

# Structure of the exact wave function. V. Iterative configuration interaction method for molecular systems within finite basis

Hiroshi Nakatsuji<sup>a)</sup> and Masahiro Ehara

*Department of Synthetic Chemistry and Biological Chemistry, Graduate School of Engineering,  
Kyoto University, Sakyo-ku, Kyoto 606-8501, Japan*

(Received 28 December 2001; accepted 30 April 2002)

The iterative configuration interaction (ICI) method is applied to molecular systems within finite basis using only few (1–3) variables and shown to give the exact results that are identical to the full CI (FCI) ones. Since each iteration step of ICI is variational, the ICI converges monotonically to the exact solution from above. The diagonalization in ICI is so slight as the number of variables is so small, in contrast to the huge number of variables of FCI. We calculated the molecular ground states of various spin-space symmetries using minimal basis and double zeta basis. The number of iterations for convergence was small for minimal basis but moderate for double zeta basis, considering that only 1–3 variables are optimized in each iteration step. © 2002 American Institute of Physics. [DOI: 10.1063/1.1487830]

## I. INTRODUCTION

Solving the Schrödinger equation (SE) is a central theme of theoretical chemistry, since it constitutes a basic principle of chemistry.<sup>1</sup> In this series of studies,<sup>2–6</sup> we have investigated the structure of the exact wave function, and proposed the iterative configuration interaction (ICI) method<sup>2,3</sup> and the extended coupled cluster (ECC) method<sup>4</sup> as the methods to calculate the exact wave function. This has been confirmed numerically in the applications to harmonic oscillator<sup>4,5</sup> and hydrogen atom,<sup>6</sup> both involving only one-particle operator. Horn and Weinstein, Kosloff and Tal-Ezer, Cioslowski, and others considered to solve the time-dependent SE on the imaginary time axis.<sup>7</sup> Huang *et al.*<sup>8</sup> proposed the surplus function method for the variational Monte Carlo calculations. Nooijen<sup>9</sup> considered to solve the generalized coupled cluster wave function by the density equation<sup>10</sup> that is equivalent with the SE. Van Voorhis and Head-Gordon<sup>11</sup> showed numerically that the generalized coupled cluster certainly reproduces the full CI energy for Ne and N<sub>2</sub>.

In a finite basis set space, the full CI (FCI) method gives the best possible approximation of the exact solution of the SE within the linear expansion treatment. The ICI and ECC methods should therefore give the results identical to the FCI, when applied using the same basis set. We show in this paper that this is certainly the case for the ICI method.

The number of the variables of the ICI method,  $N_D$ , is related to the way how we define the Hamiltonian operator:  $N_D$  is the dividing number of the Hamiltonian operator. When the Hamiltonian is given by

$$H = \sum_i -\frac{1}{2}\Delta_i - \sum_i \sum_A Z_A / r_{Ai} + \sum_{i>j} 1/r_{ij}, \quad (1.1)$$

we may divide it into the kinetic operator, nuclear attraction operator, and electron repulsion operator. Then,  $N_D$  is three.

We may further divide the nuclear attraction operator into those due to nucleus A, nucleus B, etc., and then,  $N_D$  is  $N_N + 2$  where  $N_N$  is the number of the nuclei involved in the molecule. When we use the Hamiltonian in the second quantized form, it may be divided into singles and doubles operators and then  $N_D$  is the general singles and doubles (GSD) number. If necessary, the singles part may be further divided into kinetic part and nuclear attraction part. In this paper, we use the division associated with Eq. (1.1) and the nuclear attraction part is not divided, so that  $N_D$  is three for all the molecules, independent of the quality of the basis set. We have also performed the simplest ICI (SICI), where the Hamiltonian is not divided, namely  $N_D = 1$ . The number of variables in the present ICI is, therefore, only one or three, in contrast to the huge ones of the FCI.

## II. ICI METHOD

We briefly explain the ICI method.<sup>2,3</sup> We divide the Hamiltonian into  $N_D$  parts as

$$H = \sum_{I=1}^{N_D} H_I, \quad (2.1)$$

and corresponding to this division, we define the variable operator  $S$  by

$$S = \sum_{I=1}^{N_D} C_I H_I. \quad (2.2)$$

The ICI wave function is defined by the recurrence,

$$\psi_n = (1 + S_n) \psi_{n-1}, \quad (2.3)$$

where  $n$  is the iteration number. The variable coefficients  $C_I$  in the  $i$ th iteration are written as  $C_{I,n}$  and are calculated by the variational principle using the secular equation,

$$\langle \psi_n | H - E_n | \psi_{n-1} \rangle = 0, \quad (2.4)$$

$$\langle \psi_n | (H - E_n) H_I | \psi_{n-1} \rangle = 0, \quad (I = 1, \dots, N_D). \quad (2.5)$$

<sup>a)</sup>Fax: +81-75-753-5910. Electronic mail: hiroshi@sbchem.kyoto-u.ac.jp

Note that the energy of the  $n$ th iteration,  $E_n$ , satisfies

$$\langle \psi_n | H - E_n | \psi_n \rangle = 0 \quad (2.6)$$

as well as Eq. (2.4).

It is convenient to introduce a unity operator by  $H_0$  ( $H_0=1$ ), and assign  $C_0$  to this operator instead of unity in Eq. (2.3). Then, the ICI wave function is written as

$$\psi_n = \sum_{I=0}^{N_D} C_{I,n} H_I \psi_{n-1}, \quad (2.7)$$

and the above secular equation is rewritten as

$$\sum_{J=0}^{N_D} \langle \psi_{n-1} | H_I (H - E_n) H_J | \psi_{n-1} \rangle C_{J,n} = 0, \quad (I=0, \dots, N_D). \quad (2.8)$$

Note that this is a very small secular equation of the dimension of 2 and 4 in the present calculations. All that is necessary for the ICI calculation is the evaluation of the few integrals involved in Eq. (2.8), namely

$$H_{IJ} = \langle \psi_{n-1} | H_I H_J | \psi_{n-1} \rangle \quad (2.9)$$

and

$$S_{IJ} = \langle \psi_{n-1} | H_I H_J | \psi_{n-1} \rangle. \quad (2.10)$$

The essence of the finite basis approximation adopted in FCI is that these integrals are calculated within the space of this finite basis set. We denote the full configuration space of this finite space by  $\{|k\rangle\}$ , which is complete within this finite space, namely

$$\sum_k |k\rangle \langle k| = 1. \quad (2.11)$$

Then, the integrals  $H_{IJ}$  and  $S_{IJ}$  within this finite space are evaluated by using the so-called resolution of identity (RI) procedure as

$$H_{IJ} = \sum_{k,l} \langle \psi_{n-1} | H_I | k \rangle \langle k | H | l \rangle \langle l | H_J | \psi_{n-1} \rangle \quad (2.12)$$

and

$$S_{IJ} = \sum_k \langle \psi_{n-1} | H_I | k \rangle \langle k | H_J | \psi_{n-1} \rangle. \quad (2.13)$$

An important notice here is that if we calculate the integrals  $H_{IJ}$  and  $S_{IJ}$  by analytical procedure or some equivalent method, the solution of the ICI will be the true solution of the SE. In general, Eq. (2.11) does not hold, as long as the basis  $\{|k\rangle\}$  is not truly complete. The finite space approximation adopted above is a crude approximation and its accuracy is limited by the quality of the basis set initially adopted. A dilemma of FCI is that to improve the quality, we have to use very large basis set, but it is impractical since the dimension of FCI soon reaches an astronomical figure.

The ICI calculation of FCI is very simple: it was performed by modifying the related parts of the GAMESS program.<sup>12</sup> All we have to do is an evaluation of the ten or three integrals of  $H_{IJ}$  and  $S_{IJ}$ , depending on ICI3 ( $N_D=3$ ) or SICI ( $N_D=1$ ). In the RI formula, the configuration  $|k\rangle$

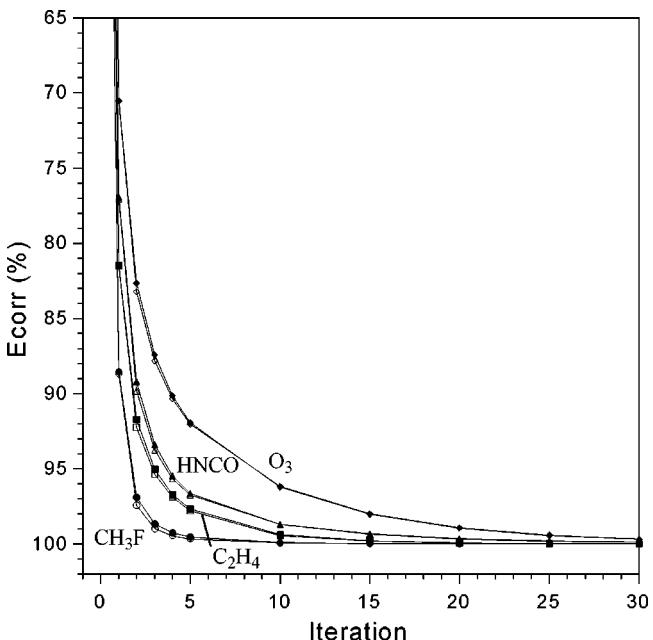


FIG. 1. Convergence of SICI and ICI3 calculations with the minimal STO-6G basis sets. Full and open symbols are SICI and ICI3 results, respectively.

runs over the full configurations. However, since the operator  $H_I$  is one or two electron operator, the integral  $\langle \psi_{n-1} | H_I | k \rangle$  is zero if  $|k\rangle$  is different from the elements of  $|\psi_{n-1}\rangle$  by more than two orbital indices. Using such a criterion, we can make the evaluation of the integrals  $H_{IJ}$  and  $S_{IJ}$  efficient.

Since each step of ICI is variational, we can calculate the ground state of each spin-space symmetry by starting from an appropriate initial guess function  $|\psi_0\rangle$ . The higher excited states of the same symmetry as the ground state can be calculated by Method A described in Paper II.<sup>3</sup> For this purpose, all the lower states of the desired excited states must exist within the eigenvalues of the small matrix of the dimension  $N_D+1$  in the iteration process, even though the ground and other lower excited states are approximate except for the excited state under consideration. In the present very small  $N_D$  case, this requirement was difficult to be satisfied even for the lowest excited state, so that we calculated only the ground state of each spin-space symmetry. The calculations of the excited states by Method B,<sup>3</sup> which is ICI-CI in the notation of Paper IV,<sup>5</sup> were also not performed by the same reason.

### III. RESULTS

We performed ICI calculations using minimal STO-6G basis<sup>13</sup> and double-zeta basis.<sup>14</sup> Minimal basis calculations were done for CO, H<sub>2</sub>O, HCN, acetylene (C<sub>2</sub>H<sub>2</sub>), ethylene (C<sub>2</sub>H<sub>4</sub>), CH<sub>3</sub>F, HCHO, CO<sub>2</sub>, O<sub>3</sub> and HNCO. Double zeta calculations were done for Be, LiH, CH<sup>+</sup>, HF, <sup>1</sup>A<sub>1</sub> and <sup>3</sup>B<sub>1</sub> states of CH<sub>2</sub>, and H<sub>2</sub>O in various spin-space symmetries (<sup>1</sup>A<sub>1</sub>, <sup>1</sup>A<sub>2</sub>, <sup>1</sup>B<sub>1</sub>, doublet ionized <sup>2</sup>A<sub>1</sub>, <sup>2</sup>B<sub>1</sub>, <sup>2</sup>B<sub>2</sub>, and triplet <sup>3</sup>A<sub>1</sub>, <sup>3</sup>A<sub>2</sub>, <sup>3</sup>B<sub>1</sub> states). In the minimal basis calculations, the 1s orbitals of the second-row atoms were kept as frozen cores.

Figure 1 shows the convergence behavior of the SICI

TABLE I. FCI, SICI, and ICI3 results for the minimal basis sets. The initial guess is Hartree–Fock.

Molecule	Active space occ.×unocc.	FCI			SICI			ICI3						
		Hartree–Fock Energy (au)	Dimension	Energy (au)	Iter. $10^{-5}$ au accuracy		Energy (au)	$10^{-3}$ au accuracy	Iter. $10^{-4}$ au accuracy	Dim.	$10^{-5}$ au accuracy	Energy (au)	$10^{-3}$ au accuracy	Iter. $10^{-4}$ au accuracy
					Dim.	accuracy								
H <sub>2</sub> O	4×2	−75.676 51	37	−75.727 91	2	10	−75.727 91	2	5	4	7	−75.727 91	2	4
HCN	5×4	−92.573 46	1436	−92.741 21	2	29	−92.741 21	8	18	4	27	−92.741 21	8	17
C <sub>2</sub> H <sub>2</sub>	5×5	−76.602 41	2640	−76.775 87	2	30	−76.775 87	9	19	4	28	−76.775 87	9	18
C <sub>2</sub> H <sub>4</sub>	6×6	−77.826 60	29248	−77.991 65	2	34	−77.991 65	11	21	4	33	−77.991 65	10	21
CH <sub>3</sub> F	7×4	−138.472 33	16542	−138.570 67	2	23	−138.570 67	4	10	4	20	−138.570 67	4	9
HCHO	6×4	−113.440 29	3644	−113.584 52	2	34	−113.584 52	9	21	4	33	−113.584 52	9	21
CO <sub>2</sub>	8×4	−186.852 49	9189	−187.065 94	2	46	−187.065 94	16	30	4	44	−187.065 94	15	29
O <sub>3</sub>	9×3	−223.415 85	4067	−223.679 98	2	67	−223.679 98	29	48	4	68	−223.679 98	29	48
HNCO	8×5	−167.105 76	216249	−167.323 97	2	68	−167.323 97	18	39	4	69	−167.323 97	19	40

and ICI3 with the minimal basis. The initial guess is Hartree–Fock. The energy decreases monotonically and converges to the FCI energy from above as theoretically predicted since each iteration process of ICI is variational. The ICI3 gives initially more rapid convergence than the SICI, because the number of variables is larger in ICI3. However, since the difference is only two, the overall behaviors of ICI3 and SICI are similar. Generally speaking, the convergence of ICI is very rapid in the initial stage; for example, at 5 iterations, 99.66% of the correlation energy is achieved for CH<sub>3</sub>F. The convergence of O<sub>3</sub> is slower initially than the other molecules, since the ground state of O<sub>3</sub> has quasi-degenerate biradical character and so the Hartree–Fock initial guess is poor. For the ICI with the double-zeta basis, similar convergence behavior was observed, though the convergence was slower than the minimal basis case.

Table I shows the SICI and ICI3 results, together with the FCI results, for nine molecules calculated with the minimal basis. The ICI converges exactly to the FCI energy without any difficulty. The dimension of the present ICI is only 2 or 4 and is remarkably small in comparison with the FCI dimension (29 248 for ethylene and 216 249 for HNCO). We show the number of iterations for obtaining the accuracy of  $10^{-3}$  au (0.6 kcal/mol),  $10^{-4}$  au (0.06 kcal/mol), and  $10^{-5}$  au (0.006 kcal/mol). For ordinary calculations in chemistry,  $10^{-3}$  au accuracy would be sufficient. Comparing the iteration times necessary for different accuracies, we see that the ICI converges quite nicely at the initial stage and then, to

improve only the order of 0.1–0.01 kcal/mol, many iterations are done. For example, for HNCO, the molecule having largest FCI dimension, the iteration time for the  $10^{-3}$  au accuracy is 19, that for  $10^{-4}$  au accuracy is 40 and that for the  $10^{-5}$  au accuracy is 69. Similar behavior is also observed for other molecules. For obtaining millihartree of accuracy, 10–20 iterations are enough, except for ozone.

Table II shows the results for the molecules calculated with the double-zeta basis. Again, we confirm that the SICI and ICI3 energies converge to the FCI energy from above. Here, we need larger number of iterations than the minimal basis case, but the convergence was very smooth and monotonic, never oscillating, for all the cases. For obtaining a chemical accuracy, i.e.,  $10^{-3}$  au accuracy, we need about 60 (average value) iterations. The iteration numbers show again that the convergence is very slow in the final stage to improve only 0.01 kcal/mol. For example, for singlet methylene, it took 260 times of iterations to improve only about  $10^{-5}$  au (0.006 kcal/mol). A reason is clearly a very small number of variables in the present calculations, which was kept to 2 to 4 irrespective of the quality of the basis set and the size of the molecules.

Table III shows the ICI3 results for water with the double zeta basis. We calculated the ground states of many different spin-space symmetries. Since every step of the present ICI is totally symmetric, we just used the initial guess function of appropriate spin-space symmetry. We used

TABLE II. FCI, SICI, and ICI3 results for the double zeta basis sets. The initial guess is Hartree–Fock except for CH<sub>2</sub>(<sup>3</sup>B<sub>1</sub>), for which it is a single dominant spin-adapted configuration.

Molecule	Active space occ.×unocc.	Hartree–Fock			FCI			SICI			ICI3						
		or single config.	Energy (au)	Dimension	Energy (au)	Dim.	Iter. $10^{-5}$ au accuracy		Energy (au)	$10^{-3}$ au accuracy	Iter. $10^{-4}$ au accuracy	Dim.	Iter. $10^{-5}$ au accuracy		Energy (au)	$10^{-3}$ au accuracy	Iter. $10^{-4}$ au accuracy
							10 <sup>-5</sup> au accuracy	10 <sup>-3</sup> au accuracy					10 <sup>-5</sup> au accuracy	10 <sup>-3</sup> au accuracy			
Be	2×2	−14.568 53	20	−14.582 69	2	24	−14.582 69	4	13	4	24	−14.582 69	3	12			
LiH	2×10	−7.981 09	620	−8.008 68	2	89	−8.008 68	11	42	4	86	−8.008 68	9	38			
CH <sup>+</sup>	3×9	−37.885 84	4582	−37.969 40	2	286	−37.969 40	89	181	4	329	−37.969 40	101	208			
HF	5×7	−100.021 97	44 388	−100.160 29	2	198	−100.160 29	59	112	4	201	−100.160 29	63	119			
CH <sub>2</sub> ( <sup>1</sup> A <sub>1</sub> )	4×10	−38.861 52	71 881	−38.962 44	2	618	−38.962 44	145	374	4	679	−38.962 44	162	419			
CH <sub>2</sub> ( <sup>3</sup> B <sub>1</sub> )	4×10	−38.874 92	111 600	−38.998 14	2	251	−38.998 14	107	178	4	288	−38.998 14	120	203			

TABLE III. FCI and ICI3 results for the singlet, triplet, and ionized states of H<sub>2</sub>O with double zeta basis set.<sup>a</sup>

State	Hartree–Fock or single config. Energy (au)	FCI			ICI3		
		Dimension	Energy (au)	Dim.	Iter. $10^{-5}$ au accuracy	Energy (au)	Iter. $10^{-3}$ au accuracy
<b>Singlet</b>							
<sup>1</sup> A <sub>1</sub>	−76.009 56	256 473	−76.157 87	4	214	−76.157 87	67
<sup>1</sup> A <sub>2</sub>	−75.567 61	245 000	−75.761 02	4	362	−75.761 02	130
<sup>1</sup> B <sub>1</sub>	−75.643 27	245 776	−75.838 26	4	334	−75.838 26	117
<b>Triplet</b>							
<sup>3</sup> A <sub>1</sub>	−75.613 83	440 475	−75.797 15	4	825	−75.797 15	169
<sup>3</sup> A <sub>2</sub>	−75.584 53	437 640	−75.779 89	4	421	−75.779 89	144
<sup>3</sup> B <sub>1</sub>	−75.672 20	437 520	−75.867 48	4	330	−75.867 48	111
<b>Ionized</b>							
<sup>2</sup> A <sub>1</sub>	−75.456 87	232 968	−75.655 68	4	319	−75.655 68	92
<sup>2</sup> B <sub>1</sub>	−75.507 36	223 282	−75.720 23	4	290	−75.720 23	85
<sup>2</sup> B <sub>2</sub>	−75.292 61	232 068	−75.460 08	4	356	−75.460 08	78

<sup>a</sup>For initial guess, see the text.

Hartree–Fock for the <sup>1</sup>A<sub>1</sub> ground state, single dominant spin-adapted configuration for the singlet and triplet excited states, and Koopmans state for the doublet ionized states. Again, the ICI of only 2 to 4 dimensions has reproduced the FCI results of the dimensions as large as 223 282–440 475. The convergence rate is different from state to state but the average iteration number for obtaining  $10^{-3}$  accuracy is 110.

#### IV. CONCLUDING REMARKS

This paper gives a numerical proof that the ICI method certainly gives the exact wave function for molecular systems within finite basis. The exact nature of the ICI is thus confirmed, together with the theoretical proof given in Papers I and II.

It is necessary to study the computational algorithm best suited to the ICI method for practical utilities in molecular calculations. As this method is *different* from the *conventional* FCI, the suitable algorithm may be different from the FCI ones. Use of the general singles and doubles (GSD) number of variables is also interesting, because it would result in better convergence and simultaneous calculations of excited states with the ICI–CI method<sup>5</sup> (Method B<sup>3</sup>). Such study is in progress in our laboratory.

#### ACKNOWLEDGMENT

This study has been supported financially by the special fund from the Ministry of Education, Science, Culture, and Sports of the Government of Japan.

- <sup>1</sup>P. A. M. Dirac, Proc. R. Soc. London, Ser. A **123**, 714 (1929).
- <sup>2</sup>H. Nakatsuji, J. Chem. Phys. **113**, 2949 (2000).
- <sup>3</sup>H. Nakatsuji and E. R. Davidson, J. Chem. Phys. **115**, 2000 (2001).
- <sup>4</sup>H. Nakatsuji, J. Chem. Phys. **115**, 2465 (2001).
- <sup>5</sup>H. Nakatsuji, J. Chem. Phys. **116**, 1811 (2002).
- <sup>6</sup>H. Nakatsuji, Phys. Rev. A **65**, 052122 (2002).
- <sup>7</sup>D. Horn and M. Weinstein, Phys. Rev. D **30**, 1256 (1984); R. Kosloff and H. Tal-Ezer, Chem. Phys. Lett. **127**, 223 (1986); T. J. Park and J. C. Light, J. Chem. Phys. **85**, 5870 (1986); J. Cioslowski, Phys. Rev. Lett. **58**, 83 (1987).
- <sup>8</sup>H. Huang, Q. Xie, Z. Cao, Z. Li, Z. Yue, and L. Ming, J. Chem. Phys. **110**, 3703 (1999); H. Huang, X. Zeng, and L. Ming, *ibid.* **112**, 5257 (2000).
- <sup>9</sup>M. Nooijen, Phys. Rev. Lett. **84**, 2108 (2000).
- <sup>10</sup>H. Nakatsuji, Phys. Rev. A **14**, 41 (1976).
- <sup>11</sup>T. Van Voorhis and M. Head-Gordon, J. Chem. Phys. **115**, 5033 (2001).
- <sup>12</sup>M. W. Schmidt, K. K. Baldridge, J. A. Boatz *et al.*, J. Comput. Chem. **14**, 1347 (1993).
- <sup>13</sup>W. J. Hehre, R. F. Stewart, and J. A. Pople, J. Chem. Phys. **51**, 2657 (1969).
- <sup>14</sup>S. Huzinaga, J. Chem. Phys. **42**, 1293 (1965); T. H. Dunning, Jr., *ibid.* **53**, 2823 (1970).