AD-A103 095 CALIFORNIA UNIV SAN DIEGO LA JOLLA DEPT OF CHEMISTRY F/G 7/4
A COMPUTER SIMULATION METHOD FOR THE CALCULATION OF EQUILIBRIUM--ETC(U)
AUG 81 W C SWOPE, H C ANDERSEN, P H BERENS N00014-78-C-0325
UNCLASSIFIED TR-4 NL

1 OF 1
AD A
103095

END
DATE
FILMED
9-81
DTIC

ADA103095

7

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

REPORT DOCUMENTATION PAGE	READ INSTRUCTIONS BEFORE COMPLETING FORM
1. REPORT NUMBER 4 (6	2. GOVT ACCESSION NO. AD-A103095
3. RECIPIENT'S CATALOG NUMBER 9	
4. TITLE (and Subtitle) A COMPUTER SIMULATION METHOD FOR THE CALCULATION OF EQUILIBRIUM CONSTANTS FOR THE FORMATION OF PHYSICAL CLUSTERS OF MOLECULES: APPLICATION TO SMALL WATER CLUSTERS.	5. TYPE OF REPORT & PERIOD COVERED Technical r 4
	6. PERFORMING ORG. REPORT NUMBER 15
7. AUTHOR(s) (10 William C. Swope [redacted] Hans C./Andersen Peter H. Berens [redacted] Kent R. Wilson	8. CONTRACT OR GRANT NUMBER(s) ONR-N00014-78 C-0325
9. PERFORMING ORGANIZATION NAME AND ADDRESS Department of Chemistry University of California, San Diego La Jolla, CA 92093	10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS
11. CONTROLLING OFFICE NAME AND ADDRESS Office of Naval Research Arlington, VA 22217 (11	12. REPORT DATE Aug[redacted]981
	13. NUMBER OF PAGES 66
14. MONITORING AGENCY NAME & ADDRESS (if different from Controlling Office) (12) 76	15. SECURITY CLASS. (of this report) Unclassified
	15a. DECLASSIFICATION DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT (of this Report)

This document has been approved for public release and sale; its distribution is unlimited.

14 TR-4

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

DTIC
ELECTE
AUG 20 1981
S
A

18. SUPPLEMENTARY NOTES

Submitted to the Journal of Chemical Physics for publication

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)

Computer simulation	molecular dynamics
water	clusters
equilibrium constants	statistical mechanics

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

We present a molecular dynamics computer simulation method for calculating equilibrium constants for the formation of physical clusters of molecules. The method is based on Hill's formal theory of physical clusters. In the method, a molecular dynamics calculation is used to calculate the average potential energy of a cluster of molecules as a function of temperature, and the equilibrium constants are calculated from the integral of the energy with respect to reciprocal temperature. The method is illustrated by calculations

DTIC FILE COPY

DD FORM 1 JAN 73 1473 EDITION OF 1 NOV 65 IS OBSOLETE
S/N 0102-LF-014-6601

400410

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

81 8 19 004

Block 20. Abstract (continued from front page)

of the equilibrium constants for the formation of clusters of two to five water molecules that interact with each other by an intermolecular potential devised by Watts. The method is compared with other procedures for calculating the thermodynamic properties of clusters.

OFFICE OF NAVAL RESEARCH

Contract N00014-78 C-0325

TECHNICAL REPORT NO. 4

A COMPUTER SIMULATION METHOD FOR THE CALCULATION OF EQUILIBRIUM CONSTANTS FOR THE FORMATION OF PHYSICAL CLUSTERS OF MOLECULES: APPLICATION TO SMALL WATER CLUSTERS

BY

William C. Swope and Hans C. Andersen
Department of Chemistry
Stanford University
Stanford, CA 94305

AND

Peter H. Berens and Kent R. Wilson
Department of Chemistry
University of California, San Diego
La Jolla, CA 92093

Prepared for Publication
in
The Journal of Chemical Physics

A

I. Introduction

The thermodynamic stability of clusters of small numbers of atoms or molecules can be described quantitatively by the equilibrium constants for the formation of the clusters. For the reaction

$$n\ A = A_n,$$

the equilibrium constant is

$$K_n = (A_n)/(A)^n$$

where parentheses about a species name denotes the number density of the species. A knowledge of these equilibrium constants is important in a number of areas of chemical physics. First, the values of K_2 and K_3 are related to the second and third virial coefficients of the substance. Second, the theory of nucleation of the liquid phase (or the solid phase) in a supersaturated vapor uses these equilibrium constants and their closely related free energies. Third, it has been suggested that the variable continuous absorption of infrared radiation by the atmosphere arises from these clusters.[1] A reliable and practical method for the theoretical calculation of cluster formation equilibrium constants would have consequences for each of these three areas.

The calculation of the equilibrium population distribution of water clusters is of particular current importance because of the controversy surrounding the recent hypothesis[2] that water clusters are much more prevalent in the atmosphere than had been assumed.

Equilibrium constants for reactions such as

$$n\ A + B = BA_n ,$$

where A is water and B is an ion can be measured by mass spectrometric techniques.[3] A theory for the calculation of such constants would enable the experiments to be interpreted quantitatively in terms of the basic interactions between ions and water and between water molecules.

The classical way to evaluate equilibrium constants for clusters of identical molecules treats the cluster as a droplet whose free energy contains bulk and surface contributions.[4,5] Such a theory makes a connection with the macroscopic thermodynamic properties of the material but does not address the relationship between intermolecular interactions and the stability of clusters.

The partition functions for clusters of atoms and molecules can in principle be calculated, leading to a statistical mechanical theory of cluster formation constants that is similar to the usual statistical mechanical theory of equilibrium constants for chemical reactions. Hill's physical cluster theory[6] provides a rigorous formalism for defining the cluster equilibrium constants and relating them to the thermodynamics of a gas.

If one applies to clusters the same approximations that are used in calculating partition functions for molecules in the gas phase, such as rigid rotor and harmonic oscillator approximations, then partition functions and their associated free energies and equilibrium constants can be calculated. Such calculations are most appropriate for clusters if the temperature is low, if the configuration is solid-like, and if there is a unique lowest energy configuration that dominates the thermodynamic properties at low temperatures. (See Ref. 7 for a discussion of the single configuration approximation.)

Computer simulation methods, such as the molecular dynamics and Monte Carlo methods, can be used to eliminate statistical mechanical approximations. Many studies of atomic and molecular clusters have been made using these methods.[8-10] A fundamental problem associated with using simulation methods for the evaluation of cluster properties is that the simulation of a cluster at a particular temperature cannot be used to evaluate the equilibrium constant or cluster free energy at that temperature. The quantity that is most easily calculated is the average energy of a cluster as a function of temperature. Most of the computer studies listed above were in fact calculations of the energy and not the free energy. The energy is related to the temperature derivative of the free energy, and so calculation of the energy for a range of temperatures can lead to free energies by numerical integration, but the results contain an unknown constant of integration that represents the free energy in the state at which the integration is started.

There are several ways around this difficulty. One is to start the integration at a very low temperature state and use the harmonic approximation and the single configuration approximation to evaluate the free energy of that state. A second, due to Lee *et al.*,[9] makes use of the arbitrariness in the definition of a cluster. They developed a way of calculating the derivative of the free energy with respect to the radius of the hypothetical containment sphere used to define the cluster. For infinite containment spheres, the atoms in the cluster are mostly far from each other and analytical methods can be used to calculate the free energy. Then by numerical integration with respect to the sphere radius, they are able to find the absolute free energy of a cluster defined with a finite sphere containment radius. A third method was developed by Mruzik et al.[10] They used a coupling parameter method and evaluated the average potential energy of interaction of one molecule with the n - 1 other molecules in the cluster calculated as if the interaction of the first molecule with the others were multiplied by a coupling parameter between 0 and 1. This average energy is in fact the derivative of the cluster free energy with respect to the coupling parameter. Integration of the average energy with respect to coupling parameter then gives the difference in free energy of clusters of n - 1 molecules and clusters of n molecules (when an additional correction is made for the difference in the size of the containment sphere used in the definition of n molecule clusters and n - 1 molecule clusters).

The first method depends for its accuracy on the validity of the assumptions used in the analytical calculation of the free energy of the low temperature cluster. The second and third methods rely solely on the computer simulation techniques and thus are clearly to be preferred. The cost and difficulty of performing such simulations has decreased dramatically in the last decade, and in the future computer simulation methods will be the most practical and accurate ways of evaluating cluster free energies.

In this paper we present a new method of evaluating the free energies and equilibrium constants of clusters of atoms and molecules. The method is based on evaluating the energy of a cluster as a function of temperature using a variation of the molecular dynamics computer simulation technique and then integrating with regard to temperature. The state of infinite temperature is used to evaluate the constant of integration analytically without any approximations. Provided that the intermolecular or interatomic potential satisfies certain conditions, namely that the potentials be finite at all nonzero distances and diverge no more strongly than r^{-3} at short distances, the integrand in the temperature integration does not diverge as the temperature approaches infinity, and thus the integration to the infinite temperature state can easily be performed. The method is applied to the calculation of the equilibrium constants for the formation of clusters containing two to five water molecules.

In Section II, we discuss the theory of the method, using Hill's physical cluster theory as a starting point. In Section III we discuss the theory of the computer simulation algorithm we use. Section IV discusses the intermolecular and intramolecular potentials for water that we used. Section V discusses the methods we used to analyze the results and verify their correctness. Section VI contains the results for clusters of two to five water molecules. Section VII contains a brief discussion of the method.

II. Formal Theory of Equilibrium Constants for Cluster Formation

A. Hill's general theory

Hill's physical cluster theory[6] provides the statistical mechanical basis for calculating equilibrium constants for cluster formation. His theory provides no unique choice of definitions for clusters. Instead it provides a set of easily satisfied conditions that the cluster definitions must satisfy, and for any such set of conditions it gives precise expressions for the cluster equilibrium constants. We will apply his general formulation to the problem of water clusters.

We use classical mechanics to describe the nuclear motions in the water molecule. Let $\underset{\sim}{x}_i$ denote a complete set of position and momentum variables for a single water molecule. The canonical partition functions for N water molecules in volume V is

$$Q_N(\beta,V) = (N!2^N h^{9N})^{-1} \int_V d\underset{\sim}{x}^N \exp[-\beta H(\underset{\sim}{x}^N)]. \qquad (2.1)$$

The factors of two are the symmetry numbers for the molecules. To apply Hill's theory, one must find a way to decompose Q_N so that:

$$Q_N = \Sigma'_{\underset{\sim}{\underline{N}}} Q_{\underset{\sim}{\underline{N}}}. \qquad (2.2)$$

Here $\underset{\sim}{N}$ is a set of nonnegative integers N_1, N_2, ..., N_N. N_i refers to the number of clusters of i molecules. These numbers must satisfy the condition

$$\Sigma_{s=1}^{N} sN_s = N. \tag{2.3}$$

The prime in the sum in Eq. (2.2) denotes that the sum is over all sets of nonnegative values of $\underset{\sim}{N}$ that satisfy condition (2.3). There are many ways to make such a decomposition. The physical meaning of $Q_{\underset{\sim}{N}}$ is that it represents the canonical partition function of a system that contains N molecules that exist as N_1 monomers, N_2 dimers, ... N_s clusters of s molecules, ..., etc.

The partition functions for systems that contain only one cluster play an important role in the theory. Departing from Hill's notation, let us define

$$\begin{aligned} Q^{(1)} &= Q_{100\ldots}(\beta,V) \\ Q^{(2)} &= Q_{010\ldots}(\beta,V) \\ Q^{(n)} &= Q_{00\ldots1\ldots}(\beta,V), \qquad n \geq 3 \end{aligned} \tag{2.4}$$

$Q^{(n)}$ is the canonical partition function for one cluster of n molecules. According to most reasonable definitions of clusters, n molecules will be in a cluster only if they are separated by microscopic distances that are independent of the size of the volume V. This leads to $Q^{(n)}$ being of order V for large V.

Let ρ_n denote the average number density of clusters of n molecules. Hill's theory provides a formula for ρ_n that is an expansion in powers of the density. To lowest order in density, his result is

$$\rho_n/(\rho_1)^n = K_n(\beta)$$

where

$$K_n(\beta) = (Q^{(n)}/V)/(Q^{(1)}/V)^n. \tag{2.5}$$

The important assumptions made in deriving this result are that the partition function can be decomposed according to Eq. (2.2) for all N and that $Q^{(n)}/V$ approaches a limit as $V \to \infty$.

B. One realization of Hill's theory

One way of accomplishing the decomposition in Eq. (2.2) is to decompose phase space for each value of N into disjoint regions such that in each region the states have a well defined set of values of the cluster numbers $\underset{\sim}{N}$. In particular, if we can define the set of functions $C_{\underset{\sim}{N}}(\underset{\sim}{x}^N)$ for all N and $\underset{\sim}{N}$ that satisfy Eq. (2.3) such that

$$\begin{aligned} C_{\underset{\sim}{N}}(\underset{\sim}{x}^N) &= 1 \quad \text{if the phase point } \underset{\sim}{x}^N \text{ has cluster} \\ & \text{numbers } \underset{\sim}{N} \\ &= 0 \quad \text{otherwise} \end{aligned} \tag{2.6}$$

and such that

$$\Sigma'_{\underset{\sim}{N}}\, C_{\underset{\sim}{N}}(\underset{\sim}{x}^N) = 1, \tag{2.7}$$

for all N and $\underset{\sim}{x}^N$ then Eq. (2.2) is satisfied if we let

$$Q_{\underset{\sim}{N}} = (N!2^N h^{9N})^{-1} \int_V d\underset{\sim}{x}^N C_{\underset{\sim}{N}}(\underset{\sim}{x}^N)\exp[-\beta H(\underset{\sim}{x}^N)]. \tag{2.8}$$

To define these functions it is necessary to have a procedure for assigning a unique set of cluster numbers $\underset{\sim}{N}$ to every phase point $\underset{\sim}{x}^N$ for an N molecule system.

Let us write

$$c^{(1)}(\underset{\sim}{x}^1) = 1$$

$$c^{(2)}(\underset{\sim}{x}^2) = C_{01}(\underset{\sim}{x}^2) \tag{2.9}$$

$$c^{(n)}(\underset{\sim}{x}^n) = C_{00\ldots1}(\underset{\sim}{x}^n), \quad n \geq 3. \tag{2.10}$$

The procedure we use for defining the functions in (2.6) is the following. We imagine that a necessary (but not sufficient) condition for n molecules to form a cluster of n molecules is that the center of mass of each of the molecules be less than a distance R_n from their mutual center of mass. We decide on the set of values of R_n for all n. We use the following algorithm for deciding, for a phase point $\underset{\sim}{x}^N$, how many clusters of each size are present.

1. Let n = N.
2. If n = 1, let the molecule be regarded as a cluster of 1 and stop.
3. If n > 1, see if there are any sets of n molecules that are all within a distance R_n of their mutual center of mass.
4. If there is no such set, let n = n - 1 and go to 2.
5. If there is one such set, call these molecules a cluster of n, remove them from further consideration, let n = n - 1 and go to 2.
6. If there is more than one such set, find the set of n that is most compact, in the sense that the largest distance of a molecule from the mutual center of mass is smaller for that set than for any other set. Call this most compact set a cluster of n, remove them from further consideration, keep n unchanged and go to 3.

Except for a set of phase points of measure zero (those points where there are groups of the same n that are equally compact), this algorithm gives a precise way of defining how many clusters of each type exist for any phase point, and thus it provides a way of partitioning N molecule phase space in a way needed for the validity of Eqns. (2.6) and (2.7).

It follows from this procedure that

$$C^{(n)}(\underset{\sim}{x}^n) = 1 \quad \text{if the center of mass of each molecule is less than } R_n \text{ from the center of mass of the set of n molecules}$$

$$= 0 \quad \text{otherwise.} \qquad (2.11)$$

Combining Eqs. (2.4), (2.5), (2.8), (2.9), (2.10), and (2.11), we obtain

$$K_n(\beta) = I_n(\beta)/[I_1(\beta)]^n \qquad (2.12)$$

where

$$I_n(\beta) = (n!)^{-1} \int_V d\underset{\sim}{x}^n \, C^{(n)}(\underset{\sim}{x}^n)\exp[-\beta H(\underset{\sim}{x}^n)]. \qquad (2.13)$$

C. Temperature dependence of the equilibrium constants

The integrals $I_n(\beta)$ cannot be evaluated directly by computer simulation, but their logarithmic derivative with respect to β can be evaluated. From Eq. (2.13) it follows that

$$d \ln I_n/d\beta = -\langle H(\underset{\sim}{x}^n)\rangle_N$$

where the average on the right side is defined by

$$\langle f(\underset{\sim}{x}^n)\rangle_n = \int_V d\underset{\sim}{x}^n f(\underset{\sim}{x}^n)\exp[-\beta H(\underset{\sim}{x}^n)]C^{(n)}(\underset{\sim}{x}^n)$$

$$\times \left[\int_V d\underset{\sim}{x}^n \exp[-\beta H(\underset{\sim}{x}^n)]C^{(n)}(\underset{\sim}{x}^n)\right]^{-1}$$

It follows from Eq. (2.12) that

$$d \ln K_n(\beta)/d\beta = -\langle H(\underset{\sim}{x}^n)\rangle_n + n\langle H(\underset{\sim}{x}^1)\rangle_1. \qquad (2.14)$$

The Hamiltonian H is a sum of kinetic energy T and potential energy U,

$$H(\underset{\sim}{x}^n) = T(\underset{\sim}{x}^n) + U(\underset{\sim}{x}^n)$$

where U includes both intermolecular and intramolecular potential energy. The contributions of T to the two terms on the right of (2.14) cancel each other. If we define

$$U_n(\beta) = \langle U(\underset{\sim}{x}^n)\rangle_n \qquad (2.15)$$

we have the following result for the temperature derivative of $\ln K_n$

$$d \ln K_n(\beta)/d\beta = -U_n(\beta) + nU_1(\beta).$$

Hence,

$$\ln K_n(\beta) = \ln K_n(0) - \int_0^\beta d\beta'[U_n(\beta') - nU_1(\beta')]. \qquad (2.16)$$

The first term on the right can be evaluated analytically, and the integrand on the right can be evaluated by performing molecular dynamics simulations. This is the basic expression we use for the evaluation of the cluster formation equilibrium constants.

To evaluate the high temperature limiting behavior of ln K_n, we note that at high temperature the Boltzmann factor can be replaced by unity and we have

$$\begin{aligned} K_n(0) &= I_n(0)/[I_1(0)]^n \\ &= (n!V)^{-1} \int_V d\underset{\sim}{x}^n C^{(n)}(\underset{\sim}{x}^n)/[V^{-1} \int_V d\underset{\sim}{x}_1]^n. \end{aligned}$$

For each molecule we convert the variables of integration to the momentum variables, a center of mass variable, and a set of internal coordinates. The $C^{(n)}$ function depends only on the center of mass variables. The momentum and internal coordinate integrations factor out of the numerator and denominator and cancel each other, leaving

$$K_n(0) = (n!V)^{-1} \int_V d\underset{\sim}{r}^n C^{(n)}(\underset{\sim}{r}^n)$$

where the $\underset{\sim}{r}^n$ are the center of mass variables. This integral has been evaluated by Lee <u>et al</u>.[9] The result is

$$K_n(0) = a(n)n^{3/2}[4\pi R_n^3]^{n-1}/n! \qquad (2.17)$$

where a(n) is a function of n whose values range from 2.437 to 2.974. For n = 2 to 5, the values are 2.828, 2.436, 2.590, and 2.661, respectively. (The values reported in Table I of Reference 9 are incorrect.)

A potential complicating feature of the beta integration in Eq. (2.16) is that for some intermolecular potentials the integrand diverges as $\beta \to 0$. (The divergence is integrable, however.) If the integrand is very large, it is difficult to evaluate accurately by computer simulations, since in general the statistical noise in the answer will also be large. This divergence exists for potentials that increase as $r^{-(3+n)}$, as the interatomic or intermolecular distance r decreases to zero, for $n \geq 0$. For the Lennard-Jones potential, for example, the integrand would diverge. The potentials we use for water diverge only as 1/r for small r, and the integrand approaches a finite value as $\beta \to 0$.

D. Choice of the R_n and the relationship between cluster formation constants and virial coefficients

According to the definition of cluster that we use, a necessary condition for n molecules being a cluster of n molecules is that they all lie within a sphere of radius R_n centered at their mutual center of mass. We will refer to this sphere as the "cluster containment sphere." The choice of the containment sphere radius for each value of n is arbitrary; for any set of choices, Hill's formal theory provides expressions for the equilibrium constants for cluster formation. When the various R_n satisfy certain conditions, it can be shown that the equilibrium constants for dimer and trimer formation are simply related to the second and third virial coefficients. Here we discuss our choice of the R_n and the relationship of equilibrium constants and virial coefficients.

For most of our calculations, we chose R_n using the criterion of Lee _et al_.[9] i.e. we required that the volume of the sphere for clusters of n molecules be five times the volume occupied by n molecules in the liquid at standard temperature and pressure. Thus

$$R_n = 3.292\ n^{1/3}\ \text{Å}.$$

The factor of five is admittedly arbitrary, but Lee et al. demonstrated that the free energy of clusters of atoms is independent of the volume of the containment sphere for low enough temperatures. For clusters of three water molecules, we have also performed calculations with a different choice of R_3 and have verified that the equilibrium constants are insensitive to the volume of the containment sphere at low temperatures.

The equilibrium constant for dimer formation can be related to the second virial coefficient for the gas under certain conditions. When two molecules are in a cluster, they are less than a distance $2R_2$ apart. Thus the dimer constant is not affected by the form of the intermolecular potential for separations larger than this. If the potential is zero at such large distances, it is straightforward to show, using Eqns. (2.1) and (2.8) that

$$Q_2 = Q^{(2)} + \left[V - \frac{4\pi}{3}(2R_2)^3\right]Q_1^2/2V.$$

Using the standard relationship between the second virial coefficient and the partition functions for one and two molecules[11]

$$B_2 = -V\left[(Q_2/Q_1^2) - \frac{1}{2}\right]$$

we find

$$B_2 = 16\pi R_2^3/3 - K_2. \tag{2.18}$$

Thus, if K_2 is calculated for a particular choice of R_2 and a particular intermolecular potential, this equation shows that the result can be used to obtain the second virial coefficient for a gas whose intermolecular potential is zero for separations greater than $2R_2$ and equal to the potential used in the calculation for separations less than $2R_2$.

A similarly simple relationship for the third virial coefficient cannot be obtained for an arbitrary choice of R_3. However, if

$$R_3 = 2R_2 \tag{2.19}$$

and if the range of the potential is equal to or less than $2R_2$, we can obtain a simple result. (A simple result can also be obtained if $R_3 > 2R_2$, but we shall not discuss this case.) When these restrictions are satisfied, if three molecules are located so that there are two (or three) intermolecular distances less than $2R_2$, the three molecules must form a trimer. The points in three molecule configuration space can then be divided into the following sets: A. points in which the three molecules form a trimer; B. points in which the three molecules do not form a trimer but two molecules form a dimer; and C. points in which the three molecules do not form any dimer and no intermolecular distance is less than $2R_2$. In set C, all the intermolecular interactions are zero. In set B, the molecule that is not in the dimer must be separated from the center of mass of the other two by at least $(3/2)R_3$, which is equal to $3R_2$, to avoid formation of a trimer. Since the other two are separated by a distance of less than $2R_2$, this latter condition also guarantees that the third molecule is further than R_2 from each of the others and hence does not interact with the other two. It is then straightforward to show that

$$Q_3 = Q^{(3)} + [V - (4\pi/3)(3R_2)^3]Q_1Q^{(2)}/V + \alpha Q_1^3$$

where the three terms on the right correspond to the regions A, B, and C, respectively. The quantity α is an integral that depends on V and on R_2 but is independent of β. Using the standard relationship between the third virial coefficient and the two and three molecule partition functions,[11]

$$B_3 = -V^2[2(Q_3/Q_1^3) - 4(Q_2^2/Q_1^4) + 2(Q_2/Q_1^2) - \tfrac{1}{3}],$$

and Eq. (2.18), we find

$$B_3 = -2K_3' + 4B_2^2 - 72\pi R_2^3B_2 + \left\{\left(\frac{1}{3} - 2\alpha\right)V^2 - 32\pi R_2^3V/3 + 384\pi^2R_2^6\right\}$$

where the prime in the trimer equilibrium constant denotes that it is defined using the nonstandard trimer containment sphere radius given in Eq. (2.19). The integral α can be evaluated tediously. A simpler procedure is to use the fact that for $\beta = 0$ the virial coefficients are zero and the equilibrium constants are given by Eq. (2.17). Then we obtain

$$B_3 = -2K_3' + 4B_2^2 - 72\pi R_2^3 B_2 + 480\pi^2 R_2^6. \qquad (2.20)$$

This relationship between the second and third virial coefficient and the dimer and trimer equilibrium constants is valid when Eq. (2.19) holds and when the range of the intermolecular potential is equal to or less than $2R_2$.

III. Theory of the Simulation Algorithm

A. Description and justification of the algorithm

To evaluate $U_n(\beta)$, defined in Eq. (2.15), we must evaluate the average of the n molecule potential energy over a distribution of n molecule phase points, where the distribution function in n molecule phase space is proportional to

$$\exp[-\beta H(\underset{\sim}{x}^n)]C^{(n)}(\underset{\sim}{x}^n). \tag{3.1}$$

The function $C^{(n)}$ can be regarded as the Boltzmann factor for a hard potential that keeps the n molecules within the region of space in which they form an n molecule cluster. If we define

$$\begin{aligned} W^{(n)}(\underset{\sim}{x}^n) &= 0 \qquad \text{if } C^{(n)}(\underset{\sim}{x}^n) = 1 \\ &= \infty \qquad \text{if } C^{(n)}(\underset{\sim}{x}^n) = 0, \end{aligned}$$

then the distribution function in (3.1) can be written as

$$\exp[-\beta\{H(\underset{\sim}{x}^n) + W^{(n)}(\underset{\sim}{x}^n)\}], \tag{3.2}$$

which is a Boltzmann factor for the combination of the molecular Hamiltonian and the wall potential. From (2.11) it can be seen that the wall potential is a function of the center of masses of the n molecules and is independent of their internal coordinates. Since the wall potential depends only on the relative positions of the centers of mass, it is a momentum conserving interaction. When the molecules are within the cluster containment sphere of radius R_n centered at their mutual center of mass, the potential is zero, but it becomes nonzero and infinite when the center of mass of any of the molecules touches the surface of the cluster containment sphere. When this happens, the relative velocity of the center of mass of that molecule with respect to the mutual center of mass of the remaining $n - 1$ molecules is reversed in direction and unchanged in magnitude. In effect, the molecule that hits the surface of the containment sphere is specularly reflected from the surface and the velocity of each of the other molecules is changed by the amount required to conserve total momentum. Each of the other molecules suffers the same change in center of mass velocity, and the positions, internal coordinates, and internal velocities of all the molecules are unaffected. The collisions with the wall conserve kinetic energy.

To generate a set of phase points for an n molecule cluster that are distributed according to the Boltzmann factor in (3.2), we use the following procedure, which is a combination of the molecular dynamics and Monte Carlo simulation methods. We calculate a molecular dynamics trajectory for a system of n molecules subject to the Hamiltonian H and to the wall potential W. At regular intervals of time, we replace the momenta of each of the atoms by a momentum chosen at random from the Boltzmann distribution appropriate for the mass of the atom and for the temperature of interest. This latter procedure can be regarded as subjecting the atoms to stochastic collisions with a heat bath that has the temperature of interest. Using the theory of Markov processes, it is possible to show that this combination of Hamiltonian dynamics and stochastic collisions is a Markov process and that a trajectory for the process has the property that the time average of a mechanical property over a trajectory is equal to the statistical average of the same property over a distribution function proportional to (3.2), provided the trajectory is long enough. (The details of the proof are similar to those of a proof in Reference 12 and they will be omitted here.)

Three important assumptions must be made to apply the theorem. First it must be assumed that under the motion generated by the Markov process a system can in principle get from any phase point to any other phase point in a finite amount of time. (In other words, the motion must be ergodic.) The second assumption is that the trajectory used in the calculation is long enough so that the trajectory actually samples all the important parts of phase space. The third assumption is that the trajectory is long enough that the statistical noise in the calculation is averaged away.

For dense liquids it is commonly believed that the first assumption is correct, and it is very reasonable to expect that it is correct for this type of simulation of a cluster. The second assumption can be tested by starting two trajectories in very different parts of phase space and seeing if they both give the same time average. The statistical error in a calculation can be estimated on the basis of physical reasoning or from the results of the simulation. In the next part of this section we discuss how the statistical error can be estimated.

B. Estimation of statistical error

Let A represent a dynamical variable. Ensemble averages of any quantity over the correct distribution will be denoted by angular brackets. Time averages over a particular trajectory will be denoted by overbars. The time average of A is

$$\overline{A} = T^{-1} \int_0^T dt\, A(t),$$

where A(t) is the value of A at time t on the particular trajectory. The theorem mentioned above states that under the appropriate conditions

$$\bar{A} \to \langle A \rangle \text{ as } t \to \infty.$$

Let

$$\Delta(t) = A(t) - \langle A \rangle.$$

Then the error made in assuming $\bar{A}$ and $\langle A \rangle$ are equal is

$$\bar{A} - \langle A \rangle = T^{-1} \int_0^T dt\, \Delta(t).$$

The square of the error is

$$(\bar{A} - \langle A \rangle)^2 = T^{-2} \int_0^T dt_1 \int_0^T dt_2\, \Delta(t_1)\Delta(t_2).$$

We can estimate the square of the error by calculating the ensemble average of both sides of this equation; i.e. we calculate the average over an ensemble of trajectories starting at all points in phase space and appropriately weighted by the probability distribution of the initial state. We obtain

$$\langle (\bar{A} - \langle A \rangle)^2 \rangle = T^{-2} \int_0^T dt_1 \int_0^T dt_2 \langle \Delta(t_1)\Delta(t_2) \rangle.$$

The average in the integrand is the ensemble averaged two time correlation function of the fluctuation of A from its ensemble average. For the ensemble, such a correlation function depends only on the time interval $|t_1 - t_2|$, and we expect the correlation function to vanish as $|t_1 - t_2| \to \infty$. Thus we can write

$$\langle\Delta(t_1)\Delta(t_2)\rangle = \langle\Delta^2\rangle g(t_1 - t_2)$$

where

$$g(0) = 1$$

$$g(t) \to 0 \text{ as } t \to \pm\infty.$$

We obtain

$$\langle(\bar{A} - \langle A\rangle)^2\rangle = 2\langle\Delta^2\rangle\tau_c/T \tag{3.3}$$

where

$$\tau_c = \int_0^\infty dt\, g(t)$$

In obtaining Eq. (3.3) we have assumed that T is large compared with the times for which g(t) is nonzero. The quantity τ_c can be regarded as the correlation time for fluctuation of A about its ensemble average. (If g(t) were an exponential function of t, τ_c would be the time constant for the exponential.) Let us define

$$N_c = T/\tau_c.$$

N_c is the number of correlation times contained within the duration, T, of the trajectory. Hence

$$\langle (\bar{A} - \langle A \rangle)^2 \rangle^{1/2} = \langle \Delta^2 \rangle^{1/2} (2/N_c)^{1/2}. \tag{3.4}$$

This is an estimate of the root mean square error made by assuming that $\bar{A}$ obtained from one trajectory is equal to $\langle A \rangle$.

To estimate the right side of this equation it is reasonable to use the results of the one trajectory and assume

$$\langle \Delta^2 \rangle = \langle (A - \langle A \rangle)^2 \rangle \approx \overline{(A - \bar{A})^2} \tag{3.5}$$

and

$$\begin{aligned} \langle \Delta^2 \rangle g(t) &= \langle \Delta(0) \Delta(t) \rangle \\ &\approx (T - t)^{-1} \int_0^{T-t} dt_1 [A(t_1) - \bar{A}][A(t_1 + t) - \bar{A}] \end{aligned} \tag{3.6}$$

i.e. to assume that the trajectory averaged correlation function of fluctuations of A from its trajectory average is approximately equal to the ensemble average.

The time average $\overline{(A - \bar{A})^2}$ is easy to evaluate by performing the indicated integration of fluctuations of A. (Alternatively, when A is a potential energy, as it is in our case, the ensemble average squared fluctuation in the potential energy is simply related to the temperature derivative of the average potential energy, which can be estimated from numerical differentiation of the averages with regard to temperature.) The correlation function on the right side of the last part of (3.6) is more difficult to evaluate, but it can be obtained from the results of the simulation. Alternatively one could use physical reasoning to estimate τ_c.

The quantity τ_c can be interpreted as the time over which fluctuations in A are correlated rather than uncorrelated. If A represents intermolecular potential energy, as in the problem at hand, then there are some parts of phase space where A is large and negative (namely in hydrogen bonding configurations), some parts where A is small (namely when the molecules are far apart), and other parts where A is large and positive. At any temperature, a certain range of energies is likely to be important in the equilibrium distribution. Then at any temperature, τ_c should be approximately the time scale for moving from the low energy to the high energy parts of phase space that are important at that temperature.

C. Choice of interval between stochastic collisions

An important parameter in the calculation is the time interval between the stochastic collisions suffered by the molecules. For infinitely long trajectories, the value of the time interval is irrelevant: correct average potential energies will be obtained for any choice of interval. For trajectories of finite length, the value of the interval has an important effect on the statistical error of the calculation, and thus it must be chosen carefully.

The stochastic collisions equilibrate the kinetic energy and the total energy of the water molecules to the values typical of the temperature of interest. Taking only this into account, one might be tempted to make the collision frequency very large to insure rapid equilibration. However, this would impede the motion of the molecules in configuration space. In effect, all motion would be diffusion limited. As discussed above, it is important that the molecules be able to visit all the appropriate high energy and low energy parts of configuration space in order to obtain good statistical averages from the trajectory. If the stochastic collisions are too frequent, τ_c will be too large, thereby making the statistical error large. If we can estimate τ_c in the absence of stochastic collisions, it is worthwhile to choose the time interval between stochastic collisions to be no shorter than this estimate of τ_c. Stochastic collisions that are this infrequent will not inhibit the motion of the system from high to low energy parts of configuration space. Thus, in our calculations on a particular system at a particular temperature, we make an estimate (on the basis of physical reasoning or previous calculations) of what τ_c will be for that system and we choose the interval between stochastic collisions to be that estimate.

IV. The Molecular Interactions

For this simulation of water clusters the intramolecular potential was assumed to be a sum of atom-atom potentials. For the two intramolecular atom-atom potentials, we used harmonic potentials with equilibrium bond lengths and spring constants which correspond to the minima and curvatures of the improved central force potentials of Stillinger and Rahman[13]:

$$V_{OH}(r) = \frac{1}{2}(1147.6)(r - 0.9584\ \text{Å})^2 \text{ kcal/mole}$$

$$V_{HH}(r) = \frac{1}{2}(257.3)(r - 1.5151\ \text{Å})^2 \text{ kcal/mole.}$$

The minima of these two potentials are those which reproduce the correct geometry of an isolated water molecule. The curvatures were selected to reproduce the asymmetric stretch frequency in D_2O and to equalize the fractional errors in the frequencies of the other two vibrational modes of the same molecule.

To define the intermolecular potential, let $\underset{\sim}{r}_{ij}$ represent the position of the ith atom on the jth molecule. The first atom on each molecule is the oxygen atom and the others are hydrogen atoms. We define a set of intermolecular atom-atom interactions $V_{OO}(r)$, $V_{OH}(r)$, and $V_{HH}(r)$, and a switching function S(r). Then the intermolecular interactions are of the form:

$$u(\underset{\sim}{r}_{11},\underset{\sim}{r}_{21},\underset{\sim}{r}_{31},\underset{\sim}{r}_{12},\underset{\sim}{r}_{22},\underset{\sim}{r}_{32})$$

$$= S(|\underset{\sim}{r}_{11} - \underset{\sim}{r}_{12}|) \sum_{i,j=1}^{3} V_{ij}(|\underset{\sim}{r}_{i1} - \underset{\sim}{r}_{j2}|),$$

where the subscripts on the V_{ij} functions should be OO, OH, or HH, as is appropriate to the nature of the atoms i and j. For these atom-atom intermolecular interactions, we chose to use those of Watts[14] which were optimized to reproduce the temperature dependence of the second virial coefficient of water in the temperature range for which experimental data was available. The switch function, S(r), is designed to switch the Watts potential off smoothly as a pair of molecules separates. The switch function has been used before[15] with excellent results. The form of the S(r) is:

$$\begin{aligned} S(r) &= 1 && \text{for } 0 < r < r_L \\ &= s(r^2) && \text{for } r_L < r < r_U \\ &= 0 && \text{for } r_U < r. \end{aligned}$$

where $s(r^2)$ is a fifth-order polynomial function in r^2 which is designed to have zero first and second derivatives at r_L^2 and r_U^2, and which lets S(r) be continuous. (We chose 6.0 Å and 5.5 Å for r_U and r_L, respectively.) The use of switch functions is preferable to a simple truncation (or truncation and shift) of the atom-atom potentials especially when the values of these potentials or the forces are large where the truncation is to occur. This is the case for the water-water potentials used here because the three atom-atom components are dominated by a long-range Coulombic term at large distances even though their sum, the full Watts potential, is dipolar-dipolar at these distances. The use of a continuous switch function in the potential is also preferable to truncation of the force for intermolecular distances beyond a certain cutoff distance. Truncation of the force leads to a force that is not the derivative of a potential and hence leads to lack of conservation of energy and a secular heating of the sample.

The choice of a purely harmonic intramolecular potential was motivated by a desire for simplicity and by the expectation that cluster equilibrium constants are more sensitive to intermolecular interactions than to intramolecular interactions. The Watts intermolecular potential was chosen for these cluster studies because it has the correct second virial coefficient. The switching off of the Watts potential at long distances is expected to have a small effect on the properties of the small clusters we are concerned with here.

V. Computational Procedure, Error Analysis and Consistency Checks

Clusters of one to five molecules were simulated by molecular dynamics using the potentials described in the previous section. For each cluster size, the system was equilibrated to several different temperatures through the administration of stochastic collisions on every atom at regular intervals. The potential energy, U_n, of each n molecule cluster was calculated at each molecular dynamics time step and averaged over the length of a run. Finally, $U_n - nU_1$ was plotted and integrated as a function of β $(= 1/k_BT)$ to yield $\ln K_n(\beta)$. This section describes some aspects of the computational procedure, states the parameters of the calculation, and discusses several consistency checks used to verify the accuracy of the calculation.

To choose the time intervals between stochastic collisions, we estimated the correlation time for potential energy fluctuations using physical reasoning and choose the interval to be that estimate. For monomers, the correlation time for potential energy fluctuations was estimated to be the period of the slowest vibrational mode, the bending vibration, 2×10^{-14} sec. For clusters, we considered two possibilities for the time scale for energy fluctuations. The first is the time for a molecule to cross the containment sphere. At high temperatures, the molecules are not bound to each other and the time scale for potential energy fluctuations is approximately the time between collisions of a molecule, which is approximately the sphere traversal time. The second is the intermolecular vibrational period for a cluster, which we estimated as 0.09 psec. This is an appropriate estimate of the correlation time at low temperatures. We actually used the geometric mean of these two estimates as the interval between stochastic collisions for our simulations of clusters at all temperatures. Typical values of the interval were 0.3 to 0.6 psec.

The time duration of each simulation was chosen on the basis of the amount of statistical error that would be tolerated. To estimate the statistical error as a function of duration, Eq. (3.4) was used. The correlation time needed for the right side of Eq. (3.4) was estimated to be equal to the geometric mean of the sphere traversal time and the cluster breathing period. The mean square fluctuation of the potential energy was estimated using Eq. (5.1) below and an estimate of $(27n - 6)k_B/2$ as the heat capacity of an n-mer. (Note that the statistical error estimated in this way was done in advance of the simulation and the estimate was used only for deciding on the length of the simulation. The statistical errors reported in the results section below were obtained from correlation times and mean square energy fluctuations obtained from the simulation data.) The durations varied from five nanoseconds (for the dimer at 2000° K) to 400 picoseconds (for the monomer at 50° K).

The statistical errors in the resulting potential energy averages were calculated by computing the mean square fluctuation in the potential energy for each simulation and the correlation function of the fluctuations for some of the simulations and using Eqs. (3.4)-(3.6). Typical values of the error were 0.02, 0.2, 0.2, 0.3, and 0.4 kcal/mole for $U_1(\beta)\ldots U_5(\beta)$, respectively.

The choice of time step for the numerical integration of the equations of motion is crucial for obtaining accurate results. The step must be small enough that the integration algorithm gives an accurate description of the trajectory, at least for the purpose of calculating potential energy averages. Too small a value is undesirable, because for a given amount of computer time the number of correlation times in the trajectory is inversely proportional to the time step. We chose the time step in the following way. We chose the durations of the runs so that the statistical error in the potential energy averages would be about 0.1 kcal/mole. We wanted the time step to be small enough that the trajectory calculated for times of the order of the correlation time would give average potential energies that are in error by no more than 0.05 kcal/mole. We performed various test calculations starting at the same initial mechanical state and integrating the equations of motion for the same length of time using different time intervals, and we determined what the time step should be to give the desired accuracy. As the result of these tests, we used a timestep of 0.2×10^{-15} sec in all calculations except those at 4000° K, which used 0.1×10^{-15} sec.

The equations of motion were integrated using a version of the Verlet algorithm discussed in the appendix.

U_1, the monomer energy, as a function of T, was fit very well by a straight line of slope $\frac{3}{2} k_B$. This is not surprising since the atoms within a molecule interact by purely harmonic potentials. There was substantial deviation from the straight line behavior only at the one high temperature of 4000° K. (This may be due to rotational-vibrational coupling.) For all lower temperatures, the linear fit was used to obtain nU_1.

For each cluster size the data consisted of $U_n - nU_1$ at 12 to 15 values of the temperature (see Figure 1). Since it is the area under this curve as a function of $\beta = 1/k_BT$ that gives $d \ln K_n/d\beta$, the temperature values selected were more or less evenly spaced in β and ranged from $\beta = 0.25$ $(\mathrm{kcal/mole})^{-1}$ (~ 2000° K) to $\beta = 3.0$ $(\mathrm{kcal/mole})^{-1}$ (~ 170° K).

For each cluster size, the function $U_n - nU_1$ was fitted to an analytic form to facilitate the integration of $\ln K_n(\beta)$. The fitting procedure called for a simple analytic form, so that the integration would be easy to perform, as well as some sort of smoothing, to decrease the effect of statistical noise.

The fitting procedure chosen used a cubic spline. For the m values of x_i and y_i, the fit is determined by the $m - 1$ sets of the four coefficients that define the cubic polynomial in each interval. The cubic spine fit is further required to be continuous and have continuous first and second derivatives at each x_i. Most cubic spline fits further require the fit to pass through all the data points. To allow for smoothing, this last condition was relaxed. Given a set of error estimates δy_i for each data point, the procedure that was used[16] produced the unique cubic spline, $S(x)$, such that the quantity

$$\int_{x_i}^{x_m} dx[S''(x)]^2$$

was minimized subject to the constraint that

$$\sum_{i=1}^{m} \left[\frac{S(x_i) - y_i}{\delta y_i}\right]^2 \leq P.$$

If P = 0, the cubic spline goes through the data points. For very large P, the spline approaches a linear least squares fit.

The fitted function S(x) is unique for each choice of P. Following earlier work,[17] P is determined to be the value which allows the fit to go "between" the data points rather than consistently above or below them. This idea is made more precise by defining the residual

$$R_i = y_i - S(x_i)$$

and a correlation

$$Q = \sum_{i=1}^{m-1} R_i R_{i+1}$$

Q is a measure of the amount of smoothing. For large values of P, adjacent data points tend to fall on the same side of the fit and Q is positive. For small values of P, Q is invariable negative. Furthermore, Q is usually a well-behaved function of P. The value of P that yields Q = 0 is taken as the value that yields a fit that is smoothed enough to remove statistical noise without removing the features of interest.

The values used for δy_i were the statistical error in the U_n. Note that an acceptable fit should also satisfy $R_i \lesssim \delta y_i$ for all i. This condition was usually satisfied.

Once the data was plotted and the fit was determined, it was quite easy to integrate the d ln $K_n/d\beta$ curve from $\beta = 0$ to any other value. As explained in Section II, there is no divergence in $U_n - nU_1$ near $\beta = 0$.

The remainder of this section will discuss several consistency checks that were made by comparing different calculations in order to detect various types of systematic and random errors.

One kind of error possible in this procedure arises from there being two or more low-energy regions of phase space with different values of the potential energy and separated by a large potential barrier. At low enough temperatures, the rate of barrier crossing will vanish and average values of U_n will be different depending upon the side of the barrier on which the system remains trapped. (One possible example might be the existence of two stable n-molecule clusters, one in an n-membered ring and one in an n-1 membered ring with a branch.) One way to test for the existence of such a problem is to do two or more simulations from different starting configurations and notice if the results obtained are consistent with each other--i.e., if each lies within the error ranges of the others. We performed such tests at low, medium and high temperatures for clusters of all sizes considered to check that the U_n computed was the same at each temperature for all starting configurations. If there were low energy but mutually inaccessible regions of phase space not sampled during our simulations they either did not show up in our tests or the values of the energy U_n in each region were very close. To obtain the plots of $U_n - nU_1$ which were fitted and integrated, the values of U_n from different initial conditions were averaged together.

A second consistency check is to see if the fluctuations in potential energy at a given temperature are correctly related to the heat capacity at that temperature. For a canonical ensemble,

$$\overline{U_n^2} - \overline{U_n}^2 = -d\overline{U_n}/d\beta. \tag{5.1}$$

To make this check U_n^2 was calculated and averaged at each time step so that the left side of Eq. (5.1) could be determined. $d\overline{U_n}/d\beta$ was determined from the cubic spline fit of $U_n - nU_1$ as a function of β and the linear fit of U_1 as a function of T. The results of this test for dimers and pentamers are displayed in Figure 2, and those for trimers and tetramers are similar. In this figure the curves are more accurate than the dots, since each curve is the smoothed result of calculations of the average energy for all simulations of an n molecule cluster, whereas each dot is the result of a calculation of the energy fluctuations for just one simulation. If the dots fell exactly on the curve, this would mean that each simulation was long enough for the mean squared fluctuations in the potential energy to be calculated accurately. The statistical errors in average values is usually smaller than the errors in mean squared fluctuations, and thus agreement of the points and the curve would imply that the average potential energies are subject to very little statistical error. If the points fall near but not exactly on the line, it is still likely that the trajectories are long enough to give accurate averages even though their fluctuations are in error.

At high or low temperature, where the system is behaving mostly as n noninteracting molecules or as an n molecule cluster, respectively, the agreement of the points and the curve is excellent. The agreement is much worse at intermediate temperatures where there is a local maximum in the cluster specific heat. In fact, the correlation times evaluated from the simulations in the intermediate region are generally longer than the prior estimates based on the sphere traversal time and the cluster intermolecular vibrational frequency. This probably means that the time for formation and breakup of clusters determines the correlation time. For these temperatures, the runs were not long enough to achieve the same statistical accuracy achieved at lower and higher temperatures.

The volume of the containment sphere should have no effect on either $U_n - nU_1$ or K_n at low temperatures, provided the volume is large enough not to interfere with the motion of the n molecule cluster. To test that this was so, we performed a new series of simulations for the trimer using a sphere that is 5.33 times larger in volume than the standard trimer containment sphere. Although there should be differences in U_n and K_n at high temperature, these should disappear at temperatures low enough that the cluster is compact and the molecules stay away from the wall of the containment sphere. In Figure 1 it can be seen that at low temperature the energy of the cluster is indeed independent of the size of the containment sphere, to within the statistical uncertainty of the calculation. The equilibrium constants are shown in Figure 3. It is easy to show from Eqns. (2.12) and (2.13) that, at each β, $K_n(\beta)$ is an increasing function of the radius of the containment sphere. Thus, in Figure 3 the dashed curve should always be above the solid curve, if both were calculated exactly. In fact, the curves cross, but the difference between the two curves at low temperatures is approximately equal to the estimated statistical error in the two calculations. Thus, to within the statistical uncertainty in our calculations, the equilibrium constants are the same for both choices of the containment sphere radius.

VI. Results

The results for the logarithms of the equilibrium coefficients are displayed in Figure 4. In each case, these are graphed relative to ln $K_n(0)$, the infinite temperature result [see (2.17)].

The equilibrium coefficients for dimer and trimer formation may be compared indirectly with experiment as a test of the method and of the intermolecular potentials. In Section II we discuss the relationships that exist between the second and third virial coefficients, $B_2(\beta)$ and $B_3(\beta)$, and the second and third equilibrium coefficients, $K_2(\beta)$ and $K_3(\beta)$. These allow the calculation of a theoretical $B_2(\beta)$ from $K_2(\beta)$, as well as the calculation of an experimental ln $K_3(\beta)$ from experimental $B_2(\beta)$ and $B_3(\beta)$.

Figure 5 provides the comparison of experimental[18-20] and theoretically derived second virial coefficients. The experimental values are shown over the entire range of temperatures for which the experiments were performed. The theoretically derived curve for B_2 came from Eq. (2.18) plus small corrections for the effect of the long-ranged dipolar part of the intermolecular potential. (The second virial coefficient can be expressed as the integral of the Mayer f function over all relative positions and orientations of two molecules. Equation (2.18) for B_2 was derived assuming that the interaction, and hence the Mayer f function, is zero for intermolecular distances larger than $2R_2$. Thus, to obtain B_2 for the full Watts potential, it is necessary to add the contribution from the integral for distances larger than $2R_2$. The latter was calculated analytically[21] assuming that the Watts potential was dipolar at those large distances.)

The agreement between the experimental and theoretical second virial coefficients is seen to be quite good. Over the temperature range for which experimental information is available, the differences between the theory and experiment are only about as large as the differences among the various experiments. It should be noted, however, that the Watts potential was constructed to give second virial coefficients in agreement with experiment. The agreement between experiment and theory as shown in Figure 5 should therefore be interpreted as confirmation of our method for computing phase integrals. The procedure used by Watts[14] to compute second virial coefficients for various water potentials kept the water molecules rigid at their equilibrium internuclear configurations. The present calculations make no such restriction, and so the present method for calculating dimer formation constants of vibrating molecules gives a convenient way of calculating second virial coefficients of vibrating molecules.

In the case of water, however, the effect of vibrations in this purely classical calculation of the second virial coefficient is small.

Figure 6 compares K_3', the trimer equilibrium constant obtained from simulations using the larger than standard trimer containment sphere, with an experimental value of K_3' calculated from experimental second and third virial coefficients using Eq. (2.20). (In order to be consistent with this equation, we have subtracted from the experimental second virial coefficients an estimate of the long ranged dipolar contribution, since K_2 and K_3' were calculated for a potential that had no such long ranged part. In principle, the corresponding correction to the experimental third virial coefficient should also be made, but that is much more difficult and is not necessary since the experimental third virial coefficient makes only a small contribution to the experimental curve in this figure.) The difference between the experimental and theoretical results is of about the same size as the difference between different experiments and the statistical uncertainties in the calculations.

In principle the third virial coefficient can be calculated from K_2 and K_3' using Eq. (2.20). Unfortunately, this is not possible in practice because of the magnitude of the statistical error in K_3' and because Eq. (2.20) gives the third virial coefficient as the small difference between much larger numbers.

VII. Discussion

In this paper we have presented a practical procedure for evaluating the equilibrium constants for the formation of clusters in the gas phase using equilibrium classical statistical mechanics and assumptions about the intermolecular interactions. The calculations were based on a specific model potential for water and on the assumption that the energy of a collection of molecules is the sum of the interactions between pairs of molecules. The Watts intermolecular potential is known to be significantly different from the true water potential.[22] The assumption of pairwise additivity of the potential is also known to be not quantitatively accurate for water. Moreover, for water the internal motions cannot be regarded as classical. Because of these three facts, the calculations presented here should not be regarded as predictions of the equilibrium constants for water; rather they are example calculations designed to show the feasibility of the method. The method could easily be applied to a different choice of potentials, pairwise nonadditive potentials, clusters of ions and molecules, and other situations. The restriction to classical mechanics is intrinsic to the molecular dynamics method.

The method is an alternative to the methods of Lee _et_ _al_.[9] and of Mruzik _et_ _al_.[10] It differs from that of Lee _et_ _al_. by using an integration to infinite temperature to establish the absolute free energy of a cluster rather than using an integration over the value of the size of the containment sphere. It differs from the method of Mruzik _et_ _al_. in that the latter calculates the free energy difference for clusters differing by one molecule and relies on the fact that the absolute free energy of the single molecule cluster can be evaluated exactly.

As a byproduct of this method, we have a simulation method for calculating the second virial coefficient of nonrigid molecules.

Appendix: Velocity Form of the Verlet Algorithm

In this appendix we state the velocity form of the Verlet algorithm,[23] which we use for integrating Newton's equations of motion.

Assuming, for notational simplicity, that there is one degree of freedom, Newton's equations can be written as

$$\ddot{r} = f(r)$$

in which f(r) is the force divided by the mass. Let h be the time step for the numerical integration, and let

$$t_n = nh$$

$$r_n = r(t_n)$$

$$\dot{r}_n = \dot{r}(t_n).$$

The Verlet algorithm is based on the following approximations:

$$\dot{r}_n \approx (r_{n+1} - r_{n-1})/2h \tag{A.1}$$

$$\ddot{r}_n \approx (r_{n+1} - 2r_n + r_{n-1})/h^2. \tag{A.2}$$

The Verlet algorithm is obtained by treating Eq. (A.2) as an equality to get

$$r_{n+1} = 2r_n - r_{n-1} + h^2 f(r_n). \tag{A.3}$$

This can be the basis for a recursive procedure for calculating all subsequent r_n from r_0 and r_1. If the velocity is needed we define

$$v_n = (r_{n+1} - r_{n-1})/2h \qquad \text{(A.4)}$$

and using Eq. (A.1) we interpret v_n as $\dot{r}_n$. In the numerical analysis literature, the Verlet algorithm is known as the "explicit central difference method."[24]

A better way of implementing the Verlet algorithm on a computer with finite precision is to use the "summed form." Defining

$$z_n = (r_{n+1} - r_n)/h$$

it is easy to show that Eq. (A.3) is equivalent to

$$\begin{aligned} r_n &= r_{n-1} + hz_{n-1} \\ z_n &= z_{n-1} + hf(r_n). \end{aligned}$$

These equations can be iterated to obtain all subsequent values of r_n and z_n from r_0 and z_0. If velocities are needed, they can be obtained from

$$v_n = (z_n + z_{n-1})/2.$$

These equations are mathematically equivalent to the Verlet algorithm, but they are not numerically equivalent and are superior on a computer of finite precision.[24]

The velocity form of the Verlet algorithm is one in which v_n appears directly in the equations to be iterated. It is straightforward to show that the following equations are equivalent to Eqs. (A.3) and (A.4).

$$r_{n+1} = r_n + hv_n + h^2 f(r_n)/2$$

$$v_{n+1} = v_n + h[f(r_{n+1}) + f(r_n)]/2.$$

These equations retain the superior numerical precision of the summed form. Since they are a way of directly getting the position and velocity at the end of the time step from the position and velocity at the beginning of the step, they provide an easy way of grafting stochastic collisions into the algorithm. The effect of a stochastic collision is merely to change the value of v_n just before r_{n+1} is to be calculated.

References

1. H. R. Carlon, Infrared Physics 19, 49 (1979); and H. R. Carlon, Applied Physics 20, 726 (1981).
2. H. R. Carlon, Applied Physics 19, 1776 (1980); and H. R. Carlon, Applied Physics 20, 1316 (1981).
3. P. Kebarle, S. K. Searles, A. Zolla, J. Scarborough, and M. Arshadi, Adv. Mass Spectrom. 4, 621 (1968); P. Kebarle in Mass Spectrometry in Inorganic Chemistry, edited by J. L. Margrave (Adv. Chem. Ser. 72, American Chemical Society, Washington, D.C., 1968), p. 24; P. Kebarle in Ions and Ion Pairs in Organic Reactions, Volume 1, edited by M. Szware (Wiley-Interscience, New York, 1972) p. 27; P. Kebarle in Ion-Molecule Reactions, Volume 1, edited by J. L. Franklin (Plenum, New York, 1972) p. 315; and, P. Kebarle in Modern Aspects of Electrochemistry, Volume 9, edited by B. E. Conway and J. Bockris (Plenum, New York, 1974) p. 1.
4. J. E. McDonald, Am. J. Physics 30, 870 (1962); and J. J. Burton in Modern Theoretical Chemistry, edited by B. J. Berne (Plenum, New York, 1977), Volume 5A, p. 195.
5. F. F. Abraham, Homogeneous Nucleation Theory (Academic Press, New York, 1974).
6. T. L. Hill, Statistical Mechanics (McGraw-Hill, New York, 1956) p. 152.

7. D. J. McGinty, Chem. Phys. Lett. 13, 525 (1972).

8. C. L. Briant and J. J. Burton, Nature London Phys. Sci. 243, 100 (1973); C. L. Briant and J. J. Burton, J. Chem. Phys. 60, 2849 (1974); 63, 2045 (1975); 63, 3327 (1975); 64, 2888 (1976); C. L. Briant and J. J. Burton, J. Atmospheric Sci. 33, 1357 (1976); W. D. Kristensen, E. J. Jensen, and R. M. J. Cotterill, J. Chem. Phys. 60, 4161 (1974); D. J. McGinty, J. Chem. Phys. 58, 4733 (1973); F. F. Abraham, J. Chem. Phys. 61, 1221 (1974).

9. J. K. Lee, J. A. Barker and F. F. Abraham, J. Chem. Phys. 58, 3166 (1973).

10. M. R. Mruzik, F. F. Abraham, D. E. Schreiber, and G. M. Pound, J. Chem. Phys. 64, 481 (1976).

11. T. L. Hill, Introduction to Statistical Thermodynamics (Addison-Wesley, Reading, Massachusetts, 1960) p. 261.

12. H. C. Andersen, J. Chem. Phys. 72, 2384 (1980).

13. F. H. Stillinger and A. Rahman, J. Chem. Phys. 68, 666 (1978).

14. R. O. Watts, Chem. Phys. 26, 367 (1977).

15. T. Andrea and H. C. Andersen, unpublished results.

16. C. H. Reinsch, Numer. Math. 10, 177 (1967); 16, 451 (1971). A computer program using this approach is SMOOTH which is available as part of the Argonne Scientific Program Package.

17. B. K. Janousek, Ph.D. thesis, Stanford University, Stanford, CA (1979).

18. F. G. Keyes, L. B. Smith, and H. T. Gerry, Proc. Am. Acad. Arts Sci. 70, 319 (1936).

19. G. S. Kell, G. E. McLaurin, and E. Walley, J. Chem. Phys. 48, 3805 (1968).

20. M. P. Vukalovich, M. S. Trakhtengerts, and G. A. Spiridonov, Teploenergetika 14, 65 (1967).

21. This exercise is similar to that of determining the second virial coefficient of point dipoles imbedded in hard spheres. See J. O. Hirschfelder, C. F. Curtiss and R. B. Bird, Molecular Theory of Gases and Liquids (Wiley, New York, 1954) p. 210.

22. M. D. Morse and S. A. Rice, to be published.

23. L. Verlet, Phys. Rev. 159, 98 (1967).

24. G. Dahlquist and A. Björck, Numerical Methods (Prentice Hall, New York, 1974) p. 353.

Figure Captions

Figure 1. $U_n - nU_1$ for n = 2 to n = 5. The dashed line shows $U_3 - 3U_1$ with U_3 calculated using the larger than standard trimer containment sphere. (Vertical lines are error estimates of ±1 standard deviation.)

Figure 2. Dimer (n = 2) and pentamer (n = 5) potential energy heat capacity, $-dU_n/d\beta$ (solid lines), and fluctuation in potential energy, $\langle U_n^2 \rangle - \langle U_n \rangle^2$ (circles), as a function of β. See the text for the use of this graph as a consistency check on the calculation.

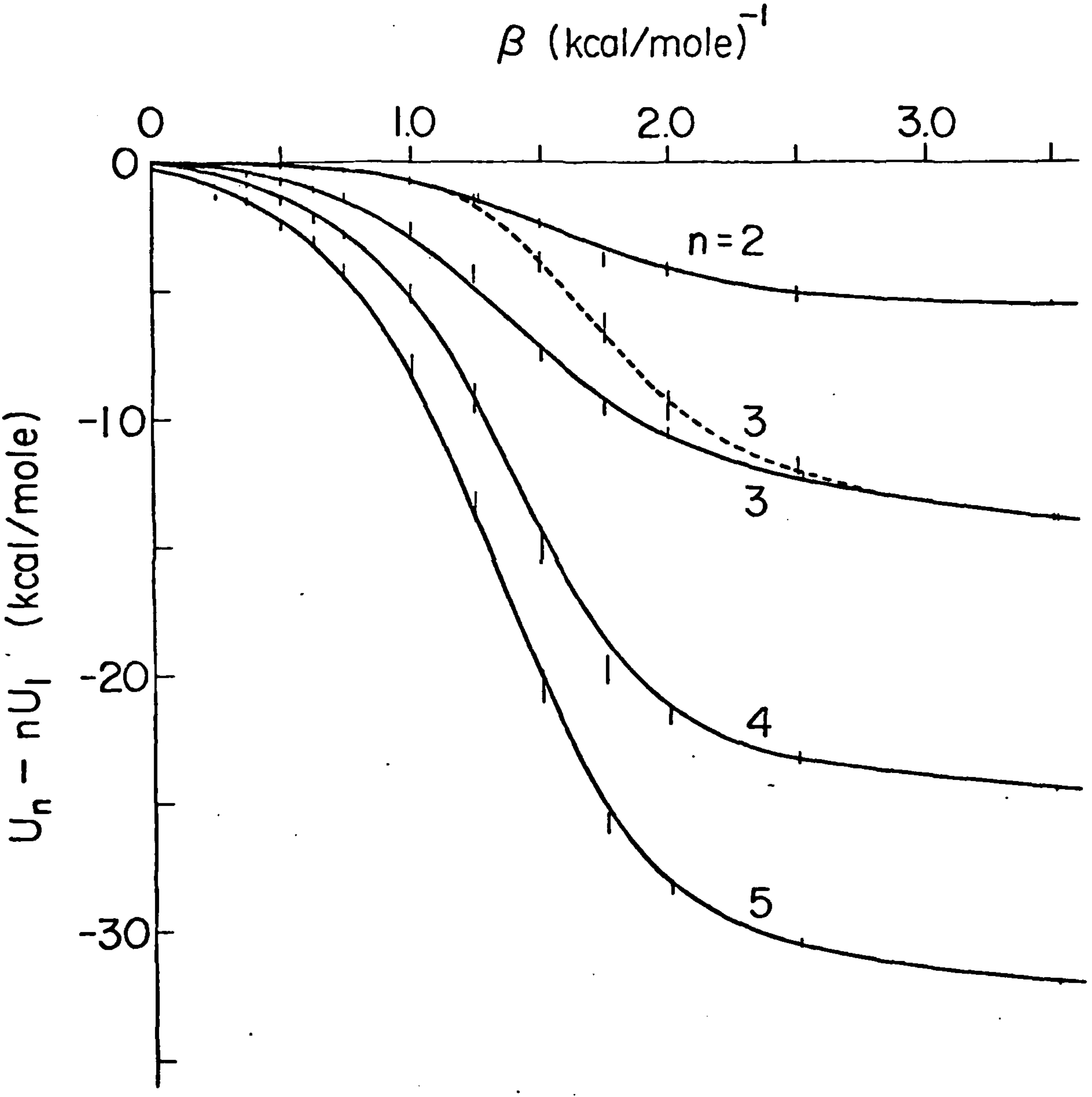
Figure 3. $\mathrm{Ln}[K_3(\beta)/K_3(0)]$ calculated using both the standard containment sphere size (solid curve) and the larger than standard containment sphere size (dashed curve). $K_3(0)$, to which both curves are referenced is that of the standard trimer containment sphere.

Figure 4. $\mathrm{Ln}[K_n(\beta)/K_n(0)]$ for n = 2 to n = 5.

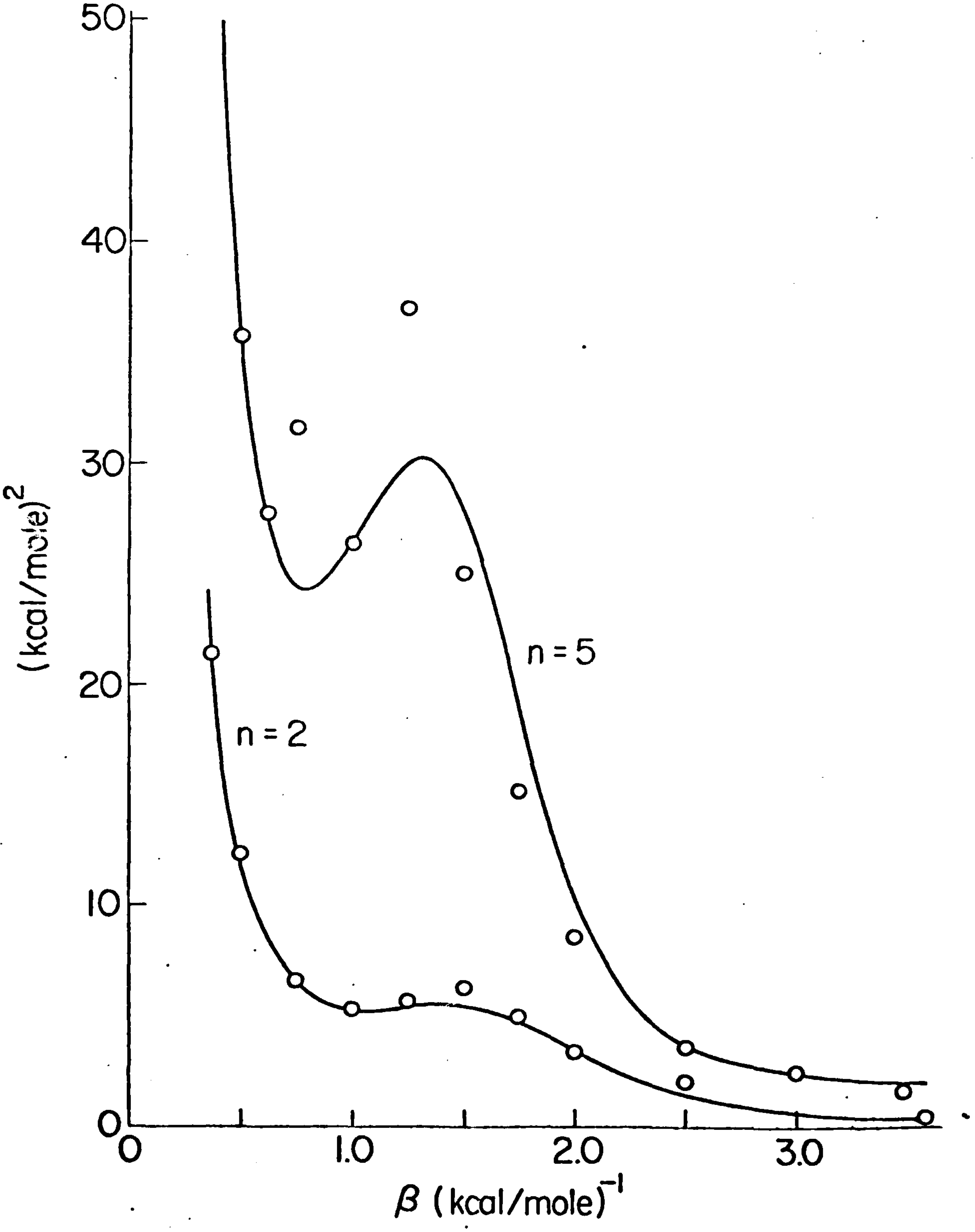
Figure 5. The second virial coefficient, $B_2(\beta)$, from experiment and theory. The solid line shows B_2 calculated from $K_2(\beta)$. The dashed line is from the data of Keyes, et al. (Reference 18); the dotted line is from the data of Kell et al. (Reference 19); the

dot-dashed line is from the data of Vukalovich _et_ _al_. (Reference 20).

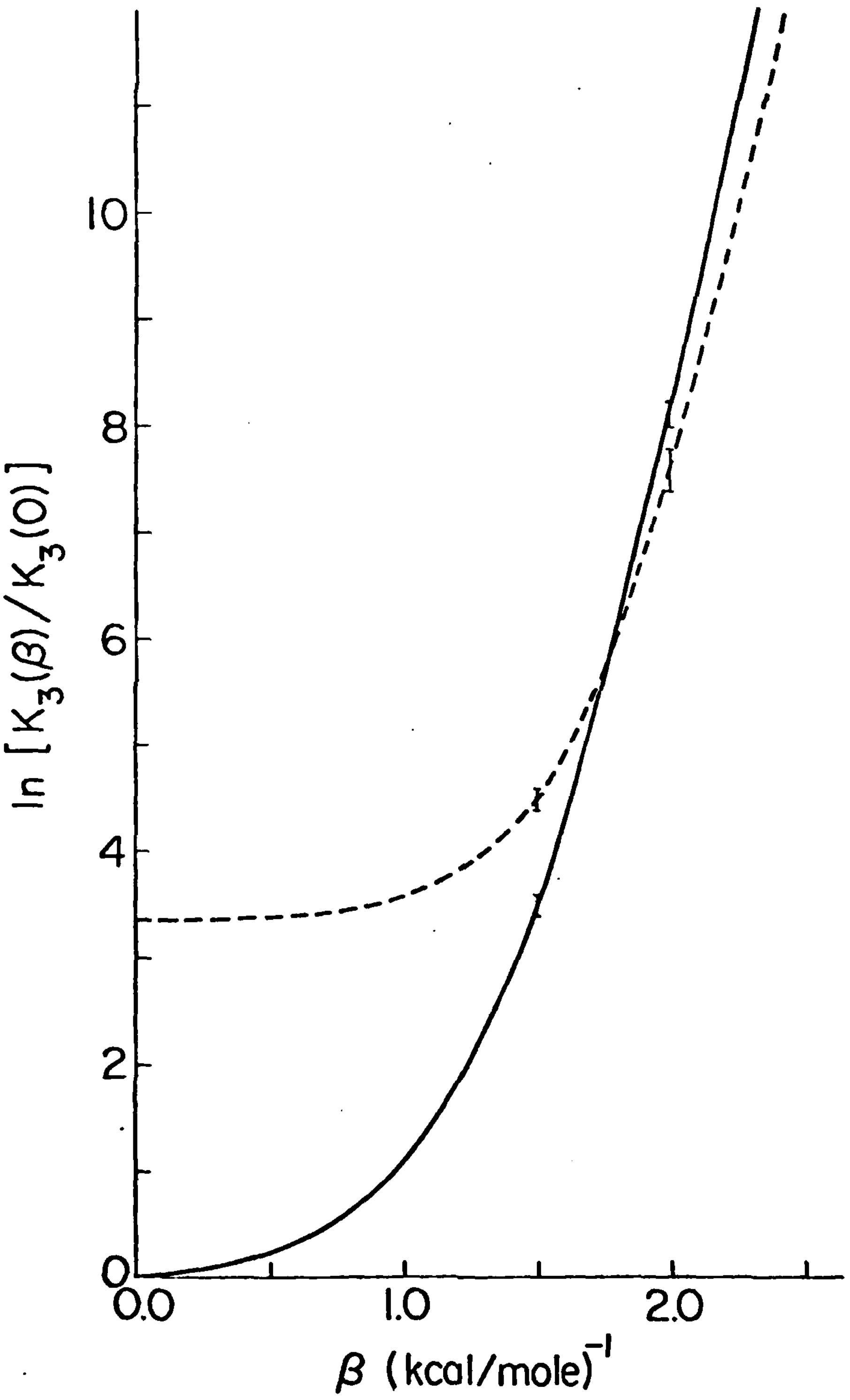
Figure 6. Ln[$K_3'(\beta)/K_3'(0)$] as computed from theory (solid curve) and from experiments. The theoretical curve was obtained from molecular dynamics results using the larger than standard trimer containment sphere. The experimental curves were obtained using Eq. (2.20) and second and third virial coefficient data of Kell _et_ _al_. (Reference 19) and of Vukalovich _et_ _al_. (Reference 20). The data of Kell _et_ _al_. determine the dotted curve and that of Vukalovich _et_ _al_. determine the dot-dashed curve.



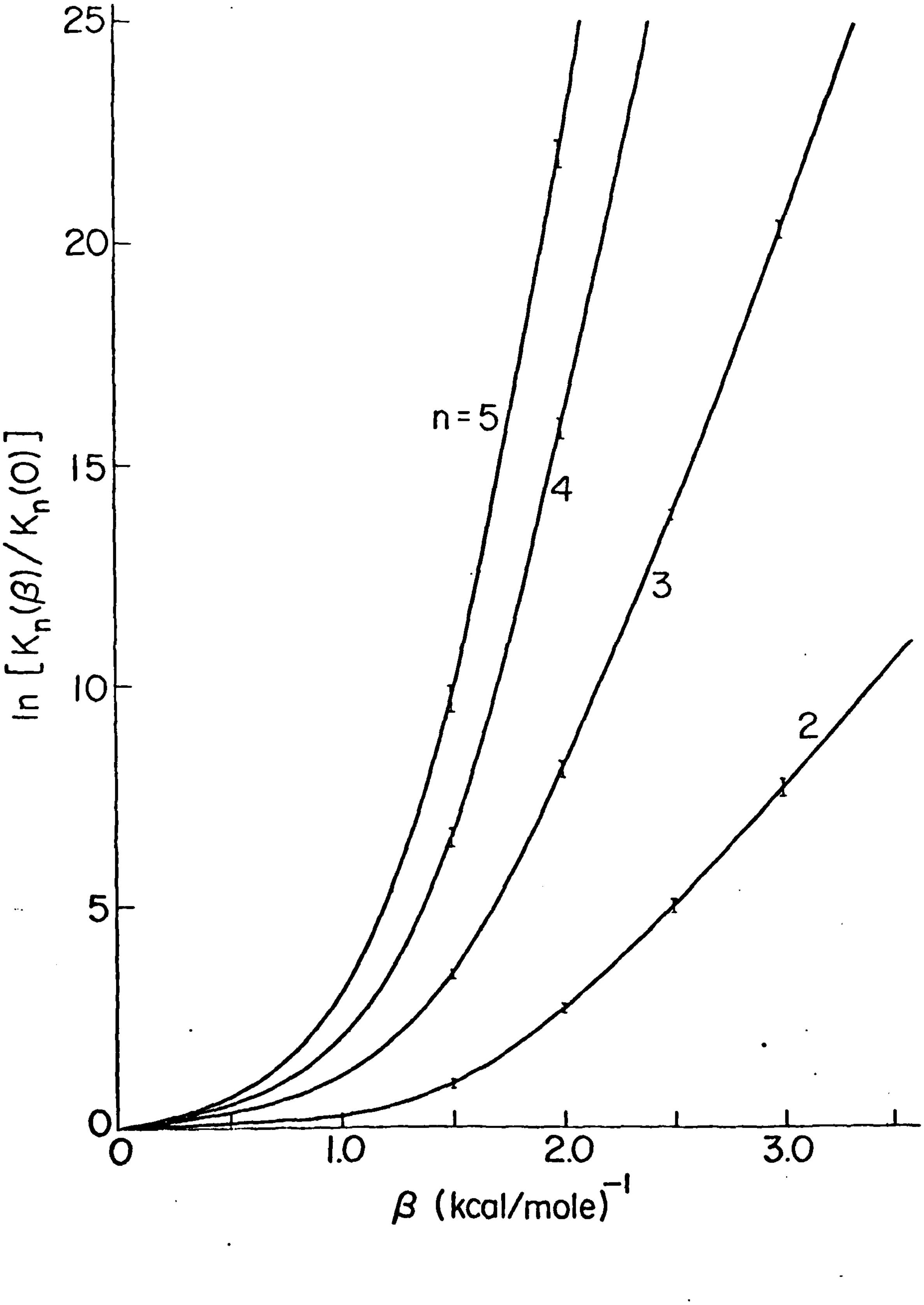
β (kcal/mole)$^{-1}$
0
1.0
2.0
3.0
0
-10
-20
-30
$U_n - nU_1$ (kcal/mole)
n = 2
3
3
4
5



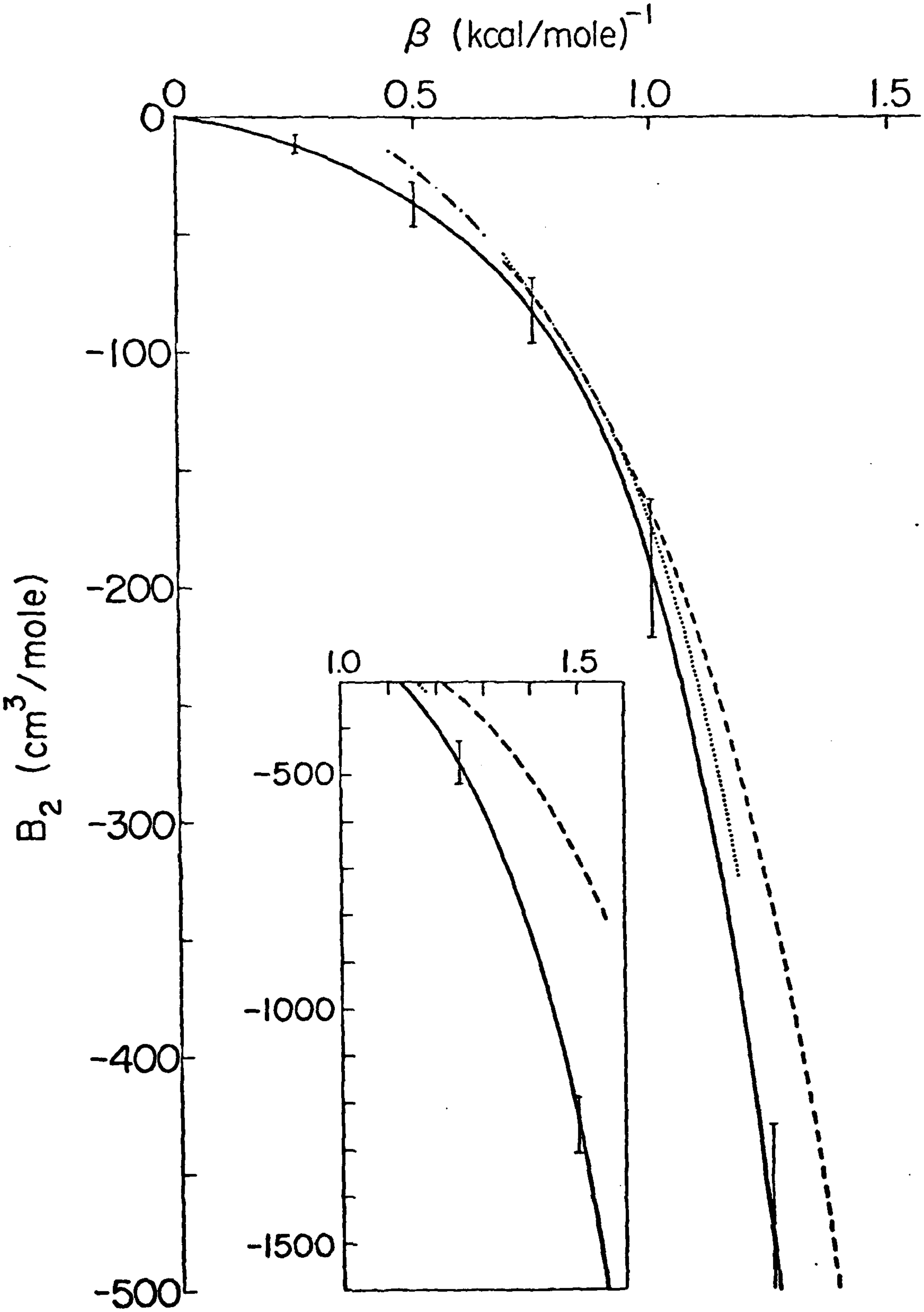
50
40
30
20
10
0
(kcal/mole)^2
n = 5
n = 2
0
1.0
2.0
3.0
β (kcal/mole)$^{-1}$

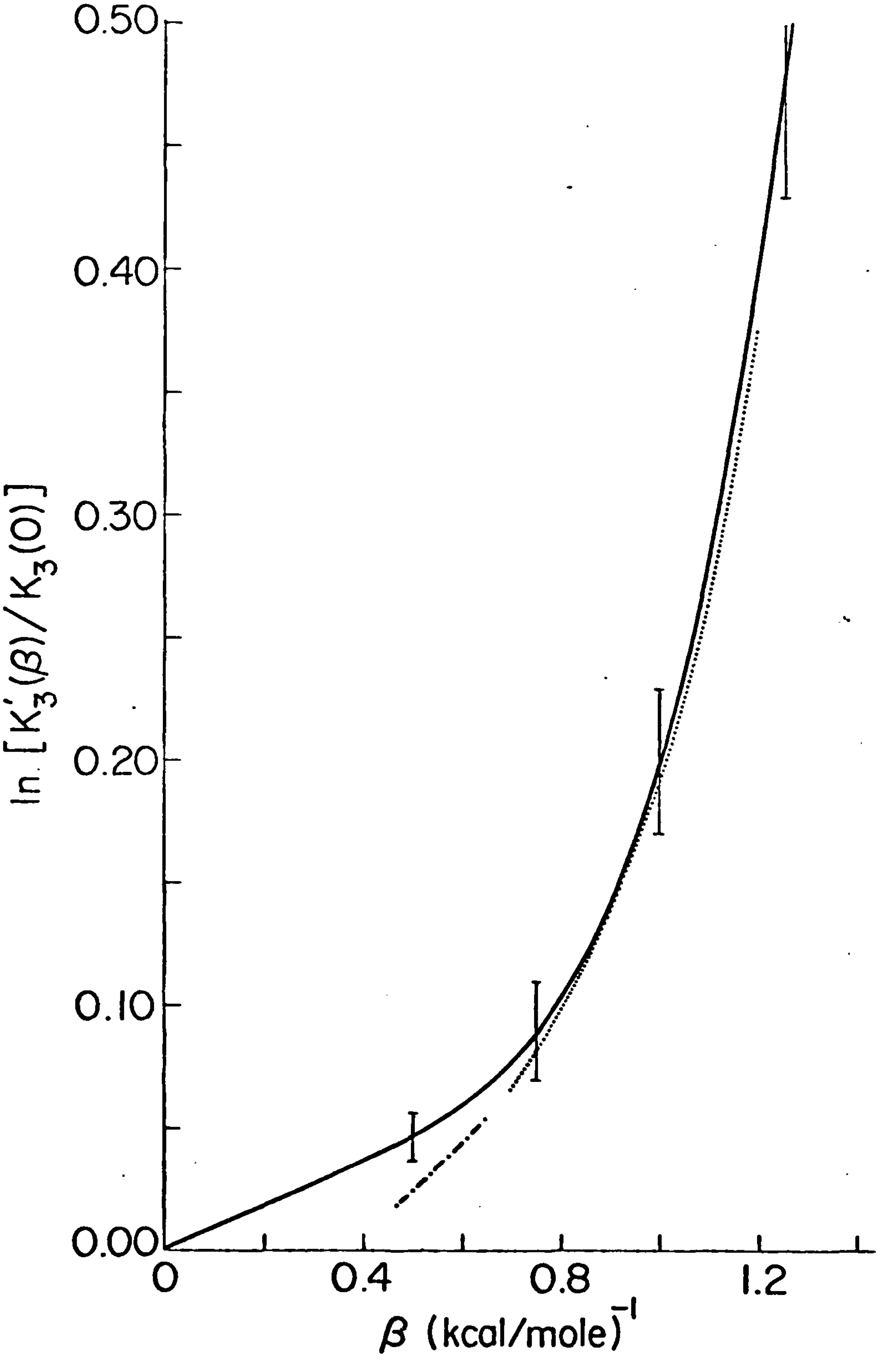


ln $[K_3(\beta)/K_3(0)]$
10
8
6
4
2
0
0.0
1.0
2.0
β (kcal/mole)$^{-1}$



25
20
15
10
5
0
ln [K_n(β)/K_n(0)]
n = 5
4
3
2
0
1.0
2.0
3.0
β (kcal/mole)^-1


β (kcal/mole)$^{-1}$
0
0.5
1.0
1.5
0
-100
-200
-300
-400
-500
B_2 (cm^3/mole)
1.0
1.5
-500
-1000
-1500



0.50
0.40
0.30
0.20
0.10
0.00
ln $[K_3'(\beta)/K_3(0)]$
0
0.4
0.8
1.2
β (kcal/mole)$^{-1}$

TECHNICAL REPORT DISTRIBUTION LIST, GEN

	No. Copies
Office of Naval Research Attn: Code 472 800 North Quincy Street Arlington, Virginia 22217	2
ONR Branch Office Attn: Dr. George Sandoz 536 S. Clark Street Chicago, Illinois 60605	1
ONR Branch Office Attn: Scientific Dept. 715 Broadway New York, New York 10003	1
ONR Branch Office 1030 East Green Street Pasadena, California 91106	1
ONR Branch Office Attn: Dr. L. H. Peebles Building 114, Section D 666 Summer Street Boston, Massachusetts 02210	1
Director, Naval Research Laboratory Attn: Code 6100 Washington, D.C. 20390	1
The Assistant Secretary of the Navy (R,E&S) Department of the Navy Room 4E736, Pentagon Washington, D.C. 20350	1
Commander, Naval Air Systems Command Attn: Code 310C (H. Rosenwasser) Department of the Navy Washington, D.C. 20360	1
Defense Documentation Center Building 5, Cameron Station Alexandria, Virginia 22314	12
Dr. Fred Saalfeld Chemistry Division Naval Research Laboratory Washington, D.C. 20375	1
U.S. Army Research Office Attn: CRD-AA-IP P.O. Box 1211 Research Triangle Park, N.C. 27709	1
Naval Ocean Systems Center Attn: Mr. Joe McCartney San Diego, California 92152	1
Naval Weapons Center Attn: Dr. A. B. Amster, Chemistry Division China Lake, California 93555	1
Naval Civil Engineering Laboratory Attn: Dr. R. W. Drisko Port Hueneme, California 93401	1
Department of Physics & Chemistry Naval Postgraduate School Monterey, California 93940	1
Dr. A. L. Slafkosky Scientific Advisor Commandant of the Marine Corps (Code RD-1) Washington, D.C. 20380	1
Office of Naval Research Attn: Dr. Richard S. Miller 800 N. Quincy Street Arlington, Virginia 22217	1
Naval Ship Research and Development Center Attn: Dr. G. Bosmajian, Applied Chemistry Division Annapolis, Maryland 21401	1
Naval Ocean Systems Center Attn: Dr. S. Yamamoto, Marine Sciences Division San Diego, California 91232	1
Mr. John Boyle Materials Branch Naval Ship Engineering Center Philadelphia, Pennsylvania 19112	1

TECHNICAL REPORT DISTRIBUTION LIST, GEN

	No. Copies
Dr. Rudolph J. Marcus Office of Naval Research Scientific Liaison Group American Embassy APO San Francisco 96503	1
Mr. James Kelley DTNSRDC Code 2803 Annapolis, Maryland 21402	1

472:CAN:716:tam
78u472-608

TECHNICAL REPORT DISTRIBUTION LIST, 051A

	No. Copies
~~Dr. M. A. El-Sayed~~ Stet ~~University of California,~~ ~~Los Angeles~~ ~~Department of Chemistry~~ ~~Los Angeles, California 90024~~	1
Dr. M. W. Windsor Washington State University Department of Chemistry Pullman, Washington 99163	1
Dr. E. R. Bernstein Colorado State University Department of Chemistry Fort Collins, Colorado 80521	1
Dr. C. A. Heller Naval Weapons Center Code 6059 China Lake, California 93555	1
Dr. J. R. MacDonald Naval Research Laboratory Chemistry Division Code 6110 Washington, D.C. 20375	1
Dr. G. B. Schuster University of Illinois Chemistry Department Urbana, Illinois 61801	1
Dr. E. M. Eyring University of Utah Department of Chemistry Salt Lake City, Utah 84112	1
Dr. A. Adamson University of Southern California Department of Chemistry Los Angeles, California 90007	1
Dr. M. S. Wrighton Massachusetts Institute of Technology Department of Chemistry Cambridge, Massachusetts 02139	1

	No. Copies
Dr. M. Rauhut American Cyanamid Company Chemical Research Division Bound Brook, New Jersey 08805	1
Dr. J. I. Zink University of California, Los Angeles Department of Chemistry Los Angeles, California 90024	1
Dr. B. Schechtman IBM San Jose Research Center 5600 Cottle Road San Jose, California 95143	1
Dr. John Cooper Code 6130 Naval Research Laboratory Washington, D.C. 20375	1

TECHNICAL REPORT DISTRIBUTION LIST, 051B

	No. Copies
~~Professor K. Wilson~~ ~~Department of Chemistry,~~ ~~B-014~~ ~~University of California,~~ ~~San Diego~~ ~~La Jolla, California 92093~~	1
Professor C. A. Angell Department of Chemistry Purdue University West Lafayette, Indiana 47907	1
Professor P. Meijer Department of Physics Catholic University of America Washington, D.C. 20064	1
Dr. S. Greer Chemistry Department University of Maryland College Park, Maryland 20742	1
Professor P. Delahay New York University 100 Washington Square East New York, New York 10003	1
Dr. T. Ashworth Department of Physics South Dakota School of Mines & Technology Rapid City, South Dakota 57701	1
Dr. G. Gross New Mexico Institute of Mining & Technology Socorro, New Mexico 87801	1
Dr. J. Kassner Space Science Research Center University of Missouri - Rolla Rolla, Missouri 65401	1
Dr. J. Telford University of Nevada System Desert Research Institute Lab of Atmospheric Physics Reno, Nevada 89507	1

	No. Copies
Dr. B. Vonnegut State University of New York Earth Sciences Building 1400 Washington Avenue Albany, New York 12203	1
Dr. Hank Loos Laguna Research Laboratory 21421 Stans Lane Laguna Beach, California 92651	1
Dr. John Latham University of Manchester Institute of Science & Technology P.O. Box 88 Manchester, England M601QD	

TECHNICAL REPORT DISTRIBUTION LIST, 051C

No. Copies

Dr. M. B. Denton
Department of Chemistry
University of Arizona
Tucson, Arizona 85721 1

Dr. R. A. Osteryoung
Department of Chemistry
State University of New York
at Buffalo
Buffalo, New York 14214 1

Dr. B. R. Kowalski
Department of Chemistry
University of Washington
Seattle, Washington 98105 1

Dr. S. P. Perone
Department of Chemistry
Purdue University
Lafayette, Indiana 47907 1

Dr. D. L. Venezky
Naval Research Laboratory
Code 6130
Washington, D.C. 20375 1

Dr. H. Freiser
Department of Chemistry
University of Arizona
Tuscon, Arizona 85721

Dr. Fred Saalfeld
Naval Research Laboratory
Code 6110
Washington, D.C. 20375 1

Dr. H. Chernoff
Department of Mathematics
Massachusetts Institute of Technology
Cambridge, Massachusetts 02139 1

~~Dr. K. Wilson~~
~~Department of Chemistry~~
~~University of California, San Diego~~
~~La Jolla, California~~ 1

Dr. A. Zirino
Naval Undersea Center
San Diego, California 92132 1

No. Copies

Dr. John Duffin
United States Naval Postgraduate
School
Monterey, California 93940 1

Dr. G. M. Hieftje
Department of Chemistry
Indiana University
Bloomington, Indiana 47401 1

Dr. Victor L. Rehn
Naval Weapons Center
Code 3813
China Lake, California 93555 1

Dr. Christie G. Enke
Michigan State University
Department of Chemistry
East Lansing, Michigan 48824

Dr. Kent Eisentraut, MBT
Air Force Materials Laboratory
Wright-Patterson AFB, Ohio 45433 1

Walter G. Cox, Code 3632
Naval Underwater Systems Center
Building 148
Newport, Rhode Island 02840 1

DATE

ILME