Transition probability approach for direct calculation of coefficients of configuration interaction wave function

Arijit Bag*

Division of Chemical Sciences, Indian Institute of Science Education and Research Kolkata, Mohanpur 741 246, India

To reduce the computation cost of configuration interaction (CI) method, a novel technique is used to calculate the coefficients of doubly excited determinants directly from orbital energies, orbital overlap matrix and electron population obtained from Hartree–Fock level run. This approach to approximate the coefficients of CI wave function is termed as transition probability approximated CI (TPA-CI). In principle, calculated dynamical electron correlation energies of TPA-CI and full CI (FCI) are equivalent. It is observed that computed TPA-CI correlation energies of hydrogen, water, ammonia and ozone are close to the FCI values, within 5% error. The potential energy curve of hydrogen molecule is also studied, and it is found that the energy is minimum at its equilibrium bond length.

Keywords: Configuration interaction, potential energy curve, transition probability approximation, wave function.

AMONG the post Hartree–Fock (HF) methods[1,2], configuration interaction (CI)[3,4] and coupled cluster (CC)[5–7] methods are most popular and considered as a benchmark for all computational calculations. Dynamical electron correlation has been treated properly in CI formulation. Yet we require good alternatives as CI is computationally too expensive to calculate full CI (FCI) energy even for a tri- or tetra-atomic molecule with reasonable basis set. On the other hand, CC and Møller–Plasset (MP) perturbation theory are two good alternatives[8,9]. Though FCI and full CC are equivalent, truncated CC has advantages over truncated CI as the latter is not size-consistent, unlike the former. Also, due to the presence of exponential character of the CC wave function, truncated CC takes care of the contribution of higher excited determinants. Thus, truncated CC is better than truncated CI at the same level of truncation. Presently, CC methods are frequently used for property calculation of small molecules[10–14]. However, due to requirement of large computing facilities, neither CI nor CC is used for computer-aided molecular modelling, drug designing or reaction mechanism studies, where currently only density functional theory (DFT)-based methods are used.

Energy calculation using CI only depends on the coefficients of doubly excited determinants. So far, there is no method to calculate coefficients of doubly excited determinants of CI wave function as the equations used to solve them are coupled with other classes (singly excited, triply excited, etc.) of determinants. In general, we use iterative method to solve these equations, which requires huge computational facilities. Any technique to obtain values of coefficients of doubly excited determinants of CI wave function would help implement the CI method for larger systems, where currently only DFT-based methods are applicable.

In the present approach, called transition probability approximated CI (TPA-CI), the coefficients of doubly excited determinants are calculated using transition probability of electrons from the occupied orbitals to virtual orbitals. The transition probabilities are calculated from electron population, orbital energy and overlap integrals.

The CI wave function is as follows:

$$|\Psi\rangle = C_0\,|\Psi_0\rangle + \sum_{a,r} C_a^r\,|\Psi_a^r\rangle + \sum_{a<b,r<s} C_{ab}^{rs}\,|\Psi_{ab}^{rs}\rangle + \cdots. \quad (1)$$

Now, from Brillouin's theorem we get $\langle\Psi_0|H|\Psi_a^r\rangle = 0$. The triple and higher excitations do not mix with $|\Psi_0\rangle$, as they differ from it by more than two orbitals. Thus, the expression for CI correlation energy (E_{corr}) is

$$E_{\text{corr}} = \sum_{a,b,r,s} C_{ab}^{rs}\langle\Psi_0|H|\Psi_{ab}^{rs}\rangle \quad (2)$$

where a and b are occupied orbitals, and r and s are virtual orbitals. In eq. (2) different coefficients of doubly excited determinants like C_{ab}^{rs} are unknown variables. In CI, these unknown variables are evaluated by an iterative method which requires diagonalization and renormalization. Also, doubly excited determinants are coupled with other excited determinants. Thus, computation time and storage requirement increase exponentially with increase in the number of basis sets. In the present study, an alternative method is used to find the values of these coefficients. In this approach, the coefficients of doubly excited determinants like C_{ab}^{rs} are calculated from the transition probability of electrons from a, b orbitals to r, s orbitals. This is because the excited determinants in CI wave function are constructed by replacing one or more occupied atomic orbitals (AO) depending upon the order of excitations, with the same number of virtual orbitals. Thus, the coefficient of any specific order excited determinant would be the transition probability of that excitation. In the present approach (TPA-CI), it is considered that the transition probability of electrons from orbitals a and b (occupied orbitals) to orbitals r and s (unoccupied orbitals) is proportional to the electron density at a and b, orbital overlap between a and r (S_a^r), and b and s (S_b^s), as well as partition of electron distribution probability

---

*e-mail: bagarijit@gmail.com

between a and r, and b and s. Partition of electron distribution probability is taken as an exponential function of the orbital energy difference, i.e. $e^{(E_a-E_r)}$ and $e^{(E_b-E_s)}$. For non-degenerate states, the transition probability is inversely proportional to their orbital energy difference. If we consider proportionality constant as unity we get an expression for the coefficients of doubly excited determinants as follows

$$C_{ab}^{rs} = \frac{\rho_a \rho_b S_a^r S_b^s \, e^{-\Delta E_{ar} \Delta E_{bs}}}{\Delta E_{ar} \Delta E_{bs}} \quad (3)$$

where ρ_a and ρ_b are the electron densities of orbitals a and b; ΔE_{ar} and ΔE_{bs} are the orbital energy difference between a and r and b and s respectively. Thus, the correlation energy expression of TPA-CI is as follows

$$E_{\text{corr}} = \frac{1}{8} \sum_{a,b,r,s} C_{ab}^{rs} \langle rs \parallel ab \rangle. \quad (4)$$

As values of coefficients of doubly excited determinants are evaluated using eq. (3), we can say, in principle, that calculated correlation energy using eq. (4) is an approximation to FCI correlation energy.

In this study GAMESS (Version 6 SEP 2001 (R5), Iowa State University)[15] is used for HF calculation. Mulliken charge population is used as electron density of the respective atomic orbitals. Total energy is the sum of HF energy and TPA-CI correlation energy. Overlap matrix is taken from the GAMESS output file. Two-electron integrals in molecular orbital basis are used in eq. (4). Calculation of coefficients of doubly excited determinants and correlation energy is done using in home code.

In TPA-CI computation, first, we have to run HF, similar to other post-HF methods. Thus, computation time and requirement up to this point are the same as CI, CC and perturbation methods. After the HF run, we have to compute two-electron integrals in molecular orbital basis from those in atomic basis and MO coefficients. This is the most time-consuming part in TPA-CI calculation. However, time and computation cost are the same as in other methods. Calculation of coefficients of doubly excited determinants and correlation energy is done in four nested 'do' loops among which the two outer loops run from 1 to m, where m is the number of occupied orbitals and the two inner loops run from $m + 1$ to N where N is the total number of basis sets. Thus, at this stage we do not need any storage for the calculation of correlation energy, though, we can store the coefficients of doubly excited determinants for further calculations. In TPA-CI, we get these coefficients directly. There is no need for diagonalization or reconstruction, as is required in CI calculation. Time required for the calculation of correlation energy in TPA-CI is proportional to $m^2(N - m)^2$, which is much less than N^4 for large values of N. On the other hand, in general, the time required for energy calculation in CI and CC is proportional to N^5.

The TPA-CI correlation energy of hydrogen molecule in the four different basis sets is calculated and compared with FCI results to test the efficacy of the present method. The potential energy curve of hydrogen molecules is studied to check whether the present theory is consistent with distorted geometry or not. The TPA-CI correlation energy of ozone, water and ammonia are also computed and compared with FCI correlation energy to test the applicability of the present method for polyatomic molecules.

Hydrogen molecule is very small and hence FCI calculation in different basis sets is possible. Thus, this molecule is used for comparison of TPA-CI and FCI correlation energies. STO-3G, 631-G, cc-pVDZ and cc-pVTZ basis sets are used for this purpose. An equilibrium bond length, i.e. 1.401 a.u. is used for all calculations[16]. Table 1 presents results of the calculation. TPA-CI correlation energies agree well with FCI correlation energies, except for STO-3G basis set, which is 16.57% higher than the corresponding FCI correlation energy. It is observed that TPA-CI correlation energies of STO-3G, 631-G and cc-pVDZ basis sets are higher than the corresponding FCI correlation energies, though the percentage of error decreases with increase in the number of basis sets. This may be due to higher electron density in the occupied orbitals when the number of basis sets is less. However, the TPA-CI correlation energy of hydrogen molecules in cc-pVTZ basis set is less than the corresponding FCI correlation energy; also the error is only 2.74%. Thus, we may conclude that for larger basis sets TPA-CI and FCI yield equivalent results.

To test the accuracy of the present approach at different distorted geometries, the potential energy curve of hydrogen molecules is studied taking 13 different H–H bond lengths, on both sides of its equilibrium bond length. cc-pVDZ basis set is used for this purpose. Calculated results are plotted and presented in Figure 1. It is observed that the potential energy curve of HF is above the TPA-CI and FCI curves, which is expected. However, the TPA-CI curve is above the FCI curve for distorted bond length and below the FCI curve around the equilibrium bond length. The nature of the TPA-CI curve is also different from HF and FCI curves. The latter are smooth all along the plot, while the former has a pocket-like structure around equilibrium bond length. The reason for this is still not known. It may be due to the contribution of singly, triply or higher-order excited determinants to the ground state energy at any distorted geometry. In the present work, TPA-CI ground state energy is calculated taking only the coefficient of doubly excited determinant. Thus, at equilibrium geometry both FCI and TPA-CI yield nearly the same energy. However, for distorted geometries while FCI calculation includes contributions from

Table 1. TPA-CI energy of hydrogen molecule in different basis sets

Basis set	RHF energy (a.u.)	TPA-CI correlation energy (a.u.)	FCI correlation energy (a.u.)	Error in correlation energy (%)
STO-3g	−1.1167	0.02396	0.02056	16.57
6-31G	−1.1267	0.02611	0.02494	4.72
cc-pVDZ	−1.1287	0.03593	0.03469	3.55
cc-pVTZ	−1.1330	0.03839	0.03947	2.74

Table 2. TPA-CI correlation energy of ozone, water and ammonia

Molecule	Basis set	TPA-CI correlation energy (a.u.)	FCI correlation energy (a.u.)	Error in correlation energy (%)
O_3	DZ (dunning)	0.22416	0.23807	5.84
NH_3	631-G	0.12353	0.12846	3.84
H_2O	cc-pVDZ	0.06955	0.07168	2.97

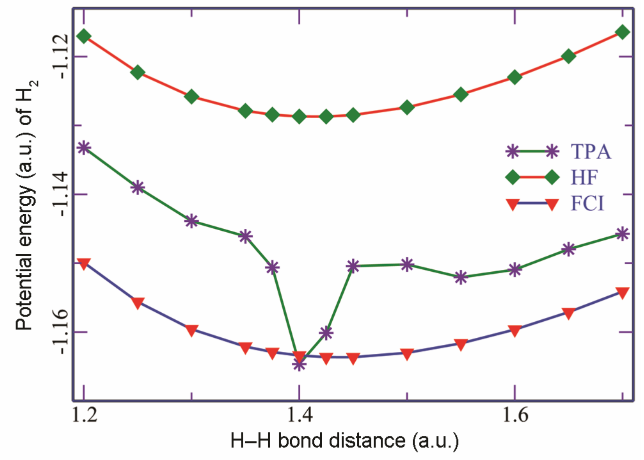

Figure 1. Potential energy curve of hydrogen molecule.

other excited determinants, TPA-CI calculation takes only the contribution of doubly excited determinants. Thus, we observe discrepancies between FCI and TPA-CI energies for distorted geometries. Equilibrium bond length calculated from the potential energy curve of TPA-CI (1.4 au) matches well with the experimental value (1.401 au)[16] compared to HF (1.425 au) and FCI (1.45 au).

TPA-CI is also implemented for polyatomic molecules. Ozone, water and ammonia are used for this purpose. For all three molecules, first, geometry optimization is done at the HF level with the same basis and then TPA-CI and FCI calculation is performed separately. Table 2 provides results of the calculation. For all three molecules, error in TPA-CI correlation energy with respect to FCI is very small. We can get 94–97% correlation energy using TPA-CI technique. Here we observe that TPA-CI correlation energy is less than the FCI correlation energy, though a small basis set is used. This may be due to the use of optimized geometry which may be different from the exact geometry. We have also observed that at non-equilibrium geometry TPA-CI correlation energy is less than the FCI correlation energy, irrespective of size of the basis set.

Thus, we can conclude that the TPA-CI approach is applicable for electronic structure calculation of molecules. Its accuracy in dynamical electron correlation energy calculation is equivalent to FCI at moderate size of basis sets. As the computational requirements for TPA-CI technique are much less than FCI, we can use it for larger molecules. Also, its application limit is the same as DFT. Thus, this approach would promote the use of CI method for real problems. However, we need further development in computation (coding) for property calculation using this approach.

---

1. Fock, V., Näherungsmethode zur Lösung des quantenmechanischen Mehrkörperproblems. *Z. Physik.*, 1930, **61**, 126–148.
2. Pauli, W., Uber den Zusammenhang des Abschlusses der Elektronengruppen im Atom mit der Komplexstruktur der Spektr. *Z. Physik.*, 1925, **31**, 756–783.
3. Boys, S. F., Electronic wave functions II. A calculation for the ground state of the beryllium atom. *Proc. R. Soc. London, Ser. A*, 1950, **201**, 125–137.
4. Pople, J. A. and Binkley, J. S., Theoretical models incorporating electron correlation. *Int. J. Quant. Chem. Symp*., 1976, **10**, 1–19.
5. Cizek, J., On the correlation problem in atomic and molecular systems. Calculation of wave function components in Ursell-type expansion using quantum-field theoretical methods. *J. Chem. Phys*., 1966, **45**, 4256–4266.
6. Cizek, J., On the use of the cluster expansion and the technique of diagrams in calculations of correlation effects in atoms and molecules. *Adv. Chem. Phys*., 1969, **14**, 35–89.
7. Purvis, D. and Bartlett, R. J., A full coupled-cluster singles and doubles model: the inclusion of disconnected triples. *J. Chem. Phys*. 1982, **76**, 1910–1918.
8. Møller, C. and Plasset, M. S., Note on an approximation treatment for many-electron systems. *Phys. Rev*., 1934, **46**, 618.
9. Leininger, M. L., Allen, W. D., Schaefer, H. F. and Sherrill, C. D., Is Möller–Plesset perturbation theory a convergent *ab initio* method? *J. Chem. Phys*., 2000, **112**, 9213–9222.
10. Bag, A., Manohar, P. U., Vaval, N. and Pal, S., First- and second-order electrical properties computed at the FSMRCCSD level for

excited states of closed-shell molecules using the constrained-variational approach. *J. Chem. Phys.*, 2009, **131**, 024102.
11. Bag, A., Manohar, P. U. and Pal, S., Analytical dipole moments and dipole polarizabilities of oxygen mono-fluoride and nitrogen dioxide: a constrained variational response to Fock-Space multi-reference coupled-cluster method. *Comput. Lett.*, 2007, **3**, 351(2–4).
12. Manohar, P. U., Shamasundar, K. R., Bag, A., Vaval, N. and Pal, S., On some aspects of Fock-space multi-reference coupled-cluster singles and doubles energies and optical properties. In *Challenges and Advances in Computational Chemistry and Physics*, 2010, vol. 11, pp. 375–393.
13. Bag, A., Bhattacharya, S. and Pal, S., Static hyper-polarizability of open shell molecules computed at the FSMRCCSD level using constrained variational approach. In *Recent Advances in Spectroscopy: Theoretical, Astrophysical and Experimental Perspectives* (ed. Chaudhury, R. K. *et al.*), 2010, p. 99–109.
14. Bag, A., *Linear Response Approach in FSMRCC for Molecular Property*, LAMBERT Academic Publishing, Deutschland, Germany, 2016.
15. GAMESS, General atomic and molecular electronic structure system. *J. Comput. Chem.*, 1993, **14**, 1347–1363.
16. Kolos, W. and Wolniewicz, L., Improved theoretical ground-state energy of the hydrogen molecule. *J. Chem. Phys.*, 1968, **49**, 404–410.

ACKNOWLEDGEMENTS. I thank IISER Kolkata for financial support and Dr Pradip Ghorai for permission to publish this work independently. I also thank the reviewer for his valuable comments to improve this article.

Received 10 November 2015; revised accepted 22 July 2017

doi: 10.18520/cs/v113/i12/2325-2328

Satellite-based mapping and monitoring of heavy snowfall in North Western Himalaya and its hydrologic consequences

Bhaskar R. Nikam*, Vaibhav Garg, Prasun K. Gupta, Praveen K. Thakur, A. Senthil Kumar, Arpit Chouksey, S. P. Aggarwal, Pankaj Dhote and Saurabh Purohit

Indian Institute of Remote Sensing, 4, Kalidas Road,
Dehradun 248 001, India

Snow cover is one of the most important land surface parameters in global water and energy cycle. Large area of North West Himalaya (NWH) receives precipitation mostly in the form of snow. The major share of discharge in rivers of NWH comes from snow and glacier melt. The hydrological models, used to quantify this runoff contribution, use snow-covered area (SCA) along with hydro-meteorological data as essential inputs. In this context, information about SCA is essential for water resource management in NWH region. Regular mapping and monitoring of snow cover by traditional means is difficult due to scarce snow gauges and inaccessible terrain. Remote sensing has proven its capability of mapping and monitoring snow cover and glacier extents in these area, with high spatial and temporal resolution. In this study, 8-day snow cover products from MODIS, and 15-daily snow cover fraction product from AWiFS were used to generate long-term SCA maps (2000–2017) for entire NWH region. Further, the long term variability of 8-daily SCA and its current status has been analysed. The SCA mapped has been validated using AWiFS derived SCA. The analysis of current status (2016–17) of SCA has indicated that the maximum extent of snow cover in NWH region in last 17 years. In 2nd week of February 2017, around 67% of NWH region was snow covered. The comparison of SCA during the 1st week of March and April in 2016–17 against 2015–16 indicates 7.3% and 6.5%, increased SCA in current year. The difference in SCA during 1st week of March 2017 and 1st week of April 2017 was observed to be 14%, which indicates that the 14% SCA has contributed to the snow melt during this period. The change in snow water equivalent retrieved using SCATSAT-1 data also validates this change in snow volume.

Keywords: AWiFS, MOD10A2, North Western Himalaya, snow cover area, SCATSAT-1.

SPATIO-temporal snow cover of North West Himalayas (NWH) river basins has large variations due to difference in topography and hydro-meteorological characteristics of each basin. It is to be noted that heavy snowfall was recorded in many parts of NWH region including the Kashmir Valley during the winter season of 2016–17. According to reports in various national and local agencies and media, unusual variations in temperature and occurrence of high snowfall in early spring season in many parts of NWH broke the last 30-year record[1]. The heavy snowfall and persistent positive temperature in early spring season may cause rapid melting of snow and associated hydro-meteorological hazards in lower parts of the river valleys. The recent flood-like situation in the Kashmir valley and occurrence of avalanches in the upper reaches in Kashmir have proven the hazards associated with heavy snowfall. Therefore, the present study was made to map and monitor, current and historical snow cover area (SCA) in NWH region.

Regular mapping and monitoring of seasonal snow cover by traditional ground-based methods is difficult due to scanty snow gauges, inaccessible terrain during

---

*For correspondence. (e-mail: bhaskarnikam@iirs.gov.in)