INTERNATIONAL ATOMIC ENERGY AGENCY
UNITED NATIONS EDUCATIONAL, SCIENTIFIC AND CULTURAL ORGANIZATION

INTERNATIONAL CENTRE FOR THEORETICAL PHYSICS
34100 TRIESTE (ITALY) - P.O.B. 586 - MIRAMARE - STRADA COSTIERA 11 - TELEPHONES: 224281/2/3/4/5/6
CABLE: CENTRATOM - TELEX 460392-1

SMR/111 - 23

SECOND SUMMER COLLEGE IN BIOPHYSICS

30 July - 7 September 1984

Lecture 1: Liquid Water: 2,3 and 4 body interactions potentials from quantum mechanics.

Lecture 2: Liquid Water: Monte Carlo and molecular dynamics simulations.

G. CORONGIU
IBM Corporation
IS/TG
Poughkeepsie, New York 12602
U.S.A.

1.2 DEFINITION OF CHEMICAL COMPLEXITY

The complexity of a chemical system, considered as a discrete and numerable ensemble of particles, is proportional to the number of mutually dependent variables characterizing a given aspect of the system. In our implicit definition of complexity, we use two criteria:

1. The number of atoms or molecules present in the system and
2. The size of the largest molecule in the system. The size is defined as the number of atoms in the largest molecule of the system.

When one deals with biological systems, the above conditions are necessary to define the system, but probably not sufficiently, since in defining a biological system one must include conditions, whereby, for a limited interval of time, structural forms evolve from previous forms at little energy cost.

Let us consider two examples of a complex system: in the first, complexity results from the number of molecules, in the second, from the size of the largest molecule.

A single water molecule is here defined as a simple system, a cluster of about ten water molecules represents a case of intermediate complexity and liquid water is an example of a complex system. Here complexity is due to the number of the components. The mutually dependent variables are the positions and orientations of the water molecules. The sugar molecule (or a phosphate group), the sugar-phosphate-sugar complex, and a protein are examples of simple, intermediate, and complex chemical systems, where complexity is due to the size of the largest component of the system. The mutually dependent variables are the angles of internal rotations.

The two criteria, number and size, cover a large number of chemically complex systems (see Figure 1). In a simple system, the appropriate model (quantum mechanics) assumes as particles the nuclei and the electrons; thus the dominant statistics is the one of Fermi-Dirac. In complex systems the appropriate particles are atoms or group of atoms; the model is classical in the limit and the Boltzmann statistics (and time fluctuations) are essential for a proper description. It follows, as a corollary, that intramolecular forces play a dominant role in simple systems, whereas intermolecular forces, and torsional and rotational barriers are dominant in complex systems. As a second corollary, a valid goal in small systems is the determination of the exact geometrical relationship between the component atoms. On the contrary, the probabilistic nature of the distribution of the atoms or groups of atoms is of basic importance in a complex system. The notion of probability distribution brings in most naturally the need for the concept of entropy. This concept has been somewhat overlooked in quantum chemistry, since subtly built into the Schroedinger wave mechanical representation, whereas it is more easily grasped in classical chemical particles.

In conclusion, depending on the complexity of the system, one uses either quantum mechanics, statistical mechanics, or fluid mechanics and thermodynamics. In biological systems an amino acid, a protein, and a living cell provide an example of systems appropriate to the above three models.

One can safely predict that important advances in the description of complex chemical systems will be obtained by attempting to connect quantum mechanics with statistical mechanics, and statistical mechanics with fluid dynamics and thermodynamics. As noted, the need is not as

much i:. "formal" connection, that essentially is available, but mainly in a "operational" connection, for example, by means of computational methods.

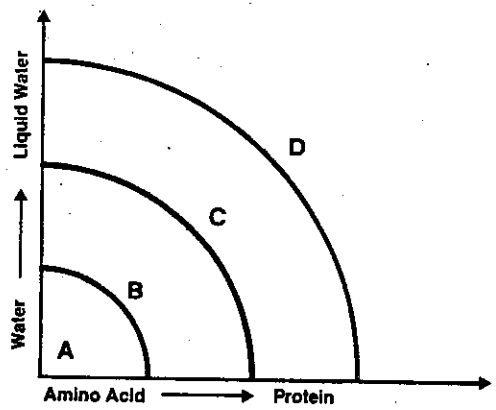


Figure 1. Simple, to intermediate, to complex chemical systems (A,B,C, respectively): the two criteria are Size of the Largest Molecule and Number of Molecules. At the limit (region D) starts the continuous description of matter.

1.4 A GENERAL METHOD FOR SIMULATIONS OF A COMPLEX CHEMICAL SYSTEM

The most basic selection one makes to describe a system with some given model, relates to the choice of the adopted statistic; this choice brings about a specific definition of the nature of the "objects" (particles or medium) and concomitant model-equations. "Natural objects" for chemical systems are either nuclei and electrons or atoms and molecules or some continuous medium (rigid or nonrigid). The corresponding statistics are either the Fermi-Dirac statistic or the classical statistic (inclusive of Boltzmann distribution); the corresponding equations are either the Schroedinger equation or Newton's equations (adopted for discrete particles or for continuous medium). In the latter case, of particular chemical interest are aspects dealing with the thermodynamic of reversible and irreversible systems, problems of linearity and nonlinearity, single or multiple solution for stationary states, etc.

Today, quantum chemistry has sufficiently evolved so as to allow us to consider realistically complex systems as an object for numerical simulations. As previously pointed out, however, it is basic to reject the indiscriminate use of quantum chemistry (intended as approximate solutions of the Schroedinger equation) as the only tool to be used for a complex system's description. Such indiscriminate use brings about necessarily gross over-simplifications that could be avoided by using quantum-chemistry as a first step of a many-step methodology aimed at realistic simulations of complex chemical systems.

We briefly summarize an operational procedure to simulate complex chemical systems. Most of the theoretical foundations needed in our approach were known long ago (21). However, little has been done to operationally link different methods that are traditionally kept independent one from the other, despite the fact they represent the successive step of our approach or of any physically reasonable approach. Details on the proposed method will be given in later chapters; here we wish only to present a very brief account of what will follow. Let us start by considering some appropriate subsystem of our complex system. The simplest and most immediate subsystems are the individual, separated molecules. As previously indicated, the size of a molecule for which "decent" quantum chemical simulations are feasible is constantly increasing, either because of new methods or because of the increased performance of computers. "Decent" simulations are those that make use of adequate basis sets and that include electronic correlation corrections, when needed.

The general method is presented as a five-step technique; the output of step (i-1) constitutes the input to step (i). This constraint is basic to develop an operational procedure.

Step 1: Quantum Chemistry

Ab initio quantum chemical simulations are performed either on one or few molecules. If the complex system is composed of only one macromolecule, then we shall use methods described in the second chapter. Alternatively, the complex chemical system might be composed of many molecules. If ab initio computations are performed in order to obtain inter molecular interactions, then we proceed at first by obtaining the two-body interactions. If three-body interactions are required, then the same procedure used to obtain the two-body correction is followed. Details are given in the fourth chapter. The introduction of the reaction field is assumed under the heading of "quantum chemistry", not because it is a generally accepted procedure, but because it is a necessary interaction in most problems dealing with complex systems.

Step 2: Construction of Interaction Potentials

From the numerical potentials of Step 1, one obtains analytical potentials. This second step constitutes a most critical aspect to operationally connect quantum chemistry to statistical mechanics. The potentials must be constrained in such a way as to be of an analytically simple form, fast for computational use, transferable from molecule to molecule, standardized both in the form and reliability and, finally, amenable to gradual refinements and extensions. By design we have neglected the possibility to use experimental data as the starting parameters to obtain the potentials. One reason is that often there are not sufficient experimental data available; a second reason is that it is often arbitrary to extract two- and three-body contributions from an experiment obtained at conditions corresponding to a full n-body potential.

Basic to this step is an analysis of how an atom can be characterized when in a molecule. From the valance concepts of Lewis and Langmuir, the valence bond approximation was derived by Heitler, London, Slater and Pauling; an important concept in that language is the valence state concept. Alternatively, from the Lewis-Langmuir concepts, we can arrive at the Mulliken-Hartree-Fock Molecular Orbitals approximation. Heuristically, an important concept connected to the MO model is that partitioning of the electronic density known as the electron population analysis. From the valence bond approximation, we have translated into the MO theory the equivalent of the concept of Valence State and called it "molecular orbital valence state", MOVS. An atom in a molecule can be characterized by its value of the MOVS energy and by the

value of its net charge: hybridization, charge transfer and nearest neighbors are included in this characterization. By definition, we characterize an atom of a given atomic number and in a specific electronic environment with a label referred to as "class" for that atomic specification. In this way, atoms of the same atomic number and in the same molecular environment belong to the same class. Hence, by construction, the identification of a "class" for an atom is transferable from molecule to molecule, as long as the atom has the same Z value and class label. As a corollary, atom-atom pair potentials are transferable from molecule to molecule. Historically, we can note the following evolution in the characterization of atoms in molecules: the first and gross characterization is the one corresponding to the atomic number, (for example, for the carbon atoms, Z=6); then the hybridization characterization (for example, sp, sp^2 and sp^3) and the valence state concepts; now the "class" characterization is added, providing a finer characterization. For example, the carbon atoms in CH_4,CH_3-CH_3, CH_3-NH_2, CH_3-OH, CH_3-SH, have all the same valence, but each one of the carbon atoms belongs to a different class, as clear from the different values of the two parameters, namely the MOVS energy and the net charge value.

The intermolecular interaction potentials must be limited to the two-body Hartree-Fock equivalent interaction, but must include dispersion corrections at the two-body level and induction corrections at the many body level. As later explained in detail, such corrections can be obtained both easily and accurately for intermolecular interactions.

Step 3: Static Properties

The availability of atom-atom pair potential allows us to easily pass to statistical thermodynamics. Static properties are first analyzed. A basic tool in this step is the Monte Carlo, MC, method (22) that allows us to introduce temperature averaging in the complex system. We generally use the MC method at constant volume with periodic boundary conditions and at constant pressure for clusters studies. The introduction of temperature eliminates for example, the unphysical (but currently used) approximation to study solutions at zero temperature. As known, the reactivity of a system is related to its free energy and not only to the internal energy; in this regard we note that the simulation of entropy is now becoming more and more feasible using MC techniques.

Step 4: Dynamical Properties

In this step the time parameter is introduced, for example, in the standard form of molecular dynamics MD (23). Prerequisite for this step is the availability of pair-potentials from Step 1. As an initial condition to describe the system, we use the final configuration obtained from Step 3. Work is in progress on this step in our group. We note, however, that from the current and past literature on applications of MD one could easily obtain a realistic estimate on the importance of this step. The transport aspect is one among several that can be simulated from the time parameter introduction.

Step 5: Continuoum Representation

In this step, the basic coefficients needed to solve, for example, the diffusion equations of a flow are obtained from Steps 3 and 4. As in the case of Step 4, we are only beginning work at this problem, but the very ample literature in fluido-dynamics coupled with the ample literature on biological dissipative systems, and time or/and space fluctuations (24) should be enough to let one understand how much rewarding it

will be to unify the scope of Step 5 with the outputs of Steps 3 and 4. Traditionally, this step is the less operationally connected to quantum mechanics and quantum chemistry.

In our opinion, the theory and simulations of complex chemical systems have suffered because of the lack of a general framework, as for example, the one above outlined. Essentially, as with physico-chemical experiments, one can analyze either simple or complex systems without artificial limitations (like the problem of having too many electrons, etc.) so we wish to reach the same type of operative freedom in theoretical simulations. We wish to solve for Newton's equations of comple systems at different levels of resolution. If the particles of the system are electrons and point-charge nuclei (high resolution level) than we impose on Newton's equations the constraint of the quantum numbers (forced on by the Fermi-Dirac statistic). If there are no "particles" in the system, but a continuous distribution of matter (low resolution level), then we impose on Newton's equations the Rayleigh numbers (or its equivalent) as constraints, representing boundary conditions and conservations laws. When we still talk of discreate distribution of particles, but these are atoms and molecules rather than electrons and nuclei, then we are at the intermediate resolution level. Different aspects of a given chemical problem require different levels of resolution hence different statistics and, therefore, different constraints to the equations of motion.

1.8) Choice of the Basis Set for SCF-LCAO-MO Functions - Let us briefly comment on the selection criteria for the basis sets. In general, the choice depends on what one wishes to obtain from a wave function. For example, one might wish low accuracy in the total energy, but fair accuracy in the valence electrons, or high accuracy in the inner shells and low accuracy in the valence electrons. Clearly, the strategy in selecting wavefunction accuracy should be a function of the assumed goals. We shall consider the molecule of water as an example.

The basis set for the hydrogen and oxygen atoms are those computed recently by F. van Duijneveldt[29] and are not reported here since too many tables would be required. We have used from 2 to 7 s functions for the hydrogen atom (designated as H/2,3,...7) and for the oxygen atom the smallest basis set is 4/2 namely 4 s and 2 functions of p_x, p_y, p_z type) up to 13/8 (namely 13 s functions and 8 functions of p_x, p_y and p_z type).

With the basis set for atomic hydrogen and oxygen atoms, a number of computations for the water molecule were performed. The molecular geometry for these computations is the experimental equilibrium geometry. In Table 1 the Total Energy (in a.u.) is given for 68 different basis sets; in this table the hydrogen and the oxygen functions have been used without contraction. From the table it is clear that the Hartree-Fock limit for a basis set containing only s and p functions on the oxygen and only s function on the hydrogen is about -76.026 a.u. In addition, if we wish an accuracy of 0.001 a.u. (which is adequate for Hartree-Fock computations) a basis set of 6s functions on the hydrogen and 11s functions plus 7p functions on the oxygen (11s, 7p, 6s) is sufficient. If one is interested in an accuracy of 0.01 a.u., then a set

with 9s, 5p for oxygen and a 5s for hydrogen seems to be adequate.

Table 1

TOTAL ENERGY FOR H_2O (IN a.u.) WITHOUT POLARIZATION*

H / O	2	3	4	5	6	7
4/2	-75.15051	-75.17415	-75.17880			
5/2	-75.43926	-75.45778	-75.45501			
6/2	-75.50145	-75.52124	-75.52585			
5/3	-75.74707	-75.77026	-75.77419			
6/3	-75.80813	-75.83298	-75.83798			
7/3	-75.86667	-75.88745	-75.89138	-75.91008	-75.91171	-75.91212
8/3			-75.90618	-75.97500	-75.97614	-75.97602
7/4			-75.97129	-75.99073	-75.99187	-75.99175
8/4			-75.98710	-75.99489	-75.99603	-75.99592
9/4			-75.99126	-76.01467	-76.01567	-76.01545
9/5			-76.00153	-76.01721	-76.01818	-76.01796
10/5			-76.01419	-76.02118	-76.02189	-76.02178
10/6			-76.01823	-76.02234	-76.02313	-76.02292
11/6			-76.01946	-76.02267	-76.02345	-76.02325
12/6			-76.01982	-76.02303	-76.02382	-76.02362
11/7			-76.02068	-76.02381	-76.02457	-76.02435
12/7			-76.02112	-76.02424	-76.02486	-76.02466
13/7			-76.02120	-76.02438	-76.02492	-76.02471
13/8				-76.02499	-76.02533	-76.02552

*The first column X/Y gives the number X of s functions and the number Y of p functions used on the oxygen atom; the numerals 2, 3, ...7 in the first row gives the number of s functions used on each hydrogen atom.

These comments should be taken with care however; there is no good physical reason, really, for selecting an 11s, 7p, 6s set rather than a 12s, 6p, 6s or an 11s, 6p, 6s set. Regarding computational time, however, a 12s, 6s, 6s set is cheaper than an 11s, 7p, 6s set, since in molecule the p functions are used three times ($2p_x$, $2p_y$, $2p_z$). It should be noted that for accurate computations, a larger basis set optimized on the atom does not necessarily guarantee a better S.C.F. energy when used in a molecule, than a somewhat smaller basis set optimized on the atom.

In Table 2 the computed binding energy for H_2O is reported (in a.u.).

Table 2

COMPUTED BINDING ENERGY FOR H_2O (IN a.u.)*

H / O	2	3	4	5	6	7
3/2	.2037	.2054	.2052			
5/2	.1992	.2010	.2007			
6/2	.1832	.1811	.1807			
5/3	.2225	.2238	.2227			
6/3	.1941	.1930	.1920			
7/3	.1940	.1925	.1917			
8/3			.1915	.1943	.1957	.1960
7/4			.2063	.2090	.2098	.2096
8/4			.2068	.2094	.2102	.2100
9/4			.2066	.2091	.2100	.2098
9/5			.2120	.2141	.2148	.2145
10/5			.2126	.2146	.2152	.2150
10/6			.2131	.2150	.2156	.2153
11/6			.2133	.2151	.2156	.2154
12/6			.2133	.2151	.2156	.2153
11/7			.2138	.2157	.2162	.2159
12/7				.2157	.2161	.2158
13/7				.2156	.2161	.2158
13/8				.2162	.2166	.2164

*These values are obtained by subtracting from the total energy given in Table 10, the atomic oxygen total energy (^3P) and twice the hydrogen total energy (^2S) as given in reference (Sc).

The reported values are the difference between the Total Energy of H_2O (Table 1) and the sum of the atomic energies of the hydrogen and oxygen atom, computed with the same basis as in the molecule. The resulting binding energy is nearly constant for all basis sets. Thus from this example, it follows that the computed binding energy is nearly constant for very different basis sets as long as (1) the binding energy is defined as above, (2) no polarization functions are introduced, and (3) an equally balanced set of functions for all atoms in the molecules is selected. Equally balanced set means: poor atomic set for all atoms, medium set for all atoms, good atomic set for all the atoms in the molecule.

Table 3

POPULATION ANALYSIS AND DIPOLE MOMENT (IN a.u.) FOR H_2O WITHOUT POLARIZATION

Case	H charge	O charge	Dipole moment
0:4s,2p;H:4s	0.7824	8.4352	0.7965
0:5s,2p;H:4s	0.7791	8.4411	0.7991
0:6s,2p;H:4s	0.7537	8.4925	0.7833
0:5s,3p;H:4s	0.6906	8.6188	0.9757
0:6s,3p;H:4s	0.6446	8.7107	0.9628
0:7s,3p;H:4s	0.6427	8.7145	0.9623
0:8s,3p;H:4s	0.6426	8.7147	0.9617
0:7s,4p;H:4s	0.6275	8.7450	1.0123
0:8s,4p;H:4s	0.6271	8.7458	1.0138
0:9s,4p;H:4s	0.6275	8.7450	1.0129
0:9s,5p;H:4s	0.6251	8.7498	1.0408
0:10s,5p;H:4s	0.6339	8.7221	1.0399
0:10s,6p;H:4s	0.6293	8.7491	1.0526
0:11s,6p;H:4s	0.6351	8.7298	1.0531
0:12s,6p;H:4s	0.6377	8.7246	1.0532
0:11s,7p;H:4s	0.6303	8.7394	1.0590
0:12s,7p;H:4s	0.6320	8.7358	1.0582
0:12s,7p;H:5s	0.6259	8.7522	1.0609
0:12s,7p;H:6s	0.6280	8.7610	1.0611
0:12s,7p;H:7s	0.6185	8.7629	1.0618

In Table 3 we continue our analysis for the wave functions reported in Table 1. This time the gross atomic charges (from the electronic analysis) and the dipole moment (in a.u.) are reported for a subset of the cases given in Table 1. One can notice that with exception of the cases where only two 2p Gaussians were used for the oxygen atom, the dipole moment oscillates between 0.96 a.u. and 1.06 a.u. This variation of 10% is systematic and with an increased set, one quickly approaches the limit value of 1.06 a.u. The gross charges converge to a value of 0.63 on the hydrogen atom and 8.74 on the oxygen atom. Again, the convergency is regular and fast.
Thus, in selecting a basis set, if one whishes to predict dipole moments and charges, a rather large set must be used in order to converge to a limit.

We now introduce polarization functions. A $2p_x$, $2p_y$, $2p_z$ basis with orbital exponent 0.75 was added to the hydrogen atoms, and a $3d_{xx}$, $3d_{yx}$, $3d_{yy}$, $3d_{zx}$, $3d_{zy}$, $3d_{zz}$ was added to the oxygen atom (with orbital exponent 1.0).

Table 4

EFFECT OF POLARIZATION FUNCTIONS ON H_2O TOTAL ENERGY.*

Case	sp basis	sp + polarization	Difference
0:4s,2p;H:4s	-75.17800	-75.29351	.11461
0:5s,2p;H:4s	-75.45778	-75.54615	.08837
0:6s,2p;H:4s	-75.52585	-75.57691	.05106
0:5s,3p;H:4s	-75.77419	-75.85941	.08522
0:6s,3p;H:4s	-75.83298	-75.88177	.01879
0:7s,3p;H:4s	-75.89138	-75.93602	.04464
0:8s,3p;H:4s	-75.90618	-75.91976	.01358
0:7s,4p;H:4s	-75.97129	-76.01274	.04145
0:8s,4p;H:4s	-75.98710	-76.02657	.03947
0:9s,4p;H:4s	-75.99126	-76.03032	.03906
0:9s,5p;H:4s	-76.01153	-76.04867	.03713
0:10s,5p;H:4s	-76.01419	-76.05047	.03628
0:10s,6p;H:4s	-76.01824	-76.05498	.03674
0:11s,6p;H:4s	-76.01946	-76.05597	.03651
0:12s,6p;H:4s	-76.01982	-76.05630	.03648
0:11s,7p;H:4s	-76.02088	-76.05778	.03690
0:12s,7p;H:4s	-76.02120	-76.05811	.03691
0:12s,7p;H:5s	-76.02412	-76.05952	.03540
0:12s,7p;H:6s	-76.02486	-76.05973	.03487
0:12s,7p;H:7s	-76.02485	-76.05985	.03580

*The added polarization functions are $2p_x$, $2p_y$, $2p_z$ on the hydrogen atoms (with exponent 0.75) and $3d_{xx}$, $3d_{yx}$, $3d_{xy}$, $3d_{zx}$, $3d_{zy}$, $3d_{zz}$ on the oxygen (with exponent 1.0).

In Table 4, the gain in total energy is reported.

In Table 5, we report gross atomic charges and the dipole moment for the new basis set (with polarization functions).

Table 5

POPULATION ANALYSIS AND DIPOLE MOMENT FOR H_2O (BASIS SET WITH POLARIZATION)

Case	H charge	O charge	Dipole moment (a.u.)
0:4s,2p;H:4s	0.8154	8.3692	0.5688
0:5s,2p;H:4s	0.8160	8.3681	0.5636
0:6s,2p;H:4s	0.8061	8.3878	0.5963
0:5s,3p;H:4s	0.8153	8.3695	0.7615
0:6s,3p;H:4s	0.8065	8.3870	0.7691
0:7s,3p;H:4s	0.8025	8.3950	0.7716
0:8s,3p;H:4s	0.8026	8.3948	0.7714
0:7s,4p;H:4s	0.7480	8.5021	0.8286
0:8s,4p;H:4s	0.7434	8.5132	0.8310
0:9s,4p;H:4s	0.7442	8.5116	0.8301
0:9s,5p;H:4s	0.7098	8.5804	0.8570
0:10s,5p;H:4s	0.7046	8.5909	0.8561
0:10s,6p;H:4s	0.7080	8.5840	0.8662
0:11s,6p;H:4s	0.7037	8.5926	0.8664
0:12s,6p;H:4s	0.7073	8.5854	0.8706
0:11s,7p;H:4s	0.7038	8.5923	0.8664
0:12s,7p;H:4s	0.7067	8.5866	0.8707
0:12s,7p;H:5s	0.7082	8.5836	0.8694
0:12s,7p;H:6s	0.7093	8.5814	0.8693
0:12s,7p;H:7s	0.7066	8.5868	0.8694

From Table 4 we learn that the introduction of polarization functions brings about a gain in the binding energy of about 0.035 a.u. (or nearly 20% of the binding computed in Table 2). One can notice that the difference in total energy between the set with and without polarization decreases from 0.114 a.u. to 0.035 a.u. This is a clear indication that the polarization functions added to a small basis set, even if balanced, operate in two complementary ways. On one hand, the polarization functions introduce polarization effects. On the other hand, the polarization functions supplement part of the s and p basis set which is lacking in the small set. (It is noted that a combination of functions like

$$(3d_{xx} + 3d_{yy} + 3d_{zz})$$

are equivalent to an s function). From Table 4 one can easily conclude that molecular computations at the Hartree-Fock level with basis sets which include polarization functions but have insufficient atomic sets are not too meaningful if used to predict the Hartree-Fock binding energy. The dipole moment and atomic gross charges (Table 5) present a rather systematic trend. If one wishes to compute a dipole moment within 0.02 a.u. (or 0.05 Debye units) from the Hartree-Fock limit, then one should use at least a 9s, 5p set for oxygen, and 4s for hydrogen plus polarization functions. The gross charges do not fluctuate as long as one uses a basis set both balanced and larger or equal to a 9s, 5p on oxyge, and 4s on hydrogen. It is noted that the Hartree-Fock model predicts dipole moments accurate within 0.25 or 0.4 debyes. Computations of dipole moments reaching an agreement with experimental data better than the above limit are often either fortuitous or are artifacts obtained by conveniently varying the orbital exponent of the polarization functions, so as to obtain agreement with experimental data.

Table 6

TOTAL ENERGY FOR H_2O WITH POLARIZATION*

Polarization Function Exponents Od	Of	Hp	Hd	Energy (a.u.)	Energy Gain (a.u.)
				-76.02563	0.0
1.00		0.75		-76.06027	0.03464
0.40		0.75		-76.06186	0.03623
2.00					
0.40		0.30		-76.06400	0.03837
2.00		1.50			
0.40		0.30		-76.06454	0.03891
1.00		1.50			
2.00					
0.40	1.00	0.30		-76.06503	0.03940
2.00		1.50			
0.40	1.00	0.30	0.75	-76.06587	0.04024
1.00		1.50			
2.00					

In Table 6 H_2O energies (in a.u.) computed with large basis sets are

reported. We started with an (O 13s, 8p; H 6s) basis set and found that contracting the six O s functions, the four O p functions and the three H s functions with the largest exponents resulted in an energy loss of only 0.00010 a.u. compared with the energy of -76.02573 a.u. in Table 1 for the completely uncontracted basis set. Consequently we used a (O 8s, 5p; H 4s) contracted basis set and augmented it with the polarization functions given in Table 6. The best energy of -76.06587 a.u. was obtained by using three sets of d functions and a set of f functions on the oxygen and two sets of p functions and a set of d functions on the hydrogens. A consideration of the energies reported in Tables 4 and 6 leads us to believe that the above energy is about 0.002 a.u. from the Hartree-Fock limit for H_2O with the experimental equilibrium geometry.

Part 2 Structure of Liquid Water as a Test Case

2.1 Introduction - In the previous chapter we have analyzed how a wave-function can be used to describe either a single molecule of water or a molecule of water interacting with ions or with a second molecule of water. In particular, we have stressed the chemical and physical interpretation of the quantities such as electronic density, total energy and orbital energies that constitute the traditional output of a quantum mechanical computation. In this section we shall be confronted with another problem: how to describe a system of many molecules such as a liquid. Clearly, if we fail to describe a liquid as water, there are reasonable doubts on the possibility to describe the structure of an ion or of a molecule when surrounded by the many molecules of water considered as a solvent. Therefore, our aim in this chapter is to produce the structural information (X-rays and neutron diffraction) today available for liquid water, considered as a solution composed of one molecule of water surrounded by many molecules of water. In so doing, we shall find the need of another quantity, previously not even mentioned, namely the temperature and (as its consequence) several aspects of statistical thermo-dynamics. Therefore we shall have a new "many body systems" with reference no longer to electrons and nuclei but molecules and ions, and accordingly we shall pass from the Fermi statistics to the Boltzmann statistics.

2.2 Approximate Hartree-Fock Potential for the Water Dimer - In this section we report a study of the Hartree-Fock potential for the water dimer[1] obtained with the restriction that the two monomers are kept rigid (each with the experimental H_2O geometry; the calculations were carried out using the H_2O basis set given in reference[2]).

We note that the hydrogen bond strength in the water dimer is about 5 kcal/mole (0.008 a.u.). Thus particular care should be used in selecting the basis set. An insufficiently small basis set will yield different accuracy for different water-water separations. At shorter distances the basis set of one molecule will tend to compensate for the deficiencies in the basis set of the other molecule, thereby yielding a lower energy and a larger biding energy[3]. To ensure that our results are of sufficient accuracy, we have carried out a calculation near the equilibrium configuration using a very large basis set that yields a total energy close to the Hartree-Fock limit.

Calculations on the water dimer have been reported[3] for 190 different geometries. As noted previously, the geometry of each water molecule in the dimer is kept rigid and the first water molecule is placed in a fixed position. It lies in the xy-plane with its symmetry axis along the x-axis. The second water molecule is moved around the first one.

In the vicinity of the energy minimum an additional 26 points on the energy surface have been calculated. We have attempted to find an analytical expression for the energy surface by fitting the 216 calculated points. As a starting point we tried several models proposed for the H_2O-H_2O potentials, [4,5,6]. The analytical formula that best combines numerical accuracy with mathematical simplicity can be considered as resulting from an H_2O charge distribution similar to that proposed by Bernal and Fowler[4]. We wish to stress that the point charge model is only a convenient way to simplify the choice of an analytical expression for the potential. We attach no significance to the values of the numerical parameters of the point charge model given in Fig. 1. The parameters obtained are the ones that best fit our numerical Hartree-Fock data.

The analytical potential that we obtained is given by the expression (in a.u.):

$$
\begin{aligned}
E = \; & q^2 (1/r_{13} + 1/r_{14} + 1/r_{23} + 1/r_{24}) \\
& + 4q^2/r_{78} \\
& - 2q^2 (1/r_{18} + 1/r_{28} + 1/r_{37} + 1/r_{47}) \\
& + a_1 \exp(-b_1 r_{56}) \\
& + a_2 [\exp(-b_2 r_{13}) + \exp(-b_2 r_{14}) + \exp(-b_2 r_{23}) + \exp(-b_2 r_{24})] \\
& + a_3 [\exp(-b_3 r_{16}) + \exp(-b_3 r_{26}) + \exp(-b_3 r_{35}) + \exp(-b_3 r_{45})]
\end{aligned}
$$

where: $a_1 = 582.277054$; $a_2 = 0.143789$; $a_3 = 5.470184$; $b_1 = 2.520593$; $b_2 = 1.221756$; $b_3 = 1.936626$; $q^2 = 0.449387$ and the O_5–M_7 (or O_6–M_8) distance is 0.436 a.u. The labelling of the water dimer nuclei and M points is given in Fig. 1.

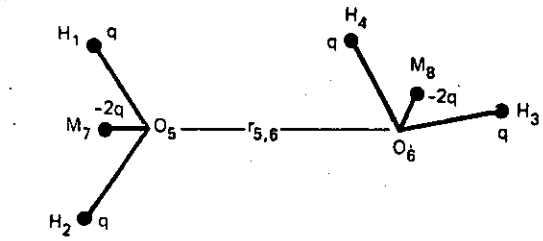


Figure 1 Point charge model used in the derivation of the analytical water-water interaction potential.

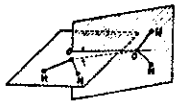
The accuracy of this fit is indicated by the following values of the standard deviation. The standard deviation is 0.0002 a.u. for the 80 points with negative E (i.e., that are binding), 0.0004 a.u. for the 165 points with E less than +5 kcal/mole, 0.0013 a.u. for the 208 points with E less than 1 eV, and 0.0018 a.u. for all 216 points.

Thus our fit is very good for attractive regions of the potential, but is somewhat less accurate for the repulsive regions. Of course this is what we want, since repulsive configurations are not important (because of the Boltzmann factor) in the study of systems at room temperature.

The Hartree-Fock potential can be compared with the empirical potentials proposed by Rowlinson[5] and Ben-Naim and Stillinger[6]. We note that these authors were mainly interested in deriving an effective pair potential for the condensed phase rather than a potential for the water dimer.

It is of particular interest to present the energy surfaces where the constraint on the orientation of the symmetry axis of molecule 2 is relaxed. For a given oxygen-oxygen distance, the symmetry axis of molecule 2 is positioned so that the dimer energy is at its minimum value. The resulting surfaces are given in Fig. 2, 3 and 4. All three potential show nearly linear hydrogen bonds for the most stable configurations. The oxygen-oxygen equilibrium separation decreases on going from Fig. 2 to Fig. 4. By counting contours one determines that the Hartree-Fock minimum is not as deep as that for the empirical potentials. In addition, for the empirical potential the hydrogen bonding is more localized near the potential minimum. The linear hydrogen bond orientations are much more favoured than the orientations where the symmetry axes of the two water molecules are parallel.

The optimum angle between the line joining the two oxygen nuclei and the symmetry axis of molecule 2 is, shown in the right half of Figs. 2-4. The qualitative behaviour is similar for all three potentials. For the Hartree-Fock potential, in a given region of space there is hardly any reorientation of the symmetry axis of molecule 2, i.e., the value of α is approximately zero. This is less true for the Rowlinson potential and even



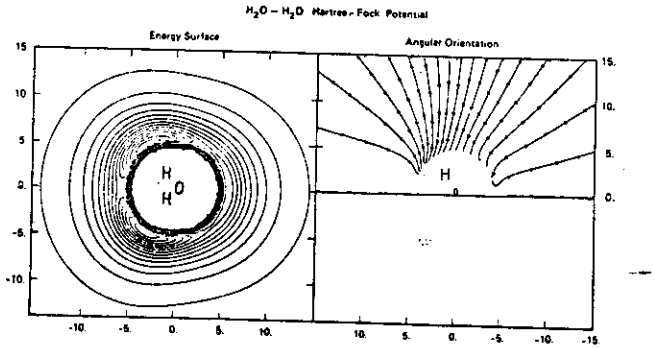


Figure 2 Hartree-Fock potential surface for the illustrated configuration type as a function of the oxygen-oxygen distance and angle α between the symmetry axes of molecule 2 and the line joining the two oxygen nuclei. For the plot on the left, the contour interval is 0.0005 a.u. The oxygen atom is at the origin and the distance scale is in a.u. For the plot on the right, the contour interval is 10°. The horizontal line to the left of the origin is the 0° contour. Only the upper part of the symmetric plot is given.

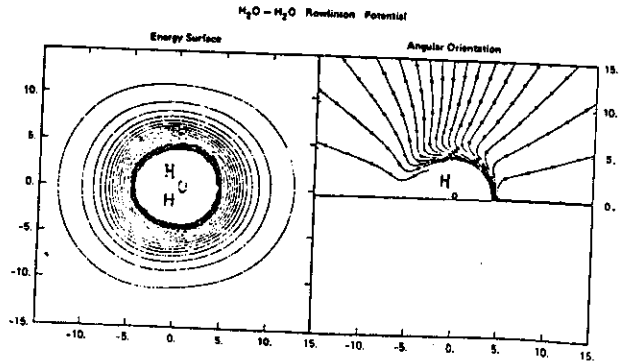


Figure 3 Rowlinson potential surface for the illustrated configuration type

TABLE 1 Water Dimer Binding Energy Computed by Various Authors.

Authors	Uncontracted[a] Basis Set	Contracted[a] Basis Set	H_2O Energy (a.u.)	$(H_2O)_2$ Energy (a.u.)	$(H_2O)_2$ Binding Energy (Kcal/mole)
Morokuma & Pedersen (1968) (α = 0°)	(5,3/3)	(5,3/3)	-75.54939	-151.11887	12.6
Kollman & Allen (1969) (α = 25°)	(10,5/5)	(3,1/1)	-75.97468	-151.95768	5.25
Del Bene & Pople (1970) (α = 58°)	(8,4/4)	(2,1/1)	-75.50013	-151.00998	6.09
Hankins, Moskowitz & Stillinger (1970) (α = 40°)	(10,5,1/4,1)	(5,3,1/2,1)	-76.04158	-152.09069	4.72
Diercksen (1971) (α = 0°)	(11,7,1/6,1)	(5,4,1/3,1)	-76.05326	-152.11167	4.83
Present work (α = 0°)	(11,7,1/6,1)	(4,3,1/2,1)	-76.05525	-152.11747	4.37
Present work (α = 30°)[b]	(11,7,1/6,1)	(4,3,1/2,1)	-76.05525	-152.11784	4.60
(α = 0°)	(11,7,2/6,2)	(4,3,2/2,2)	-76.05990	-152.12561	3.65
Present work (α = 0°)	(13,8,1,1/6,1,1)	(8,5,1,1/4,1,1)	-76.06255	-152.13206	4.37
(α = 30°)[b]	(13,8,1,1/6,1,1)	(8,5,1,1/4,1,1)	-76.06255	-152.13244	4.60
Present work (α = 0°)[b] (α = 30°)[c]	(13,8,2,1/6,2,1)	(8,5,2,1/4,2,1)	-76.06598	-152.13781	3.67 3.90

(a) The notation (13,8,2,1/6,2,1) indicates that the O atom basis set consists of 13 s-, 8 p-, 2 d- and 1 f-type Gaussian and the H atom basis set consists of 6 s-, 2 p- and 1 d-type Gaussians.

(b) For the definition of α see Table I?

(c) Extrapolated value.

less for the Ben-Naim-Stillinger potentials. This results from the pronounced tendency of the empirical potentials towards nearly tetrahedral orientation of neighboring molecules.

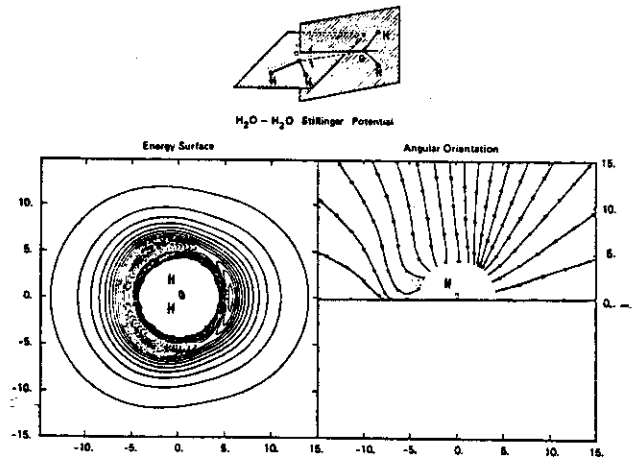


Figure 4 Ben-Naim and Stillinger potential surface for the illustrated configuration type

In Table 1 we compare a number of computations in the SCF-LCAO-MO approximation to determine the binding energy of the water dimer; those of Diercksen et al. and Hankins et al. correspond to references 7 and 8 respectively.

2.3 Polymers of Water

The more a potential for two molecules is spherical, likely the less meaningful, in a thermodynamical sense, is it to discuss the exact geometry since we are in a situation described by polytopic-type bonds and geometries[9]. Therefore, in the study of clusters of water we expect the existence of many configurations with nearly the same energy, not only at finite temperatures, but also near zero temperature. As a consequence, in the study of the structure of water complexes, the emphasis should be on the probability distribution of the various configurations. We wish to point out that this situation occurring for the water complexes is not unique: since it has long been recognized that in the study of the structure of polymeric chains, the chemically significant problem is not the exact determination of the energetically lowest possible conformation, but rather the study of the probability distribution of all those conformations that are within an energy range statistically compatible with the temperature of the system in consideration[10]. We further note that the energy barriers in polymeric chains, the main factor controlling the distribution of conformations, are in the same energy range as the water-water binding energies (namely of the order of 3

to 6 kcal/mole).

The basic technique found useful in the conformational study of polymers can be adapted in the study of the conformations of the water complexes (or of ion-water complexes), namely an analytical potential of relatively simple form is constructed and iso-energy contour maps are obtained to aid in the selection of the most probable conformations.

Previous theoretical studies on the structure of water complexes[11,12] have, however, followed the traditional approach commonly used for the quantum chemical characterization of small molecules in the gaseous phase, where one attempts to determine the energetically most stable single geometry for the molecule. We are of the opinion that this commonly adopted approach might not be sufficiently meaningful since: (a) a number of competing structures will be neglected and (b) likely it will be practically impossible to obtain the lowest configuration since there are too many degrees of freedom to analyze, and presently a full search of the potential surface is too expensive for direct quantum mechanical computations.

Therefore, in this section we present a study of the water-water complexes[13] where use is made of a simple analytical potential obtained from direct computations in the Hartree-Fock approximation; with such a potential, assuming pair-wise additivity, a full search of the potential surface for the $(H_2O)_n$ complexes is performed. For the complexes $(H_2O)_3$ and $(H_2O)_4$, direct Hartree-Fock computations are presented to check the validity of the pair-wise addivity assumption. Previously we have reported the iso-energy contour maps for the Hartree-Fock interaction of two molecules of water, referred below as molecule A and molecule B, subjected to two geometrical constraints.

First, the geometry of each molecule was kept rigid (and chosen to be the known experimental geometry at equilibrium internuclear distances).

Second, both molecular planes were constrained to be perpendicular to each other, with the oxygen of the molecule B lying in the plane of the molecule A. This contour map can be constructed using the available analytical fit to the Hartree-Fock sampling of the interaction potential for A and B, and considering three variable geometrical parameters. If we fix rigidly the molecule A so that the oxygen atom is at the cartesian origin and the O-H bonds of A are in the x-y plane, then the variables are the x and y coordinates of the oxygen for B, and (for each pair of value of the variables) the rotational angle around an axis perpendicular to the molecular plane of A and passing through the oxygen nucleus of B. The number of configurations to be considered is high, in the order of several thousands, even with the assumptions implied in the above mentioned constraints.

We shall now relax some of the above constraints. For each position of the molecule B, its hydrogen is free to be either in or out of the plane. In other words, the rotational axis, passing through the oxygen nucleus of the molecule B, is not constrained to be perpendicular to the molecular plane of A, but can assume any direction. As known, any rotation can be obtained by performing three successive rotations with pre-established order and sense (for example, one can use Eulerian angles). The sampling of the interaction potential needed to obtain iso-energetic maps is now sharply increased, since there are no longer three geometrical variables, but five (the x and y coordinates of the oxygen atom of the molecule B and the three rotational angles).

This process can be repeated with the oxygen atom of the molecule B no longer constrained to be in the molecular plane defined by the molecule A, but free to be in any plane parallel to the one defined by the molecule A. This way only one constraint remains: the molecule is not allowed to vibrate.It is not a trivial task to find a practical and satisfactory solution to the problem of displaying the enormous number of configurations analyzed. In Fig. 5 we present a graphical solution that takes advantage of two computer programs kindly made available to us by D.A. Schreiber[14] and by C.K. Johnson[15]. For each oxygen-oxygen distance $R_{O\text{-}O}$, only the configuration corresponding to the lowest energy for the three angular rotations is reported. A master coordinate axis x, y, z is located at the oxygen nucleus of the fixed

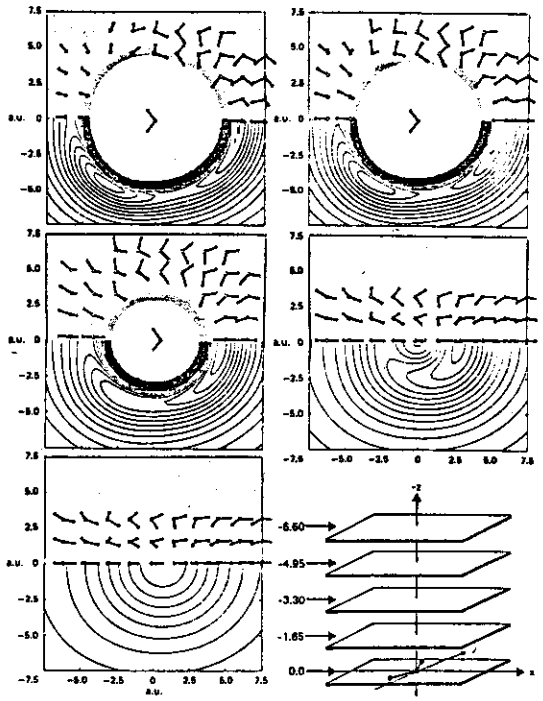


Figure 6 Minimal energy contour maps for the interaction of two molecules of water. One molecule is kept fixed at the origin of the coordinate system (see bottom right insert). The second molecule has its oxygen nucleus constrained parallel to the xy plane at z = 0.0 a.u. (top left insert), at z = 1.65 a.u. (top right insert), at z = 3.3 a.u. (middle left insert), at z = 4.95 a.u. (middle right insert), at z = 6.6 a.u. (bottom left insert). The contour interval is 0.0005 a.u. The lowest contour (minimum) is at -4.4 kcal/mole with the exception of the bottom left insert, where the lowest contour is at -2.8 kcal/mole. The orientations of the second molecule are energy optimized and some of these orientations are given in a perspective view with the O-H distances being given only in half of their actual length and the viewer being perpendicular above the planes.

molecule, A, lying in the xy-plane, the x axis being, for example, oriented as to coincide with the C_2 rotational axis of molecule A (a water molecule has C_{2v} symmetry at the equilibrium configuration of the ground state). The oxygen atom of the molecule B is constrained to be either in the plane of xy (z=0.0) or in a number of planes parallel to the xy-plane. The planes we have considered are defined by z=0.00 a.u., -1.65 a.u., -3.30 a.u., -4.95 a.u., -6.6 a.u. The iso-energy contours in each plane have intervals of 0.0005 a.u. (about 0.3 kcal/mole). A computer program computes for a grid of points in each plane the energy of the A-B system, using the analytical fit to the Hartree-Fock potential, optimizes the molecular orientation based on a minimum energy criterion, and then transmits these data to the contour program[14]. The angular orientation of the hydrogen atoms (or of the O-H bonds) for the molecule B, corresponding to the lowest energy for the chosen $R_{O\text{-}O}$ distance, is graphically given in Fig. 5 by presenting a three dimensional schematic representation of the B molecule. The O-H bond is not given in scale, but the O-H distance is represented by exactly one half of the correct one in order to display the data in a comprehensible way.

The display of the B molecules is restricted to a sufficient number so as to convey a visual "feeling" of the way the molecule B orients itself in space when it moves from point to point in a given plane. The main information we wish to convey is of qualitative nature (the quantitative information is in the analytical expression for the Hartree-Fock potential). From the information displayed in Fig. 5, a cross section of the energy surface, one can obtain a good representation of the surface. In the following we shall briefly comment on the cross section presented in this section. First, we have a repulsive region and an attractive region. The repulsive region can be identified very simply in the -0.0 a.u., in the -1.65 a.u. and in the -3.3 a.u. planes, since here the repulsion is very rapidly increasing when one brings the molecule B close to A: the contours are very near to one another, forming a ring, whose last contour represents a cut of 3.4 kcal/mole. We have not displayed more repulsive contours, since they are not too important for the study of systems near room temperature.

The attractive regions of the potentials are more important: the three minima of the 0.0 a.u. plane, corresponding to the open for of the dimer with nearly linear hydrogen bonds, are still present in the -1.65 a.u. plane, but not all of them appear in the -3.3 a.u. plane. The remaining minimum moves towards the x=y=0 position in the -4.95 a.u. plane and is nearly at this position in the -6.6 a.u. plane. It is noted that all the contours of the -6.6 a.u. plane are attractive. These are the main energetic features emerging from the Hartree-Fock potential. Fig. 5 contains a number of important details that will be discussed later. In Fig. 5 the energy of the lowest contours is -4.4 kcal/mole (with the exception of the lowest contour in the -6.6 a.u. plane, where the energy is -2.8 kcal/mole).

The orientation of the water B, relative to the fixed molecule of water A, is now briefly considered. Let us fix our attention on a molecule B on the C_{2v} axis, to the left of the molecule A, with the oxygen at about x=5.5 a.u., y=0.0 a.u. In the 0.0 a.u. plane, we can see only one O-H bond (looking from above the plane in the -z direction) since the second one is masked and exactly below. If we move this molecule along the -z axis (from the 0.0 a.u. plane to the -6.6 a.u. plane) and let the hydrogen find the energetically optimal configuration relative to the molecule A, we can learn from Fig. 5 that the molecule slowly rotates around an axis passing through the oxygen and parallel to the y axis; in the plane 0.0 a.u. the hydrogen points away from the A molecule, in the plane -6.6 a.u. the hydrogen points to the 0.0 a.u. plane, and one to the oxygen of the A molecule (the molecule A is not shown in the -4.95 a.u. and - 6.6 a.u. planes).

Let us now consider in more detail the -3.30 a.u. plane. In Fig. 5 there is only one minimum; but one can see a rather extended region enclosed between the third and fourth lowest contour. In this flat region there is a very shallow minimum that can be obtained, for example, by selecting a smaller contour interval than the one of 0.0005 a.u. used for Fig. 5. To this shallow minimum corresponds a second form for a water dimer, designated as "closed form", containing two hydrogen bonds[13'16]. The closed form of the dimer is of interest as the simplest prototype of cyclic polymers of water (closed rings), whereas the open form is the prototype of the open chain polymers of water.

Finally it is noted that previous studies on the dimer like those of Marokuma and Pederson[3], Kollman and Allen[17], Del Bene and Pople[3] (all done in the SCF-LCAO-MO approximation) are not sufficiently accurate so as to be of little help in this problem. Other computations by Hankins, Moskowitz and Stillinger[8], by Diercksen[7], Lentz and Scheraga[12], have not scanned the potential space sufficiently as to be in a position to study the closed dimer structure.

Hankins, Moskowitz and Stillinger[11] have reported a Hartree-Fock study (with a sufficiently large basis set as to be near the Hartree-Fock limit) on the water trimer, and have concluded that triplets of water molecules with sequential hydrogen bonds stabilize significantly the condensed phase which incorporate such structures. A total of eight configurations have been analyzed by these authors. Since these authors have considered structures related to the cubic and the hexagonal forms of ice, from this work it is not easy to extract a general conclusion for the stable form of the water trimer.

Lentz and Scheraga[12] have recently examined the structure of the water trimer and the water tetramer with the aim to compare the binding energy of the cyclic and open form of these complexes. Their conclusion is that the cyclic trimer is less stable than the non-cyclic one, in contradiction with the results obtained by Del Bene and Pople[19]. It is noted that Lentz and Scheraga's computations have been performed in the Hartree-Fock model with adequately large basis sets, whereas Del Bene and Pople's basis set is sufficiently truncated as to give up to 2 kcal/mole error in the stabilization energy of small complexes, like dimers[3].

As remarked in the introduction to this chapter, and as evidenced by the two works summarized above, a full optimization of the geometry for a system containing several energy minima is a complex task. Clementi et al.[13] have chosen to use the analytical fit to the Hartree-Fock sampling (a-f-H-F) for the $(H_2O)_2$ complex to determine the optimal configuration on an energy criterium for the $(H_2O)_n$ (n=3,...,8) complexes. This might constitute a severe assumption, which however has been tested (see below) by considering the magnitude of the three and four body effects obtained by computations on the $(H_2O)_3$ complex and computations on the $(H_2O_4$ complex. For the accuracy required in an approximative structure determination of the $(H_2O)_n$ complexes, the previous additivity assumption appears satisfactory as shown below.

For the $(H_2O)_3$ complex, the best configurations optimization yielded a geometry given in Fig. 6. Thousands of possible configurations have been generated with the a-f-HF, translating and rotating the three molecules in many possible ways. The resolution of this search depends on the magnitudes of the steps between two successive geometries. The optimization method used a modified version of a Simplex computer program kindly supplied by J.P. Chandler[18]. Examining Fig. 6, we learn that the most stable geometry obtained has nearly C_s symmetry, a result in agreement with Del Bene and Pople[19]; however, the dimension of the ring is significantly larger: our values for the oxygen-oxygen distances (see Fig.6 for notation) are $R(O_1,O_2)$ =2.96A, $R(O_2,O_3)$=2.96A, $R(O_1,O_3)$ =2.97A to be compared with Del Bene and Pople's distance of $R(O_1\text{-}O_2)$=$R(O_2\text{-}O_3)$ =$R(O_1\text{-}O_3)$=2.56A[18].

The hydrogen bond angles in our computation are bent outwards by about 22°. The energy of stabilization for this structure obtained with the a-f-HF is 0.0196 a.u. (12.3 kcal/mole).

Using the same technique previously described for obtaining the best (most stable) configuration of the water trimer, the water tetramer was found to have the structure given in Fig. 6. The constraint imposed on the tetramer was to leave all oxygen atoms lying on the perimeter of a circle.

The geometrical characterization of this tetramer is as follows: O(1)–O(2) =2.95A, O(1)–O(3) = 4.17A, O(1)–O(4) = 2.95A, O(2)–O(3) = 2.95A, O(2)–O(4) = 4.15A, O(3)–O(4) = 2.95A; the angular characterization is given in reference[13].

As an introductory analysis to larger complexes of water, we have analyzed[13] a very restricted class of polymers of water, which have been discussed amply in literature but not studied in a quantitative way. We feel that the use of the addivity approximation will yield reliable indication on the best structure and

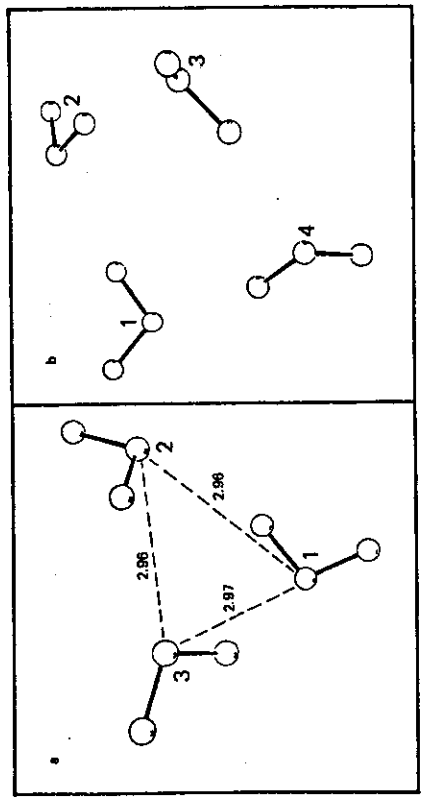


Figure 6 Most stable configurations for the trimer (insert a)) and the tetramer (insert b)) of water at T = 0°K (see text for additional explanations about the geometry).

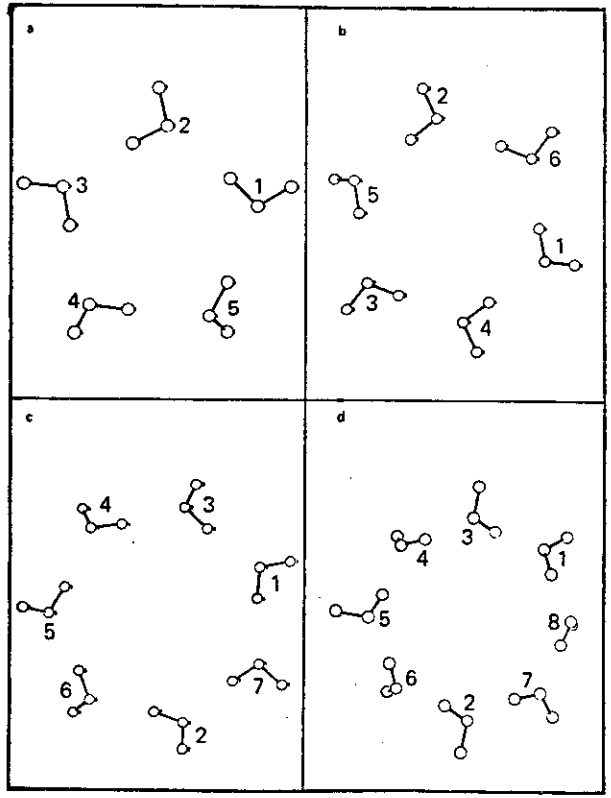


Figure 7. Most stable configurations for $(H_2O)_5$, $(H_2O)_6$, $(H_2O)_7$ and $(H_2O)_8$ obtained under the constraint that the oxygen nuclei have to lie on the periphery of a ring. The hydrogens are free to assume any orientation.

stabilization energy. With the constraint that the oxygen atoms should be on the periphery of a circle, that means we require all oxygens to be in the same plane, and with no constraint for the hydrogen orientation, we have searched for the configuration of minimum energy for $(H_2O)_5$, $(H_2O)_6$, $(H_2O)_7$, and $(H_2O)_8$. The resulting structures are displayed in Fig. 7. For all these rings one hydrogen for a given molecule is hydrogen-bonded and the second one tends to minimize the hydrogen-hydrogen repulsion. We shall not discuss at length these structures, but only some that characterize the result given in Fig. 7.

For the pentamer, $(H_2O)_5$ the geometrical characterization is: O(1)–O(2) = 2.97A, O(1)–O(3) = 4.80A, O(1)–O(4) = 4.79A, O(1)–O(5) = 2.96A, O(2)–O(3) = 2.97A, O(2)–O(3) = 4.80A, O(3)–O(4) = 4.80A, O(3)–O(4) = 2.97A, O(3)–O(5) = 4.80A, O(4)–O(5) = 2.95A.

The hydrogens that are hydrogen bonded are characterized by the angles: $\alpha = 4° \pm 2°$, $\beta = 3° \pm 1°$, $\gamma = 1° \pm 1°$. One can notice that the hydrogen bonds are somewhat more straight than in the tetramer case: this is expected since there is less strain the larger the ring.

For the hexamers, $(H_2O)_6$ the geometrical characterization is as follows: O(1)–O(2) = 5.13A, O(1)–O(3) = 5.13A, O(1)–O(4) = 2.96A, O(1)–O(5) = 5.92A, O(1)–O(6) = 2.96A, O(2)–O(3) = 5.12A, O(2)–O(4) = 5.92A, O(2)–O(5) = 2.96A, O(2)–O(6) = 2.96A, O(3)–O(4) = 2.96A, O(3)–O(5) = 2.96A, O(3)–O(6) = 5.92A, O(4)–O(5) = 5.12A, O(4)–O(6) = 5.13A, O(5)–O(6) = 5.12A. The value of the angle is $3 \pm 1°$ with exception of α ($H_5O_5O_3$) = 7°; β is equal to $2° \pm 1°$; and γ is equal to $3° \pm 1°$ with exception of γ ($H_5O_5H_5'$) = 7°.

For the heptamer, $(H_2O)_7$ the following geometrical characterization is given: O(1)–O(2) = O(1)–O(4) = O(2)–O(5) = O(3)–O(5) = O(3)–O(7) = O(4)–O(6) = O(6)–O(7) = 5.32 ± 0.01A; O(1)–O(3) = O(1)–O(7) = O(2)–O(6) = O(2)–O(7) = O(3)–O(4) = O(4)–O(5) = O(5)–O(6) = 2.95 ± 0.01A, and O(1)–O(5) = O(1)–O(6) = O(3)–O(6) = O(4)–O(7) = O(5)–O(7) = 6.65 ± 0.01A. The values of the angles are $\beta = 6° \pm 1°$, $\beta = 5° \pm 1°$ and $\gamma = 2° \pm 1°$ with exception of γ ($H_3O_3H_3$) = 5°.

The octamer, $(H_2O)_8$ has the following geometrical characterization: the nearest neighbor has (O_i)–(O_j) = 2.96, $\alpha = 8° \pm 1°$, $\beta = 8° \pm 2°$ and $\gamma = 3° \pm 1°$. The next nearest neighbor has a (O_i)–(O_j) distance of 5.46A, the next nearest neighbors have (O_i)–(O_j) = 7.13A.

In a large number of calculations we relaxed the symmetry constraints imposed on the positions of the oxygen nuclei. At first we required the oxygen nuclei to lie on the periphery of a sphere, thus admitting non-planar ring structures. The resulting structures differ not too much in energy and geometry from the planar rings, however the larger clusters show puckered non-planar rings. Relaxing all constraints we obtain for $(H_2O)_4$ and $(H_2O)_5$ ring structures, but for the larger clusters we obtain for a given cluster size various irregular structures with multiple branching of H-bonds on one molecule (corresponding to open H-bonds) and considerably different geometries but quite similar energies. For instance, in Fig. 8 three forms for the $(H_2O)_8$ complex are given. The corresponding stabilization energies starting from the right are 47.30 kcal/mole, 46.95 kcal/mole, and 46.76 kcal/mole. This proves the necessity to study the statistical distribution of configurations and thus reduces the importance of individual geometries.

In Fig. 9 we display the energy stabilization and the stabilization energy per molecule as a function of the polymer size n. For the planar rings, the spherical rings, and the irregular structures the stabilization energies are quite similar, the spread is not large, but within this spread a large number of different structures are existing, increasingly larger the larger the polymer size. The binding energy per molecule is increasing with increasing cluster size, this indicates the stability of the larger clusters against dissociation in any combination of smaller ones, at least at T=0°K. The binding energy per molecule is much smaller than the corresponding value of a large system of water molecules (for instance at T=273°K Monte Carlo calculations with the a-f-H-F potential yield a value of 7.1 kcal/mole). The relatively small stabilization energy per molecule and the fact that the open structures are more stable for larger clusters indicates that it does not seem to be possible to describe the energetic properties of water as resulting from an ideal mixture of small regular clusters[19]. This conclusion would not be changed even when we take into account

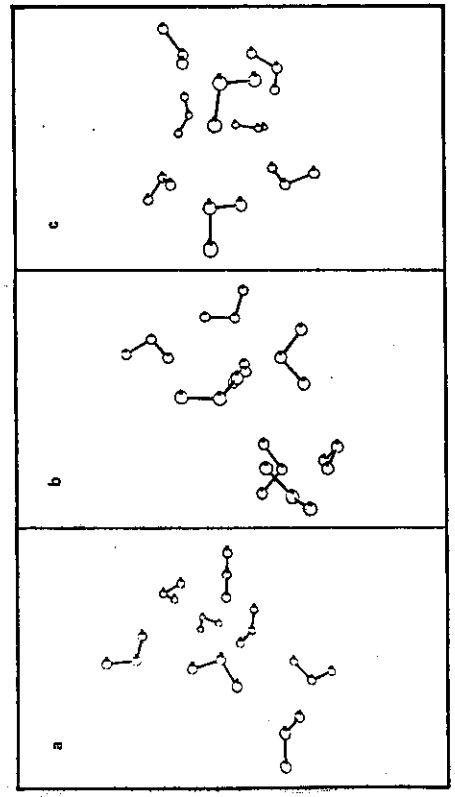


Figure 8 Different irregular structures for the energy optimized $(H_2O)_{12}$ complex, with no constraints imposed. Starting from the left the corresponding stabilization energies are 47.30 kcal/mole, 46.95 kcal/mole and 46.76 kcal/mole.

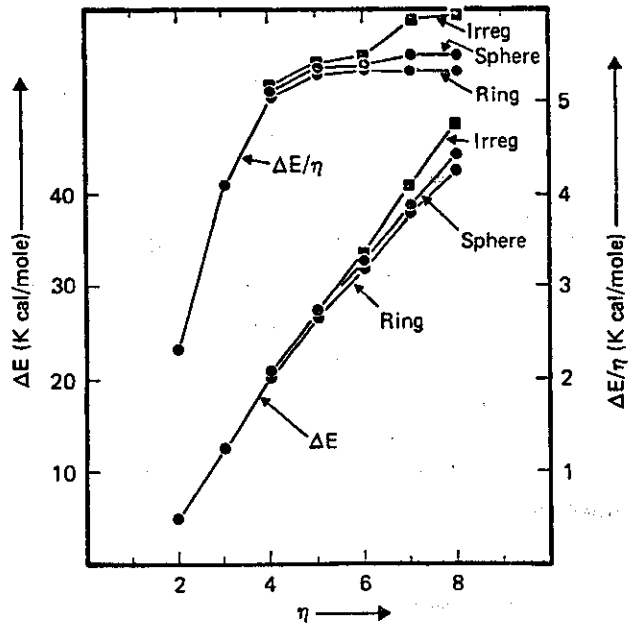


Figure 9 Stabilization energies for the $(H_2O)_n$ clusters as a function of the cluster size n. The stabilization energies have been obtained by imposing three different constraints. In the first case the oxygen nuclei have been required to be on the periphery of a planar ring (designated by "ring" in Figure 10). In the second the oxygen nuclei have been required to be on the periphery of a sphere (designated by "sphere" in Figure 10). In the third there have been no constraints imposed (designated as "irreg" in Figure 10).

three-body effects for the smaller clusters. This is somewhat in contradiction with the results obtained by Pople and Del Bene[19], which would favour a mixture model, with the trimers being the dominant structure[12]. As noted before, Pople and Del Bene's result is probably caused by the use of a too much truncated basis set. As pointed out by Ben Naim[20] any liquid can be formally regarded as a mixture of quasicomponents, for instance clusters, but according to our results this does not seem to be very convenient for liquid water, since the mixture would be strongly non-ideal and one would probably need a large amount of different clusters, differing both in size and geometry. But again, as noted before, one has to be very careful in transferring this conclusion to systems at finite temperature, since there entropy effects may be very important.

We concluded that the stabilization energy per molecule shows no special stability of an individual cluster and is also considerably smaller than for liquid water. This seems to be in contrast with mixture-models of liquid water, describing the properties of water as resulting from an ideal mixture of small regular polymers. Due to many energetically nearly equivalent but geometrically quite different structures for the larger clusters, the exact geometrical structure is less important. The probability distribution of different cluster structures at finite temperatures is likely a more physically meaningful parameter. In addition it has to be noted that entropy effects have an important influence on the relative stability of different clusters.

2.5 The Structure of Liquid Water using an Accurate Potential

Recently, Diercksen, Kraemer, and Ross[27] reported extensive CI calculations for the water dimer. Although their calculations are limited to a relatively small portion of the potential surface in the vicinity of the equilibrium geometry, comparison with their results and with the dispersion correction results calculated by the perturbation technique[28] provides a valuable check on the accuracy of our correlation energy results.

However, a potential sufficiently extended in space, and not restricted around the minimum energy of the

dimer, is now available[29], as below described. The electronic wavefunctions and energies for the ground state of the water dimer have been calculated using the configuration-interaction method. The ground state wavefunction is expanded in a form

$$\Psi = C_{SCF}\,\Phi_{SCF} + \sum_i C_i\,\Phi_i^{INTER} + \sum_i C_i\,\Phi_i^{INTRA} \qquad (7)$$

where C's are variationally determined coefficients, and Φ are orthonormal configuration state functions (CSF). Since we were interested in mapping a potential surface for the water dimer in any possible conformation, we treated the system without reference to point symmetry consideration. Thus, each Φ is linear combination of Slater determinants (SD) constrained only to be a singlet state function. The SD's are built from an orthonormal one-particle basis set of spacial orbitals, which in turn are expanded in terms of a set of elementary basis functions (EBF). In equation (7) the wavefunction is written as a sum of three classes of Φ's. The first class consists of the SCF state function, Φ_{SCF}, which is taken as the reference CSF in this work. The second and third classes consist of Φ^{INTER} and Φ^{INTRA}, which contribute mainly to either the inter- or intra-molecular correlation energy, respectively. Later in this section we shall give precise definitions of these classes of CSF and the details on how these CSF's are constructed.

The elementary basis set we have used consists of contracted Gaussian-type (CGTF) centered at each atom; the type, the orbital exponent and contraction coefficient for each CGFT are listed in reference[29]. This essentially the same as the one used in the previous SCF water dimer potential calculations, and consists of (11,7,1/4,3,1)-functions (which means 11 s-, 7 p-, and 1 d-type functions contracted to 4 s-, 3 p-, and 1 d-type functions) centered on each oxygen atom and (6,1/2,1)-functions on each hydrogen atom.

A SCF calculation with the elementary basis set just described yields 10 canonical SCF occupied orbitals and 48 virtual orbitals. In our CI calculations we used as the core and internal orbitals, a set of orbitals obtained by localizing these canonical SCF orbitals on each monomer following the method developed by Edmiston and Ruedenberg[30]. This transformation was performed in order to ensure a meaningful partitioning of the dimer correlation energy to inter- and intra- parts. In addition, this transformation provides a reasonable way of selecting a limited number of external orbitals to be used in constructing CSF's.

We write the SCF state function as

$$\Phi_{SCF} = 1\phi_A^2\,2\phi_A^2\,3\phi_A^2\,4\phi_A^2\,5\phi_A^2\,6\phi_A^2\,7\phi_A^2\,8\phi_A^2\,9\phi_A^2\,10\phi_A^2 \qquad (8)$$

where the subscripts A and B designate the water monomer on which the orbitals are localized; $1\phi_A$ and $2\phi_B$ are essentially the 1s oxygen orbitals that we shall call the core orbitals. The remaining 8 localized orbitals will be designated as the internal orbitals.

With this order of the core and internal orbitals, we define two classes of CSF as follows:

$$\bar{\phi}^{INTER} \equiv \left\{\bar{\phi}_{ij,\alpha}^{k\ell}\right\} \qquad i = 3, ..., 6;\ j = 7, ..., 10;\ k\ell = 11, ..., 58, \qquad (9)$$

$$\bar{\phi}^{INTRA} \equiv \left\{\bar{\phi}_{ij,\alpha}^{k\ell}\right\} \qquad i,j = 3, ..., 6, \text{ or } 7, ..., 10;\ k\ell = 11, ..., 58, \qquad (10)$$

where $\{\ \}$ means a CSF with state α constructed by exciting a pair of electron from i and j internal orbitals into k and ℓ SCF virtual orbitals. All the possible states arising from an orbital configuration which interact with the ϕ_{SCF} are included; the number of CSF in $\bar{\phi}^{INTER}$ and $\bar{\phi}^{INTRA}$ are 36,864 and 37,057, respectively.

In order to obtain an overall description of the water-water potential function, it is necessary to cover a wide range of geometrical configurations of the water dimer.

In selecting geometrical configurations we use the following approach. First, we fix a water molecule in a coordinate frame with O_1 at the origin and with H_1 and H_2 in the xy plane with the x axis bisecting the HOH angle (see Fig. 14). We note here that the geometry for the water molecule is always kept constant with the experimental structural parameters of H_2O (R_{OH}= 0.9572 A and <HOH= 104.52°) [31]. We then

let the second water molecule approach the first from different directions, and these directions are used in this work to define different types of geometrical configurations. We have selected 10 directions (see Fig.14). From the geometries of Fig. 15, the B curve is clearly a potential curve for the bifurcated dimer.

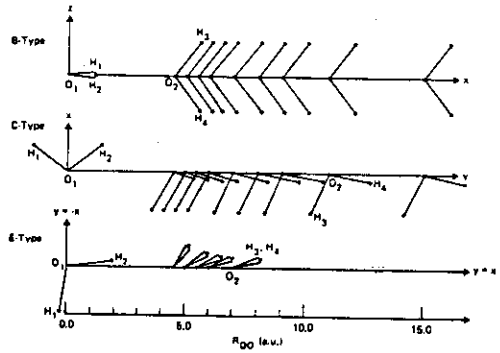


Figure 13a. Geometries of the H_2O dimer for various R_{OO}.
H_1 and H_2 are in the xy plane for all types.
B-type: H_3 and H_4 are in the xz plane.
C-type: H_3 and H_4 are off and opposite sides of the xy plane at small R_{OO}.
E-type: The xy plane is the symmetry plane of the second H_2O molecule.

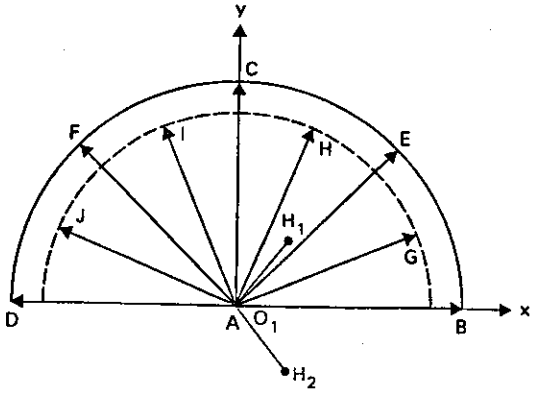


Figure 14 Geometry of the first H_2O molecule and 10 directions adopted to define geometrical configuration types of the H_2O dimer (see also Table 3).

Similarly, the E curve may be identified as that for the linear dimer, and its geometry at R_{OO}= 5.5 a.u. indicates it is very close to the optimum geometrical configuration for the water dimer. The geometry for the C curve is not exactly the symmetric cyclic structure, but it is close enough to be classified as a somewhat distorted cyclic structure in the vicinity of equilibrium. Since this region of the potential surface appears to be rather flat, evidenced by the closeness of the curves C, H and I, we shall take the C curve to represent the cyclic dimer.

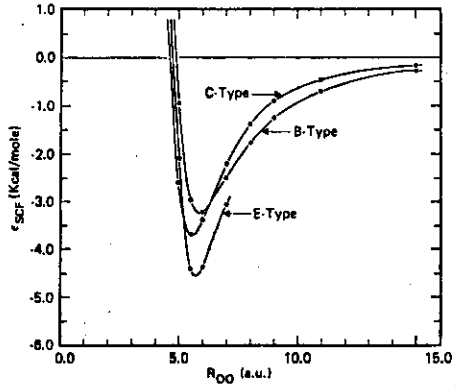


Figure15b. SCF H_2O dimer potential curves for B-, C-, and E-type geometrical configurations.

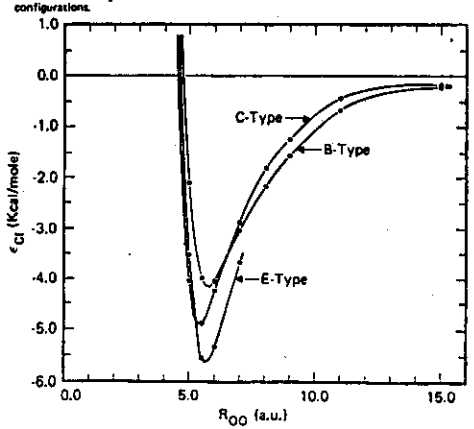


Figure15c. CI H_2O dimer potential curves for B-, C-, and E-type geometrical configurations.

Thus, the potential minima for the linear, cyclic, and bifurcated dimers are estimated to be respectively -5.6 (2.98 A), -4.9 (2.87 A), and -4.2 (3.01 A) kcal/mole (the corresponding O—O distances are given in parentheses).

The form of the potential function used is

$$\epsilon = q^2\left[\frac{1}{r_{13}} + \frac{1}{r_{14}} + \frac{1}{r_{23}} + \frac{1}{r_{24}}\right] + \frac{4q^2}{r_{78}} - 2q^2\left[\frac{1}{r_{18}} + \frac{1}{r_{28}} + \frac{1}{r_{37}} + \frac{1}{r_{47}}\right]$$
$$+ a_1 \exp(-b_1 r_{56})$$
$$+ a_2 \left[\exp(-b_2 r_{13}) + \exp(-b_2 r_{14}) + \exp(-b_2 r_{23}) + \exp(-b_2 r_{24})\right]$$
$$+ a_3 \left[\exp(-b_3 r_{16}) + \exp(-b_3 r_{26}) + \exp(-b_3 r_{35}) + \exp(-b_3 r_{45})\right]$$
$$- a_4 \left[\exp(-b_4 r_{16}) + \exp(-b_4 r_{26}) + \exp(-b_4 r_{35}) + \exp(-b_4 r_{45})\right] \quad (11)$$

The set of constants obtained by Matsuoka et al. is, in A and kcal/mole:

$a_1 = 1.088.213$ $a_2 = 666.3373$ $a_3 = 1455.427$ $a_4 = 273.5954$ $b_1 = 5.152712$
$b_2 = 2.760.844$ $b_3 = 2.961895$ $b_4 = 2.233264$ $q^2 = 170.9389$.

While stressing that any constant appearing in the analytical potential is only a fitting parameter, we may note that the dipole moment calculated using the above parameters is 2.19 D, to be compared with the dipole moment for a single water molecule, 1.85 D.

With the above potential, a Monte Carlo simulation on the liquid water structure was recently performed[32]. In this study, 343 water molecules are considered in a cube with periodic boundary conditions, the assumed water density is equal to the experimental density at 25°C (0.03334 molecules/A^3). The size of the cube is thus $(21.73)^3$ A^3 and is believed to be big enough that the long range water-water interactions are sufficiently accounted for. In the simulation the total number of configurations generated (after rejecting the initial 500.000 configurations) is 600.000; about half of them are rejected in the Markov chain[33].

The average distribution of molecules in a liquid is usually represented by radial distribution functions. The computed oxygen-oxygen radial distribution functions, $g_{OO}(r)$, $\Gamma_{OO}(r) = 4\pi\rho r^2 g_{OO}(r)$, and $N_{OO}(r) = \int_0^r \Gamma_{OO}(R)dR$ are shown in Fig. 16, together with the experimental results of Narten et al.[34]. Here is to be noted that, in deriving three radial distributions, g_{OO} g_{OH} and g_{HH}, from experimental x-ray and neutron scattering data, a model for the average orientation of pairs of near neighbouring molecules had to be assumed. (Narten had at hand only two methods involving sufficiently different values for the atomic scattering amplitudes).

It is clear from Fig. 16 that the agreement between the simulated and experimental g_{OO} is very satisfactory. This is clearly shown by analyzing the height and position of the first peak : the first peak in the simulations occurs at an oxygen-oxygen distance 2.83 A with height value of 2.48, to be compared with the x-ray data of 2.82 A and 2.31 for distance and height, respectively. Lie and Clementi[35] obtained in their Monte Carlo simulation, using only radial correlated pair potentials, the peak 2.78 at 2.78 A, whereas Stillinger and Rahman obtained, using an improved empirical potential in a molecular dynamics simulation, 3.09 at 2.85 A[36]. The two broad maxima observed experimentally at ~4.5 A and ~6.5 A are also well reproduced in the simulation. The agreement between the simulated and the experimental result is even more evident by taking into account the volume element, as can be seen from Figs. 16(b) and 16(c). It is interesting to note that all simulations of liquid water carried out up to this time, including the present work, fail to reproduce exactly the distinct, albeit small, peak observed experimentally at 3.75 A, an indication that the peak is probably a cut off ripple caused by the finite termination of the upper convolution limit.

In the radial distribution functions of Narten and Levy[34], the maximum of the first peak shifts gradually

from 2.82 to 2.24 A as the temperature is increased from 4 to 200°C. A similar trend is expected to be found in our simulation since, as noted previously, the minimum energy configuration for the pair potential used in the present work is 2.87 A, whereas the nearest neighboring oxygen-oxygen distance found in liquid at 25°C is 2.83 A.

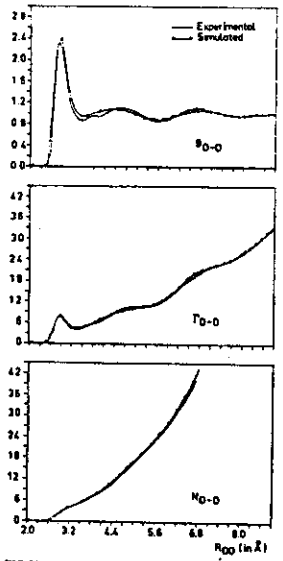


FIG.16 Oxygen–oxygen radial distribution functions.

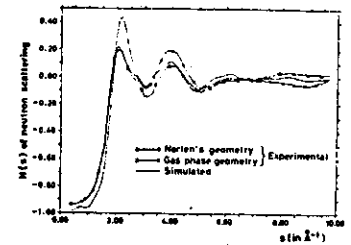


FIG.17 Structure function $H(s)$ from neutron scattering (see text for the explanation of the two different geometries used for the water molecule in constructing the experimental curves).

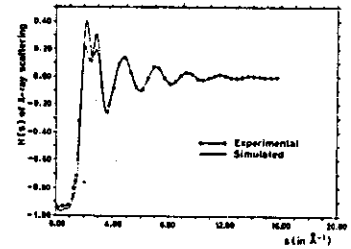


FIG. 18 Structure function $H(s)$ from x-ray scattering.

In Table 2 we numerically compare the radial function obtained with the results of the most recent, accurate molecular dynamic simulation of Stillinger and Rahman. It should be noted that any difference rests mainly in the differences in the potential functions used, but not the simulation techniques. The potential used by Stillinger and Raham, called ST2, is of empirical nature and predicts also a nearly linear hydrogen-bonded structure with bonding energy -6.8 kcal/mole as the most stable dimer configuration. All results of the ST2 potential reported in Table 2 are interpolated from the original paper between 4 and 41 °C.

Table 2 - Comparison of radial distribution functions at 25° C

		1st max		1st min		2nd max		2nd min	
g_{OO}	ST2	2.85	3.09	3.53	0.72	4.70	1.13	5.8	0.80
	CI	2.83	2.46	3.53	0.94	4.25	1.08	5.6	0.89
	Exptl	2.83	2.31	3.45	0.85	4.53	1.12	5.6	0.86
g_{OH}	ST2	1.90	1.38	2.50	0.31	3.40	1.60	4.60	0.92
	CI	1.90	1.08	2.55	0.28	3.35	1.68	4.85	0.90
	Exptl	1.90	0.80	2.45	0.50	3.35	1.70	–f	–f
g_{HH}	ST2	2.50	1.50	3.10	0.78	4.00	1.15	5.40	0.96
	CI	2.50	1.40	3.10	0.86	3.90	1.20	5.50	0.93
	Exptl	2.35	1.04	3.00	0.47	4.00	1.08	–f	–f

Since the radial distribution functions are indirect experimental data, it is desirable to calculate the x-ray and neutron scattering intensities from our simulated radial distribution functions and thus to be in a position to compare directly with the corresponding original experimental data.

The coherent scattering intensity I(s) for x-ray, neutron, or electron scattering can be treated in the same way and, neglecting the surface scattering, is given, in number of electrons per molecule, by

$$I(s) = \langle F^2 \rangle + \left(\sum_{\alpha} \sum_{\beta} f_{\alpha} f_{\beta} \right) \int_0^{\infty} 4\pi r^2 \rho \{ g_{\alpha\beta}(r) - 1 \} \frac{\sin(sr)}{sr} \, dr,$$

where

$$\langle F^2 \rangle = \sum_i \sum_j f_i f_j \exp(-b_{ij} s^2) \frac{\sin(sr_{ij})}{sr_{ij}}$$

is the scattering intensity from one independent molecule averaged over all orientations, depending on the average interatomic distances r_{ij} and their mean-square variations $2b_{ij}$. Note that indices i and j are used to refer to atoms belonging to the same molecule, while α and β to different molecules. f_{α} (or f_i) is the static coherent scattering amplitude for atom α (or i). β is the experimental (bulk) density of water and $g_{\alpha\beta}(r)$ the radial distribution function for the pair α and β. s is related to the scattering angle 2θ and the wavelength λ of the incident beam by the following equation:

$$s = 4\pi \sin\theta / \lambda .$$

The scattering amplitude f_a is independent of s in the case of neutron scattering and is a function of s for x-ray scattering.

In studying the structure of liquid, however, it seems more meanful to subtract out the intensity due to the scattering of a single independent molecule; hence, we can introduce a structure function H(s) defined as

$$H(s) = \frac{I(s) - \langle F^2 \rangle}{\langle F^2 \rangle}$$

The denominator is used for normalization, so that H(s) approaches 0 as s-0 if there is no structure (i.e., idea gas) and -1 if the structure is highly ordered (e.g., perfect rigid crystal). We note that the definition used here is different from that of Narten et al.[34], and is believed to be more instructive in presenting a structure function for nonsimple liquids, since all interference patterns in H(s) are then solely caused by the deviation of g from unity. At large angles only independent molecular scattering is observed, and hence H(s) approaches zero at large s.

The neutron atomic scattering amplitudes used in this work are $f_O = 0.566 \times 10^{-12}$ cm[37] and $f = 0.67 \times 10^{-12}$ cm[38]. We assume that the structure of liquid H_2O is the same as that of liquid D_2O in order to compare the simulated neutron intensity with experiment. The atomic x-ray form factors, a function of s, are extracted and interpolated from the tabulation of Narten and Levy[34]. Fig. 17 compares the simulated x-ray structure function with the experimental results. The agreement seems to us to be very satisfactory. The unique double peaks observed experimentally in the range $s \cong 2$ A^{-1} and $s \cong 3$ A^{-1} have been found also in the simulation, although the computed left peak is too high, whereas the right one is too low. Comparing with the simulated results of Lie and Clementi[35], where no angular correlations were included in the potential, we can notice not only qualitative but also quantitative improvements in the present work. The structure functions constructed from the neutron scattering intensities of the simulation and the experiment[34] are compared in Fig. 18. The agreement is less satisfactory but still moderately good. One reason for the discrepancy may stem from the fact that the neutron scattering lengths for the oxygen and deuterium atoms are about the same, hence a precise knowledge of the dimensions of the water molecule is necessary in untangling the intramolecular scattering from the experimental total intensity. This difficulty is practically not present in the x-ray scattering of liquid water, since the scattering is dominated by the oxygen atom. Two different geometries for the water molecule are used to subtract the single molecule scattering from the experimental results of Narten in Fig. 16: the first one is that obtained from the gas phase by Benedict et al.[39] with $b_{OH} = 0.0022$ A^2 and $b_{HH} = 0.0066$ A^2 from Shibata and Bartell[40], the second that obtained by Narten[34] by least squares fitting to the observed data and with the assumption

of a model. It should therefore be noted that the OD distance found by Narten is 0.02 A shorter than the equilibrium distance in D_2O vapor, whereas the DD distance is 0.02 A longer. However, Narten suggested that differences should not be accepted as real unless confirmed by a more refined neutron experiment. Another source of discrepancy between the simulation and experiment may lie in the following different definitions of the "structure" as "observed" by the simulation and the x-ray and neutron experiments. As is well known, the molecules in a liquid undergo rotational and translational displacements and more frequent oscillations. The structure observed in the coherent x-ray or neutron scattering is called, according to Eisenberg and Kauzmann[41], diffuse-averaged or D structure, which can be regarded as either a time average or a space average of different vibration-averaged structures. In our simulation, however, no vibrational average of each displacement generated by a Markov chain was taken into account.

Despite all these problems in interpretation and comparison it seems clear from Fig. 16 that the intramolecular scattering affects almost exclusively the region for $s > 2.5$ A^{-1} and that the simulated positions of the scattering maxima and minima agree well with experiment, but the simulated intensities less so. Since the zero angle scattering is connected with the isothermal compressibility, we postpone its discussion to the next section.

We are of the opinion that starting from first principles, one can readily simulate equilibrium properties of a quite complex system, like liquid water, to a high degree of accuracy. Although three-body (or even four-body) interactions (and possibly quantum effects) have frequently been suggested to be important in the study of the structure of water, it seems, judging from the results presented here, that they contribute only nominally to the pair distribution functions and heat capacity. We are presently working towards a quantitative definition of the importance of three-body effects.

(34)

Monte Carlo Study of Liquid Water with Two- and Three-Body *ab Initio* Potentials

E. CLEMENTI AND G. CORONGIU
IBM Corporation, IS/TG, Poughkeepsie, New York 12602, U.S.A.

Abstract

An *ab initio* three-body potential (called CC) is presented, discussed, and applied. Using the *ab initio* two-body potential of Matsuoka–Clementi–Yoshimine (1975) and the CC potential, we have carried out Metropolis-Monte Carlo simulation at 298 K for an (N, V, T) ensemble with N = 64, 125, and 343 water molecules. Improvements, relative to the use of only the two-body potential, have been obtained for the g(O—O), g(O—H), g(H—H), g(N) correlation functions, for the X-ray and neutron beam scattering intensities, and for the enthalpy.

Introduction

As is known, the many-body correction is among the most celebrated problems in classical and quantum mechanics. In this work, we present a study on the three-body interaction energy for water trimers and a computer experiment on liquid water. As proposed over 10 years ago [1], we obtain the intermolecular potentials from *ab initio* computations and we use these potentials in statistical mechanical computer experiments on pure solvents or solutions [2a,3]. Our main goal is to assess the reliability degree in predictions obtained for *ab initio* solvents and solutions relative to laboratory solvents and solutions. However, since exact potentials are as unattainable as exact wavefunctions, the concomitant goal is to derive *an ordered set of approximations leading to an increasingly realistic description of the* ab initio *liquid*. We recall that a rather standard alternative routine is the use of experimental data for the determination of the values of parameters present in an intermolecular potential [2b]. We recall, in addition, that this latter avenue can be used only when the experimental data are available in *sufficient number and quantity*. For the case of water, the experimental wealth of data is rather enormous [4]. This notwithstanding, the "experimentally derived potentials" have been subjected to criticism [5]. This clearly illustrates that the task of obtaining a reliable intermolecular potential from experimental data is not a trivial one, even when many and accurate laboratory data are available.

The CC Three-Body Potential

For an N body system, the standard expression to represent the total interaction energy system, $E(1, \ldots, N)$, is the N term series expansion

$$E(1, \ldots, N) = \sum \varepsilon(i, j) + \sum \varepsilon(i, j, K) + \cdots + \varepsilon(1, \ldots, N), \quad (1)$$

International Journal of Quantum Chemistry: Quantum Biology Symposium 10, 31–41 (1983)

CCC 0360-8832/83/010031-11$04.00

where the summations are extended over all possible doublets (i, j), triplets (i, j, k), etc. In the above system, 1, . . . ,N are atoms or molecules which are considered classically. From standard quantum mechanical notions, we know that the interaction energy of the above system—considered not as an ensemble of atoms and molecules but as an ensemble of nuclei and electrons—can be decomposed into the Hartree–Fock energy E_0 (namely, the energy corresponding to the one-electron approximation model) and the electronic correlation energy E'. Thus, we can write

$$E(1, \ldots, N) = E_0(1, \ldots, N) + E'(1, \ldots, N), \tag{2}$$

where

$$E_0(1, \ldots, N) = \sum \varepsilon_0(i, j) + \sum \varepsilon_0(i, j, k) + \cdots + \varepsilon_0(1, \ldots, N), \tag{3}$$

and

$$E'(1, \ldots, N) = \sum \varepsilon'(i, j) + \sum \varepsilon'(i, j, k) + \cdots + \varepsilon'(1, \ldots, N). \tag{4}$$

Whereas E_0 can be computed for relatively large systems (recently, up to 80 to 100 atom molecules have been reported [6]), E' can be computed reasonably well only for systems with 20 to 40 electrons at most. (The latter part of this statement is unfortunately true, especially when we consider variational techniques, such as MCSCF, CI, etc.; however, perturbational technique and density functional methods, which have been available for some time, seem to yield reliable energies [3].) In considering eqs. (3) and (4), it has been noted [7] that $\varepsilon_0(i, j, k) + \varepsilon'(i, j, l) = \sim\varepsilon_0(i, j, k)$, since the major contribution to (i, j, k) comes from the induction energy. Numerical verifications of the above proposition have been presented, for example, for systems such as $(H_2O)_3$, $(H_2O)_2$—Li^+, (H_2O)—Li^+—F—; see Refs. 8, 9, and 10, respectively. In particular, in a recent numerical experiment [11], $\varepsilon_0(1, 2, 3)$ has been compared with $\varepsilon(1, 2, 3)$ for a number of water trimers yielding the results reported in the left inset of Figure 1. For each trimer geometry, a point in the x,y plane is associated, x being the $\varepsilon(1, 2, 3)$ value and y, the $\varepsilon_0(1, 2, 3)$ value. One can readily see that the points fell very near to the diagonal, where $x = y$.

On the basis of the above arguments [7–11], we have computed the Hartree–Fock energy for 173 different geometrical arrangements of trimers of water. The basis set on the oxygen atom consists of 13 s functions (contracted to 6), 8 $2p(x)$ functions (contracted to 3, and equally for the $2p(y)$ and $2p(z)$ functions), and one set of $3d$ polarization functions. The basis set for the hydrogen atoms has 6 s function (contracted to 2) and a set of p polarization functions (see Ref. 12). By indexing three water molecules as 1, 2, and 3, respectively, the Hartree–Fock three-body energy is defined as

$$\varepsilon_0(1, 2, 3) = E_0(1, 2, 3) - E_0(1, 2) - E_0(1, 3) \sim E_0(2, 3) + E_0(1) + E_0(2) + E_0(3), \tag{5}$$

where $E_0(1) = E_0(2) = E_0(3)$. As is known, the use of a large basis set ensures that the basis set superposition error will not be too large. However, for accurate

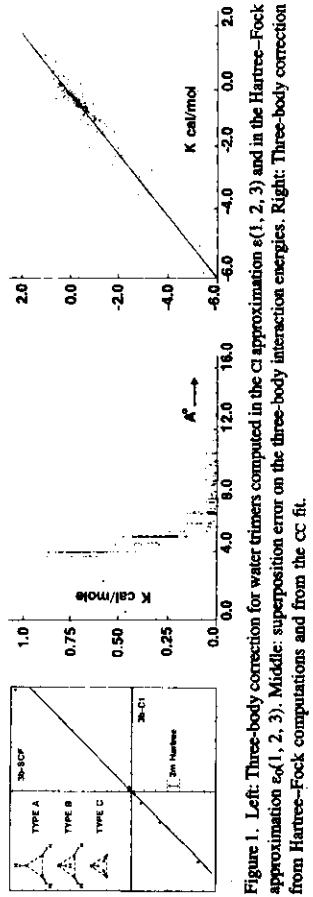


Figure 1. Left: Three-body correction for water trimers computed in the CI approximation $\varepsilon(1, 2, 3)$ and in the Hartree–Fock approximation $\varepsilon_0(1, 2, 3)$. Middle: superposition error on the three-body interaction energies. Right: Three-body correction from Hartree–Fock computations and from the CC fit.

computations, as in our case, the superposition error is too large to be neglected; therefore, a correction was computed using the counterpoised technique [13] yielding the three-body energy as defined below:

$$\varepsilon^*(1, 2, 3) = E_0(1, 2, 3) - E^*(1, 2) - E^*(1, 3) - E^*(2, 3) + E^*(1) + E^*(2) + E^*(3), \quad (6)$$

with $E^*(1) \neq E^*(2) \neq E^*(3)$. In Figure 1 (center), we report the value of $\Delta E(i, j) = E^*(i, j) - E_0(i, j)$ in kcal/mol versus R(O—O), the oxygen–oxygen distance (in angstrom units). Note that the error can be of the order of 1 kcal/mol. In addition, note that $\Delta E(i) = E^*(i) - E_0(i)$ can be as large as $\Delta E(i, j)$, namely, the superposition error for the monomers can be as large as the error for the dimers. Inspection of the distribution of the $\Delta E(i, j)$ values tells us that computations of three-body corrections for the water trimer without corrections for the basis set superposition error should be viewed with scepticism.

As usual in our work, the potential is obtained by fitting an analytical expression to the computed *ab initio* energies, namely, the 173 three-body interaction energy values. Since the induction energy is known to be a good approximation of the three-body nonadditivity correction [3], its expression will represent an obvious *starting* choice for an analytical representation. The *final* expression is the one reported below:

$$\varepsilon_0 \text{ (three-body)} = -\tfrac{1}{2} \sum_{i \neq k} [\alpha (\boldsymbol{\xi}_i \cdot \boldsymbol{\xi}_i) + \delta (\boldsymbol{\xi}_i \cdot \mathbf{e}_i)^2], \quad (7)$$

where α and δ can be considered fitting constants (with values 3.91 and 1.42, respectively), i is the index for the ith bond in the kth molecule, e_i is the unit vector in the direction of the bond i, and $\boldsymbol{\xi}_i$ is a quantity defined for the midpoint of bond i by the relation

$$\boldsymbol{\xi}_i = \sum_j (q(j)^* R_{ij}^3 / |R_{ij}^3 - C_j|) * \mathbf{R}_{ij}/R_{ij}^3, \quad (8)$$

where the summation extends over all atoms with point charges $q(i)$ of the other molecules in the trimer different from K, and R_{ij} denotes the radius vector from the atom j to the midpoint of the bond. The fitting constants C_j have the values 3.038 and 0.00 (in units of Å^3) for oxygen and hydrogen, respectively. Alternatively stated, we have found that there is no need for a direct dependence on C_j for the hydrogen atoms. Note that eq. (8) is written for Monte Carlo simulations where it is used with a cutoff for the O—O distance; indeed, it should be evident that the equation needs corrective terms for small R_{ij} values. The values of $q(i)$ are -0.6436 and 0.3218 for oxygen and hydrogen, respectively. One can easily verify that eqs. (7) and (8) are closely related to the expression for the induction energy (see Ref. 3, page 81).

In Figure 1 (right inset), the three-body corrections for the 173 trimers are reported as obtained from the *ab initio* computation (x axis) and from the fitting equation (y axis); one can see that the accuracy of the fit is very satisfactory, as also indicated by the mean standard deviation value, $\sigma = 0.21$ kcal/mol. The three-body correction given by the expression in eq. (7) is hereafter referred to as the CC three-body correction.

Comments on the MCY Two-Body Potential

The two-body potential we shall use in this work is the MCY potential [14], which was obtained by fitting *ab initio* interaction energies obtained from extended CI computations. A most interesting feature of this potential is that it yields a surprisingly good agreement with a number of experimental data on liquid water, despite the fact that it represents only the *first term* in the expansion series of eq. (1). The surprise is actually somewhat lessened by realizing that the first term in eq. (2) alone can reproduce the main gross structural features of liquid water as shown in Ref. 15. In addition, a rather rough approximation to $\varepsilon'(i, j)$ is known to yield immediately a *notable* improvement in the simulation of the pair correlation functions and X-ray scattering data for liquid water (see Ref. 16).

However, other features of liquid water were not satisfactorily reproduced, as was pointed out in our earlier papers on the MCY potential [17,18]. For example, the pressure is too high (even keeping in mind the very large error in general associated with computations of the pressure), the second virial coefficient is not well represented (errors up to 50% were reported in Ref. 18), and the enthalpy is too small. Other quantities related to the above one are also (and necessarily) affected. A number of articles have been written pointing out in more detail [19] these shortcomings. Despite all this, the MCY potential is among the most reliable in the literature, *especially for structural information*. It should be recalled that the number of *ab initio* CI computations for geometrical configurations of the dimer was small [17] relative to the number of parameters in need of fitting. For this reason, the MCY potential was refitted, starting from a perturbation potential [16], which in turn was obtained by refitting the previously obtained Hartree–Fock potential [15] where many *ab initio* points were computed.

As is known, the selection of an analytical form (model) is a rather critical step. For example, more recently a rather large number of water dimers have been computed in the CI approximation [20], where we experimented with a number of different ansatzes. A "very good" fit was obtained, on the basis of the mean standard deviation (as a criterion); however, the new potential was very unsatisfactory on the basis of the second virial coefficients (about 100% in error), even if it predicted the equilibrium geometry for the dimer well. This discussion is presented in order to recall that a good fit (that is, small mean standard deviation) is needed, but it is in itself far from sufficient. These comments holds also for another recent note [21]. In addition, one should always remember that a "good fit" for a set $S(1)$ of points might turn out to be a somewhat "poor" fit for a second and larger set $S(2)$ containing $S(1)$.

In conclusion, on the one hand, we are working at obtaining a more accurate *ab initio* two-body potential for water–water interactions [20]. On the other hand, we shall keep on using the MCY potential until we have derived a truly superior two-body potential.

In the following section, we compare the use of the MCY potential with and without the three-body correction CC.

Monte Carlo Simulations

The Metropolis-Monte Carlo simulation [22] was carrier out with an (N, V, T) ensemble for T = 298 K and N = 64, 125, or 343 water molecules at the experimental density ρ = 0.998 g/cm^3. Cubic boundary conditions with a minimum image cutoff have been used [23]. As usual [22b], we started the Monte Carlo simulation with an ice-type structure. Then, we annealed the sample to T = 2000 K to ensure an initial random distribution; and after some time, we cooled the sample back to T = 298 K. The initial part of the computation is spent to obtain equilibration and is discarded in the statistics; we consider our sample equilibrated when the fluctuation (error) in the enthalpy is about ±0.02 kcal/mol per water molecule. The optimized step size for our computation is

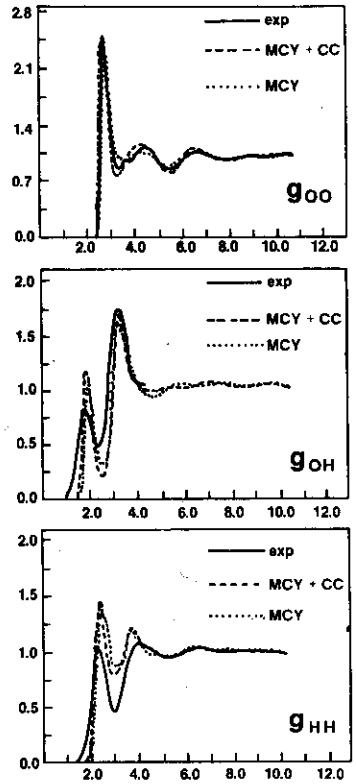


Figure 2. Pair correlation functions obtained by modeling X-ray diffraction data (Narten) or simulated with either the MCY or the (MCY + CC) potentials: $g_{OO}(r)$ (top), $g_{OH}(r)$ (middle), $g_{HH}(r)$ (bottom).

about 0.30 ± 0.05 Å, and the optimized angle is about 20 ± 5°. The collection of statistical data starts at this stage, and about 10^6 steps were carried out.

We recall that, in general, different properties have a different dependence on N, unless N is really large: we *consider the range 64 to 100 water molecules as submarginal*. Indeed, this can be easily recognized by recalling (1) that a trimer can occupy a volume with a length of 8 to 10 Å; (2) that the half-side of the cubic box containing 64, 125, or 343 water molecules is 6.2, 7.6, or 10.5 Å, respectively; and (3) that a correct description of dimers, trimers, and tetramers of water molecules is essential to any realistic description of liquid water.

The main results are collected in Figures 2 to 4 and in Tables I and II. In Figure 2, we compare the pair correlation functions g(O—O), g(O—H), and g(H—H) obtained with MCY either with or without CC (designated as MCY + CC or MCY, respectively). The correlation functions are reported for N = 343 and compared with those obtained by Narten modeling X-ray and neutron beam diffraction data [24]. In Figure 3, we report the computed structure function $H(s)$, defined as

$$H(s) = [I(s) - \langle F^2 \rangle]/\langle F^2 \rangle, \tag{9}$$

where s is related to the scattering angle 2ϑ and the wavelength λ of the incident beam by the relation $s = 4\pi \sin \vartheta/\lambda$, $I(s)$ is the coherent scattering intensity, and $\langle F^2 \rangle$ is the scattering intensity for one independent molecule averaged over

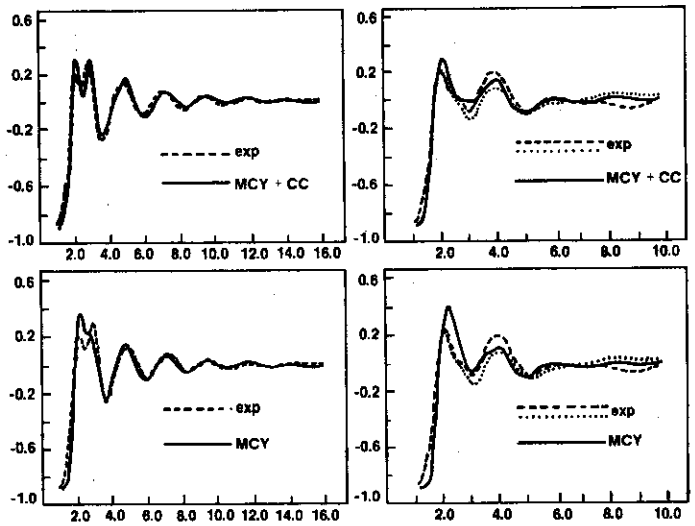


Figure 3. Comparison of X-rays structured function $H(s)$ from experiments (Narten) and from simulations with either MCY or (MCY + CC) potentials; same for neutron base scattering.

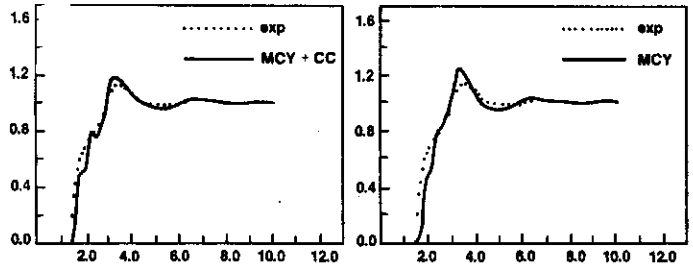


Figure 4. Comparison of Dore's neutron data with simulated $g_N(r)$ obtained either with MCY or (MCY + CC) potentials.

all orientations. For details on these computations, we refer to our earlier report [17].

The insets of Figure 4 compare our weighted sums of g(D—D), g(O—D), and g(O—O) with the laboratory data by Dore [25] reporting recently obtained neutron diffraction experiments. We should point out that the X-ray and neutron data are not in full agreement with the electron diffraction data [26]; this fact might advise some prudence in discussing the agreement with "experimental" data.

Discussion and Conclusions

From Figure 2, it can be seen that the (MCY + CC) potential yields correlation functions which are more structured than those obtained with the MCY potential. The previously obtained agreement [17] was rather good, and now it remains essentially the same, which can be seen also by inspection of Table I. In Table I, we have compared our results for $N = 343$ with those obtained with $N = 64$ and $N = 125$. As one can note, there are differences in increasing N from 64 to 343, but these are not as large as those shown by comparing the average enthalpy (see below). In this table, we have also reported Stillinger and Rahman's results with the ST2 [27] and with the SD [28,29] potentials. We recall that these empirical potentials were derived from the BNS [30] and Rowlinson [31] potentials.

Narten's X-ray and neutron beam data (see Fig. 3) are more closely reproduced by the "MCY + CC" potential than by the MCY potential; note in particular the intensity agreement in the first split peak (X-ray) and in the first peak (neutrons). The more recent neutron data by Dore [25] are compared in Figure 4 with both MCY and MCY + CC; again, the latter is an improvement on the former. We recall that Stillinger and Rahman's simulation [29] yielded a rather poor agreement with Dore's data (for details see Ref. 25). We recall, in addition, that the electron diffraction data [26] seem to yield correlation functions more structured than those reported from the neutron beam data. Finally, simply by in-

TABLE I. Intensity I at the radial distance R of the maxima (M) and minima (m) of pair correlation functions.

g_{OO}								
Source	1st M		1st m		2nd M		2nd m	
	R	I	R	I	R	I	R	I
Narten[a]	2.83	2.31	3.45	0.85	4.53	1.12	5.6	0.86
MCY[b]	2.83	2.46	3.53	0.94	4.25	1.08	5.6	0.89
ST2[c]	2.85	3.09	3.53	0.72	4.70	1.13	5.8	0.89
SD[d]	2.85	3.10	3.25	0.63	4.60	1.17	5.75	0.80
(MCY + CC)-64[e]	2.80	2.68	3.40	0.70	4.50	1.22	5.65	0.80
(MCY + CC)-125[e]	2.80	2.69	3.40	0.73	4.53	1.17	5.65	0.80
(MCY + CC)-343[e]	2.81	2.66	3.37	0.75	4.50	1.17	5.65	0.82
g_{OH}								
Source	1st M		1st m		2nd M		2nd m	
	R	I	R	I	R	I	R	I
Narten[a]	1.90	0.80	2.45	0.50	3.35	1.70	-	-
MCY[b]	1.90	1.08	2.55	0.28	3.35	1.68	4.85	0.90
ST2[c]	1.90	1.38	2.50	0.31	3.40	1.60	4.60	0.92
SD[d]	1.83	1.47	2.42	0.13	3.13	1.50	3.96	0.98
(MCY + CC)-64[e]	1.85	1.28	2.55	0.17	3.31	1.63	4.60	0.97
(MCY + CC)-125[e]	1.88	1.30	2.50	0.19	3.30	1.64	4.62	0.94
(MCY + CC)-343[e]	1.87	1.28	2.50	0.19	3.28	1.62	4.65	0.94
g_{HH}								
Source	1st M		1st m		2nd M		2nd m	
	R	I	R	I	R	I	R	I
Narten[a]	2.35	1.04	3.00	0.47	4.00	1.08	-	-
MCY[b]	2.50	1.40	3.10	0.86	3.90	1.20	5.50	0.93
ST2[c]	2.50	1.50	3.10	0.78	4.00	1.15	5.40	0.96
SD[d]	2.17	1.81	3.00	0.69	3.58	1.17	-	-
(MCY + CC)-64[e]	2.40	1.49	3.00	0.75	3.78	1.27	-	-
(MCY + CC)-125[e]	2.41	1.49	2.97	0.76	3.80	1.18	5.45	0.93
(MCY + CC)-343[e]	2.40	1.46	3.00	0.77	3.80	1.22	5.36	0.94

[a]Ref. 24.
[b]Ref. 17.
[c]Ref. 27.
[d]Ref. 29.
[e]This work with $N = 64$, 125, and 343.

TABLE II. Contributions to the internal energies and enthalpies for $N = 64$, 125, and 343.

N	U(2-body)	U(3-body)	U(2 + 3-body)	U - 3 kT
64	-9.194±0.014	-1.109±0.012	-10.303±0.024	-8.527
125	-8.748±0.021	-0.891±0.010	-9.639±0.027	-7.863
343	-8.620±0.012	-0.858±0.002	-9.477±0.013	-7.701

spection of the simulated $g_{OH}(r)$ and $g_{HH}(r)$ results, one can expect that additional improvement will be gained by relaxing the rigidity constraint in the water molecules (this topic will be discussed in a forthcoming report where vibrations are allowed). The main corrections which need to be added to the MCY + CC model are, therefore, (1) vibrational freedom and (2) higher-order corrections, in particular the four-body correction which has been estimated to be nonnegligible [32] and of the order of a few tenths of a kilocalorie per mole for water tetramers near equilibrium.

The computed internal energy (see Table II) varies with N. For $N = 64$, 125, and 343, the internal energy converges to a value of about -7.70 ± 0.05 kcal/mol, which should be compared with -8.1 kcal/mol from experimental data and with the value of -6.8 ± 0.2 kcal/mol obtained from the MCY potential for $N = 343$. From these data, it appears that the remaining corrections (neglected in this work) will bring about a contribution of about 0.5 kcal/mol. Note that one could use a small value for N and, by adding *ad hoc* corrections to the potential, one can obtain an energy close to the one corresponding to a large value of N. We have not used this avenue for the reason expressed in the previous section.

Bibliography

[1] (a) E. Clementi, in *Physics of Electronic and Atomic Collisions*, T. R. Govers and F. J. Heer, Eds. (North-Holland, Amsterdam, 1971), pp. 399–426; (b) E. Clementi and H. Popkie, J. Chem. Phys. **57,** 1077 (1972).

[2] (a) See, for example, the monographic works by E. Clementi, Determination of Liquid Water Structure, Coordination Numbers for Ions and Solvation for Biological Molecules, *Lecture Notes in Chemistry*, Vol. 2 (Springer-Verlag, Berlin, New York, 1976); (b) P. Shuster, W. Jacubetz, W. Marius, and S. A. Rice, *Structure of Liquids* (Springer-Verlag, Berlin, New York, 1975).

[3] See, for example, the monographic work by E. Clementi, Computational Aspects for Large Chemical Systems, *Lecture Notes in Chemistry*, Vol. 19 (Springer-Verlag, Berlin and New York, 1980).

[4] See, for example, the five volumes *Water: A Comprehensive Treatise* F. Franks, Ed. (Plenum, New York, 1973).

[5] See for example, 1978 issue of the *Faraday Discussions of the Chemical Society*, No. 66, Structure and Motion of Molecular Liquids (Burlington House, London, 1978).

[6] E. Clementi and G. Corongiu, Int. J. Quantum Chem. **9,** 213 (1982).

[7] (a) P. Claverie, in *Intermolecular Interactions: From Diatomic to Biopolymers*, B. Pullman, Ed. (Wiley, New York, 1978); (b) W. Kolos, Theor. Chim. Acta, Berlin, **51,** 219 (1979); (c) B. M. Axilrod and E. Teller, J. Chem. Phys. **11,** 299 (1943); (d) A. D. Buckingham, Adv. Chem. Phys. **12,** 107 (1967); (c) W. Kolos, in *New Horizons of Quantum Chemistry*, P. O. Lowin and B. Pullman, Eds. (D. Reidel, Boston, 1983), pp. 243–277.

[8] E. Clementi, W. Kolos, G. C. Lie, and G. Ranghino, Int. J. Quantum Chem. **17,** 377 (1980); H. Kistenmacher, G. C. Lie, H. Popkie, and E. Clementi, J. Chem. Phys. **61,** 546 (1974).

[9] E. Clementi, H. Kistermacher, W. Kolos, and S. Romano, Theor. Chim. Acta, Berlin, **55,** 257 (1980); H. Kistenmacker, H. Popkie, and E. Clementi, J. Chem. Phys. **61,** 799 (1974).

[10] J. W. Kress, E. Clementi, J. J. Kozak, and M. E. Schwartz, J. Chem. Phys. **63,** 3907 (1975).

[11] P. Habitz, P. Bagus, P. Siegbahn, and E. Clementi, Int. J. Quantum Chem. **23,** 1803 (1983).

[12] The basis set for the trimer is available in a IBM Research Report available upon request to the authors of this article.

[13] S. F. Boys and F. Bernardi, Mol. Phys. **19,** 553 (1970).

[14] O. Matsuoka, E. Clementi, and M. Yoshimine, J. Chem. Phys. **64,** 1351 (1976).

[15] H. Popkie, H. Kistenmacher, and E. Clementi, J. Chem. Phys. **59,** 1325 (1973); H. Kistenmacher, H. Popkie, E. Clementi, and R. O. Watts, J. Chem. Phys. **60,** 4455 (1974).

[16] G. C. Lie and E. Clementi, J. Chem. Phys. **62,** 2195 (1975).

[17] G. C. Lie, E. Clementi, and M. Yoshimine, J. Chem. Phys. **64,** 2314 (1976).

[18] G. C. Lie and E. Clementi, J. Chem. Phys. **64,** 5308 (1976).

[19] M. Mezei, S. Swaminathan, and D. L. Beveridge, J. Chem. Phys. **71,** 1773 (1979); R. W. Impey, M. L. Klein, and I. R. McDonald, J. Chem. Phys. **74,** 647 (1981); J. C. Owicik and H. A. Scheraga, J. Am. Chem. Soc. **99,** 7403 (1977); I. R. McDonald and M. Klein, J. Chem. Phys. **68,** 4875 (1978).

[20] E. Clementi and P. Habitz, J. Phys. Chem. **87,** 2815 (1983).

[21] D. G. Bounds, Chem. Phys. Lett. **96,** 604 (1983).

[22] (a) N. Metropolis, A. W. Rosenbluth, A. H. Teller, and E. Teller, J. Chem. Phys. **21,** 1078 (1953). (b) Our first Monte Carlo program has been documented by J. Fromm, E. Clementi, and R. Watts, *IBM Research Report* (San Jose, California, 1973); our second was documented by R. Barsotti and E. Clementi, *Montedison Technical Report* (Istituto G. Donegani, Novara, Italy, 1976); our third version, by S. Romano and E. Clementi, Gazz. Chim. Ital. **108,** 506, 319 (1978).

[23] See, for example, page 150 of J. P. Vallau and S. G. Whittington, in *Statistical Mechanics—Part A: Equilibrium Techniques*, B. J. Berne, Eds. (Plenum, New York, 1977).

[24] A. H. Narten and H. A. Levy, J. Chem. Phys. **55,** 2263 (1971); A. H. Narten, J. Chem. Phys. **56,** 5861 (1972); A. H. Narten, *ORNL-4578 Report* (Oak Ridge Natl. Lab., Chemistry Div. Oak Ridge, Tennessee (1970).

[25] J. C. Dore, Faraday Disc. Chem. Soc. **66,** 82 (1978).

[26] G. Palinkas, E. Kalman, and P. Kovacs, Mol. Phys. **34,** 525 (1977).

[27] F. H. Stillinger and A. Rahman, J. Chem. Phys. **60,** 1545 (1974).

[28] F. H. Stillinger and C. W. David, J. Chem. Phys. **69,** 1473 (1978).

[29] F. H. Stillinger and A. Rahman, J. Chem. Phys. **68,** 666 (1978); F. H. Stillinger, J. Chem. Phys. **71,** 1547 (1979).

[30] A. Ben-Naim and F. H. Stillinger, in *Structure and Transport Processes in Water*, R. H. Horne, Ed. New York, Wiley-Interscience, 1972.

[31] J. S. Rowlinson, Trans. Faraday Soc. **47,** 120 (1951).

[32] H. Kistenmacher, G. C. Lie, H. Popkie, and E. Clementi, J. Chem. Phys. **61,** 545 (1974).

Received April 7, 1983

Parallelism in Computational Chemistry: Applications in Quantum and Statistical Mechanics

by

E. Clementi, G. Corongiu, J. H. Detrich, H. Kahnmohammadbaigi,
and
S. Chin, L. Domingo, A. Laaksonen, H. L. Nguyen

IBM Corporation IS&TG
Department D55/Building 701
P. O. Box 390
Poughkeepsie, New York 12602
and
National Foundation for Cancer Research
Poughkeepsie Laboratory

V. Parallelism in Monte-Carlo Simulations.

For some time now we have been deriving intermolecular potentials from ab initio computations and using these potentials in statistical

mechanical computer experiments on pure solvents or solutions(26), with the goal of deriving an ordered set of approximations leading to an increasingly realistic description of the ab initio liquid. An important step along this road is the recent work of Clementi and Corongiu(27), which obtains a three-body interaction potential for water from ab initio calculations and applies this in Metropolis-Monte Carlo simulations of liquid water. In the last section of this paper we shall consider this problem once more, this time using Molecular dynamics.

Here we briefly review this potential, and its generalization to include a four-body intermolecular interaction potential as well, (details are available in a recent paper by J. H. Detrich, et al(28)). This lays the concrete groundwork for our main concern here, which is the development of a practical computational scheme to meet the very heavy demands made, for example, by our increasingly realistic computer simulations of liquid water.

For an N body system we can represent the total intermolecular interaction energy, U, in terms of a series of the form

$$U = \sum_i \sum_{j \neq i} U^{(2)}_{i,j} + \sum_i \sum_{j \neq i} \sum_{k \neq i,j} U^{(3)}_{i,j,k} + \sum_i \sum_{j \neq i} \sum_{k \neq i,j} \sum_{\ell \neq k,i,j} U^{(4)}_{i,j,k,\ell} + \cdots \qquad (1)$$

where the summations range from 1 to N, and the various $U^{(k)}$ are defined so that they are necessarily zero in case $N < k$.

Clementi and Corongiu base their expression for $U^{(3)}$ on a classical bond polarizibility model. The water molecules are taken to be rigid, but polarizible, so that a dipole $\vec{\mu}_{bi}$ is induced at the midpoint of each OH bond according to the relation

$$\vec{\mu}_{bi} = \alpha \vec{E}_{bi} + \delta \vec{e}_{bi} (\vec{e}_{bi} \cdot \vec{E}_{bi}) \qquad (2)$$

here and are the polarizibility constants for the water OH bonds, i is the index denoting the i-th water molecule, b is the index used to distinguish the two bonds in a given molecule, $\vec{e}_{bi}$ is the unit vector in the direction of the bond bi, and $\vec{E}_{bi}$ is the electric field at the midpoint of the bond bi. The electric field $\vec{E}_{bi}$ originates from the assumed permanent electrostatic moments of the surrounding water molecules, and is naturally expressed as a sum

$$\vec{E}_{bi} = \sum_{j \neq i} \vec{E}_{bi,j} \qquad (3)$$

where $\vec{E}_{bi,j}$ is the electric field component at the bond midpoint bi originating from the permanent electrostatic moments of the j-th water molecule.

To first order in the polarizibility, this model yields the energy given by

$$U_{P1} = -\frac{1}{2} \sum_{bi} \vec{\mu}_{bi} \cdot \vec{E}_{bi} \qquad (4)$$

If we substitute Eqs. (2) and (3) in this expression, we obtain a

triple sum which shows the three-body character of our interaction energy.

Continuing now to second order in the polarizibility, we find an interaction energy which is conveniently expressed in terms of dipole-dipole interactions, namely

$$U_{P2} = \sum_{bi} \sum_{b'j, j \neq i} \left[\vec{\mu}_{bi} \cdot \vec{\mu}_{b'j} / r^3_{bi,b'j} - 3 (\vec{r}_{bi,b'j} \cdot \vec{\mu}_{bi})(\vec{r}_{bi,b'j} \cdot \vec{\mu}_{b'j}) / r^5_{bj,b'j} \right] \qquad (5)$$

where $\vec{r}_{bj,b'j}$ is the distance vector pointing from bond midpoint bi to bond midpoint b'j. We can again substitute Eqs. (2) and (3) here to obtain an expression with a four-fold summation which clearly shows the four-body character of the interaction energy in Eq. (5). In practice, the interaction terms are computed by first evaluating Eq. (3) and then Eq. (2). Hence our four-body term requires a single loop followed by a double loop, instead of a four-fold loop; this economy is very important. Even so, liquid water simulations incorporating the four-body term are far more demanding than those with only two- and three-body interactions.

The actual parameters used by Clementi and Corongiu in realizing their expression for the three-body interaction term are derived from a fit to ab initio SCF calculations on the water trimer at 173 different geometries; to ensure accuracy, each of these calculations uses a large basis set and corrects for superposition error by means of the

counterpoise technique(29). We have seen that the four-body term we propose is already implicit in the model adopted by Clementi and Corongiu, so we are obliged to carry over their parameters as well. It remains to verify that our model does in fact account for the four-body interaction, and we have carried out several ab initio SCF calculations for the water tetramer at different geometries as a check. These calculations support the use of our model, and also indicate that the three-body terms found in Eq. (5) are helpful in giving a good description of the three-body interaction.

The production calculations described here have a unit cell containing 512 water molecules, and we find that, even with the economies permitted by our model for the three- and four-body interaction terms, our calculations require two orders of magnitude more run time than analogous calculations with two-body interactions only.

Let us now briefly outline our programming solution for the migration of our Monte Carlo codes from sequential to parallel (on the LCAP). The practical implementation of this system requires that the main program (which runs on the host) be able to closely control the subtasks running on the AP's, periodically transferring data to them, starting them, and subsequently gathering up each set of subtask output.

Our Metropolis-Monte Carlo liquid simulation programs can all be broken down according to the following scheme:

(1) initial input and setup,

(2) generate tables for energy evaluation,

(3) evaluate initial energies,

(4) generate trial move coordinates,

(5) generate trial move energy arrays,

(6) accept or reject move,

(7) update coordinates,

(8) update energy arrays,

(9) final results and output.

Items (1),(2) and (9) occur only once for a particular run, but the remaining steps take place inside loops over the desired number of moves. This includes item (3), since the energies should be recomputed from time to time in order to avoid the accumulation of roundoff error. The number of moves in a run is typically very large (many thousands), so virtually all the run time resides in items (3-8). For a single move, handling the trial move coordinates and the calculation to accept or reject the move are not demanding; by far the heaviest computational demands come from the evaluation of the energy differences associated with the move.

From this breakdown it is fairly clear which parts of the program are worthwhile to modify so they run in parallel on the AP's. The initial input and setup and the final results and output, items (1) and (9), run best on the host, while the evaluation of the move energies should be put on the AP's in parallel; this includes not only items (3) and (5), but also the table generation, item (2).

We thus arrive at a plan for the program where input and run setup is handled on the host, after which a copy of the initial coordinates is transmitted to each AP; each AP thereupon generates its own set of tables for future use and then proceeds with the initial energy evaluation. The energy parameters required by the host are then gathered from the AP's and the loops associated with move generation are entered; these loops are under the control of the host. Within them, the generation of the trial move coordinates takes place on the host, which then transmits a copy of these coordinates to each AP to use in its portion of the generation of the trial move energy arrays. The AP's then transmit back to be gathered up by the host the energies associated with the move. On this basis, the host decides whether to accept the move, and transmits this decision to each of the AP's, so they may update their coordinate and energy arrays if necessary. Control then passes back to the host, and the loops are recycled or exited for final results and output.

We have added only a few refinements to this program plan, in order to minimize the time taken up with transmission between host and AP's (28).

The AP's must retain the old trial move coordinates in memory when control is passed back to the host. In addition to these coordinates, we require retention of the tables that were set up at the beginning of the run, all the intermediate energy arrays, and the set of coordinates for all of the molecules. It is an enormous savings in transmission overhead to retain this data in AP memory throughout the

run. Fortunately, this capability is standard with the APEX software normally supplied with the FPS-164: provided a new program image is not loaded, material in the FPS-164 memory is undisturbed unless specifically altered by an APEX software call(30).

It is well worth the trouble to minimize host-AP transmission overhead. With only a two-body interaction, evaluation of a single move (on one AP) takes place on the order of several hundredths of a second, and host-AP move transmission time can be an appreciable fraction of this if one is careless.

The remaining task in migrating our Monte Carlo programs to parallel mode is the actual splitting of the move energy evaluation among the various AP's. We shall not be too concerned with details here, since these naturally vary according to the specifics of a particular program; instead, we have presented elsewhere (28) organizational features that tend to emerge in most of our programs.

We recall that many intermediate arrays are maintained for move energy evaluation. An obvious recipe for migration to parallel execution is to divide the arrays among the various AP's, with each AP responsible only for the evaluation of its particular portion of the array (28).

We note that the strategies we have implemented for handling the arrays not only support parallel use of the AP CPU's, but also the AP memories: Since the arrays are distributed over the different AP's, without much duplication, we have the sum of all the AP memories

available to the run. Thus our parallel strategy not only allows practical execution of runs requiring more CPU time, but also those with larger memory requirements, or both.

VI. Liquid Water M.C. Simulation with up to Four-Body Interactions.

The previous section intended to give some idea of how we actually accomplished migration of our programs to parallel execution mode. We note that the idea of parallel execution actually has a rather long history; we cannot claim much originality in this area. The distinction we do claim is that we have put these ideas to practical test in production calculations of genuine scientific interest. As a test for our scheme in this section, we discuss the Metropolis-Monte Carlo code we are using for liquid simulation, incorporating the three- and four-body interactions described above. This calculation is very heavy in its computational demands, and can hardly be handled with any normal computer resources, so that it fulfills every requirement as a good practical test of the LCAP system. Furthermore, it is of considerable interest, since the most sophisticated calculations performed heretofore(27) yield an enthalpy which is about 1.674 KJ/mole above the experimental value, 41.3 KJ/mole at room temperature and one obvious explanation of this discrepancy is the omission of four-body interactions.

The calculations we are pursuing to investigate this question are not yet to the point where we can present conclusive results: Our current runs include about 1,000,000 moves, which is enough to achieve

equilibriation, and nearly enough to support good statistics. On a preliminary basis, however, we appear to be getting an enthalpy about 3.35 KJ/mole below the results incorporating only the three-body terms. This suggests that the four-body interactions are indeed significant in getting a good liquid water simulation. One notices that our results actually fall below experiment; presumably this is due to our omission of the quantum correction, which is estimated to be of the right sign and magnitude (3 to 4 KJ/mole) to bring our results into better agreement with experiment (31). We shall return to this point in the last section of the paper. Here we wish to comment that this exceptional agreement obtained by using ab initio potentials, could in itself justify our efforts in assembling the LCAP system.

In Figure 8 (top) we report our simulated pair correlation function g(O-O) obtained with two-body (left), two- and three-body (middle) and two-, three- and four-body interaction potentials (right). The full line is for the experimental data. In the same figure we report the simulated X-ray and neutron-beams scattering intensities. The agreement with the experimental data and the "convergence" of the many-body series are most gratifying. Notice, in particular, how the many-body brings more and more structure into the g(O-O) and into the beam scattering intensity. After about ten years of systematic work, the intensity in the split first peak is finally correct. These results point out that now it is reasonable for us to introduce vibrations into our model (of course, ab-initio, and without any empiricism).

We may conclude that LCAP, our experimental system, is in fact already achieving one of its principal goals even at this rather early stage. It is providing us with a means of accomplishing scientific calculations that are essentially out of reach for a more conventional computer configuration.

VII. Parallelism in Molecular Dynamics.

As it is known, the Molecular Dynamics, MD, method was first developed for rigid spheres fluids by Alder and Wainwright (32). This computer modeling method was later extended to more realistic Lennard-Jones potentials for simple liquids by Rahman (33), Levesque and Verlet (34) and others. The MD method, as standarly used, has two major drawbacks, i) accurate potential-energy functions for real systems are not readily available, ii) it requires numerical solution of vast number of coupled nonlinear differential equations which must be solved many times.

The first drawback is being confronted with increasingly accurate potentials obtained from quantum mechanical computations as indicated in the previous section and elsewhere (20). This section attempts to confront the second drawback, by considering our own brand of parallelism, as put forward in the first two sections of this paper.

We recall that molecular dynamics is used to calculate the observed classical trajectory for an assembly of N (usually several hundreds to thousands) interacting molecules in a box of side L. The Hamiltonian