ARTICLE

Comparison of First Principles Quantum Chemical Calculations Method for Methane Molecule

Hafiz Aji Aziz[1], Hendrawan[1], Fitri Khoerunnisa[1]*
[1]Chemistry Study Program, Unversitas Pendidikan Indonesia
Koresponden: E-mail: ha.aziz@upi.edu

A B S T R A C T

First principles quantum chemical calculation methods are groups of electronic structure calculations that are entirely derivable Schrodinger equation. Each method is defined by the approximations used and the level of correction included in the calculation. In this study, we compared the performance of the methods in term of accuracy and computational complexity. We computed the structure and vibrational spectra of methane using Hartree-Fock, Configuration Interaction, Moller-Plesset Perturbation Theory, and Coupled Clusters. Calculation results show that Coupled Cluster methods tend to have the best accuracy, but the slowest calculation, with Hartree-Fock showing an acceptably similar results provided a sufficiently large basis sets.

Kata Kunci: First principle, Hartree-Fock, Configuration Interaction, Moller-Plesset Perturbation Theory, Coupled Cluster

A B S T R A K

Metode perhitungan ab initio (first principles) merupakan metode perhitungan struktur elektronik yang dapat diturunkan langsung dari persamaan Schrodinger. Tiap metode dibedakan oleh pendekatan dan koreksi yang digunakan dalam perhitungannya. Dalam penelitian kami membandingkan performa masing-masing metode dalam wujud waktu dan akurasi perhitungan. Perhitungan dilakukan terhadap molekul metana dengan metode Hartree-Fock, Configuration Interaction, Moller-Plesset Perturbation Theory, and Coupled Clusters. Hasil perhitungan menunjukkan bahwa metode Coupled Cluster merupakan metode paling akurat, namun cenderung memerlukan waktu perhitungan yang lama, dengan metode HF memberikan hasil yang sebanding namun dengan basis set yang cukup besar.

Keywords: ab initio, Hartree-Fock, Configuration Interaction, Moller-Plesset Perturbation Theory, Coupled Cluster

INTRODUCTION

Computational chemistry is a branch of physical chemistry that employ efficient algorithm in a computer to calculate and/or predict the behaviour of a chemical system based on physical model of the system. Due to this, computational chemistry is very dependent on the physical model that was used for the system. There are basically 2 model, which are the molecular mechanics methods and the electronic structure methods. The difference between these models is whether there is a substructure in atoms or not. Molecular mechanics model in general treat atoms as solid sphere with perfectly harmonic spring as bonds. Electronic structure methods on the other hand are based on the quantum mechanical model of the atoms, and electrons are modelled using Schrodinger equation. Electronic structure methods are further divided into semiempirical methods, density functional theory-based methods, and first principles (ab initio) methods [1].

First principles methods are considered to be the 'purest' form of electronic structure methods due to their lack of need for experimental parameters as well as their use of only wavefunction or density matrix formulation in the calculations. The simplest method that was derived directly from the Schrodinger equation is the Hartree-Fock (HF) method. Hartree-Fock method was derived via a series of approximations including Born-Oppenheimer approximation, mean field approximation, and single determinant approximation. These approximations, while it reduces the computational complexity, it also introduces error in to the calculation results. The error was then denoted as electronic

correlation due to the fact that mean field approximation is one of the approximations that was used [1-2].

To improve on these inaccuracies in the calculation, post-HF methods was introduced. Post-HF methods are first principles methods that was also derivable from Schrodinger equation, but did not include some of the approximations on the HF, or alternatively reduced the impact of the approximations by introducing new terms into HF calculation. One of the methods was Moller-Plesset Perturbation Theory (MPn) that introduced perturbative term into the Hamiltonian to correct for the mean field approximation. Other methods are Configuration Interactions (CI) and Coupled Clusters (CC) that include excited state Slater determinant to correct for single determinant approximations. These corrections, while it can improve on the calculation results, it can also significantly increase the computational complexity of the calculation, making it less desirable than HF [2-3].

In this study, we will perform first principles calculations on methane molecule and compared the performance of the different first principles methods both in accuracy and calculation time. Methane was chosen due to its size as well as its available experimental data for comparison with the computational results. The simple structure of methane also removed the possibility of the calculation so be trapped in local minima.

METHODS
COMPUTATIONAL DETAILS
The computational methods that were used in this study are Hartree-Fock, MP2, CISD, CCSD, and CCSD(T) [4-5]. The basis sets that were used in the calculation are STO-3G, 6-31G, and LANL2DZ. These basis sets were chosen due to their stark difference in sizes [6-11]. The properties extracted from the calculation results are bond length, energy level of the highest occupied molecular orbital (HOMO), and the vibrational spectra of methane.

All calculations are performed using Orca 4.2.1 on Lenovo Idea Pad S145 personal computer with specification of Intel Processor i5 (8 core, @ 1.60 GHz), 8 GB RAM, and 512 GB SSD Internal Storage [12-13]. Visualization of the orbitals and molecule, are performed using Avogadro 1.2.0 [14].

RESULT AND DISCUSSION
First principles computational chemistry methods are a type of calculations that was derived directly from Schrodinger equation. The accuracy and complexity of the calculation is directly connected to the level of the theory – which in this study are HF, MP2, CISD, CCSD, CCSD(T) – as well as the basis sets – which in this study are STO-3G, 6-31G, and LANL2DZ.

By performing the calculation with combination of all methods and basis sets, complexity of the calculation methods can be gauged and compared with one another. Calculation time needed for molecule of methane with the same starting coordinate was shown in Table 1.

Table 1. Calculation time in seconds

Methods	Basis Sets		
	STO-3G	6-31G	LANL2DZ
HF	7.021	14.068	7.092
MP2	243.972	264.247	242.199
CISD	4897.081	4984.818	5516.029
CCSD	4450.524	5813.557	5364.565
CCSD(T)	7161.692	7299.692	7515.305

It can be seen from Table 1 that in general the order of the complexity is HF, MP2, CISD, CCSD, and CCSD(T), which is consistent with the level of theory. HF is the fastest due to lack of electron correlation. MP2 is slower than HF, but faster than other methods because MP2 is the first level correction in the perturbation series, making the calculation relatively simpler. CISD and CCSD are second order correction in their respective class, however CI methods in general are faster than CC due to the use of cluster operator in the CC calculations. CCSD(T) is the slowest due to high level of correction included in the calculations.

It can also be seen that calculation time in general are increasing as the basis set increase from minimum basis set (STO-3G), to split valence basis sets (6-31G), to multiple zeta basis sets (LANL2DZ). However, this tendency did not hold true for all calculations. The exceptions – such as HF/6-31G – calculation is due the fact that calculation time not only depend on the but also on the initial structure. Within a given methods, exceptions happened when the optimized structure is slightly further away from the initially provided structure as shown in HF, MP2, and CCSD.

Structure
Methane is a very simple molecule and very symmetric with symmetry group of T_d. The bond length of C-H in methane is given in Table 2.

Table 2. C-H bond length in Angstrom

Methods	Basis Sets		
	STO-3G	6-31G	LANL2DZ
HF	1.083	1.082	1.083
MP2	1.099	1.095	1.101
CISD	1.106	1.098	1.103
CCSD	1.108	1.100	1.106
CCSD(T)	1.108	1.101	1.107

Due to its symmetry, the molecular structure of methane is entirely characterized by its C-H bond length. Experimental measurement of the bond length had been performed and yielding various value including 1.085 Å [15], 1.086 Å [16], 1.087 Å [17], and 1.093 Å [18] with 1.087 Å usually generally accepted as the most accurate value of the bond length. It can be seen that all calculations actually produced bond length

value that is within less 2% from than 1.087 Å. HF tend to underestimate the bond length while other methods tend to overestimate the bond length. This difference is probably due to the electronic correlation that is absence in HF formalism. The effect of basis sets sizes however, still unpredictable as tend to be the best in post-HF method, but it is the worst in HF, indicating that this is probably related to electronic correlations.

Geometry optimization calculation that produced the molecular structure of the molecule also produced the electronic structure. This electronic structure can be represented in multitude of ways, however in this study will focus on the molecular orbital and HOMO energy level. The shapes of the molecular orbitals up to the second energy level that was unoccupied is provided in Figure 1.
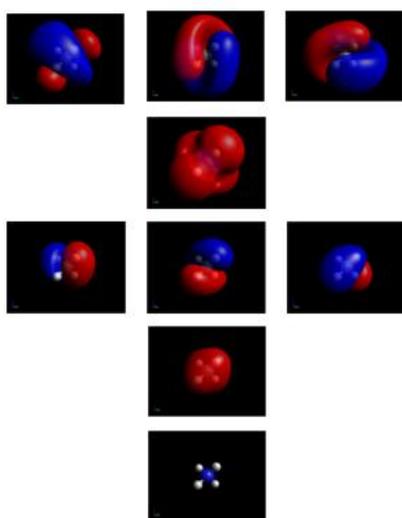
Figure 1. Shapes of molecular orbitals of methane

Shapes of molecular orbitals up to second energy levels that was unoccupied from all calculations are similar with difference in sizes accommodating the bond length of C-H. This is due to the fact that the shapes of the orbitals are actually derivable from Symmetry-Adapted Linear Combinations (SALC) approximations. The energy levels however, would be different, as shown in Table 3.

Table 3. Ionization potential of methane in eV

Methods	Basis Sets		
	STO-3G	6-31G	LANL2DZ
HF	14.16	14.83	14.81
MP2	13.99	14.74	14.69
CISD	13.92	14.72	14.68
CCSD	13.90	14.71	14.66
CCSD(T)	13.90	14.70	14.66

While the true value of orbitals energy levels is very difficult to measure experimentally, it is possible to estimate it by employing Koopman's theorem [1-3] which stated that

HOMO energy levels of RHF calculation at basis sets limit will be equal to the vertical ionization potentials. There are 2 reported measurement of methane ionization potential, which produced value of 13.7 eV [19] and 12.6 eV [20]. It can be seen that the calculation results are quite similar to the experimental results from reference [19] with difference of 1.3 eV or less. It can also be seen that the best performance is from CCSD(T) with 6.01% average error, while HF has the worst performance – albeit not by far – with average error of 7.37%, with other methods in between in order of complexity.

Vibrational Spectra

Vibrational spectra are arguably the simplest of the measurement for methane. Methane produced 9 vibrational modes, 8 of which are degenerate with degeneracy level of 2, 3, and 3, those producing 4 vibrational peaks. 2 of the vibrational peaks (A_1 and E) is not IR active due to non-polarity of the molecule during vibration. Comparison between the measured value and the calculated value is given in Table 4.

Table 4. Vibrational frequencies of methane in cm^{-1}

Methods		Vibrational Frequency			
		A_1	E	T_2	T_2
Experimental [21]		2917	1534	3019	1306
HF	STO-3G	3526	1904	3786	1675
	6-31G	3182	1709	3296	1517
	LANL2DZ	3177	1697	3307	1506
MP2	STO-3G	3342	1809	3600	1577
	6-31G	3063	1620	3193	1441
	LANL2DZ	3015	1580	3167	1413
CISD	STO-3G	3282	1768	3506	1549
	6-31G	3037	1604	3151	1432
	LANL2DZ	2998	1567	3125	1405
CCSD	STO-3G	3247	1756	3479	1538
	6-31G	3009	1592	3125	1420
	LANL2DZ	2964	1553	3097	1392
CCSD(T)	STO-3G	3244	1756	3477	1537
	6-31G	2996	1589	3115	1416
	LANL2DZ	2954	1550	3088	1389

From Table 4, it can be seen that all the computational methods overestimate the frequencies of the vibration to some degree. This behavior is actually expected from the formulation of the methods. All the methods however reproduced the order of the frequencies but with shifted value, multiplied value, or combination of both. To measure this correlation, we can actually find the linear correlation between the calculation results and experimental results, as well as measure the average percentage error for each calculation methods. The resulting data from that is given in Table 5.

Table 5. Statistical parameters of the linear regression

Methods		Parameters			
		Error	Slope	Intercept	R^2
HF	STO-3G	24.66	1.2064	75.9296	0.9970
	6-31G	11.44	1.0484	125.6512	0.9996
	LANL2DZ	11.09	1.0580	100.4860	0.9994
MP2	STO-3G	18.13	1.1504	58.1482	0.9968
	6-31G	6.67	1.0299	69.6268	0.9993
	LANL2DZ	4.86	1.0281	38.1273	0.9987
CISD	STO-3G	15.63	1.1215	65.7706	0.9978
	6-31G	5.67	1.0155	77.9922	0.9993
	LANL2DZ	4.01	1.0148	47.4110	0.9989
CCSD	STO-3G	14.69	1.1093	71.1642	0.9974
	6-31G	4.80	1.0058	79.7713	0.9993
	LANL2DZ	3.01	1.0036	49.7184	0.9989
CCSD(T)	STO-3G	14.64	1.1082	72.1666	0.9974
	6-31G	4.46	1.0010	82.6189	0.9993
	LANL2DZ	2.74	0.9998	51.7798	0.9988

From definitions of the linear regression, it can be determined that the ideal results should be error of 0, slope of 1, intercept of 0, and R^2 of 1. While all calculations produced R^2 that is close to 2, other measure are not, with the best performance from CCSD(T). These parameters also shown that accuracy tends to increase as level of theory and the basis sets increase.

However, the increase in accuracy for post-HF/LANL2DZ calculations from one level to another is not very significant indicating that it actually has reach near its limits. This should be taken as discouraging high level of calculation, but rather a consideration for further uses. While some higher level of theory would give a very accurate results, increase the theory level and basis set size after a certain point would only give a diminishing return.

CONCLUSIONS

We show how HF, MP2, CISD, CCSD, and CCSD(T) compared to each other in term of performance in accuracy and computational time. In term of accuracy, CCSD(T) in general are the most accurate methods with accuracy of CCSD(T)/LANL2DZ for prediction of vibrational spectra less than 3%, while HF tends to produce the least accurate estimate. However, the computational time for HF is almost 3 orders of magnitude lower than CCSD(T), with calculation results that not very different from each other. We thus conclude that, unless the calculation is for high level accuracy prediction of vibrational spectra – which would need CCSD(T)/LANL2DZ, HF/LANL2DZ can provide an acceptable result, with much less computational time.

AUTHORS CONTRIBUTION

HAA, HH, and FK all involved in the conceptualization of the paper. HAA performed the calculation and writing.

REFERENCES

[1] F. Jensen, Introduction to Computational Chemistry, West Sussex: John Wiley and Sons, 2007.

[2] C.J. Cramer, Essentials of Computational Chemistry, West Sussex: John Wiley and Sons, 2004.

[3] D.C. Young, Computational Chemistry, New York: John Wiley and Sons, 2010.

[4] V. Staemmler, "Introduction to Hartree-Fock and CI Methods", John von Neumann Institute for Computing, vol. 31, 1-18, 2006.

[5] C. Hattig, "Beyond Hartree Fock: MP2 and Coupled Clusters Methods for Large Systems", John von Neumann Institute for Computing, vol. 31, 245-278, 2006.

[6] T. Muller, "Basis Sets, Accuracy, and Calibration in Quantum Chemistry", John von Neumann Institute for Computing, vol. 31, 19-43, 2006.

[7] W.J. Hehre, R.F. Stewart, and J.A. Pople, "Small Gaussian Expansions of Slater-Type Orbitals", The Journal of Chemical Physics, vol. 52, no. 1, 431–438, 1970.

[8] R. Ditchfield, W.J Hehre, and J.A. Pople, "Self-consistent molecular-orbital methods. IX. An extended Gaussian-type basis for molecular-orbital studies of organic molecules

[9] P.J. Hay and W.R. Wadt," Ab initio effective core potentials for molecular calculations. Potentials for the transition metal atoms Sc to Hg", The Journal of Chemical Physics, vol. 82, no. 1, 270-283, 1985.

[10] P.J. Hay and W.R. Wadt, "Ab initio effective core potentials for molecular calculations. Potentials for main group elements Na to Bi", The Journal of Chemical Physics, vol. 82, no. 1, 284-298, 1985.

[11] P.J. Hay and W.R. Wadt, "Ab initio effective core potentials for molecular calculations. Potentials for K to Au including the outermost core orbitals", The Journal of Chemical Physics, vol. 82, no. 1, 299-310, 1985.

[12] F. Neese, F. Wennmohs, U. Becker, and C. Riplinger, "The ORCA quantum chemistry program package", The Journal of Chemical Physics, vol. 152, pp. 224108, 2020.

[13] F. Neese, "Software update: the ORCA program system, version 4.0", Wiley Periodicals, vol. 8, 2018.

[14] M.D. Hanwell, D.E. Curtis, D.C. Lonie, T. Vandermeersch, E. Zurek, and G.R. Hutchison, "Avogadro: an advanced semantic chemical editor, visualization, and analysis platform", Journal of Cheminformatics, vol. 4, no. 17, 2012.

[15]   K. Kuchitsu and L.S Bartell, "Effect of anharmonic vibrations on the bond lengths of polyatomic molecules. II. Cubic constants and equilibrium bond lengths of methane", The Journal of Chemical Physics, vol. 36, no. 9, 2470-2481, 1962.

[16]   C.C. Costain and B.P. Stoicheff, "Microwave spectrum, molecular structure of vinyl cyanide and a summary of CC, CH bond lengths in simple molecules", The Journal of Chemical Physics, vol. 30, no. 3, 777-782, 1959.

[17]   E. Hirota, "Anharmonic potential function and equilibrium structure of methane", Journal of Molecular Spectroscopy, vol. 77, 213-221, 1979

[18]   D.M. Dennison, "The Infra-Red Spectra of Polyatomic Molecules. Part II", Reviews of Modern Physics, vol. 12, no. 3, 175-214, 1940.