NCC 2-371
IN-72-CR
158955
P-14

On the bond distance in methane

Philippa Bowen-Jenkins, Lars G.M. Pettersson
and Per Siegbahn
Institute of Theoretical Physics
University of Stockholm, Vanadisvägen 9
S-11346 Stockholm, Sweden

Jan Almlöf
Department of Chemistry and Supercomputer Institute
University of Minnesota
Minneapolis, Minnesota 55455

and

Peter R. Taylor
ELORET Institute†
Sunnyvale, California 94087

Abstract

The equilibrium bond distance (r_e) in methane has been optimized using coupled-pair functional and contracted CI wave functions, and a Gaussian basis that includes $g-$type functions on carbon and $d-$type functions on hydrogen. The resulting bond distance, when corrected for core-valence correlation effects, agrees with the experimental value of 2.052 a_0 to within the experimental uncertainty of 0.002 a_0. The main source of error in the best previous studies, which showed discrepancies with experiment of 0.007 a_0, is shown to be basis set incompleteness. In particular, it is important that the basis set be close to saturation, at least for the lower angular quantum numbers.

---

† Mailing address: NASA Ames Research Center, Moffett Field, CA 94035

I. Introduction

Methane is the simplest saturated hydrocarbon and contains the prototype carbon-hydrogen bond. It is therefore not surprising that numerous attempts have been made to compute the bond distance in the ground state as accurately as *ab initio* quantum chemical methods allow. Given that the CH_4 ground state is well described at the Hartree-Fock level, it would be expected that a single reference configuration correlation treatment would be adequate for this purpose. However, the best calculations performed to date have shown a discrepancy with experiment of 0.01 a_0 or more: considerably larger than would have been expected. Several suggestions as to the source of this discrepancy have been made, and it is convenient to discuss these as part of a review of previous calculations and of the analysis of experiment. We restrict this review to calculations which attempt to account for a substantial fraction of the correlation energy.

The first indication of a possible disagreement between experiment and theory arose from the work of Meyer. In his classic 1973 paper on the PNO-CI method [1], Meyer reported an r_e value of 2.062 a_0, 0.01 a_0 larger than the then experimental estimate [2]. As Meyer had used an accurate treatment of electron correlation, the coupled electron-pair approximation (CEPA), and a very large basis, including two d sets and an f set on carbon, this discrepancy was disappointingly large. For example, the same approach (that is, the same $n-$particle and a similar $1-$particle space treatment) gave much smaller differences between theory and experiment for the first-row diatomic hydrides and water [3]. Meyer therefore suggested [1] that the larger discrepancy in CH_4 might result from the use of an incorrect estimate of anharmonic effects in deducing r_e from the experimental r_0 in the early work of Kuchitsu and Bartell [2]. This suggestion seemed very plausible given that the cubic anharmonicity for the CH stretch from Ref. 2 was some four times larger than that in the molecule CH. However, while later studies by Pulay *et al.* [4] indicated that the cubic anharmonicity itself had indeed been overestimated in Ref. 2, this had little effect on the computed bond length.

Pulay *et al.* [4] used *ab initio* force field data (including cubic anharmonic effects) and electron diffraction and spectroscopic information to infer an r_e value of 2.051 a_0. This value also agreed well with a newer experimental estimate from Gray and Robiette [5] and a newer analysis by Bartell and Kuchitsu [6]. Pulay *et al.* suggested that the most likely explanation for the discrepancy between theory and experiment was some inadequacy in the theoretical calculations, possibly due to basis set deficiencies or to incorrect accounting for the effects of higher excitations.

In view of the findings in Refs. 1 and 4 it was somewhat surprising that in 1980 Yamaguchi and Schaefer [7] obtained an r_e value in perfect agreement with experiment using a single reference single and double excitation CI (SDCI) method. Yamaguchi and Schaefer attributed their better agreement with experiment than Meyer [1] had obtained to a better basis, but this was apparently based on a misunderstanding of the work presented in Ref. 1. In fact, the excellent agreement between the results of Ref. 5 and experiment arises from a cancellation of errors between the truncation of the configuration space to SDCI and the limited basis set used. The basis set limit SDCI r_e value is discussed explicitly below.

More recently, Siegbahn [8] explored the effect on r_e of expanding both the $n-$particle and $1-$particle spaces. Four different extended basis sets were used in conjunction with complete active space SCF (CASSCF) [9] and multireference contracted CI (CCI) [10] wave functions. The effects of higher excitations were estimated using a multireference analog of Davidson's correction [11,12], and in some calculations core correlation was explicitly included. The largest basis set used gave an r_e value in essentially perfect agreement with that of Meyer [1]. There are no obvious deficiencies in the CCI calculations: all important reference states were included in the CI expansion (several different choices of reference states were investigated) and a large basis set was used. In fact, the only real improvement of these calculations over those of Meyer was the estimate of a reduction of 0.003 a_0 in r_e as a result of core-valence correlation effects. This correction for core-valence correlation is also used in the present work (see below).

Further investigation of the effect of expanding the $1-$particle space has been performed very recently by Handy *et al.* [13]. These authors have investigated convergence of the MP2 r_e value to the basis set limit, and demonstrated that for H p sets with exponents larger than 1.0 are required. By this criterion there is a problem with Meyer's basis [1], as its largest H exponent is only 0.75, although the largest basis sets used by Siegbahn [5] are acceptable. However, while this is an important observation on basis set problems for r_e in CH_4, it seems very unlikely that an MP2 treatment can, in itself, eliminate the discrepancy between theory and experiment. For example, although all electrons were correlated in these calculations, no basis functions suitable for core correlation [8] were included. It is unlikely that such calculations recover much of the core-valence correlation effect on the bond length, and any additional contraction of the bond as a result of correcting for these effects would worsen the agreement with experiment, possibly giving a final error close to 0.01 a_0.

Two possible sources of error in the calculations reviewed here suggest themselves. The first is a deficiency in the basis sets. Although the sets used in Refs. 1, 8 and 13 are large, they are probably not completely saturated in the polarization spaces used, and they contain no higher than f-type functions on C. Recent work using atomic natural orbital (ANO) basis sets [14] has indicated the importance of saturated polarization spaces. A second problem is the multireference Davidson correction for unlinked clusters used in Ref. 8. During recent years there has been increased evidence that this correction is unreliable for geometries (see, for example, Ref. 15). In view of these possibilities, we have undertaken a redetermination of the methane r_e value using ANO basis sets and employing both the CCI and the (size-consistent) coupled-pair functional (CPF) [16] methods. The following section contains details of the computational methods, our results are presented in section III and conclusions in section IV.

II. Methods

The carbon atomic basis is derived from the ($14s$ $9p$) primitive set of Huzinaga *et al.* [17], which has the same exponents for the p primitives and the outermost nine s primitives. This basis was augmented with d, f and g polarization sets: the d exponents were taken from the third through eighth highest p exponents and the f exponents were taken from the fifth through seventh highest p exponents; the g exponents were 1.3661 and 0.5464. This ($14s$ $9p$ $6d$ $3f$ $2g$) primitive set was contracted using ANOs obtained from an SDCI calculation on the ^5S state of carbon atom. The final contracted set used was [$4{+}1s$ $3{+}1p$ $3d$ $2f$ $1g$], that is, the s and p spaces comprise respectively 4 and 3 ANOs plus the most diffuse primitive uncontracted. The hydrogen basis is derived from van Duijneveldt's ($8s$) primitive set [18], augmented with six p sets (with exponents forming an even-tempered sequence with geometric mean 1.0 and an internal ratio of 2.5) and three d sets (exponents 6.25, 1.75, 0.40), contracted to [$4s$ $3p$ $1d$] based on NOs for H_2, as described in Ref. 14. Previous studies have shown that such sets are saturated in the s, p and d spaces and almost saturated in the f and g spaces. To investigate the contribution of the higher angular functions, a smaller basis was constructed by omitting the g ANO from C and the d ANO from H, and using 5 s and 3 p ANOs on C, giving a [$5s$ $3p$ $2d$ $1f/4s$ $3p$ $1d$] basis. Subsets of these basis sets were used to investigate aspects of basis set superposition error and basis set saturation. The largest basis set used comprised 146 contracted Cartesian Gaussians, or 127 contracted functions after the $3s$, $4p$ and $5s$, $5d$ components of the d, f and g sets have been eliminated.

Several types of wave function were generated using these basis sets: SCF, full valence CASSCF, CPF and multireference CCI. SDCI calculations were also performed in the smaller basis. The reference space for the CCI calculations consisted of all occupations giving rise to at least one configuration with a coefficient greater than 0.05 in magnitude in the CASSCF calculation. Only the eight valence electrons were correlated in all calculations.

The calculations using the largest basis sets were performed on the FPS-164 in the Institute of Theoretical Physics in Stockholm. Integrals were generated using the MOLECULE program [19], and a vectorized integral-driven code [20] was used for the CPF calculations. Some of the smaller calculations and superposition investigations were carried out on the CRAY X–MP/48 at NASA Ames, using the MOLECULE-SWEDEN program system [19,21] and the vectorized CPF code [20].

III. Results and Discussion

The computed r_e results are reported in Table 1, together with the results of other calculations. The values from the present work are determined from a parabolic fit in $1/r$ to three distances: 2.00, 2.05 and 2.10 a_0. Perhaps the most interesting result is that the CPF r_e value, when corrected by the core-valence contribution determined in Ref. 8, agrees with experiment to within 0.001 a_0. The CCI+Q result is some 0.003 a_0 longer, but this is still in better agreement with experiment than any other large basis calculation. A comparison of the two basis sets used in this work shows that the effect of high angular momentum functions is rather small: the inclusion of g functions on C and d functions on H reduces the computed bond length by only 0.001 a_0.

The best r_e value from Ref. 8 was in error by 0.007 a_0 when core-valence correlation effects were included. The results of the present work can be combined with those earlier results to identify contributions to the error in the bond length. The first source of error referred to above concerns the 1−particle basis: by comparing the CCI results of the present work with those of Ref. 8 it appears that basis set improvements decrease r_e by 0.004 a_0. This comparison demonstrates the power of the ANO contraction procedure — the largest basis used in Ref. 8 contained 111 contracted functions, while the small basis in the present work contains only 98, yet gives a bond 0.002 a_0 shorter by effectively exhausting a much larger primitive basis. This primitive set is much closer to saturation in each l value than the set used in Ref. 8.

Further information on the effects of basis set saturation can be obtained from the results given in Table 2. These results were obtained using MCPF [22], rather than CPF, wave functions, but the difference in computed r_e values between the two methods is less than 0.0002 a_0. It is clear that the elimination of the f set on C and the d set on H scarcely affects the computed r_e, and even more substantial reductions in the size of the basis set have little effect on r_e, except for the lengthening observed when the second p set on H is removed. This confirms the contention of Handy and co-workers that saturation of the p space on H is an important factor in obtaining the correct bond length. Interestingly, one of the smaller basis sets used in Ref. 8 (basis C) featured p sets on H with an exponent of 2.0. This set actually gave a bond distance 0.002 a_0 shorter than the largest set used in that work. The p space on H in the ANO set used in this work contains primitives with exponents as large as 3.95 and 9.88. It is a particular advantage of the ANO contraction method that saturated function spaces such as these can be used without the size of the contracted basis set becoming unmanageable.

An additional factor to consider in the context of basis set effects is that of basis set superposition error (BSSE). This has been investigated for several of the sets used in this work. The [$5s$ $3p$ $2d$ $1f/4s$ $3p$] basis results, if corrected for BSSE using the counterpoise method, are increased by less than 0.001 a_0 in the bond length. (The full ghost basis is used in the counterpoise calculation). This BSSE is computed for the C ^3P state: if the ^5S state is used, the BSSE effect is halved. As C in CH_4 is much closer to the ^5S state we may regard 0.001 a_0 as a pessimistic upper bound to the BSSE. For the large basis the BSSE will be even smaller. For the smaller basis results given in Table 2 the BSSE is somewhat larger: for the [$4{+}1s$ $3{+}1p$ $2d$ $1f/3s$ $2p$] set, the same pessimistic upper bound would be a 0.003 a_0 increase in r_e. As this would roughly cancel with the core-valence contribution, it can be seen that these smaller sets would all display an error of $0.002 - 0.003$ a_0 in r_e.

The second source of error in earlier calculations of r_e discussed above was the question of the multireference Davidson correction and contracted CI. The difference between the r_e value predicted by CPF and CCI+Q is 0.003 a_0 for both the large and small basis. This is some four times larger than the difference between CCI and CPF in these basis sets. Previous experience suggests that at least half the difference between CPF and CCI+Q derives from the use of the multireference Davidson correction, and the remainder from the external contraction procedure. For well-behaved systems (that is, those dominated by a single reference configuration) the Davidson correction to a multireference CCI result will usually be an overestimate of the effect of higher excitations.

An additional source of error in the work of Ref. 8 was the use of a parabolic fit in r to compute r_e. If the total energies of the present work are fitted in r, instead of $1/r$ [23], the computed r_e values become uniformly longer by some 0.001 a_0.

In view of the error analyses presented above, then, it appears that we can state definitely that our best computed bond distance for CH_4 agrees with experiment to within some 0.001 a_0, and that further extension of the 1−particle or n−particle spaces will not alter this conclusion.

The bond distance in CH_4 was also optimized using the SDCI method and the [5*s* 3*p* 2*d* 1*f*/4*s* 3*p*] basis set. The resulting r_e value is too small, but is considerably improved when the single reference Davidson correction is included and is then in good agreement with the CPF result. Such agreement is common where a single configuration dominates the CI expansion, as is the case for CH_4. For such systems, the present work suggests that accurate bond lengths can be obtained with SDCI+Q, CPF or multireference CCI (without a Davidson correction) in an adequate basis. As noted above the CCI+Q result indicates that the multireference Davidson correction overestimates the effect of higher excitations on the bond length. On the other hand, the SDCI method (uncorrected for multiple excitations) gives much too short a bond: given that improving the basis set will decrease r_e (relative to the smaller basis results) by over 0.001 a_0 at the correlated level, as illustrated by the difference between the [5*s* 3*p* 2*d* 1*f*/4*s* 3*p*] and [4+1*s* 3+1*p* 3*d* 2*f* 1*g*/4*s* 3*p* 1*d*] basis results in Table 1, it is clear that the SDCI limit r_e, corrected for core-valence effects, must be smaller than 2.047 a_0. This in fact represents less than half the correlation contribution to r_e (measured as the difference between SCF and experiment). The excellent agreement between SDCI and experiment observed by Yamaguchi and Schaefer [7] therefore arises from a fortuitous cancellation of 1−particle and n−particle space errors.

As an additional measure of the quality of the present calculations, we list in Table 3 energies and harmonic frequencies ω_1 (for the symmetric stretching mode) obtained in the present work and in previous studies. The total energy obtained in the large basis at the CCI level is the lowest variational energy reported for CH_4 in which the core is not correlated, although it is only 0.0075 E_H below the best result of Ref. 8. Interestingly, the [$5s$ $3p$ $2d$ $1f/4s$ $3p$] basis of the present work yields a total energy very close to the best result of Ref. 8, yet the geometry prediction with the former is superior to the latter. The slightly lower energy obtained in Ref. 8 is probably due to the inclusion of d−type functions on the hydrogens, while these functions are less important for the geometry than saturation of the d−space on C and the p−space on H. The harmonic frequencies for the totally symmetric stretching mode follow the trend expected from the predicted bond lengths: the methods which predict too short a bond give frequencies that are somewhat too large, and vice versa. Those methods, such as CPF and CCI, that give good bond lengths also give very good frequencies. It may be noted that Meyer's PNO-CEPA value [1] is in good agreement with the current experimental estimate, whereas it differed substantially from the earlier estimate of 3143 cm^{-1} (cf Ref. 1).

IV. Conclusions

The largest calculations presented here, when corrected for core-valence correlation effects, yield an r_e value for CH_4 in essentially perfect agreement with experiment. Studies of the convergence of r_e with extension of the 1−particle space suggest that the largest ANO basis used here is effectively at the basis set limit (certainly to within less than 0.001 a_0 in r_e), and the studies of the n−particle space performed here and in Ref. 8 suggest that the predicted r_e is also converged with respect to the n−particle expansion. These results resolve the previous discrepancy between the best theoretical results and experiment, and illustrate once again the role that basis set saturation plays in achieving such agreement. ANO basis sets seem to be an efficient approach to accomplishing such basis set saturation.

Acknowledgment

PRT was supported by NASA grant NCC 2-371.

REFERENCES

1. W. Meyer, J. Chem. Phys. **58**, 1017 (1973).
2. K. Kuchitsu and L.S. Bartell, J. Chem. Phys. **36**, 2470 (1962).
3. W. Meyer and P. Rosmus, J. Chem. Phys. **63**, 2356 (1975).
4. P. Pulay, W. Meyer and J.E. Boggs, J. Chem. Phys. **68**, 5077 (1978).
5. D.L. Gray and A.G. Robiette, Mol. Phys. **37**, 1901 (1979).
6. L.S. Bartell and K. Kuchitsu, J. Chem. Phys. **68**, 1213 (1978).
7. Y. Yamaguchi and H.F. Schaefer, J. Chem. Phys. **73**, 2310 (1980).
8. P.E.M. Siegbahn, Chem. Phys. Letters **119**, 515 (1985).
9. B.O. Roos, P.R. Taylor and P.E.M. Siegbahn, Chem. Phys. **48**, 157 (1980).
10. P.E.M. Siegbahn, Intern. J. Quantum Chem. **23**, 1869 (1983).
11. E.R. Davidson, in "The world of quantum chemistry", eds. R. Daudel and B. Pullman (Reidel, Dordrecht, 1974).
12. M.R.A. Blomberg and P.E.M. Siegbahn, J. Chem. Phys. **78**, 5682 (1982).
13. N.C. Handy, J.F. Gaw and E.D. Simandiras, J. Chem. Soc. (Far. II), in press.
14. J. Almlöf and P.R. Taylor, J. Chem. Phys. **86**, 4070 (1987).
15. L.G.M. Pettersson and P.E.M. Siegbahn, J. Chem. Phys. **83**, 3538 (1985).
16. R. Ahlrichs, P. Scharf and C. Ehrhardt, J. Chem. Phys. **82**, 890 (1984).
17. S. Huzinaga, M. Klobukowski and H. Tatewaki, Can. J. Chem. **63**, 1812 (1985).
18. F.B. van Duijneveldt, IBM Research Report RJ 945 (1971).
19. J. Almlöf, MOLECULE, a vectorized Gaussian integral program.
20. M.R.A. Blomberg and P.E.M. Siegbahn, unpublished work.
21. SWEDEN is a vectorized SCF-MCSCF-direct CI-conventional CI-CPF-MCPF program, written by P.E.M. Siegbahn, C.W. Bauschlicher, Jr., B. Roos, P.R. Taylor, A. Heiberg, J. Almlöf, S.R. Langhoff and D.P. Chong.
22. D. P. Chong and S. R. Langhoff, J. Chem. Phys. **84**, 5606 (1986).
23. G. Simons, R.G. Parr and J.M. Finlan, J. Chem. Phys. **59**, 3229 (1973).
24. R.D. Amos, J. Chem. Soc. (Far. II), in press.

Table 1. Computed methane r_e results (a_0)[a].

Method	Basis	r_e
SCF	[5s 3p 2d 1f/4s 3p]	2.045
SCF	[4+1s 3+1p 3d 2f 1g/4s 3p 1d]	2.044
SCF[b]	[8s 5p 2d 1f/4s 2p]	2.047
SCF[c]	[9s 5p 4d 1f/4s 2p 1d]	2.046
SCF[d]	[8s 6p 4d 1f/6s 3p 1d]	2.044
CASSCF	[5s 3p 2d 1f/4s 3p]	2.081
CASSCF	[4+1s 3+1p 3d 2f 1g/4s 3p 1d]	2.080
MP2[e]	[8s 6p 3d 1f/6s 3p]	2.046
SDCI	[5s 3p 2d 1f/4s 3p]	2.051
SDCI[f]	[6s 4p 2d/4s 1p]	2.046
PNO-CI[b]	[8s 5p 2d 1f/4s 2p]	2.056
SDCI+Q	[5s 3p 2d 1f/4s 3p]	2.057
CCI	[5s 3p 2d 1f/4s 3p]	2.057
CCI	[4+1s 3+1p 3d 2f 1g/4s 3p 1d]	2.056
CCI[c]	[9s 5p 4d 1f/4s 2p 1d]	2.059
+CV[c,g]		2.056
CCI+Q	[5s 3p 2d 1f/4s 3p]	2.059
CCI+Q	[4+1s 3+1p 3d 2f 1g/4s 3p 1d]	2.058
CCI+Q[c]	[9s 5p 4d 1f/4s 2p 1d]	2.062
+CV[c,g]		2.060
PNO-CEPA[b]	[8s 5p 2d 1f/4s 2p]	2.062
CPF	[5s 3p 2d 1f/4s 3p]	2.056
+CV[g]		2.054
CPF	[4+1s 3+1p 3d 2f 1g/4s 3p 1d]	2.055
+CV[g]		2.052
Expt[h]		2.052±0.002

[a] All values from this work unless otherwise noted.

[b] Ref. 1.

[c] Ref. 8.

[d] Ref. 24.

[e] Ref. 13.

[f] Ref. 7.

[g] Core-valence correlation contribution ($-0.003\ a_0$) added from Ref. 8.

[h] Ref. 5.

Table 2. Basis set dependence of methane MCPF r_e (a_0).

Basis	r_e
[4+1s 3+1p 3d 2f 1g/4s 3p 1d]	2.055[a]
[5s 3p 2d 1f/4s 3p]	2.056[a]
[4+1s 3+1p 2d 1f/3s 2p 1d]	2.057
[4+1s 3+1p 2d/3s 2p 1d]	2.056
[4+1s 3+1p 2d/3s 2p]	2.056
[4+1s 3+1p 1d/3s 2p]	2.056
[4+1s 3+1p 1d/3s 1p]	2.065

[a] CPF result (see text)

Table 3. Total energies[a] (E_H) and harmonic frequencies ω_1 (cm^{-1}).

Method	Basis	Energy	ω_1
SCF	[5s 3p 2d 1f/4s 3p]	−40.215759	3143
SCF	[4+1s 3+1p 3d 2f 1g/4s 3p 1d]	−40.216614	3146
SCF[b]	[8s 5p 2d 1f/4s 2p]	−40.212896	3149
SCF[c]	[9s 5p 4d 1f/4s 2p 1d]	−40.21556	3146
SCF[d]	[8s 6p 4d 1f/6s 3p 1d]		3150
MP2[e]	[8s 6p 3d 1f/6s 3p]		3081
SDCI	[5s 3p 2d 1f/4s 3p]	−40.423058	3070
PNO-CI[b]	[8s 5p 2d 1f/4s 2p]	−40.407457	3071
SDCI+Q	[5s 3p 2d 1f/4s 3p]	−40.437692	3028
CCI	[5s 3p 2d 1f/4s 3p]	−40.429584	3026
CCI	[4+1s 3+1p 3d 2f 1g/4s 3p 1d]	−40.438970	3031
CCI[c]	[9s 5p 4d 1f/4s 2p 1d]	−40.43145	3021
CCI+Q	[5s 3p 2d 1f/4s 3p]	−40.436570	3014
CCI+Q	[4+1s 3+1p 3d 2f 1g/4s 3p 1d]	−40.446331	3018
CCI+Q[c]	[9s 5p 4d 1f/4s 2p 1d]	−40.44149	3032
PNO-CEPA[b]	[8s 5p 2d 1f/4s 2p]	−40.418021	3037
CPF	[5s 3p 2d 1f/4s 3p]	−40.434356	3036
CPF	[4+1s 3+1p 3d 2f 1g/4s 3p 1d]	−40.444403	3039
Expt[f]			3025

[a] Computed at 2.05 a_0.

[b] Ref. 1.

[c] Ref. 8.

[d] Ref. 24.

[e] Ref. 13.

[f] See Ref. 13.