energies of Be and CO *Chem. Phys. Lett.* **164** 57

8. Stanton J F and Bartlett R J 1993 The equation of motion coupled-cluster method. A systematic biorthogonal approach to molecular excitation energies, transition probabilities, and excited state properties *J. Chem. Phys.* **98** 7029
9. Krylov A I 2008 Equation-of-motion coupled-cluster methods for open-shell and electronically excited species: the hitchhiker's guide to Fock space *Annu. Rev. Phys. Chem.* **59** 433
10. Nooijen M and Bartlett R J 1995 Equation of motion coupled cluster method for electron attachment *J. Chem. Phys.* **102** 3629
11. Kowalski K and Piecuch P 2000 The active-space equation-of-motion coupled-cluster methods for excited electronic states: the EOMCCSDt approach *J. Chem. Phys.* **113** 8490
12. Bartlett R J 2012 Coupled-cluster theory and its equation-of-motion extensions *Wiley Interdiscip. Rev. Comput. Mol. Sci.* **2** 126
13. Pal S and Mukherjee D 1989 Use of cluster expansion methods in the openshell correlation problem *Adv. Quantum Chem.* **20** 291
14. Pal S, Rittby M, Bartlett R J, Sinha D and Mukherjee D 1988 Molecular applications of multireference coupled-cluster methods using an incomplete model space: direct calculation of excitation energies *J. Chem. Phys.* **88** 4357
15. Pal S 2010 Fock space multi-reference coupled-cluster method for energies and energy derivatives *Mol. Phys.* **108** 3033
16. Meissner L 1998 Fock-space coupled-cluster method in the intermediate Hamiltonian formulation: model with singles and doubles *J. Chem. Phys.* **108** 9227
17. Musial M and Bartlett R J 2008 Intermediate Hamiltonian Fock-space multireference coupled-cluster method with full triples for calculation of excitation energies *J. Chem. Phys.* **129** 044101
18. Musial M and Bartlett R J 2008 Multireference Fock-space coupled-cluster and equation-of-motion coupled-cluster theories: the detailed interconnections *J. Chem. Phys.* **129** 134105
19. Christiansen O, Koch H and Jørgensen P 1995 Response functions in the CC3 iterative triple excitation model *J. Chem. Phys.* **103** 7429
20. Kucharski S A, Woch M, Musia M and Bartlett R J 2001 Coupled-cluster theory for excited electronic states: the full equation-of-motion coupled-cluster single, double, and triple excitation method *J. Chem. Phys.* **115** 8263
21. Manohar P U and Krylov A I 2008 A noniterative perturbative triples correction for the spin-flipping and spin-conserving equation-of-motion coupled-cluster methods with single and double substitutions *J. Chem. Phys.* **129** 194105
22. Manohar P U, Stanton J F and Krylov A I 2009 Perturbative triples correction for the equation-of-motion coupled-cluster wave functions with single and double substitutions for ionized states: theory, implementation, and examples *J. Chem. Phys.* **131** 114112
23. Woch M, Lodriguito M D, Piecuch P, and Gour J R 2006 Two new classes of non-iterative coupled-cluster methods derived from the method of moments of coupled-cluster equations *Mol. Phys.* **104** 2149
24. Gour J R, Piecuch P and Wloch M 2005 Active-space equation-of-motion coupled-cluster methods for excited states of radicals and other open-shell systems: EA-EOMCCSDt and IP-EOMCCSDt *J. Chem. Phys.* **123** 134113
25. Kowalski K and Piecuch P 2002 Extension of the method of moments of coupled-cluster equations to excited states: the triples and quadruples corrections to the equation-of-motion coupled-cluster singles and doubles energies *J. Chem. Phys.* **116** 7411
26. Stanton J F and Gauss J A 1996 A simple correction to final state energies of doublet radicals described by equation-of-motion coupled cluster theory in the singles and doubles approximation *Theor. Chim. Acta* **93** 303
27. Kállay M and Gauss J 2004 Calculation of excited-state properties using general coupled-cluster and configuration-interaction models *J. Chem. Phys.* **121** 9257
28. Dutta A K, Pal S and Ghosh D 2013 Perturbative approximations to single and double spin flip equation of motion coupled cluster singles doubles methods *J. Chem. Phys.* **139** 124116
29. Dutta A K, Vaval N and Pal S 2013 Performance of the EOMIP-CCSD (2) method for determining the structure and properties of doublet radicals: a benchmark investigation *J. Chem. Theory Comput.* **9** 4313
30. Dutta A K, Gupta J, Pathak H, Vaval N and Pal S 2014 Partitioned EOMEA-MBPT (2): an efficient N 5 scaling method for calculation of electron affinities *J. Chem. Theory Comput.* **10** 1923
31. Dutta A K, Vaval N and Pal S 2015 EOMIP-CCSD (2)*: an efficient method for the calculation of ionization potentials *J. Chem. Theory Comput.* **11** 2461
32. Bhattacharya D, Dutta A K, Gupta J and Pal S 2015 Perturbative order analysis of the similarity transformed Hamiltonian in Fock-space coupled cluster theory: difference energy and electric response properties *Mol. Phys.* **113** 2046
33. Bartlett R J 1981 Many-body perturbation theory and coupled cluster theory for electron correlation in molecules *Annu. Rev. Phys. Chem.* **32** 359
34. Kucharski S A and Bartlett R J 1986 Fifth-order many-body perturbation theory and its relationship to various coupled-cluster approaches *Adv. Quantum Chem.* **18** 281
35. Nooijen M and Snijders J G 1995 Second order many-body perturbation approximations to the coupled cluster Greens function *J. Chem. Phys.* **102** 1681
36. Stanton J F and Gauss J 1995 Perturbative treatment of the similarity transformed Hamiltonian in equation-of-motion coupled-cluster approximations *J. Chem. Phys.* **103** 1064
37. Ghosh D 2014 Perturbative approximation to hybrid equation of motion coupled cluster/effective fragment potential method *J. Chem. Phys.* **140** 094101
38. Gwaltney S R, Nooijen M and Bartlett R J 1996 Simplified methods for equation-of-motion coupled-cluster excited state calculations *Chem. Phys. Lett.* **248** 189
39. Christiansen O, Koch H and Jørgensen P 1995 The second-order approximate coupled cluster singles and doubles model CC2 *Chem. Phys. Lett.* **243** 409
40. Schreiber M, Silva-Junior M R, Sauer S P A and Thiel W 2008 Benchmarks for electronically excited states:

- CASPT2, CC2, CCSD, and CC3 *J. Chem. Phys.* **128** 134110
41. Jacquemin D, Duchemin I and Blase X 2016 Benchmarking the Bethe–Salpeter formalism on a standard organic molecular set *J. Chem. Theory Comput.* **12** 5477
 42. Dutta A K, Vaval N and Pal S Assessment of Low Scaling Approximations to EOM-CCSD Method for Ionization Potential Communicated: <http://arxiv.org/abs/1708.01293>
 43. Byrd J N, Rishi V, Perera A and Bartlett R J 2015 Approximating electronically excited states with equation-of-motion linear coupled-cluster theory *J. Chem. Phys.* **143** 164103
 44. Weigend F, Häser M, Patzelt H and Ahlrichs R 1998 RI-MP2: optimized auxiliary basis sets and demonstration of efficiency. *Chem. Phys. Lett.* **294** 143
 45. Whitten J L 1973 Coulombic potential energy integrals and approximations *J. Chem. Phys.* **58** 4496
 46. Feyereisen M, Fitzgerald G and Komornicki A 1993 Use of approximate integrals in ab initio theory. An application in MP2 energy calculations *Chem. Phys. Lett.* **208** 359
 47. Vahtras O, Almlöf J, Feyereisen M W 1993 Integral approximations for LCAO-SCF calculations *Chem. Phys. Lett.* **213** 514
 48. Neese F 2003 An improvement of the resolution of the identity approximation for the formation of the Coulomb matrix *J. Comput. Chem.* **24** 1740
 49. Jung Y, Sodt A, Gill P M W and Head-Gordon M 2005 Auxiliary basis expansions for large-scale electronic structure calculations *Proc. Natl. Acad. Sci. USA* **102** 6692
 50. Distasio R A, Steele R P, Rhee Y M, Shao Y and Head-Gordon M 2007 An improved algorithm for analytical gradient evaluation in resolution-of-the-identity second-order Møller–Plesset perturbation theory: application to alanine tetrapeptide conformational analysis *J. Comput. Chem.* **28** 839
 51. Beebe N H F and Linderberg J 1977 Simplifications in the generation and transformation of two-electron integrals in molecular calculations *Int. J. Quantum Chem.* **12** 683
 52. Koch H, de Merás A and Pedersen T B 2003 Reduced scaling in electronic structure calculations using Cholesky decompositions *J. Chem. Phys.* **118** 9481
 53. Aquilante F, Lindh R and Pedersen T B 2007 Unbiased auxiliary basis sets for accurate two-electron integral approximations *J. Chem. Phys.* **127** 114107
 54. Pedersen T B, Aquilante F and Lindh R 2009 Density fitting with auxiliary basis sets from Cholesky decompositions *Theor. Chem. Acc.* **124** 1
 55. Aquilante F, Gagliardi L, Pedersen T B and Lindh R 2009 Atomic Cholesky decompositions: a route to unbiased auxiliary basis sets for density fitting approximation with tunable accuracy and efficiency *J. Chem. Phys.* **130** 154107
 56. Weigend F, Kattannek M and Ahlrichs R 2009 Approximated electron repulsion integrals: Cholesky decomposition versus resolution of the identity methods *J. Chem. Phys.* **130** 164106
 57. Aquilante F, Boman L, Boström J, Koch H, Lindh R, de Merás A S and Pedersen T B 2011 Cholesky decomposition techniques in electronic structure theory. In Zalesny R, Papadopoulos M G, Mezey P G and Leszczynski J (Eds.) *Linear-Scaling Techniques in Computational Chemistry and Physics SE-13, Volume 13 of Challenges and Advances in Computational Chemistry and Physics* (Berlin: Springer) pp. 301–343
 58. Bozkaya U 2014 Analytic energy gradients and spin multiplicities for orbital-optimized second-order perturbation theory with density-fitting approximation: an efficient implementation *J. Chem. Theory Comput.* **10** 4389
 59. Rendell A P and Lee T J 1994 Coupled-cluster theory employing approximate integrals: an approach to avoid the input/output and storage bottlenecks *J. Chem. Phys.* **101** 400
 60. Boström J, Pitoák M, Aquilante F, Neogrady P, Pedersen T B and Lindh R 2012 Coupled cluster and Møller–Plesset perturbation theory calculations of noncovalent intermolecular interactions using density fitting with auxiliary basis sets from Cholesky decompositions *J. Chem. Theory Comput.* **8** 1921
 61. Pitoák M, Aquilante F, Hobza P, Neogrady P, Noga J and Urban M 2011 Parallelized implementation of the CCSD (T) method in MOLCAS using optimized virtual orbitals space and Cholesky decomposed two-electron integrals *Collect. Czechoslov. Chem. Commun.* **76** 713
 62. DePrince A E and Sherrill C D 2013 Accuracy and efficiency of coupled-cluster theory using density fitting/Cholesky decomposition, frozen natural orbitals, and a t1-transformed Hamiltonian *J. Chem. Theory Comput.* **9** 2687
 63. Gyrrffy W, Shiozaki T, Knizia G and Werner H J 2013 Analytical energy gradients for second-order multireference perturbation theory using density fitting *J. Chem. Phys.* **138** 104104
 64. Hättig C 2003 Geometry optimizations with the coupled-cluster model CC2 using the resolution-of-the-identity approximation *J. Chem. Phys.* **118** 7751
 65. Köhn A and Hättig C 2003 Analytic gradients for excited states in the coupled-cluster model CC2 employing the resolution-of-the-identity approximation *J. Chem. Phys.* **119** 5021
 66. Schütz M, Werner H J, Lindh R and Manby F R 2004 Analytical energy gradients for local second-order Møller–Plesset perturbation theory using density fitting approximations *J. Chem. Phys.* **121** 737
 67. Ledermüller K and Schütz M 2014 Local cc2 response method based on the laplace transform: analytic energy gradients for ground and excited states *J. Chem. Phys.* **140** 164113
 68. Bozkaya U 2014 Derivation of general analytic gradient expressions for density-fitted post-Hartree–Fock methods: an efficient implementation for the density-fitted second-order Møller–Plesset perturbation theory *J. Chem. Phys.* **141** 124108
 69. Bozkaya U 2014 Orbital-optimized second-order perturbation theory with density-fitting and cholesky decomposition approximations: An efficient implementation *J. Chem. Theory Comput.* **10** 2371
 70. Epifanovsky E, Zuev D, Feng X, Khistyayev K, Shao Y and Krylov A I 2013 General implementation of the resolution-of-the-identity and Cholesky representations

- of electron repulsion integrals within coupled-cluster and equation-of-motion methods: Theory and benchmarks *J. Chem. Phys.* **139** 134105
71. Dutta A K, Neese F and Izsk R 2016 Speeding up equation of motion coupled cluster theory with the chain of spheres approximation *J. Chem. Phys.* **144** 034102
72. Hättig C and Weigend F 2000 CC2 excitation energy calculations on large molecules using the resolution of the identity approximation *J. Chem. Phys.* **113** 5154
73. Hirao K and Nakatsuji H 1982 A generalization of the Davidson's method to large nonsymmetric eigenvalue problems *J. Comput. Phys.* **45** 246
74. Löwdin P O 1963 Studies in perturbation theory *J. Mol. Spectrosc.* **10** 12
75. Aidas K, Angeli C, Bak K L, Bakken V, Bast R, Boman L, Christiansen O, Cimiraglia R, Coriani S, Dahle P, Dalskov E K, Ekström U, Enevoldsen T, Eriksen J J, Ettenhuber P, Fernández B, Ferrighi L, Fliegl H, Frediani L, Hald K, Halkier A, Hättig C, Heiberg H, Helgaker T, Hennum A C, Hetttema H, Hjertenæs S E, Høst S, Høyvik I M, Iozzi M F, Jansík B, Jensen H J A, Jónsson D, Jørgensen P, Kauczor J, Kirpekar S, Kjærgaard T, Klopper W, Knecht S, Kobayashi R, Koch H, Kongsted J, Krapp A, Kristensen K, Ligabue A, Lutnæs O B, Melo J I, Mikkelsen K V, Myhre R H, Neiss C, Nielsen C B, Norman P, Olsen J, Olsen J M H, Osted A, Packer M J, Pawłowski F, Pedersen T B, Provasi P F, Reine S, Rinkevicius Z, Ruden T A, Ruud K, Rybkin V V, Saek P, Samson C C M, de Merás A S, Saue T, Sauer S P A, Schimmelpfennig B, Sneskov K, Steindal A H, Sylvester-Hvid K O, Taylor P R, Teale A M, Tellgren E I, Tew D P, Thorvaldsen A J, Thøgersen L, Vahtras O, Watson M A, Wilson D J D, Ziolkowski M and Ågren H 2014 The Dalton quantum chemistry program system *Wiley Interdiscip. Rev. Comput. Mol. Sci.* **4** 269
76. Furche F, Ahlrichs R, Hättig C, Klopper W, Sierka M and Weigend F 2014 Turbomole *Wiley Interdiscip. Rev. Comput. Mol. Sci.* **4** 91
77. Neese F 2012 The ORCA program system *Wiley Interdiscip. Rev. Comput. Mol. Sci.* **2** 73
78. Shao Y, Gan Z, Epifanovsky E, Gilbert A T B, Wormit M, Kussmann J, Lange A W, Behn A, Deng J, Feng X, Ghosh D, Goldey M, Horn P R, Jacobson L D, Kaliman I, Khaliullin R Z, Kusn T, Landau A, Liu J, Proynov E I, Rhee Y M, Richard R M, Rohrdanz M A, Steele R P, Sundstrom E J, Woodcock III H L, Zimmerman P M, Zuev D, Albrecht B, Alguire E, Austin B, Beran G J O, Bernard Y A, Berquist E, Brandhorst K, Bravaya K B, Brown S T, Casanova D, Chang C M, Chen Y, Chien S H, Closser K D, Crittenden D L, Diedenhofen M, DiStasio Jr RA, Do H, Dutoi A D, Edgar R G, Fatehi S, Fusti-Molnar L, Ghysels A, Golubeva-Zadorozhnaya A, Gomes J, Hanson-Heine M W D, Harbach P H P, Hauser A W, Hohenstein E G, Holden Z C, Jagau T C, Ji H, Kaduk B, Khistyayev K, Kim J, Kim J, King R A, Klunzinger P, Kosenkov D, Kowalczyk T, Krauter C M, Lao K U, Laurent A D, Lawler K V, Levchenko S V, Lin C Y, Liu F, Livshits E, Lochan R C, Luenser A, Manohar P, Manzer S F, Mao S P, Mardirossian N, Marenich A V, Maurer S A, Mayhall N J, Neuscammann E, Oana C M, Olivares-Amaya R, O'Neill D P, Parkhill J A, Perrine T M, Peverati R, Prociuk A, Rehn D R, Rosta E, Russ N J, Sharada S M, Sharma S, Small D W, Sodt A, Stein T, Stück D, Su Y C, Thom A J W, Tsuchimochi T, Vanovschi V, Vogt L, Vydrov O, Wang T, Watson M A, Wenzel J, White A, Williams C F, Yang J, Yeganeh S, Yost S R, You Z Q, Zhang I Y, Zhang X, Zhao Y, Brooks B R, Chan G K L, Chipman D M, Cramer C J, Goddard III W A, Gordon M S, Hehre W J, Klamt A, Schaefer III H F, Schmidt M W, Sherrill C D, Truhlar D G, Warshel A, Xu X, Aspuru-Guzik A, Baer R, Bell A T, Besley N A, Chai J D, Dreuw A, Dunietz B D, Furlani T R, Gwaltney S R, Hsu C P, Jung Y, Kong J, Lambrecht D S, Liang W, Ochsenfeld C, Rassolov V A, Slipchenko L V, Subotnik J E, VanVoorhis T, Herbert J M, Krylov A I, Gill P M W and Head-Gordon M 2015 Advances in molecular quantum chemistry contained in the Q-Chem4 program package *Mol. Phys.* **113** 184
79. Dutta A K, Sengupta T, Vaval N and Pal S 2015 Electron attachment to DNA and RNA nucleobases: an EOMCC investigation *Int. J. Quantum Chem.* **115** 753
80. Huels M A, Hahndorf I, Illenberger E and Sanche L 1998 Resonant dissociation of DNA bases by subionization electrons *J. Chem. Phys.* **108** 1309
81. Wheldon T E 2000 Radiation physics and genetic targeting: new directions for radiotherapy *Phys. Med. Biol.* **45** R77
82. Schuster G B 2000 Long-range charge transfer in DNA: transient structural distortions control the distance dependence *Acc. Chem. Res.* **33** 253
83. Holmlin R E, Dandliker P J and Barton J K 1997 Charge transfer through the DNA base stack *Angew. Chem. Int. Ed. Engl.* **36** 2714
84. Crespo-Hernández* C E, Close D M, Gorb L and Leszczynski J Determination of redox potentials for the Watson-Crick base pairs, DNA nucleosides, and relevant nucleoside analogues 2007 *J. Phys. Chem. B* **111** 5386
85. Gu J, Xie Y and Schaefer H F 2006 Understanding electron attachment to the DNA double helix: the Thymidine Monophosphate-Adenine pair in the gas phase and aqueous solution *J. Phys. Chem. B* **110** 19696
86. Li X, Cai Z and Sevilla M D 2002 DFT Calculations of the electron affinities of nucleic acid bases: dealing with negative electron affinities *J. Phys. Chem. A* **106** 1596
87. Vera D M A and Pierini A B 2004 Species with negative electron affinity and standard DFT methods *Phys. Chem. Chem. Phys.* **6** 2899
88. Roca-Sanjuán D, Merchán M, Serrano-Andrés L and Rubio M 2008 Ab initio determination of the electron affinities of DNA and RNA nucleobases *J. Chem. Phys.* **129** 095104
89. Dedíková P, Demovič L, Pitoák M, Neogrady P and Urban M 2009 CCSD(T) calculations of the electron affinity of the uracil molecule *Chem. Phys. Lett.* **481** 107
90. Bravaya K B, Kostko O, Dolgikh S, Landau A, Ahmed M and Krylov A I 2010 Electronic structure and spectroscopy of nucleic acid bases: ionization energies, ionization-induced structural changes, and photoelectron spectra *J. Phys. Chem. A* **114** 12305
91. Bravaya K B, Epifanovsky E and Krylov A I 2012 Four bases score a run: ab initio calculations quantify a cooperative effect of H-bonding and π -stacking on the ionization energy of adenine in the AATT tetramer *J. Phys. Chem. Lett.* **3** 2726

92. Golubeva A A and Krylov A I 2009 The effect of [small pi]-stacking and H-bonding on ionization energies of a nucleobase: uracil dimer cation *Phys. Chem. Chem. Phys.* **11** 1303
93. Bravaya K B, Kostko O, Ahmed M and Krylov A I 2010 The effect of [small pi]-stacking, H-bonding, and electrostatic interactions on the ionization energies of nucleic acid bases: adenine-adenine, thymine-thymine and adenine-thymine dimers *Phys. Chem. Chem. Phys.* **12** 2292
94. Aflatooni K, Gallup G A, and Burrow* P D 1998 Electron attachment energies of the DNA bases *J. Phys. Chem. A* **102** 6205
95. Periquet V, Moreau A, Carles S, Schermann J P and Desfrancois C 2000 Cluster size effects upon anion solvation of N-heterocyclic molecules and nucleic acid bases *J. Electron Spectrosc. Relat. Phenom.* **106** 141
96. Harinipriya S and Sangaranarayanan M V 2003 Estimation of the electron affinities of nucleobases using thermochemical data and structural considerations *J. Mol. Struct.* **644** 133
97. Gu J, Leszczynski J and Schaefer I I I H F 2012 Interactions of electrons with bare and hydrated biomolecules: from nucleic acid bases to DNA segments *Chem. Rev.* **112** 5603