Optimization of Trial Wavefunctions for Use in Quantum Monte Carlo with Application to LiH

Hongtao Chen, B.Sc. Chem.

A thesis
submitted to the Department of Chemistry
in partial fulfilment of the requirements for the degree of
Master of Science

December 1988
Brock University
St. Catharines, Ontario

Acknowledgements:

I would like to express my gratitude to Dr. S.M. Rothstein, and Dr. J. Vrbik for their patient but strict instruction during my research, as well as their noble manner.

I would also like to thank Dr. Rothstein for his careful reading and correcting of my thesis, and Danny Legare for his proofreading of it.

Table of Contents

Lists of Tables

1. Tables from the Text

2. Tables from the Appendices

List of Figures

Abstract

Methods for both partial and full optimization of wavefunction parameters are explored, and these are applied to the LiH molecule. A partial optimization can be easily performed with little difficulty. But to perform a full optimization we must avoid a wrong minimum, and deal with linear-dependency, timestep-dependency and ensemble-dependency problems. Five basis sets are examined. The optimized wavefunction with a 3-function set gives a variational energy of **-7.998 ± 0.005** a.u., which is comparable to that $(-7.990 \pm 0.003)^1$ of Reynold's unoptimized Ψ_{II} (a double-ζ set of eight functions). The optimized wavefunction with a double-ζ plus $3d_{Z^2}$ set gives an energy of **-8.052 ± 0.003** a.u., which is comparable with the fixed-node energy $(-8.059 \pm 0.004)^1$ of the Ψ_{II}. The optimized double-ζ function itself gives an energy of **-8.049 ± 0.002** a.u. Each number above was obtained on a Bourrghs 7900 mainframe computer with 14 -15 hrs CPU time.

I. Introduction

It has been clearly shown[1-6] theoretically and practically that the employment of the so-called importance sampling technique greatly reduces the statistical error in the calculations of Quantum Monte Carlo (QMC). For this purpose a trial wavefunction (Ψ_T), which is also called a guiding function, is introduced. Although much work has been done on the construction of a suitable Ψ_T [1,7-13] and the investigation of a basis set[14-15] used for it, how to optimize the parameters of Ψ_T, and how to efficiently choose a basis set for a system still needs to be explored. These issues are related to that of a further improvement in the variational energy.

The trial wavefunction is mathematically a function of a set of adjustable parameters, and the optimization of these parameters is based on the so-called Variational Theorem[16-17] which states that, for any normalized, well-behaved trial wave function Ψ_T obeying the boundary conditions of a system, the following integral equation is true,

$$\int \psi_T^* \mathbf{H} \psi_T \, d\tau = E_{var} \geq E_0 \qquad (I-1)$$

where E_0 is the exact lowest eigenvalue of **H** and E_{var} is the variational energy. This theorem allows us to obtain an approximation to the ground-state energy of a system by calculating an upper bound without solving the Schrödinger equation. One could, using a numerical method, determine a trial wave function by optimizing its parameters to get the lowest variational energy. How small the variance of the local energy($E_L \equiv \mathbf{H}\psi_T / \psi_T$) is

depends on how close the trial function is to the exact one: a small variance usually implies a lower energy as well. Since a smaller variance indicates a more precise estimate, directly minimizing the variance rather than the variational energy is also an acceptable strategy.

Research in our group has focused on Diffusion QMC (DQMC) (A description of DQMC appears in Appendix A). In the calculation of energy expectation values in DQMC, the trial wavefunction is chosen to locate the nodes of the simulated exact wavefunction (fixed-node approximation). This affects the variance of the energy and not so much the energy itself[1], whereas the expectation values of other properties strongly depend on the form of the trial wavefunction[13,18,19]. Thus, the closer the trial function is to the exact one, the more correct the nodes will be, and the greater the accuracy of the estimates of physical properties which are obtained from it.

So far, only Wilson *et al.*[20] have given a detailed investigation into the method of optimization (using Green's function Monte Carlo, not DQMC). By minimizing a weighted variational energy for several two-electron atoms and ions, and for the Be atom, they obtained both smaller statistical errors and better estimates of the ground-state energy than those obtained from commonly used methods (configuration iteraction (CI)). The formula[20] used is shown below, minimizing

$$\sigma^2_{opt} = \frac{\sum_{i=1}^{N_{opt}} [H\psi(i)/\psi(i) - E_g]^2 w(i)}{\sum_{i=1}^{N_{opt}} w(i)} \qquad (I - 2)$$

where $w(i) = |\psi(i)/\psi_0(i)|^2$, E_g is a "guess-timate" value of the energy, and the sum is taken over all the N_{opt} configurations in the ensemble generated using ψ_0 for importance sampling. $\psi(i)$ is the trial wavefunction being optimized, which keeps changing during the optimization procedure, and ψ_0 is believed to be the best trial wave function before an optimization. The objective is to obtain the smallest variance σ^2_{opt}.

The objective of this work is to obtain, using the Variational Quantum Monte Carlo (VQMC) method, the smallest possible individual iteration variance and the lowest possible variational energy, which is comparable with the fixed-node energy obtained in DQMC (which uses a process called branching). Thus we can avoid the complications of having to do branching, and only one or a few timestep(s) may be required in a variational calculation rather than several.

VQMC is an alternative form of Diffusion Quantum Monte Carlo, and the only difference is that in VQMC the simulated Green's function has no branching term, and this function is given by

$$G(\mathbf{R} \rightarrow \mathbf{R}', \tau) = (2\pi\tau)^{-3N/2} \times \exp\{-[\mathbf{R}' - (\mathbf{R} + \tau\mathbf{F}(\mathbf{R}))]^2/2\tau\} \qquad (I - 3)$$

where **R**, **R'**, and **F(R)** represent the original and displaced three-dimensional coordinates in small timestep τ, and the quantum velocity ($\nabla\Psi_T/\Psi_T$), respectively. **F** introduces "importance sampling" into the simulation: it is larger when the trial wavefunction is small. Eq.(I - 3) is actually a multivariate normal distribution function relative to $\mathbf{R}+\tau\mathbf{F}(\mathbf{R})$, with a mean equal to 0 and a variance equal to τ. A random number drawn from this distribution is noted as $N(0,\sqrt{\tau})$.

In DQMC, the simulated Green's function is interpreted as "a transition probability for moving the set of coordinates from **R** to **R'** in time τ".[1,18] Thus the propagation of the set of coordinates (or configurations) in VQMC is determined by

$$\mathbf{R'} = \mathbf{R} + \tau\mathbf{F}(\mathbf{R}) + N(0,\sqrt{\tau}) \qquad (I-4)$$

where the step + $\tau\mathbf{F}(\mathbf{R})$ is called **Drift(D)** and + $N(0,\sqrt{\tau})$ **Jump(J)**. Eq. (I - 4) is hence called the **DJ algorithm**, which is represented as follows:

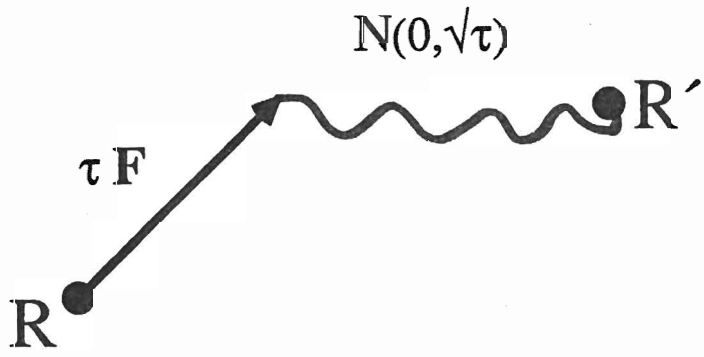


Fig. I. 1. DJ Algorithm.

If the guiding function is $\Psi_T(\mathbf{R})$, when an equilibrium state is achieved (at an "infinite" time duration for the simulation) the density $f_\infty(\mathbf{R})$ is equal to $|\Psi_T(\mathbf{R})|^2$, and the average of local energies sampled from this distribution is then equal to the variational energy.

$$\begin{aligned}\langle E_L(\mathbf{R}) \rangle &= \int E_L(\mathbf{R})\, f_\infty(\mathbf{R}) d\mathbf{R} / \int f_\infty(\mathbf{R}) d\mathbf{R} \\ &= \int (\mathbf{H}\Psi_T / \Psi_T)(|\Psi_T|^2) d\mathbf{R} / \int (|\Psi_T|^2) d\mathbf{R} \\ &= \langle \Psi_T | \mathbf{H} | \Psi_T \rangle \, / \langle \Psi_T | \Psi_T \rangle \\ &= E_{var} \end{aligned} \qquad (I-5)$$

Note that the more elaborate DQMC method has also been used in some preliminary investigations, where there is one more step, **branching** (**B**), in the propagation. The algorithm is called **DJB**, as shown below (where the point of branching is denoted by a circle)

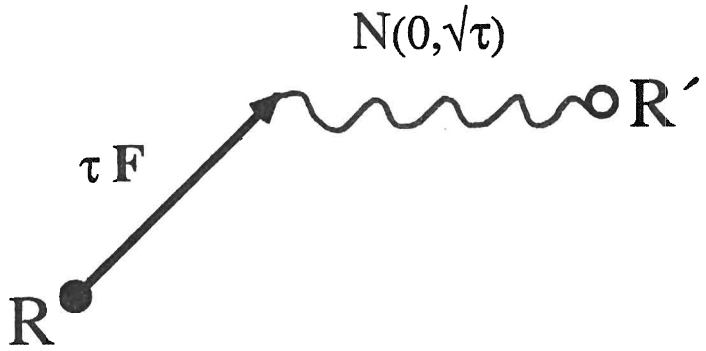


Fig. I. 2. DJB Algorithm.

When Wilson *et al* published their paper, my own work had been in

progress for one year. It was found that some conclusions basically agree with each other even though they were derived using different Monte Carlo methods.

This thesis specifically points out the success of efficiently choosing a basis set with 3 AO'S in the trial wavefunction of the LiH molecule. The variational ground-state energy and its variance obtained from this simple set is comparable with that of Reynolds', which is obtained from a double-ζ set.[1] Using an optimized double-ζ set in the trial function, the *variational* energy obtained here is comparable with the unoptimized *fixed-node* result .

It is obvious that the smaller the basis set, the smaller the CPU time required. Thus, optimizing a simple basis set to obtain a comparably low variational energy and variance is a matter of considerable interest. The objective of this work is to do a general exploration of both partial and full optimization of the parameters of a trial wavefunction and the application of this to the LiH molecule. Examinations are performed on several basis sets.

In this work the following problems were uncovered: a wrong minimum caused by minimizing the variance, a linear dependence of some optimized parameters, an ensemble dependence and a time-step dependence of the results. After the solution of these problems an optimization method is developed, with which five basis sets were examined by optimization and the best results are obtained.

II. Theory

A. Trial Wavefunction and Basis Sets Used

1. Trial Wavefunction

The trial wavefunction $\Psi_T(\mathbf{R})$ used here is a form frequently adopted,[1,14,15,21,22]

$$\Psi_T(\mathbf{R}) = \text{Det}\,|\,D_{ij}^{\alpha}\,|\,\text{Det}\,|\,D_{ij}^{\beta}\,|\,e^{\sum_{i>j} u_{ij}} \qquad (\text{II.A - 1})$$

where the full Slater determinant has reduced to a product of a spin-up (α) determinant $\text{Det}\,|\,D_{ij}^{\alpha}\,|$, and a spin-down ($\beta$) determinant $\text{Det}\,|\,D_{ij}^{\beta}\,|$ as shown below for LiH:

$$\text{Det}\,|\,D_{ij}^{\alpha}\,| = \begin{vmatrix} \Phi_1(\mathbf{r}_1) & \Phi_1(\mathbf{r}_2) \\ \Phi_2(\mathbf{r}_1) & \Phi_2(\mathbf{r}_2) \end{vmatrix} = \Phi_1(\mathbf{r}_1)\,\Phi_2(\mathbf{r}_2) - \Phi_2(\mathbf{r}_1)\Phi_1(\mathbf{r}_2)$$

$$\text{Det}\,|\,D_{ij}^{\beta}\,| = \begin{vmatrix} \Phi_1(\mathbf{r}_3) & \Phi_1(\mathbf{r}_4) \\ \Phi_2(\mathbf{r}_3) & \Phi_2(\mathbf{r}_4) \end{vmatrix} = \Phi_1(\mathbf{r}_3)\,\Phi_2(\mathbf{r}_4) - \Phi_2(\mathbf{r}_3)\Phi_1(\mathbf{r}_4)$$

where $\mathbf{r}_j$ represents the coordinates of the jth electron (here there are four electrons altogether), and Φ_k is the kth molecular orbital formed from linear combinations of AO basis functions. These are centered at the nuclei (Li or H), and are given by

$$\Phi_1(\mathbf{r}) = \sum_{j=1,N_{AO}} C_{1j}\,\chi_j \qquad (\text{II.A - 2})$$

$$\Phi_2(\mathbf{r}) = \sum_{j=1,N_{AO}} C_{2j}\,\chi_j \qquad (\text{II.A - 3})$$

where N_{AO} is the total number of the AO functions χ_j's used in a basis set. The C_{ij}'s are the linear MO coefficients to be optimized. The AO functions χ_j's are shown in Table II.1 below. The exponents ζ in the functions will also be optimized.

Table II. 1. The AO functions used.

AO	Functions χ
1s	$e^{-\zeta r}$
2s	$r\,e^{-\zeta r}$
$2p_z$	$z\,e^{-\zeta r}$
3s	$(3 - 6\zeta r + 2\zeta^2 r^2)\,e^{-\zeta r}$
$3p_z$	$z(2 - \zeta r)\,e^{-\zeta r}$
$3d_{z^2}$	$(2z^2 - x^2 - y^2)\,e^{-\zeta r}$

The last factor in Eq.(II.A - 1) is a Jastrow factor[1], where u_{ij} is defined by,

$$u_{ij} = a r_{ij}/(1+b r_{ij}) \qquad \text{(II.A - 4)}$$

where a and b are constants. When considering the a and b related to r_{ij} (the distance between two electrons e_i and e_j), one can divide them into **antiparallel** a_a, b_a (e_i and e_j are of opposite spins) and **parallel** a_p, b_p (e_i and e_j are of same spins). This factor is always positive, and any adjustments to the constants a and b or a_a, b_a and a_p, b_p don't change the nodes of Ψ_T. Thus, the effort to adjust the constants can decrease the variance only, but not lower the fixed-node energies.[1,18] The variational energy will lower, however, as these constants are adjusted. In this work the a and b, or a_a, b_a and a_p, b_p are considered as adjustable parameters to be optimized.

2. Basis Sets

There are five AO (Table II.1) basis sets explored in this work. They are a 3 - function Set (see Table II. 2.), 5 - Function Set(Table II. 3.), Double - ζ Set (Table II. 4), Double - ζ plus $3d_z^2$ Set (Table II. 5.), and a 9 - Function Set (Table II. 6.).

Table II. 2. 3 - function Set

Index (χ_i)	Nucleus	AO	Function χ_i
1	Li	1s	$e^{-\zeta_1 r_{Li}}$
2	Li	2s	$r_{Li}\, e^{-\zeta_2 r_{Li}}$
3	H	1s	$e^{-\zeta_3 r_H}$

Table II. 3. 5 - function Set

Index (χ_i)	Nucleus	AO	Function χ_i
1	Li	1s	$e^{-\zeta_1 r_{Li}}$
2	Li	2s	$r_{Li} e^{-\zeta_2 r_{Li}}$
3	Li	$2p_z$	$z_{Li} e^{-\zeta_3 r_{Li}}$
4	H	1s	$e^{-\zeta_4 r_H}$
5	H	$2p_z$	$z_H e^{-\zeta_5 r_H}$

Table II. 4. Double - ζ Set

Index (χ_i)	Nucleus	AO	Function χ_i
1	Li	1s	$e^{-\zeta_1 r_{Li}}$
2	Li	1s	$e^{-\zeta_2 r_{Li}}$
3	Li	2s	$r_{Li} e^{-\zeta_3 r_{Li}}$
4	Li	$2p_z$	$z_{Li} e^{-\zeta_4 r_{Li}}$
5	Li	$2p_z$	$z_{Li} e^{-\zeta_5 r_{Li}}$
6	H	1s	$e^{-\zeta_6 r_H}$
7	H	1s	$e^{-\zeta_7 r_H}$
8	H	$2p_z$	$z_H e^{-\zeta_8 r_H}$

Table II. 5. Double - ζ plus $3d_{z^2}$ Set

Index (χ_i)	Nucleus	AO	Function χ_i
1	Li	1s	$e^{-\zeta_1 r_{Li}}$
2	Li	1s	$e^{-\zeta_2 r_{Li}}$
3	Li	2s	$r_{Li} e^{-\zeta_3 r_{Li}}$
4	Li	$2p_z$	$z_{Li} e^{-\zeta_4 r_{Li}}$
5	Li	$2p_z$	$z_{Li} e^{-\zeta_5 r_{Li}}$
6	H	1s	$e^{-\zeta_6 r_H}$
7	H	1s	$e^{-\zeta_7 r_H}$
8	H	$2p_z$	$z_H e^{-\zeta_8 r_H}$
9	Li	$3d_{z^2}$	$(2z^2-x^2-y^2) e^{-\zeta_9 r_{Li}}$

Table II. 6. 9 - function Set

Index (χ_i)	Nucleus	AO	Function χ_i
1	Li	1s	$e^{-\zeta_1 r_{Li}}$
2	Li	3s	$(3 - 6\zeta_2 r_{Li}+2\zeta_2{}^2 r_{Li}{}^2) e^{-\zeta_2 r_{Li}}$
3	Li	2s	$r_{Li} e^{-\zeta_3 r_{Li}}$
4	Li	$2p_z$	$z_{Li} e^{-\zeta_4 r_{Li}}$
5	Li	3p	$z_{Li} (2-\zeta_5 r_{Li}) e^{-\zeta_5 r_{Li}}$
6	H	1s	$e^{-\zeta_6 r_H}$
7	H	2s	$z_H e^{-\zeta_7 r_H}$
8	H	$2p_z$	$z_H e^{-\zeta_8 r_H}$
9	Li	$3d_{z^2}$	$(2z^2-x^2-y^2) e^{-\zeta_9 r_{Li}}$

B. VQMC estimate of the variational energy.

Before beginning a VQMC simulation one must choose the trial wavefunction ψ_T, timesteps τ, and the size of an ensemble (several hundred configurations) M_c. After first arbitrarily distributing the electrons $f(\mathbf{R},t=0)$, each configuration **Drifts**, and then **Jumps** according to Eq. (I - 3) and Eq. (I - 4). Thus, for a single electron

$$\mathbf{r'}_j^m = \mathbf{r}_j^m + \tau \mathbf{F}(\mathbf{r}_j^m) + N(0,\sqrt{\tau}) \qquad (\text{II.B - 1})$$

where $\mathbf{r}_j^m$ is the set of 3 Cartesian coordinates of the jth electron in the mth configuration and $\mathbf{r'}_j^m$ is that after a time τ. When all N_e electrons ($N_e = 4$ in the LiH molecule) and all M_c configurations have been taken once, a new ensemble ($f(\mathbf{R'},t=\tau)$) is created and one **iteration** is finished. One repeats the iterations until a target time value is reached (a large number of iterations have been completed) and the approximate $f_\infty(\mathbf{R})$ of an equilibrium state is given.

Suppose this equilibrium state is reached for a selected time step τ. One may continually run N_{IT} iterations to complete one **block**. The nth iteration mean of energy $E^n_{av}(\tau)$ is given by taking the average of the local energies $E_L(\mathbf{R})$ over all M_c configurations in the nth iteration with a variance $\mathrm{Var}(E^n_{av}(\tau))$, i.e.

$$\begin{aligned} E^n_{av}(\tau) &= \sum_{m=1,M_c} E_L(\mathbf{R}^m)/M_c \\ &= \sum_{m=1,M_c} (\mathbf{H}\,\psi_T(\mathbf{R}^m) / \psi_T(\mathbf{R}^m))/M_c \end{aligned} \qquad (\text{II.B - 2})$$

$E^n_{av}(\tau)$ has a τ-related bias[23] according to

$$E^n_{av}(\tau) = E_{var} + a\tau + b\tau^2 + ... \qquad (II.B - 3)$$

Note that the dominant reason for the introduction of this timestep bias is due to the Green's function being correct only at $\tau = 0$. The $Var(E^n_{av}(\tau))$ (or just $Var(E_{av}(\tau))$) is called the **individual iteration variance**(or just **iteration variance**), defined by

$$Var(E^n_{av}(\tau)) = \sum_{m=1,M_c} (E^n_{av}(\tau) - E_L(\mathbf{R}^m))^2/M_c \qquad (II.B - 4)$$

Of course, this variance is averaged over the iterations to obtain the **grand mean** of the iteration variance

$$Var_{gm}(E_{av}(\tau)) = < Var(E^n_{av}(\tau)) >$$
$$= \sum_{n=1,N_{IT}} Var(E^n_{av}(\tau))/N_{IT} \qquad (II.B - 5)$$

Since the smaller the grand mean of the iteration variance is, the more accurate the trial wavefunction is, and since it is also τ–dependent, we will use its extrapolated form ($\tau = 0$) as one of the two important criterions for determining an optimized trial wavefunction.

Similarly, taking the average of $E^n_{av}(\tau)$ over all the N_{IT} iterations, we get the **grand mean** of the local energy,

$$E_{gm}(\tau) = < E^{n}_{av}(\tau) > = \sum_{n=1,N_{IT}} E^{n}_{av}(\tau)/N_{IT} \qquad (II.B - 6)$$

The variance of $E_{gm}(\tau)$ is still τ–dependent[22], given by

$$Var(E_{gm}(\tau)) = \sigma^2(\tau)/N_{IT} \qquad (II.B - 7)$$

and

$$\sigma^2(\tau) = \sigma^2/\tau + ... \qquad (II.B - 8)$$

To reduce the statistical errors of the estimates at any timestep τ, several more such grand means are usually required to obtain an average $<E_{gm}(\tau)>$ and $<Var_{gm}(E_{av}(\tau))>$, i.e. several more blocks are needed. Furthermore, there are at least 30 iterations "thrown away" between two blocks to avoid serial correlation. If N_b blocks are taken for timestep τ, we have the block means $E_b(\tau)$ and $Var_b(E_{av}(\tau))$ as shown below,

$$E_b(\tau) = < E_{gm}(\tau) > = \sum_{j=1,N_b} E_{gm}(\tau)/N_b \qquad (II.B - 9)$$

$$Var_b(E_{av}(\tau)) = < Var_{gm}(E_{av}(\tau)) > = \sum_{j=1,N_b} Var_{gm}(E_{av}(\tau))/N_b \qquad (II.B - 10)$$

Finally, to remove the τ-related bias, the whole procedure described above is repeated for several distinct values of τ, and then these $E_b(\tau)$ and $Var_b(E_{av}(\tau))$ are extrapolated to τ=0 by means of statistical regression[1,18,23-25]. Thus, we ultimately get the ground state variational energy E_{var} and the extrapolated individual iteration variance

$Var(E_{av}(\tau \rightarrow 0))$ without the τ-related bias, by which we can evaluate how good the trial wavefunctions are.

C. Optimization

In the optimization, the Levenberg-Marquardt algorithm[26-29] was employed by calling the ZXSSQ routine from the IMSL Library[26] on the Burroughs 7900 mainframe.

Before doing the optimization, several sets of equilibrium configurations (ensembles) must be created by simulation with the initial trial wavefunction Ψ_0. The ensembles created by a simulation with the DJ algorithm are called **DJ ensembles**, or **DJB ensembles** with the DJB algorithm. By sampling these equilibrium ensembles with a total size of M_c configurations, we get M_c nonlinear functions $SQ_1, SQ_2, ..., SQ_{Mc}$ of a set of J parameters $\{P_j\}$ to be estimated. In this work the functions SQ_i are defined either as,

$$SQ_i (P_{j, j=1,J}) = [H\psi_T(\mathbf{R}^m)/\psi_T(\mathbf{R}^m) - E_g]^2$$
$$= [E_L(\mathbf{R}^m) - E_g]^2 \qquad (II.C - 1)$$

or,

$$SQ_i (P_{j, j=1,J}) = [E_L(\mathbf{R}^m) - E_{av}(\tau)]^2 \qquad (II.C - 2)$$

where E_g may be taken as E_{var} or the best estimate obtained. The latter is the standard function approprate for the variance. The sum of these functions

SSQ ($P_{j,\,j=1,J}$) will be minimized over the space $\{P_j\}$, where

$$SSQ\ (P_{j,\,j=1,J}) = \sum_{m=1,Mc} [E_L(\mathbf{R}^m) - E_g]^2 \qquad (II.C - 3)$$

or,

$$SSQ\ (P_{j,\,j=1,J}) = \sum_{m=1,Mc} [E_L(\mathbf{R}^m) - E_{av}(\tau)]^2 \qquad (II.C - 4)$$

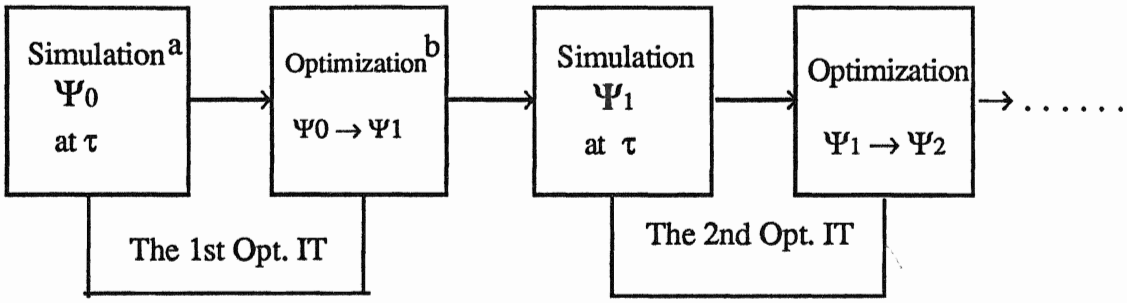
It is worth noting that $E_{av}(\tau)$ varies with the change of $\{P_j\}$, whereas E_g is a constant. To explore the significance of the choice is an objective of this thesis.

If weighted optimization[20] is being used, the SSQ is

$$SSQ\ (P_{j,\,j=1,J}) = \sigma^2_{opt} = \frac{\sum_{i=1}^{M} SQ_i(P_{j,\,j=1,J})w(i)}{\sum_{i=1}^{M} w(i)} \qquad (II.C - 5)$$

where w(i) is the same as in Eq.(I - 2).

This procedure, simulation with the Ψ_0, and then optimization starting with the ψ_0, is called an **optimization iteration**, or abbreviated as **Opt. IT**. More than one such iteration is always needed to get a set of equilibrium parameters $\{P_j\}$. Here the optimization is repeated using the optimized Ψ_1 from the previous **Opt. IT** as Ψ_0. The logic route of this procedure is shown in the Fig. II. 1, where there are two such Opt. IT's.



[a]The simulation at τ with the trial wavefunction Ψ_0.
[b]The optimization uses ensemble generated from Ψ_0 and yields the "best" function Ψ_1.

Fig. II. 1. The logic route of the optimization iterations.

III. Exploration of Optimization

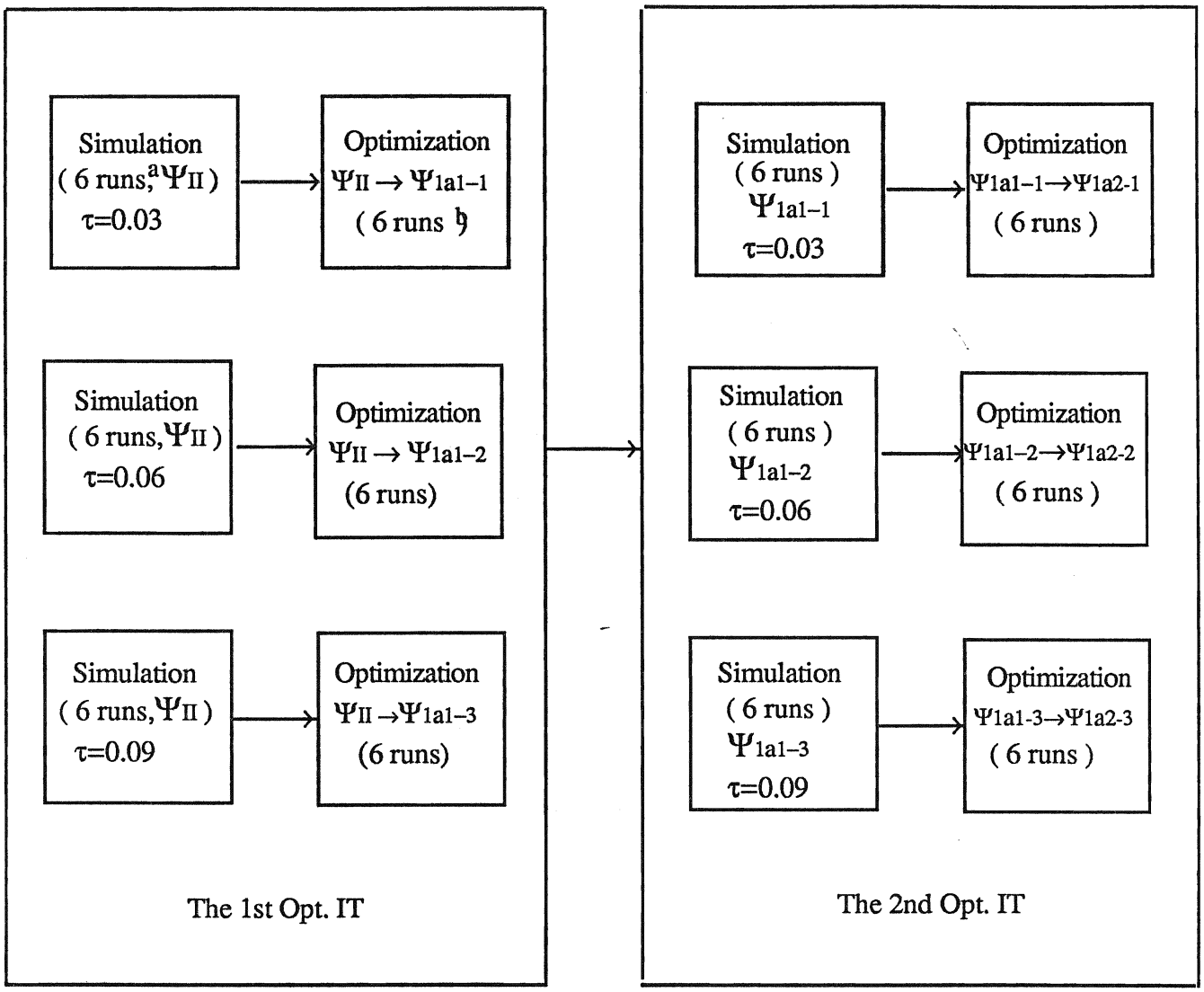
A. Preliminary Investigations by Partial Optimizations.

The preliminary investigations involve the optimizations of some (but not all) of the exponents, the split Jastrow constants, and the geometry in the trial wavefunction with the double-ζ set, sampling from the DJB or DJ ensembles, for the ground state system of LiH.

1. Optimization of the "dominant" parameters sampling from the DJB ensembles at three timesteps.

Work began with the assumption that the "dominant" AO functions were comprised of one function of each type in the double-ζ set, with the duplicated functions being "non-dominant". Also, the Jastrow constants **a** and **b** were considered "dominant" because these introduce electron correlation which lowers the variational energy and gives a smaller variance. So there are altogether 7 "dominant" parameters, ζ_1, ζ_3, ζ_4, ζ_6, ζ_8, Jast. **a** and **b**. (See Table II.4). The optimization of these parameters first started with Reynold's Ψ_{II} (see Table 1 in App. B), and Eq.(II.C - 4) (optimization without weights) was used to directly minimize the individual iteration variance. Fig. III.A.1.1. shows the logic route of this optimization procedure.

With Reynold's Ψ_{II} and DJB algorithm, six equilibrium ensembles (each one has 600 configurations) were obtained by six simulation runs at each timestep, which were then used to do the optimizations independently.



[a]Each run is a block of 300 iterations with 600 configurations in the ensemble, which are the same as in Fig. III.A.2.1-2 and Fig. III.A.3.1-2.
[b]Each run involves the optimization sampling the final equilibrium ensemble created by a block.

Fig. III.A.1.1. Logic route of the optimizations of all the "dominant" parameters.

In the simulation as well as in optimization the "truncation" technique[24] was used, where the singularities of $E_L(\mathbf{R})$ and $F(\mathbf{R})$ were truncated in a τ-neighborhood of each singularity.

One finds from Table III.A.1.2.-3 that after the 1st Opt. IT, the individual iteration variances are more than twice as small as before, and the energies are more reliable with smaller standard error and closer values at the different timesteps. A more interesting thing is to compare their simulation results. See Table.III.A.1.1 and 4, and Fig. III.A.1.2-5. Though the variational energies are a little different, the $E_b(\tau)$ vs. τ curve with the optimized trial wavefunctions Ψ_{1a1}'s ($\Psi_{1a1\text{-}1,2,3}$) is more flat. In any case, the great change is in the variances as expected, the result obtained by using Ψ_{1a1}'s being less than half that using the Reynold's Ψ_{II}.

Since in the calculation of the energy in DQMC, the trial wavefunction is chosen to locate the nodes of the exact wavefunction only, and it only affects the variance of the energy but not the energy itself, **the optimized Ψ_{1a1} has a much more accurate nodal surface than that of unoptimized Ψ_{II}, and the Levenberg - Marquardt routine used is very powerful for doing this optimization.**

Note that the extrapolated energies given by these two wavefunctions here are all *lower* than the exact one (-8.070 a. u.)[30]. That is because the calculations here are "rough": only three timesteps with six blocks at each of these timesteps, so that the extropolations are unreliable.

By examining the optimized parameters, we find some of them are

Table III. A. 1. 1. DQMC estimates with Reynold's Ψ_{II}.

τ	$E_b(\tau)$[a]	$Var_b(E_{av}(\tau))$
0.03	-8.060±0.001	0.285±0.0020
0.06	-8.018±0.002	0.273±0.0010
0.09	-7.949±0.001	0.268±0.0004

Extrapolated[b] to τ=0: $E_0 = -8.075 \pm .0068$, $Var(E_{av}(\tau \rightarrow 0)) = .304 \pm .0067$

[a]The block mean at each timestep is taken from 6 blocks, and there are 40 iterations in a block, and 600 configurations in an ensemble.
[b]The regression model is quadratic. See Fig.III.A.1.2 - 3.

Table III. A. 1. 2. Iteration means in the optimization of the **1st Opt. IT.**

τ	$E_{av,ini}$[a]	$E_{av,fin}$[b]	$Var(E_{av})_{ini}$[c]	$Var(E_{av})_{fin}$[d]
0.03	-8.080 ± .020	-8.066 ± .006	.308 ± .020	.109 ± .008
0.06	-8.257 ± .137	-8.056 ± .013	.541 ± .225	.126 ± .019
0.09	-8.096 ± .006	-8.061 ± .007	.301 ± .009	.103 ± .002

[a,c]The mean of the individual iteration means in the six runs before optimization, "initial".
[b,d]The mean of the individual iteration means in the six runs after optimization, "finial".

Table III. A. 1. 3. Changes of the iteration means in the optimization of the **1st Opt. IT.**

τ	ΔE_{av}[a]	$\Delta Var(E_{av})$[b]
0.03	0.014 ± 0.021	-0.199 ±0.022
0.06	0.201 ± 0.138	-0.415 ±0.226
0.09	0.035 ± 0.009	-0.198 ±0.009

[a,b]The changes of the individual iteration means in the optimization of the 1st Opt. IT, which are obtained directly from Table III.A.1.2.

Table III. A. 1. 4. DQMC estimates with Ψ_{1a1}.

τ	$E_b(\tau)$	$Var_b(E_{av}(\tau))$
0.03	-8.065±0.001	0.113±0.0010
0.06	-8.047±0.001	0.107±0.0010
0.09	-8.022±0.001	0.106±0.0003
Extrapolated[b] to τ=0:	E_0 = -8.076 ± .0044,	$Var(E_{av}(\tau\rightarrow 0))$ = .124 ± .0043

[a]The block mean at each timestep is taken from 6 blocks, and there are 40 iterations in a block, and 600 configurations in an ensemble.
[b]The regression model is quadratic. See Fig.III.A.1.4-5.

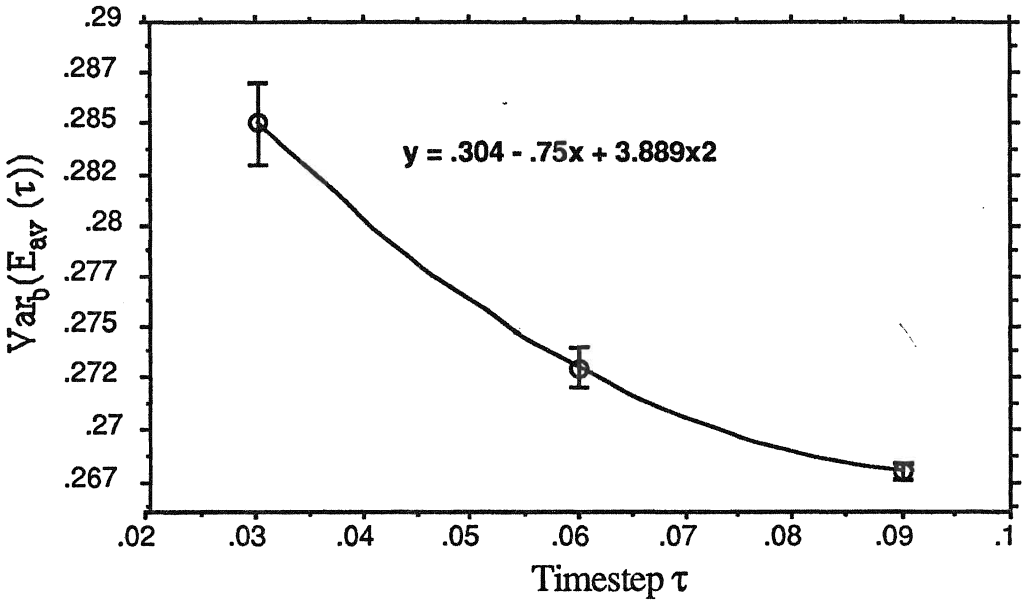


Fig. III. A.1. 2. DQMC estimate(rough) of Var(Eav(τ→0)) with Reynold's Ψ_{II}.

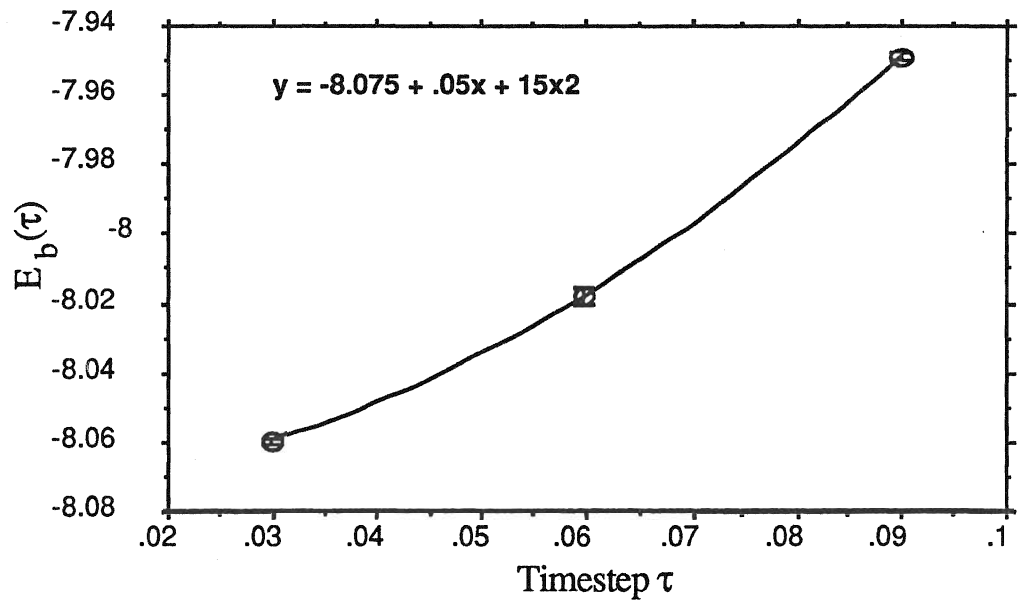


Fig. III. A.1. 3. DQMC estimate (rough) of E_0 with Reynold's Ψ_{II}.

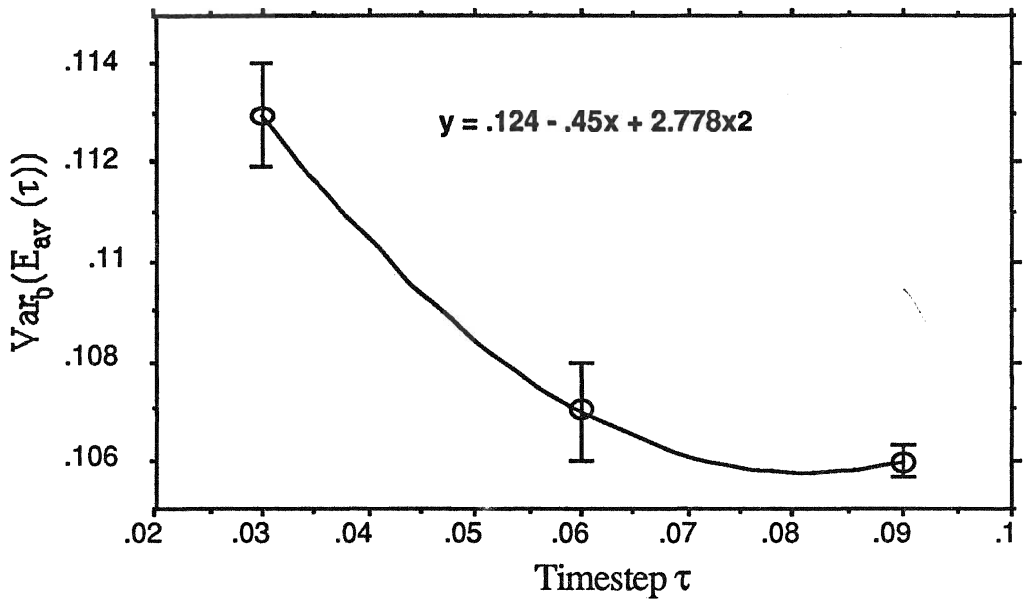


Fig. III. A.1. 4. DQMC estimate(rough) of $Var(E_{av}(\tau \rightarrow 0))$ with Ψ_{1a1}.

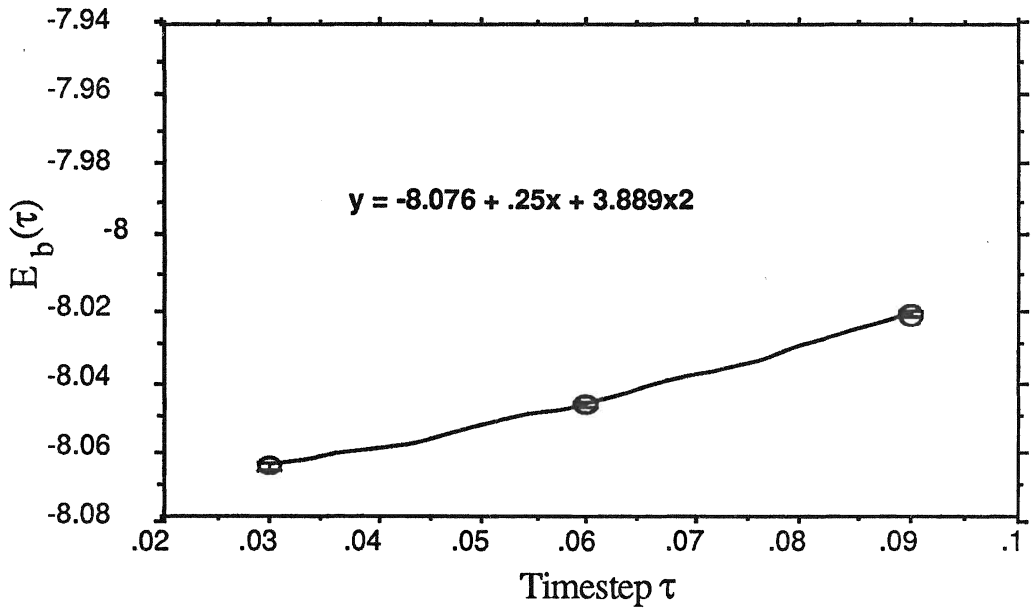


Fig. III. A.1. 5. DQMC estimate(rough) of E_0 with Ψ_{1a1}.

different at the different timesteps (see Table A.1.5. and Fig.III.A.1.6 -12, which show the curves of parameters vs. τ). It was naturally assumed that **the optimizations were timestep-dependent**, and some of the parameters were very sensitive, while some were not. Since the SSQ to be minimized is a function of the whole set of all parameters $\{P_j\}$, it is not necessary to test which ones are more or less sensitive. The important task is to verify further if the optimization is indeed timestep-dependent just as is the case in the simulation.

One more Opt. IT. was done starting with Ψ_{1a1}, where the individual iteration average energies and their variances didn't change much, compared with that of the first Opt. IT (See Table III.A.1.6-7 and 2-3).

Table III.A.1.8 and Fig. III.A.1.13-19 show once again the timestep behaviours of these parameters. But some of the behaviors are different from before. For example, $\zeta 3$ increases as τ here, but it doesn't in the 1st Opt. IT (See Fig. III.A.1.14 and 7). Thus it was doubtful whether the timestep behaviours of the parameters was just a fluctuation due to the optimizations' sampling from different ensembles. More investigations are required.

Table III. A. 1. 5. "Best" "dominant" parameters[a] in the1st Opt. IT .

τ	ζ_1	ζ_3	ζ_4	ζ_6	ζ_8	Jast a	Jast b
0.03	2.583 ± 0.007	0.870 ± 0.025	0.794 ± 0.070	0.943 ± 0.030	1.820 ± 0.133	0.545 ± 0.014	1.014 ± 0.032
0.06	2.593 ± 0.006	1.103 ± 0.236	1.099 ± 0.068	0.885 ± 0.041	1.773 ± 0.051	0.537 ± 0.017	0.944 ± 0.028
0.09	2.603 ± 0.005	0.855 ± 0.019	1.042 ± 0.076	0.935 ± 0.008	1.789 ± 0.065	0.519 ± 0.007	0.909 ± 0.012

[a]These values are the averages taken from the six runs.

Table III. A. 1. 6. Iteration means in the 2nd Opt. IT.

τ	$E_{av,ini}$	$E_{av,fin}$	$Var(E_{av})_{ini}$	$Var(E_{av})_{fin}$
0.03	-8.060 ± .004	-8.069 ± .005	.117 ± .003	.111 ± .003
0.06	-8.068 ± .005	-8.079 ± .004	.124 ± .005	.116 ± .004
0.09	-8.076 ± .004	-8.077 ± .004	.111 ± .001	.107 ± .002

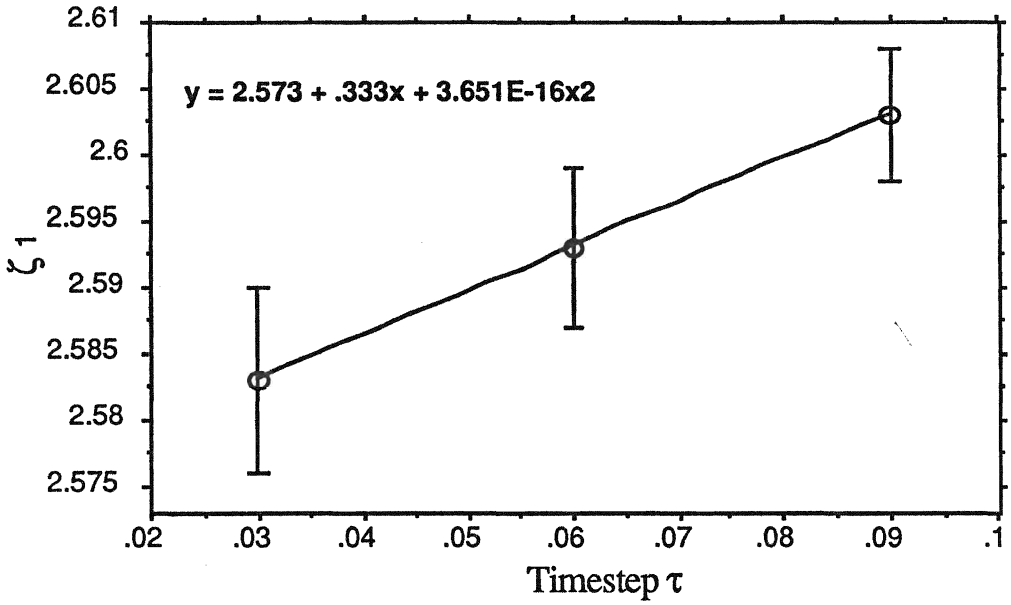


Fig. III. A. 1. 6. "Best" ζ1 vs. τ in the 1st Opt. IT.

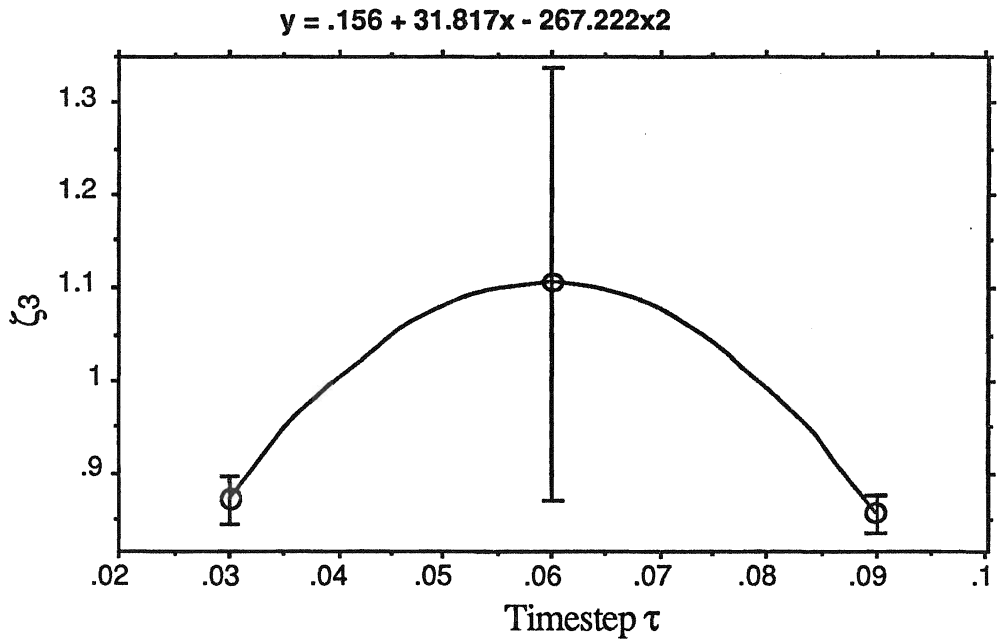


Fig. III. A. 1. 7. "Best" ζ3 vs. τ in the 1st Opt. IT.

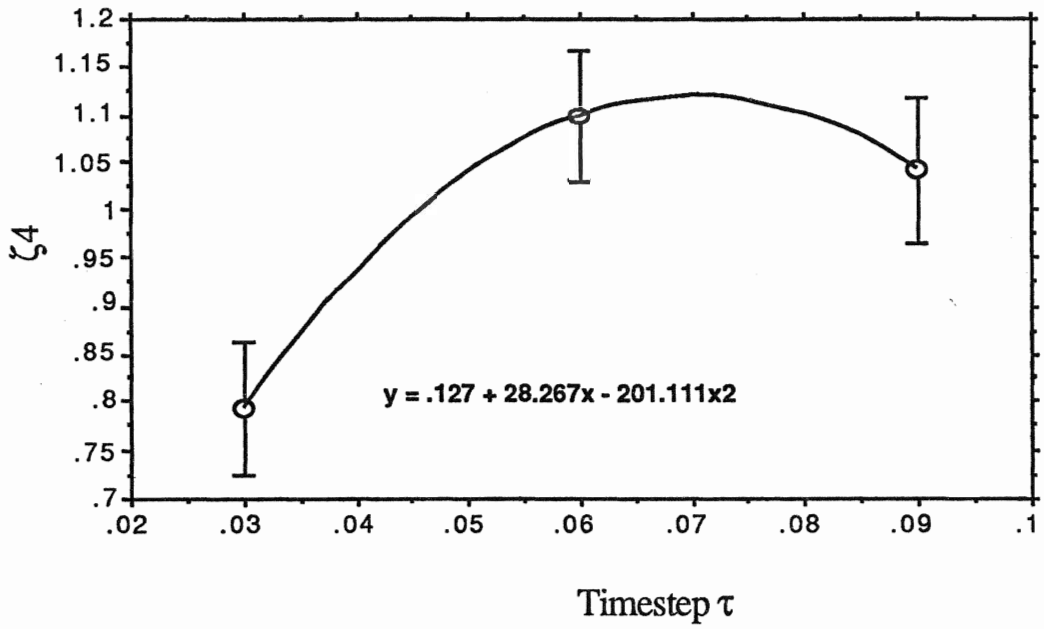


Fig. III. A. 1. 8. "Best" ζ4 vs. τ in the 1st Opt. IT.

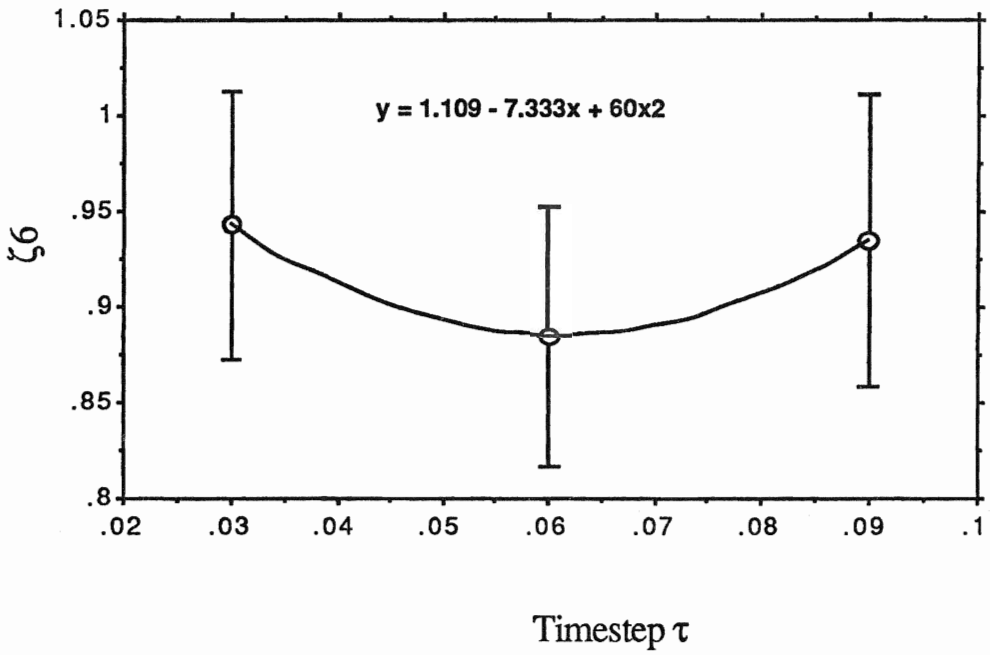


Fig. III. A. 1. 9. "Best" ζ6 vs. τ in the 1st Opt. IT.

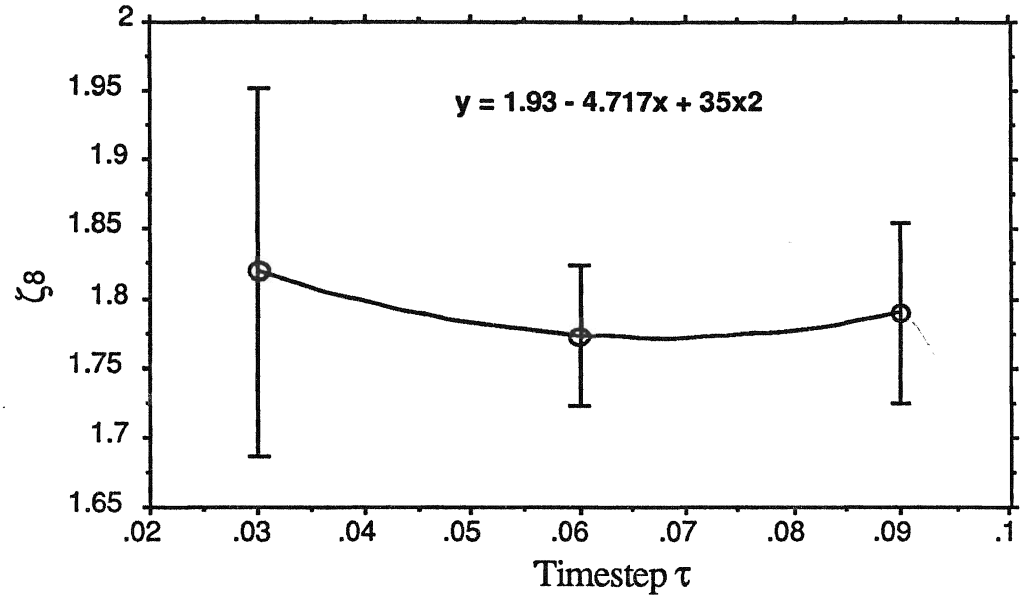


Fig. III. A. 1. 10. "Best" ζ8 vs. τ in the 1st Opt. IT.

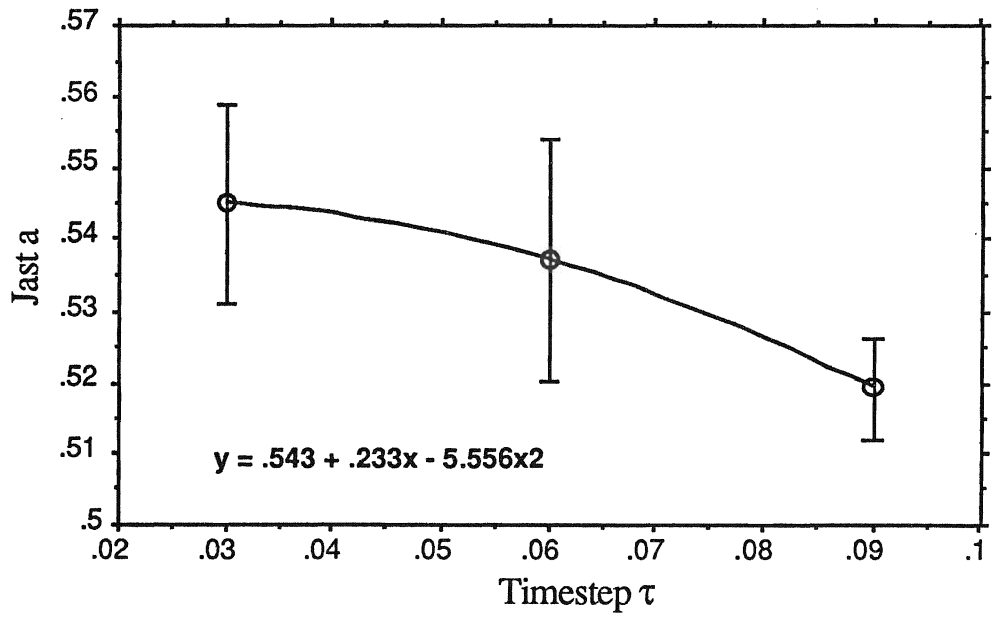


Fig. III. A. 1. 11. "Best" Jastrow a vs. τ in the 1st Opt. IT.

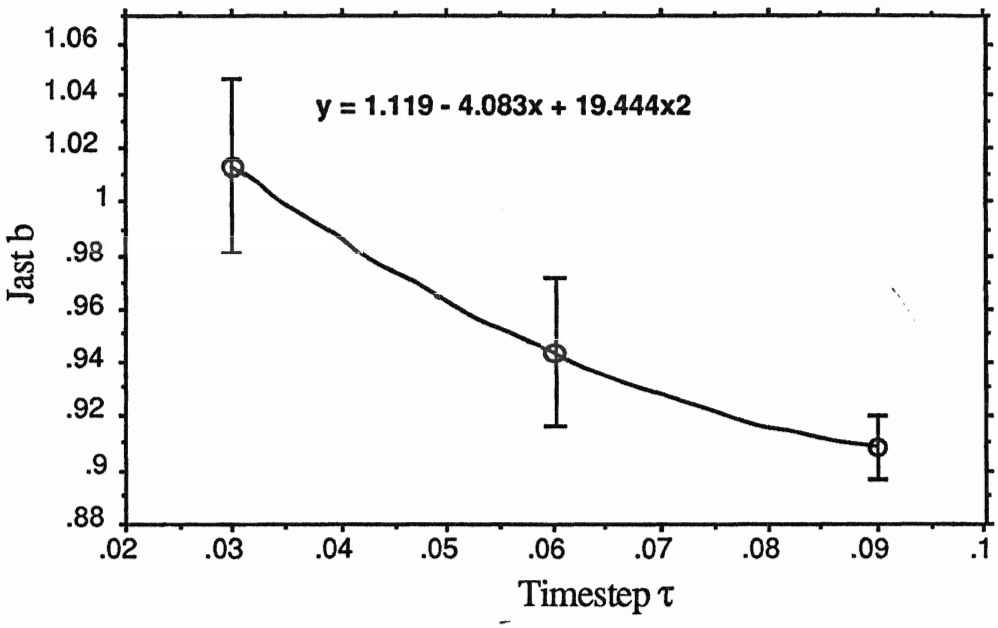


Fig. III. A. 1. 12. "Best" Jastrow b vs. τ in the 1st Opt. IT.

Table III. A. 1.7. Changes of the iteration means in the optimization of the **2nd Opt. IT.**

τ	ΔE_{av}[a]	$\Delta Var(E_{av})$[b]
0.03	-0.009 ± 0.006	-0.006 ±0.004
0.06	-0.011 ± 0.006	-0.008 ±0.006
0.09	-0.001 ± 0.006	-0.004 ±0.002

[a,b]The changes of the individual iteration means in the optimization of the 1st Opt. IT, which are obtained directly from Table III.A.1.6.

Table III. A. 1. 8. "Best" "dominant" parameters[a] in the **2nd Opt. IT.**

τ	ζ_1	ζ_3	ζ_4	ζ_6	ζ_8	Jast a	Jast b
0.03	2.579 ± 0.003	0.783 ± 0.039	0.992 ± 0.043	0.923 ± 0.008	1.732 ± 0.094	0.536 ± 0.010	1.000 ± 0.025
0.06	2.577 ± 0.005	0.823 ± 0.033	0.952 ± 0.028	0.916 ± 0.010	1.803 ± 0.065	0.521 ± 0.005	0.978 ± 0.013
0.09	2.612 ± 0.005	0.869 ± 0.020	1.033 ± 0.033	0.903 ± 0.007	1.712 ± 0.038	0.522 ± 0.008	0.937 ± 0.022

[a]These values are the averages taken from the six runs.

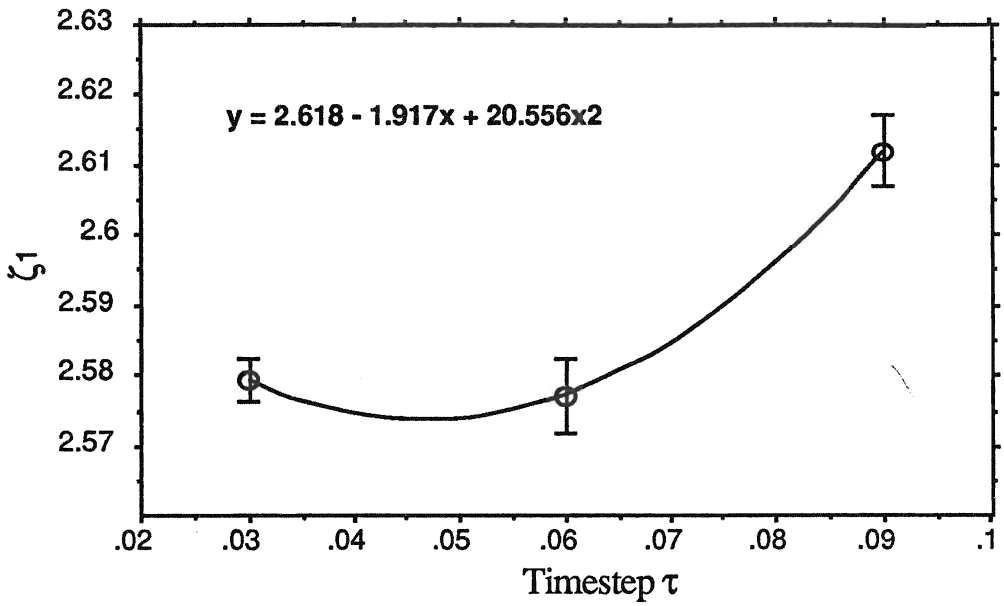


Fig. III. A.1.13. "Best" ζ1 vs. τ in the **2nd Opt. IT.**

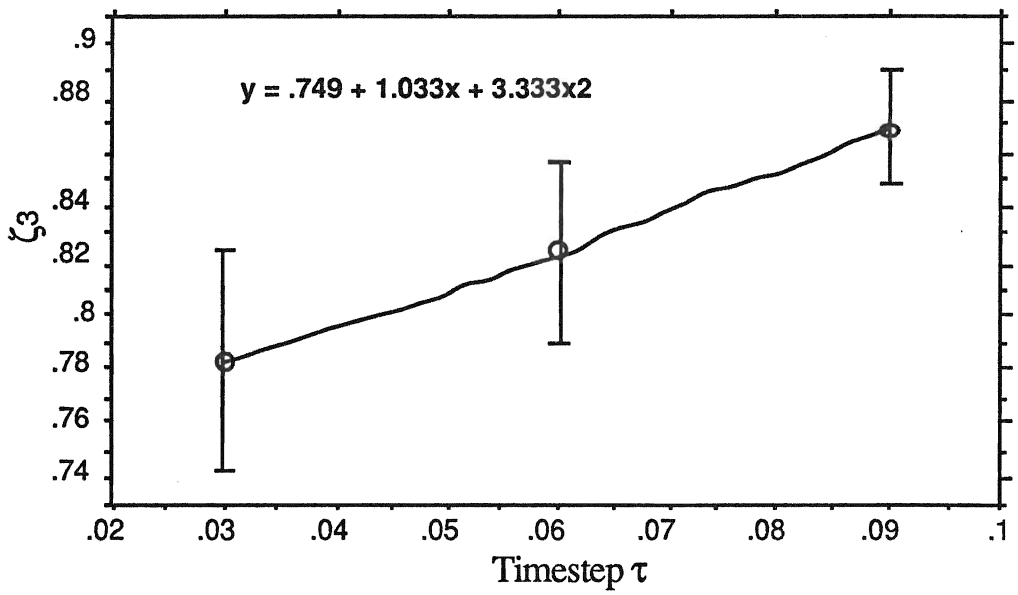


Fig. III. A.1.14. "Best" ζ3 vs. τ in the **2nd Opt. IT.**

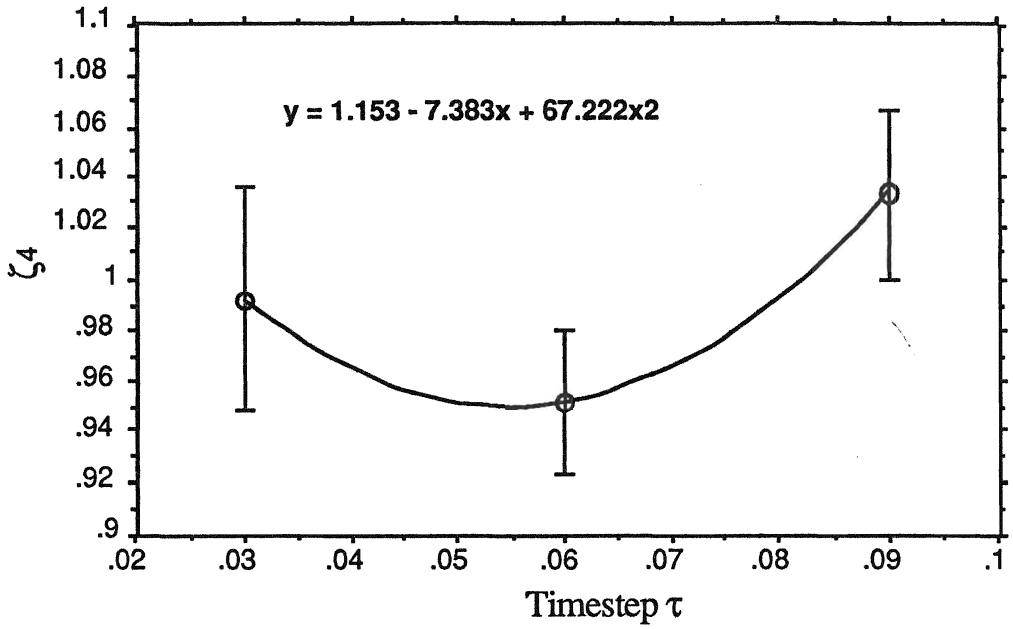


Fig. III. A.1.15. "Best" ζ4 vs. τ in the **2nd Opt. IT.**

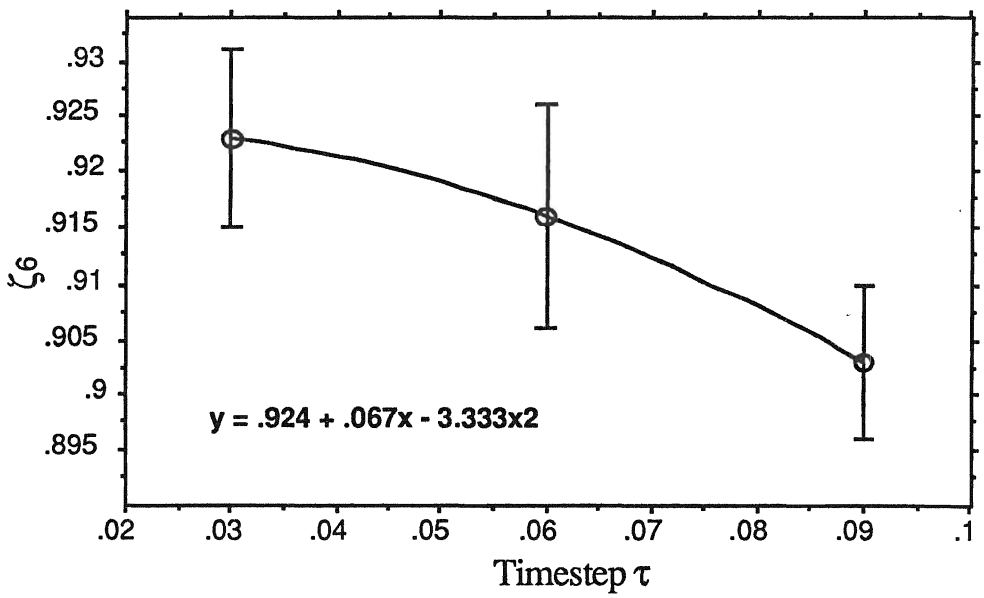


Fig. III. A.1.16. "Best"ζ6 vs. τ in the **2nd Opt. IT.**

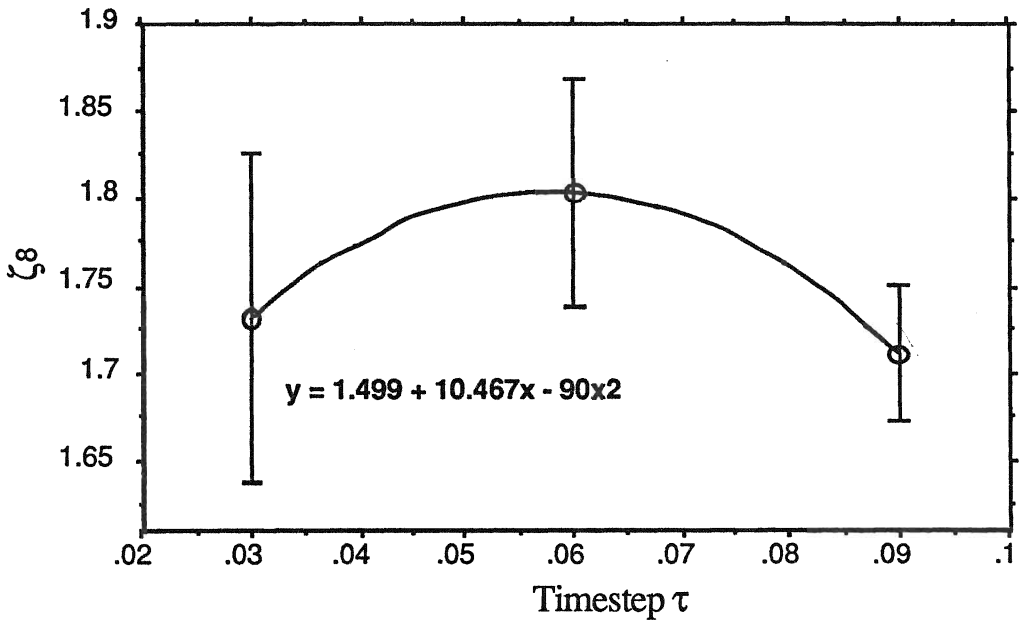


Fig. III. A.1.17. "Best" ζ8 vs. τ in the **2nd Opt. IT.**

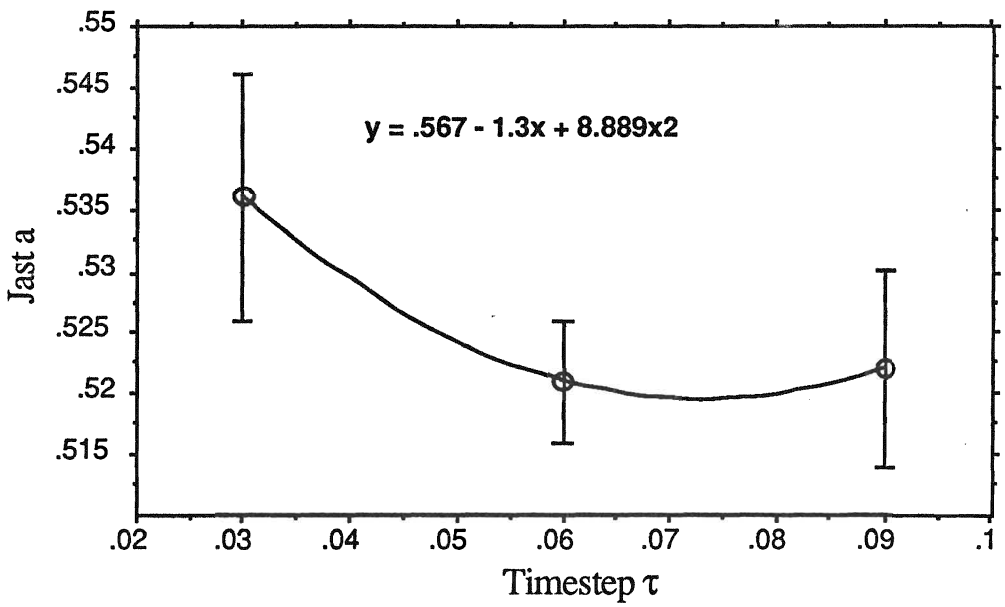


Fig. III. A.1.18. "Best" Jastrow a vs. τ in the **2nd Opt. IT.**

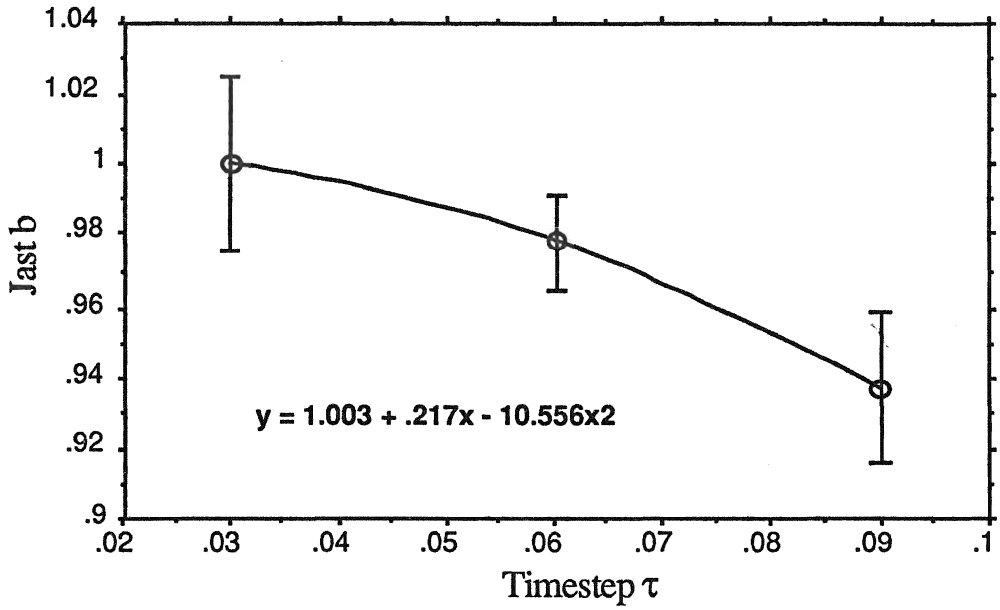


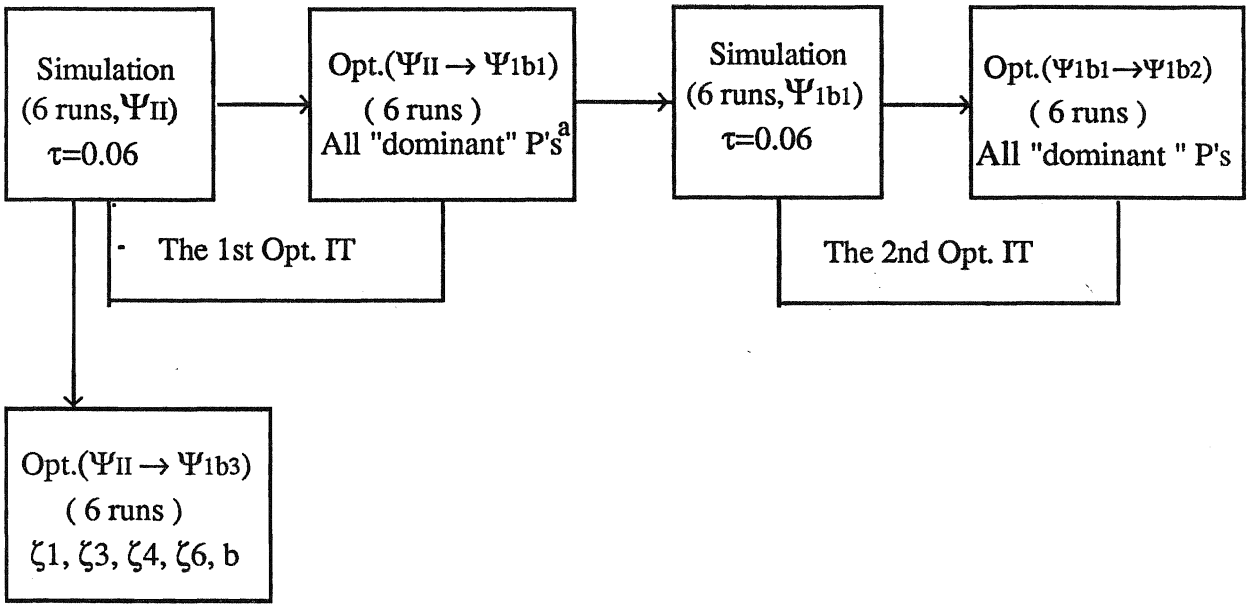
Fig. III. A.1.19. "Best" Jast b vs. τ in the **2nd Opt. IT.**

2. Optimization of the "dominant" parameters sampling from the DJ ensembles at $\tau = 0.06$.

A direct optimization sampling from the distribution of $|\Psi_T|^2$ (the DJ ensemble) should be theoretically more efficient than that of $\phi_0\Psi_T$ (the DJB ensemble). That is because we are minimizing the variance (see Eq.II.C - 3 and 4), and the computations of $E_L(\mathbf{R})$ rely on the Ψ_T only. Thus a lot of singularities due to the differences between ϕ_0 and Ψ_T will be avoided if one samples $|\Psi_T|^2$ instead of $\phi_0\Psi_T$, and furthermore, there is no need to employ the truncation on $E_L(\mathbf{R})$, which will delay the convergence time. On the other hand, since a lot of singularities have been avoided for DJ configurations, **the DJ ensembles are in some sense "larger" than the DJB, being used more efficient**. We will show later that, the larger the size of the ensemble used in optimization, the better the wavefunction which will be obtained. One other obvious fact is that, because of the lack of branching the DJ algorithm is much simpler than that of the DJB. So, the **partial** optimization by directly sampling the distribution of $|\Psi_T|^2$ is more acceptable.

These points will be proved by comparing the results of the optimizations sampling from the two kinds of ensembles. For this purpose the procedure was repeated for $\tau = 0.06$, as in the last section, but starting with the **DJ ensembles** of Ψ_{II} and **no truncation on the local energies** in the optimization. See the logic route on the top of Fig. III.A.2.1.

Table III.A.2.3 (also Table 3 and 4 in App. B) presents the "best" optimized parameters after the 1st, and the 2nd Opt. IT. The two optimized



[a]The parameters inside the box are those being optimized, and similarly in the following figures.

Fig. III. A. 2. 1. Logic routes of the optimizations of the "dominant" parameters.

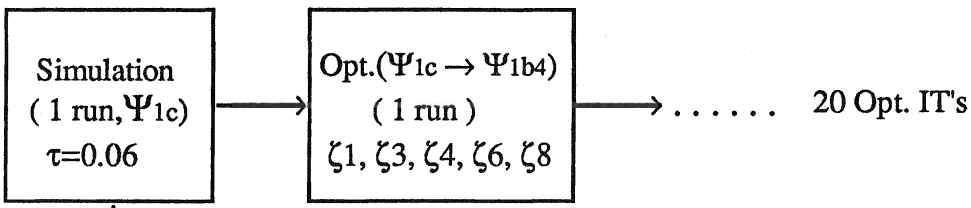


Fig. III. A. 2. 2. Logic route of the optimizations of the five "dominant" parameters.

functions, Ψ_{1b1} and Ψ_{1b2}, gave almost the same individual iteration energies and variances in optimization (Table III.A.2.1-2), and also the same block means in simulation (Table III.A.2.4). That means one only needs to do one Opt. IT to get the "best" wavefunction by sampling the DJ ensembles, while at least two Opt. IT's are required when sampling the DJB ensembles (see Table III.A.1.6-7, where the ΔE_{av} and $\Delta Var(E_{av})$ are still big).

A strong argument for the previous statement is made by comparing the block means of two different wavefunctions obtained by optimizations using DJ and DJB ensembles, respectively. The first function is Ψ_{1b} (Table 5), whose "dominant" parameters are the means of that in Ψ_{1b1} and Ψ_{1b2}. The second is Ψ_{1a2} (Table 2), whose "dominant" parameters are obtained by extrapolating to $\tau = 0$ the values (in $\Psi_{1a2\text{-}1}$, $\Psi_{1a2\text{-}2}$, and $\Psi_{1a2\text{-}3}$, see Table III.A.1.8) obtained at the three timesteps in the 2nd Opt. IT. of last section. The results are tabulated in Table III.A.2.4, from which we find that Ψ_{1b} gives a much lower block energy and smaller variance than Ψ_{1a2}.

Summarizing the discussion above, one may conclude that **a partial optimization can be performed by sampling the DJ ensembles.**

Now, we are interested in improving further the efficiency of the optimization procedure. One possible way is to optimize fewer parameters and still get a wavefunction as good as those above, if some "dominant" parameters are related.

An attempt was made to optimize two, three, four and five of the

Table III. A. 2. 1. Iteration means[a] in the optimizations of III.A.2.

Optimization	$E_{av,ini}$	$E_{av,fin}$	$Var(E_{av})_{ini}$	$Var(E_{av})_{fin}$
$\Psi_{II} \rightarrow \Psi_{1b1}$[b]	-7.873 ± .030	-8.034 ± .010	.310 ± .012	.120 ± .012
$\Psi_{1b1} \rightarrow \Psi_{1b2}$[c]	-8.015 ± .009	-8.012 ± .010	.127 ±.005	.121 ± .004
$\Psi_{II} \rightarrow \Psi_{1b3}$[d]	-7.873 ± .030	-8.037 ± .009	.310 ± .012	.123 ± .013

[a]These values are the averages taken from the six runs.
[b,c]The results of the optimization of all the seven "dominant" parameters.
[d]The results of the optimization of the five "dominant" parameters.

Table III. A. 2. 2. Changes of the iteration means in the optimizations of III.A.2.

Optimization	ΔE_{av}[a]	$\Delta Var(E_{av})$[b]
$\Psi_{II} \rightarrow \Psi_{1b1}$	-0.161 ± 0.032	-0.298 ± 0.017
$\Psi_{1b1} \rightarrow \Psi_{1b2}$	0.003 ± 0.005	-0.006 ± 0.006
$\Psi_{II} \rightarrow \Psi_{1b3}$	-0.164 ± 0.031	-0.187 ± 0.018

[a,b]The changes of the individual iteration means in the optimization, which are obtained directly from Table III.A.2.1.

Table III. A. 2. 3. "Best" "dominant" parameters[a] in III.A.2.

Ψ_T[b]	ζ_1	ζ_3	ζ_4	ζ_6	ζ_8	Jast a	Jast b
Ψ_{1b1}	2.610 ± 0.016	.990 ± .044	1.080 ± 0.039	.951 ± .014	1.791 ± 0.231	.557 ± .007	.965 ± .031
Ψ_{1b2}	2.620 ± 0.008	.895 ± .025	.990 ± .036	.933 ± .015	1.819 ± 0.104	.567 ± .002	.980 ± .043
Ψ_{1b3}	2.611 ± 0.016	1.003 ± 0.042	1.068 ± 0.038	.952 ± .012			0.852 ± 0.023

[a]These values are the averages taken from the six runs.
[b]The trial wavefunctions whose "dominant" parameters have been fully or partially optimized.

Table. III. A. 2. 4. Block means of the five trial wavefunctions.

Mean[a]	Reynld's Ψ_{II}	Ψ_{1b1}	Ψ_{1b2}	Ψ_{1b}	Ψ_{1a2}
E_b	- 7.888 ±.020	-8.022 ±.002	-8.023 ±.003	-8.050 ±.001	-8.037 ±.001
$Var_b(E_{av})$	.315 ±.007	.126 ±.002	.124 ±.003	.110 ±.001	.113 ±.001

[a]The block means at $\tau = 0.06$ are taken from 10 blocks, and there are 300 iterations in a block, and 600 configurations in an ensemble.

"dominant" parameters with different groups. Most results were poor, except for those given by the wavefunction Ψ_{1b3}. Ψ_{1b3} was obtained by optimizing only five parameters, ζ_1, ζ_3, ζ_4, ζ_6, and Jastrow **b**, and starting with Ψ_{II} (see also Fig. III.A.2.1). Table III.A.2.1 shows Ψ_{1b3} is comparable to Ψ_{1b1} and Ψ_{1b2}, giving almost the same individual iteration average energy as well as variances. These five parameters are hence considered **dominant.**

In the final part of this section, the exploration of how the iteration (after optimization) and block means (in simulation) behave as the Opt. IT number is increased, where five "dominant" parameters, ζ_1, ζ_3, ζ_4, ζ_6, ζ_8, were optimized (see the logic in Fig. III.A.2.2). In the simulation procedure there are 600 configurations in an ensemble (DJ), but only 40 iterations in a block. The optimization started with Ψ_{1c} (Table 6) which was obtained in the optimization described in the next section. The data collected in Table III.A.2.5-6 is that starting from the 5th Opt. IT where the optimized trial wavefunction has converged, so that the first Opt. IT which appears in the graphs and tables is actually the 5th.

It is found from Table III.A.2.5 that **the simulation results** (the means of grand means of local energies and the iteration variances) **agree with that of the optimization** (the means of iteration means). This is an important conclusion, which suggests a way to solve the so-called "ensemble dependency problem", to be discussed later.

Fig. III.A.2.3 - 9 show how the energies, variances and the "dominant" parameters **fluctuate** with Opt. IT. As Opt. IT increases, the grand and

Table III. A. 2. 5. Grand means and iteration means vs. Opt. IT.

Opt. IT[a]	E_{gm}[b]	E_{av}[c]	$Var_{gm}(E_{av})$[d]	$Var(E_{av})$[e]
1	-8.063	-8.064	0.115	0.114
2	-8.025	-8.033	0.119	0.115
3	-8.047	-8.044	0.112	0.114
4	-8.045	-8.047	0.110	0.112
5	-8.059	-8.056	0.117	0.116
6	-8.028	-8.039	0.104	0.105
7	-8.032	-8.031	0.109	0.105
8	-8.058	-8.062	0.120	0.109
9	-8.034	-8.032	0.106	0.111
10	-8.024	-8.026	0.112	0.112
11	-8.031	-8.044	0.103	0.118
12	-8.016	-8.030	0.089	0.118
13	-8.010	-8.006	0.109	0.116
14	-8.014	-8.032	0.114	0.115
15	-8.023	-8.005	0.126	0.117
16	-8.045	-8.019	0.101	0.115
Mean	-8.035 ± 0.004	-8.036 ± 0.004	0.110 ± 0.002	0.113 ± 0.001

[a]The procedure started with the trial wavefunction Ψ_{1c} (see Table 6). In each Opt. IT, there are only 40 iterations in a block and 600 configurations in an ensemble.
[b]The grand means of the local energies after simulation.
[c]The individual iteration means of the local energies after optimization.
[d]The grand means of the individual iteration variances after simulation.
[c]The individual iteration means of the local energy variances after optimization.

Table III. A. 2. 6. Five "dominant" ζ's vs. Opt. IT.

Opt. IT	ζ_1	ζ_3	ζ_4	ζ_6	ζ_8
1	2.604	0.851	0.987	0.948	1.558
2	2.583	0.820	0.956	0.975	1.561
3	2.596	0.818	0.780	0.964	1.418
4	2.600	0.939	0.863	0.939	1.883
5	2.602	0.819	0.740	0.927	1.824
6	2.616	0.918	0.842	0.955	1.751
7	2.591	0.769	0.731	0.952	1.860
8	2.596	0.851	0.811	1.005	1.536
9	2.608	0.845	0.790	0.962	1.915
10	2.603	0.808	0.961	0.957	1.809
11	2.614	0.826	1.076	0.982	1.474
12	2.643	0.848	0.898	0.931	1.383
13	2.610	0.898	0.848	0.931	1.555
14	2.622	1.025	1.012	0.952	1.867
15	2.609	0.879	0.760	0.981	1.582
16	2.599	0.870	0.830	0.968	1.671
Mean	2.606 ± 0.004	0.861 ± 0.015	0.868 ± 0.026	0.958 ± 0.005	1.665 ± 0.045

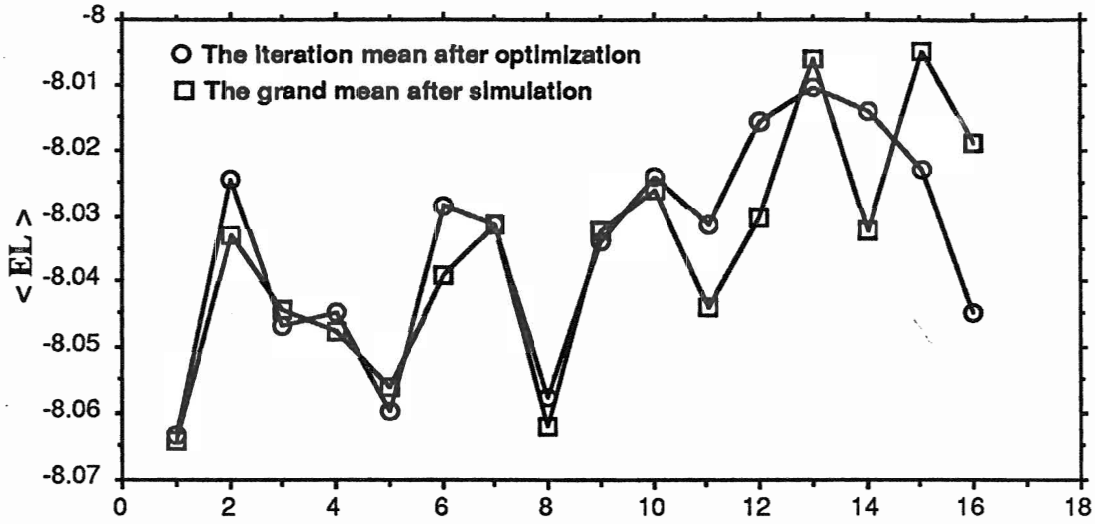


Fig. III. A. 2. 3. Iteration mean after optimization and the grand mean in simulation of the local energy vs. Opt. IT

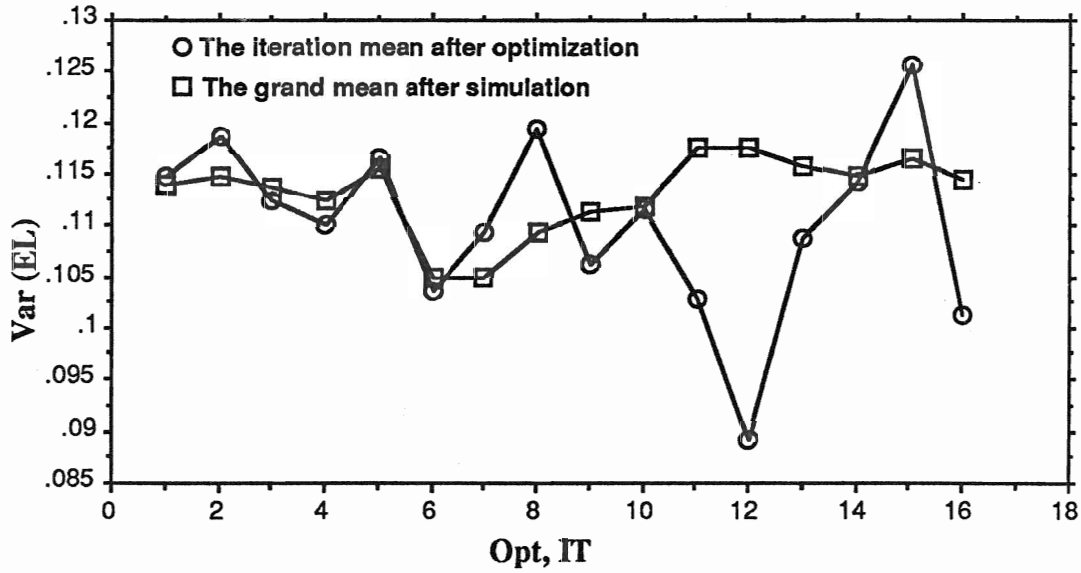


Fig. III. A. 2. 4. Iteration mean after optimization and grand mean in simulation of the iteration variance vs. Opt. IT

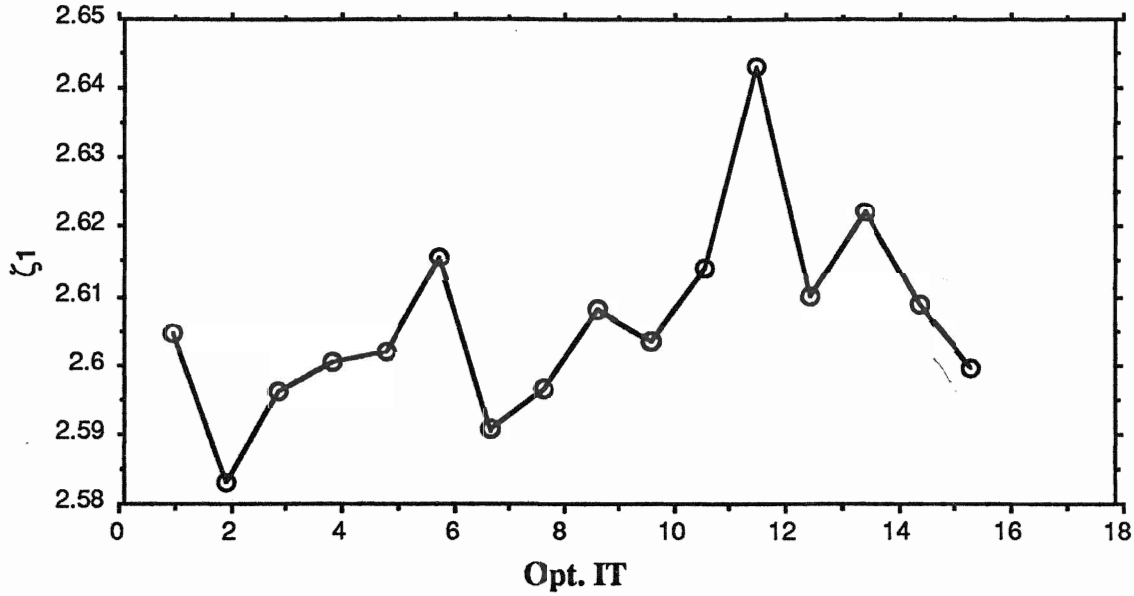


Fig. III. A. 2. 5. ζ1 vs. Opt. IT

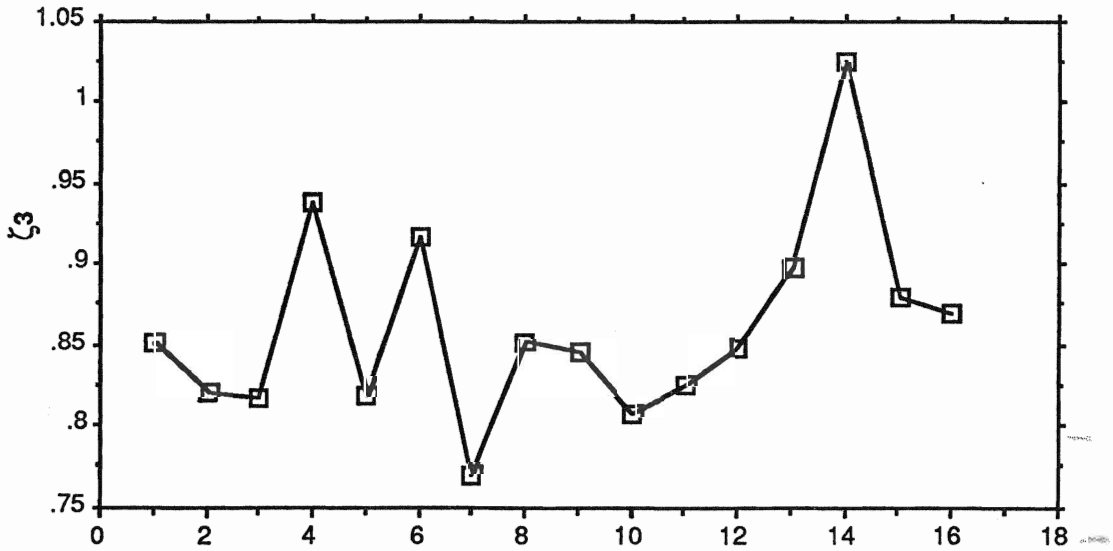


Fig. III.C. A. 2. 6. ζ3 vs. Opt. IT

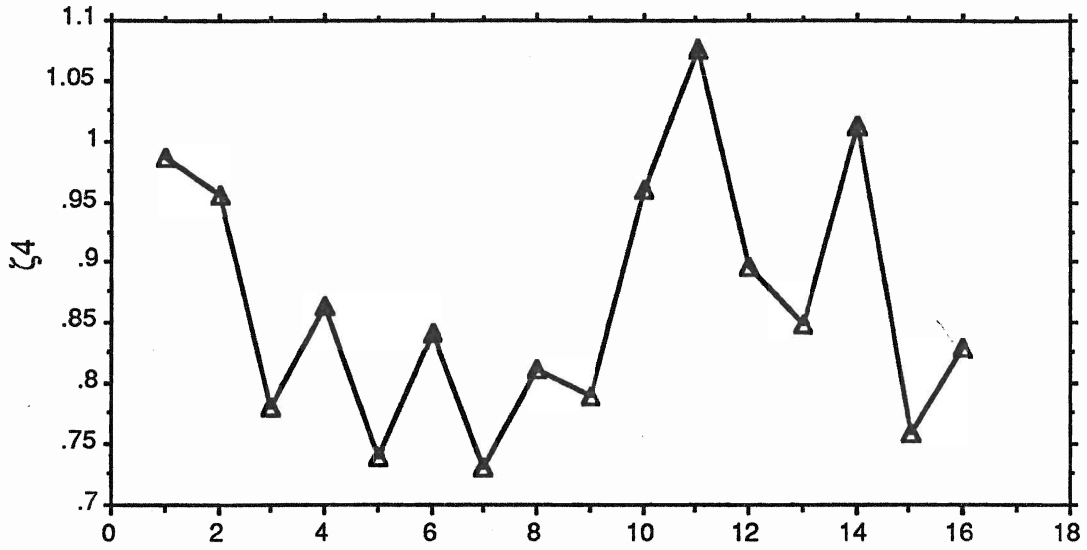


Fig. III. A. 2. 7. ζ4 vs. Opt. IT

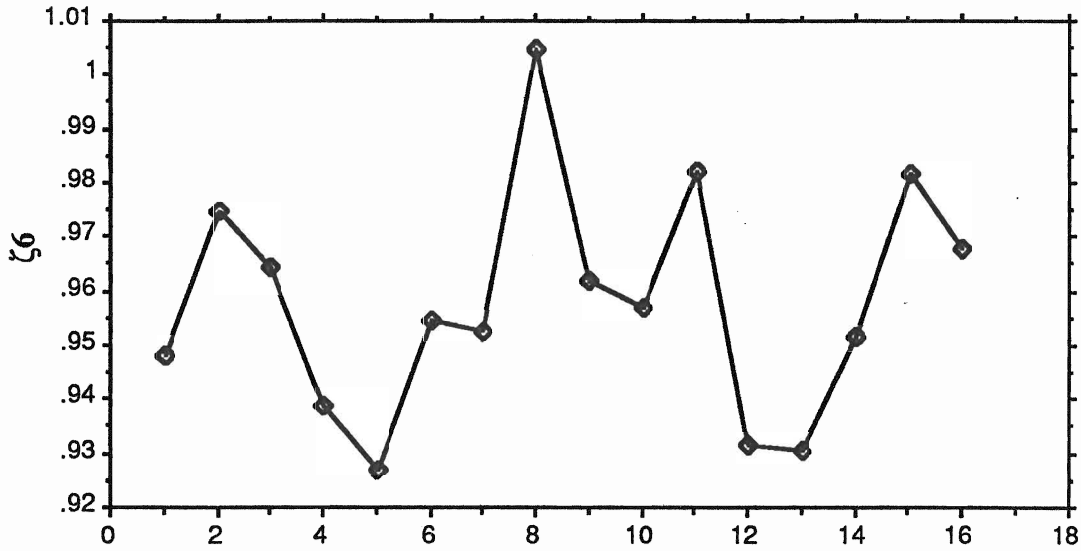


Fig. III. A. 2. 8. ζ6 vs. Opt. IT

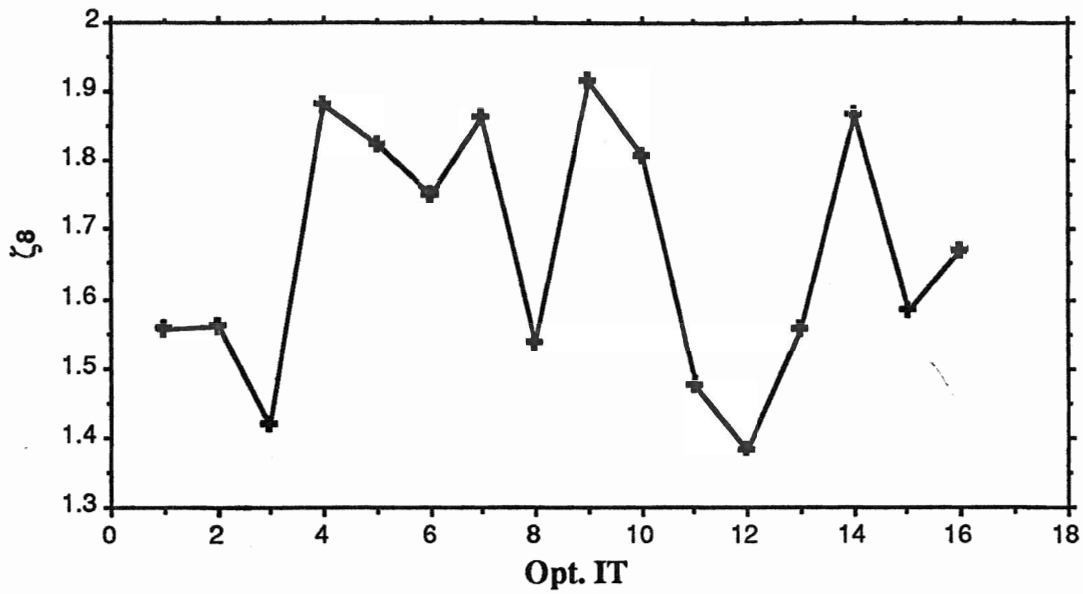


Fig. III. A. 2. 9. ζ8 vs. Opt. IT

iteration means of the energy become higher and higher. This is because the algorithm used is:

$$SSQ\ (P_{j,\ j=1,J}) = \sum_{m=1,Mc} [E_L(\mathbf{R}^m) - E_{av}(\tau)]^2$$

and whose minimum, which is different from that of $\sum_m [E_L(\mathbf{R}^m) - E_g]^2$, will change as $\{P_j\}$ **due to the change of** $\mathbf{E_{av}(\tau)}$. In other words, there are many possible ways of producing the same value for $\sum_m [E_L(\mathbf{R}^m) - E_{av}(\tau)]^2$, but having different values of $E_{av}(\tau)$. Thus, the more Opt. IT's are done, the further removed the $E_{av}(\tau)$ is from that of E_g, even though the variance (SSQ/M_c) fluctuates around a constant (see Fig. III.A.2.4). This is the so-called **wrong minimum** problem and will be encountered again later.

3. Optimization of the split Jastrow constants, geometry and all the exponents , sampling from the DJ ensembles at $\tau = 0.06$.

Considering that there are two different kinds of correlations between two electrons in a system, i.e. a Fermi correlation, when the two correlated electrons are in parallel-spin states and a Coulomb correlation, when they are in antiparallel-spin states, the Jastrow constants **a** are split into parallel $\mathbf{a}_p$ and antiparallel $\mathbf{a}_a$, **b** into parallel $\mathbf{b}_p$ and antiparallel $\mathbf{b}_a$. They are optimized in this section. On the other hand, the distance $r_{Li\text{-}H}$ between the two nuclei (Li and H) is an important factor influencing the nature of the wavefunction, and it is of interest to optimize it. When optimizing the geometry $r_{Li\text{-}H}$ Eq.II.C - 3 was used instead of Eq. II.C-4 in order to force the optimization go along the "correct" path that leads to the smallest variance as well as the lowest energy.

The split Jastrow constants were first optimized with the "non-dominant" exponents, ζ_2, ζ_5, ζ_7, starting with Ψ_{II}, to see if a better function than Ψ_{1b1} could be obtained. Next, starting with the optimized wavefunction Ψ_{1b}, only the split constants and the geometry were independently optimized, in order to see the individual roles they play in the wavefunction, and then to see what kind of improvement in these parameters could be made. Furthermore, optimization was done on the geometry together with all the "dominant" parameters, and then the geometry and the split constants together with all the "dominant" exponents. Finally, beginning with an altered version of the trial wavefunction Ψ_{1b} (the alteration consisted of a random fluctuation in several of the parameters), all the exponents were optimized, as well as the split Jastrow constant and the geometry, to see if the

energies and variance would converge to a lower minimum. This whole procedure is summarized in Fig. III.A.3.1 and 2, and the results are shown in Table. III. A. 3.1 - 7, respectively .

The first entries in Table III. A.3.1 and 2 show that by optimizing the split Jastrow constants with the "non-dominant" exponents there is very little change in E_{av}. Although the change in $Var(E_{av})$ is significant, the iteration variance given by Ψ_{1c1} obtained is much worse than that of Ψ_{1b1}(see Table III. A.2.1). Therefore one can assume that the "non-dominant" exponents do not play a significant role in the trial wavefunction.

See Table III. A.3.1 and 2. The optimization of the split constants only did not result in a significant improvement. The same conclusion was obtained by optimizing the geometry only (these results are not tabulated). In any case, significant improvements were achieved by the optimization of $r_{Li\text{-}H}$ together with all the "dominant" parameters (the results for the change in iteration energy are not tabulated). Also, improvement from the optimization of the geometry and the split constants together with all the "dominant" exponents were achieved.

One finds from Table III.A.3.3 - 4 and 6 - 7 that the standard errors for a_p and b_p are rather large, which means that they are unimportant because their great fluctuations don't significantly affect the results (this is expected since the Fermi correlation is handled by the Slater determinant). Therefore **we can set them equal to zero, but since splitting the a and b does not bring about a significantly better wavefunction, we can use their single forms.**

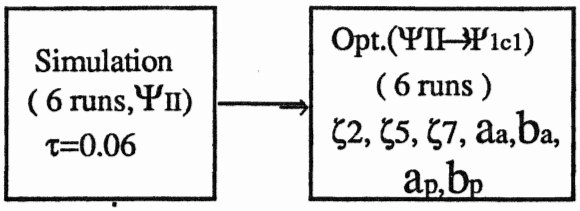


Fig. III.A.3.1. Logic routes of the optimizations of the split Jastrow constants and the "non-dominant" parameters.

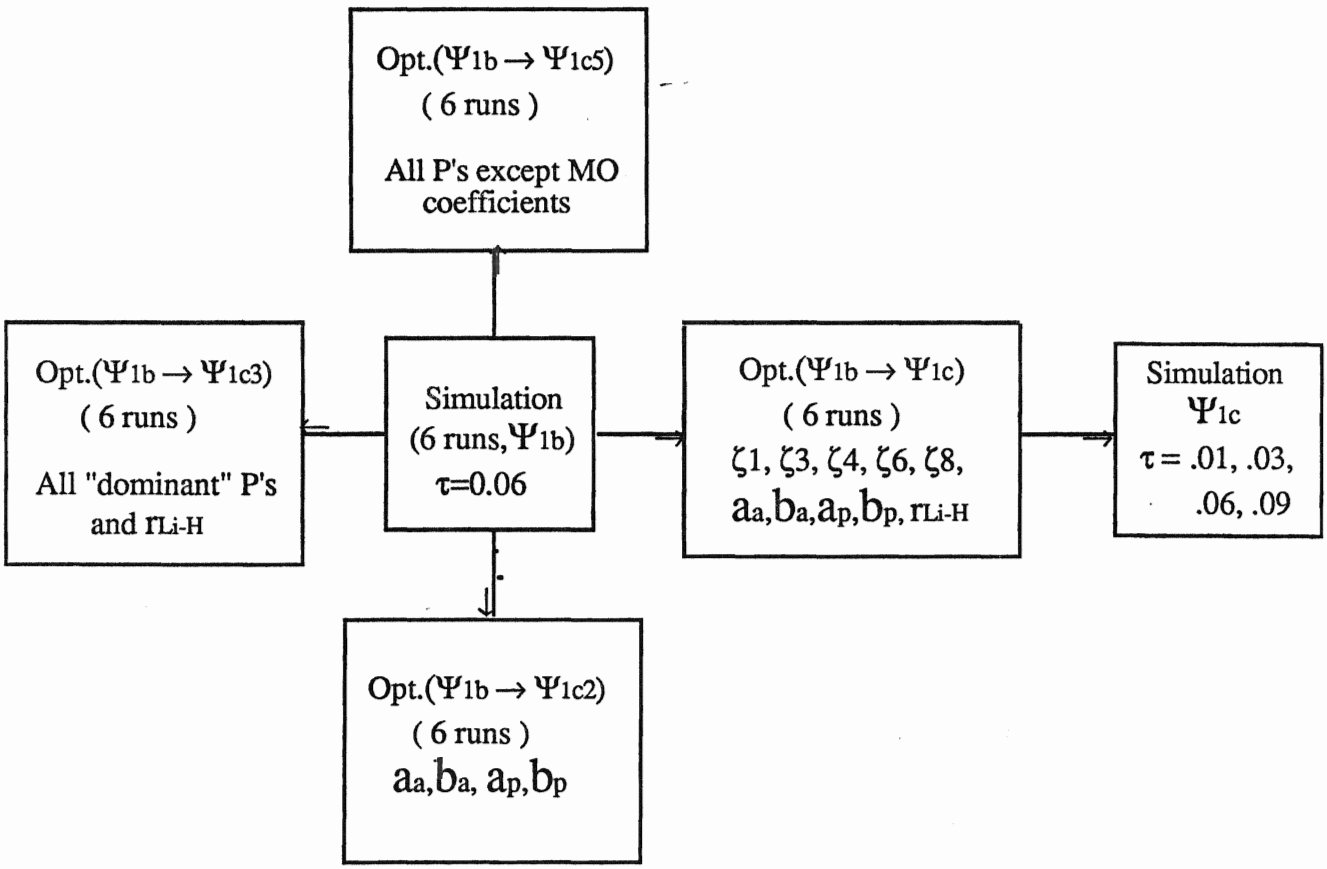


Fig. III.A.3.2. Logic routes of the optimizations of the split Jastrow constants, geometry and all exponents.

Table III. A. 3. 1. Individual means[a] in the optimizations of III.A.3.

Optimization	$E_{av,ini}$	$E_{av,fin}$	$Var(E_{av})_{ini}$	$Var(E_{av})_{fin}$
$\Psi_{II} \rightarrow \Psi_{1c1}$	-7.873 ± .030	-7.938 ± .014	.310 ± .012	.163 ± .007
$\Psi_{1b} \rightarrow \Psi_{1c2}$	-8.047 ± .008	-8.052 ± .008	.126 ±.009	.116 ± .009
$\Psi_{1b} \rightarrow \Psi_{1c3}$			.138 ± .020	.108 ± .006
$\Psi_{1b} \rightarrow \Psi_{1c(1c4)}$	-8.012 ± .009	-8.034 ± .008	.126 ±.005	.107 ± .007
$\Psi_{1b} \rightarrow \Psi_{1c5}$	-7.976 ± .036	-8.101 ± .042	.628 ± .119	.096 ± .008

[a]These values are the averages taken from the six runs except that $\Psi_{1b} \rightarrow \Psi_{1c5}$ where only four runs have converged.

Table III. A. 3. 2. Changes[a] of the iteration means in the optimizations of of III.A.2.

Optimization	ΔE_{av}	$\Delta Var(E_{av})$
$\Psi_{II} \rightarrow \Psi_{1c1}$	-0.065 ± 0.033	-0.147 ± 0.014
$\Psi_{1b} \rightarrow \Psi_{1c2}$	-0.005 ± 0.011	-0.010 ± 0.016
$\Psi_{1b} \rightarrow \Psi_{1c3}$		-0.030 ± 0.021
$\Psi_{1b} \rightarrow \Psi_{1c(1c4)}$	-0.022 ±0.012	-0.019 ± .0.009
$\Psi_{1b} \rightarrow \Psi_{1c5}$	-0.125 ± 0.055	-0.532 ± 0.119

[a]The changes in the individual iteration means in the optimization, which were obtained directly from Table III.A.3.1.

Table III. A. 3. 3. "Best" split Jastrow constants and the "non-dominant" parameters.

Ψ_T	ζ_2	ζ_5	ζ_7	a_p	b_p	a_a	b_a
Ψ_{1c1}	4.828 ± 0.029	1.481 ± 0.292	1.590 ± 0.037	1.510 ± 0.631	4.051 ± 0.584	0.622 ± 0.037	1.158 ± 0.055

[a]These values are the averages taken from the six runs, as in Table III.A.3.4 - 6. The values in Table III.A.3.7 are taken only from the four converged runs.

Table III. A. 3. 4. "Best" split Jastrow constants.

Ψ_T	a_p	b_p	a_a	b_a
Ψ_{1c2}	2.558 ± .865	3.799 ± .860	.519 ± .007	.818 ± .012

Table III. A. 3. 5. "Best"Jastrow constants,the "dominant" parameters and $r_{Li\text{-}H}$.

Ψ_T	ζ_1	ζ_3	ζ_4	ζ_6	ζ_8	Jast a	Jast b	$r_{Li\text{-}H}$
Ψ_{1c3}	2.595 ± 0.004	0.730 ± 0.038	1.041 ± 0.046	0.913 ± 0.005	1.728 ± 0.042	0.551 ± 0.010	0.996 ± 0.022	1.498 ± 0.009

Table III. A. 3. 6. "Best" r_{Li-H}, split Jastrow constants and "dominant" parameters .

Ψ_T	ζ_1	ζ_3	ζ_4	ζ_6	ζ_8	a_p	b_p	a_a	b_a	r_{Li-H}
Ψ_{1c}	2.644 ± 0.008	.787 ± .062	0.954 ± 0.074	.937 ± .012	1.764 ± 0.099	2.149 ± 0.874	4.843 ± 0.537	.517 ± .013	.782 ± .035	1.536 ± 0.018

Table. III. A. 3. 7. "Best" parameters excepting MO coefficients.

Ψ_T	ζ_1	ζ_2	ζ_3	ζ_4	ζ_5	ζ_6	ζ_7	ζ_8	a_p	b_p	a_a	b_a	r_{Li-H}
Ψ_{1c5}	2.704 ± 0.022	4.450 ± .030	.865 ± .068	1.190 ± 0.103	8.518 ± 0.176	1.103 ± 0.111	0.937 ± 0.118	2.006 ± 0.302	3.491 ± 0.823	4.846 ± 0.176	.492 ± .009	.645 ± .034	1.448 ± 0.119

The case for r_{Li-H} is different. It was observed that during the optimization, although the change of r_{Li-H} did not cause much change in the energies and variances, **a small change in geometry caused a big change in the other parameters when they are all optimized together, resulting in lower energies and smaller variances;** however, there is still no apparent way at this point to force the estimated geometry to fluctuate about the "correct" one (the experimental value is 3.014 a.u.)[1], and the optimization causes it to go further away from that value (See Table. III. A.3.5 - 7).

Finally, the 13 parameters (8 exponents, 4 split Jastrow constant and the geometry) were optimized, along with all parameters except the linear MO coefficients. We began with an alternate form of Ψ_{1b}, where the only changes in parameters are in the Jastrow constant: $\mathbf{a_p}=\mathbf{b_p}=0.562$, $\mathbf{a_a}=\mathbf{b_a}=0.972$. Although only two parameter values ($\mathbf{b_p}$ and $\mathbf{a_a}$) have actually been changed, and among them $\mathbf{a_a}$ is unimportant, the optimizations were carried out with some difficulty. These took much more CPU time than before, with only four out of six runs converging, and with not only $\mathbf{a_p}$ and $\mathbf{b_p}$ fluctuating greatly but also r_{Li-H} and other parameters. These difficulties were even greater and caused some runs to render nonsensical values when more initial parameters were "fluctuated". This is the so-called **linear dependency** problem and will be discussed in the next section. The last lines of Table III.A.3.1-2 and Table III.A.3.7. show the results of the averages taken from the four converged runs.

In my experience, if one optimizes no more than ten parameters (5

"dominant" parameters, plus a_a, b_a, a_p, b_p, and $r_{Li\text{-}H}$, in terms of the double - ζ set) at once, one can usually obtain a reasonable set of results and the optimizations can proceed very smoothly no matter how one fluctuates the initial parameters. More than ten such runs were performed and comparably good wavefunctions were obtained. If one intends to optimize more parameters, including the MO coefficients, one must solve the linear dependency problem, which is one of the topics of the next section.

Before finishing this preliminary investigation, a test was performed to determine how much improvement had been achieved at this point. The best trial wavefunction Ψ_{1c} obtained in this section was chosen to do the VQMC calculation. The results are shown in Table III.A.3.8 and the statistical regression curves are plotted in Fig.III.A.3 - 4. A similar calculation with Reynold's Ψ_{II} was done for comparison (See Table II.A.3.9. and Fig.III.A.5-6). An outstanding result was obtained for the variational energy, and this result is: **-8.048 ± .006 a.u** and the **individual iteration variance is: .126 ± .005**, which are not only much better than that of Reynold's, but it is also the best value for the variational calculation reported so far in any currently available literature. Further efforts are being made to improve on even this result.

Table III. A. 3. 8. VQMC estimates with Ψ_{1c}.

τ	$E_b(\tau)$[a]	$Var_b(E_{av}(\tau))$
.01	-8.040 ± .0025	.128 ± .002
.03	-8.030 ± .0018	.127 ± .002
.06	-8.023 ± .0017	.122 ± .002
.09	-8.017 ± .0028	.130 ± .002

Extrapolated[b] to τ=0: E_{av} = -8.048 ± .0058, $Var(E_{av}(\tau=0))$ = .126 ± .0052

[a]There are 10 blocks at each τ, 300 iterations in a block, and 600 configurations in the ensemble.
[b]The regression model is cubic. See Fig.III.A.3.3-4.

Table III. A. 3. 9. VQMC estimates with Ψ_{II}.

τ	$E_b(\tau)$[c]	$Var_b(E_{av}(\tau))$
.01	-7.965 ± .0018	.287 ± .0010
.03	-7.912 ± .0015	.278 ± .0008
.06	-7.847 ± .0011	.271 ± .0004
.09	-7.787 ± .0015	.267 ± .0008

Extrapolated[d] to τ=0: E_{av} = -7.991 ± .0025, $Var(E_{av}(\tau=0))$ = .293 ± .0013

[c]There are 10 blocks at each τ, 300 iterations in a block, and 600 configurations in the ensemble.
[d]The regression model for the energy is quadratic and cubic for the variance. See Fig.III.A.3.5-6.

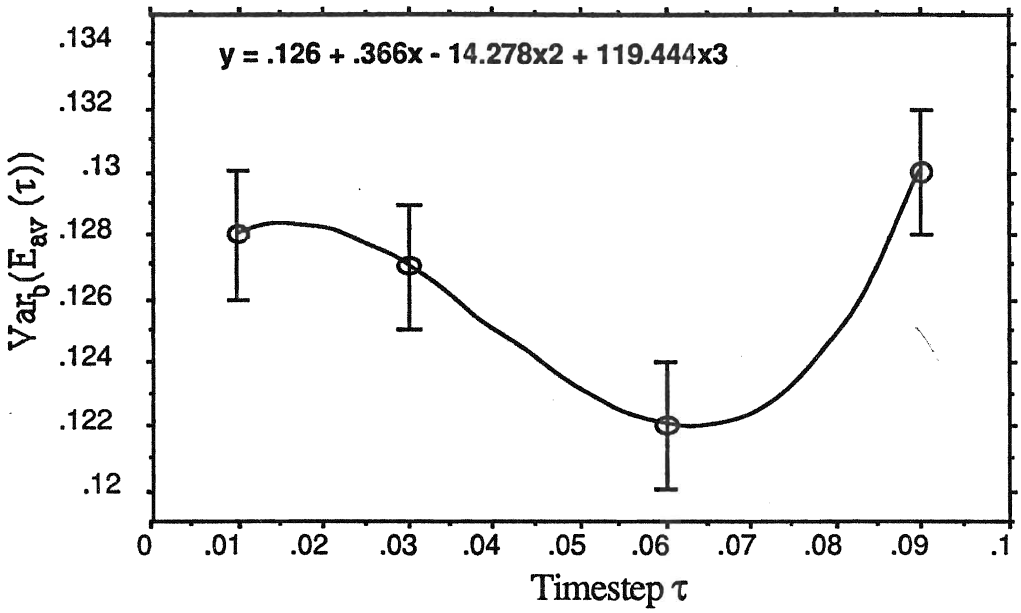


Fig. III. A. 3. 3. VQMC estimate of $Var(E_{av}(\tau \rightarrow 0)$ with Ψ_{1c}.

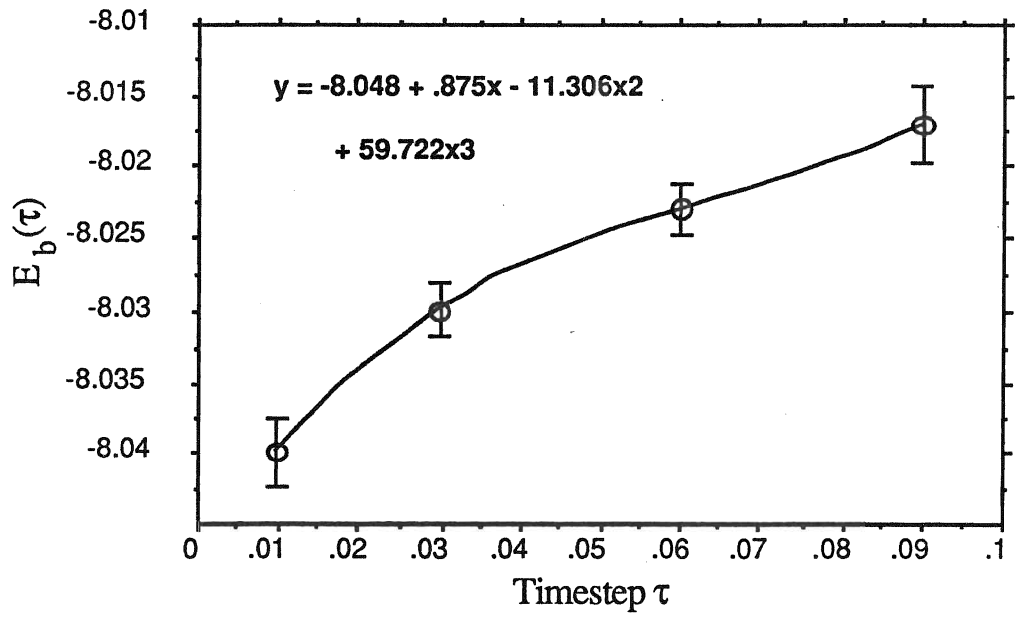


Fig. III. A. 3. 4. VQMC estimate of E_{var} with Ψ_{1c}.

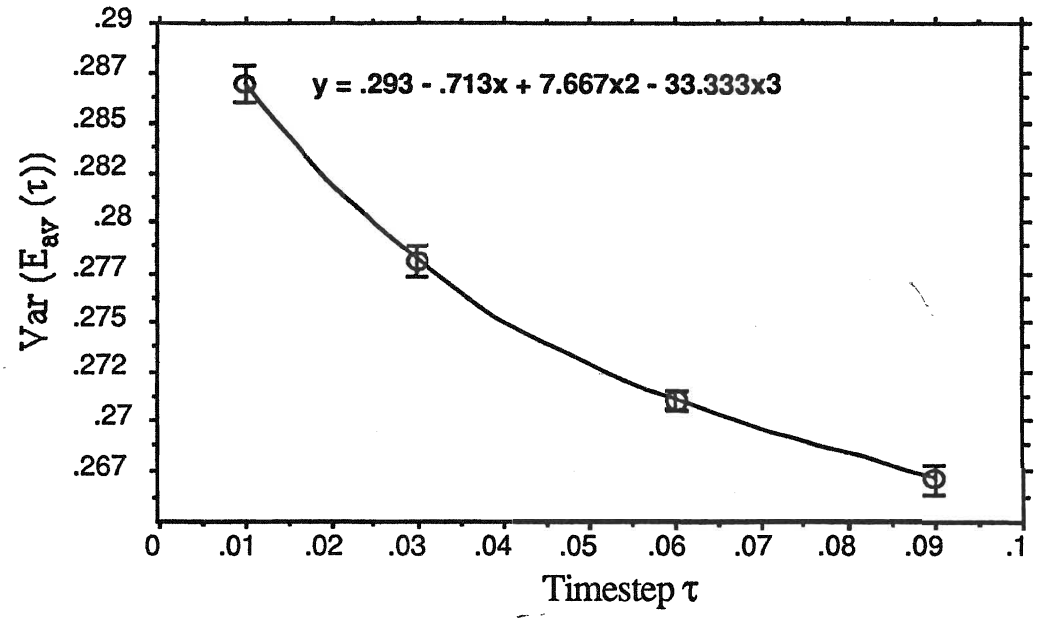


Fig. III. A. 3. 5. VQMC estimate of $Var(E_{av}(\tau \rightarrow 0)$ with Ψ_{II}.

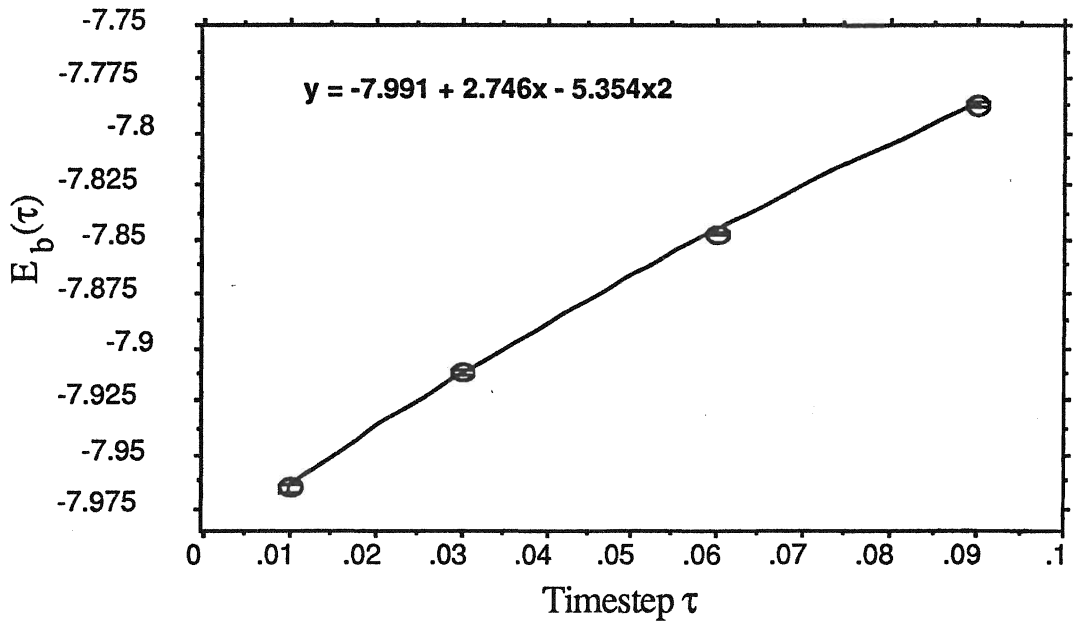


Fig. III. A. 3. 6. VQMC estimate of E_{var} with Ψ_{II}.

B. Investigation by Full Optimization

During an optimization, making full use of the flexibility of the trial wavefunction is strongly recommended from a mathematical viewpoint. The optimization of all parameters, exponents, linear MO coefficients and Jastrow constants for a wavefunction was performed. Because of the large change in the wavefunction during the optimization process, this was a task of great difficulty. The problems encountered become progressively more serious as the optimization was carried forward with progressively more complicated trial wavefunctions. These problems were resolved, and an available method was developed, which is summarized at the end of this section.

To perform a full optimization, one must first solve the linear dependency difficulties inherent in optimizing: (1) the parameters in the functions with the same mathematical forms (e.g. the double-ζ set), and; (2) the linear dependency of the MO coefficients.

In a double-ζ set, the ζ's and the coefficients appear in a duplicated fashion in the functions $1s_{Li}$, $2p_{Li}$ and $1s_H$. Each pair of these parameters has the potential to "exchange" their values while being optimized. Hence they do not always converge within a reasonable amount of CPU time. In order to keep this from happening, all the parameters were divided into five groups so that the pairs were separated and put in different groups for optimization. Then logical loops were traversed as shown below, where the parameters in the box are those being optimized (note that the geometry $r_{Li\text{-}H}$ is always fixed, for reasons mentioned in the previous section), i.e.

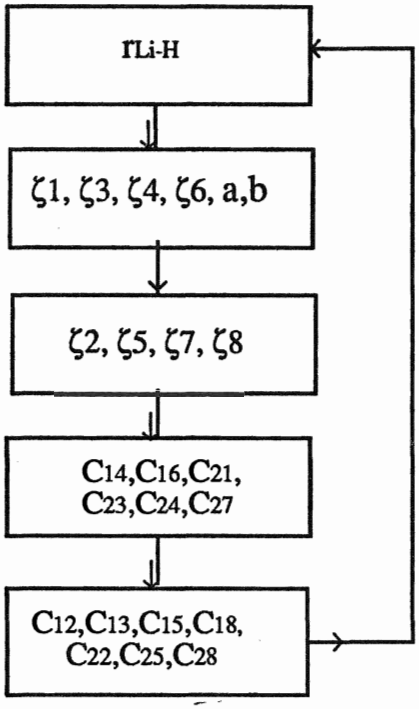


Fig. III.B. Logic route of the loops in the full optimization of a double-ζ set.

The difficulty in optimizing the linear MO coefficients arises out of the nonnormalized trial wavefunction Ψ_T. The multiplication of the whole set of coefficients by any constant k does not change the value of the local energy. Thus, there should be an infinite number of sets of coefficients $\{kC_{ij}\}$'s which satisfy the minimum of SSQ, if it can be reached. The fluctuations in these coefficients actually make the optimization very difficult to converge, and even though a convergence is obtained, the coefficients may be nonsensical! The only way to get over this problem is by fixing some of the coefficients. What was done for a double-ζ set was the fixing of C_{11}, C_{17} and C_{26}, the biggest and smallest absolute numbers; or C_{11}, C_{21}, C_{16} and C_{26}, the coefficients of the most important AO functions. In practice, there is not much difference between these two operations.

After these problems had been solved, the optimizations were successful when sampling the DJ ensemble with **600** configurations. The optimized wavefunctions using this fixed ensemble gave a very small iteration variance (around 0.07 to 0.08) as well as a very low iteration energy (–8.06 to –8.07). But, unfortunately in the simulations (ensembles now change in time) with these optimized wavefunctions the block means of the variances were 0.15 to 0.18 and that of the energies were –8.00 to –8.02. I attribute this to a "**ensemble dependency**" problem, described in the next paragraph.

Suppose the **600** configurations sampled were to be generated with Ψ_0, and the "best" wavefunction yielded was Ψ_T, and further suppose that the individual iteration average energies and their variances after the optimization were to disagree with that of the block means after the simulation. The most important reason for this disagreement is that the **two sets of results** (optimization and simulation) **are obtained by sampling two different ensembles,** $|\Psi_0|^2$ **and** $|\Psi_T|^2$. If Ψ_0 is close to Ψ_T, the results should also be close (Recall Table III.A.2.5, which shows a Ψ_0 which is approximately equal to Ψ_T, and the means of the grand means after simulation is the same as that of the iteration means after optimization). If Ψ_0 differs from Ψ_T by only a small amount, then if one wishes to use a large ensemble size for optimization, then more of these configurations are likely to reappear in the new ensemble of the optimized wavefunction, and thus **the closer the resemblance of these two ensembles**; however, if Ψ_0 is quite different from Ψ_T, fewer configurations resemble each other, and more serious problems will result. Other techniques are then required to deal with this, something which will be discussed at a later point.

One other reason for the difference in optimization and simulation results is that the SSQ is not only a function of the ensemble but also of {Pj}. When the minimum SSQ was reached, the best set of parameters, $\{Pj\}_{best}$, was obtained. Any change in these parameters, or their regrouping by taking averages over several runs (six runs were done for each full optimization), will unfortunately raise the value of SSQ. In the case of partial optimization this seems to matter little because only a few parameters were being optimized, and also the fluctuations in these parameters were small in the various runs. But now, all parameters except the few which were fixed were optimized and the fluctuations were great. Therefore, averaging the parameters from several runs led to worse results in the simulations.

1. Optimization starting with the Ψ_{II}.

After understanding the influential factors above, a further exploration was done for a double-ζ set with sampling more configurations, and with some linearly dependent parameters fixed; this procedure is outlined in Fig. III.B.1.1. Starting with Reynold's Ψ_{II} and **6000** configurations, with C_{17}, C_{11}, and C_{26} fixed, Eq.(II.C - 4) was employed for the optimizations from the 1st Opt. IT to the 3rd Opt. IT, and Eq.(II.C - 3) for the 3rd Opt. IT, where $E_g = -8.070$. The ensembles used in optimization were generated at τ=0.06. Since taking the average from several runs will lead to a larger SSQ, here only one run was performed in the optimization.

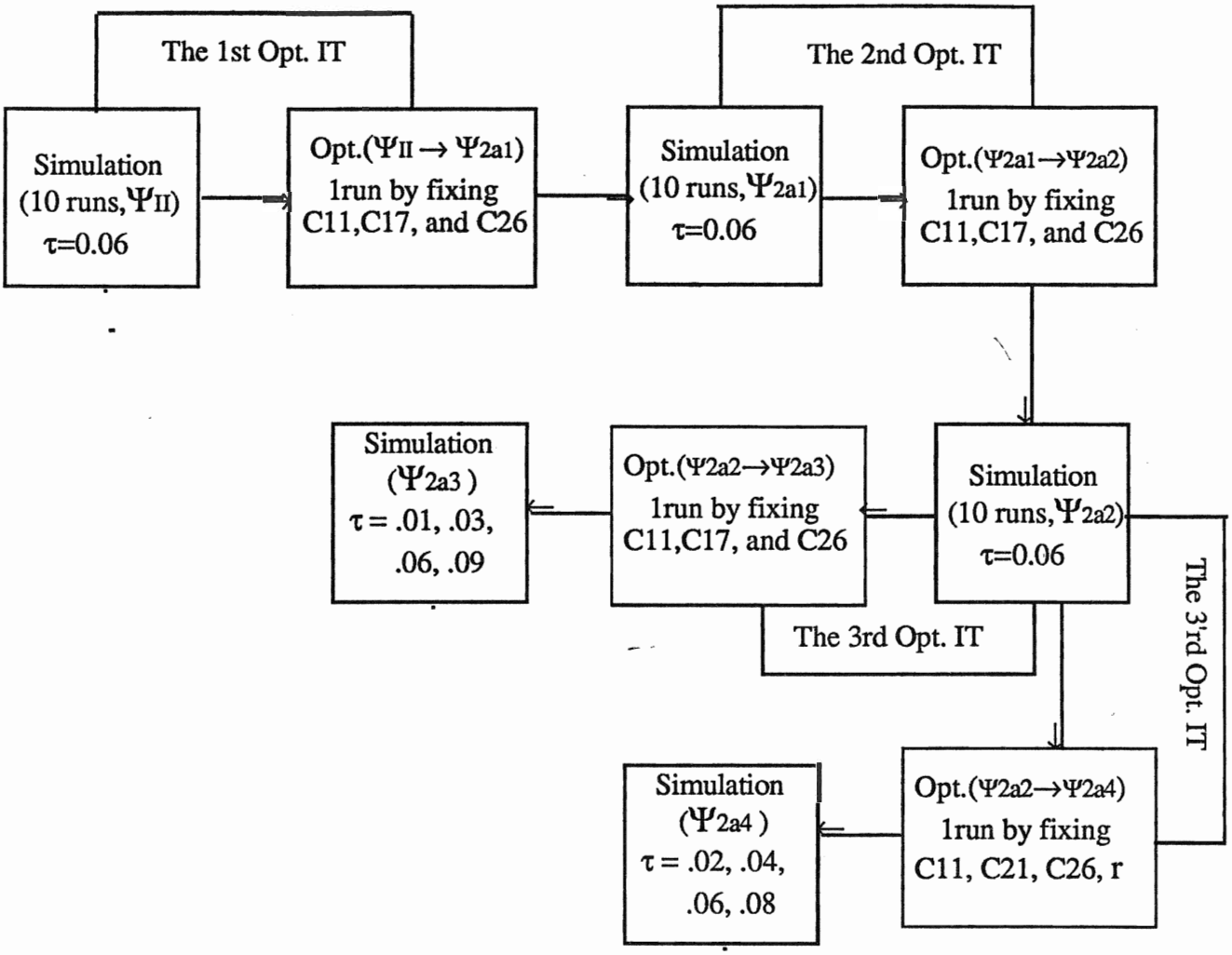


Fig. III.B.1.1. Logic route of the optimizations starting with Ψ_{II}.

The changes in the individual iteration energies and their variances are shown in Table III.B.1.1. Basically consistent with the conclusion derived from the partial optimization, the trial wavefunction converged in the 2nd Opt. IT, where the ΔE_{av} and $\Delta Var(E_{av})$ are small. An additional Opt. IT. with Eq.(II.C - 4) didn't result in a significant improvement, but it did with Eq.(II.C - 3) .

Table III.B.1.2 is a tabulation of the block means computed with different wavefunctions optimized in the 4 Opt. IT's. Comparing this with Table III.B.1.1, one finds that the ensemble dependency problem still exists; the block means of the energy from the simulation are higher than those of the iteration means from the optimization, and the variance is larger when simulating than when optimizing, even when **6000** configurations were used for the optimization.

When plotting these results vs. Opt. IT in Fig. III.B.1.2-3, it is seen that the energy rises as the number of Opt. IT's increases, even though the variance changes little. This situation agrees with that in Fig. III.A.2.3-4 (here there are only 40 iterations in each simulation procedure). This means the optimization with Eq.(II.C - 4) leads to a wrong minimum, and that with Eq.(II.C - 3) forces it back to a correct path. More strong evidence of this fact is shown by comparing the two sets of E_{var}'s and $Var(E_{av}(\tau=0))$ (See Table III.B.1.3-4 and Fig.B.1.4-7). The results of Ψ_{2a4} are much better than that of Ψ_{2a3}. This suggests that **Eq.(II.C - 3) should be used in the full optimization to avoid approaching a wrong minimum**; however, Ψ_{2a4} is still not better than Ψ_{1c}, even if at $\tau=0.06$ Ψ_{2a4} gives a lower energy($-8.045 \pm .001$) and smaller variance($.088 \pm .0009$) than that of Ψ_{1c}

Table III. B. 1. 1. Changes of the iteration means in the optimizations[a] of III.B.1.

Optimization	$E_{av,ini}$	$E_{av,fin}$	$Var(E_{av})_{ini}$	$Var(E_{av})_{fin}$	ΔE_{av}	$\Delta Var(E_{av})$
$\Psi_{II} \rightarrow \Psi_{2a1}$[b]	-7.873	-8.086	0.310	0.086	-0.213	-0.224
$\Psi_{2a1} \rightarrow \Psi_{2a2}$[b]	-8.065	-8.073	0.110	0.103	-0.008	-0.007
$\Psi_{2a2} \rightarrow \Psi_{2a3}$[b]	-8.045	-8.036	0.100	0.097	0.009	-0.003
$\Psi_{2a2} \rightarrow \Psi_{2a4}$[c]	-8.040	-8.068	0.111	0.089	-0.028	-0.022

[a]There is only one optimization run at each Opt. IT, but the ensemble used at each run is made up of 6000 configurations.
[b]The optimizations with the algorithm Eq.(II.C - 4).
[c]The optimization with the algorithm Eq.(II.C - 3), where Eg = -8.070.

Table. III. B. 1. 2. Block means[a] in simulation vs. the **Opt. IT**.

Opt. IT	Ψ_T	E_b	$Var_b(E_{av})$
0	Ψ_{II}	-7.888 ± 0.020	0.315 ± 0.007
1st	Ψ_{2a1}	-8.057 ± 0.002	0.115 ± 0.003
2nd	Ψ_{2a2}	-8.044 ± 0.003	0.102 ± 0.001
3rd	Ψ_{2a3}	-8.024 ± 0.003	0.105 ± 0.001
3'rd	Ψ_{2a4}	-8.045 ± 0.001	0.088 ± 0.001

[a]Each block mean is taken from ten blocks and there are 300 iterations in a block, and 600 configurations in the ensemble. The block means are computed at τ =.06 with the trial wavefunction listed in the same line. See also Fig. III.B.1.2-3.

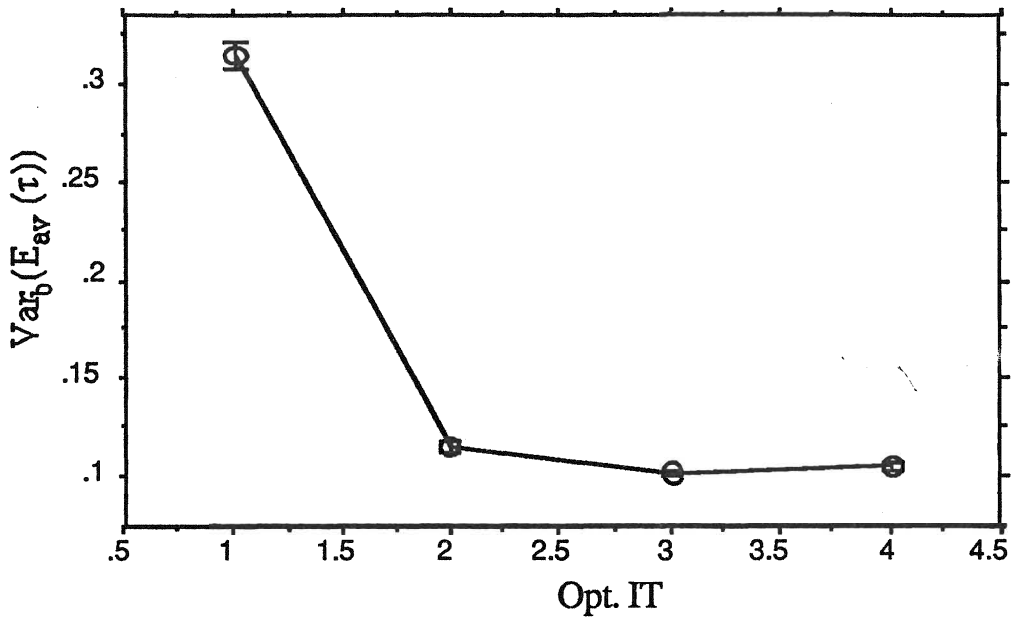


Fig. III.B.1.2. Block mean of the iteration variances vs. Opt. IT[a].

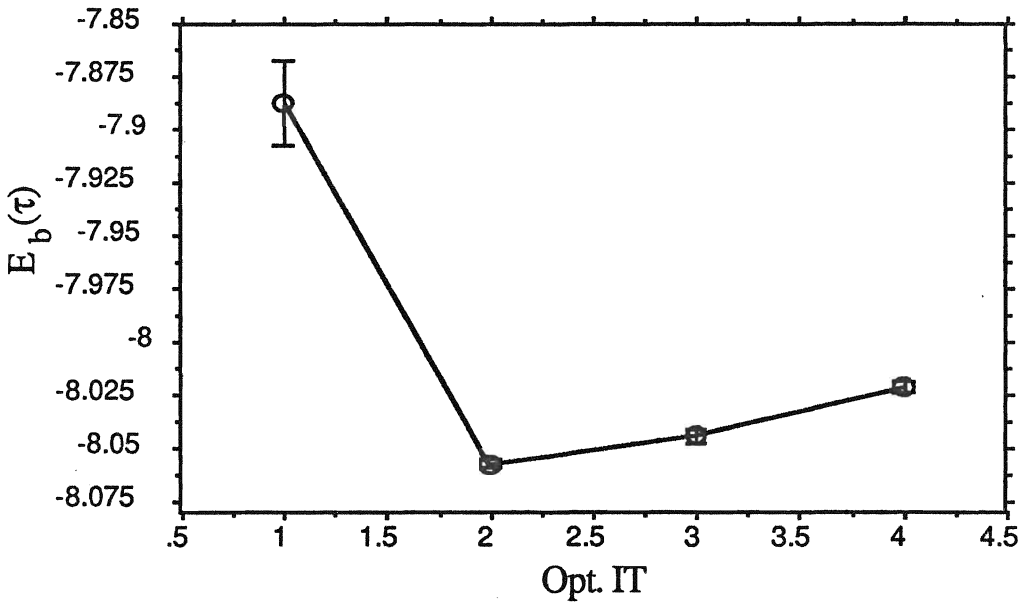


Fig. III.B.1.3. Block mean of the iteration energies vs. Opt. IT[b].

[a,b]Note that, to see the changes of the energy and the variance upon optimization, the Ψ_{II} results are also plotted in the figures. So the Opt. IT numbers there are added by 1.

Table III.B.1.3. VQMC estimates with Ψ_{2a3}.

τ	$E_b(\tau)$[a]	$Var_b(E_{av}(\tau))$
.01	-8.037 ± .002	.104 ± .0019
.03	-8.036 ± .002	.102 ± .0016
.06	-8.024 ± .003	.105 ± .0014
.09	-8.031 ± .002	.106 ± .0015

Extrapolated[b] to τ=0: E_{av} = -8.031 ± .0054, $Var(E_{av})$ = .107 ± .0033

[a]There are 10 blocks at each τ, 300 iterations in a block, and 600 configurations in an ensemble. Similarly for the next table.
[b]The extrapolation regression model is cubic. See Fig. III.B.1.4-5.

Table III.B.1.4. VQMC estimates with Ψ_{2a4}.

τ	$E_b(\tau)$	$Var_b(E_{av}(\tau))$
.01	-8.043±.0010	.094±.0010
.03	-8.045±.0004	.089±.0007
.06	-8.045±.0010	.088±.0009
.09	-8.043±.0004	.086±.0008

Extrapolated[a] to τ=0: E_{av} = -8.042 ± .0015, $Var(E_{av})$ = .099 ± .0023

[a]The regression model is cubic for the variance and quadratic for the energy. See Fig. III.B.1.6-7.

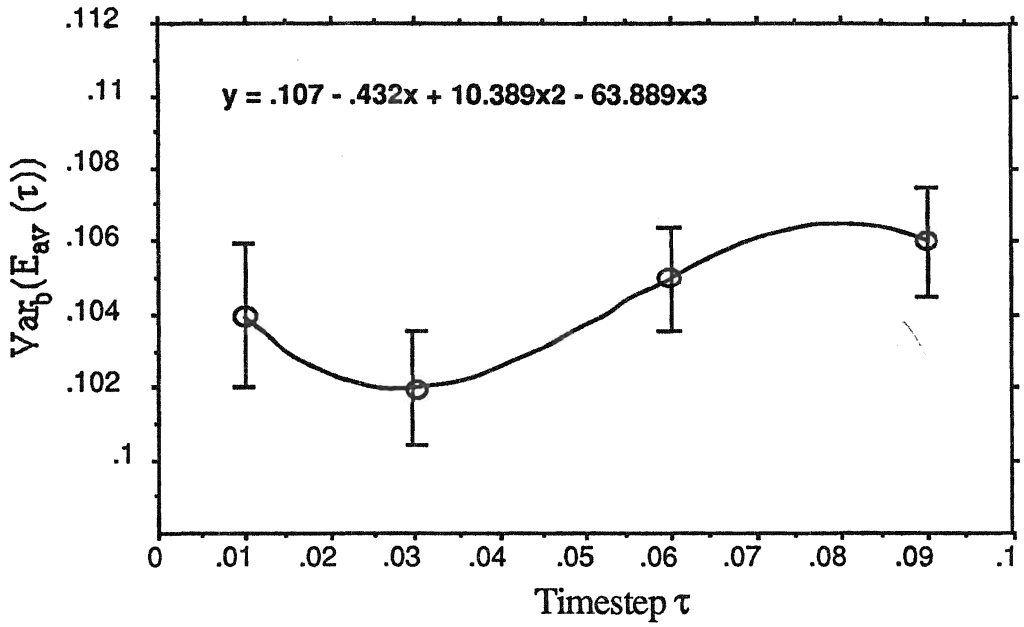


Fig. III.B.1.4. VQMC estimate of $Var(E_{av}(\tau \rightarrow 0)$ with Ψ_{2a3}.

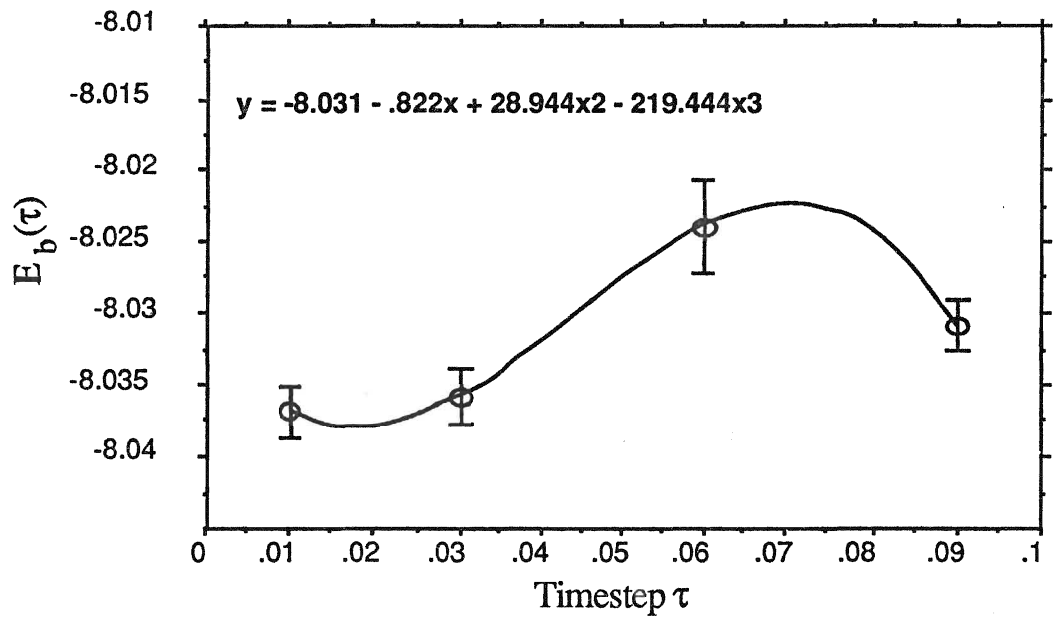


Fig. III.B.1.5. VQMC estimate of E_{var} with Ψ_{2a3}.

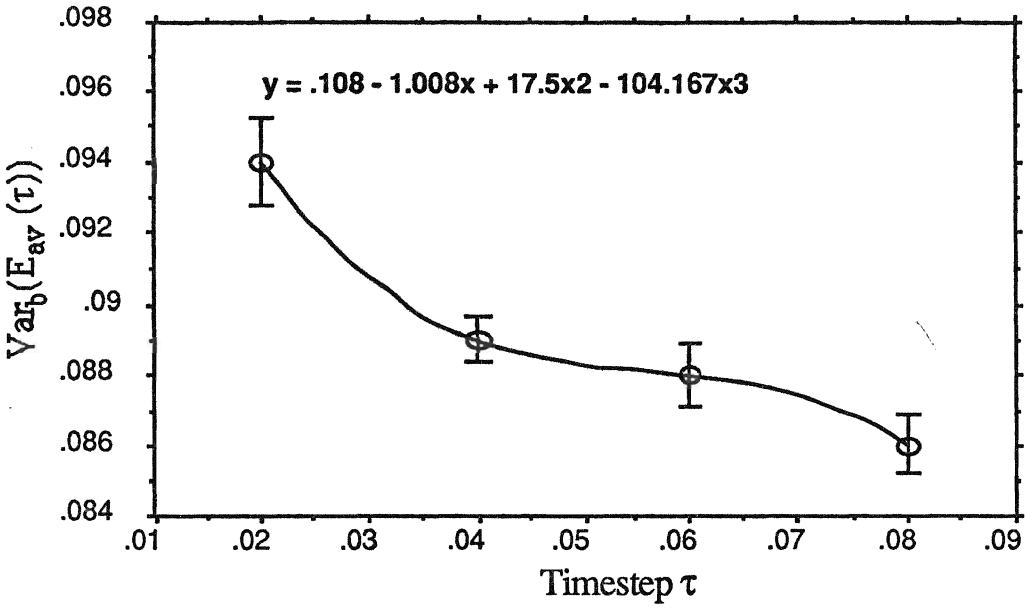


Fig. III.B.1.6. VQMC estimate of $Var(E_{av}(\tau \rightarrow 0)$ with Ψ_{2a4}.

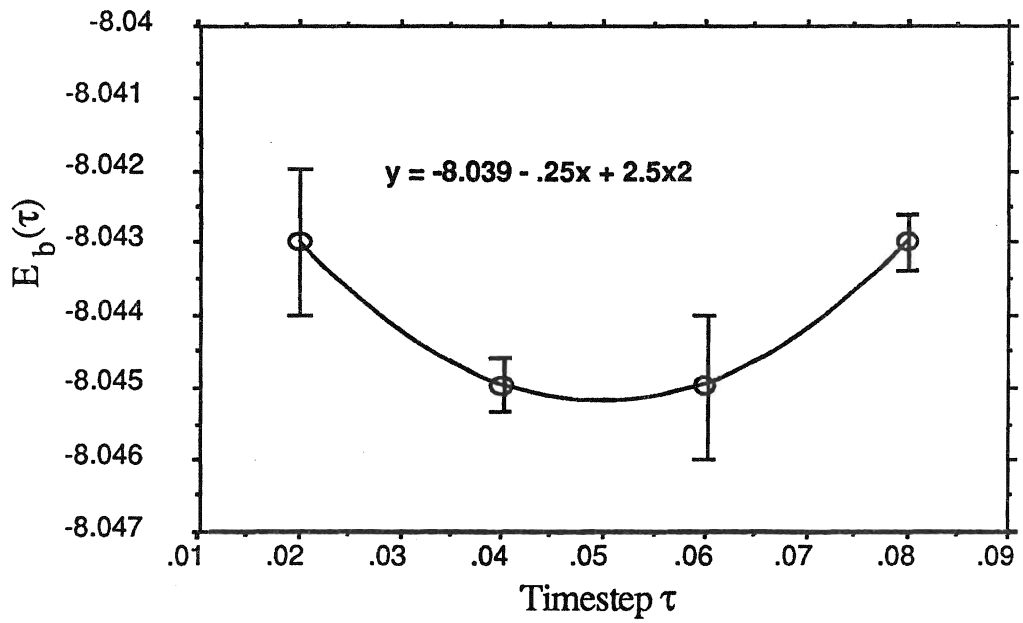


Fig. III.B.1.7. VQMC estimate of E_{var} with Ψ_{2a4}.

(–8.023 ± .0017 and .122 ± .002). The former has an extrapolated energy of -8.042 ± .002, and the latter of -8.048 ± .006 (See also Table III.A.3.8). The main reason for this is that the optimization was done at a certain timestep, τ=0.06, so that the result at that point was favoured and the curve of $E_b(\tau)$ vs. τ became lower at the center and higher on both sides. Thus the extrapolated energy cannot be lowered by much (see Fig. III.B.1.4-7 and also Fig. III.A.3.3-4, where the variance is minimized instead of the energy). This is the so-called **timestep dependency** problem.

To solve this problem, **we can use a mixture of ensembles generated from several timesteps, or we can use for the optimization the ensembles generated at a small timestep**, say τ=0.01, since this will favour the extrapolated energy. This would improve the double-ζ set further.

2. Optimization starting with a "guessed" Ψ_T.

All of the optimizations discussed above started with Reynold's unoptimized Ψ_{II}, considered to be a reasonable wavefunction. A more challenging and general question is: can we start with a "guessed" Ψ_T ? The investigation of this problem began with a guessed wavefunction Ψ_{3fg} (Table11 in App. B) with a minimum basis set, the 3-function set. Also, the search for other sets began at the same time. Fig. III.B.2.1 is a diagram of the whole procedure, where all optimizations were done by **one run** with the algorithm of **Eq.(II.C - 3)**, and sampling of the DJ ensembles was done with **6000** configurations generated at $\tau = 0.01$. The results are shown in Table III. B.2.1 - 5 and Fig. III.B. 2.2 -7.

A great deal of improvement seems to have been achieved after the optimizations from the 3-function set to 5-function and double - ζ sets, the variances being at about 7 times lower and the energies being lowered by about 0.2 atomic units (see Table III. B.2.1). But the simulations with these optimized functions gave unsatisfactory results, listed in Table III. B.2.2. Although the variances have been reduced, there is almost no change in the block energy for Ψ_{2b1}, the optimized function with tbe 5-function set, and that of Ψ_{2b2}, the optimized function with a double - ζ set, is even worse than before. Also note that Ψ_{2b1} gave lower iteration variance than Ψ_{2b2}.

A further verification of the fact that optimizations for wavefunctions with more adjustable parameters gave worse results was provided by computing the variational energy and the extrapolated iteration variance (See Table III. B.2.3-5 and Fig. III. B.2.2-7). (The optimization of wavefunctions

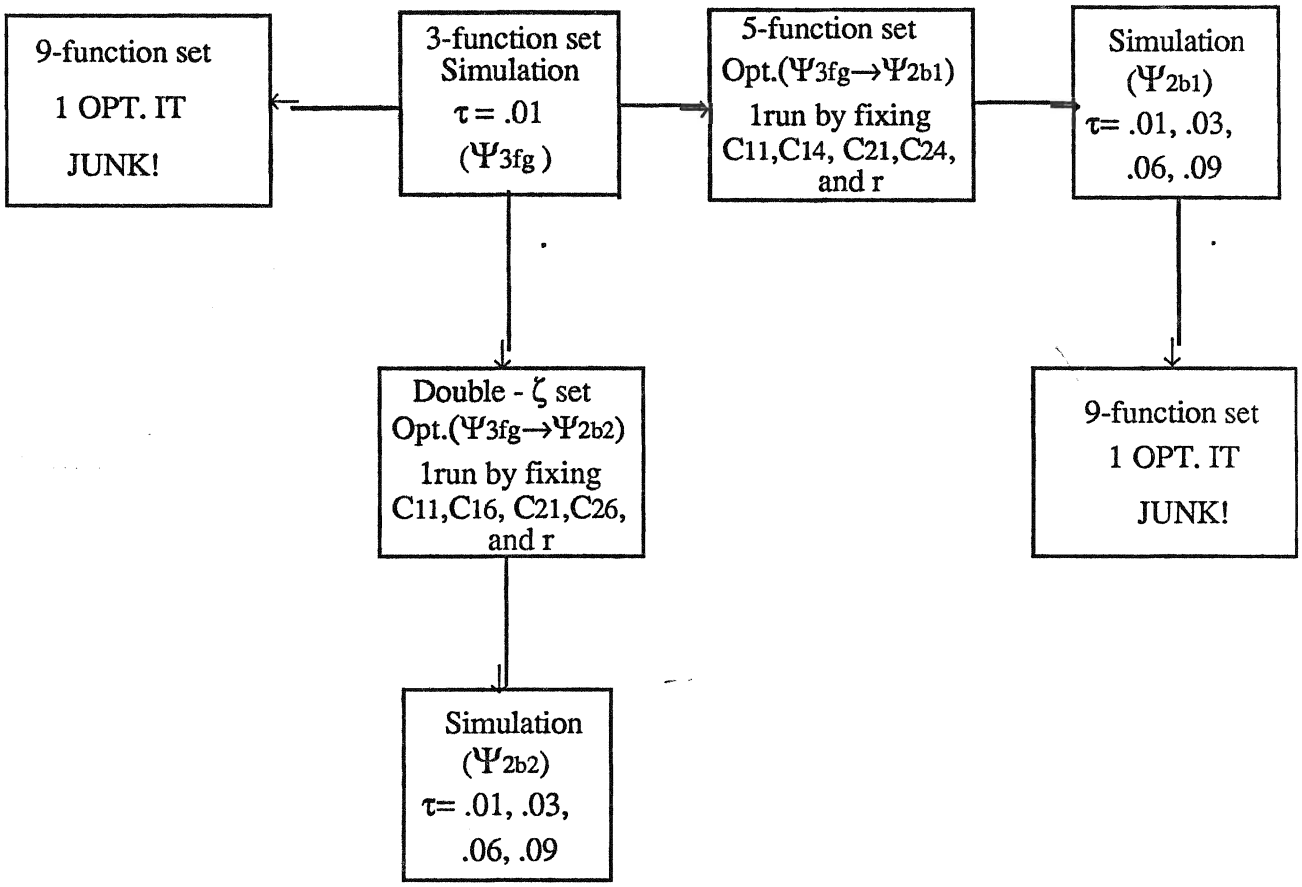


Fig. III.B.2.1. Logic route of the optimizations starting with a guessed Ψ_T.

Table III. B. 2. 1. Changes of the iteration means in the optimizations of III.B.2.

Optimization	$E_{av,ini}$	$E_{av,fin}$	$Var(E_{av})_{ini}$	$Var(E_{av})_{fin}$	ΔE_{av}	$\Delta Var(E_{av})$
$\Psi_{3fg} \rightarrow \Psi_{2b1}$[a]	-7.942	-8.037	0.785	0.095	-0.213	-0.603
$\Psi_{3fg} \rightarrow \Psi_{2b2}$[b]	-7.947	-8.046	0.785	0.099	-0.213	-0.224
$\Psi_{3fg} \rightarrow$ 9-function set[c]			JUNK !			
$\Psi_{2b1} \rightarrow$ 9-function set[d]			JUNK !			

[a,b]The optimizations with the algorithm Eq.(II.C - 3), where Eg = -8.070.
[c,d]The optimizations with the algorithm Eq.(II.C - 5), where Eg = -8.070.

Table. III. B. 2. 2. Block means[a] of three wavefunctions in III.B.2.

Ψ_T	E_b	$Var_b(E_{av})$
Ψ_{3fg}	-7.968 ± 0.003	0.925 ± 0.011
Ψ_{2b1}	-7.975 ± 0.003	0.250 ± 0.004
Ψ_{2b2}	-7.838 ± 0.005	0.519 ± 0.011

[a]Each block mean is taken from ten blocks and there are 300 iterations in a block, and 600 configurations in the ensemble. The block means are computed at τ =.01 with the trial wavefunction listed in the same line.

Table III.B.2.3. VQMC estimates with Ψ_{3fg}.

τ	$E_b(\tau)$[a]	$Var_b(E_{av}(\tau))$
.01	-7.968 ± .003	.929 ± .011
.03	-8.004 ± .001	.787 ± .005
.06	-8.056 ± .002	.643 ± .003
.09	-8.107 ± .001	.544 ± .002

Extrapolated[b] to τ=0: E_{av} = -7.950 ± .0040, $Var(E_{av})$ = .890 ± .0059

[a]There are 10 blocks at each τ, 300 iterations in a block, and 600 configurations in an ensemble. Similarly for the tables below.
[b]The extrapolation regression model is quadratic for the variance and linear for the energy. See Fig. III.B.2.2-3.

Table III.B.2.4. VQMC estimates with Ψ_{2b1}.

τ	$E_b(\tau)$	$Var_b(E_{av}(\tau))$
.01	-7.975 ± .003	.250 ± .004
.03	-7.969 ± .001	.248 ± .003
.06	-7.958 ± .002	.269 ± .003
.09	-7.954 ± .003	.293 ± .003

Extrapolated[a] to τ=0: E_{av} = -7.976 ± .0060, $Var(E_{av})$ = .259 ± .0090

[a]The extrapolation regression model is cubic. See Fig. III.B.2.4-5.

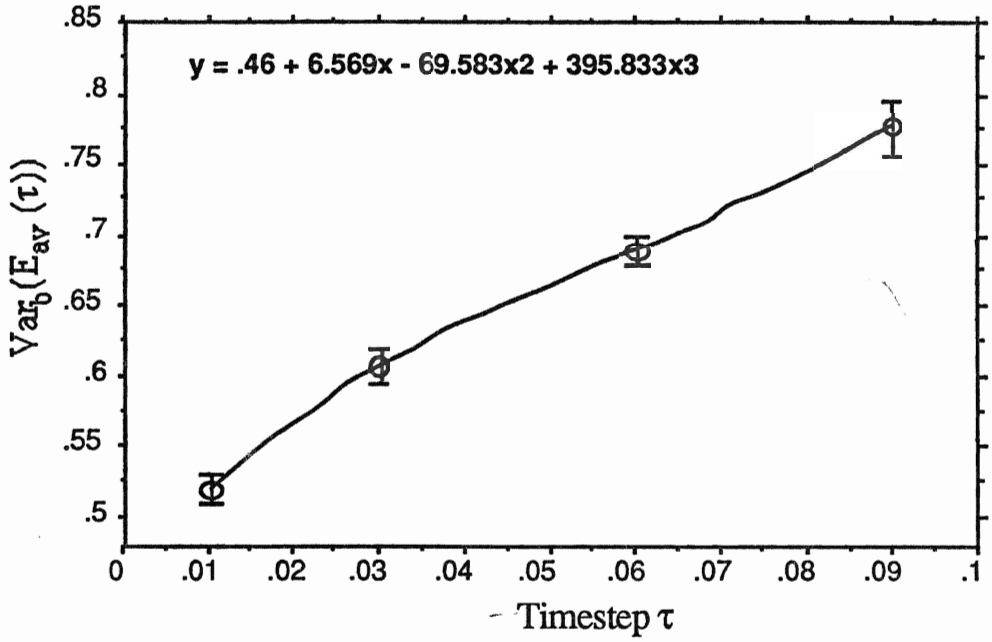


Fig. III. B. 2. 2. VQMC estimate of $Var(E_{av}(\tau \rightarrow 0)$ with Ψ_{3fg}.

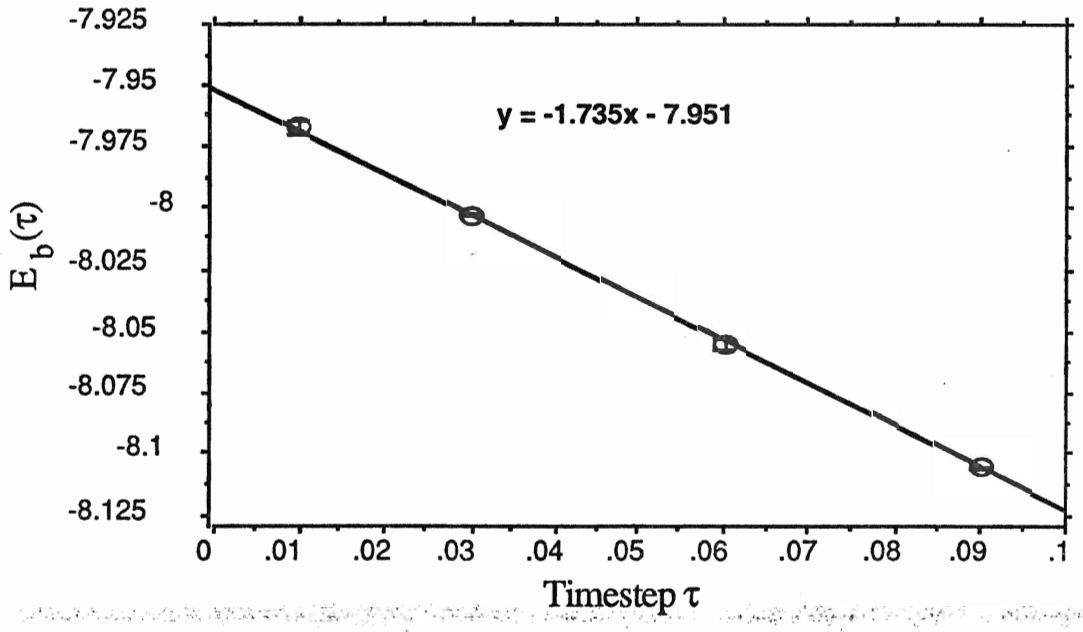


Fig. III. B. 2. 3. VQMC estimate of E_{var} with Ψ_{3fg}.

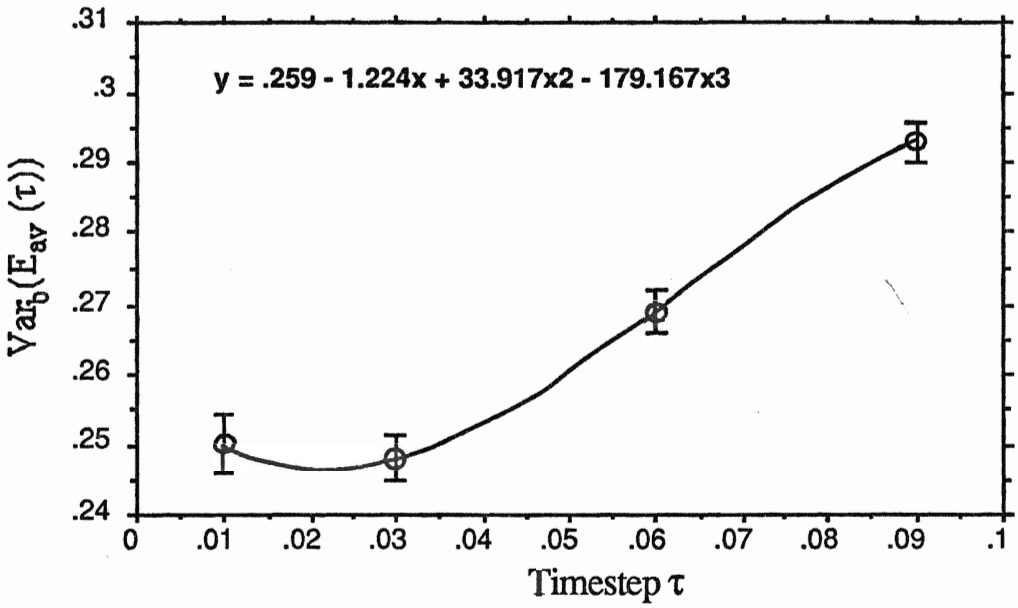


Fig. III. B. 2. 4. VQMC estimate of $Var(E_{av}(\tau \rightarrow 0)$ with Ψ_{2b1}.

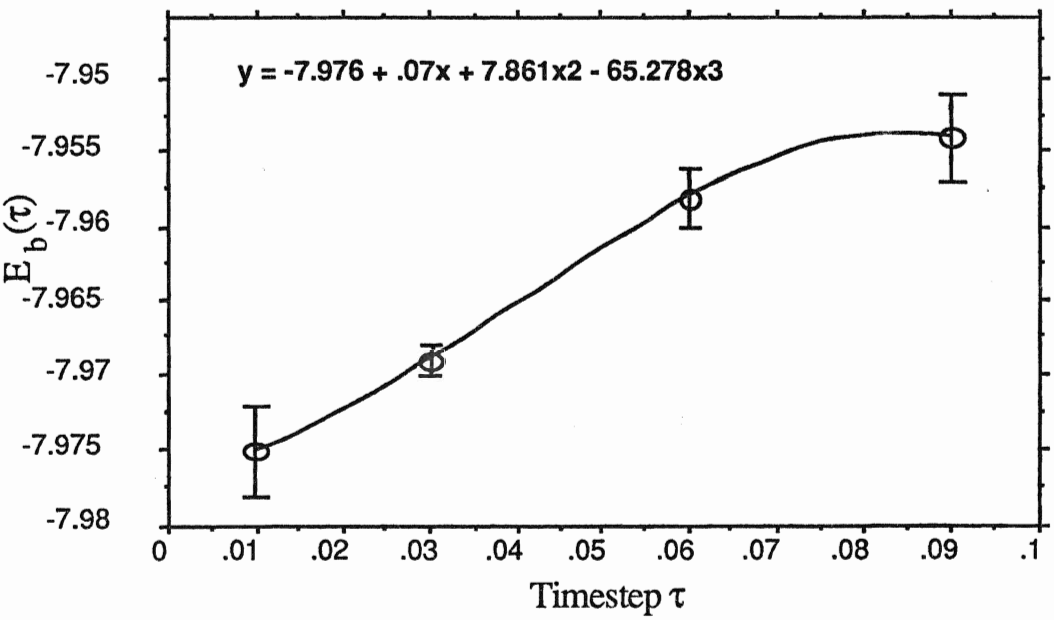


Fig. III. B. 2. 5. VQMC estimate of E_{var} with Ψ_{2b1}.

Table III.B.2.5. VQMC estimates with Ψ_{2b2}.

τ	$E_b(\tau)$	$Var_b(E_{av}(\tau))$
.01	-7.838±.005	.519±.011
.03	-7.793±.005	.605±.012
.06	-7.746±.005	.689±.011
.09	-7.787±.006	.776±.019

Extrapolated[a] to τ=0: $E_{av} = -7.857 \pm .013$, $Var(E_{av}) = .460 \pm .029$

[a]The regression models are cubic. See Fig. III.B.2.6-7.

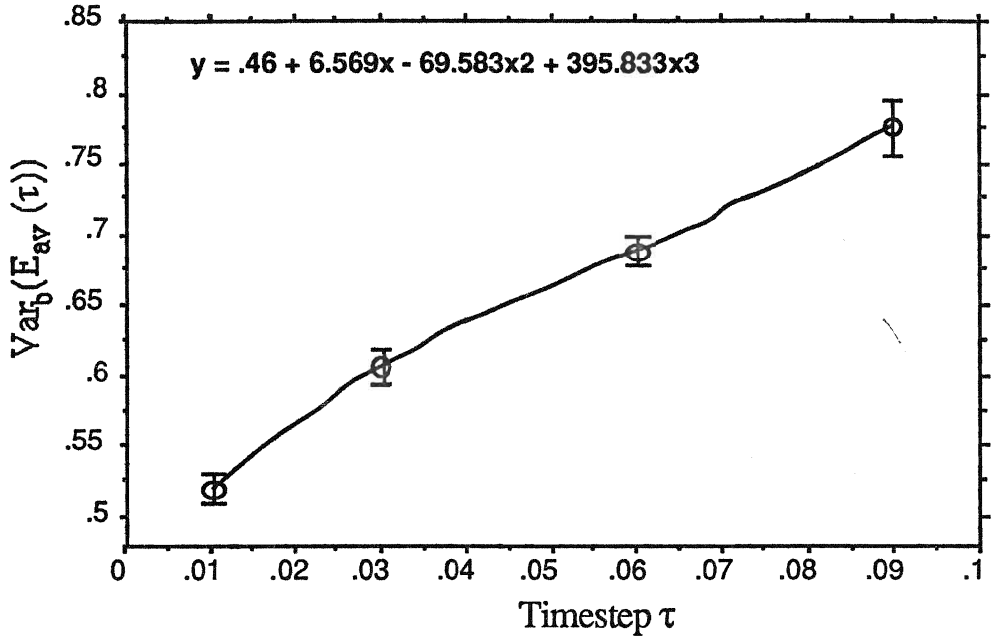


Fig. III. B. 2. 6. VQMC estimate of $Var(E_{av}(\tau \rightarrow 0)$ with Ψ_{2b2}.

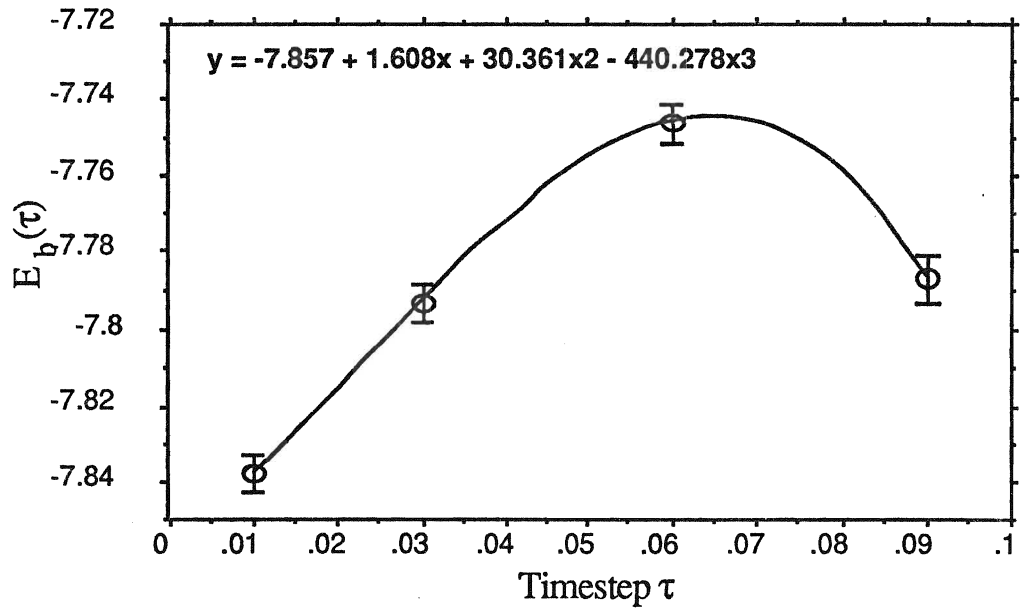


Fig. III.B. 2. 7. VQMC estimate of E_{var} with Ψ_{2b2}.

with more adjustable parameters should result in a better wavefunction, rather than a worse one) Similar contradictory results were also obtained from a group of other runs with these two sets (5-function and double-ζ) as well as the 9 -function set. The reason for this problem is still the ensemble dependency, described as follows:

The trial wavefunction keeps changing because of the change in its parameters during optimization, but the ensemble used is kept fixed. Here the ensemble is sampled from $|\Psi_{3fg}|^2$, which is very different from $|\Psi_{2b1}|^2$, generated by the optimized Ψ_{2b1} in the simulation. This difference in the ensembles is even greater than before because of the drastic change in the basis set (3 - function set $\rightarrow$ 5 - function set). Thus the optimization iteration means (obtained by sampling $|\Psi_{3fg}|^2$) greatly differed from the simulation block means (obtained by sampling $|\Psi_{2b1}|^2$). Therefore, the case of the change of basis set from 3 - function set $\rightarrow$ double - ζ set provides even worse results, as previously mentioned. **An optimization performed by fixing the ensemble with the algorithm of Eq.(II.C - 3) or Eq.(II.C - 4) cannot lead to a better wavefunction if the initial wavefunction is very different from the final one.**

To solve this problem the so-called "SO" algorithm (simulation $\rightarrow$ optimization $\rightarrow$ simulation $\rightarrow$ optimization $\rightarrow$ ...) is suggested, where the ensembles change while the wavefunctions change. This involves inserting a simulation procedure between every two groups of parameters to be optimized in a loop such as that shown in Fig. III.B. For example, for a double-ζ set the logic route is shown in Fig. III.B.2. 8.

Table III.A.2.6 was actually obtained from such a SO algorithm but only one group (the five "dominant" parameters) in the loop shown in the Fig. III.B.2.8. From this table, one can determine that with this algorithm the optimization results will agree with that of the simulation when the wavefunction has converged. In any case, if the change of the wavefunction is great compared with the full optimization within a same set, but not too great, the ensemble-fixing method can still be used, but one must use Wilson's weighted algorithm, **Eq.(I - 2) or Eq.(II.C - 5).**

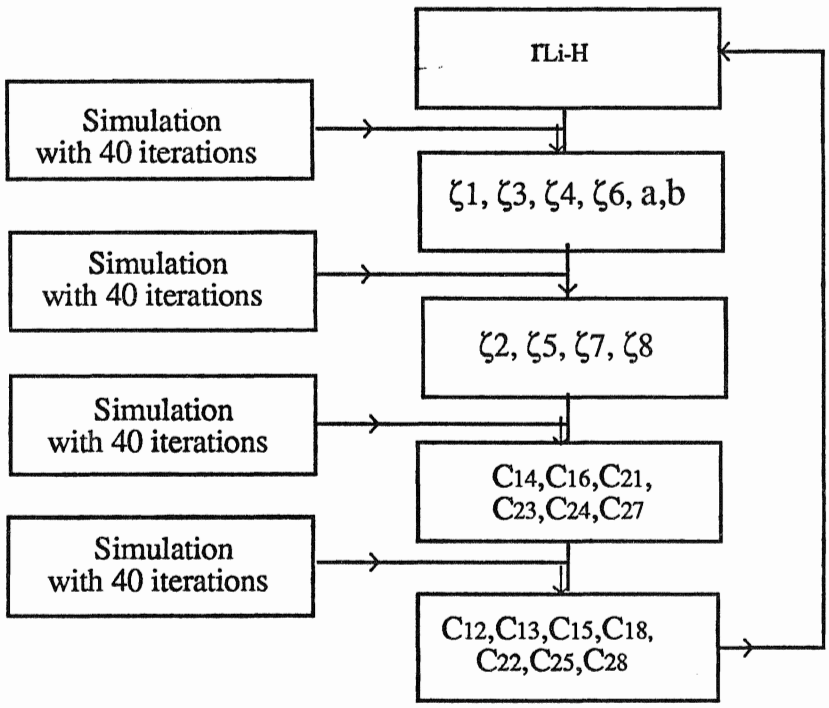


Fig. III.B.2. 8. Logic route of the loops in the full optimization of a double-ζ set with the SO algorithm.

The optimizations just mentioned were repeated with the algorithm of

Eq.(II.C - 5), and they were successful. The final results will be discussed in Chapter IV. Here it is shown that two other runs with this algorithm for the 9-function set did not work well, giving almost zero variances and exact energies (-8.070) after the optimizations, but nonsensical simulation results, denoted by the word "JUNK"! in Table III.B.2.1, and Fig. III.B.2.1. The only obvious reason for this is that too great a change in the wavefunction has occurred in going from the 3- and 5-function sets to the 9-function set. In this case we must employ the SO algorithm to perform a full optimization.

Otherwise, all that can be done is a partial optimization. By fixing the best 5-function set parameters the others were optimized for the wavefunction with a 9-function set, and by fixing the best double-ζ set parameters the procedure was repeated for the wavefunction with a double-ζ plus $3d_{z^2}$ set. All of these results appear in Chapter IV.

IV. Results and Discussion

Following the logical route in Fig.IV.1, the wavefunctions with the five different basis sets were finally optimized partially or fully, and then the variational energies and the individual iteration variances were estimated with the best wavefunction from each set. The results are listed in Table IV.1, where the results of Reynold's Ψ_{II}, Ψ_{3fg}, and Ψ_{1c} are also listed for the convenience of comparision. Each ensemble used in the optimizations had 6000 configurations, which were generated at $\tau = 0.01$(for $\Psi_{3.1}$, $\Psi_{3.2}$, $\Psi_{3.5}$, and $\Psi_{3.6}$) or they were a mixture(for $\Psi_{3.3}$ and $\Psi_{3.4}$) of ten small ensembles generated at the four different timesteps (three from each of $\tau =$ 0.01, and 0.03, and two from each of $\tau = 0.06$ and 0.09, which has 600 configurations). The algorithms used in the optimization are based on Eq.(I - 2), the weighted sum of squares.

It can be seen from Table IV.1 that the optimized $\Psi_{3.1}$ (a 3-function basis set) when used in a simulation produced a result even better than that of Reynold's 8-function set. When the basis set was expanded from the 3-function set to the 5-function set $\Psi_{3.2}$, an even greater improvement is obtained.

Continuing to expand the 3-function set to the double-ζ set, the function $\Psi_{3.3}$ is obtained. It gave smaller iteration variance but not a lower energy, compared with Ψ_{1c}. One of the reasons for this is because the ensemble used to perform the optimization is a mixing one instead of that generated at $\tau = 0.01$, and thus the timestep-dependency problem cannot be dealt with thoroughly, resulting in a higher energy for small timesteps (See

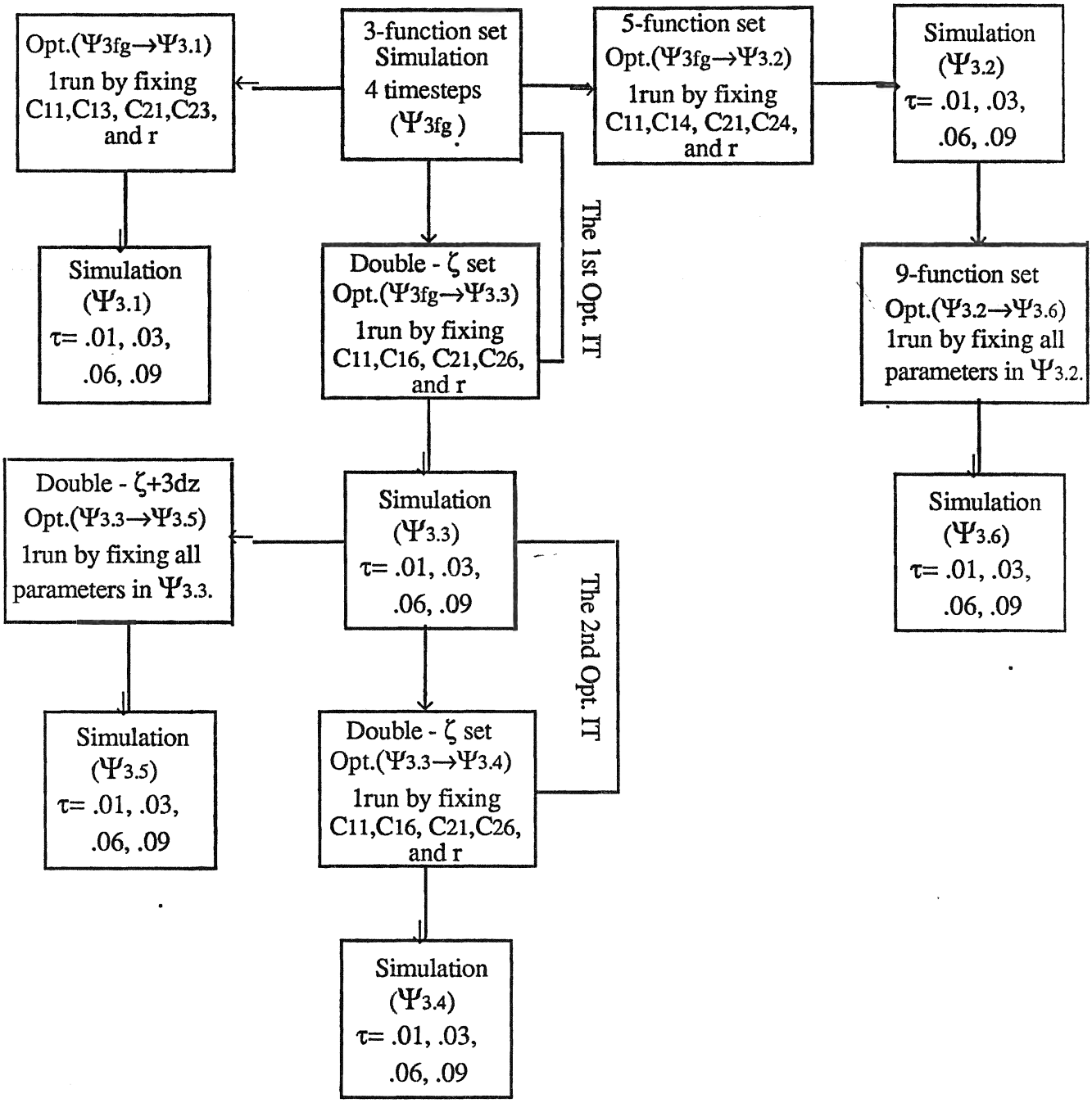


Fig. IV.1. Logic route of the final calculation.

Table IV.1. Final results of different basis sets

Ψ_T (basis set)	E_{var} [a]	$Var(E^n_{ave}(\tau\rightarrow 0))$[b]	Model[c]	Comment
Ψ_{II} (Double - ζ)[d]	**-7.990±..0025**	**.291±..0013**	**Quadratic**	**Unopt.[e]**
Ψ_{3fg} (3-function)[f]	-7.950±.0040	.890±.0059	Linear	Unopt.
$\Psi_{3.1}$ (3-function)[g]	**-7.998±.0045**	**.226±.0044**	**Cubic**	**Opt.[h]**
$\Psi_{3.2}$ (5-function)[g]	-8.027±.0020	.113±.0012	Quadratic	Opt.
Ψ_{1c} (Double - ζ)[d]	-8.048±.0058	.126±.0052	Cubic	Opt.
$\Psi_{3.3}$ (Double - ζ)[g]	-8.047±.0026	.106±.0051	Cubic	Opt.
$\Psi_{3.4}$ (Double - ζ)[g]	-8.049±.0024	.106±.0025	Cubic	Opt.
$\Psi_{3.5}$(Double-ζ+3d_{z^2})[g]	**-8.052± 0.0026**	**.094±.0026**	**Cubic**	Opt.
$\Psi_{3.6}$ (9-function)[g]	-8.049±.0009	.094±.0010	Cubic	Opt.

[a] Extrapolated variational energy.
[b] Extrapolated individual iteration variance.
[c] The polynormial models extrapolating $\tau \rightarrow 0$.
[d] The values listed in the same line were obtained in Section III.A.3.
[e] The wavefunction listed in the same line is unoptimized one.
[f] The values listed in the same line were obtained in Section III.B.2.
[g] The values listed in the same line were obtained in this section.
[h] The wavefunction listed in the same line is optimized one.

Table IV.2 and Fig. IV.2-3). One more Opt. IT. was also run for the set, but the results were the same, implying the function had converged.

There is also quite a large improvement in expanding the 5-function set to the 9-function set. But the latter $\Psi_{3,6}$ did not provide better results than that of $\Psi_{3,4}$. This result was expected, because only a partial optimization (with all parameters in the 5-function set fixed) was carried out in going from the 5-function set to the 9-function set. The results would be much better if a full optimization with the SO algorithm were to be performed.

The best results here are that of the double-ζ plus d_{z2} set function $\Psi_{3,5}$, which was obtained by a partial optimization of $\Psi_{3,3}$ (with all parameters in the double-ζ set of $\Psi_{3,3}$ fixed). The energy, $-8.052\pm .0026$, was comparable with the fixed-node energy (-8.059 ± 0.004)[1] of Reynold's unoptimized Ψ_{II}, and the variance was $0.094\pm.0026$. It is obvious that $\Psi_{3,5}$ can also be much improved if one does a full optimization with the SO algorithm by starting with $\Psi_{3,3}$.

Summarizing the above discussion and also that of the previous sections, one finds that a partial optimization can be easily performed with little difficulty because there are only a few parameters to be optimized; when performing a full optimization, one must avoid falling into a wrong minimum, and one must deal with the linear - dependent, timestep - dependent and ensemble - dependent problems.

To avoid a wrong minimum, the algorithm of Eq.(II.C - 3) should be used. As shown in section III.B.1 the optimization can be forced to go along

Table IV.2. VQMC estimates with $\Psi_{3,3}$.

τ	$E_b(\tau)$	$Var_b(E_{av}(\tau))$
.01	-8.048±.001	.105±.001
.03	-8.051±.001	.103±.002
.06	-8.056±.001	.101±.002
.09	-8.060±.001	.099±.002

Extrapolated[a] to τ=0: E_{av} = -8.047 ± .0026, $Var(E_{av})$ = .106 ± .0051

[a]The regression models are cubic. See Fig. IV.2-3.

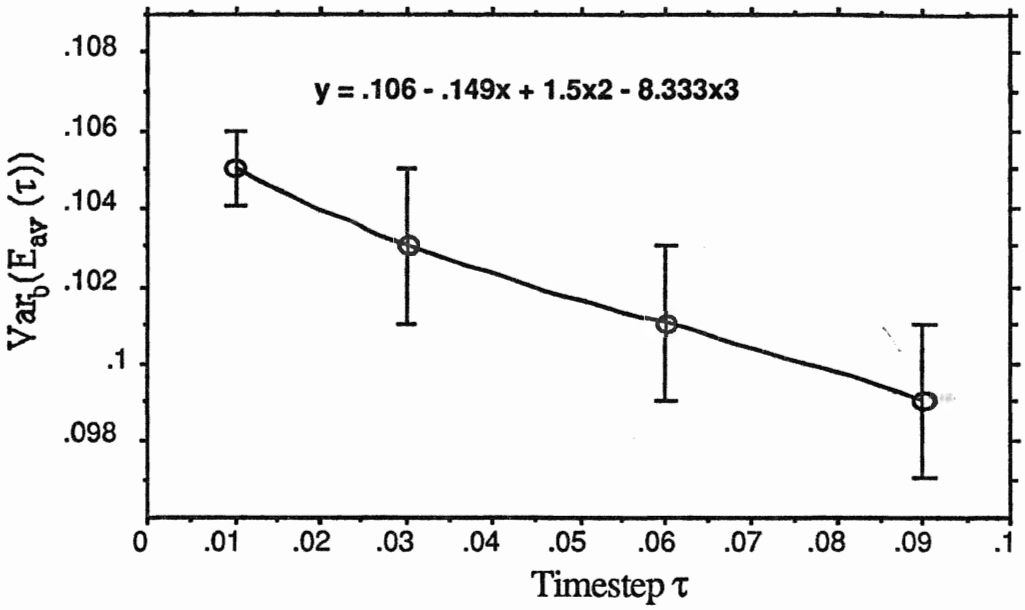


Fig. IV.2. VQMC estimate of $Var(E_{av}(\tau \rightarrow 0)$ with $\Psi_{3.3}$.

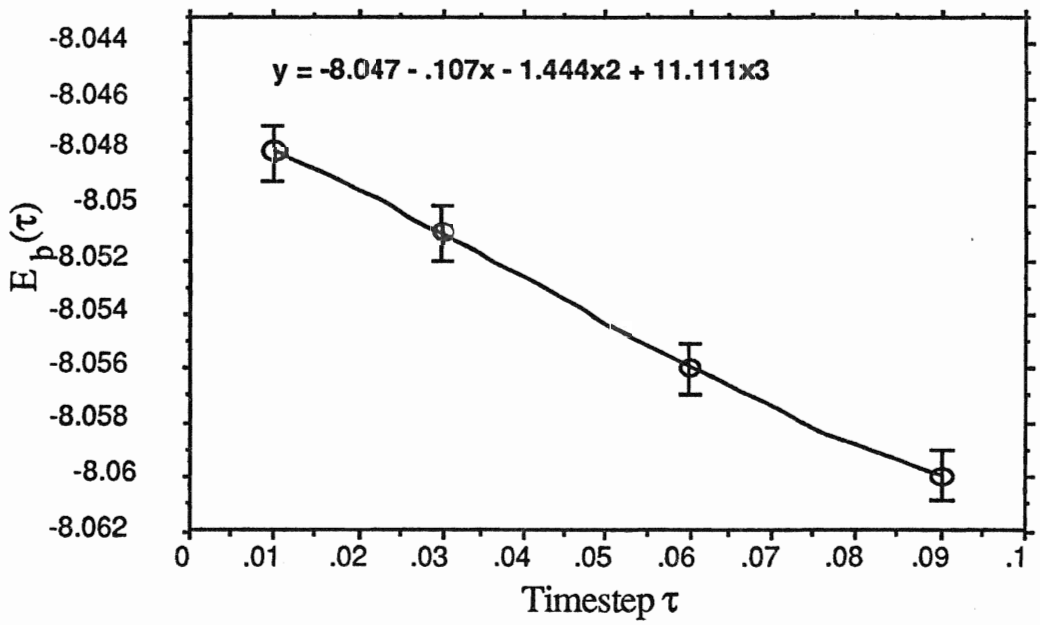


Fig. IV.3. VQMC estimate of E_{var} with $\Psi_{3.3}$.

a correct path. But relative to using Eq.(II.C - 4), this algorithm is more sensitive to the timestep at which the ensemble is generated (recall Fig. III.B.1.4.-7). This timestep dependency problem can be solved by sampling a mixture of ensembles generated at several timesteps, or with ensembles generated only at a single small timestep only, say τ=0.01. We have found that the limiting case for the mixing ensemble, i.e., all ten 600-configuration ensembles generated at τ=0.01, is the best that we have found for removing this dependency (recall Fig. III.B.2.4.-7). As discussed in the front part of section III.B, dividing all the parameters to be estimated into several groups so that the linearly-related ones can be separated is the only way we have found to deal with the linear - dependency problem.

The difficulties in dealing with the ensemble dependency problem become greater with more complicated wavefunctions. Depending on how complicated the wavefunction to be optimized is, one can just use Eq.(II.C - 3) or Eq.(II.C - 4) and more configurations (600-10000) which are all fixed. In more complicated case, one also can include the use of Eq.(II.C - 5). In the extreme case, use of the SO algorithm may also be included.

The wavefunctions described above were obtained by optimization using the **DJ** ensembles. These ensembles are adequate for partial optimization because the wavefunction being optimized is changed little. However, I believe the **DJB** ensembles are preferred for the **more complicated** full optimization. This should solve the ensemble dependency problem. That is because in the full optimization the final wavefunction Ψ_{fin} can more closely approach the exact one Φ_0, and thus the ensemble sampling from $\Phi_0\Psi_{ini}$ is much closer to $|\Psi_{fin}|^2$ than that sampling from the initial

distribution $|\Psi_{ini}|^2$. But for this, the truncation of both F and E_L should be used in optimization as well as in simulation.

Therefore, to efficiently perform an optimization one should

1) use Eq.(I - 2) to minimize the weighted energy;
2) divide all the parameters to be optimized into several groups, and also fix some dominant coefficients (usually a pair for each MO);
3) use a DJ ensemble in partial optimization or a DJB ensemble in full optimization (yet to be verified), which is generated at as small a timestep as possible (say $\tau = 0.01$ for LiH);
4) use a large ensemble size (6000-10000 or more configurations);
5) use a more complicated straightforward Opt. IT approach, the SO algorithm, for the case where the change in the wavefunction from Ψ_{ini} to Ψ_{fin} is very large.

Proposals 1) and 4) were also suggested by Wilson *et al.*[20]

Appendix A. Principles of Diffusion Quantum Monte Carlo[1,18,21]

Diffusion Quantum Monte Carlo is a statistical simulation method for solving the time-independent Schrödinger equation, where one simulates a so-called Green's function by letting the electrons in a system, under the influence of a certain "guiding" function, walk randomly or diffuse through configuration space. When a stationary state, or a dynamic equilibrium of the system is reached, values of the ground-state energy and other properties may be estimated.

1. Theory

Consider the Schrödinger equation in atomic units,

$$[-1/2\nabla^2 + V(\mathbf{R})\,]\,\Phi(\mathbf{R}) = E\,\Phi(\mathbf{R}) \qquad (A-1)$$

where $V(\mathbf{R})$ is the potential energy, and $\mathbf{R}$ represents the 3N Cartesian coordinates of N electrons. Assume that the trial solutions to this equation, $\phi_i(\mathbf{R})$ (i = 1, 2, ...,) could infinitely closely approach the exact wavefunction $\Phi(\mathbf{R})$ during the course of a calculation. These functions thus become operationally "time"- dependent, becoming more and more accurate as time progresses. The same thing occurs in Diffusion QMC. Introducing such a time t (called imaginary time) to Eq. (A - 1), and after some algebraic manipulation we obtain,

$$-\partial\Phi(\mathbf{R},t)/\partial t = -1/2\nabla^2\Phi(\mathbf{R},t) + (V(\mathbf{R}) - E_0)\Phi(\mathbf{R},t) \qquad (A-2)$$

where E_0 is the ground-state energy of the system considered. As mentioned above, $\Phi(\mathbf{R},t)$ is an approximate solution to Eq. (A - 1) at time t, and $|\Phi(\mathbf{R},t)|^2$ is the corresponding probability density for the electrons in the system.

Eq. (A - 2) is the Fick's diffusion equation if the second term on the right-hand side, called the "**branching**" term, is omitted. If the second term on the right-hand side was present alone, the equation would be a rate equation, describing the change in the population of particles. Here this population is almost proportional to the potential $V(\mathbf{R})$(when it is large enough), which is unbounded and can diverge to infinity. If one simulates this equation, large fluctuations in the population could possibly be caused, which is difficult to control, and this results in large statistical errors. Also, it will slow the convergence of the averages $< V(\mathbf{R}) - E_0 >$, from which one expects to valuate the estimate of E_0. Therefore we do not directly simulate this equation, but rather use its elegant alternative, where the **importance sampling** is employed. Introducing a new density function $f(\mathbf{R},t)$ by multiplying a known trial wavefunction $\Psi_T(\mathbf{R})$ by the original one

$$f(\mathbf{R},t) \equiv \Phi(\mathbf{R},t)\psi_T(\mathbf{R}) \qquad (A - 3)$$

and substituting this into Eq.(A - 2), we obtain

$$-\partial f(\mathbf{R},t)/\partial t = -1/2\nabla^2 f(\mathbf{R},t) + (E_L(\mathbf{R}) - E_0)f(\mathbf{R},t) + \nabla(f(\mathbf{R},t)F(\mathbf{R})) \qquad (A - 4)$$

where $E_L(\mathbf{R})$ is called the **local energy** given by

$$E_L(\mathbf{R}) = \mathbf{H}\Psi_T(\mathbf{R}) / \Psi_T(\mathbf{R}) \qquad (A-5)$$

Now, the branching term is proportional to $< E_L(\mathbf{R})-E_0 >$, which is much better-behaved than the original one, since the local energy $E_L(\mathbf{R})$ is much smoother than $V(\mathbf{R})$, given a sensible choice of $\Psi_T(\mathbf{R})$. Moreover, the closer the $E_L(\mathbf{R})$ is to E_0, the smoother the branching term will be.

The term $\nabla(f\mathbf{F}(\mathbf{R}))$ is called the "**drift**" term because it acts to impose a directed **drift** velocity $\mathbf{F}(\mathbf{R})$ on the diffusion, where the velocity is

$$\mathbf{F}(\mathbf{R}) = \nabla \Psi_T(\mathbf{R}) / \Psi_T(\mathbf{R}) \qquad (A-6)$$

According to Eq.(A - 6), $\mathbf{F}(\mathbf{R})$ is larger where $\Psi_T(\mathbf{R})$ is smaller, and hence any diffusers reaching such a region are driven away faster, leaving density $f(\mathbf{R},t)$ smaller. Thus, the trial function $\Psi_T(\mathbf{R})$ also plays the role of a "guide", guiding the electrons diffusing within the volumes bounded by the nodes of $\Psi_T(\mathbf{R})$.

By expanding $f(\mathbf{R},t)$ in a complete set of eigenfunctions $\Phi_i(\mathbf{R})$ of the Hamiltonian $\mathbf{H}$, we have

$$f(\mathbf{R},t) = \sum_i N_i \Phi_i(\mathbf{R}) \qquad (A-7)$$

substituting it into Eq.(II.1 - 4) and solving the equation, we get

$$f(\mathbf{R},t) = \sum_i N_i \exp(-(E_i - E_0)\, t)\, \Phi_i(\mathbf{R}) \qquad (A-8)$$

where E_i is the eigenvalue of the corresponding eigenfunction $\Phi_i(\mathbf{R})$ and the N_i are the coefficients. It is apparent that, when $t \rightarrow \infty$,

$$\exp(-(E_i - E_0)\, t) \rightarrow 0, \text{ if } i \neq 0 \qquad (A-9)$$

$$\exp(-(E_i - E_0)\, t) \rightarrow 1, \text{ if } i = 0 \qquad (A-10)$$

Thus, all terms except the one with a lowest energy E_0 in Eq.(A - 8) are cancelled out and Eq.(A - 8) becomes,

$$f(\mathbf{R},t\rightarrow \infty) = f_\infty(\mathbf{R}) = N_0 \Phi_0(\mathbf{R}) \Psi_T(\mathbf{R}) \qquad (A-11)$$

where $\Phi_0(\mathbf{R})$ is the exact ground state wavefunction. The solution obtained here is called a steady-state solution, since the density $f_\infty(\mathbf{R})$ is the time-independent, and thus a steady state has been reached. Taking the average of $E_L(\mathbf{R})$ over $f_\infty(\mathbf{R})$,

$$\begin{aligned} \langle E_L(\mathbf{R}) \rangle &= \int E_L(\mathbf{R})\, f_\infty(\mathbf{R}) d\,\mathbf{R} / \int f_\infty(\mathbf{R}) d\,\mathbf{R} \\ &= \int (\mathbf{H}\, \Psi_T / \Psi_T)(N_0\, \Phi_0\, \Psi_T) d\,\mathbf{R} / \int (N_0\, \Phi_0\, \Psi_T) d\,\mathbf{R} \\ &= \int E_0\, \Phi_0(\mathbf{R})\, \Psi_T(\mathbf{R})\, d\,\mathbf{R} / \int \Phi_0(\mathbf{R})\, \Psi_T(\mathbf{R})\, d\,\mathbf{R} \\ &= E_0 \end{aligned} \qquad (A-12)$$

Therefore, as soon as the density $f_\infty(\mathbf{R})$ is obtained the ground state

energy E_0 is given from the above integral. In fact, since the density is stable at a steady state, then the population of the electrons should remain unchanged. Thus the branching term or the rate term should be equal to zero on the average, i.e.

$$< E_L(\mathbf{R}) - E_0 > = < E_L(\mathbf{R}) > - E_0 = 0$$

and therefore,

$$< E_L(\mathbf{R}) > = E_0 \qquad (A\text{-}13)$$

We cannot solve Eq.(A - 4) exactly, but by using a small timestep τ, we can solve the equation to within an error which is of $O(\tau)$ or higher. This solution is called the DQMC **Green's function**, given by

$$\begin{aligned} G(\mathbf{R} \rightarrow \mathbf{R}', \tau) = (2\pi\tau)^{-3N/2} & \\ & \times \exp\left[-\tau\{[\, E_L(\mathbf{R}) + E_L(\mathbf{R}')]/2 - E_0\}\right] \\ & \times \exp\left\{-[\, \mathbf{R}' - (\mathbf{R} + \tau F(\mathbf{R}))\,]^2/2\tau\right\} \end{aligned} \qquad (A\text{-}14)$$

where $\mathbf{R}'$ represents the 3N coordinates of the N electrons in time τ. The Green's function satisfies the boundary condition,

$$G(\mathbf{R} \rightarrow \mathbf{R}', \tau = 0) = \delta(\mathbf{R} \rightarrow \mathbf{R}') \qquad (A\text{-}15)$$

then get directly,

$$f(\mathbf{R}') = \int \delta(\mathbf{R}\rightarrow\mathbf{R}')\, f(\mathbf{R})\, d\mathbf{R}$$
$$= \int G(\mathbf{R}\rightarrow\mathbf{R}', \tau = 0)\, f(\mathbf{R})\, d\mathbf{R} \qquad (A - 16)$$

or within the τ-related error,

$$G(\mathbf{R}\rightarrow\mathbf{R}', \tau_{small}) = \delta(\mathbf{R}\rightarrow\mathbf{R}') + O(\tau) \qquad (A - 17)$$

$$f(\mathbf{R}',\tau) \approx \int G(\mathbf{R}\rightarrow\mathbf{R}',\tau)\, f(\mathbf{R})\, d\mathbf{R} + O(\tau) \qquad (A - 18)$$

So, the Green's function is actually "a transition probability for moving the set of coordinates from $\mathbf{R}\rightarrow\mathbf{R}'$ in time τ" within the $O(\tau)$ error. Similarly,

$$f(\mathbf{R}',\tau(1+1)) \approx \int G(\mathbf{R}\rightarrow\mathbf{R}',\tau)\, f(\mathbf{R},\tau)\, d\mathbf{R} + O(\tau)$$

$$f(\mathbf{R}'',\tau(1+2)) \approx \int G(\mathbf{R}'\rightarrow\mathbf{R}'',\tau)\, f(\mathbf{R}',(1+1)\tau)\, d\mathbf{R} + O(\tau)$$

. . . . . .

$$f(\mathbf{R}^n,\tau(1+n)) \approx \int G(\mathbf{R}^{n-1}\rightarrow\mathbf{R}^n,\tau)\, f(\mathbf{R}^{n-1},n\tau)\, d\mathbf{R} + O(\tau) \qquad (A - 19)$$

When $n \rightarrow \infty$, we indeed get $f_\infty(\mathbf{R})$ with an error of $O(\tau)$. The above iterations are done by simulating the Green's function.

2. Simulation

Suppose there are N electrons in the molecule. Choose M replicates (**configurations**) of the molecule, and have the N×M electrons walk randomly in a 3N-dimensional configuration space **R**. If at time t we take a "snapshot" of them, and repeat for an infinite duration of time, an infinite number of such snapshots could be obtained. An electron "cloud" could be formed by overlaping all of these snapshots. This "cloud" is called an **ensemble**, distributed as the density $f(\mathbf{R},t)$.

In the Green's function (Eq. (A -14)), there are mainly two parts in exponential forms. If the origin point in the space is **shifted** to $\mathbf{R} + \tau\mathbf{F}$, the second part is a normal distribution with a mean equal to 0 and a variance equal to τ, denoted by $N(0, \sqrt{\tau})$. Thus, the moving of the electrons in the mth configuration in small time τ gives,

$$\mathbf{R'}^{m} = \mathbf{R}^{m} + \tau\mathbf{F} + N(0, \sqrt{\tau}) \qquad (A - 20)$$

The drift caused by $\tau\mathbf{F}$ is called **Drift** (D) and the move caused by $N(0, \sqrt{\tau})$ is called **Jump**(J).

On the other hand, the first exponential part of the Green's function could be taken as a parameter of the $N(0, \sqrt{\tau})$. One simulates it at point **R'** by eliminating or copying the configuration according to

$$M = \text{int}\,(M_m + \zeta) \qquad (A - 21)$$

where ζ is a random number which may be chosen between 0 and 1, and M_m is the first exponent of the Green's function(see Eq.(A - 14)),

$$M_m = \exp(-\tau((E_L(\mathbf{R}) + E_L(\mathbf{R}'))/2 - E_0)) \qquad (A - 22)$$

If M = 0, the mth configuration is "killed"; if M =1, it is kept and unchanged; if M> 1, M-1 additional copies are made. This step is called **Branching** (B) (This step is eliminated in the Variational Quantum Monte Carlo).

When all configurations in the ensemble are taken through these steps once, one **iteration** is finished, and a new ensemble is created. So, one **iteration** consists of three steps: **Drift**, **Jump** and **Branching**. The algorithm to simulate these three steps is called the **DJB** algorithm, shown in Fig. I. 2.

One repeats the iterations until a target time value is reached (after a large number of iterations have been completed); at this point, the steady state is reached and the approximate density $f_\infty(\mathbf{R})$ is given. By taking the average of $E_L(\mathbf{R})$ sampled from the distribution $f_\infty(\mathbf{R}) = N_0\Phi_0(\mathbf{R})\Psi_T(\mathbf{R})$, the ground state energy $E_0(\tau)$ with an $O(\tau)$ error is then obtained according to Eq.(A - 12). Several $E_0(\tau)$'s are obtained by repeating the procedure at several distinct values of τ, and one eventually obtains an exact ground-state energy E_0 by extrapolating those $E_0(\tau)$'s to $\tau \to 0$.

Note that the electron density f(**R**, t) must be non-negative, or

$$f(\mathbf{R},t) \equiv \Phi(\mathbf{R},t)\Psi_T(\mathbf{R}) \geq 0 \qquad (A-23)$$

Therefore, it is required that $\Phi(\mathbf{R},t)$ and $\Psi_T(\mathbf{R})$ have the same sign everywhere. One simulates that by "killing" the configurations when they diffuse to a boundary of the nodally-bounded volume of $\Psi_T(\mathbf{R})$ and Eq.(A - 4) is solved in each such volume separately. It means the wave function of the system concerned is forced to satisfy the boundary condition, $\Phi(\mathbf{R},t) = 0$ at and only at the nodal surface of the trial wave function $\Psi_T(\mathbf{R})$, i. e., the nodes of $\Phi(\mathbf{R},t)$ are fixed by that of $\Psi_T(\mathbf{R})$. Unless this constraint is relaxed, applying DQMC is also denoted as **fixed-node** DQMC.

Appendix B: Reynold's Ψ_{II} and the Optimized Trial Wavefunctions

Table 1.[a] Reynold's trial wavefunction Ψ_{II}[b].

STO	ζ	C_{1j}[c]	C_{2j}[d]
$1s_{1,\,Li}$	2.521	0.894	-0.128
$1s_{2,\,Li}$	4.699	0.103	-0.004
$2s_{Li}$	0.797	-0.003	0.346
$2p_{z1,\,Li}$	0.737	-0.001	0.176
$2p_{z2,\,Li}$	1.200	-0.004	0.046
$1s_{1,\,H}$	0.888	0.007	0.601
$1s_{2,\,H}$	1.566	0.000	0.100
$2p_{z,\,H}$	1.376	0.002	0.017

Geometry: $r_{Li\text{-}H} = 3.015$ a.u; Jastrow Constants: a = 0.5, b = 0.5

[a]Note that the parameters in Table 1- 9 are matched with the old "Jan" subroutine, where the STO functions appear with coefficients, and those in Table 10-19 with the new subroutine, where the STO functions appear without coefficients (see Fig. II.1).

[b]From Reynold *et. at.* in Ref. 1.

[c]The linear MO coefficients of ϕ_1.

[d]The linear MO coefficients of ϕ_2.

Table. 2. Trial wavefunction Ψ_{1a2}.

STO	ζ	C_{1j}	C_{2j}
$1s_{1,Li}$	**2.618**[a]	0.894	-0.128
$1s_{2,Li}$	4.699	0.103	-0.004
$2s_{Li}$	**0.749**	-0.003	0.346
$2p_{z1,Li}$	**1.153**	-0.001	0.176
$2p_{z2,Li}$	1.200	-0.004	0.046
$1s_{1,H}$	**0.924**	0.007	0.601
$1s_{2,H}$	1.566	0.000	0.100
$2p_{zH}$	**1.499**	0.002	0.017
r_{Li-H} = 3.015 a.u.;		a = **0.567**	b = **1.003**

[a]The bold numbers are what optimized. Same as in Table 3-6.

Table 3. Trial wavefunction Ψ_{1b1}.

STO	ζ	C_{1j}	C_{2j}
$1s_{1,Li}$	**2.610**	0.894	-0.128
$1s_{2,Li}$	4.699	0.103	-0.004
$2s_{Li}$	**0.990**	-0.003	0.346
$2p_{z1,Li}$	**1.080**	-0.001	0.176
$2p_{z2,Li}$	1.200	-0.004	0.046
$1s_{1,H}$	**0.951**	0.007	0.601
$1s_{2,H}$	1.566	0.000	0.100
$2p_{zH}$	**1.791**	0.002	0.017
r_{Li-H} = 3.015 a.u.;		a = **0.557**	b = **0.965**

Table 4. Trial wavefunction Ψ_{1b2}.

STO	ζ	C_{1j}	C_{2j}
$1s_{1,Li}$	**2.620**	0.894	-0.128
$1s_{2,Li}$	4.699	0.103	-0.004
$2s_{Li}$	**0.895**	-0.003	0.346
$2p_{z1,Li}$	**0.990**	-0.001	0.176
$2p_{z2,Li}$	1.200	-0.004	0.046
$1s_{1,H}$	**0.933**	0.007	0.601
$1s_{2,H}$	1.566	0.000	0.100
$2p_{zH}$	**1.819**	0.002	0.017
r_{Li-H} = 3.015 a.u.;		a = **0.567**	b = **0.980**

Table 5. Trial wavefunction Ψ_{1b}.

STO	ζ	C_{1j}	C_{2j}
$1s_{1,Li}$	**2.615**	0.894	-0.128
$1s_{2,Li}$	4.699	0.103	-0.004
$2s_{Li}$	**0.943**	-0.003	0.346
$2p_{z1,Li}$	**1.036**	-0.001	0.176
$2p_{z2,Li}$	1.200	-0.004	0.046
$1s_{1,H}$	**0.942**	0.007	0.601
$1s_{2,H}$	1.566	0.000	0.100
$2p_{zH}$	**1.805**	0.002	0.017
r_{Li-H} = 3.015 a.u.;		a = **0.562**	b = **0.972**

Table 6. Trial wavefunction Ψ_{1c}.

STO	ζ	C_{1j}	C_{2j}
$1s_{1,Li}$	**2.644**	0.894	-0.128
$1s_{2,Li}$	4.699	0.103	-0.004
$2s_{Li}$	**0.787**	-0.003	0.346
$2p_{z1,Li}$	**0.954**	-0.001	0.176
$2p_{z2,Li}$	1.200	-0.004	0.046
$1s_{1,H}$	**0.937**	0.007	0.601
$1s_{2,H}$	1.566	0.000	0.100
$2p_{zH}$	**1.764**	0.002	0.017

Jastrow constants: a_p = **2.149**, b_p = **4.843**, a_a = **0.517**, b_a = **0.782**

Geometry: r_{Li-H} = **3.072** a.u.

Table 7. Trial wavefunction Ψ_{2a1}.

STO	ζ	C_{1j}	C_{2j}
$1s_{1,\,Li}$	2.6682	**0.894**[a]	-0.0544
$1s_{2,\,Li}$	4.4067	0.1119	-0.0146
$2s_{Li}$	0.9194	-0.0036	0.1719
$2p_{z1,\,Li}$	1.2334	-0.0008	0.0691
$2p_{z2,\,Li}$	1.0092	-0.0048	0.0409
$1s_{1,\,H}$	0.9684	0.0284	**0.601**
$1s_{2,\,H}$	1.4567	**0.000**	0.0455
$2p_{z,\,H}$	1.2130	0.0009	0.0622

$r_{Li\text{-}H} = 3.044$ a.u; $a_a = b_a = 0$, $a_p = 0.4963$, $b_p = 0.7132$

[a]The bold numbers are what fixed in the optimization. Same as in Table 8.- 19.

Table 8. Trial wavefunction Ψ_{2a2}.

STO	ζ	C_{1j}	C_{2j}
$1s_{1,\,Li}$	2.6658	**0.894**	-0.1715
$1s_{2,\,Li}$	4.3453	0.1190	-0.0006
$2s_{Li}$	0.8870	-0.0035	0.3221
$2p_{z1,\,Li}$	1.0999	-0.0026	0.0264
$2p_{z2,\,Li}$	1.3650	-0.0049	0.0578
$1s_{1,\,H}$	0.9322	0.0095	**0.601**
$1s_{2,\,H}$	1.6219	**0.000**	0.0505
$2p_{z,\,H}$	1.1178	0.0007	0.0592

$r_{Li\text{-}H} = 3.1118$ a.u; $a_a = b_a = 0$, $a_p = 0.4952$, $b_p = 0.7154$

Table 9. Trial wavefunction Ψ_{2a3}.

STO	ζ	C_{1j}	C_{2j}
$1s_{1,\,Li}$	2.6686	**0.894**	-0.1762
$1s_{2,\,Li}$	4.3101	0.1229	-0.0001
$2s_{Li}$	0.8322	-0.0050	0.2913
$2p_{z1,\,Li}$	1.0570	0.0001	0.0187
$2p_{z2,\,Li}$	1.3198	-0.0023	0.0601
$1s_{1,\,H}$	0.9513	-0.0361	**0.601**
$1s_{2,\,H}$	1.6434	**0.000**	0.0471
$2p_{z,\,H}$	1.4110	0.0003	0.0541

$r_{Li\text{-}H} = 3.0956$ a.u; $a_a = b_a = 0$, $a_p = 0.4978$, $b_p = 0.7016$

Table 10. Trial wavefunction Ψ_{2a4}.

STO	ζ	C_{1j}	C_{2j}
$1s_{1,\,Li}$	2.6309	**3.8912**	**-0.7465**
$1s_{2,\,Li}$	3.9572	1.5782	-0.0165
$2s_{Li}$	1.0258	0.0047	0.1423
$2p_{z1,\,Li}$	2.8027	-0.2655	0.0954
$2p_{z2,\,Li}$	1.4672	0.0127	-0.2092
$1s_{1,\,H}$	1.0729	-0.0086	**0.5409**
$1s_{2,\,H}$	0.4533	-0.0049	0.0340
$2p_{z,\,H}$	1.1836	0.0008	0.0873

Geometry: $r_{Li\text{-}H} =$ **3.015** a.u; Jastrow constants: $a = 0.5044$, $b = 0.7260$

Table 11. Trial wavefunction Ψ_{3fg}.

STO	ζ	C_{1j}	C_{2j}
$1s_{1,\,Li}$	3.0	1.0	0.0
$2s_{Li}$	1.0	0.0	0.3333
$1s_{1,\,H}$	1.0	0.0	0.6667

$r_{Li\text{-}H} = 3.015$ a.u; $a = 1.0$, $b = 1.0$

Table 12. Trial wavefunction Ψ_{2b1}.

STO	ζ	C_{1j}	C_{2j}
$1s_{Li}$	2.9896	**1.0000**	**0.0000**
$2s_{Li}$	1.3511	0.0467	0.1359
$2p_{zLi}$	1.2176	-0.0165	0.1064
$1s_{H}$	0.9169	**0.0000**	**0.6667**
$2p_{zH}$	0.9071	-0.0039	
0.0914			

$r_{Li\text{-}H} =$ **3.015** a.u; $a = 0.4934$, $b = 0.7254$

Table 13. Trial wavefunction Ψ_{2b2}.

STO	ζ	C_{1j}	C_{2j}
$1s_{1,\,Li}$	3.0949	**1.0000**	**0.0000**
$1s_{2,\,Li}$	0.8316	0.0702	-0.1007
$2s_{Li}$	0.7185	-0.0394	0.2284
$2p_{z1,\,Li}$	0.9672	-0.0069	-0.1814
$2p_{z2,\,Li}$	0.6027	0.0111	0.0269
$1s_{1,\,H}$	1.0037	**0.000**	**0.6667**
$1s_{2,\,H}$	1.1509	-0.0944	0.4982
$2p_{z,\,H}$	1.4151	-0.0278	0.1472

$r_{Li\text{-}H}$ = **3.015** a.u; a = 0.5170, b = 0.8500

Table 14. Trial wavefunction $\Psi_{3.1}$.

STO	ζ	C_{1j}	C_{2j}
$1s_{1,\,Li}$	2.9254	**1.0**	**0.0**
$2s_{Li}$	0.7837	0.0015	0.1292
$1s_{1,\,H}$	1.0383	**0.0**	**0.6667**

$r_{Li\text{-}H}$ = **3.015** a.u; a = 0.5077, b = 0.7055

Table 15. Trial wavefunction $\Psi_{3.2}$.

STO	ζ	C_{1j}	C_{2j}
$1s_{Li}$	3.1503	**1.0000**	**0.0000**
$2s_{Li}$	2.3372	0.1566	-0.0970
$2p_{zLi}$	2.8418	-0.0650	-0.2360
$1s_{H}$	0.9131	**0.0000**	**0.6667**
$2p_{zH}$	1.0062	-0.0006	0.1156

$r_{Li\text{-}H}$ = **3.015** a.u; a = 0.4859, b = 0.6952

Table 16. Trial wavefunction $\Psi_{3.3}$.

STO	ζ	C_{1j}	C_{2j}
$1s_{1, Li}$	3.0673	**1.0000**	**0.0000**
$1s_{2, Li}$	0.9825	0.0541	-0.1501
$2s_{Li}$	1.0091	-0.0201	0.2265
$2p_{z1, Li}$	2.9886	-0.0700	-0.1306
$2p_{z2, Li}$	0.9361	0.0020	-0.1922
$1s_{1, H}$	0.9890	**0.0000**	**0.6667**
$1s_{2, H}$	1.7579	-0.0050	0.0838
$2p_{z, H}$	1.3532	-0.0033	0.1246

$r_{Li\text{-}H}$ = **3.015** a.u; a = 0.4945, b = 0.6780

Table 17. Trial wavefunction $\Psi_{3.4}$.

STO	ζ	C_{1j}	C_{2j}
$1s_{1,\,Li}$	3.0714	**1.0000**	**0.0000**
$1s_{2,\,Li}$	0.9697	0.0581	-0.1570
$2s_{Li}$	1.0731	-0.0242	0.2323
$2p_{z1,\,Li}$	3.0817	-0.0706	-0.1183
$2p_{z2,\,Li}$	0.9809	0.0008	-0.2395
$1s_{1,\,H}$	0.9628	**0.0000**	**0.6667**
$1s_{2,\,H}$	1.8832	0.0005	0.0987
$2p_{z,\,H}$	1.3216	-0.0002	0.1333

$r_{Li\text{-}H}$ = **3.015** a.u; a = 0.4972, b = 0.6973

Table 18. Trial wavefunction $\Psi_{3.5}$.

STO	ζ	C_{1j}	C_{2j}
$1s_{1,\,Li}$	3.1809	**1.0000**	**0.0000**
$1s_{2,\,Li}$	0.8873	0.1332	-0.2162
$2s_{Li}$	1.1408	-0.0949	0.3163
$2p_{z1,\,Li}$	3.2089	-0.0765	-0.0540
$2p_{z2,\,Li}$	1.1805	0.0002	-0.3174
$1s_{1,\,H}$	0.8306	**0.0000**	**0.6667**
$1s_{2,\,H}$	1.6294	-0.0002	0.2549
$2p_{z,\,H}$	1.1645	-0.0003	0.1332
$3d_{zLi}$	0.3121	0.0000	-0.0001

$r_{Li\text{-}H}$ = **3.015** a.u; a = 0.4988, b = 0.7271

Table 19. Trial wavefunction $\Psi_{3.6}$.

STO	ζ	C_{1j}	C_{2j}
$1s_{Li}$	3.1503	**1.0000**	**0.0000**
$3s_{Li}$	1.5392	-0.0026	0.0907
$2s_{Li}$	2.3372	0.1566	-0.0970
$2p_{zLi}$	2.8418	-0.0650	-0.2360
$3p_{zLi}$	1.6877	0.0026	-0.2889
$1s_{H}$	0.9131	**0.0000**	**0.6667**
$2s_{H}$	1.6008	-0.0008	-0.1005
$2p_{H}$	1.0062	-0.0006	0.1156
$3d_{zLi}$	3.0918	0.0375	0.3296

$r_{Li\text{-}H}$ = **3.015** a.u; a = 0.4859, b = 0.6592

References

[1] P. J. Reynolds, D. M. Ceperley, B. J. Alder, and W. A. Lester, Jr., *J. Chem. Phys.*, **77**, 5593(1982).

[2] M. H. Kalos, D. Levesque, and L. Verlet, *Phys. Rev. A* **9**, 2178(1974).

[3] R. C. Grimm and R. G. Storer, *J. Comp. Phys.*, **7**, 134(1971).

[4] F. Mentch and J. B. Anderson, *J. Chem. Phys.*, **74**, 6307(1981).

[5] M. H. Kalos, Monte Carlo Methods in Quantum Problems, 1984, p.19-31.

[6] J. W. Moskowitz , K. E. Schmidt, M. A. Lee, and M. H. Kalos, *J. Chem. Phys.* **77**, 349(1982).

[7] D. M. Ceperley and B. J. Alder, *J. Chem. Phys.*, **81**, 5833(1984).

[8] P. J. Reynolds, R. N. Barnett, and W. A. Lester, Jr., *Int. J. Quantum Chem. Symp.* **18**, 709(1984).

[9] P. J. Reynolds, M. Dupuis, and W. A. Lester, Jr., *J. Chem. Phys.*, **82**, 1983(1985).

[10] R. N. Barnett, P. J. Reynolds, and W. A. Lester, Jr., *J. Chem. Phys.*, **82**, 2700(1985).

[11] R. N. Barnett, P. J. Reynolds, and W. A. Lester, Jr., *J. Chem. Phys.* **84**, 4992(1986).

[12] P. J. Reynolds, R. N. Barnett, B. L. Hammond, R. M. Grimes, and W. A. Lester, Jr., *Int. J. Quantum Chem.*, **29**, 589(1986).

[13] R. J. Harrison and N. C. Handy, *Chem. Phys. Lett.*, **113**, 257(1985).

[14] J. W. Moskowitz and M. H. Kalos, *Int. J. Quantum Chem.*, **20**, 1107(1981).

[15] J. W. Moskowitz, K. E. Schmidt, M. A. Lee, and M. H. Kalos, *J. Chem. Phys.*, **76**, 1064(1982).

[16] John P. Lowe, Quantum Chemistry, Orlando, 1978, p.145.

[17] R. Daudel, G. Leroy, D. Peeters, and M. Sana, Quantum Chemistry, New York, 1983, p. 50.

[18] B. H. Wells, Methods in Computational Chemistry, *Vol. 1*, New York, 1987, p. 311.

[19] D. A. Legare, S. M. Rothstein, and J. Vrbik, "Quantum Montum Monte Carlo Studies of LiH: Ground State Physical Properties", preprint.

[20] C. J. Umrigar, K. G. Wilson, and J. W. Wilkins, *Phys. Rev. Lett.* **60**, 1719(1988).

[21] D. Ceperley and B. Alder, *Science*, **231**, 555(1986).

[22] S. M. Rothstein and J. Vrbik, *J. Compt. Phys*, **74**, 127(1988).

[23] J. Vrbik and S. M. Rothstein, *Int. J. Quantum. Chem.*, **29**, 461(1986).

[24] J. Vrbik and S. M. Rothstein, *J. Compt. Phys*, **63**, 130(1986).

[25] S. M. Rothstein, N. Patil, and J. Vrbik, *J. Comput. Chem.*, **8**, 412(1987).

[26] IMSL Lib., The IMSL Library, User's Manual, **4**, Houston, 1984, p.ZXSSQ-1.

[27] K. M. Brown, and J. E. Dennis, *Numerische Mathematik.*, **18**, 289(1972).

[28] K. Levenberg, *Quart. Appl. Math.*, **2**, 164(1944).

[29] D. W. Marquardt, *J. Siam.*, **11 (2)**, (1963).

[30] M. F. DePasquale, S. M. Rothstein, and J. Vrbik, *J. Compt. Phys*, **89**, 3629(1988).