HISTOGRAM FILTERING AS A TOOL IN VARIATIONAL MONTE CARLO OPTIMIZATION

By

Martin Šnajdr

B. Sc. (Physics) Faculty of Mathematics and Physics of Comenius University in Bratislava, Slovakia

A THESIS SUBMITTED IN PARTIAL FULFILLMENT OF

THE REQUIREMENTS FOR THE DEGREE OF

MASTER OF SCIENCE

in

THE FACULTY OF MATHEMATICS AND SCIENCE

DEPARTMENT OF PHYSICS

We accept this thesis as conforming

to the required standard

.....

.....

.....

.....

BROCK UNIVERSITY

1999

In presenting this thesis in partial fulfilment of the requirements for an advanced degree at Brock University, I agree that the Library shall make it freely available for reference and study. I further agree that permission for extensive copying of this thesis for scholarly purposes may be granted by the head of my department or by his or her representatives. It is understood that copying or publication of this thesis for financial gain shall not be allowed without my written permission.

Department of Physics
Brock University
500 Glenridge Avenue
St. Catharines, Ontario
Canada L2S 3A1

Date:

Abstract

Optimization of wave functions in quantum Monte Carlo is a difficult task because the statistical uncertainty inherent to the technique makes the absolute determination of the global minimum difficult. To optimize these wave functions we generate a large number of possible minima using many independently generated Monte Carlo ensembles and perform a conjugate gradient optimization. Then we construct histograms of the resulting nominally optimal parameter sets and "filter" them to identify which parameter sets "go together" to generate a local minimum. We follow with correlated-sampling verification runs to find the global minimum. We illustrate this technique for variance and variational energy optimization for a variety of wave functions for small systems. For such optimized wave functions we calculate the variational energy and variance as well as various non-differential properties. The optimizations are either on par with or superior to determinations in the literature. Furthermore, we show that this technique is sufficiently robust that for molecules one may determine the optimal geometry at the same time as one optimizes the variational energy.

Table of Contents

List of Tables

List of Figures

Acknowledgement

First of all, I would like to thank my supervisor and teacher professor Stuart M. Rothstein. He provided a great help and guidance. I would also like to thank the members of my graduate committee for their comments on my research. I am also grateful to my fellow student Jason R. Dwyer for the numerous discussions we had.

Also, I would like to thank Mr. Bernard Harrison for his support in form of the Thompson-Harrison Graduate Scholarship, and Mr. and Mrs. Penner who supported me with the Edgar and Irmgard Penner Scholarship.

Finally, I gratefully acknowledge a generous grant of computer time at the University of Alberta, without which this work would not have been possible.

Chapter 1

Variational Monte Carlo

Variational Monte Carlo (VMC) is a method which allows one to obtain an approximate analytical solution of the time independent Schrödinger equation for one or multi-electron systems. In our case the system is either a single atom or a diatomic molecule, and we are interested only in the ground state. Throughout we assume the so-called Born-Oppenheimer approximation, i.e., we treat the nuclei as infinitely heavy. The theoretical basis for VMC is the variational theorem and the numerical method used for evaluating many dimensional integrals is Monte Carlo (MC).

1.1 Variational theorem

Suppose we have Hamiltonian $\hat{H}$ and are looking for the solution of the time-independent Schrödinger equation

$$\hat{H}\Psi_G = E_G\Psi_G,$$

where Ψ_G is the ground state wave function and E_G is the ground state energy; i.e., the lowest eigenvalue of the Hamiltonian.

The Variational theorem states that the E_G is a lower bound to the average energy for any possible physical state Ψ. The proof can be done as follows: First, we expand the wave function Ψ in a series of orthonormal eigenfunctions $\Psi_1, \Psi_2, \ldots$ of $\hat{H}$ ($\hat{H}\Psi_i = E_i\Psi_i$)

$$\Psi = \sum_i c_i\Psi_i \ .$$

Then

$$\begin{aligned}\langle E\rangle_\Psi &= \langle\Psi|\hat{H}|\Psi\rangle = \sum_{i,j}\langle c_i\Psi_i|\hat{H}|c_j\Psi_j\rangle = \sum_{i,j} c_i^* c_j\langle\Psi_i|\hat{H}\Psi_j\rangle \\ &= \sum_{i,j} c_i^* c_j E_j\langle\Psi_i|\Psi_j\rangle = \sum_{i,j} c_i^* c_j E_j\delta_{ij} = \sum_i |c_i|^2 E_i.\end{aligned}$$

Because $|c_i|^2$ are positive numbers and $E_G \leq E_i$

$$\langle E\rangle_\Psi = \sum_i |c_i|^2 E_i \geq \sum_i |c_i|^2 E_G = E_G\sum_i |c_i|^2 = E_G$$

QED

The variational method is based on this theorem and can be described in following manner.

1. Choose a wave function depending on a set of so-called variational parameters $\bar{\lambda} = (\lambda_1, \lambda_2, \ldots, \lambda_k)$

2. Obtain the multi-variable function

$$E(\bar{\lambda}) = \frac{\langle\Psi(\bar{\lambda})|\hat{H}|\Psi(\bar{\lambda})\rangle}{\langle\Psi(\bar{\lambda})|\Psi(\bar{\lambda})\rangle} = \frac{\int_\Omega \Psi^*(\bar{\lambda},\mathbf{R})\hat{H}\Psi(\bar{\lambda},\mathbf{R})d\Omega}{\int_\Omega \Psi^*(\bar{\lambda},\mathbf{R})\Psi(\bar{\lambda},\mathbf{R})d\Omega} \tag{1.1}$$

3. Find the parameters corresponding to the global minimum of $E(\bar{\lambda})$, $\bar{\lambda}^*$, and the corresponding energy, $E(\bar{\lambda}^*)$.

The wave function $\Psi(\bar{\lambda}^*)$ is then the so-called variational solution of the Schrödinger equation. We can use it for calculating an arbitrary property of the ground state such as dipole moment, average electron distance etc..

The procedure described above is very easy in theory, but in practice we face various problems in every step.

The first step is crucial. The form of the wave function we choose has to satisfy various quantum mechanical conditions (e.g., it has to be antisymmetric for the fermionic system)

and should incorporate as much qualitative information about the system as possible (e.g., it should reflect the symmetry of the problem). This can often significantly decrease the number of variational parameters and save a lot of computer time.

The problem which arises in the next step is due the fact that with the exception of the simplest systems (and simplest wave functions), the analytical form of $E(\bar{\lambda})$ is not known. Therefore we must use numerical methods for evaluating the energy. This introduces an error which is present in every numerical method. For an N-electron system the corresponding configuration space is $3N$-dimensional. Beginning with $N = 2$ (6-dimensional integral) the best known numerical method is MC.

The last step is to find a global minimum of a multi-variable function. This is a difficult task. Despite much effort no general algorithm is available. There are many completely different approaches, and no one is better than the others in every aspect. To choose a good algorithm (in a sense of reliability and time requirements) for a specific problem is therefore very important. In addition we have another difficulty related to the unavailability of the analytical formula for the $E(\bar{\lambda})$. Because every evaluation of $E(\bar{\lambda})$ is accompanied with an error (which is random in nature for VMC) the function is not a well-defined mathematical object. For example, when we evaluate the function at the same point (in parameter space) two times, we will get two different values. We can imagine such a function as a superposition of the "true" function $E(\bar{\lambda})$ and a random "noise" function $\delta(\bar{\lambda})$. Therefore, such optimization is sometimes called optimization with the presence of noise.

Fortunately, there is a way to circumvent this problem and obtain a continuous and differentiable approximation of the $E(\bar{\lambda})$, so one can use conventional optimization methods.

1.2 Monte Carlo integration

In this section we will explain MC integration, one of the most powerful methods for evaluation of multi-dimensional integrals. We will be particularly interested in evaluation of integrals of the type

$$\int_{\Omega} f(\mathbf{R})\rho(\mathbf{R})d\Omega \ , \tag{1.2}$$

where $\rho(\mathbf{R})$ is a non-negative function with a property

$$\int_{\Omega} \rho(\mathbf{R})d\Omega = 1 \ .$$

Hence we can write it in the form

$$\rho(\mathbf{R}) = \frac{\Psi^2(\mathbf{R})}{\int_{\Omega} \Psi^2(\mathbf{R})d\Omega} \ .$$

This function can be considered a probability density distribution of a random vector $\mathbf{R}$. If $\mathbf{R}$ is sampled from $\rho(\mathbf{R})$, the probability of finding $\mathbf{R}$ in a volume element $d\Omega$ of the vector space $\Re^N$ is $P(\mathbf{R}) = \rho(\mathbf{R})d\Omega$.

By definition, the integral 1.2 can be written as

$$\int_{\Omega} f(\mathbf{R})\rho(\mathbf{R})d\Omega = \lim_{N\to\infty} \sum_{i=1}^{N} f(\mathbf{R}_i)\rho(\mathbf{R}_i)\Delta\Omega_i \ ,$$

where

$$\Delta\Omega_i \subset \Omega, \ \|\Delta\Omega_i\|_{\Re^N} \overset{N\to\infty}{\longrightarrow} 0, \ \bigcup_i \Delta\Omega_i = \Omega \ (\Delta\Omega_i \bigcap \Delta\Omega_j = \emptyset, i \neq j).$$

This can be processed further as follows:

$$\begin{aligned} \lim_{N\to\infty} \sum_{i=1}^{N} f(\mathbf{R}_i)\rho(\mathbf{R}_i)\Delta\Omega_i &= \lim_{N\to\infty} \sum_{i=1}^{N} f(\mathbf{R}_i)P(\mathbf{R}_i) = \lim_{N\to\infty} \frac{1}{N} \sum_{i=1}^{N} f(\mathbf{R}_i)NP(\mathbf{R}_i) \\ &= \lim_{N\to\infty} \frac{1}{N} \sum_{i=1}^{N} f(\mathbf{R}_i)N_i = \lim_{N\to\infty} \frac{1}{N} \sum_{\{\mathbf{R}_i\}_\rho} f(\mathbf{R}_i) \ , \end{aligned}$$

where $N_i = NP(\mathbf{R}_i)$ is the number of samples (later called walkers) in the volume element $\Delta\Omega_i$. In summary, we can write in more conventional notation:

$$\int_\Omega f(\mathbf{R})\rho(\mathbf{R})d\Omega = \frac{\int_\Omega f(\mathbf{R})\Psi^2(\mathbf{R})d\Omega}{\int_\Omega \Psi^2(\mathbf{R})d\Omega} \approx \frac{1}{N}\sum_{\{\mathbf{R}_i\}_{\Psi^2}} f(\mathbf{R}_i) \ ,$$

where $\{\mathbf{R}_i\}_\rho$ denotes a random sequence of vectors $\mathbf{R}_i$ generated from the distribution ρ.

1.3 Metropolis algorithm

In this section we will describe how to generate sequences of random vectors $\mathbf{R}_i$ sampled from arbitrary distribution Ψ^2. The method was first described by Metropolis *et al.* [15] in 1953, hence the name Metropolis algorithm.

First, we give the general description of the algorithm and the appropriate formulas will be derived afterwards. The basic object used in the Metropolis algorithm is called a walker. In general, a walker represents a state of some system, in our case, a point in $3N$-dimensional space $\Re^{3N}$. The algorithm can be described as follows:

1. Create a set (ensemble) of M walkers $\{\mathbf{R}_i^1\}$ randomly placed in the configuration space $\Re^{3N}$ (or for the practical purposes in the domain $\Omega \subset \Re^{3N}$ where the function $\Psi^2(\mathbf{R})$ is non-zero).

2. For each walker propose a step $\mathbf{R}_i^k \to \mathbf{R}_i^{'}$ with probability $T(\mathbf{R}_i^{'} \leftarrow \mathbf{R}_i^k)$ and accept this move with probability $A(\mathbf{R}_i^{'} \leftarrow \mathbf{R}_i^k)$. If the step is accepted $\mathbf{R}_i^{k+1} = \mathbf{R}_i^{'}$, otherwise $\mathbf{R}_i^{k+1} = \mathbf{R}_i^k$.

3. Step 2 completes an iteration. Repeat step 2 until equilibrium is reached. This means that (in the limit $M \to \infty$) in every region of configuration space $\Delta\Omega$ we would find $M \int_{\Delta\Omega} \Psi^2(\mathbf{R})d\Omega / \int_\Omega \Psi^2(\mathbf{R})d\Omega$ walkers. Further iterations retain that property so there is no macroscopic change of the density of walkers. For M finite there are of course statistical fluctuations.

This algorithm has one interesting property. If we take only one walker and move it through the configuration space (iterate it), we would see that the walker visits every region in space and the "time" spent in each region is exactly proportional to Ψ^2. Thus instead of taking the "space" average of function $f(\mathbf{R})$ over (equilibrated) whole ensemble (the sum over index i below), we can take the "time" average following one walker (the sum over index k below) and those results will be equal[1] (in statistical limit). We can write this as follows:

$$\lim_{M\to\infty} \frac{1}{M} \sum_{i=1}^{M} f(\mathbf{R}_i^k) = \lim_{M\to\infty} \frac{1}{M} \sum_{k=1}^{M} f(\mathbf{R}_i^k) \ .$$

Although there is no difference between those two averages in theory, in practice every approach has its advantages and disadvantages. To get a sufficiently accurate estimates of the MC integral one has to use a large number of walkers (often $10^6 - 10^8$). However, to store the position of each walker is impossible (for double precision calculations one needs $8 \times 3N$ bytes for every walker), and equilibration of such ensemble would be extremely time consuming. The problem with taking only one walker and using the "time"average is due the fact that the random walk is correlated, in other words, the $(i+1)$-th position is not independent on the i-th position. This so-called serial correlation decreases the number of effective iterations (often by factor of 10 or more), so the precision due to serial correlation is lower than the theoretical estimate.

In practice, we take the combined average

$$\langle f \rangle = \frac{1}{N_C} \sum_{i=1}^{N_C} \frac{1}{N_I} \sum_{k=1}^{N_I} f(\mathbf{R}_i^k) \ , \tag{1.3}$$

where $\mathbf{R}_i^k$ is the position of i-th walker after k iterations (after equilibrium is reached) and N_C, N_I are the number of configurations (ensemble size) and the number of iterations,

[1] Process which satisfies this property is called ergodic. A necessary (but not sufficient) condition for ergodicity is that there is a non zero probability for the walker to visit every region in configuration space.

respectively. In addition this allows us to estimate the statistical error correctly. More details will be presented in the next section.

Let us now explore the conditions for $T(\mathbf{R}' \leftarrow \mathbf{R})$ and $A(\mathbf{R}' \leftarrow \mathbf{R})$ which result from the equilibrium condition. Because in equilibrium there is no macroscopic "drift" of walkers, the number of walkers which leave some domain has to be equal to the number of incoming walkers from other regions. Suppose our ensemble is already equilibrated, i.e., the walkers are distributed according to Ψ^2. The decrease of the density of walkers at some point $\mathbf{R}'$ is

$$\Delta_-(\mathbf{R}') = \int \Psi^2(\mathbf{R}')T(\mathbf{R} \leftarrow \mathbf{R}')A(\mathbf{R} \leftarrow \mathbf{R}')d\Omega - \Psi^2(\mathbf{R}')T(\mathbf{R}' \leftarrow \mathbf{R}')A(\mathbf{R}' \leftarrow \mathbf{R}')d\Omega \ .$$

The increase of the density of walkers due to walkers incoming from other regions is

$$\Delta_+(\mathbf{R}') = \int \Psi^2(\mathbf{R})T(\mathbf{R}' \leftarrow \mathbf{R})A(\mathbf{R}' \leftarrow \mathbf{R})d\Omega - \Psi^2(\mathbf{R}')T(\mathbf{R}' \leftarrow \mathbf{R}')A(\mathbf{R}' \leftarrow \mathbf{R}')d\Omega \ .$$

The equilibrium condition dictates that $\Delta_-(\mathbf{R}') - \Delta_+(\mathbf{R}') = 0$ which gives

$$\int [\Psi^2(\mathbf{R})T(\mathbf{R}' \leftarrow \mathbf{R})A(\mathbf{R}' \leftarrow \mathbf{R}) - \Psi^2(\mathbf{R}')T(\mathbf{R} \leftarrow \mathbf{R}')A(\mathbf{R} \leftarrow \mathbf{R}')]d\Omega = 0 \ .$$

An obvious way to satisfy the above condition is to require the so-called detailed balance condition

$$\Psi^2(\mathbf{R})T(\mathbf{R}' \leftarrow \mathbf{R})A(\mathbf{R}' \leftarrow \mathbf{R}) = \Psi^2(\mathbf{R}')T(\mathbf{R} \leftarrow \mathbf{R}')A(\mathbf{R} \leftarrow \mathbf{R}') \ ,$$

which yields an expression for the acceptance probability ratio

$$\frac{A(\mathbf{R}' \leftarrow \mathbf{R})}{A(\mathbf{R} \leftarrow \mathbf{R}')} = \frac{\Psi^2(\mathbf{R}')T(\mathbf{R} \leftarrow \mathbf{R}')}{\Psi^2(\mathbf{R})T(\mathbf{R}' \leftarrow \mathbf{R})} \ .$$

The most common form used for $A(\mathbf{R}' \leftarrow \mathbf{R})$ is to set

$$A(\mathbf{R}' \leftarrow \mathbf{R}) = \min(1, \frac{\Psi^2(\mathbf{R}')T(\mathbf{R} \leftarrow \mathbf{R}')}{\Psi^2(\mathbf{R})T(\mathbf{R}' \leftarrow \mathbf{R})}) \ .$$

As we can see one needs to evaluate only the ratio of the density function at two different points. Therefore the density function does not need to be normalized, which is very convenient for practical calculations. Although any form of the transition probability $T(\mathbf{R}' \leftarrow \mathbf{R})$ is theoretically suitable, the efficiency of the Metropolis algorithm depends strongly on that choice. Appendix A deals with the issue of the transition probability choice in more detail.

1.4 Monte Carlo estimators

As we have mentioned in the previous section to calculate the error of a MC estimate $\langle f \rangle$ is not completely trivial. In our work we need to evaluate expressions of the type

$$\langle \mathcal{O} \rangle = \frac{\langle \Psi | \hat{\mathcal{O}} | \Psi \rangle}{\langle \Psi | \Psi \rangle} = \frac{\int \mathcal{O}_L(\mathbf{R}) \Psi^*(\mathbf{R}) \Psi(\mathbf{R}) d\Omega}{\int \Psi^*(\mathbf{R}) \Psi(\mathbf{R}) d\Omega} , \tag{1.4}$$

where $\mathcal{O}_L(\mathbf{R})$ is defined as

$$\mathcal{O}_L = \frac{\hat{\mathcal{O}}\Psi}{\Psi} .$$

If the operator $\hat{\mathcal{O}}$ is the Hamiltonian $\hat{H}$ the quantity $\mathcal{O}_L$ is called local energy and is usually denoted as E_L.

To estimate the expectation value we use the formula 1.3

$$\langle \mathcal{O} \rangle = \frac{1}{N_C} \sum_{i=1}^{N_C} \frac{1}{N_I} \sum_{k=1}^{N_I} \mathcal{O}_L(\mathbf{R}_i^k) . \tag{1.5}$$

To estimate the error of the expectation value $\langle \mathcal{O} \rangle$ one could naively calculate the variance

$$s^2 = \frac{1}{N_C} \sum_{i=1}^{N_C} \frac{1}{N_I} \sum_{k=1}^{N_I} (\mathcal{O}_L(\mathbf{R}_i^k) - \langle \mathcal{O} \rangle)^2 ,$$

and use the standard error formula

$$s_{\langle \mathcal{O} \rangle} = \sqrt{\frac{s^2}{N_C N_I - 1}} .$$

However, the error calculated in such way would be underestimated due the serial correlation of the quantities $\mathcal{O}_L(\mathbf{R}_i^k), \mathcal{O}_L(\mathbf{R}_i^{k+1}) \ldots$. In order to avoid such a bias we can look at the equation 1.5 in a different way

$$\langle \mathcal{O} \rangle = \frac{1}{N_C} \sum_{i=1}^{N_C} \langle \mathcal{O} \rangle_i \ ,$$

where $\langle \mathcal{O} \rangle_i$ is the average local energy taken over the i-th configuration. Because $\mathcal{O}_L(\mathbf{R}_1^k), \mathcal{O}_L(\mathbf{R}_2^k), \ldots, \mathcal{O}_L(\mathbf{R}_{N_C}^k)$ are not correlated the configuration averages $\langle \mathcal{O} \rangle_i$ are not correlated either. We can use now the standard formula and estimate the error of the average

$$s_{\langle \mathcal{O} \rangle} = \sqrt{\frac{s_c^2}{N_C - 1}} \ ,$$

where $s_c^2 = \frac{1}{N_C} \sum_{i=1}^{N_C} (\langle \mathcal{O} \rangle_i - \langle \mathcal{O} \rangle)^2$.

It is useful to have some quantity which measures the strength of the serial correlation. Introduction of such quantity can be motivated by following consideration. Without any correlation we can write

$$\frac{s^2}{N_C N_I - 1} = \frac{s_c^2}{N_C - 1} \ ,$$

and because $N_C, N_I \gg 1$ we can ignore the factor -1 in each denominator. After this we get

$$\frac{s^2}{N_I} = s_c^2 \ .$$

The above relation can be rewritten as

$$T_{corr} \frac{s^2}{N_I} = s_c^2 \ ,$$

where $T_{corr} = N_I s_c^2 / s^2$ is the so-called *correlation time*, and is equal to one when there is no correlation. This quantity may be used to find the optimal choice for the transition probability $T(\mathbf{R} \leftarrow \mathbf{R}')$ which affects the efficiency of the MC algorithm. It is important to note that the correlation time is different for each quantity. It is most commonly calculated (and optimized) for the energy.

1.5 Correlated sampling

Correlated sampling is a technique which makes it possible to estimate the value of a MC integral $\langle f \rangle_{\Phi^2}$ by using an ensemble sampled from a different distribution Ψ^2. The derivation can be done by starting from its definition by performing some simple algebraic manipulations:

$$\frac{\int f\Phi^2 d\Omega}{\int \Phi^2 d\Omega} = \frac{\frac{\int f\frac{\Phi^2}{\Psi^2}\Psi^2 d\Omega}{\int \Psi^2 d\Omega}}{\frac{\int \frac{\Phi^2}{\Psi^2}\Psi^2 d\Omega}{\int \Psi^2 d\Omega}} .$$

We see that the numerator and denominator have exactly the form we need for using MC integration, and we can estimate the MC integral as follows:

$$\langle f \rangle_{\Phi^2} = \frac{\sum_{\{\mathbf{R}_i\}_{\Psi^2}} f(\mathbf{R}_i) w(\mathbf{R}_i)}{\sum_{\{\mathbf{R}_i\}_{\Psi^2}} w(\mathbf{R}_i)} ,$$

where the $w_i = w(\mathbf{R}_i) = \Phi^2(\mathbf{R}_i)/\Psi^2(\mathbf{R}_i)$ are called "weights". The above formula resembles the weighted average of the function f.

If the two distributions are not very different the weights are not far from unity and the estimate is reliable. However, for two significantly different distributions some weights can reach very large values and dominate in the average. In this case the effective size of the ensemble is drastically reduced and the estimate is not reliable anymore. As a measure of the effective ensemble size we introduce a so-called *index* [1] which is defined as follows:

$$index = \frac{(\sum_{i=1}^{N_C} w_i)^2}{\sum_{i=1}^{N_C} w_i^2} .$$

The index ranges from 1 (for one weight dominating) to N_C (all weights are unity).

What are the features of the correlated sampling for which we find this method so appealing ? To answer this let us take the expression for the variational energy 1.1. Suppose our ensemble is equilibrated according to $\Psi^2(\bar{\lambda}_0)$. Using correlated sampling we

can write the variational energy function as

$$E(\bar{\lambda}) = \frac{\sum_{\{\mathbf{R}_i\}_{\Psi^2(\bar{\lambda}_0)}} E_L(\mathbf{R}_i, \bar{\lambda}) w(\mathbf{R}_i, \bar{\lambda})}{\sum_{\{\mathbf{R}_i\}_{\Psi^2(\bar{\lambda}_0)}} w(\mathbf{R}_i, \bar{\lambda})} ,$$

where $w(\mathbf{R}_i, \bar{\lambda}) = \Psi^2(\mathbf{R}_i, \bar{\lambda})/\Psi^2(\mathbf{R}_i, \bar{\lambda}_0)$. The first thing we notice is that we can obtain the variational energy for different parameters $\bar{\lambda}$ without any need of re-equilibration. Second, and perhaps even more important feature is that such defined $E(\bar{\lambda})$ is a smooth function of the parameters $\bar{\lambda}$. This is easy to see because for fixed $\{\mathbf{R}_i\}$ and $\bar{\lambda}_0$ both $E_L(\mathbf{R}_i, \bar{\lambda})$ and $w(\mathbf{R}_i, \bar{\lambda})$ are smooth functions of $\bar{\lambda}$, and the product, sum and ratio of smooth functions is again a smooth function. This is the key property which allows us to use conventional methods for optimizing $E(\bar{\lambda})$. The only problem which could arise is if for some $\mathbf{R}_i$ the corresponding value of $\Psi^2(\mathbf{R}_i, \bar{\lambda}_0)$ is zero. However, this is not possible because the probability of finding a walker in such region is zero.

One immediate application is the numerical calculation of the first derivatives of $E(\bar{\lambda})$ with respect to the individual parameters

$$\frac{\partial E(\bar{\lambda})}{\partial \lambda_i} \approx \frac{E(\lambda_1, \ldots, \lambda_i + h, \ldots, \lambda_k) - E(\lambda_1, \ldots, \lambda_i - h, \ldots, \lambda_k)}{2h} .$$

(The symmetric form for numerical derivatives is chosen for better numerical behavior.) Knowledge of first derivatives is necessary for the conjugate gradient optimization method. In practical applications we prefer using analytical formulas for derivatives, but this is not always possible.

One of the drawbacks of this method is the previously mentioned index problem. If during the optimization we go too far[2] from the initial value $\bar{\lambda}_0$ and the index becomes too small, we have to stop it and start a new one. More about this and other optimization related topics will be presented in the section on optimization.

[2]the meaning of the word *far* is not well defined and is used very loosely here

1.6 Trial wave functions

Quantum mechanics demands that many-body wave functions exhibit certain fundamental properties. For fermionic systems such as electrons, the corresponding wave functions have to be antisymmetric with respect to exchange of any pair of electrons.

Generally, in our work we take

$$\Psi = \Psi_{\mathcal{A}} J \ ,$$

where $\Psi_{\mathcal{A}}$ is completely antisymmetric, and the so-called Jastrow factor J is completely symmetric with respect to the interchange of any pair of coordinates. Hence the product as a whole is completely antisymmetric.

$\Psi_{\mathcal{A}}$ has the form of a Slater determinant

$$\Psi_{\mathcal{A}} = \begin{vmatrix} \Phi_1(1) & \Phi_1(2) & \dots & \Phi_1(n) \\ \Phi_2(1) & \Phi_2(2) & \dots & \Phi_2(n) \\ \vdots & \vdots & \ddots & \vdots \\ \Phi_n(1) & \Phi_n(2) & \dots & \Phi_n(n) \end{vmatrix} , \tag{1.6}$$

where $\Phi_i(j)$ denotes i-th *spin-orbital* as a function of j-th electron coordinates. Our Hamiltonians do not include any spin operators, so the spins of the electrons, once set, do not change in time. This gives us a tool for distinguishing between two types of electrons: spin-up and spin-down.

Motivated by this reasoning we require that the wave functions be antisymmetric only with respect to interchange of electrons of the same spin. As a consequence the Slater

determinant 1.6 reduces to a product of two Slater determinants of *spatial* orbitals

$$\Psi_{\mathcal{A}} = \begin{vmatrix} \phi_1(1) & \phi_1(2) & \dots & \phi_1(n^\uparrow) \\ \phi_2(1) & \phi_2(2) & \dots & \phi_2(n^\uparrow) \\ \vdots & \vdots & \ddots & \vdots \\ \phi_{n^\uparrow}(1) & \phi_{n^\uparrow}(2) & \dots & \phi_{n^\uparrow}(n^\uparrow) \end{vmatrix} \begin{vmatrix} \phi_{n^\uparrow+1}(n^\uparrow+1) & \phi_{n^\uparrow+1}(n^\uparrow+2) & \dots & \phi_{n^\uparrow+1}(n) \\ \phi_{n^\uparrow+2}(n^\uparrow+1) & \phi_{n^\uparrow+2}(n^\uparrow+2) & \dots & \phi_{n^\uparrow+2}(n) \\ \vdots & \vdots & \ddots & \vdots \\ \phi_n(n^\uparrow+1) & \phi_n(n^\uparrow+2) & \dots & \phi_n(n) \end{vmatrix}$$
$$= \Psi_{\mathcal{A}}^{\uparrow}\Psi_{\mathcal{A}}^{\downarrow} ,$$

where $n^\uparrow$ is the number of spin-up electrons, and n is the total number of electrons. This decomposition can be justified mathematically and the proof is given in appendix C.

Let us now have a look at the symmetric part of the wave function—the Jastrow factor J . The purpose of that factor is to explicitly incorporate the electron-electron correlation and sometimes other types of correlations as well. We need a mechanism which would decrease the probability of finding two electrons very near to each other. Slater determinants incorporate the Pauli exclusion principle and prevent two electrons with the same spin from occupying the same region in space (so-called Fermi correlation). However, there is no such mechanism for electrons with opposite spins (so-called Coulomb correlation).

In general

$$J = J(r_{\alpha i}, r_{ij}) ,$$

where r_{ij} is the inter-electronic distance, and $r_{\alpha i}$ is the distance between the α-th nucleus and i-th electron.

In our work we have used several different types of the Jastrow factor. The details will be given in sections dealing with specific systems.

If we look at the expression for the local energy

$$E_L(\mathbf{R}) = \frac{\hat{H}\Psi}{\Psi} = \frac{\hat{T}\Psi}{\Psi} - \sum_{\alpha=1}^{N_a}\sum_{i=1}^{N_e}\frac{Z_\alpha}{r_{\alpha i}} + \sum_{i=1}^{N_e}\sum_{j=i+1}^{N_e}\frac{1}{r_{ij}} ,$$

we see that one can run into difficulties when an electron approaches a nucleus($r_{\alpha i} \to 0$) or when two electrons approach each other ($r_{ij} \to 0$). In the above formula $\hat{T}$ is the kinetic energy operator, N_a (N_e) is the number of nuclei (electrons) in the system. To avoid singularities in the local energy one can derive a set of so-called "cusp" conditions [9] for the variational parameters which cause cancellation of divergent terms $Z_\alpha/r_{\alpha i}$ and $1/r_{ij}$ by similar terms from the kinetic part of the local energy. Cusp conditions can be directly incorporated into the trial function so that it satisfies them for all values of variational parameters. This is not always easy to achieve so the cusp conditions are often checked after optimization as a "quality indicator" since the true wave functions obey the cusp conditions. Our wave functions (except Ψ_2 for He) do not explicitly incorporate the cusp conditions.

1.7 Optimization

By optimization we mean the procedure of finding the best set of variational parameters. However we need to explain the meaning of the word "best".

There are basically two criteria used to optimize a wave function. The first, and most straightforward, is to directly minimize the energy, or to be more precise the MC estimate of the energy 1.5. The second possibility is to minimize the MC estimate of the variance of the local energy

$$s^2_{E_L} = \frac{\sum_{\{\mathbf{R}_i\}_{\Psi^2(\bar{\lambda}_0)}} [E_L(\mathbf{R}_i, \bar{\lambda}) - E_T]^2 w(\mathbf{R}_i, \bar{\lambda})}{\sum_{\{\mathbf{R}_i\}_{\Psi^2(\bar{\lambda}_0)}} w(\mathbf{R}_i, \bar{\lambda})} ,$$

where E_T is our best guess for the ground state energy. This optimization is not very sensitive to the choice of E_T. Common practice is to choose its value below the expected ground state energy which corresponds to the minimization of a combination of the variance and energy. In the ideal case, i.e., if there is a set of variational parameters for which the trial wave function becomes the true eigenfunction of the Hamiltonian, those

two methods lead to the same result — the true eigenstate. But in practice that is almost never the case and those methods give different results.

Many people prefer to optimize the variance for several reasons (see for example [23]). The lower limit of the variance is known—it is zero. Furthermore, minimizing the variance resembles least squares fitting for which powerful algorithms were developed. We decided to optimize both and compare the results afterwards.

As we explained before, correlated sampling gives a smooth approximation of the $E(\bar{\lambda})$. The larger the ensemble size the better the approximation is. An immediate conclusion might be to take a sufficiently large ensemble and do one optimization. However this rather naive approach has several drawbacks;

1. As the dimensionality of the configuration space increases the accuracy of the MC integral decreases. Therefore, we do not know in advance what number of configurations is sufficient.

2. Having only one optimal set of variational parameters we can not say how much each parameter contributes to the energy, i.e., if a small change of a parameter leads to a small or a large change in energy. This kind of information helps us better understand the qualitative characteristics of the trial wave function.

3. We might get trapped in one of many local minima and there is no way of realizing it.

There is another more or less technical problem related to the need for large memory to store all the configuration positions. But this issue is becoming less and less important nowadays when memory prices are falling rapidly and computers with hundreds and thousands MB of memory are commonly available.

The basic rule in experimental physics is to repeat measurements of some properties of some physical system as many times as possible. This allows one to use statistics for

processing the data and to get reliable estimates of the error of the measurement. It also decreases the probability of accidentally measuring some random fluctuation of the system without noticing it. Such data points known as outliers can be ignored.

Our approach is very similar: we repeat the optimization many times and use statistics to estimate the value for each parameter at the global optimum. Of course, if we used the same ensemble for every optimization we would arrive at the same minimum every time (using a deterministic minimization algorithm). So we re-equilibrate the ensemble between optimizations which randomly changes the function $E(\bar{\lambda})$. This also reduces the risk of being trapped in local minima.

We have to realize that the local minima are shallow and even a slight change of the function can help us out. There is another big advantage of this approach. The optimizations are completely independent and can be run in parallel. For example, if we have ten processors available, the time necessary for collecting all the data is reduced by factor of ten. Our task is naturally parallel, and no knowledge of parallel programming is necessary[3].

The optimization method used here can be summarized as follows:

1. Choose the best estimate of the variational parameters $\bar{\lambda}_0$ and equilibrate the ensemble according $\Psi^2(\bar{\lambda}_0)$.

2. Start from the point $\bar{\lambda}_0$ and find from a standard algorithm the minimum of the $E(\bar{\lambda})$, $\bar{\lambda}_i^*$.

3. Re-equilibrate the ensemble (again according the $\Psi^2(\bar{\lambda}_0)$) .

[3]The trend in computer industry is in parallelization. One way is to increase the number of processors in modern supercomputers and speed the communication among them. Other (and perhaps even more promising for certain tasks) is to use the tremendous computer power distributed in the Internet or even in a local university computer network. There are hundreds and sometimes thousands of computers in every university which are most of the time used for e-mail and word processing, which is like doing nothing for the processor. Experiments involving thousands of computers used for breaking the standard encryption algorithms were very impressive and led to the change of industrial security standards.

4. Repeat steps 2-3 until a sufficient number of $\bar{\lambda}_i^*$ is collected.

5. Process the collected data using the "histogram filtering" method (which will be described later).

The above description needs several comments.

To come up with a good value of the starting parameters $\bar{\lambda}_0$ is not an easy task. If we do not have any prior knowledge of the parameter values we just choose some reasonable numbers. In such case we assign the SCF orbital exponents and linear coefficients to the Slater part of the wave function, and set the Jastrow parameters to zero. Similarly, we should assign some reasonable value to the bond distance parameter. After performing a short series of optimizations, normally we get much better estimates of the starting parameters. Then, starting from the new point, we perform several other optimization runs (the number depends on how fast it converges). The usual procedure is to take the averages of the nominally optimal parameters from those runs and set those as the starting point for the main optimization. To get as close as possible to the minimum is advantageous not only for the greater accuracy of the optimization, but the time needed for convergence is decreased also. That is an important factor when several hundred optimizations are performed.

It should be pointed out that for complicated wave functions (such as Ψ_2 for LiH) it can happen that the convergence to the optimum is too slow or the number of successful[4] runs is very small. In this case we can try to do a "staged" optimization which means that in each stage we fix either the Slater determinant or the parameters in Jastrow. In the next stage we fix the previously optimized part and optimize the previously fixed part. This procedure can be applied several times until the parameters do not change

[4]By "successful" we consider those runs which converge to the optimum. It can happen that we must prematurely stop some optimizations because of the index problem or variance being too large (in the case of energy optimization). Occasionally the conjugate gradient exceeds the preset maximum number of iterations.

too much or the energy (or variance) does not drop significantly. In the last stage we can try to optimize all parameters. The above procedure is most efficient when the bond distance is fixed, because in that case we do not need to recalculate the "fixed" part (determinant or Jastrow) of the wave function during the optimization.

If we still have trouble optimizing all the parameters we have to decrease their number by fixing some of those which do not change too much or which contribute relatively little to the variational energy.

We applied the "staged optimization" scheme to the fixed geometry energy optimization of LiH molecule (Ψ_2). For geometry optimization of this molecule we fixed the Slater determinant[5].

If our trial function is based on some previously optimized trial function with known optimal values we can simply use them as our starting point.

The standard optimization algorithm of choice here is the conjugate gradient method (see appendix B). This method is deterministic and requires the first derivatives of the given function.

[5]In the beginning we have done about 30 optimizations of all parameters and we set the determinant parameters to the average values taken from those 30 runs. Thus in some sense the determinant was (at least partially) optimized.

Chapter 2

Examples of Histogram Filtering Method

In this chapter we illustrate the procedure of histogram filtering in some detail. All the numbers reported from this chapter on are in atomic units[1] The arrows in histograms indicate the values of initial parameters (lower arrow) and optimal parameters (upper arrow). If the lower arrow is missing it means that the value is out of the range. The arrows in the energy and variance histograms indicate the appropriate optimal values obtained from verification runs.

2.1 The hydrogen molecule ion

As a first example we take the energy-optimized H_2^+ molecule ion with variable geometry. The trial wave function has the simple form

$$\Psi_T = e^{-kr_a} + e^{-kr_b} ,$$

where r_a , r_b are the electron-nuclear distances, k is the orbital exponent and it is the first variational parameter. The distance between the nuclei R is the second variational parameter. For this system the variational energy $E(k, R)$ can be calculated analytically

$$E(k,R) = -\frac{1}{2}k^2 + \frac{k^2 - k - 1/R + (1+kR)e^{-2kR}/R + k(k-2)(1+kR)e^{-kR}}{1 + (1 + kR + k^2R^2/3)e^{-kR}} + \frac{1}{R}$$

and it is easy to determine the optimal values: $k = 1.23803$ and $R = 2.0033\ a_0$ with the corresponding energy $E = -0.5865065\ E_h$[2].

[1] The unit of energy is called hartree (E_h), 1 E_h = 27.2114 eV. The unit of length is called bohr (a_0), 1 a_0 = 0.529177 Å

[2] $E_{exact} = -0.6026\ E_h$

We performed both energy and variance optimizations. The variance optimization yields very poor results with R_{opt} over a factor of two larger and the accompanying variational energy uncompetitive ($E_{opt} = -0.528269\ E_h$). As we shall see later this is generally the case for variance optimization using a crude wave functions. The histograms are shown in figure 2.1.

Next, we discuss the energy optimization. To find out how the optimization depends on the choice of initial parameters, we started from three different initial values of variational parameters. Figures 2.2–2.4 show histograms for variational parameters k, R and energy at the minimum for those three different choices of initial variational parameters. The histograms shown are already free of outliers. In this case the number of outliers was small. Results are given in table 2.1. The variational energy E_{opt} is calculated from the analytic formula for the corresponding optimal parameters R_{opt}, k_{opt}. Each trial produced good agreement with the analytic values for E. For each choice of initial parameters we did 500 optimization runs. The ensemble was re-equilibrated 10 times after every optimization. This number is sufficient for the next optimization to be independent of the previous one.

k_I	$R_I/2$	k_{opt}	Δk_{opt}	$R_{opt}/2$	$\Delta(R_{opt}/2)$	E_{opt}	ΔE
0.8	1.5	1.2424(9)	4.4E-3	1.004(2)	2.4E-3	-0.5864891	1.7E-5
1.0	1.5	1.2409(8)	2.9E-3	1.003(1)	1.4E-3	-0.5864994	7E-6
1.23	1.0	1.2387(4)	7E-4	1.0021(7)	5.5E-4	-0.5865060	5E-7

Table 2.1: Variational parameters for H_2^+ molecule ion, energy optimization. R_I,k_I are the initial parameters, R_{opt}, k_{opt} are the optimal parameters.

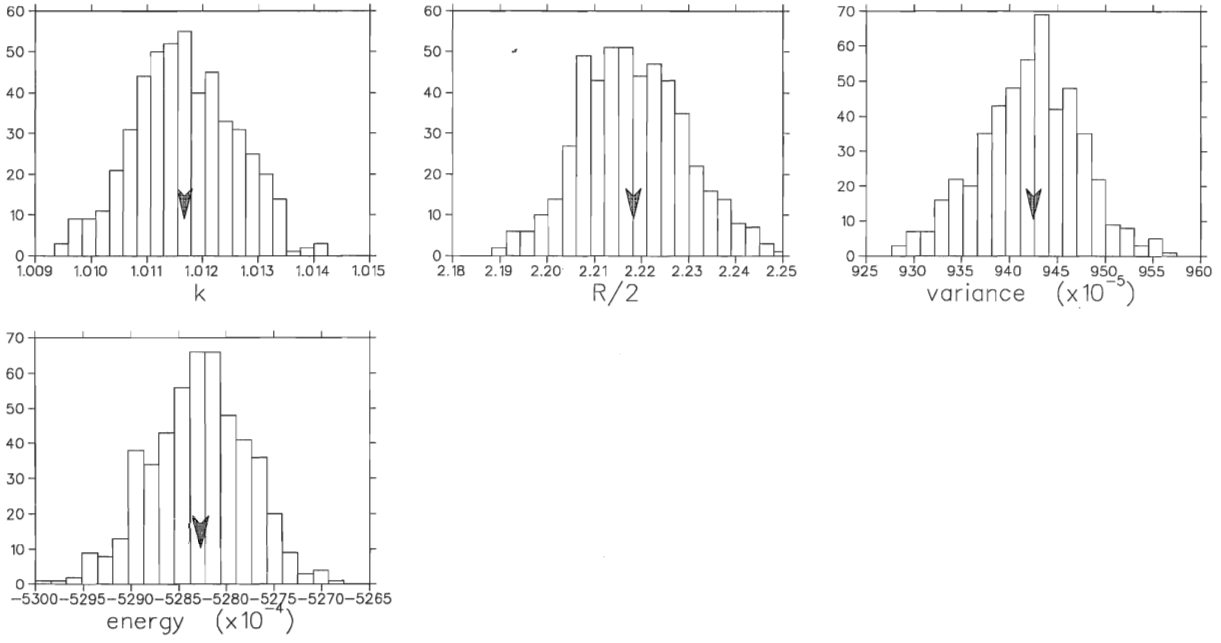

Figure 2.1: Histograms for H_2^+ molecule, variance optimization.

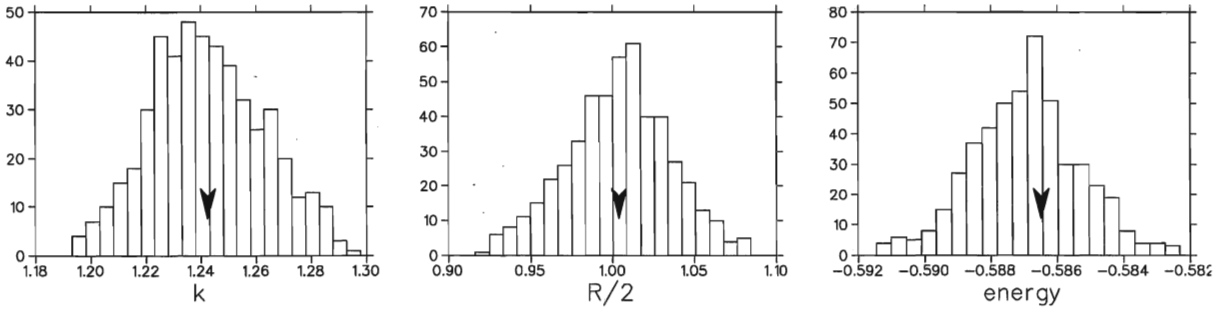

Figure 2.2: Histograms for H_2^+ molecule, $k_I = 0.8$, $R_I/2 = 1.5$.

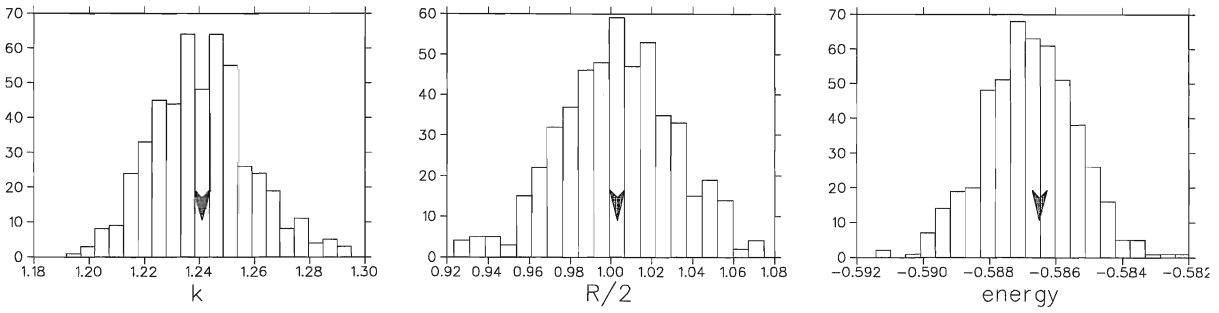

Figure 2.3: Histograms for H_2^+ molecule, $k_I = 1.0$, $R_I/2 = 1.5$.

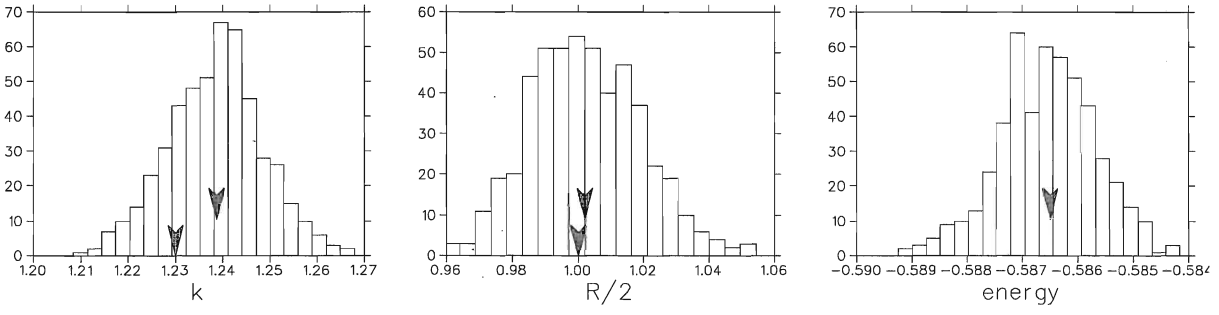

Figure 2.4: Histograms for H_2^+ molecule, $k_I = 1.23$, $R_I/2 = 1.0$.

2.2 Description of histogram filtering

In the previous section we have seen histograms consisting only of one single Gaussian-like peak for every parameter. In the general case, however, the structure of histograms is much more complicated, and we often see several (many times overlying) peaks in histograms for some variational parameters. This is very probably due to existence of several local minima.

Our data have the following form:

$$\begin{array}{l} \lambda_1^1,\ \lambda_2^1,\ \ldots\ ,\ \lambda_k^1 \\ \lambda_1^2,\ \lambda_2^2,\ \ldots\ ,\ \lambda_k^2 \\ \vdots \\ \lambda_1^i,\ \lambda_2^i,\ \ldots\ ,\ \lambda_k^i \\ \vdots \\ \lambda_1^n,\ \lambda_2^n,\ \ldots\ ,\ \lambda_k^n\ , \end{array} \tag{2.1}$$

where λ_j^i is the value of the j-th parameter in the i-th optimization. We can introduce a set of conditions for the variational parameters—the "filter"

$$\begin{array}{c} \lambda_{\min_1} < \lambda_1^i < \lambda_{\max_1} \\ \lambda_{\min_2} < \lambda_2^i < \lambda_{\max_2} \\ \vdots \\ \lambda_{\min_k} < \lambda_k^i < \lambda_{\max_k}\ . \end{array} \tag{2.2}$$

Through this so-called direct filter can *pass* only those parameter sets for which *all* the above conditions are satisfied. If *at least* one condition is invalid the set does not pass through the filter. There is another possible mode of filtering—"reverse" filtering. In this mode we *discard*[3] sets for which *all* the above conditions hold true. Thus if *at least* one condition is invalid the set *passes* through.

[3]Of course we do not physically destroy our data.

Each type of filtering is used for different purpose. In our work we have used almost always direct filtering which for now on is simply referred to as "filtering". It is a powerful tool for isolating the different local minima, i.e., for finding which peaks are mutually connected. As a result of the filtering we obtain several "candidates" for the global minimum. In many cases we can judge the candidates simply by looking at the energy (or variance) histogram of each candidate. For some of them there is often a visible shift of their energy (or variance) peak towards higher or lower values. This can be used as a first step in the process of looking for the best candidate.

To decide which minimum corresponds to the lowest energy (or variance) we have to calculate the energy (or variance) for each candidate and choose the best one. Because the energy (or variance) differences are usually very small this would require much computer time. Therefore we use correlated sampling which is much more sensitive to differences, and calculate the corresponding energies (or variances) for every candidate in one run. After the best candidate is selected, the energy and other properties are calculated using standard MC codes.

In case of energy optimization one has to be more careful with using the comparison based on the energy shifts. Many times the energy shift to the left (towards lower values) is accompanied with variance shift to the right (towards larger values). We could say that energy "exploits" the large variance to get low energy values.

How this works in practice will be seen in next section, where we will employ this technique for isolating the minima for system with a more complicated histogram structure.

2.3 The hydrogen molecule

In this section we demonstrate histogram filtering method on more complicated system—variance-optimized Ψ_3 wave function with variable geometry (see section 3.1 for details). We have used the same starting point for variance optimization as for energy optimization. This is not by far the best choice, but the optimizations (at least in this case) do not strongly depend on the choice of the starting point.

The wave function Ψ_3 has formally 12 parameters — 6 in the Slater determinant, 5 in the Jastrow factor, and one is the bond distance. Without any loss of generality we can fix one linear coefficient[4] in the molecular orbital expansion (c_1), since only the ratios of the linear coefficients are important. Thus we have together 11 free adjustable variational parameters.

Figure 2.5 shows the histograms after completing 320 optimizations. We immediately notice the very visible and well-separated double peaks for the ζ_{1s}, ζ_{2s}, c_2, c_3 and somewhat less obvious but still visible double peak structure for g_3, g_4, $R/2$ and the variance itself. To determine which peak goes with which we can use a simple direct filter for any of the "double-peaked" parameters. We have chosen c_3, and after applying the filter $c_3 < 0.09$ we get the histograms 2.6. Keeping the other peak ($0.1 < c_3$) we get 2.7. The scales were deliberately kept unchanged to make the comparison easier. In this particular example we can be certain that the 2.6 corresponds to the lower variance (there is a significant shift in the variance histogram) and in the following we will concentrate only on it.

In order to see more details it is convenient to re-scale. The figure 2.8 shows the same histograms as 2.6, only with different scales. To get rid of the outliers we set the (direct)

[4]If we did not fix this parameter there would be infinite equivalent parameter sets of optimal variational parameters and this would result in an ambiguous minimum and the optimization wouldn't be reliable.

filter:

$$\begin{array}{rcccl} & & \zeta_{1s} & < & 0.833 \\ & & \zeta_{2s} & < & 0.99 \\ & & d & < & 0.4 \\ -0.5 & < & g_2 & & \\ -0.9 & < & g_4 & & \\ 0.722 & < & R/2 & < & 0.829 \\ & & \text{variance} & < & 0.0115\ . \end{array} \tag{2.3}$$

Each of the above conditions usually serves to filter out only a few outliers.

The resulting histograms are shown in figure 2.9. The most obvious structure there is the double peak for g_3 and g_4. Keeping the left peak for g_4 ($-0.859 < g_4 < -0.191$) and filtering out couple of outliers we obtain the histograms 2.10. Taking the average for every histogram we get our first candidate (see table 2.2). We report three digits but the error for some parameters (such as g's) is already at the second decimal place.

We can go even further and try to separate the possible two-peak structure for ζ_{2s}. The filter would be $\zeta_{2s} < 0.96$ for the left peak, and $0.96 < \zeta_{2s}$ for the right one. The corresponding histograms are shown in figures 2.11 and 2.12. Taking the average for each histogram we get second and third candidates (table 2.2). To filter further is useless because now we have only 15 parameter sets and every other filtering would decrease the number even more.

Let us now return back to the figure 2.9 and isolate the right peak for g_4 ($-0.191 < g_4$). The result is on the figure 2.13. This yields the fourth and last candidate.

To check which candidate is the best, we performed a correlated sampling verification run. Thus although the true variances are related to the reported sigmas, the relative values, i.e., the ordering should be correct even if two values are within the statistical

error. The results appear in table 2.2. The second candidate has the lowest variance thus we report its parameter values as the optimal ones.

Unfortunately, we do not have always such a nice structure with so well-separated peaks. In that case the filtering is more complicated and not as straightforward. We always try to find the parameters with most separated peaks and start filtering there. If we have more possibilities, we can start from different parameters and in the end we compare all the candidates in the correlated sampling run.

	1	2	3	4
ζ_{1s}	0.793	0.782	0.800	0.804
c_1	1.0	1.0	1.0	1.0
ζ_{2s}	0.962	0.953	0.968	0.968
c_2	-0.556	-0.563	-0.551	-0.543
ζ_{2p_z}	1.686	1.670	1.697	1.684
c_3	0.073	0.073	0.073	0.0748
d	0.309	0.321	0.301	0.251
g_1	0.820	0.794	0.838	0.983
g_2	-0.180	-0.163	-0.192	-0.241
g_3	0.504	0.539	0.479	0.107
g_4	-0.453	-0.472	-0.439	-0.062
$R/2$	0.777	0.792	0.766	0.778
variance	0.011262(14)	0.011230(9)	0.011259(10)	0.011266(12)

Table 2.2: Candidates for the minima of the variance and their variances obtained by correlated verification run for H_2 with trial function Ψ_3 with variable geometry, $E_T = -1.175$.

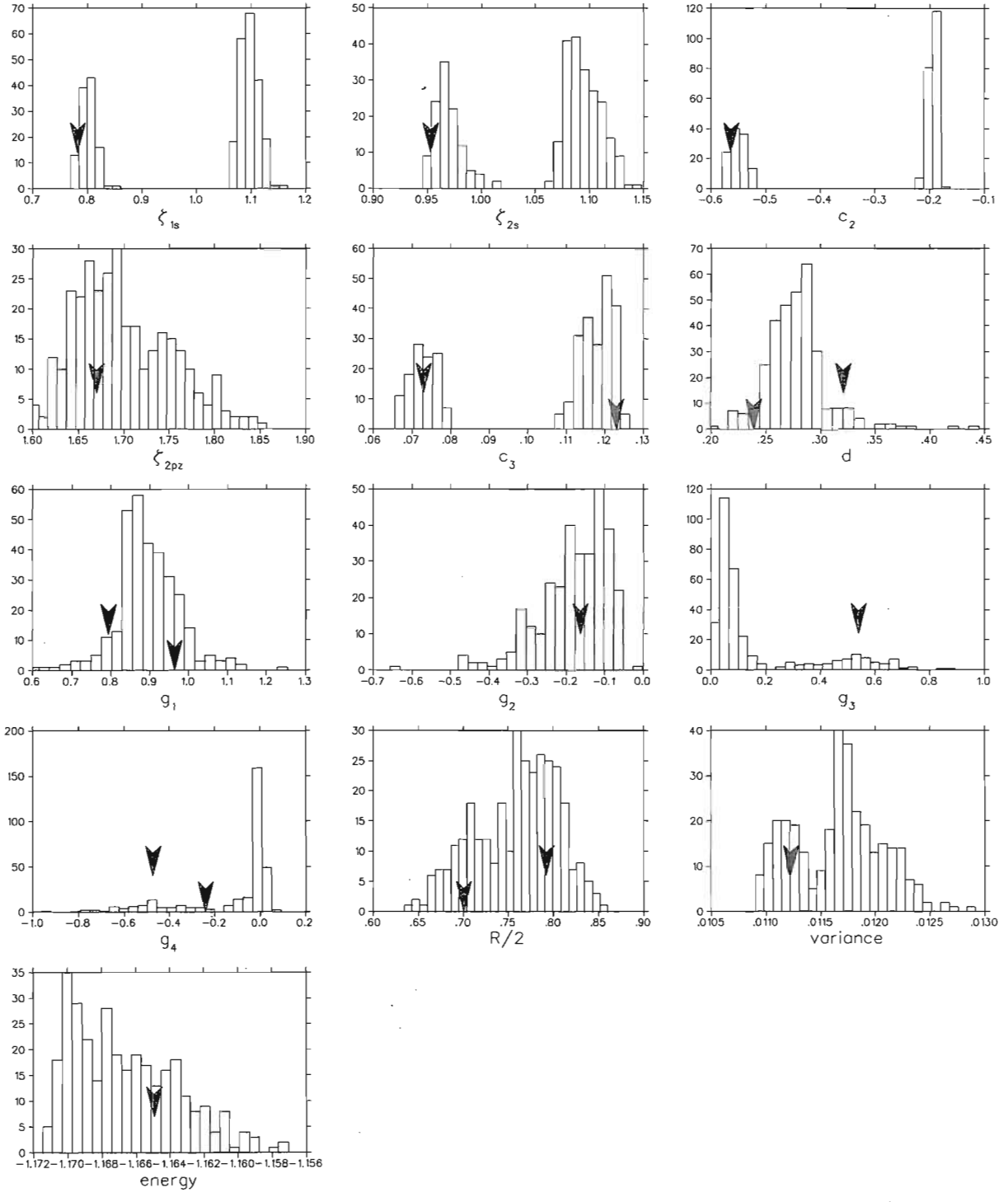

Figure 2.5: Unfiltered histograms for the variance optimization of Ψ_3 for H_2 with a variable geometry.

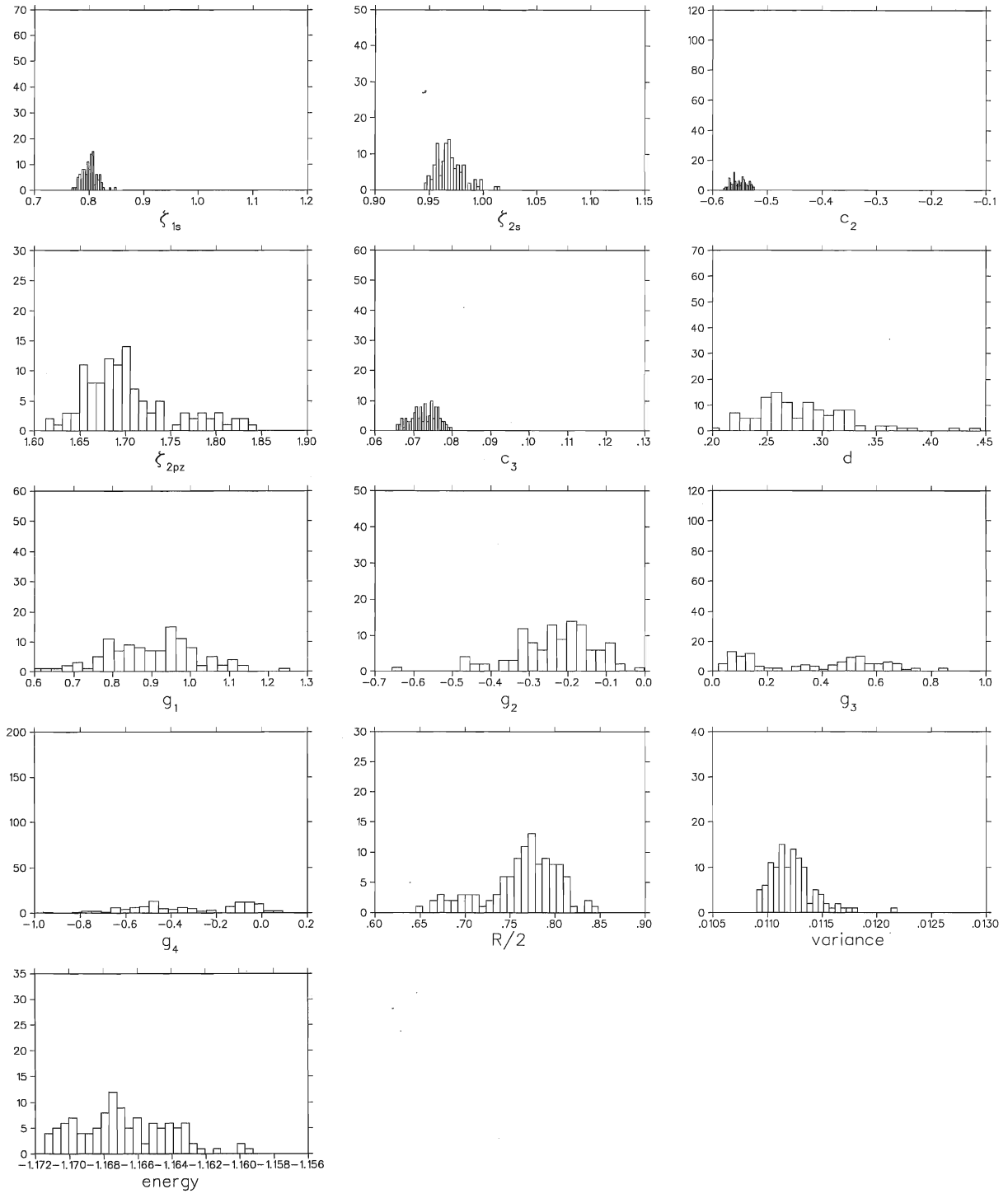

Figure 2.6: Results of filtering the left peak of c_3 in figure 2.5.

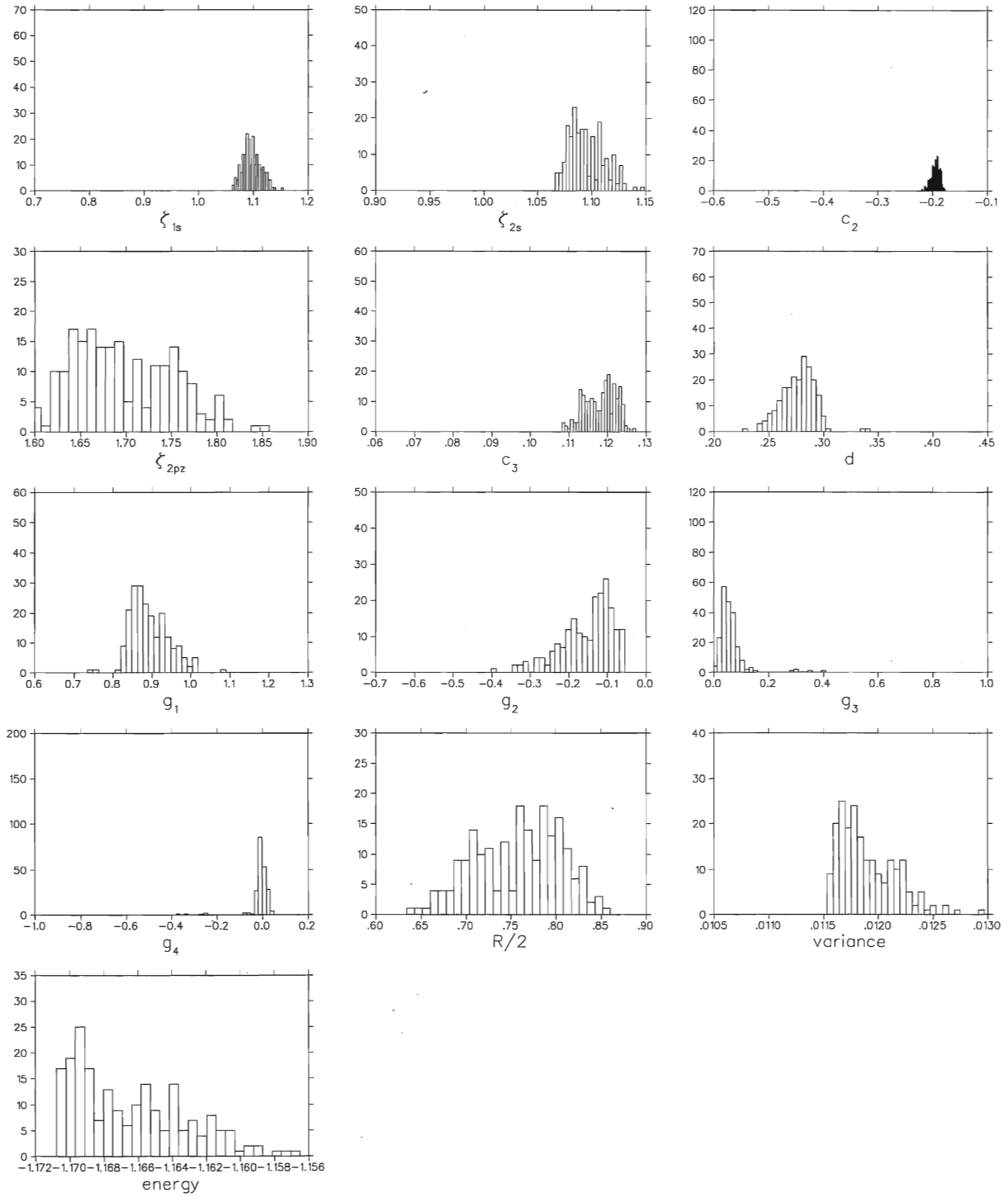

Figure 2.7: Results of filtering the right peak of c_3 in figure 2.5.

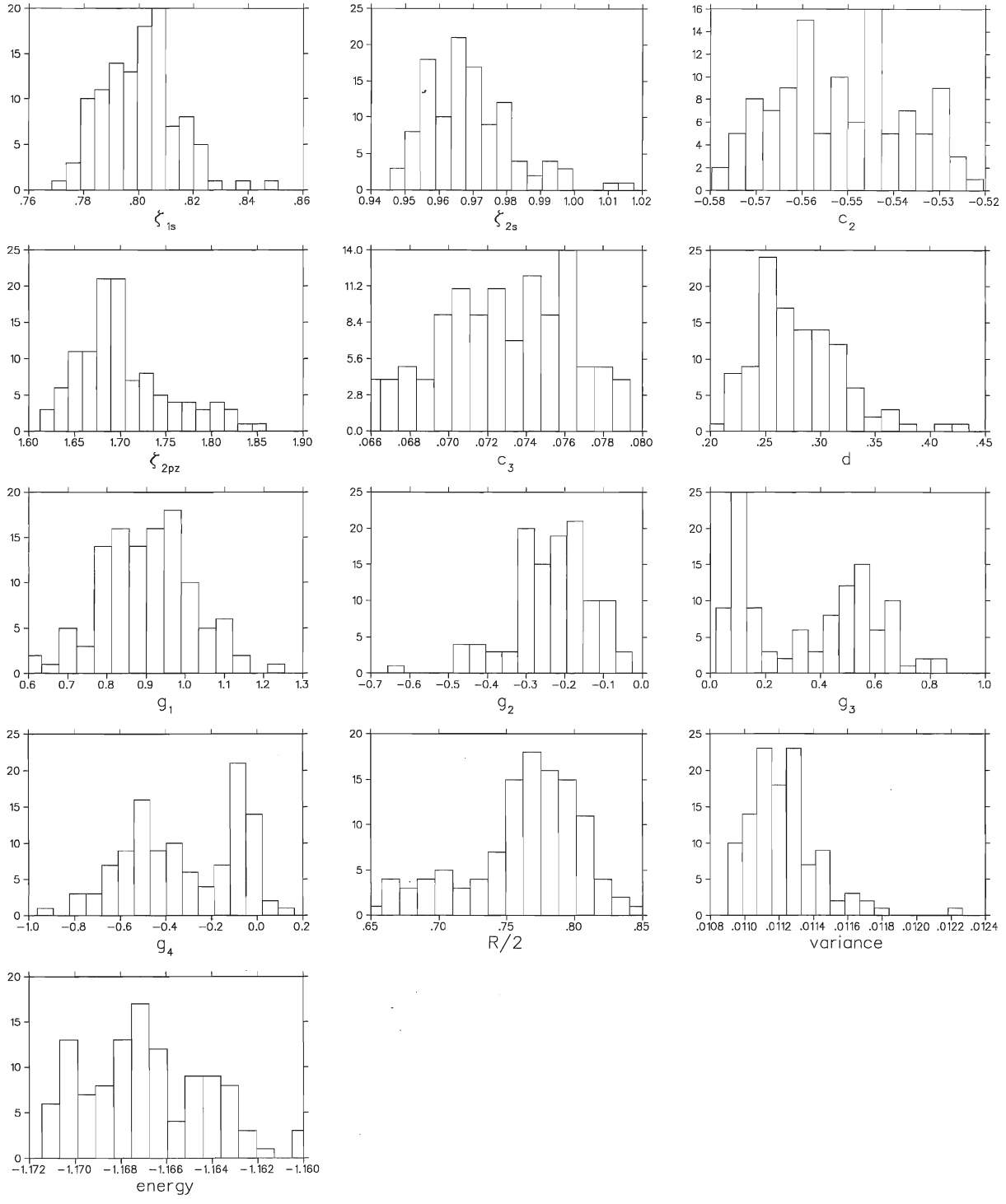

Figure 2.8: Same as 2.6, only with different scaling.

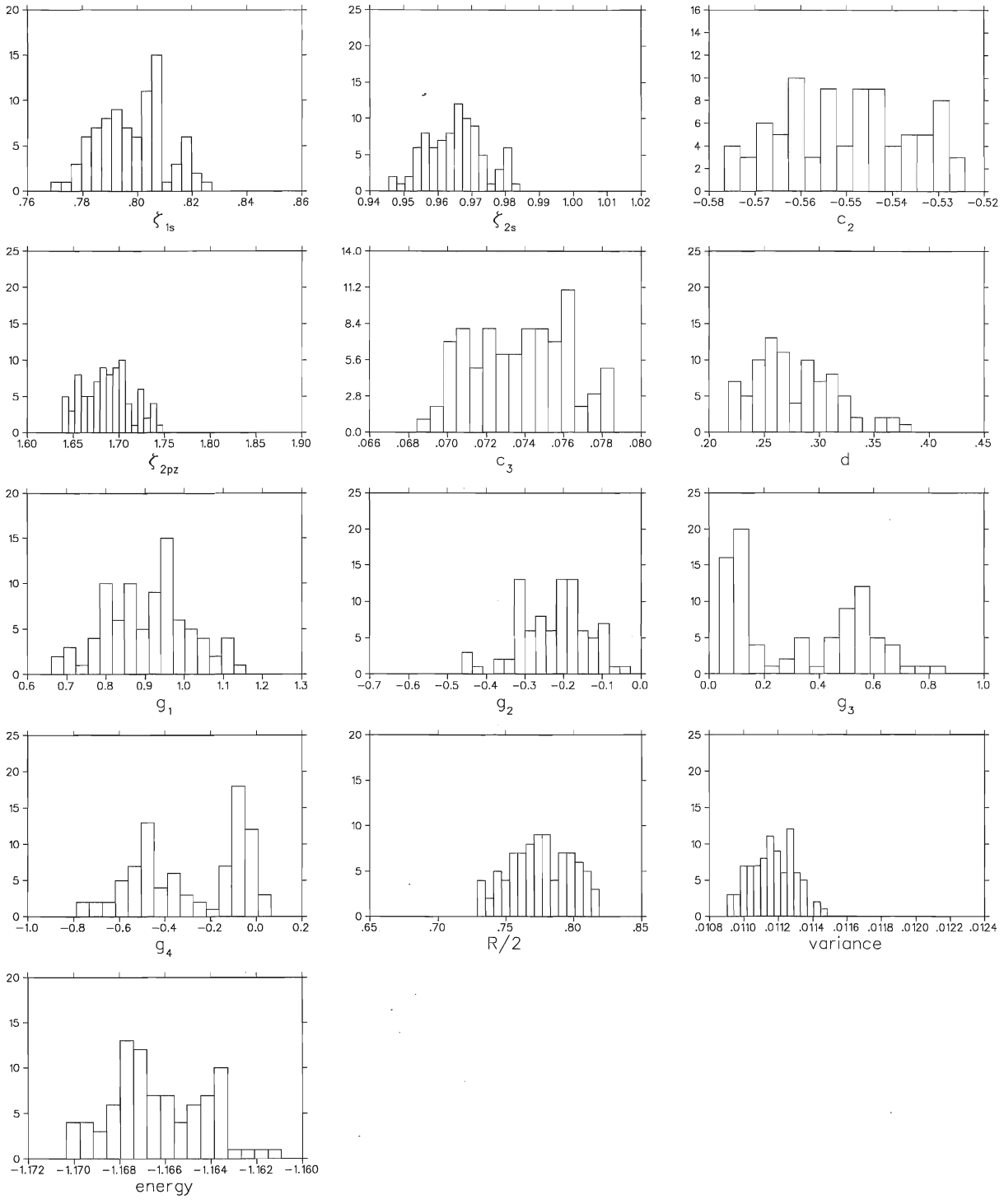

Figure 2.9: Same as 2.8, after filtering to remove outliers.

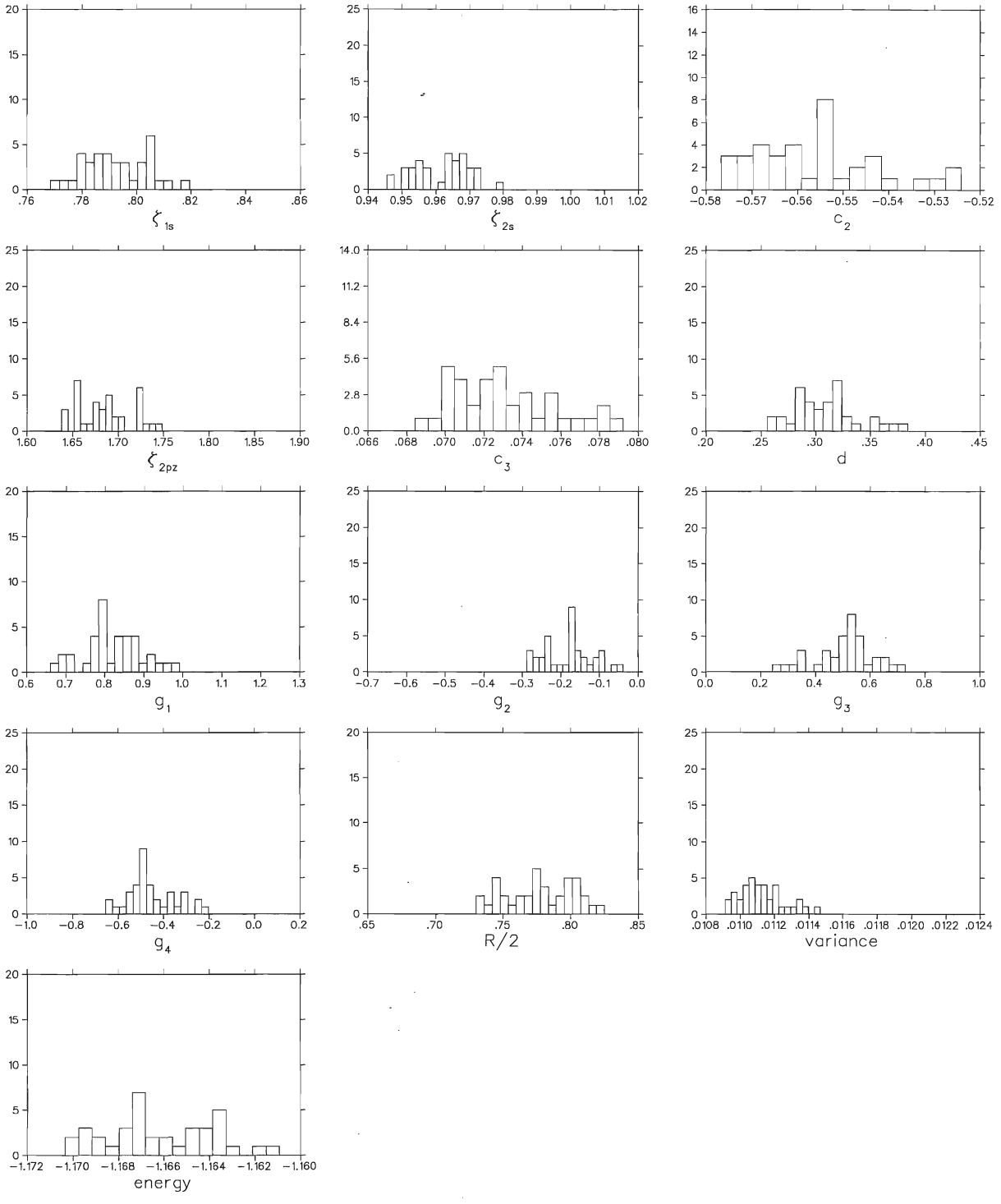

Figure 2.10: Result of filtering the left peak of g_4 in figure 2.9.

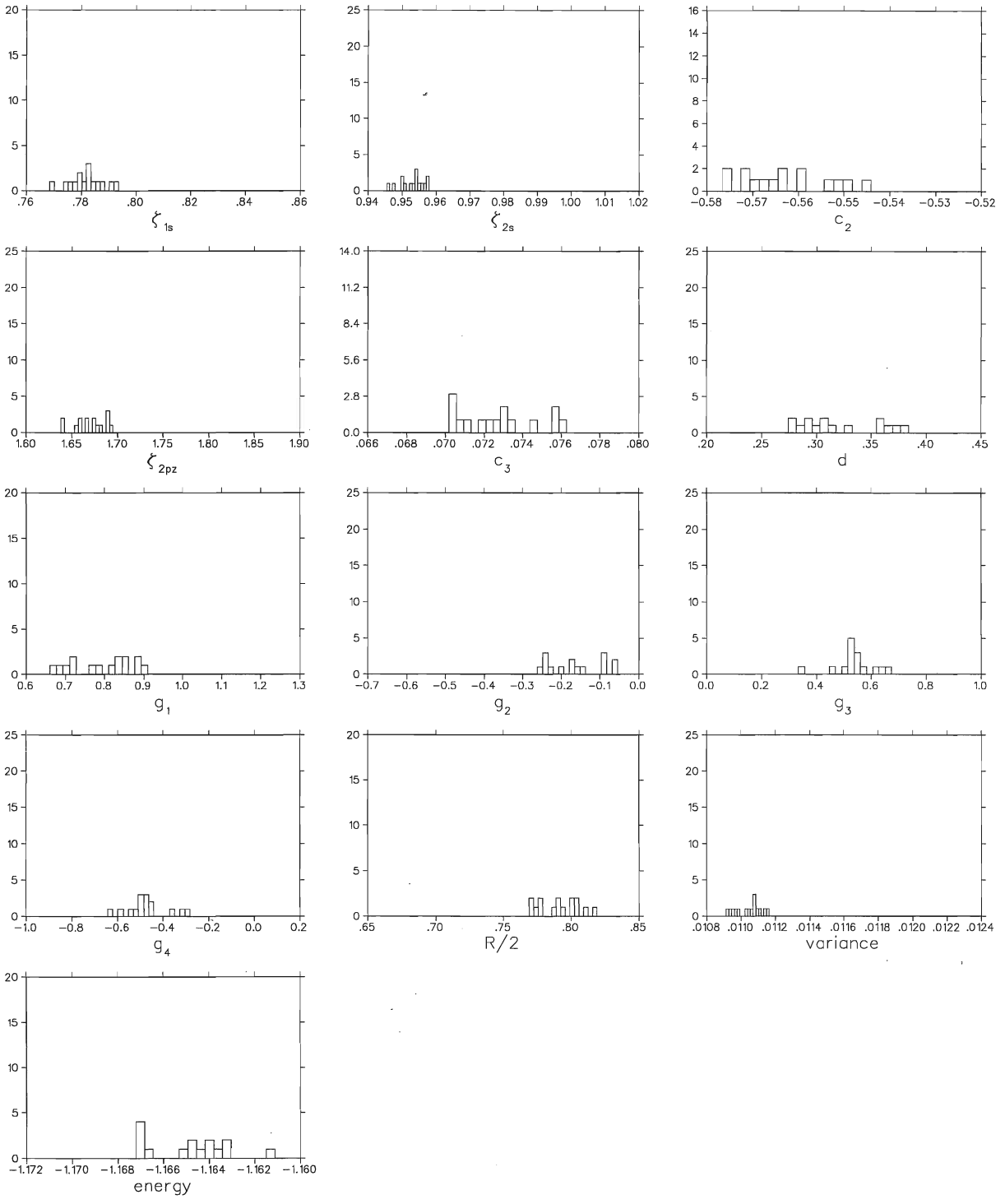

Figure 2.11: Result of filtering the left peak of ζ_{2s} in figure 2.10.

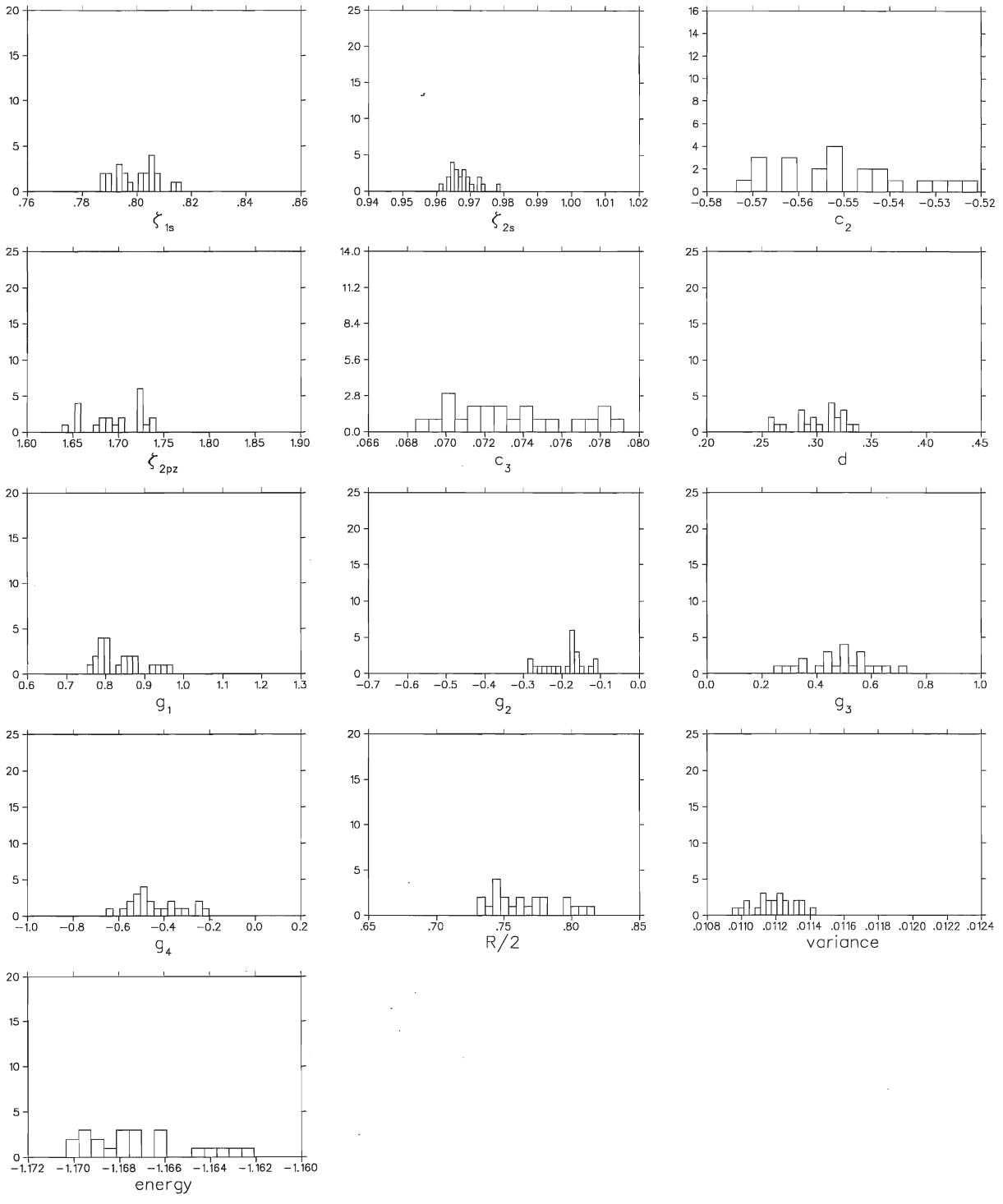

Figure 2.12: Result of filtering the right peak of ζ_{2s} in figure 2.10.

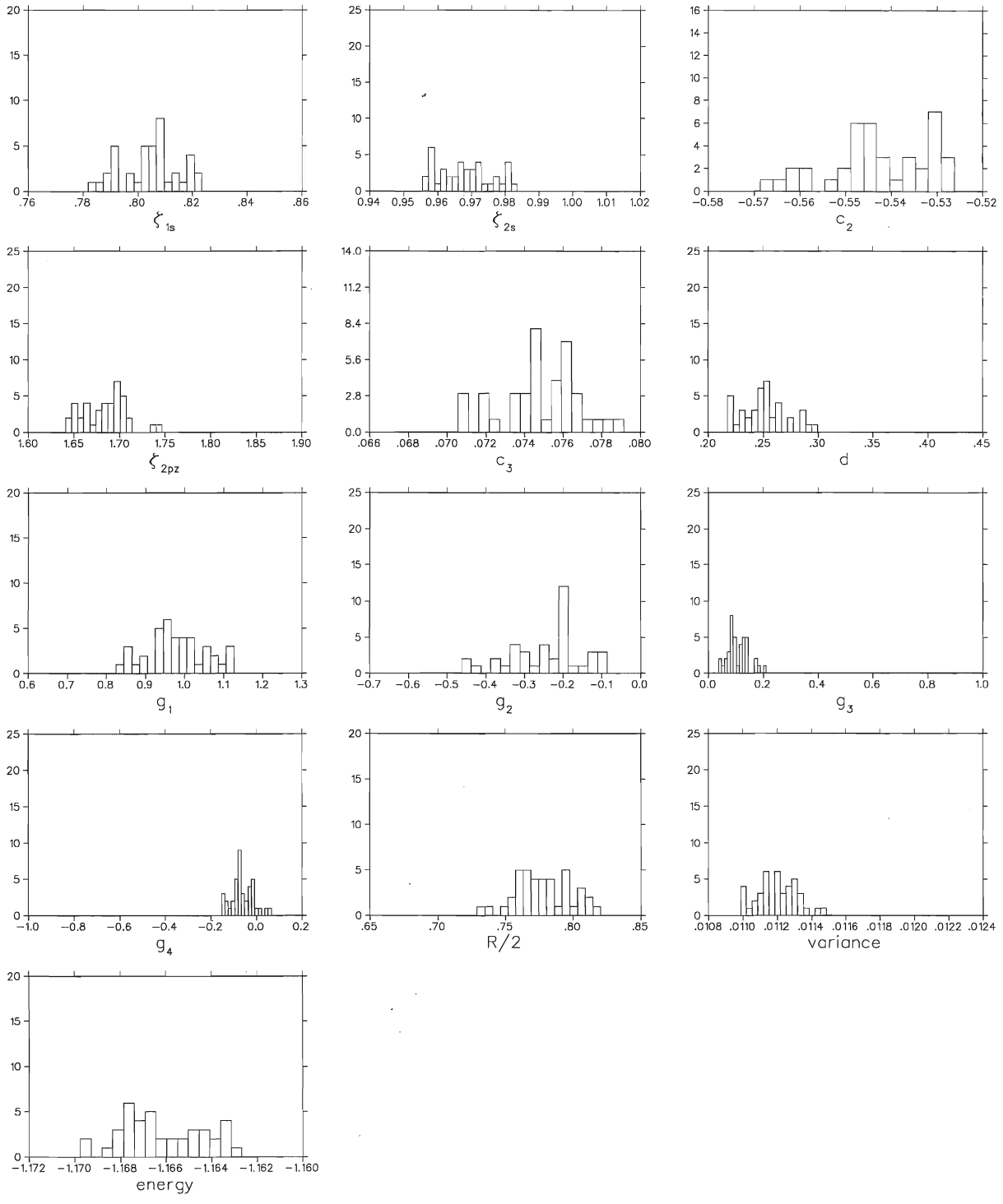

Figure 2.13: Result of filtering the right peak of g_4 in figure 2.9.

Chapter 3

Applications to Ground States of Small Systems

In this chapter we present the results of optimizations of various wave functions for three different systems—hydrogen molecule H_2, helium atom He and lithium hydride molecule LiH. For each wave function we report the initial and optimal parameters[1], variational energies and histograms ridden of outliers. The optimal parameter values were obtained in similar manner as we described in chapter 2.

For some systems the optimal values are almost the same as the initial ones. This due to the fact that the initial values are results of previous optimizations (see section 1.7).

We will use the following abbreviations to make the tables and text more readable:

EO - Energy Optimization

VO - Variance optimization

EOVG - Energy Optimization with Variable Geometry

EOFG - Energy Optimization with Fixed Geometry

VOVG - Variance Optimization with Variable Geometry

VOFG - Variance Optimization with Fixed Geometry

IP - Initial Parameters

OP - Optimal Parameters

CE - Correlation Energy

The (electronic) correlation energy is defined as the difference between the variational

[1]If there is a star by some parameter value it means that that parameter was not optimized.

energy and Hartree-Fock energy

$$\text{CE} = E_{var} - E_{HF} \ .$$

It is usually reported as the percentage of the exact CE:

$$\text{CE}[\%] = \frac{E_{var} - E_{HF}}{E_{exact} - E_{HF}} \times 100\% \ .$$

It indicates how good the variational wave function reflects the inter-electronic (Coulomb) interactions.

3.1 The hydrogen molecule

The forms of the trial functions considered here are the simplest LCAO Ψ_1 and two explicitly correlated ones, Ψ_2 and Ψ_3, given as follows:

$$\begin{aligned}
\Psi_1 &= \Phi_1(1)\Phi_1(2) \\
\Psi_2 &= \Phi_1(1)\Phi_1(2)J_1 \\
\Psi_3 &= \Phi_2(1)\Phi_2(2)J_2 \\
\\
\Phi_1 &= \phi_{1s_a}(\zeta_{1s}) + \phi_{1s_b}(\zeta_{1s}) \\
\Phi_2 &= c_1[\phi_{1s_a}(\zeta_{1s}) + \phi_{1s_b}(\zeta_{1s})] + c_2[\phi_{2s_a}(\zeta_{2s}) + \phi_{2s_b}(\zeta_{2s})] + c_3[\phi_{2pz_a}(\zeta_{2pz}) + \phi_{2pz_b}(\zeta_{2pz})] \\
J_1 &= \exp\left(\tfrac{ar_{ij}}{1+br_{ij}}\right) \\
J_2 &= \exp\left(\textstyle\sum_{k=1}^{4} g_k \bar{r}_{12}^k\right) \\
\bar{r}_{12} &= \tfrac{dr_{12}}{1+dr_{12}} \ ,
\end{aligned} \tag{3.1}$$

where ϕ's are Slater-type atomic orbitals centered on the hydrogen atom a or b.

Ψ_1 is a minimal basis set, uncorrelated wave function with a variable orbital exponent, first optimized by Wang [24] in 1928. Reynolds, Ceperley, Alder, and Lester, Jr. [20] optimized Ψ_2, which is Ψ_1 augmented by a simple electron correlation function. There

is one additional variational parameter: b in the Jastrow-Pade function, J_1. The a parameter in the Jastrow is fixed by the cusp condition, and its value is taken from the original paper. Therefore, including the equilibrium inter-nuclear distance R as an adjustable parameter, Ψ_1 and Ψ_2 have two and three variational parameters, respectively.

Ψ_3 is much more sophisticated, consisting of a double-zeta plus polarization basis set and a Schmidt and Moskowitz [21] electron-correlation factor. Including R as an adjustable parameter, and keeping the dominant MO coefficient fixed, altogether there are eleven parameters in this wave function. Basic characteristics of those wave functions are summarized in table 3.1. The initial and optimized parameters are in tables 3.2–3.5. The variational energies and CE's are reported in table 3.6. Figures 3.1–3.12 show the corresponding histograms. We show only histograms for parameters which were optimized, together with the energy and variance. In a few cases the variance histograms for energy optimization are missing because those runs were performed among the first, and the variance was not stored in the file that time.

For the energy-optimized Ψ_1 the variational energy and the optimal geometry are in excellent agreement with Coulson's [4] calculations in 1937. The variance-optimized wave function is somewhat inferior: the variational energy is 0.11 E_h higher and the optimal equilibrium bond distance is in excess of 1.24 a_0 longer.

The variational energy for energy-optimized Ψ_2 is 6 mE_h below the fixed-geometry optimizations of Reynolds[2], *et al.* [20], accounting for just over half of the electron CE. The equilibrium bond distance is within 0.023 a_0 of experiment. Again the variance-optimized wave function gives somewhat inferior results: we do not recover any of the CE, and the equilibrium geometry is 0.76 a_0 longer than experiment.

Ψ_3 is more typical of high-accuracy variational wave functions. The energy-optimized wave function recovers 93% of CE, and the optimal bond distance is within 0.007 a_0 of

[2]We used their optimal parameter values as our initial parameters for optimization.

the experimental value. The variance-optimized wave function is somewhat less accurate: 77% of CE with bond distance only within 0.18 a_0 of experiment.

	N_p	N_{Jp}	e-e correlation
Ψ_1	1+1	0	no
Ψ_2	2+1	1	yes
Ψ_3	10+1	5	yes

Table 3.1: Characteristics of the wave functions for H_2 molecule. Number of variational parameters N_p, number of variational parameters in the Jastrow factor N_{Jp}.

	Ψ_1[a]		Ψ_2			
	ζ_{1s}	$R/2$	ζ_{1s}	a	b	$R/2$
IP	1.194	0.692	1.285	0.28*	0.05	0.7005
OP	1.192	0.6933	1.296	0.28*	0.163	0.689
IP	1.19	0.7005*	1.29	0.28*	0.15	0.7005*
OP	1.1893	0.7005	1.2904	0.28*	0.1627	0.7005*

Table 3.2: Variational parameters for Ψ_1 and Ψ_2, energy optimization.

[a] Ref. [4]: ζ_{1s}=1.197, $R/2$=0.692 a_0

	Ψ_1		Ψ_2			
	ζ_{1s}	$R/2$	ζ_{1s}	a	b	$R/2$
IP	1.0	1.2	1.22	0.28*	0.162	0.7
OP	0.961	1.322	1.0955	0.28*	0.1500	1.0802
IP	1.14	0.7005*	1.29	0.28*	0.15	0.7005*
OP	1.1368	0.7005*	1.2198	0.28*	0.1679	0.7005*

Table 3.3: Variational parameters for Ψ_1 and Ψ_2, variance optimization.

	ζ_{1s}	c_1	ζ_{2s}	c_2	ζ_{2p_z}	c_3
IP	1.311	1.0*	0.587	0.008	1.937	0.127
OP	1.3108	1.0*	0.587	0.0065	1.9377	0.125
IP	1.312	1.0*	0.589	0.007	1.97	0.123
OP	1.313	1.0*	0.590	0.0066	1.968	0.123
	d	g_1	g_2	g_3	g_4	$R/2$
IP	0.241	0.959	0.117	-0.184	-0.24	0.7005*
OP	0.239	0.962	0.12	-0.183	-0.24	0.7005*
IP	0.239	0.964	0.122	-0.184	-0.242	0.7
OP	0.238	0.964	0.123	-0.185	-0.243	0.697

Table 3.4: Variational parameters for Ψ_3, energy optimization.

	ζ_{1s}	c_1	ζ_{2s}	c_2	ζ_{2pz}	c_3
IP	1.312	1.0*	0.589	0.007	1.97	0.123
OP	0.8092	1.0*	0.9844	-0.56	1.775	0.0688
IP	1.312	1.0*	0.589	0.007	1.97	0.123
OP	0.782	1.0*	0.953	-0.563	1.670	0.073
	d	g_1	g_2	g_3	g_4	$R/2$
IP	0.239	0.964	0.122	-0.184	-0.242	0.7005*
OP	0.312	0.812	-0.205	0.531	-0.51	0.7005*
IP	0.239	0.964	0.122	-0.184	-0.242	0.7
OP	0.321	0.794	-0.163	0.539	-0.472	0.792

Table 3.5: Variational parameters for Ψ_3, variance optimization.

	Ψ_1		Ψ_2		Ψ_3	
	energy	CE [%]	energy	CE [%]	energy	CE [%]
EOVG	-1.12824(5)	<0	-1.15657(3)	56	-1.17166(2)	93
EOFG	-1.12822(7)	<0	-1.15654(3)	56	-1.17163(2)	93
VOVG	-1.01814(5)	<0	-1.09681(3)	<0	-1.16492(2)	77
VOFG	-1.12546(6)	<0	-1.15184(3)	44	-1.17051(2)	90

Table 3.6: Variational energies for H_2 molecule.

E_T=-1.175 E_h, E_{exact}=-1.17447... E_h (Ref. [12]), E_{HF}=-1.1337 E_h (Ref.[11])

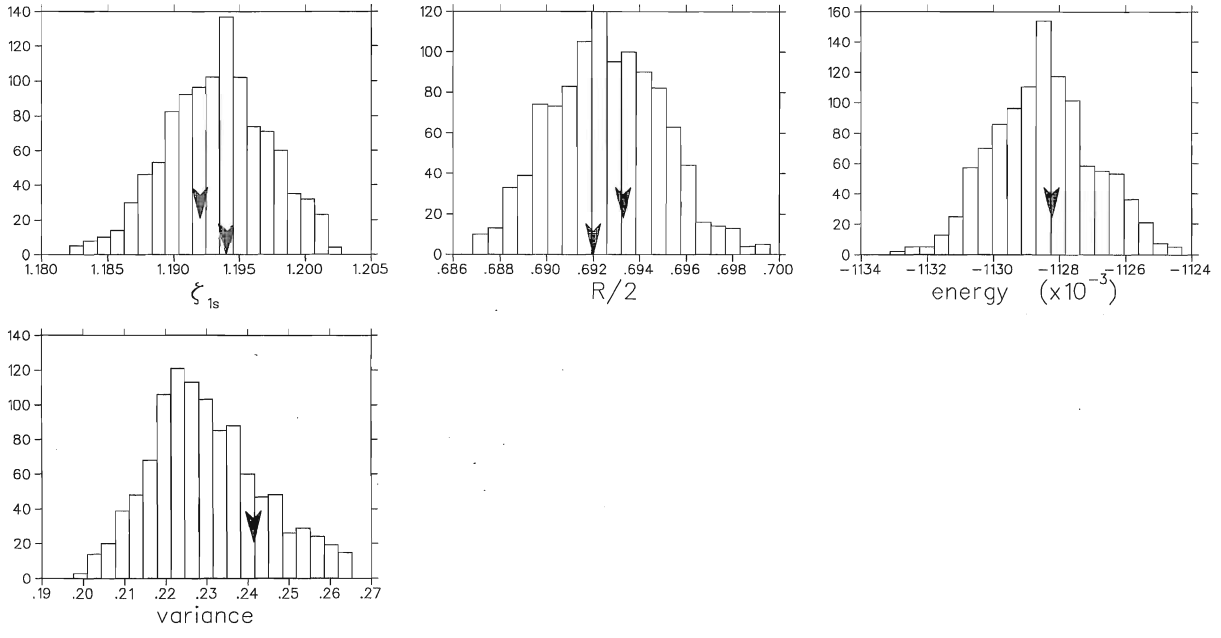

Figure 3.1: Histograms for H_2 molecule, Ψ_1, energy optimization with variable geometry.

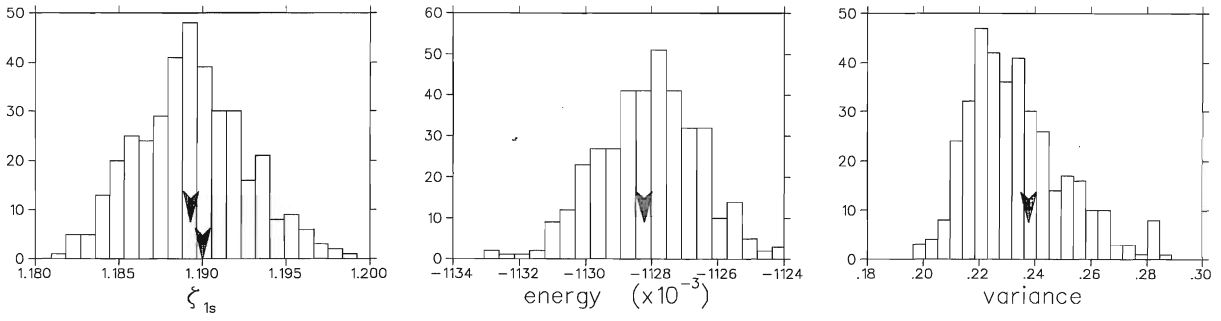

Figure 3.2: Histograms for H_2 molecule, Ψ_1, energy optimization with fixed geometry.

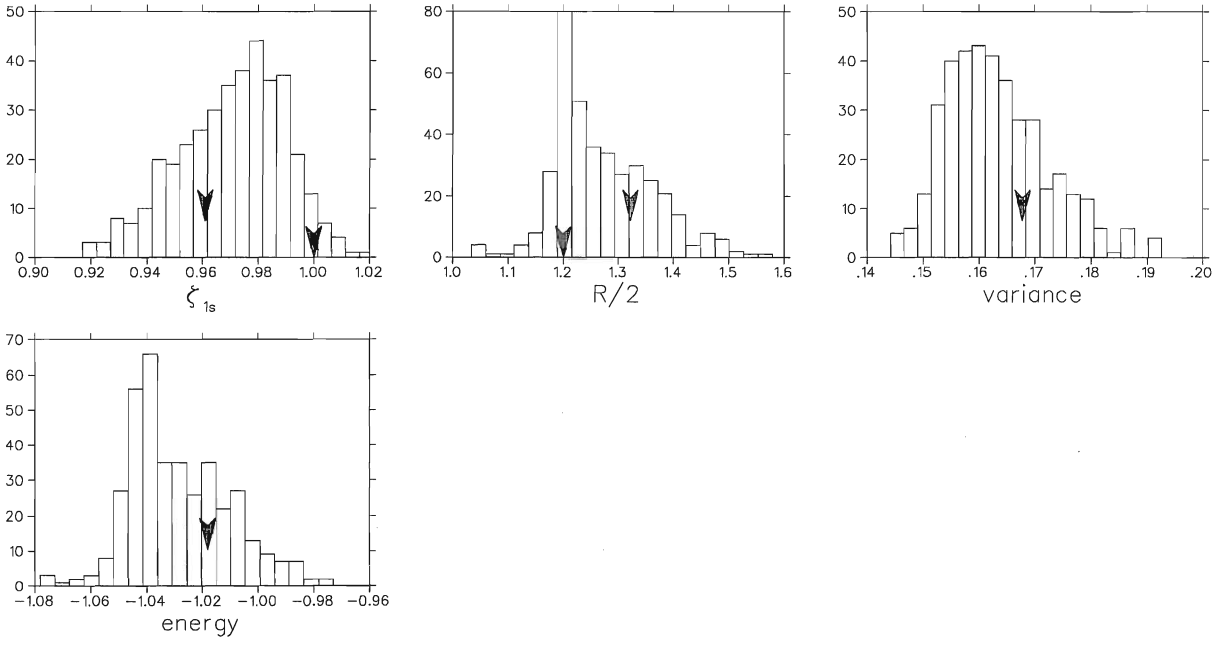

Figure 3.3: Histograms for H_2 molecule, Ψ_1, variance optimization with variable geometry.

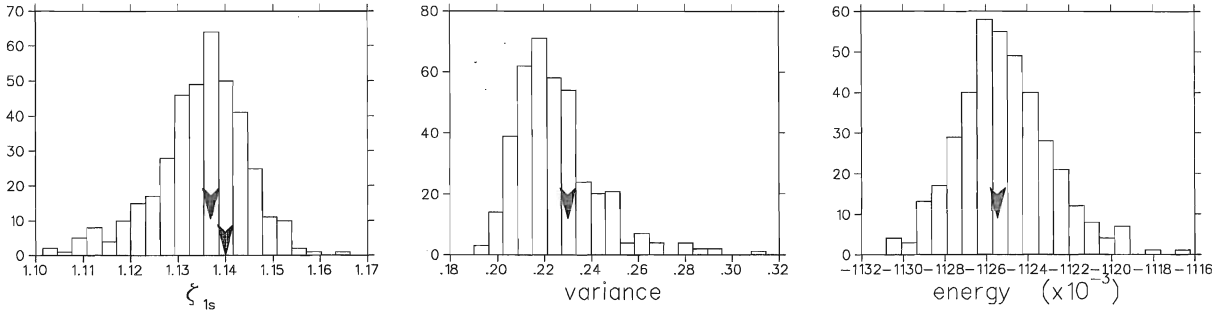

Figure 3.4: Histograms for H_2 molecule, Ψ_1, variance optimization with fixed geometry.

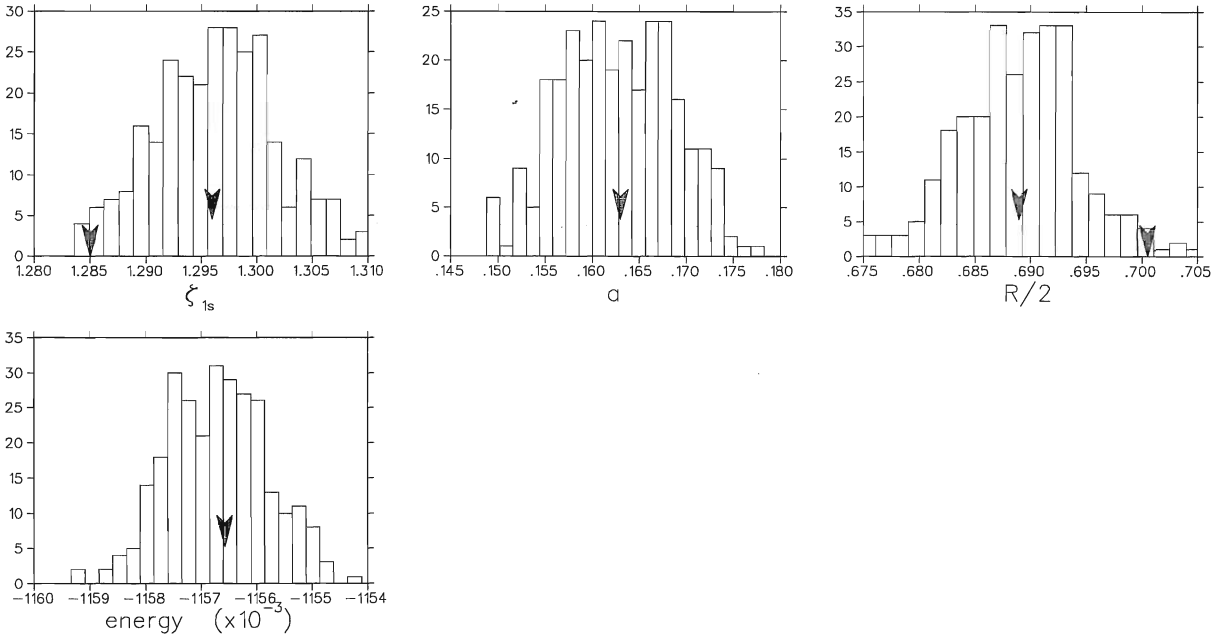

Figure 3.5: Histograms for H_2 molecule, Ψ_2, energy optimization with variable geometry.

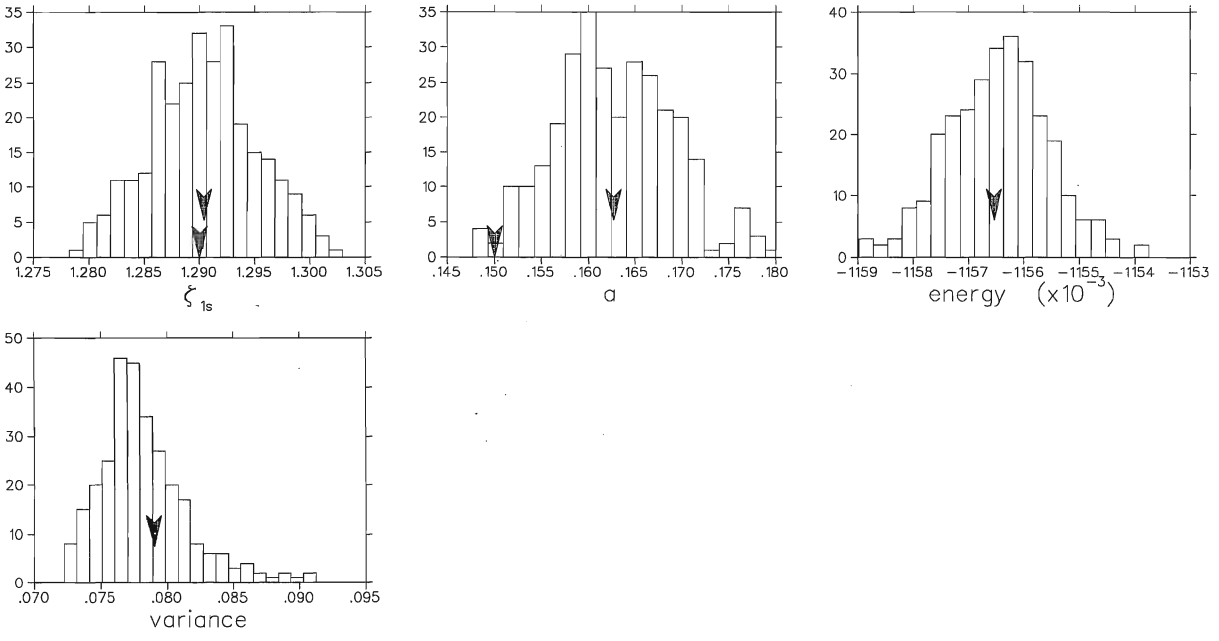

Figure 3.6: Histograms for H_2 molecule, Ψ_2, energy optimization with fixed geometry.

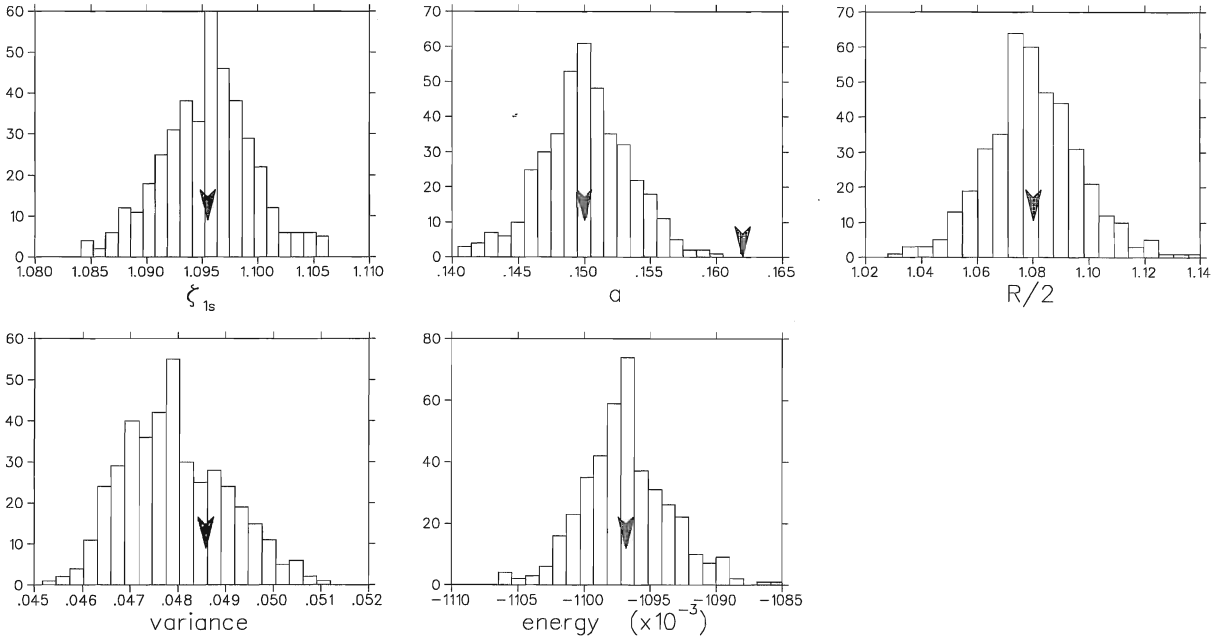

Figure 3.7: Histograms for H_2 molecule, Ψ_2, variance optimization with variable geometry.

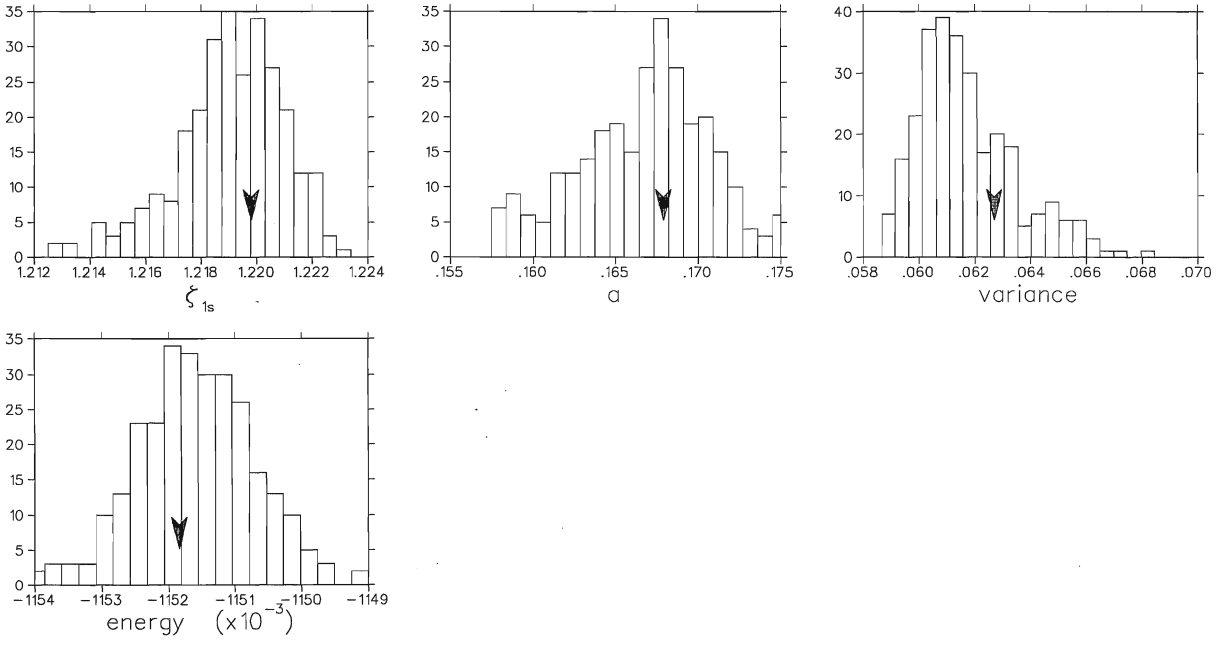

Figure 3.8: Histograms for H_2 molecule, Ψ_2, variance optimization with fixed geometry.

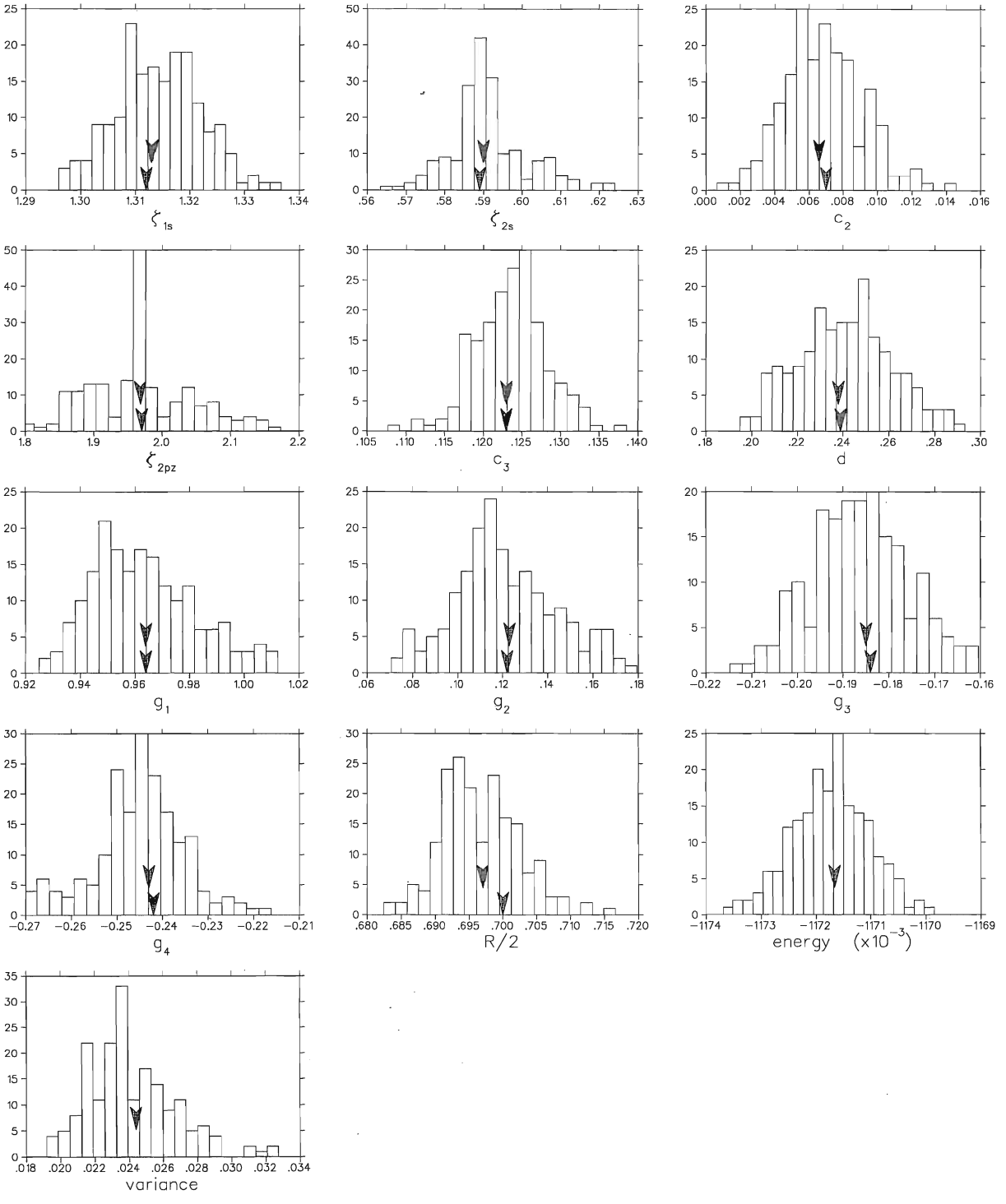

Figure 3.9: Histograms for H_2 molecule, Ψ_3, energy optimization with variable geometry.

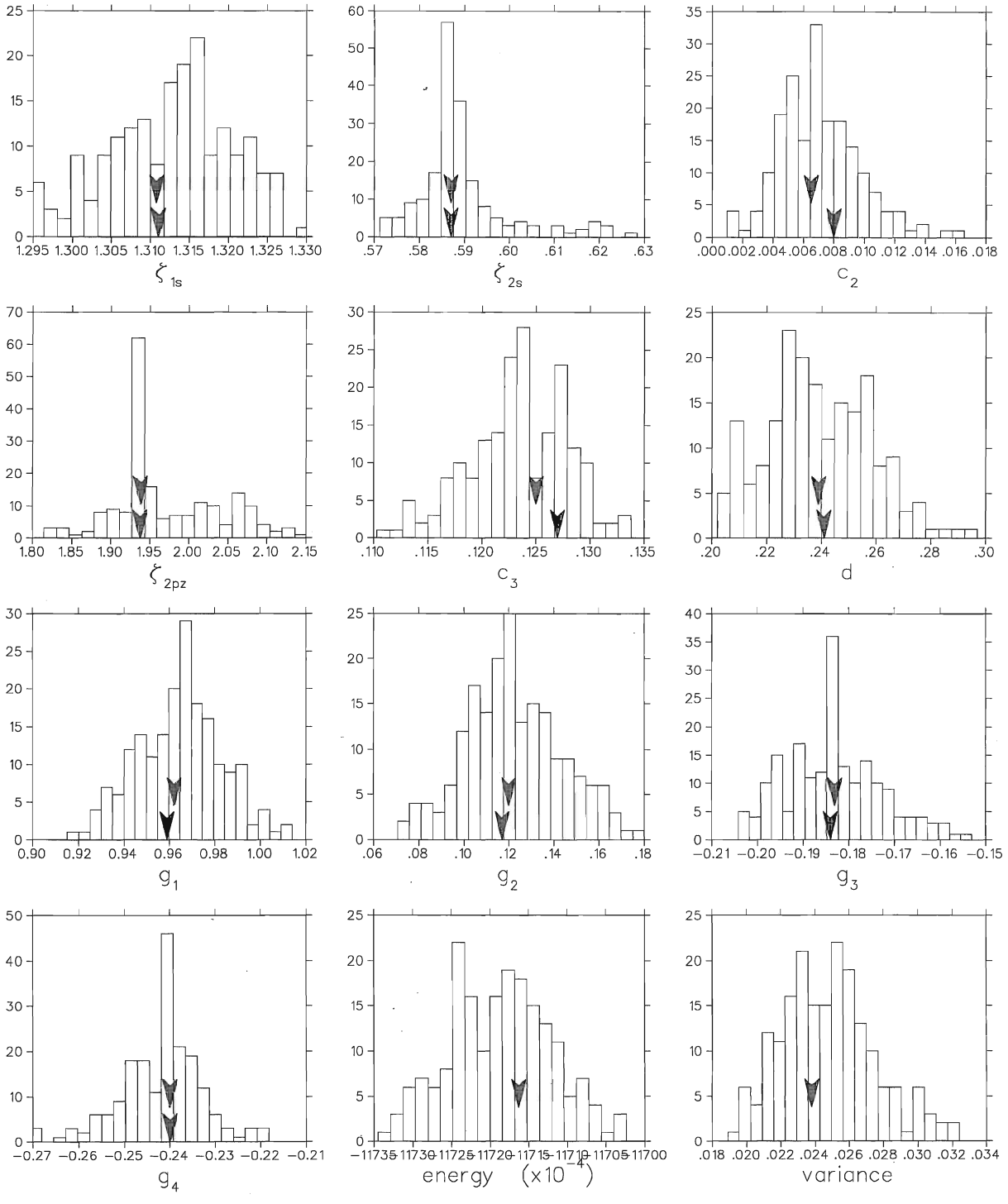

Figure 3.10: Histograms for H_2 molecule, Ψ_3, energy optimization with fixed geometry.

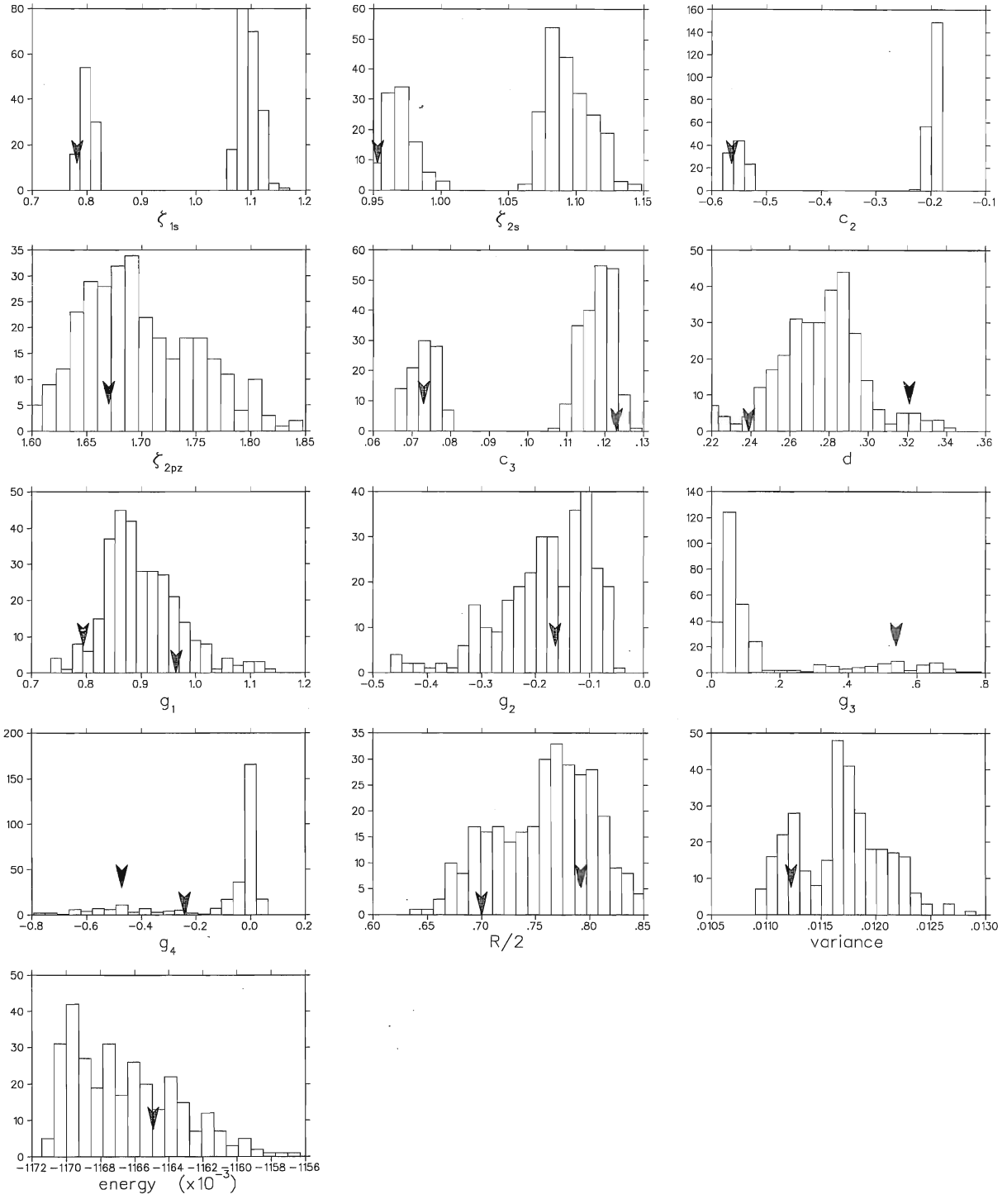

Figure 3.11: Histograms for H_2 molecule, Ψ_3, variance optimization with variable geometry.

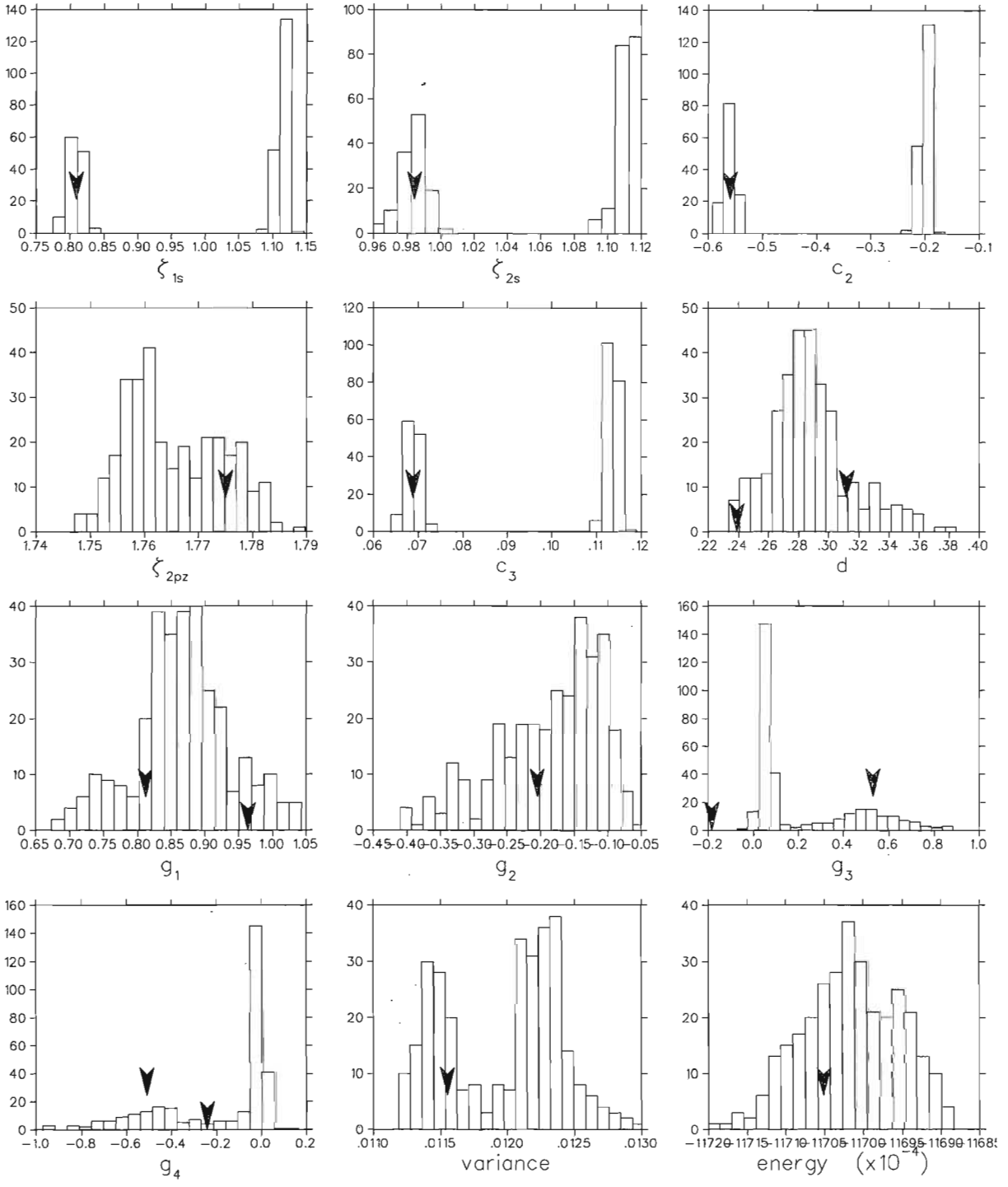

Figure 3.12: Histograms for H_2 molecule, Ψ_3, variance optimization with fixed geometry.

3.2 The helium atom

The ground state of helium atom often serves as a benchmark test for different types of calculations in quantum chemistry. In our work we have considered three variational wave functions with the following forms:

$$
\begin{aligned}
\Psi_1 &= e^{-a(r_1+r_2)}(1+br_{12}) \\
\Psi_2 &= \exp\left[-Zr_1\left(\tfrac{1+ar_1}{1+br_1}\right) - Zr_2\left(\tfrac{1+ar_2}{1+br_2}\right) + \tfrac{1}{2}\tfrac{r_{12}e^{-cr_{12}}}{1+dr_{12}}\right] \\
\Psi_3 &= \Phi(1)\Phi(2)J_{SM} \\
\Phi &= c_1\phi_{1s}(\zeta_{1s}) + c_2\phi_{2s}(\zeta_{2s_1}) + c_3\phi_{2s}(\zeta_{2s_2}) + c_4\phi_{3s}(\zeta_{3s}) \\
J_{SM} &= \exp\left[\textstyle\sum_{k=1}^{9} g_k(\bar{r}_1^{m_k}\bar{r}_2^{n_k} + \bar{r}_1^{n_k}\bar{r}_2^{m_k})\bar{r}_{12}^{o_k}\right] \\
\bar{r} &= \tfrac{br}{1+br} \\
\bar{r}_{12} &= \tfrac{dr_{12}}{1+dr_{12}} \ .
\end{aligned}
\tag{3.2}
$$

Ψ_1 is the classic two-parameter Hylleraas type wave function. Ψ_2 is a more recent four-parameter form of Kenny, Rajagopal, and Needs [10]. Ψ_3 is much more sophisticated, consisting of an extended single-particle basis set (seven variational parameters), and an eleven-parameter Schmidt-Moskowitz Jastrow electron correlation wave function. In the optimizations we set the value of g_1 to 0.5 and kept the values of b and d parameters fixed for better performance. The values of b and d were determined after several optimization runs in which they were not fixed. (Other workers usually set these parameters to unity and do not vary them at all.) The set of nine exponents m_k, n_k, o_k is given in the table 3.7. The first four terms, corresponding to $m_k = n_k = 0$, provide for electron-electron correlations, while the next three, those with $n_k = o_k = 0$, are nuclear-electronic correlation terms. The remaining two terms allow explicit electron-electron-nuclear correlation, a type of correlation shown to be important for highly-accurate wave functions [21]. Wave function characteristics are summarized in the table 3.8.

Tables 3.9–3.12 report the initial and optimal parameter values. The appropriate

histograms are shown in figures 3.13–3.18.

Parameters obtained from energy optimization of Ψ_1 (see Ref. [16]) and variance optimization of Ψ_2 (see Ref.[3] [10]) are in good agreement with the literature.

Parameters for Ψ_3 are quoted in table 3.11 (EO) and 3.12 (VO).

Table 3.13 reports the results of verification runs done using these wave functions. The Hylleraas energy is within statistical error of the accepted value. Our variance optimization of Kenny *et al.*'s wave function improves the variational energy by 0.2 mE_h, while energy-optimization lowers it by 0.7 mE_h. As was the case for the hydrogen examples, the variational energy of variance-optimized wave functions agree better with the energy-optimized ones as the quality of the wave function improves. For the most accurate wave function they agree to within statistical error, with the variance-optimized energy having a substantially smaller error-bar.

k	m_k	n_k	o_k
1	0	0	1
2	0	0	2
3	0	0	3
4	0	0	4
5	2	0	0
6	3	0	0
7	4	0	0
8	2	2	0
9	2	0	2

Table 3.7: Coefficients m_k, n_k, o_k.

[3]The columns of Table 1. in the reference are mislabeled.

	N_p	N_{Jp}	e-e correlation	n-e correlation	e-e-n correlation
Ψ_1	2	1	yes	no	no
Ψ_2	4	2	yes	yes	no
Ψ_3	15	8	yes	yes	yes

Table 3.8: Characteristics of the wave functions for He atom. Number of variational parameters N_p, number of variational parameters in the Jastrow factor N_{Jp}.

	Ψ_1[a]		Ψ_2			
	a	b	a	b	c	d
IP	1.8	0.35	0.024	0.0842	0.086	0.285
OP	1.8495	0.3658	0.02427	0.0842	0.08596	0.28549

Table 3.9: Variational parameters for Ψ_1 and Ψ_2, energy optimization.

[a]Ref. [16]: a=1.849, b=0.364

	Ψ_1		Ψ_2[a]			
	a	b	a	b	c	d
IP	1.9	0.4	0.0046	0.064	0.037	0.442
OP	1.9554	0.4162	0.00627	0.06581	0.0363	0.4431

Table 3.10: Variational parameters for Ψ_1 and Ψ_2, variance optimization.

[a]Ref. [10]: a=0.00383, b=0.0620, c=0.0316, d=0.455

	ζ_{1s}	c_1	ζ_{2s_1}	c_2	ζ_{2s_2}	c_3	ζ_{3s}	c_4	b	
IP	1.594	1.348*	1.866	-0.229	2.623	-0.0734	5.314	0.001	0.847*	
OP	1.526	1.348*	1.757	-0.366	2.88	-0.103	5.229	0.0226	0.847*	
	d	g_1	g_2	g_3	g_4	g_5	g_6	g_7	g_8	g_9
IP	0.478*	0.5*	-0.116	-0.265	-0.404	0.224	0.0375	-0.0848	-1.42	1.944
OP	0.478*	0.5*	-0.18	-0.138	-0.343	0.623	0.111	-0.155	-1.73	1.825

Table 3.11: Variational parameters for Ψ_3, energy optimization.

	ζ_{1s}	c_1	ζ_{2s_1}	c_2	ζ_{2s_2}	c_3	ζ_{3s}	c_4	b
IP	1.578	1.348*	1.852	-0.288	2.664	-0.0524	5.3	-0.0009	0.847*
OP	1.5686	1.348*	1.841	-0.2973	2.6671	-0.0524	5.298	-0.00094	0.847*

	d	g_1	g_2	g_3	g_4	g_5	g_6	g_7	g_8	g_9
IP	0.495*	0.5*	-0.267	0.102	-0.702	0.360	0.05	-0.103	-1.59	2.12
OP	0.495*	0.5*	-0.2855	0.1956	-0.80	0.353	0.0457	-0.095	-1.577	2.084

Table 3.12: Variational parameters for Ψ_3, variance optimization.

	Ψ_1[a]		Ψ_2[b]		Ψ_3	
	energy	CE [%]	energy	CE [%]	energy	CE [%]
EO	-2.89110(4)	70	-2.900033(9)	91	-2.903109(21)	99
VO	-2.88296(4)	51	-2.89957(1)	90	-2.903107(7)	99

Table 3.13: Variational energies and variances for He atom.

E_T=-2.905 E_h, E_{exact}=-2.90372... E_h (Ref. [17]), E_{HF}=-2.86179... E_h (Ref. [3])
[a]Ref. [16]: E_{var}=-2.8912 E_h
[b]Ref. [10]: E_{var}=-2.89933(1) E_h

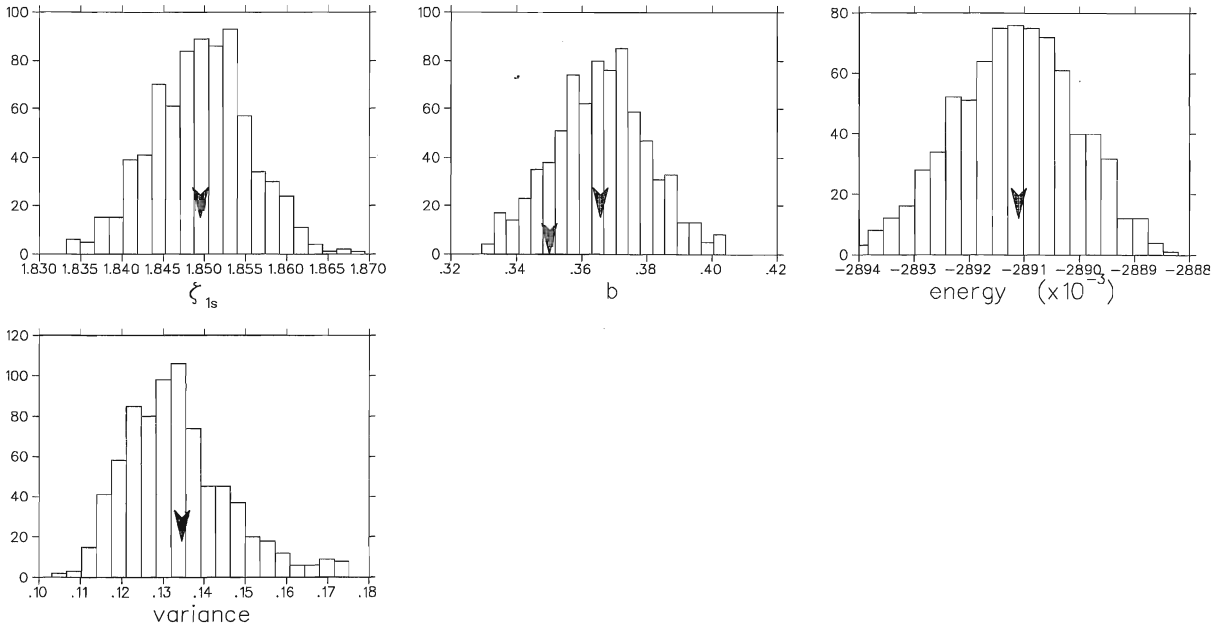

Figure 3.13: Histograms for He atom, Ψ_1, energy optimization.

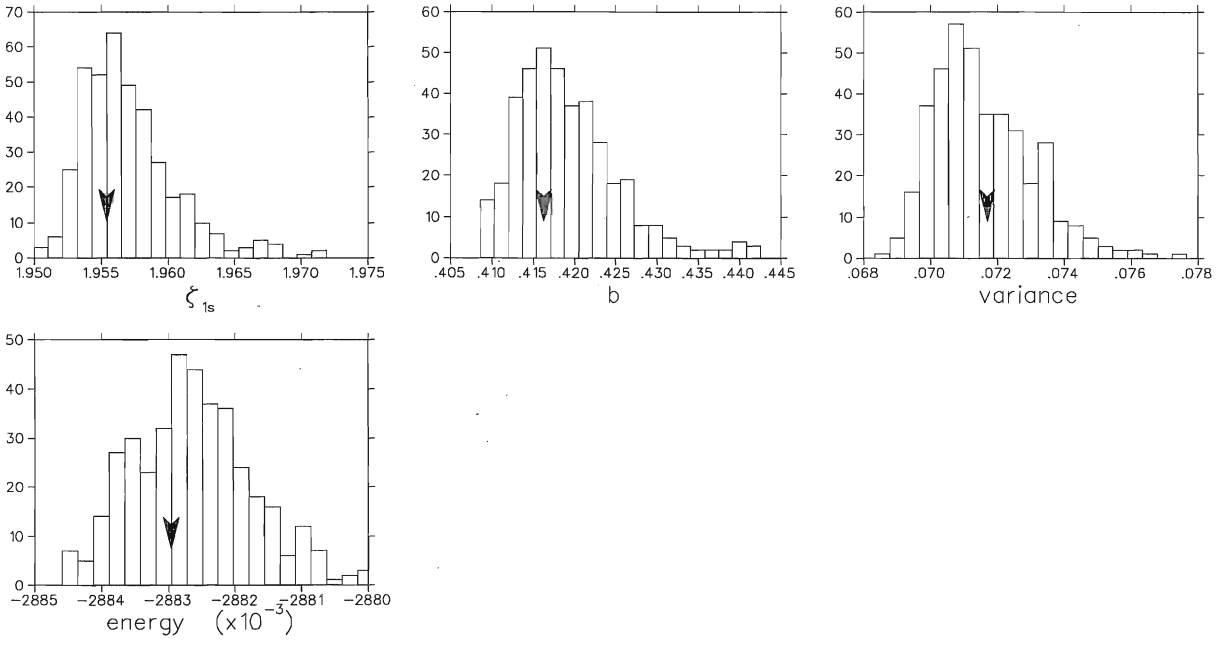

Figure 3.14: Histograms for He atom, Ψ_1, variance optimization.

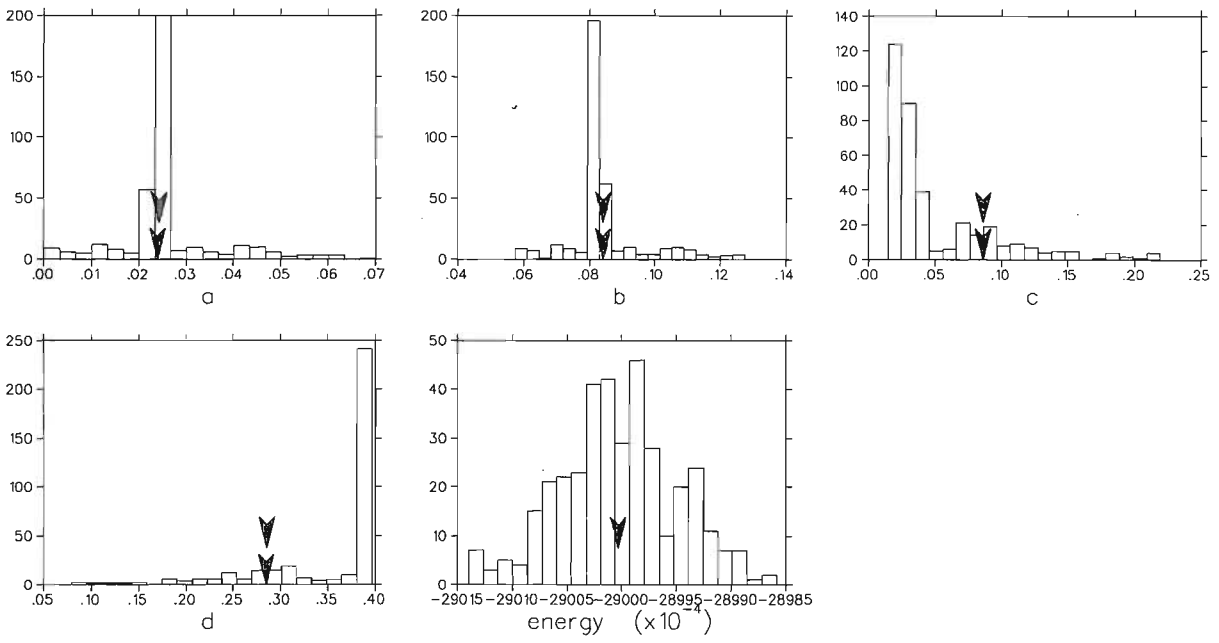

Figure 3.15: Histograms for He atom, Ψ_2, energy optimization.

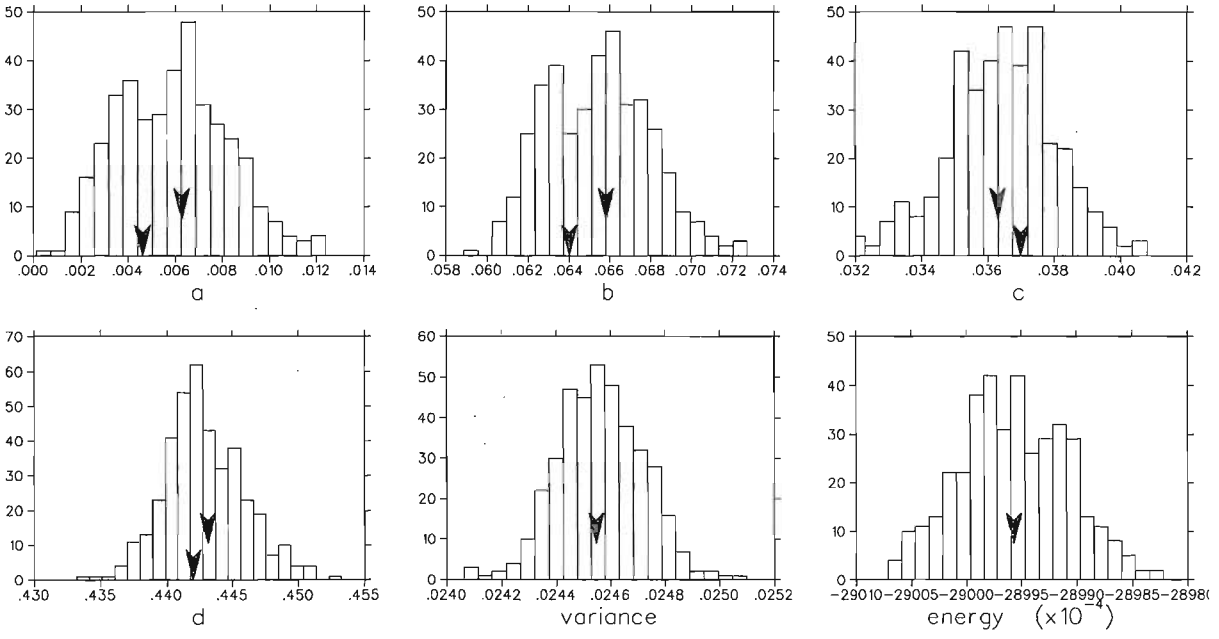

Figure 3.16: Histograms for He atom, Ψ_2, variance optimization.

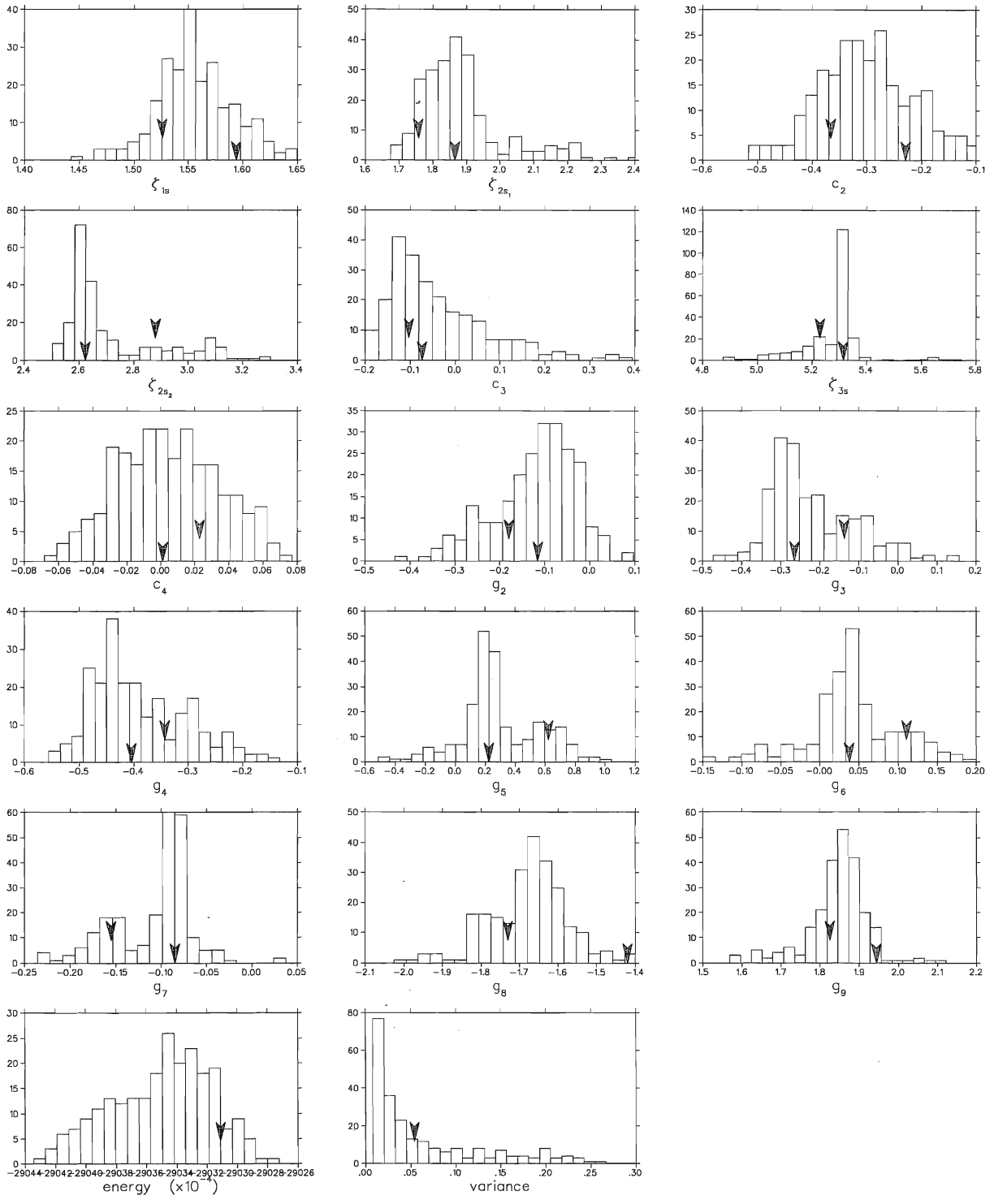

Figure 3.17: Histograms for He atom, Ψ_3, energy optimization.

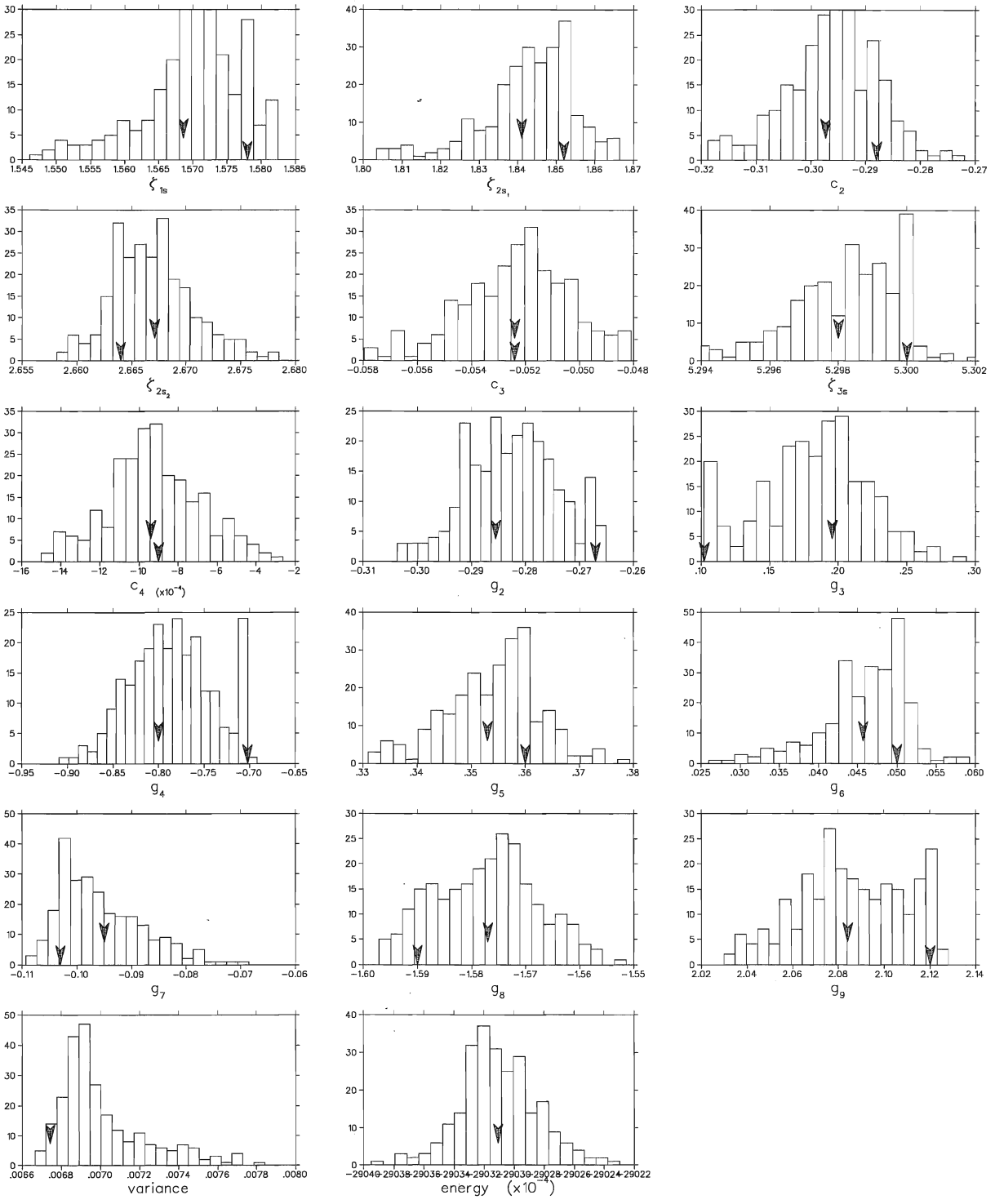

Figure 3.18: Histograms for He atom, Ψ_3, variance optimization.

3.3 The lithium hydride molecule

Lithium hydride is one of the smallest stable heteronuclear molecules in nature. We have chosen it as an example of more complicated system.

We have optimized two types of wave functions—Ψ_1 and Ψ_2. The form of these wave functions is as follows:

$$\begin{aligned}
\Psi_1 &= \det|\Phi_1^1(1)\Phi_2^1(2)|\det|\Phi_1^1(3)\Phi_2^1(4)|J_1 \\
\Psi_2 &= \det|\Phi_1^2(1)\Phi_2^2(2)|\det|\Phi_1^2(3)\Phi_2^2(4)|J_2 \\
J_1 &= \exp\left(\frac{ar_{ij}}{1+br_{ij}}\right) \\
J_2 &= \exp\left[\sum_{\alpha=1}^{2}\sum_{i=1}^{4}\sum_{j=i+1}^{4} U_{\alpha ij}\right] \\
U_{\alpha ij} &= \sum_{k=1}^{9} \Delta_k g_k(\bar{r}_{\alpha i}^{m_k}\bar{r}_{\alpha j}^{n_k} + \bar{r}_{\alpha i}^{n_k}\bar{r}_{\alpha j}^{m_k})\bar{r}_{ij}^{o_k} \\
\Delta_k &= \begin{cases} 1/2 & \text{for } k \le 4 \\ 1 & \text{otherwise} \end{cases} \\
\bar{r}_{\alpha x} &= \frac{br_{\alpha x}}{1+br_{\alpha x}} \qquad\qquad x = i, j \\
\bar{r}_{ij} &= \frac{dr_{ij}}{1+dr_{ij}} \, .
\end{aligned}$$

The Ψ_1 is relatively simple trial wave function with modest basis set and one parameter Jastrow factor optimized by Reynolds *et al.* [20]. The molecular orbitals Φ_1^1 and Φ_2^1 consist of linear combinations of 3 STO-type atomic orbitals, 2 centered on lithium and 1 on hydrogen atom. There are 4 free adjustable variational parameters in the Slater determinant (5 is fixed) so together with the bond distance the total number of variational parameters is 6. The initial and optimal parameters are reported in tables 3.14-3.17 and the corresponding histograms are shown in figures 3.19-3.22. The subscript a is used for orbitals centered on the lithium atom, and the subscript b is used for orbitals centered

on the hydrogen atom.

The Ψ_2 has larger basis set and it incorporates the Schmidt-Moskowitz Jastrow electron correlation wave function. The molecular orbitals Φ_1^2 and Φ_2^2 consist of linear combinations of 8 STO-type atomic orbitals, 5 centered on lithium and 3 on hydrogen atom. Ψ_2 has 24 parameters in the Slater determinant, 11 in the Jastrow factor (table 3.7), and 1 parameter is the bond distance. The set of nine exponents m_k, n_k, o_k is given in the table 3.7. One linear coefficient in each orbital can be fixed, and we set the linear coefficient by the $1s_b$ orbital exponent in Φ_1^1 to zero. Thus the total number of adjustable parameters is 33.

The energy optimization of Ψ_2 wave function with fixed geometry was done by stages (see also section 1.7). We started with the determinant parameters taken from [20] and set all the Jastrow parameters to zero. After performing one optimization run (to get some reasonable Jastrow parameters) we have fixed the Jastrow and optimized the determinant. We have again fixed the determinant and optimized the Jastrow and then reversed the process. 50 000 configurations were used for these optimizations. In the last stage we have performed optimization of all the 32 parameters with 200 000 configurations. The ensemble size had to be increased to improve the average number of successful runs (76%).

The energy optimization with variable geometry is harder to perform. The problem is the low number of successful runs. When all the parameters were optimized using 100 000 configurations, we had only 13% successful runs (we fixed b and d in that run so we had 31 free parameters). With 200 000 configurations it was 29%, and with 400 000 configurations 60% (in both cases b and d were optimized as well so we had 33 free parameters). The starting point for these optimizations was the same as for the final all-parameters optimization for fixed geometry, only the bond distance was set to 3.0 a_0 instead of the exact value 3.015 a_0. The results from these optimizations were used to

choose the determinant parameters for the Jastrow-only optimizations. It is interesting to note that the number of successful runs for the Jastrow-only optimization did not increase. However, the speed of the optimization was higher because there were fewer parameters to optimize (although we had to re-calculate the determinant as well because of the variable geometry).

The variance optimization of Ψ_2 was done by stages as well. Unlike the energy optimization the problem of optimizing all parameters is not the small number of successful runs but the slow convergence. To obtain starting parameters we performed two optimizations with 200 000 configurations starting from the energy-optimized parameter values. After that, we have determined the best parameter set by a correlated sampling run (we considered the average parameter values of those two runs as well). In the first stage we optimized only the Jastrow factor, in the second stage the Slater determinant and in the last third stage again only the Jastrow. The final Jastrow optimization was relatively fast and the parameters and variance did not change significantly. We have used 50 000 configurations during the stage optimizations.

We decided not to perform the VOVG. The results of the previous optimizations show that it is not competitive and the time spent would be not worth doing it. However, we did couple of optimizations to see the approximate value for the bond distance. That value lies somewhere around 3.14 a_0 (vs. 3.011 a_0 (EOVG) vs. 3.015 a_0 (experimental)).

The initial and optimal parameters for Ψ_2 are given in tables 3.18 - 3.21, and the corresponding histograms are shown in figures 3.23-3.26.

The variational energies and variances are reported in the table 3.22.

	(a,b,R/2)	STO	ζ	Φ_1^1	Φ_2^1
IP	(0.5*,0.9,1.5075*)	$1s_a$	2.8	1.0*	0.0*
		$2pz_a$	1.33	0.0*	0.29
		$1s_b$	0.88	0.0*	1.0*
OP	(0.5*,0.883,1.5075*)	$1s_a$	2.804	1.0*	0.0*
		$2pz_a$	1.315	0.0*	0.283
		$1s_b$	0.867	0.0*	1.0*

Table 3.14: Variational parameters for Ψ_1, energy optimization with fixed geometry.

	(a,b,R/2)	STO	ζ	Φ_1^1	Φ_2^1
IP	(0.5*,0.9,1.5)	$1s_a$	2.8	1.0*	0.0*
		$2pz_a$	1.33	0.0*	0.29
		$1s_b$	0.88	0.0*	1.0*
OP	(0.5*,897,1.47)	$1s_a$	2.796	1.0*	0.0*
		$2pz_a$	1.337	0.0*	0.276
		$1s_b$	0.868	0.0*	1.0*

Table 3.15: Variational parameters for Ψ_1, energy optimization with variable geometry.

	(a,b,R/2)	STO	ζ	Φ_1^1	Φ_2^1
IP	(0.5*,0.697,1.5075*)	$1s_a$	2.935	1.0*	0.0*
		$2pz_a$	1.09	0.0*	0.34
		$1s_b$	0.95	0.0*	1.0*
OP	(0.5,0.699,1.5075*)	$1s_a$	2.943	1.0*	0.0*
		$2pz_a$	1.029	0.0*	0.55
		$1s_b$	0.999	0.0*	1.0*

Table 3.16: Variational parameters for Ψ_1, variance optimization with fixed geometry.

	(a,b,R/2)	STO	ζ	Φ_1^1	Φ_2^1
IP	(0.5*,0.697,1.6)	$1s_a$	2.935	1.0*	0.0*
		$2pz_a$	1.09	0.0*	0.34
		$1s_b$	0.95	0.0*	1.0*
OP	(0.5*,0.690,1.625)	$1s_a$	2.938	1.0*	0.0*
		$2pz_a$	0.973	0.0*	0.511
		$1s_b$	0.985	0.0*	1.0*

Table 3.17: Variational parameters for Ψ_1, variance optimization with variable geometry.

	STO	ζ	Φ_1^1	Φ_2^1
IP	$1s_a$	2.414	1.0*	-0.13
	$1s_a$	4.273	0.159	0.0354
	$2s_a$	0.988	0.0153	0.35*
	$2pz_a$	1.219	0.029	0.44
	$2pz_a$	2.301	-0.039	0.018
	$1s_b$	0.751	0.0127	1.362
	$1s_b$	1.355	0.0*	-0.004
	$2pz_b$	0.838	0.002	-0.184
OP	$1s_a$	2.419	1.0*	-0.1287
	$1s_a$	4.268	0.144	0.028
	$2s_a$	0.997	0.0097	0.35*
	$2pz_a$	1.218	0.0169	0.4418
	$2pz_a$	2.304	-0.0273	0.0236
	$1s_b$	0.75	0.019	1.358
	$1s_b$	1.3554	0.0*	-0.016
	$2pz_b$	0.831	-0.0016	-0.188

Table 3.18: Variational parameters in Slater determinant for Ψ_2, energy optimization with fixed geometry.

	STO	ζ	Φ_1^1	Φ_2^1
IP	$1s_a$	2.453	1.0*	-0.129
&	$1s_a$	4.259	0.14	0.0163
OP	$2s_a$	1.036	0.0156	0.35*
	$2pz_a$	1.234	0.0272	0.442
	$2pz_a$	2.367	-0.0334	0.0305
	$1s_b$	0.768	0.0162	1.338
	$1s_b$	1.355	0.0*	-0.028
	$2pz_b$	0.835	-0.0009	-0.197

Table 3.19: Variational parameters in Slater determinant for Ψ_2, energy optimization with variable geometry. These parameters were kept fixed during the final stage of the optimization.

	STO	ζ	Φ_1^1	Φ_2^1
IP	$1s_a$	2.400	1.0*	-0.136
&	$1s_a$	3.542	0.612	0.122
OP	$2s_a$	1.290	-0.0137	0.35*
	$2pz_a$	1.180	-0.0515	0.281
	$2pz_a$	1.598	0.073	0.304
	$1s_b$	0.744	0.0214	1.163
	$1s_b$	1.158	0.0*	0.74
	$2pz_b$	0.935	-0.0081	-0.332

Table 3.20: Variational parameters in Slater determinant for Ψ_2, variance optimization with fixed geometry. These parameters were kept fixed during the final stage of the optimization.

	EOFG		EOVG		VOFG	
	IP	OP	IP	OP	IP	OP
b	1.494	1.502	1.502	1.539	1.905	1.913
d	1.094	1.0906	1.09	1.024	0.970	0.972
g_1	0.213	0.215	0.217	0.219	0.2609	0.2597
g_2	-0.342	-0.345	-0.45	-0.532	-0.6358	-0.634
g_3	0.139	0.1348	0.105	0.050	0.167	0.155
g_4	-0.123	-0.123	-0.133	-0.156	-0.3175	-0.308
g_5	0.021	0.0325	0.107	0.164	0.2653	0.2628
g_6	0.478	0.484	0.499	0.476	0.0348	0.0385
g_7	-0.586	-0.581	-0.589	-0.576	-0.0233	-0.0264
g_8	-0.474	-0.508	-0.623	-0.696	-0.5005	-0.4987
g_9	0.506	0.5224	0.652	0.777	0.7131	0.7151
R/2	1.5075*	1.5075*	1.503	1.5057	1.5075*	1.5075*

Table 3.21: Jastrow and geometry parameters for LiH molecule, Ψ_2.

	Ψ_1[a]		Ψ_2	
	energy	CE [%]	energy	CE [%]
EOVG	-8.0104(2)	28	-8.06082(14)	89
EOFG	-8.0100(2)	27	-8.0601(1)	88
VOVG	-7.9874(2)	0		
VOFG	-7.9851(2)	<0	-8.0575(1)	85

Table 3.22: Variational energies for LiH molecule.

E_T=-8.08 E_h, E_{exact}=-8.0702... E_h (Ref. [2]), E_{HF}=-7.98735 E_h (Ref. [7])
[a]Ref. [20]: E_{var}=-7.91(1) E_h

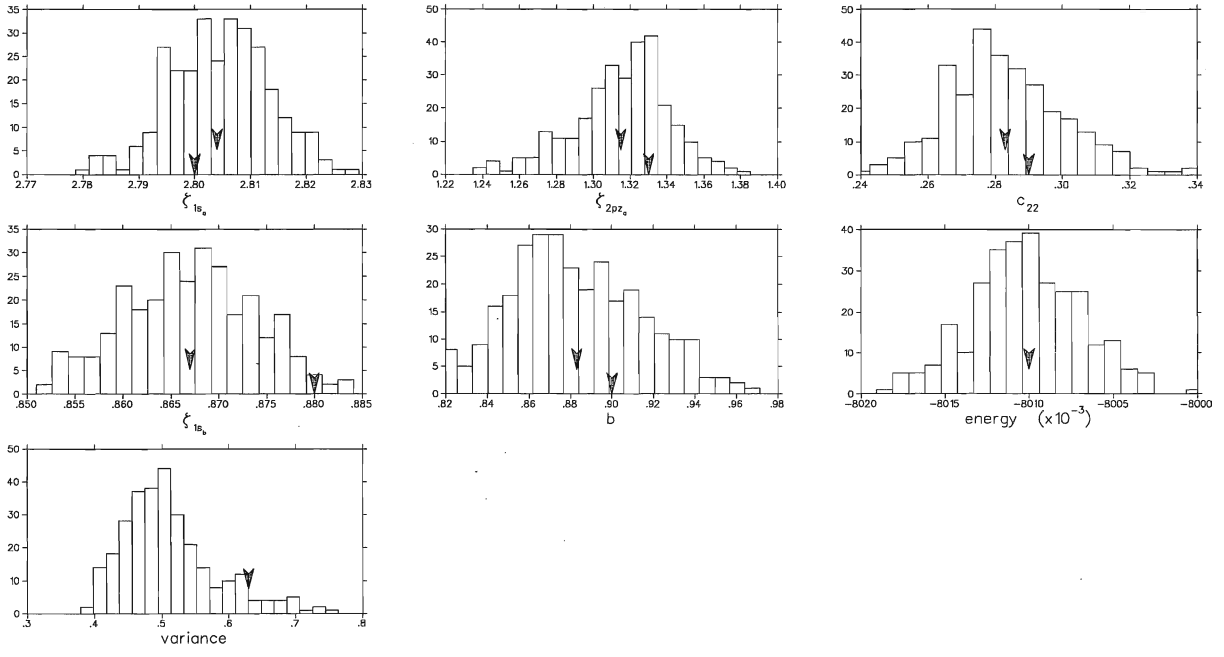

Figure 3.19: Histograms for LiH molecule, Ψ_1, energy optimization with fixed geometry.

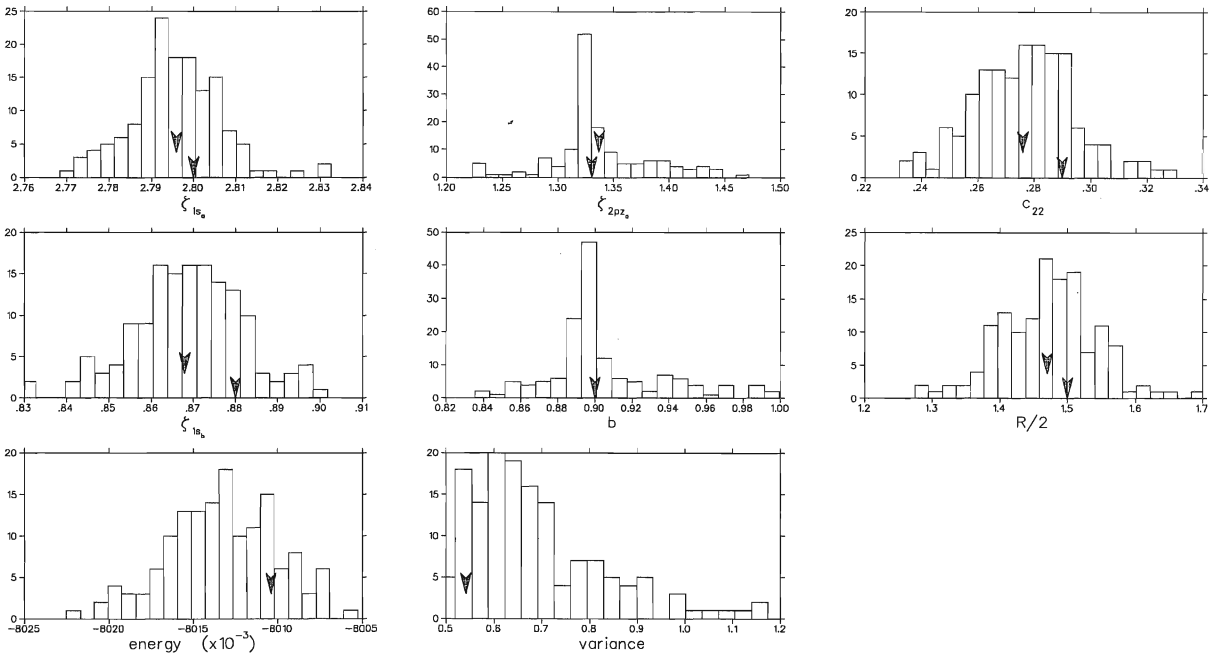

Figure 3.20: Histograms for LiH molecule, Ψ_1, energy optimization with variable geometry.

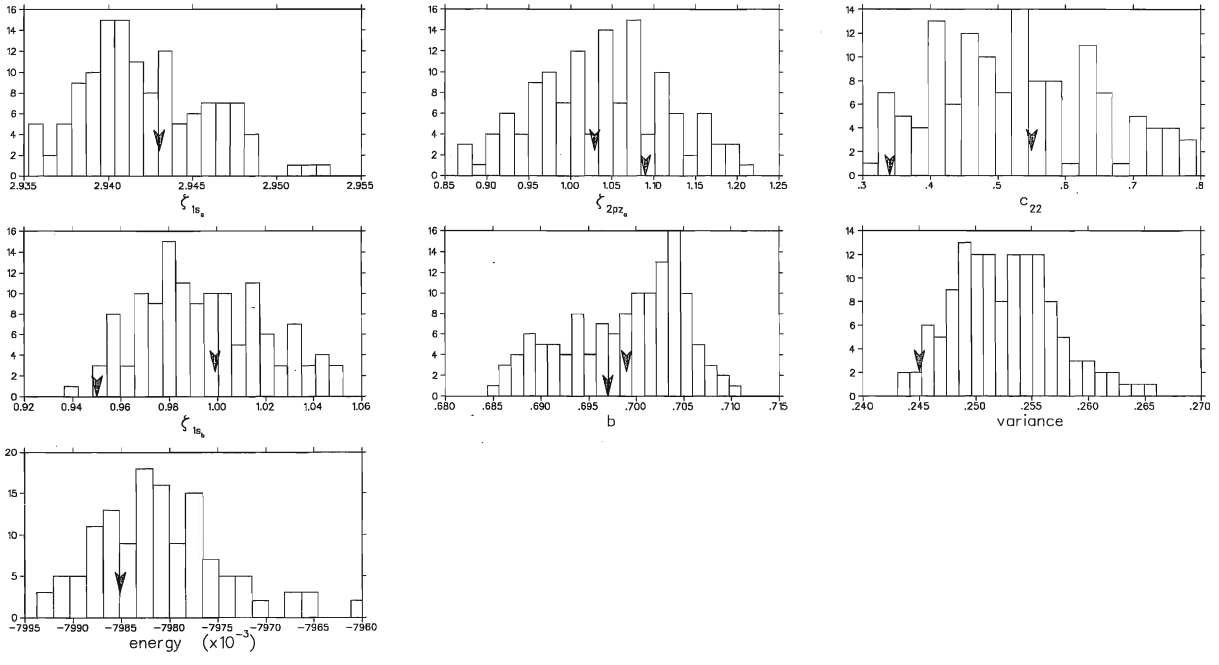

Figure 3.21: Histograms for LiH molecule, Ψ_1, variance optimization with fixed geometry.

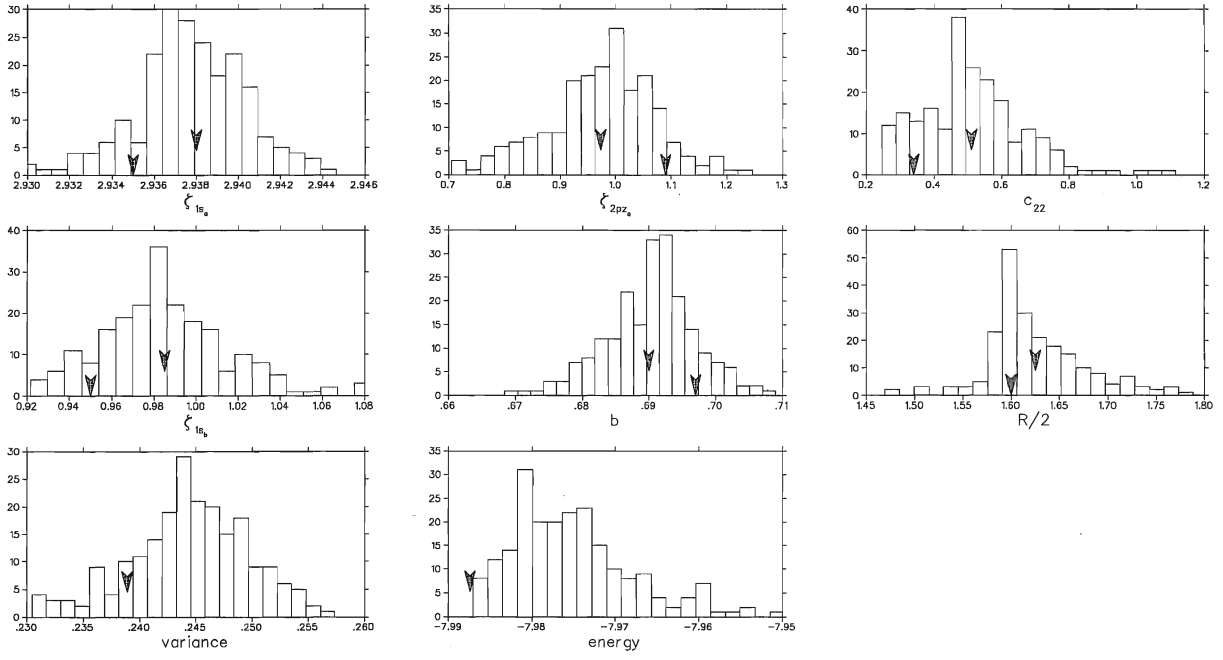

Figure 3.22: Histograms for LiH molecule, Ψ_1, variance optimization with variable geometry.

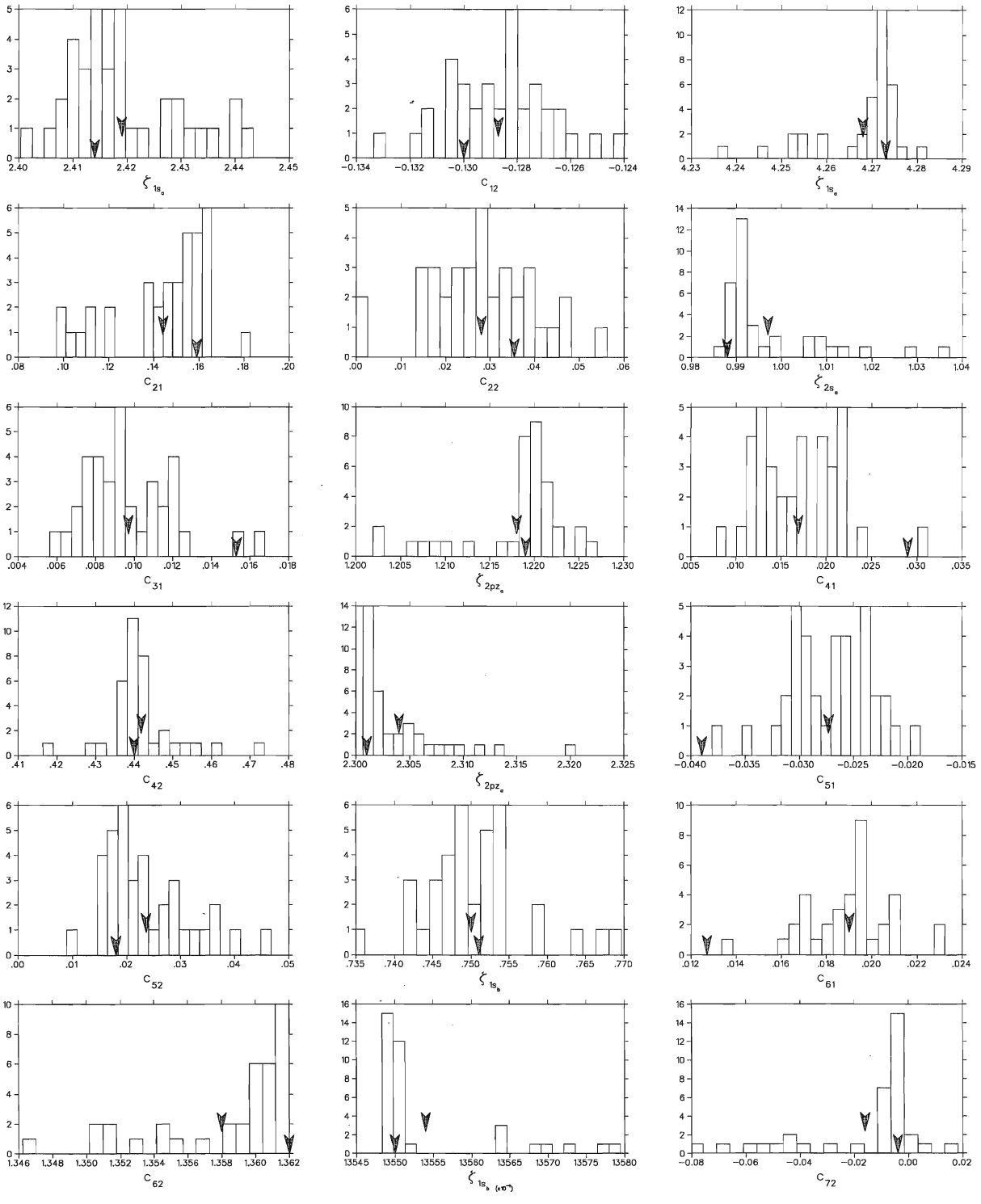

Figure 3.23: Histograms for LiH molecule, Ψ_2, energy optimization with fixed geometry.

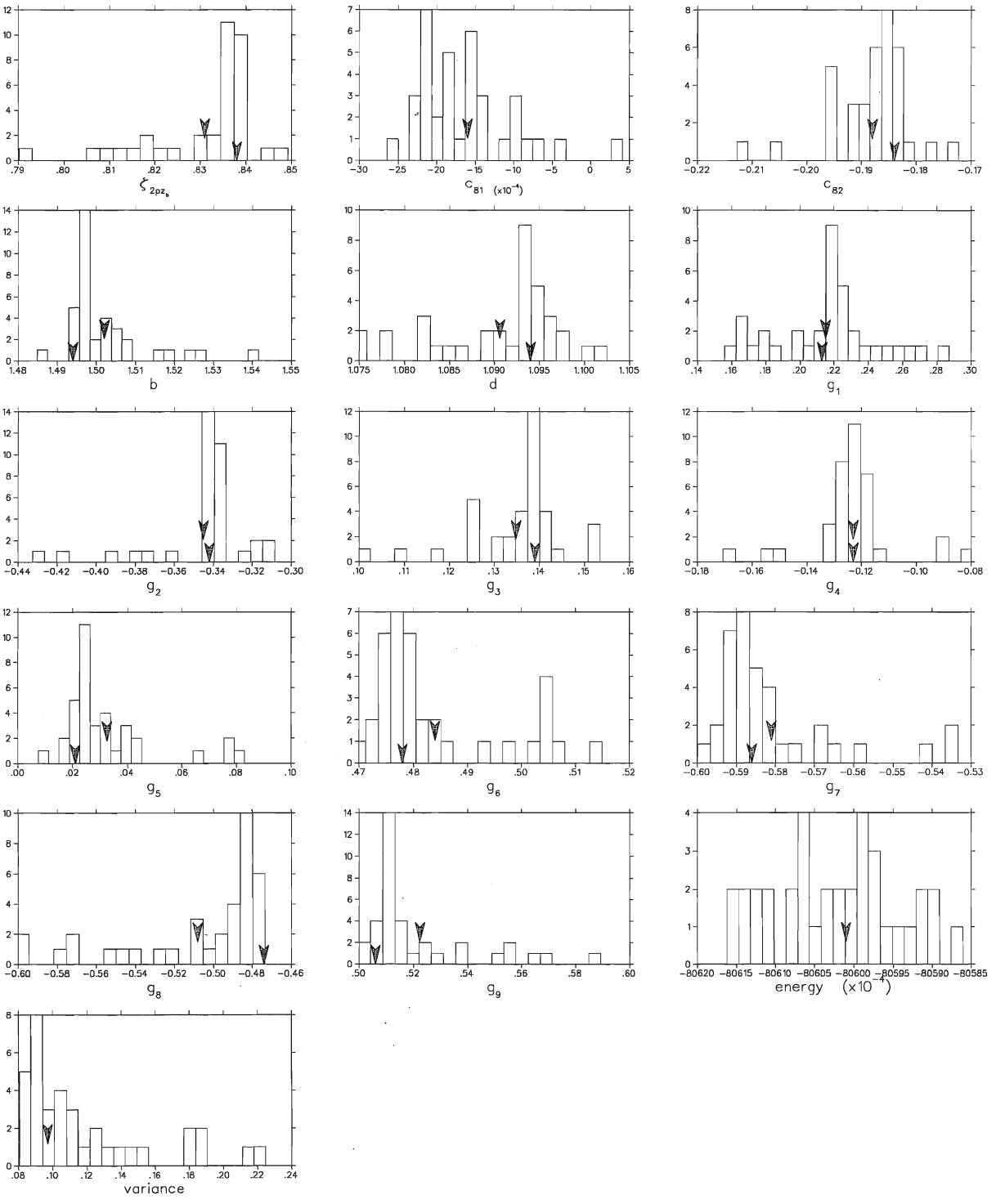

Figure 3.24: Histograms for LiH molecule, Ψ_2, energy optimization with fixed geometry (cont'd).

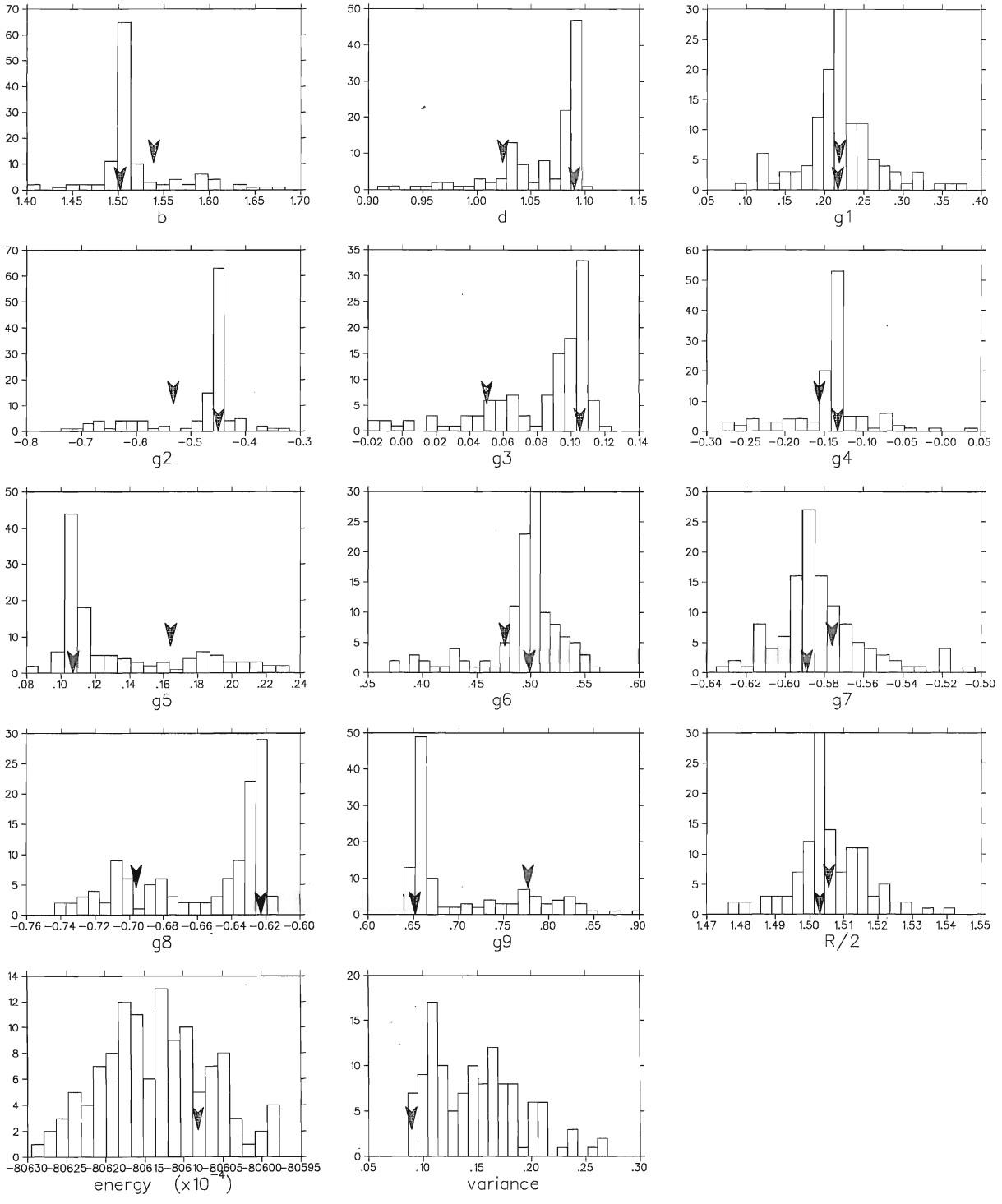

Figure 3.25: Histograms for LiH molecule, Ψ_2, energy optimization with variable geometry.

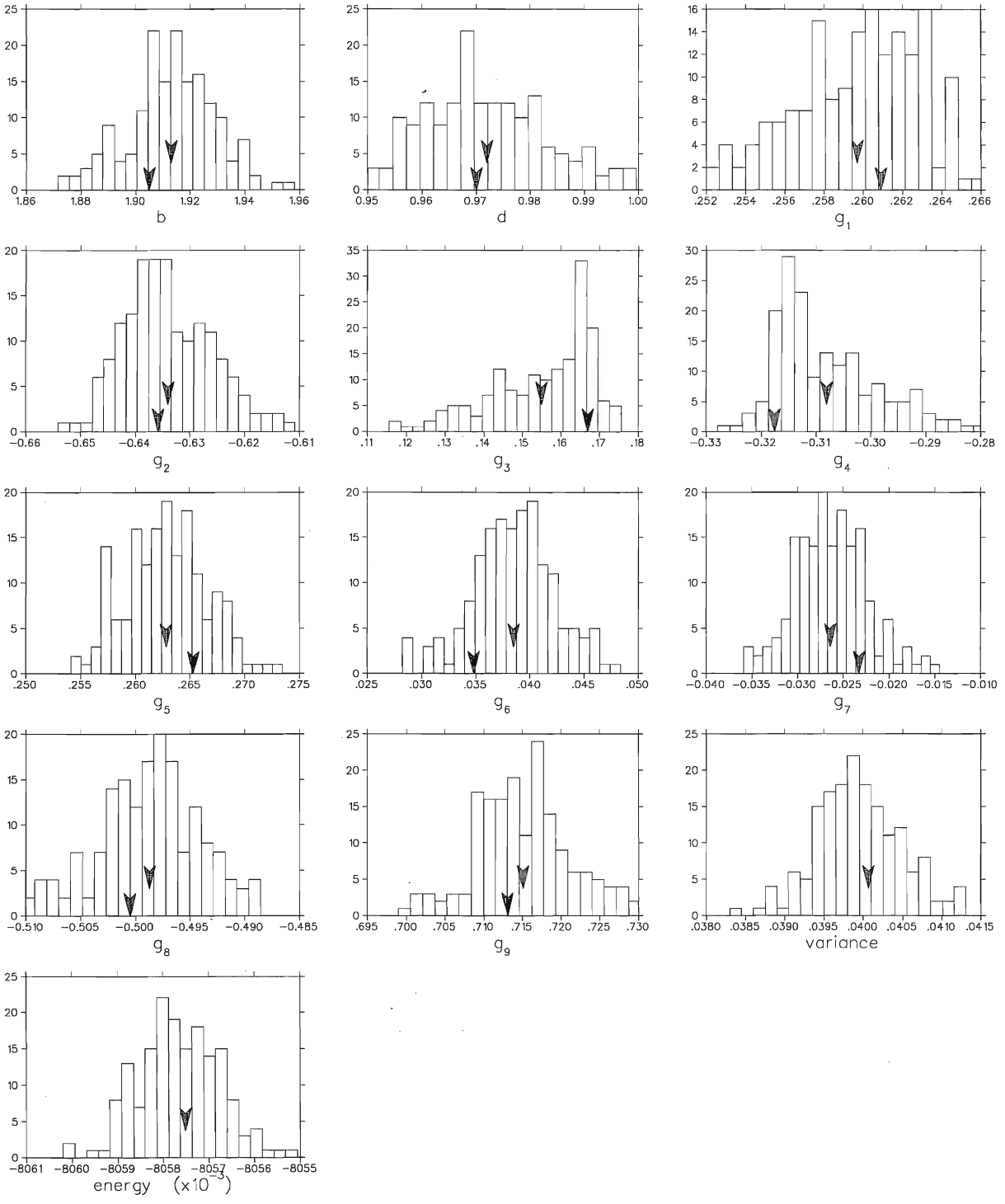

Figure 3.26: Histograms for LiH molecule, Ψ_2, variance optimization with fixed geometry.

3.4 Conclusion

Results of this chapter confirm that histogram filtering method (HFM) is able to perform well for both the variance and energy optimization. Energy optimization yields better results for ground state energy as well as equilibrium bond distance. We can see that as the quality of the trial wave functions improves the variational energies for energy and variance-optimized wave functions tend to approach each other. However, unless we have very accurate trial wave function (as Ψ_3 for He) the energy optimization yields significantly better results. For the equilibrium bond distance determination the energy optimization is clearly superior.

For wave functions where there is possible to compare results we were able to reproduce the exact parameters and energies (Ψ_1 for H_2 and He) or improve the literature values for the energy (Ψ_2 for H_2 and He, Ψ_1 for LiH). The variational energy obtained from the energy optimization of Ψ_2 for LiH is to our knowledge the lowest ever obtained from VMC calculations using explicitly correlated wave functions.

Whether to optimize the variance or the energy is not a new question. Most people in this business prefer to optimize the variance. However, to our knowledge no systematic work has been done to address this issue rigorously. The arguments are mostly qualitative and we believe the strong preference toward the variance optimization is due to the fact that the energy optimization is harder to perform.

The "variance strategy" is to use some sophisticated form of the trial wave function (often with hundreds of variational parameters) and perform a very crude variance optimization using only several thousands configurations[4]. This method yields wave functions with reasonable variational energy and is relatively fast. Usually, wave functions obtained in such way serve as trial (guiding) functions to the DMC. The DMC ground

[4]As we have seen, energy optimization of complicated wave functions requires hundreds of thousands configurations.

state energies are very close to the exact ones.

Our (HFM) approach is to optimize simpler wave functions, but optimize them more thoroughly. This requires more time and human effort, but in the end we have simpler wave functions often with lower energy than those obtained with the previous method. Although in theory only the position of the nodes of the trial functions is important for DMC, in practice, it is advantageous to have a good trial function because the variance of the DMC energy (and other properties) depends on the quality of the trial function. As well a simpler form means faster evaluation of the function values.

We certainly do not claim that the HFM is superior to other methods in every aspect. In a sense it is a complementary point of view on optimization. There is always a trade-off between the quality of a given result and the effort necessary for obtaining it.

For certain class of problems, however, HFM seems to be the natural solution. One of these is the determination of the equilibrium geometry. In the present work we have constrained ourselves to the determination of the equilibrium bond distance, but our aim is to use this technique to optimize geometrically more complicated systems. Another example could be a theoretical study of properties of various types of wave functions. Because HFM "squeezes" the maximum from the wave functions, it can be used as the objective indicator of the quality of these tested functions.

Chapter 4

Non-differential Ground State Properties of Small Systems.

In the previous chapter we presented results of energy and variance optimizations. It turned out that energy-optimized wave functions are superior with respect to both the variational energy and bond distance determination. However, there are other properties of interest which play an important role in chemistry and can be used as additional indicator of the quality of the wave functions.

This chapter contains the results of calculations of various non-differential properties for each system optimized in chapter 3. The last column of the appropriate tables shows the exact literature values for most of the properties. These are known to a very high degree of accuracy for hydrogen molecule and helium, but we do not know of any similar accurate calculations for LiH (except for the dipole moment). In that case we report the values calculated by East, Rothstein and Vrbik [6] by diffusion MC (DMC).

For easier comparison of the various properties reported there is a column in the tables labeled as Order. For example, if we have four columns of properties (as for H_2) the four numbers 2143 in the column Order mean that the property in the second column is closest to the exact value, next closest is the property in the first column, third closest is the property in the the fourth column, and the worst is the property in the the third column.

The quantity given in the last row in each table labeled as $\sum_i |\Delta_i^{rel}|$ measures the overall accuracy of the wave functions. It is the sum of the relative errors, defined as

follows:

$$\sum_i |\Delta_i^{rel}| = \sum_i \frac{|p_i - p_i^{exact}|}{|p_i^{exact}|} ,$$

where the p_i's and p_i^{exact}'s are the i-th calculated and exact property, respectively.

All numbers are in atomic units, and we take z as the inter-molecular axes.

4.1 The hydrogen molecule

Tables 4.1-4.3 show the non-differential properties for hydrogen molecule wave functions Ψ_1-Ψ_3. Most of the entries in tables are self-explanatory, and the rest is defined as follows:

$$\begin{aligned} Q_2 &= R^2/2 + x_1^2 + x_2^2 - z_1^2 - z_2^2 \\ Q_4 &= R^4/8 - [x_1^4 + x_2^4 + z_1^4 + z_2^4 - 6(x_1^2 z_1^2 + x_2^2 z_2^2)] \\ u_z^2 &= (z_1 - z_2)^2 \\ u_c^2 &= (x_1 - x_2)^2 + (y_1 - y_2)^2 \ . \end{aligned}$$

The Q_2 and Q_4 are the quadrupole and the hexadecapole moments, respectively. u_c and u_z are the transverse and longitudinal projections of the inter-electronic distance r_{12}, respectively.

	EOFG	EOVG	VOFG	VOVG	Order	Literature
r_{12}	2.0457(3)	2.0374(2)	2.1227(3)	2.8149(3)	2314	2.16895[a]
r_{12}^2	5.145(2)	5.104(1)	5.544(2)	9.665(2)	2314	5.63239[a]
r_{12}^3	15.237(8)	15.060(6)	17.072(8)	38.65(2)		
r_{12}^{-1}	0.6523(1)	0.65498(8)	0.62911(8)	0.47423(6)	2314	0.58737[a]
r_{12}^{-2}	0.704(1)	0.7108(9)	0.656(1)	0.3790(7)	3412	0.51827[b]
r_{1a}	1.5599(3)	1.5525(2)	1.6091(3)	2.2898(4)	2134	1.54880[a]
r_{1a}^2	3.061(1)	3.0330(9)	3.262(1)	6.580(2)	2134	3.03635[a]
r_{1a}^3	7.174(5)	7.083(4)	7.922(5)	22.08(1)		
r_{1a}^{-1}	0.9035(2)	0.9079(2)	0.8749(2)	0.6443(1)	2134	0.91279[a]
r_{1a}^{-2}	1.561(6)	1.588(7)	1.459(5)	0.895(3)		
$r_{1a}r_{1b}$	2.713(1)	2.6936(8)	2.918(1)	5.183(1)	1234	2.70391[a]
$r_{1a}r_{2a}$	2.4343(6)	2.4101(5)	2.5894(6)	5.245(1)	2134	2.32141[a]
$r_{1a}r_{2b}$	2.4340(6)	2.4103(5)	2.5888(6)	5.241(1)	2134	2.38484[a]
z_1z_2	-0.0004(3)	-0.00002(2)	-0.0002(2)	0.0002(6)	1324	−0.15963[a]
x_1x_2	-0.0002(2)	0.0001(1)	0.0000(2)	0.0003(2)	1324	−0.05510[a]
$(z_1^2+z_2^2)/2$	1.0786(4)	1.0671(3)	1.1436(4)	2.4881(5)	2134	1.02297[a]
$(x_1^2+x_2^2)/2$	0.7466(3)	0.7424(2)	0.8139(3)	1.1725(4)	1234	0.76169[a]
$(r_1^2+r_2^2)/2$	2.5714(8)	2.5522(6)	2.7720(7)	4.8329(9)	2134	2.54635[a]
Q_2	0.3174(8)	0.3120(6)	0.3220(8)	0.864(1)	2314	0.45684[c]
Q_4	0.174(8)	0.175(6)	0.184(8)	1.700(9)	3214	0.2826[d]
u_z^2	2.1570(4)	2.1344(5)	2.2881(5)	4.974(1)	2314	2.3652[a]
u_z	1.1643(1)	1.1579(1)	1.1974(1)	1.7938(2)	2314	1.2441[b]
u_c^2	2.9847(5)	2.9696(5)	3.2576(6)	4.691(1)	2314	3.2672[a]
u_c	1.4875(1)	1.483(1)	1.5538(1)	1.8651(2)	2314	1.5699[b]
u_c^{-1}	1.1275(2)	1.1304(2)	1.0794(1)	0.8993(2)	2314	1.0404[b]
$\sum_i \lvert\Delta_i^{rel}\rvert$	3.849	3.871	3.847	19.88	2314	

Table 4.1: Non-differential properties for H_2 molecule, Ψ_1.

[a]Nearly exact value, Ref. [13]
[b]Derived from 36 correlated Gaussian geminals, Ref. [22]
[c]Ref. [18]
[d]Ref. [8]

	EOFG	EOVG	VOFG	VOVG	Order	Literature
r_{12}	2.1855(2)	2.1696(2)	2.2895(2)	2.7998(3)	2134	2.16895[a]
r_{12}^2	5.751(1)	5.670(1)	6.316(2)	9.332(2)	2134	5.63239[a]
r_{12}^3	17.557(6)	17.198(7)	20.234(7)	35.71(1)		
r_{12}^{-1}	0.59122(6)	0.59580(7)	0.56400(6)	0.45632(5)	1234	0.58737[a]
r_{12}^{-2}	0.5450(6)	0.5532(7)	0.4965(8)	0.3212(4)	2314	0.51827[b]
r_{1a}	1.5774(2)	1.5641(2)	1.6412(2)	2.0668(3)	2134	1.54880[a]
r_{1a}^2	3.1333(9)	3.0814(9)	3.397(1)	5.361(2)	2134	3.03635[a]
r_{1a}^3	7.425(4)	7.246(4)	8.412(4)	16.343(8)		
r_{1a}^{-1}	0.9004(2)	0.9076(2)	0.8632(2)	0.7052(2)	2134	0.91279[a]
r_{1a}^{-2}	1.591(5)	1.620(7)	1.447(4)	1.041(3)		
$r_{1a}r_{1b}$	2.7734(8)	2.7350(8)	3.0412(9)	4.430(1)	2134	2.70391[a]
$r_{1a}r_{2a}$	2.4438(4)	2.4037(4)	2.6456(5)	4.1611(8)	2134	2.32141[a]
$r_{1a}r_{2b}$	2.4897(5)	2.4467(5)	2.6902(5)	4.322(1)	2134	2.38484[a]
z_1z_2	-0.1264(2)	-0.1225(2)	-0.1318(2)	-0.3103(5)	2314	−0.15963[a]
x_1x_2	-0.0532(1)	-0.0527(1)	-0.0600(1)	-0.0799(2)	1234	−0.05510[a]
$(z_1^2+z_2^2)/2$	1.1380(3)	1.1171(3)	1.2230(3)	2.0568(5)	2134	1.02297[a]
$(x_1^2+x_2^2)/2$	0.7524(2)	0.7452(2)	0.8416(3)	1.0692(4)	1234	0.76169[a]
$(r_1^2+r_2^2)/2$	2.6427(5)	2.6073(5)	2.9062(6)	4.1953(8)	2134	2.54635[a]
Q_2	0.2102(6)	0.2375(6)	0.2186(7)	0.358(1)	4231	0.45684[c]
Q_4	0.085(6)	0.123(6)	0.084(7)	0.31(2)	3241	0.2826[d]
u_z^2	2.5289(5)	2.4798(5)	2.7105(6)	4.736(1)	2134	2.3652[a]
u_z	1.2821(1)	1.2689(1)	1.3246(2)	1.7852(2)	2134	1.2441[b]
u_c^2	3.2227(6)	3.1916(6)	3.6064(7)	4.5978(9)	1234	3.2672[a]
u_c	1.5560(1)	1.5485(1)	1.6469(2)	1.8584(2)	1234	1.5699[b]
u_c^{-1}	1.0579(2)	1.0630(2)	0.9979(1)	0.8872(1)	1234	1.0404[b]
$\sum_i \lvert\Delta_i^{rel}\rvert$	2.056	1.956	3.217	11.08	2134	

Table 4.2: Non-differential properties for H_2 molecule, Ψ_2.

[a]Nearly exact value, Ref. [13]
[b]Derived from 36 correlated Gaussian geminals, Ref. [22]
[c]Ref. [18]
[d]Ref. [8]

	EOFG	EOVG	VOFG	VOVG	Order	Literature
r_{12}	2.1629(3)	2.1598(3)	2.1504(3)	2.2666(3)	1234	2.16895[a]
r_{12}^2	5.609(2)	5.594(2)	5.571(2)	6.170(2)	1234	5.63239[a]
r_{12}^3	16.885(8)	16.820(8)	16.94(1)	19.67(2)		
r_{12}^{-1}	0.58994(7)	0.59094(7)	0.59412(7)	0.56112(7)	1234	0.58737[a]
r_{12}^{-2}	0.5244(8)	0.5257(6)	0.5305(7)	0.4696(5)	1234	0.51827[b]
r_{1a}	1.5520(2)	1.5488(2)	1.5530(3)	1.6532(3)	2134	1.54880[a]
r_{1a}^2	3.040(1)	3.028(1)	3.048(1)	3.445(2)	1234	3.03635[a]
r_{1a}^3	7.158(5)	7.116(5)	7.276(8)	8.70(1)		
r_{1a}^{-1}	0.9106(2)	0.9125(2)	0.9045(2)	0.8526(2)	2134	0.91279[a]
r_{1a}^{-2}	1.608(5)	1.617(7)	1.554(5)	1.408(5)		
$r_{1a}r_{1b}$	2.709(1)	2.700(1)	2.722(1)	3.021(1)	2134	2.70391[a]
$r_{1a}r_{2a}$	2.3524(5)	2.3424(5)	2.3620(5)	2.6720(6)	2134	2.32141[a]
$r_{1a}r_{2b}$	2.4010(5)	2.3914(5)	2.4051(5)	2.7327(6)	2134	2.38484[a]
z_1z_2	-0.1226(2)	-0.1225(2)	-0.1065(2)	-0.1336(2)	2341	−0.15963[a]
x_1x_2	-0.0663(2)	-0.0661(2)	-0.0606(2)	-0.0674(2)	3214	−0.05510[a]
$(z_1^2+z_2^2)/2$	1.0175(3)	1.0152(3)	1.0050(4)	1.1398(5)	1234	1.02297[a]
$(x_1^2+x_2^2)/2$	0.7661(3)	0.7633(3)	0.7769(4)	0.8382(4)	2134	0.76169[a]
$(r_1^2+r_2^2)/2$	2.5495(6)	2.5424(7)	2.5583(8)	2.8165(9)	1234	2.54635[a]
Q_2	0.4787(7)	0.4678(7)	0.5252(8)	0.651(1)	2134	0.45684[c]
Q_4	0.34(1)	0.32(1)	0.40(2)	0.65(3)	2134	0.2826[d]
u_z^2	2.2802(9)	2.2754(6)	2.2245(7)	2.547(1)	1234	2.3652[a]
u_z	1.2150(2)	1.2134(2)	1.1961(2)	1.2863(2)	1243	1.2441[b]
u_c^2	3.330(1)	3.3184(8)	3.349(1)	3.623(1)	2134	3.2672[a]
u_c	1.5884(2)	1.5854(2)	1.5894(2)	1.6524(3)	2134	1.5699[b]
u_c^{-1}	1.0218(2)	1.0241(2)	1.0199(2)	0.9821(2)	2134	1.0404[b]
$\sum_i \lvert\Delta_i^{rel}\rvert$	0.857	0.764	1.313	3.737	2134	

Table 4.3: Non-differential properties for H_2 molecule, Ψ_3.

[a]Nearly exact value, Ref. [13]
[b]Derived from 36 correlated Gaussian geminals, Ref. [22]
[c]Ref. [18]
[d]Ref. [8]

4.2 The helium atom

Tables 4.4—4.6 show the non-differential properties for the helium atom.

	EO	VO	Order	Literature
$r_1 + r_2$	1.7939(1)	1.7055(2)	12	1.85894[a]
$r_1^2 + r_2^2$	2.1542(3)	1.9469(4)	12	2.38697[a]
$r_1^3 + r_2^3$	3.2353(7)	2.7787(8)		
$r_1^{-1} + r_2^{-1}$	3.3778(3)	3.5543(5)	12	3.37663[a]
$r_1^{-2} + r_2^{-2}$	11.60(3)	12.90(5)	12	12.0348[a]
r_{12}	1.37254(9)	1.3078(1)	12	1.42207[a]
r_{12}^2	2.3277(3)	2.1107(4)	12	2.51644[a]
r_{12}^3	4.671(1)	4.027(1)	12	5.3080[b]
r_{12}^{-1}	0.97426(7)	1.0206(1)	12	0.94582[a]
r_{12}^{-2}	1.553(3)	1.696(3)	12	1.46477[a]
$\sum_i \lvert\Delta_i^{rel}\rvert$	0.489	1.111	12	

Table 4.4: Non-differential properties for He atom, Ψ_1.

[a]Ref. [17]
[b]Ref. [5]

	EO	VO	Order	Literature
$r_1 + r_2$	1.85142(9)	1.84731(8)	12	1.85894[a]
$r_1^2 + r_2^2$	2.3589(3)	2.3601(3)	21	2.38697[a]
$r_1^3 + r_2^3$	3.8627(9)	3.902(1)		
$r_1^{-1} + r_2^{-1}$	3.3671(2)	3.3835(2)	21	3.37663[a]
$r_1^{-2} + r_2^{-2}$	11.87(2)	11.97(2)	21	12.0348[a]
r_{12}	1.41062(7)	1.40372(7)	12	1.42207[a]
r_{12}^2	2.4844(3)	2.4748(3)	12	2.51644[a]
r_{12}^3	5.244(1)	5.262(1)	21	5.3080[b]
r_{12}^{-1}	0.95447(5)	0.96447(5)	12	0.94582[a]
r_{12}^{-2}	1.487(2)	1.5263(9)	12	1.46477[a]
$\sum_i \lvert\Delta_i^{rel}\rvert$	0.090	0.124	12	

Table 4.5: Non-differential properties for He atom, Ψ_2.

[a]Ref. [17]
[b]Ref. [5]

	EO	VO	Order	Literature
$r_1 + r_2$	1.8597(1)	1.8536(2)	12	1.85894[a]
$r_1^2 + r_2^2$	2.3883(4)	2.3672(5)	12	2.38697[a]
$r_1^3 + r_2^3$	3.936(1)	3.872(2)		
$r_1^{-1} + r_2^{-1}$	3.3758(3)	3.3793(5)	12	3.37663[a]
$r_1^{-2} + r_2^{-2}$	12.11(4)	11.99(2)	21	12.0348[a]
r_{12}	1.4225(1)	1.4161(2)	12	1.42207[a]
r_{12}^2	2.5212(4)	2.4985(6)	12	2.51644[a]
r_{12}^3	5.334(2)	5.261(2)	12	5.3080[b]
r_{12}^{-1}	0.94639(8)	0.9511(1)	12	0.94582[a]
r_{12}^{-2}	1.464(1)	1.482(2)	12	1.46477[a]
$\sum_i \lvert\Delta_i^{rel}\rvert$	0.015	0.053	12	

Table 4.6: Non-differential properties for He atom, Ψ_3.

[a]Ref. [17]
[b]Ref. [5]

4.3 The lithium hydride molecule

Tables 4.7–4.8 show the non-differential properties for LiH molecule. u_c and u_z have the same meaning as described in 4.1 and μ is the dipole moment. All quantities are calculated for parallel and antiparallel electrons separately.

		EOFG	EOVG	VOFG	VOVG	Order	Literature
parallel	r_{12}^2	14.174(5)	13.728(5)	13.276(4)	14.865(3)	3214	13.2(1)[a]
antiparallel	r_{12}^2	9.663(4)	9.410(4)	9.433(3)	10.345(5)	1432	9.9(1)[a]
parallel	r_{12}	3.5724(6)	3.5124(6)	3.4715(5)	3.6843(5)	3214	3.42(1)[a]
antiparallel	r_{12}	2.6867(4)	2.6526(4)	2.6483(5)	2.7687(5)	1342	2.72(1)[a]
parallel	r_{12}^{-1}	0.31767(5)	0.32366(5)	0.32390(5)	0.30314(5)	3214	0.333(2)[a]
antiparallel	r_{12}^{-1}	0.68186(8)	0.68508(8)	0.69943(8)	0.68627(9)	1243	0.66(1)[a]
parallel	r_{12}^{-2}	0.11982(5)	0.12452(5)	0.12300(5)	0.10696(4)	3124	0.132(1)[a]
antiparallel	r_{12}^{-2}	1.218(3)	1.225(4)	1.300(2)	1.288(4)	1243	1.17(1)[a]
parallel	u_z^2	10.860(4)	10.427(4)	10.269(4)	11.749(4)	3214	8.98(6)[a]
antiparallel	u_z^2	6.306(2)	6.068(2)	6.376(4)	7.178(4)	2134	5.55(4)[a]
parallel	u_z	3.0593(6)	2.9922(6)	2.9928(5)	3.2179(6)	3124	2.718(6)[a]
antiparallel	u_z	1.9804(4)	1.9427(4)	1.9893(5)	2.1102(5)	2134	1.86(1)[a]
parallel	u_c^2	3.314(2)	3.301(2)	3.007(2)	3.116(2)	1243	4.2(1)[a]
antiparallel	u_c^2	3.357(2)	3.343(2)	3.057(2)	3.167(2)	1243	4.3(1)[a]
parallel	u_c	1.5404(4)	1.5374(4)	1.4703(4)	1.4938(4)	1243	1.70(2)[a]
antiparallel	u_c	1.4751(4)	1.4723(4)	1.4115(3)	1.4331(4)	1243	1.65(2)[a]
parallel	u_c^{-1}	1.1389(3)	1.1414(4)	1.1922(3)	1.1788(3)	1243	1.09(1)[a]
antiparallel	u_c^{-1}	1.3900(3)	1.3913(3)	1.4456(3)	1.4350(3)	1243	1.29(2)[a]
	μ	-2.972(1)	-2.915(1)	-2.514(3)	-2.751(3)	4312	−2.30[b]
$\sum_i \lvert\Delta_i^{rel}\rvert$		1.947	1.743	1.983	2.771	2134	

Table 4.7: Non-differential properties for LiH molecule, Ψ_1.

[a]Ref. [6]
[b]Ref. [7]

		EOFG	EOVG	VOFG	Order	Literature
parallel	r_{12}^2	12.966(7)	13.034(7)	13.228(7)	321	13.2(1)[a]
antiparallel	r_{12}^2	9.456(6)	9.521(6)	9.595(6)	321	9.9(1)[a]
parallel	r_{12}	3.4051(8)	3.4136(8)	3.4420(8)	213	3.42(1)[a]
antiparallel	r_{12}	2.6629(7)	2.6718(7)	2.6822(7)	321	2.72(1)[a]
parallel	r_{12}^{-1}	0.33449(7)	0.33373(8)	0.33031(8)	213	0.333(2)[a]
antiparallel	r_{12}^{-1}	0.6724(1)	0.6714(1)	0.6712(1)	321	0.66(1)[a]
parallel	r_{12}^{-2}	0.13236(7)	0.13178(7)	0.12885(7)	213	0.132(1)[a]
antiparallel	r_{12}^{-2}	1.177(3)	1.171(3)	1.185(4)	213	1.17(1)[a]
parallel	u_z^2	9.096(5)	9.123(6)	9.251(6)	123	8.98(6)[a]
antiparallel	u_z^2	5.493(3)	5.511(3)	5.531(3)	321	5.55(4)[a]
parallel	u_z	2.7518(9)	2.7562(9)	2.7797(9)	123	2.718(6)[a]
antiparallel	u_z	1.8481(5)	1.8516(5)	1.8538(5)	321	1.86(1)[a]
parallel	u_c^2	3.869(5)	3.911(5)	3.977(5)	321	4.2(1)[a]
antiparallel	u_c^2	3.963(5)	4.010(5)	4.063(5)	321	4.3(1)[a]
parallel	u_c	1.6426(8)	1.6502(8)	1.6692(8)	321	1.70(2)[a]
antiparallel	u_c	1.5836(8)	1.5919(8)	1.6042(8)	321	1.65(2)[a]
parallel	u_c^{-1}	1.1080(5)	1.1066(5)	1.0879(5)	321	1.09(1)[a]
antiparallel	u_c^{-1}	1.3447(4)	1.3438(4)	1.3326(4)	321	1.29(2)[a]
	μ	-2.356(2)	-2.373(2)	-2.438(2)	123	-2.30[b]
$\sum_i \lvert\Delta_i^{rel}\rvert$		0.477	0.4233	0.4234	312	

Table 4.8: Non-differential properties for LiH molecule, Ψ_2.

[a]Ref. [6]
[b]Ref. [7]

4.4 Conclusion

The aim of this chapter was to look at properties of wave functions other than the variational energy. Particularly, we were interested how the energy-optimized wave functions compete with the variance-optimized ones. The people who prefer to optimize the variance often suggest that although the variational energy of the variance-optimized wave functions could be worse the electron distribution and henceforth the nodal planes positions more resembles the exact ones. The argument is that for non-exact wave functions the local energy tends to diverge near the nodes and the variance optimization tries to suppress this divergence and hence to find such an electron distribution which most resembles the non-divergent exact one.

Let us first have a look at the results for hydrogen molecule and helium which are node-less. The overall performance of all the wave functions for H_2 and He clearly indicates that the EO is doing better. Ψ_3 for H_2 and He are examples where almost every property is better for the energy-optimized wave functions. It is interesting to note that the EOVG yields better results even if the bond distance is slightly different from the experimental one (for which the exact properties were calculated). Except for the very tight "victory" of the VOFG in the case of Ψ_1 for H_2 over the EOFG, the EOFG is always close behind the EOVG.

To make the comparison for LiH is more difficult. The literature values have large error bars and for many properties it is crucial to know the exact values with higher precision to make reliable comparisons. Generally, it seems that for the calculated properties the variance and energy optimization yield comparable quality results. Note, however, that all the properties for LiH are functions of the inter-electronic distances. It would be interesting to compare results of calculations for properties which are functions of nuclear-electronic distances. Again, the EOVG gives better results than the EOFG.

The calculations done on LiH can neither support nor disprove the hypothesis of the superiority of the properties derived from variance-optimized wave functions. We are working on the DMC calculations of the properties for LiH in order to get more accurate values.

Chapter 5

Comments and Suggestions

In this thesis we have shown that HFM can be successfully applied to small systems with simple geometry. However, if the method is to "survive" it has to be able to optimize more complicated systems in terms of number of electrons and geometrical structure.

The further study can be oriented to several directions. First, we can focus on standard, i.e. non-geometrical, optimization. The staged optimization proved to be a good method of dealing with many-parameter wave functions. It is really difficult to predict how many parameters can HFM potentially handle but it is clear that the human filtering by eye can not be pushed much farther than to several dozen of parameters. On the other hand, we are confident that this procedure can be (at least partially) automatized (however, short look at the histograms will be always helpful).

In our work we were extremely "picky" in our filtering and tried to inspect every possible structure suggested by a small gap, a tiny irregularity in shape etc.. We sacrificed much effort and time to lower the variational energy by every small fraction of a mE_h. We think that for a truly many-parameter wave functions such approach is neither necessary nor possible, and we can obtain high quality wave function from filtering only the most obvious structures and taking the appropriate averages.

The optimization of a geometry of systems with more than one geometrical parameter is a very challenging project. In the near future we would like to focus on the optimization of a water molecule. This system has two geometrical parameters (the O–H bond distance and the H–O–H angle) and ten electrons. There is a lot of theoretical and experimental

results available for the sake of comparison with ones we derive.

We feel that further development and tuning of the HFM itself is possible. We may, for example, investigate how the results of optimization (the histogram pattern) depend on the severity of the stopping criteria[1] of the search algorithm and the choice of the search algorithm itself. Another way of improving the HFM is to analyze the numerical instabilities which cause the "blow up" of the energy optimization. If we could circumvent this problem (without introducing a bias) the optimization efficiency would improve significantly.

We are also interested to find out if similar approaches to HFM can be applied to different problems related to optimization where some background "noise" is present. An application of HFM to problems in computational biology is in progress.

[1] In our work we have used rather strong criteria for stopping the optimization which makes it quite long in certain cases.

Appendix A

Choice of the Transition Probability

Although any form of the transition probability $T(\mathbf{R}' \leftarrow \mathbf{R})$ is theoretically suitable, the efficiency of the Metropolis algorithm depends strongly on that choice. In this section we will present one effective form of $T(\mathbf{R}' \leftarrow \mathbf{R})$ for the case of single atoms. Because the Hamiltonian in this case exhibits spherical symmetry, it is natural to use spherical polar coordinates to move the electrons. Furthermore it turns out to be useful to move only one electron at a time. In our work we have used the scheme proposed by Langfelder [14].

The probability of moving from a point (r, θ, ϕ) to a point with coordinates belonging to the intervals $\langle r', r' + dr' \rangle, \langle \theta', \theta' + d\theta' \rangle, \langle \theta', \phi' + d\phi' \rangle$ consist of two components

$$P[(r' + dr', \theta' + d\theta', \phi' + d\phi') \leftarrow (r, \theta, \phi)] = f(r' \leftarrow r) g_{r'}(\theta', \phi') dr' d\theta' d\phi' \ .$$

The probability function $g_{r'}(\theta', \phi')$ does not depend on the values of r, θ, ϕ and is chosen to be the uniform distribution over a sphere with diameter r'. Thus

$$\frac{g_{r'}(\theta', \phi') d\theta' d\phi'}{d\Sigma'} = \frac{g_{r'}(\theta', \phi') d\theta' d\phi'}{r'^2 \sin\theta' d\theta' d\phi'} = C \ ,$$

where $d\Sigma'$ is the infinitesimal area element of the surface of the sphere. The constant C can be calculated from the normalization condition $\int g_{r'}(\theta', \phi') d\theta d\phi = 1$ and equals to $1/4\pi r'^2$. The probability density $g_{r'}(\theta', \phi')$ thus has the form

$$g_{r'}(\theta', \phi') = \frac{\sin\theta'}{4\pi}$$

The function $f(r' \leftarrow r)$ has the form

$$f(r' \leftarrow r) = \frac{1}{\rho^2} r' e^{-r'^2/2\rho^2} \ ,$$

where $\rho = \rho(r)$ is a function of the electron-nuclear distance r

$$\rho(r) = \frac{r}{[\max\{\alpha, 1 + \beta(r - \gamma)\}]^{1/2}} .$$

The parameters α, β, γ were found by trial and error and are chosen as

$$\begin{aligned} \alpha &\in \langle 0.01, 0.05 \rangle \\ \beta &= \zeta_{\text{most diffuse orbital}} \\ \gamma &\in \langle 0.1, 0.5 \rangle . \end{aligned}$$

Now, the transition probability density is found by definition :

$$\begin{aligned} T((r', \theta', \phi') \leftarrow (r, \theta, \phi)) &= \frac{P[(r' + dr', \theta' + d\theta', \phi' + d\phi') \leftarrow (r, \theta, \phi)]}{dV} \\ &= \frac{f(r' \leftarrow r) g_{r'}(\theta', \phi') dr' d\theta' d\phi'}{r'^2 \sin\theta' dr' d\theta' d\phi'} \\ &= \frac{1}{4\pi r'^2} f(r' \leftarrow r) . \end{aligned}$$

The remaining question is how to generate random numbers sampled from the distributions $f(r' \leftarrow r)$ and $g_{r'}(\theta', \phi')$; i.e., how to find the corresponding functions $r'(u), \theta'(u), \phi'(u)$, where u is random number uniformly distributed over the interval $(0, 1)$. Let us consider the case of $g_{r'}(\theta', \phi')$ first. Because it does not explicitly depend on ϕ', ϕ' is chosen uniformly from interval $(0, 2\pi)$ (with probability density $1/2\pi$) thus

$$\phi' = 2\pi u .$$

The new angle θ' is generated from the probability density $\sin\theta'/2$. The cumulative probability distribution function

$$F(\theta') = \int_0^{\theta'} \frac{\sin t}{2} dt = \frac{1}{2}(1 - \cos\theta')$$

is related to the inverse function $u(\theta')$ as follows:

$$F(\theta') = u$$

and the solution is

$$\cos\theta' = 1 - 2u \ .$$

We apply the same procedure to $f(r' \leftarrow r)$ and we get

$$F(r') = \int_0^{r'} \frac{1}{\rho^2} t e^{-t^2/2\rho^2} dt = 1 - e^{-r'^2/2\rho^2} = u$$

with the solution

$$r' = \rho\sqrt{-2\ln(1-u)} = \rho\sqrt{-2\ln u} \ .$$

The last equality is due to the fact that both u and $1-u$ are random numbers uniformly distributed over the interval $(0,1)$.

For diatomic molecules the Hamiltonian does not exhibit central symmetry and the above approach is not efficient. In our work we have used the simplest form of the transition probability density—uniform moves in Cartesian coordinates

$$\begin{aligned} x' &= x + D(1-2u) \\ y' &= y + D(1-2u) \\ z' &= z + D(1-2u) \ , \end{aligned}$$

where the parameter D is chosen by trial and error and ranges from 0.5 to 1.0 .

Appendix B

Conjugate Gradient Method

This method was originally designed to minimize convex quadratic functions

$$f(\mathbf{x}) = \mathbf{x}^T\mathbf{A}\mathbf{x} - \mathbf{b}^T\mathbf{x} \; .$$

This function is minimized when its gradient $\nabla f = \mathbf{A}\mathbf{x} = \mathbf{b}$ is zero. Therefore, we solve a system of linear equations $\mathbf{A}\mathbf{x} = \mathbf{b}$, where $\mathbf{A}$ is a positive-definite matrix. The procedure, which guarantees convergence to the proper $\mathbf{x}^*$ after maximum N steps (with exact arithmetic), can be described in following steps:

- Choose an arbitrary starting point $\mathbf{x}_0$
- Set $\mathbf{h}_0 = \mathbf{g}_0 = -(\mathbf{A}\mathbf{x}_0 - \mathbf{b})$
- For $k = 0$ until convergence
  1. $\lambda_k = \frac{\mathbf{h}_k^T\mathbf{g}_k}{\mathbf{h}_k^T\mathbf{A}\mathbf{h}_k}$
  2. $\mathbf{x}_{k+1} = \mathbf{x}_k + \lambda_k\mathbf{h}_k$
  3. $\mathbf{g}_{k+1} = \mathbf{g}_k - \lambda_k\mathbf{A}\mathbf{h}_k$
  4. $\gamma_k = \frac{\mathbf{g}_{k+1}^T\mathbf{g}_{k+1}}{\mathbf{g}_k^T\mathbf{g}_k}$
  5. $\mathbf{h}_{k+1} = \mathbf{g}_{k+1} + \gamma_k\mathbf{h}_k$

The vectors $\mathbf{h}$ and $\mathbf{g}$ satisfy the orthogonality and conjugacy conditions

$$\mathbf{g}_i^T\mathbf{g}_j = 0\,; \quad \mathbf{h}_i^T\mathbf{g}_j = 0\,; \quad \mathbf{h}_i^T\mathbf{A}\mathbf{h}_j \qquad \text{for } i \neq j \; .$$

But, the $\mathbf{A}$ is unknown. However, if we take $\mathbf{g}_k = -\nabla f(\mathbf{x}_k)$ for some point $\mathbf{x}_k$, and we find the local minimum $\mathbf{x}_{k+1}$ along the direction $\mathbf{h}_k$ and set $\mathbf{g}_{k+1} = -\nabla f(\mathbf{x}_{k+1})$, then this is the same vector as constructed in step 3. We can prove it as follows:

$$\begin{aligned} \mathbf{g}_k &= -\nabla f(\mathbf{x}_k) = -\mathbf{A}\mathbf{x}_k + \mathbf{b} \\ \mathbf{g}_{k+1} &= -\nabla f(\mathbf{x}_{k+1}) = -\mathbf{A}\mathbf{x}_{k+1} + A\mathbf{b} = -\mathbf{A}(\mathbf{x}_k + \lambda\mathbf{h}_k) + \mathbf{b} = \mathbf{g}_k - \lambda\mathbf{A}\mathbf{h}_k \ . \end{aligned}$$

What is left is to find the expression for λ. Because λ is the minimum along the direction $\mathbf{h}_k$, we can write

$$\frac{\mathrm{d}}{\mathrm{d}\lambda} f(\mathbf{x}_k + \lambda\mathbf{h}_k) = 0 \ .$$

Using the fact that

$$\frac{\mathrm{d}}{\mathrm{d}\lambda} f(\mathbf{x}_k + \lambda\mathbf{h}_k) = \mathbf{h}_k^T \nabla f(\mathbf{x}_k + \lambda\mathbf{h}_k) \ ,$$

we can finish the proof:

$$0 = \mathbf{h}_k^T \nabla f(\mathbf{x}_k + \lambda\mathbf{h}_k) = \mathbf{h}_k^T(\mathbf{A}(\mathbf{x}_k + \lambda\mathbf{h}_k) - \mathbf{b}) = \mathbf{h}_k^T(\mathbf{A}\mathbf{x}_k - \mathbf{b}) + \lambda\mathbf{h}_k^T\mathbf{A}\mathbf{h}_k \ .$$

Comparing the left-most side with the right-most side we get the expression for λ:

$$\begin{aligned} -\mathbf{h}_k^T\mathbf{g}_k &= \lambda\mathbf{h}_k^T\mathbf{A}\mathbf{h}_k \\ \lambda &= \frac{-\mathbf{h}_k^T\mathbf{g}_k}{\mathbf{h}_k^T\mathbf{A}\mathbf{h}_k} \ . \end{aligned}$$

We see that this is indeed the same formula as written in step 1.

So far we were talking about minimization of exact convex quadratic form. However, this algorithm can be used for minimization of arbitrary function. The nearer to the local minimum we are, the better the approximation of convex quadratic form for the function is. It turns out that a small change of the value γ_k (proposed by Polak and Ribiere) can sometimes significantly improve the algorithm. Now we can rewrite the previous algorithm into a form suitable for our purpose:

- Choose an arbitrary starting point $\mathbf{x}_0$
- Set $\mathbf{h}_0 = \mathbf{g}_0 = -\nabla f(\mathbf{x}_0)$
- For $k = 0$ until convergence
  1. $\mathbf{x}_{k+1}$ = minimum along the line with direction $\mathbf{h}_k$
  2. $\mathbf{g}_{k+1} = -\nabla f(\mathbf{x}_{k+1})$
  3. $\gamma_k = \frac{(\mathbf{g}_{k+1}-\mathbf{g}_k)^T \mathbf{g}_{k+1}}{\mathbf{g}_k^T \mathbf{g}_k}$ (Polak-Ribiere improvement)
  4. $\mathbf{h}_{k+1} = \mathbf{g}_{k+1} + \gamma_k \mathbf{h}_k$

For line search we use the standard Brent's method (see [19]).

Very important and not trivial task is to decide when to stop the minimization process. In principle we should stop when the gradient vanishes. In practice, however, this (almost) never happens and we have to use different criterion (or combination of criteria). In our optimization we stopped when the following criterion was satisfied:

$$\text{abs}(f_i - f_{i-1}) \leq \frac{1}{2} tol[\text{abs}(f_i) + \text{abs}(f_{i+1}) + \epsilon] \ .$$

The parameter *tol* is the tolerance (we used 10^{-6}) and ϵ is a small number (typically 10^{-10}) which prevents the criterion to be too severe if the function value is close to zero.

Appendix C

Decomposition of the Slater Determinant

In section 1.6 we have given a qualitative justification of the Slater determinant decomposition. Next, we give a mathematical proof of the validity of this decomposition. What we want to prove is that the expectation value of any operator which does not include spin operators is the same for both the full Slater detarminant and the decomposed one. The operator $\mathcal{A}$ is the antisymmetrizer and is defined as follows:

$$\mathcal{A}\{\Phi_1(1)\Phi_2(2)\ldots\Phi_k(k)\} = \sum_{\pi}(-1)^{|\pi|}\Phi_{\pi(1)}(1)\Phi_{\pi(2)}(2)\ldots\Phi_{\pi(k)}(k) \ .$$

In the above expression we sum over all permutations π. The expectation value of some operator $\hat{O}$ is defined as

$$\langle\hat{O}\rangle_{\Psi} = \frac{\langle\Psi|\hat{O}|\Psi\rangle}{\langle\Psi|\Psi\rangle} \ . \tag{C.1}$$

Because the denominator has the same form as the numerator (for $\hat{O} = \hat{1}$), in the following we will focus on the expression in the numerator.

$$\begin{aligned}\langle\Psi|\hat{O}|\Psi\rangle \ = \ & \langle\mathcal{A}\{\Phi_1(1)\Phi_2(2)\ldots\Phi_{n^\uparrow}(n^\uparrow)\Phi_{n^\uparrow+1}(n^\uparrow+1)\ldots\Phi_n(n)\}|\hat{O}|\mathcal{A}\{\Phi_1(1)\Phi_2(2)\ldots \\ & \Phi_{n^\uparrow}(n^\uparrow)\Phi_{n^\uparrow+1}(n^\uparrow+1)\ldots\Phi_n(n)\}\rangle = \\ & (n!)\langle\mathcal{A}\{\Phi_1(1)\Phi_2(2)\ldots\Phi_{n^\uparrow}(n^\uparrow)\Phi_{n^\uparrow+1}(n^\uparrow+1)\ldots\Phi_n(n)\}|\hat{O}|\Phi_1(1)\Phi_2(2)\ldots \\ & \Phi_{n^\uparrow}(n^\uparrow)\Phi_{n^\uparrow+1}(n^\uparrow+1)\ldots\Phi_n(n)\rangle \ .\end{aligned}$$

The spin-orbitals labeled from 1 to $n^\uparrow$ have spin up and the rest $(n^\uparrow+1,\ldots,n)$ have spin down. The spin-up and spin-down orbitals are orthogonal (the operator $\hat{O}$ does not

affect the spin part):

$$\langle \Phi_i | \hat{O} | \Phi_j \rangle = 0, \qquad i = 1, 2, \ldots n^{\uparrow};\ j = n^{\uparrow} + 1, \ldots, n\ .$$

Because of this property only those permutations which do not assign a number greater than $n^{\uparrow}$ to a number less or equal to $n^{\uparrow}$ (and vice versa) contribute. Therefore we can break down the antisymmetrizer of the whole product into two parts:

$$\mathcal{A}\{\Phi_1(1)\Phi_2(2)\ldots\Phi_{n^{\uparrow}}(n^{\uparrow})\Phi_{n^{\uparrow}+1}(n^{\uparrow}+1)\ldots\Phi_n(n)\} \rightarrow$$
$$\mathcal{A}\{\Phi_1(1)\Phi_2(2)\ldots\Phi_{n^{\uparrow}}(n^{\uparrow})\}\mathcal{A}\{\Phi_{n^{\uparrow}+1}(n^{\uparrow}+1)\ldots\Phi_n(n)\}$$

We can now write:

$$\begin{aligned}\langle \Psi | \hat{O} | \Psi \rangle \;=\; & (n!)\langle \mathcal{A}\{\Phi_1(1)\Phi_2(2)\ldots\Phi_{n^{\uparrow}}(n^{\uparrow})\}\mathcal{A}\{\Phi_{n^{\uparrow}+1}(n^{\uparrow}+1)\ldots\Phi_n(n)\}|\hat{O}|\Phi_1(1)\Phi_2(2)\ldots \\ & \Phi_{n^{\uparrow}}(n^{\uparrow})\Phi_{n^{\uparrow}+1}(n^{\uparrow}+1)\ldots\Phi_n(n)\rangle = \\ & \frac{(n!)}{(n^{\uparrow}!)(n^{\downarrow}!)}\langle \mathcal{A}\{\Phi_1(1)\Phi_2(2)\ldots\Phi_{n^{\uparrow}}(n^{\uparrow})\}\mathcal{A}\{\Phi_{n^{\uparrow}+1}(n^{\uparrow}+1)\ldots\Phi_n(n)\}|\hat{O}| \\ & \mathcal{A}\{\Phi_1(1)\Phi_2(2)\ldots\Phi_{n^{\uparrow}}(n^{\uparrow})\}\mathcal{A}\{\Phi_{n^{\uparrow}+1}(n^{\uparrow}+1)\ldots\Phi_n(n)\}\rangle\ .\end{aligned}$$

If we substitute the above expression to the numerator and analogous one to the denominator of the C.1, the common numerical factors cancel out and we obtain the desired formula

$$\frac{\langle \Psi | \hat{O} | \Psi \rangle}{\langle \Psi | \Psi \rangle} = \frac{\langle \Psi^{\uparrow}\Psi^{\downarrow} | \hat{O} | \Psi^{\uparrow}\Psi^{\downarrow} \rangle}{\langle \Psi^{\uparrow}\Psi^{\downarrow} | \Psi^{\uparrow}\Psi^{\downarrow} \rangle}\ ,$$

where $\Psi^{\uparrow} = \mathcal{A}\{\Phi_1(1)\Phi_2(2)\ldots\Phi_{n^{\uparrow}}(n^{\uparrow})\}$ and $\Psi^{\downarrow} = \mathcal{A}\{\Phi_{n^{\uparrow}+1}(n^{\uparrow}+1)\ldots\Phi_n(n)\}$.

Bibliography

[1] S.A. Alexander, R.L. Colwell, H.J. Monkhorst, and J.D. Morgan III. *J. Chem. Phys.*, 95:6622, 1991.

[2] B. Chen and J. B. Anderson. *J. Chem. Phys.*, 102:4491, 1995.

[3] E. Clementi and C. Roetti. *Atomic Data and Nuclear Data Tables*, 14:177, 1974.

[4] C.A. Coulson. *Trans. Faraday Soc.*, 33:1479, 1937.

[5] F. Arias de Saavedra, I. Porras F.J. Gálvez, and E. Buendía. *J. Phys. B*, 28:3123, 1995.

[6] A. L. L. East, S. M. Rothstein, and J. Vrbik. *J. Chem. Phys.*, 88:4880, 1988.

[7] N.C. Handy, R.J. Harrison, P.J. Knowles, and H.F. Schaefer III. *J. Phys. Chem.*, 88:4852, 1984.

[8] G. Karl, J.D. Poll, and L. Wolniewicz. *Can. J. Phys.*, 53:1781, 1975.

[9] T. Kato. *Comm. Pure Appl. Math*, 10:151, 1957.

[10] S.D. Kenny, G. Rajagopal, and R.J. Needs. *Phys. Rev. A*, 51:1898, 1995.

[11] W. Kolos and C.C.J. Roothaan. *Rev. Mod. Phys.*, 32:219, 1960.

[12] W. Kolos, K. Szalewicz, and H.J. Monkhorst. *J. Chem. Phys.*, 84:3278, 1986.

[13] W. Kolos and L. Wolniewicz. *J. Chem. Phys.*, 43:2429, 1965.

[14] P. Langfelder. Master's thesis, Brock University, (1997).

[15] N. Metropolis, A. W. Rosenbluth, M. N. Rosenbluth, A. M. Teller, and E. Teller. *J. Chem. Phys.*, 21:1087, 1953.

[16] L. Pauling and E.B. Wilson, Jr. *Introduction to Quantum Mechanics.* McGraw-Hill, New York, N.Y., 1935. p. 224.

[17] C.L. Pekeris. *Phys. Rev.*, 115:1216, 1959.

[18] J.D. Poll and L. Wolniewicz. *J. Chem. Phys.*, 68:3053, 1978.

[19] W. H. Press, S. A. Teukolsky, W. T. Vettering, and B. P. Flannery. *Numerical recipes in Fortran.* Cambridge University Press, Cambridge, UK, second edition, 1992.

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

[21] K.E. Schmidt and J.W. Moskowitz. *J. Chem. Phys.*, 93:4172, 1990.

[22] B.S. Sharma and A.J. Thakkar. *J. Phys. B*, 17:3405, 1984.

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

[24] S.C. Wang. *Phys. Rev.*, 31:579, 1928.