

---

Konrad-Zuse-Zentrum  
für Informationstechnik Berlin

Takustraße 7  
D-14195 Berlin-Dahlem  
Germany

BURCU TUNGA<sup>1</sup> AND MARCUS WEBER

# **Free Energy Calculation Using Mayer Cluster Expansion and Fluctuation Free Integration**

---

<sup>1</sup>Istanbul Technical University, Faculty of Computer and Informatics, Maslak, 34469, İstanbul, Turkey, tungab@itu.edu.tr

Herausgegeben vom  
Konrad-Zuse-Zentrum für Informationstechnik Berlin  
Takustraße 7  
D-14195 Berlin-Dahlem

Telefon: 030-84185-0  
Telefax: 030-84185-125

e-mail: [bibliothek@zib.de](mailto:bibliothek@zib.de)  
URL: <http://www.zib.de>

ZIB-Report (Print) ISSN 1438-0064  
ZIB-Report (Internet) ISSN 2192-7782

# Free Energy Calculation Using Mayer Cluster Expansion and Fluctuation Free Integration

Burcu Tunga and Marcus Weber

12. Oktober 2012

...

## Zusammenfassung

This work aims to develop a new algorithm to calculate the free energy of water molecules by using a deterministic way. For this purpose, we assume a closed system confined to a physical volume, having water molecules in gas phase. To calculate the free energy of this system we utilized Mayer cluster expansion and the fluctuation free integration method.

## 1 Introduction

Molecular simulation methods are often used to generate statistical data in order to estimate the probability of a system to attain a certain state. For example, how probable is a receptor-ligand-binding compared to the unbound state? The answer to such questions is equivalent to computing the free energy difference between the states of the molecular system. The generation of statistical data is comparable to using a Monte Carlo approach for solving the underlying quadrature problem.

From a computational point of view, Monte Carlo methods are preferable for high-dimensional integrals, because deterministic methods suffer from the curse of dimensionality.

To compute the free energy of the systems, in general, Monte Carlo methods are often combined with sophisticated analytical reformulations of the problem [1, 2].

However, if the free energy of a system can be expressed by solving many low-dimensional integrals, purely deterministic approaches may become valuable. In this work, we do not use any simulation techniques, instead, we developed a deterministic algorithm to calculate the free energy of water in gas phase. For this purpose, we utilize some numerical methods such as Mayer Cluster Expansion[3] and Fluctuation Free Integration[4, 5].

Especially, six-dimensional integrals will play a major role in this article, if the potential energy can be expressed by only writing atom-pair-interaction terms.

## 2 Mathematical Background

We will calculate the free energy of water in gas phase. In this sense, we need to exploit the mathematical structure of partition functions, we need the Mayer Cluster Expansion method and the Fluctuation Free Integration method. These methods will help us to develop an algorithm for deterministic free energy calculations. The details of these methods are given in the followings.

### 2.1 Partition Function and Mayer Cluster Expansion

There are actually several different types of partition functions, each of them corresponds to a different type of a statistical ensemble. In this study we deal with the canonical ensemble partition function. The canonical ensemble of a system means that the system is observed at constant temperature with fixed number of particles,  $N$ . The canonical ensemble partition function is very important for statistical thermodynamics because it allows us to evaluate most of the thermodynamic properties of the system, such as the free energy, entropy, and enthalpy.

Consider a classical system inside of a finite-volumed box which consists of  $N$  particles. The canonical partition function of a gas particles system can be written in terms of a total energy  $H$  (depending on the locations and the momenta of the particles) as follows

$$Z(\beta) \equiv \frac{1}{N!h^{3N}} \int_{\mathcal{V}} d\mathcal{V} e^{-\beta H(\mathbf{p}, \mathbf{q})} \quad (1)$$

where the symbols  $h$  and the term  $e^{-\beta H}$  are known as reduced Planck constant and the Boltzmann factor respectively. The above form of the partition function is not dimensionless and to make it a dimensionless quantity, we must divide the relation by  $3N$ th power of reduced Planck constant,  $h$ . In addition, we know that it is impossible to distinguish the particles even they are actually identical particles. To this end, the partition function must be divided by  $N!$ . Moreover, here,  $\beta$  is called as reciprocal micro thermal energy and it is defined as

$$\beta \equiv \frac{1}{k_B T} \quad (2)$$

where  $k_B$  and  $T$  are Boltzmann's constant and absolute temperature, respectively. In this study, we use three dimensional physical space that is, we need  $6N$  entities formed by  $3N$  momenta and  $3N$  positions. So, in definition of the partition function shown above as (1), a  $6N$ -tuple integration is involved. Consequently,  $\mathcal{V}$  defines the volume, which is covered by momentum,  $\mathbf{p}$  and position,  $\mathbf{q}$  variables. The sets of momenta,  $\mathbf{p}$  and positions,  $\mathbf{q}$  are given through the following expression

$$\mathbf{p} \equiv \{p_1, \dots, p_{3N}\}, \quad \mathbf{q} \equiv \{q_1, \dots, q_{3N}\} \quad (3)$$

Since the equation given in (1) has  $6N$  dimensional integral, the definition of  $d\mathcal{V}$ , which corresponds to the value sets taken by  $\mathbf{p}$  and  $\mathbf{q}$  variables, can be given as follows

$$d\mathcal{V} = dp_1 \cdots dp_{3N} \cdot dq_1 \cdots dq_{3N} \quad (4)$$

Usually, a limitation is not considered for the speed of particles, but even if it is considered, it can be predicted that, a particle's momentum will not affect another particle's momentum. These predictions bring up noticable eases, otherwise major problems can be faced with abovementioned integration evaluation. Dealing with identical particles allows us to assume that the momentum of each particle will take a value from a single value set. Thus,  $V_d$  will be used as the volume parameter for each particle and the following domains will be used

$$(p_{3i-2}, p_{3i-1}, p_{3i}) \in V_d, \quad i = 1, \dots, N \quad (5)$$

Furthermore, if it is considered that all the particles are in a finite, consequently a closed volume, and this volume is symbolized with  $V$ , then the following domains are used.

$$(q_{3i-2}, q_{3i-1}, q_{3i}) \in V, \quad i = 1, \dots, N \quad (6)$$

Moreover, the mathematical expression of Hamiltonian of the system used in this work can be given as

$$H(\mathbf{p}, \mathbf{q}) \equiv \frac{1}{2\mu} \sum_{j=1}^{3N} p_j^2 + U(\mathbf{q}) \quad (7)$$

where  $\mu$  represents the mass of a particle which is same for all. However, the potential function  $U$  is not given explicitly in the above equation. It is assumed to be the sum of all possible pairwise interactions.

To this end, the partition function can be rewritten in terms of the relations given up to here as follows

$$\begin{aligned} Z(\beta) &\equiv \frac{1}{N! \hbar^{3N}} \left[ \prod_{j=1}^N \int_{V_d} dp_{3j-2} dp_{3j-1} dp_{3j} e^{\frac{-\beta}{2\mu} (p_{3j-2}^2 + p_{3j-1}^2 + p_{3j}^2)} \right] \\ &\times \int_V \cdots \int_V dq_{3j-2} dq_{3j-1} dq_{3j} e^{-\beta U(\mathbf{q})} \end{aligned} \quad (8)$$

Here, there are  $N$  number of 3-tuple integrals on the volume,  $V$ .

If the very specific nature of the Hamiltonian is used to separate the integration over momenta and positions, the following equation can be written

$$Z(\beta) \equiv \frac{1}{N!} \left( \frac{\alpha_1 \alpha_2}{\hbar} \right)^{3N} D_{mom}(\beta)^N Z_{pos}(\beta) \quad (9)$$

where

$$Z_{pos}(\beta) \equiv \frac{1}{\alpha_2^{3N}} \int_V \cdots \int_V dq_1 \cdots dq_{3N} e^{-\beta U(\mathbf{q})} \quad (10)$$

and

$$D_{mom}(\beta) \equiv \frac{1}{\alpha_1^{3N}} \int_{V_d} dp_{3j-2} dp_{3j-1} dp_{3j} e^{\frac{-\beta}{2\mu} (p_{3j-2}^2 + p_{3j-1}^2 + p_{3j}^2)} \quad (11)$$

Here,  $\alpha_1$  and  $\alpha_2$  parameters are used as scalars to make it dimensionless. It is assumed that the scaling is independent of the direction. The equations given

in (10) and (11) are called as “Position Interacted Partition Function” and “Momentum Distribution Function” respectively.

Momentum Distribution Function is determined through the idea of that the volume is completely overlapped with the whole 3-dimensional space. That is,

$$\begin{aligned}
D_{mom}(\beta) &= \frac{1}{\alpha_1^3} \int_{-\infty}^{\infty} dp_1 \int_{-\infty}^{\infty} dp_2 \int_{-\infty}^{\infty} dp_3 e^{\frac{-\beta}{2\mu}(p_1^2+p_2^2+p_3^2)} \\
&= \left( \frac{1}{\alpha_1} \int_{-\infty}^{\infty} dp e^{\frac{-\beta}{2\mu}p^2} \right)^3 = \left( \frac{2}{\alpha_1} \int_0^{\infty} \frac{dp}{\sqrt{p}} e^{\frac{-\beta}{2\mu}p} \right)^3 \\
&= \left( \frac{2\mu}{\beta\alpha_1^2} \right)^{\frac{3}{2}} \left( \int_0^{\infty} dp e^{-p} p^{\frac{-1}{2}} \right)^3 = \left( \frac{2\pi\mu k_B}{\alpha_1^2} \right)^{\frac{3}{2}} T^{\frac{3}{2}} \quad (12)
\end{aligned}$$

To this end, the Momentum Distribution Function is evaluated without depending on the interaction appeared between particles. If we use the result given in relation (12) and rewrite the equation (9) then we can reach the following equation

$$Z(\beta) \equiv \frac{1}{N!} \left( \frac{2\pi\mu\alpha_2^2}{\hbar^2\beta} \right)^{\frac{3N}{2}} Z_{pos}(\beta) \quad (13)$$

In contrast to Momentum Distribution Function, the Position Interacted Partition Function cannot be evaluated so easily. This time to evaluate that expression we can use an assumption. This assumption says that the interactions between the particles are assumed as the summation of only the binary interaction between particles.

$$U(\mathbf{q}) \equiv \sum_{j=1}^{N-1} \sum_{i=j+1}^N U_{i,j}(r_{i,j}) \quad (14)$$

Here,  $r_{i,j}$  represents the distance between  $i$ th and  $j$ th particles and it is given as follows.

$$r_{i,j} \equiv \left[ (q_{3i-2} - q_{3j-2})^2 + (q_{3i-1} - q_{3j-1})^2 + (q_{3i} - q_{3j})^2 \right]^{\frac{1}{2}}, \quad i, j = 1, \dots, N \quad (15)$$

We consider that the binary interaction potential should have the same unique structure for all pairs because of the identity of the particles in the system. Hence, we use the following relation

$$e^{-\beta U(\mathbf{q})} = \prod_{i=1}^{N-1} \prod_{j=i+1}^N e^{-\beta U_{i,j}(r_{i,j})} \quad (16)$$

and the following Mayer Function

$$\mathcal{M}_{i,j} \equiv e^{-\beta U_{i,j}(r_{i,j})} - 1, \quad i, j = 1, \dots, N \quad (17)$$

which is symbolized by  $\mathcal{M}_{i,j}$ [6, 7] to determine the canonical partition function,  $Z(\beta)$ , given in (13). Mayer Function urges us to write following relation.

$$e^{-\beta U_{i,j}(r_{i,j})} \equiv \mathcal{M}_{i,j} + 1, \quad i, j = 1, \dots, N \quad (18)$$

If we substitute this relation into the equation given in (16) the equation becomes a power series of the partition functions and it is called as Mayer Cluster Expansion[3].

$$e^{-\beta U(\mathbf{q})} = \prod_{i=1}^{N-1} \prod_{j=i+1}^N (1 + \mathcal{M}_{i,j}) = 1 + \sum_{i=1, i < j}^N \mathcal{M}_{i,j} + \sum_{i=1, i < j, k=1, k < l}^N \mathcal{M}_{i,j} \mathcal{M}_{k,l} + \dots \quad (19)$$

The first term of this expansion is for the single particle term which means that there are no any interaction between the particles, the second term corresponds to the two particle interactions, the third one is for the three particle interactions, and so on. That is, the interactions within clusters of a certain number of particles are expressed by each term. Thus, this expansion is called the cluster expansion. If this series replaced into the equation (10), the following series expansion is obtained as a result.

$$Z_{pos}(\beta) = \tilde{V}^N + \frac{N(N-1)}{2} \tilde{V}^{N-2} S_2(\beta) + \dots \quad (20)$$

Here, the first term is the most dominant term because dominance is coming from the particles which have no relations with the others like in the case of ideal gases. The second term is less dominant term and it is related to pairwise interacting two particle cases which are somehow related to Van der Waals gases. As the clustering level increases then the dominance decreases for this expansion.

The first term of the right hand side of the expansion given in (20) is symbolized by  $Z_{pos}^{(0)}$  (non-existence of particle interactions):

$$Z_{pos}^{(0)} \equiv \tilde{V}^N = \left( \frac{V}{\alpha_2^3} \right)^N = \frac{1}{\alpha_2^{3N}} \int_V \dots \int_V dq_1 \dots dq_{3N} \quad (21)$$

This term does not contain any Mayer function and its integrand is just the unit constant function. For this reason, the multivariate integral given in (21) can be easily evaluated analytically and the result is obtained as  $\tilde{V}$ . Actually,  $\tilde{V}$  expresses as the macroscopic dimensionless volume in microparticulate units.

The second term is denoted by  $Z_{pos}^{(1)}$  and it is considered as the binary interaction integral between any two particles

$$\begin{aligned} Z_{pos}^{(1)}(\beta) &\equiv \frac{N(N-1)}{2} \tilde{V}^{N-2} S_2(\beta) \\ &= \frac{1}{\alpha_2^{3N}} \int_V \dots \int_V dq_1 \dots dq_{3N} \left\{ \sum_{i=1}^{N-1} \sum_{j=i+1}^N \left( e^{-\beta U_{i,j}(r_{i,j})} - 1 \right) \right\} \end{aligned} \quad (22)$$

Here,  $S_2$  depends only on the potential between an isolated pair of molecules. If you know the type of potential and the temperature  $S_2$  can be calculated by using the following formula.

$$S_2(\beta) = \frac{1}{\alpha_2^6} \int_V \int_V dq_{3i-2} dq_{3i-1} dq_{3i} dq_{3j-2} dq_{3j-1} dq_{3j} \left( e^{-\beta U_{i,j}(r_{i,j})} - 1 \right) \quad (23)$$

The integrand of this term depends on six integration variables,  $q_{3i-2}$ ,  $q_{3i-1}$ ,  $q_{3i}$  and  $q_{3j-2}$ ,  $q_{3j-1}$ ,  $q_{3j}$ , only. The remaining integration variables are independent of the integrand hence the result is obtained as a constant that is, this integration process produce  $\tilde{V}^{N-2}$  with respect to remaining integration variables. So any particle pair independent of the other particles behaviour is characterized by the sixfold volume integral.

The higher terms can be obtained by using the same philosophy. So the third term, which includes the multiplication of two different kinds of Mayer functions, is concerned to depend on triple interaction. But this time there are ninefold integrals to be evaluated to get the third term, that is, the number of integrations is related with the clustering level. As clustering level increases the number of integrals to be evaluated increases. In order to determine the number of integrals, which belong to the clustering level, the number of interacting particles within clusters is multiplied by 3.

Finally, when we take the relations (13) and (20) into consideration, the canonical partition function given in (1) can be rewritten as follows

$$Z(\beta) \equiv \frac{1}{N!} \left( \frac{2\pi\mu\alpha_2^2}{\hbar^2\beta} \right)^{\frac{3N}{2}} \left[ \tilde{V}^N + \frac{N(N-1)}{2} \tilde{V}^{N-2} S_2(\beta) + \dots \right] \quad (24)$$

The thermal de Broglie wavelength is defined in physics as

$$\lambda_{th} = \frac{\hbar}{\sqrt{2\pi\mu k_B T}} \quad (25)$$

If we take  $\alpha_2$  parameter equal to Broglie wavelength, that is,  $\alpha_2 = \lambda_{th}$ , then the following final form is obtained as the partition function. This function is the one that is used to evaluate the free energy.

$$Z(\beta) = \frac{1}{N!} Z_{pos}(\beta) = \frac{1}{N!} \left[ \tilde{V}^N + \frac{N(N-1)}{2} \tilde{V}^{N-2} S_2(\beta) + \dots \right] \quad (26)$$

Now, we need to determine  $S_2(\beta)$  and higher terms appearing in the above relation to get the result. However, as stated before,  $S_2(\beta)$  has six-tuple integrals while  $S_3(\beta)$  has nine-tuple integrals and so on. To cope with these multiple integrals, we can use the Fluctuation Free Integration method[8, 9, 10, 11]. The details of this method are given in the followings.

## 2.2 Fluctuation Free Integration Method

The Fluctuationlessness Approximation method is a recently developed algorithm including the philosophy given by the Fluctuationlessness Theorem. The method is used to obtain numerical solutions of algebraic methods and to compose very rapidly converging univariate numerical integral schemes. These schemes can be used for both univariate and multivariate cases in various scientific and engineering research areas[4, 5, 12].

The theorem considers two infinite dimensional matrix representations of two different operators  $\hat{f}$  and  $\hat{x}$  in infinite dimensional Hilbert space  $\mathcal{H}$ . If a finite dimensional subspace  $\mathcal{H}_n$  is used to construct a matrix representation of the operators, then an approximation is obtained as follows

$$\mathcal{M}_{\hat{f}}^{(n)} \approx f(\mathbf{X}^{(n)}) \quad (27)$$



where  $n$  corresponds to the truncation level in the Hilbert space. Here,  $\mathcal{M}_{\hat{f}}^{(n)}$  is the  $n$  dimensional matrix representation of the operator  $\hat{f}$  and  $\mathbf{X}^{(n)}$  stands for the  $n$  dimensional matrix representation of  $\hat{x}$ . The explicit form of these matrices are given by using the inner product relations defined in Hilbert space as

$$\begin{aligned}\mathbf{X}^{(n)} &\equiv \begin{bmatrix} (u_1, \hat{x}u_1) & \cdots & (u_1, \hat{x}u_n) \\ \vdots & \ddots & \vdots \\ (u_n, \hat{x}u_1) & \cdots & (u_n, \hat{x}u_n) \end{bmatrix} \\ \mathcal{M}_{\hat{f}}^{(n)} &\equiv \begin{bmatrix} (u_1, \hat{f}u_1) & \cdots & (u_1, \hat{f}u_n) \\ \vdots & \ddots & \vdots \\ (u_n, \hat{f}u_1) & \cdots & (u_n, \hat{f}u_n) \end{bmatrix}\end{aligned}\quad (28)$$

where  $u_j(x)$  ( $1 \leq j \leq n$ ) form the orthonormal basis set which spans the  $n$  dimensional Hilbert subspace  $\mathcal{H}_n$ . The Fluctuationlessness Theorem says that for a proper choice of the basis functions and for  $n \rightarrow \infty$  the approximation in (27) turns into an equality. Thus, the matrix representation of an algebraic operator which multiplies its arguments by a scalar univariate function is identical to the image of the independent variable's matrix representation over the same space through the same basis set under that univariate function[8, 9, 10, 11, 12].

The elements of the above matrices are evaluated by using the following inner product of two functions from  $\mathcal{H}$  under a given weight function

$$(g, h) = \int_a^b dx w(x) g(x) h(x), \quad g(x), h(x) \in \mathcal{H} \quad (29)$$

The Fluctuationlessness Approximation method uses an arbitrary orthonormal set in its algorithm. In contrast to some other methods which need a polynomial basis set to do so. To construct an orthonormal basis set, we can use a method such as the Gram-Schmidt orthonormalization method[13].

The dimension of the matrix  $\mathbf{X}^{(n)}$  is independent of the considered problem. This dimension is directly related with the acceptable approximation quality level of the integration result. Hence, the higher  $n$  value the better the approximation. In addition, the elements of this matrix depend on only the basis set chosen. This means that the matrix structure is also independent of the considered problem. Thus, this matrix is named as a ‘‘Universal Matrix’’. If we construct this matrix for a certain  $n$  value, we can use that matrix for all problems we are dealing with.

Now, the question is how  $\mathbf{X}^{(n)}$  will be used in an algorithm to evaluate an integral. This algorithm is called ‘‘Fluctuation Free Integration Method’’[4, 5]. To explain the details of this algorithm, the following steps are listed for the integral evaluation of a univariate function under unit interval and unit weight.

Step 1. Let  $f(x)$  be an analytic function and  $\mathcal{I}$  denotes the integral of this function in the interval  $[0, 1]$ .

$$\mathcal{I} = \int_0^1 dx f(x) \quad (30)$$

- Step 2. Specify the truncation level which is denoted by  $n > 0$ . In many cases  $n = 3$  is a typical choice and provides already very good results.
- Step 3. Compose the orthonormal basis set including  $n$  number of elements. Although the Fluctuationlessness Approximation Method has no limitations on basis set selection process, there is an important exception on this flexibility in the Fluctuation Free Integration Method, that is, the first element of the selected basis set should be 1. In this work,  $u_i$  basis functions are constructed from the polynomial basis set  $1, x, x^2, \dots, x^n$  and we can make this set orthonormal by applying Gram-Schmidt orthonormalization procedure on the selected set.
- Step 4. Calculate the matrix  $\mathbf{X}^{(n)}$ , which is  $n$  dimensional matrix representation of independent variable  $x$ , by using the relation given in (28).
- Step 5. Determine eigenvalues,  $\lambda_i$  and corresponding eigenvectors,  $\boldsymbol{\xi}_i$  ( $1 \leq i \leq n$ ) of  $\mathbf{X}^{(n)}$ . The spectrum of the matrix  $\mathbf{X}^{(n)}$  is on the real axis since it is symmetric matrix. Furthermore, the spectrum must be confined into the interval  $[0, 1]$  because the basis set which is used for constructing the matrix representation of  $x$  is orthonormal over the interval  $[0, 1]$  and all eigenvalues are discrete. So, we can write the following relations

$$\mathbf{X}^{(n)} = \sum_{i=1}^n \lambda_i \boldsymbol{\xi}_i \boldsymbol{\xi}_i^T, \quad f(\mathbf{X}^{(n)}) = \sum_{i=1}^n f(\lambda_i) \boldsymbol{\xi}_i \boldsymbol{\xi}_i^T \quad (31)$$

- Step 6. Since  $u_1(x) = 1$  as stated in Step 3, this helps us to use a tricky idea in the evaluation process of the integral given in (30) as

$$\begin{aligned} \mathcal{I} &= \int_0^1 dx f(x) = \int_0^1 dx u_1(x) f(x) u_1(x) = \left( u_1, \hat{f} u_1 \right) \\ &= \mathbf{e}_1^T \mathcal{M}_{\hat{f}}^{(n)} \mathbf{e}_1 \end{aligned} \quad (32)$$

where  $\mathbf{e}_1$  is an  $n$  dimensional unit Cartesian vector whose only nonzero element its first element which is 1.

- Step 7. Take the Fluctuationlessness Theorem into consideration to rewrite the last expression of the relation given in (32) as

$$\mathbf{e}_1^T \mathcal{M}_{\hat{f}}^{(n)} \mathbf{e}_1 \approx \mathbf{e}_1^T f(\mathbf{X}^{(n)}) \mathbf{e}_1 \quad (33)$$

- Step 8. Use the spectral decomposition of  $\mathbf{X}^{(n)}$  given in (31) to rewrite the above expression.

$$\mathbf{e}_1^T f(\mathbf{X}^{(n)}) \mathbf{e}_1 = \mathbf{e}_1^T \sum_{i=1}^n f(\lambda_i) \boldsymbol{\xi}_i \boldsymbol{\xi}_i^T \mathbf{e}_1 = \sum_{i=1}^n f(\lambda_i) (\mathbf{e}_1^T \boldsymbol{\xi}_i)^2 \quad (34)$$

- Step 9. Obtain the approximate result of the integral  $\mathcal{I}$  given in (30) by using the relations (32) and (34).

$$\mathcal{I} = \int_0^1 dx f(x) \approx \sum_{i=1}^n f(\lambda_i) (\mathbf{e}_1^T \boldsymbol{\xi}_i)^2 \quad (35)$$

These steps are given for evaluation of the integral of a univariate function. This method can be extended to find the result of integral of multivariate functions[4, 5].

The methods given in this section are used to develop a new algorithm for free energy evaluation process of water molecules in gas form. In this work, we are considering a closed system having water molecules in which we have atom pairs interacting with each other through a given potential energy. The definition of this mentioned potential energy is given in the next section.

### 3 Potential Energy Definition

The free energy of a molecular system including  $N$  particles, which is confined into a physical volume  $V$ , is evaluated through the following Helmholtz Free Energy relation

$$A = -\frac{1}{\beta} \ln(Z(\beta)) \quad (36)$$

where  $Z(\beta)$  is partition function whose relation is given in (1). When we take the relation (26) into consideration, the above relation can be rewritten in terms of position interacted partition function,  $Z_{pos}(\beta)$ , as

$$A = -\frac{1}{\beta} \ln\left(\frac{1}{N!} Z_{pos}(\beta)\right) \quad (37)$$

The explicit form of  $Z_{pos}(\beta)$  is given in (20). We use the first two components of the expansion of  $Z_{pos}(\beta)$  in our free energy evaluations not to increase the computational complexity since the other components have nine-tuple and higher tuple integrals. If we take the relations given in (21) and (22) into consideration, we obtain the following relation for free energy evaluation process

$$A = -\frac{1}{\beta} \ln\left(\frac{1}{N!} \left[ \left(\frac{V}{\alpha_2^3}\right)^N + \frac{N(N-1)}{2} \tilde{V}^{N-2} S_2(\beta) \right]\right) \quad (38)$$

where the structure of  $S_2(\beta)$  is given in (23).

It is obvious that to evaluate the free energy through position interacted partition function, we need to define a binary interaction potential. The potential energy chosen in this work is as follows

$$\begin{aligned} U_{i,j}(r_{i,j}) &= \sum_{bonds} K_r(r_{i,j} - r_{eq})^2 + \sum_{angles} K_\theta(\theta_{i,j} - \theta_{eq})^2 \\ &+ \sum_{i < j} \left[ \frac{a_{i,j}}{r_{i,j}^{12}} - \frac{b_{i,j}}{r_{i,j}^6} + \frac{q_i q_j}{\epsilon r_{i,j}} \right], \quad i, j = 1, \dots, N \end{aligned} \quad (39)$$

where  $r_{eq}$  and  $\theta_{eq}$  are equilibration structural parameters while  $K_r$  and  $K_\theta$  are force constants. In addition,  $q_i$  and  $q_j$  are the electric charges in coulombs carried by  $i$ -th and  $j$ -th charges respectively and  $\epsilon$  is the electrical permittivity of space.  $r_{i,j}$ s are the distances and their mathematical expressions are given in (15).

As it is seen, the structure of the potential is composed of bond, angle, Lennard-Jones and Coulomb potentials.

## 4 A Novel Free Energy Evaluation Algorithm

This section covers the proposed algorithm to calculate the free energy of water molecules in gas phase. We are dealing with pairwise interactions between the identical particles in this work. This means that we can write the distance between the first and the second particles as follows and assume that this will be the same for the distances between other pairs

$$r_{1,2} = \left[ (x_4 - x_1)^2 + (x_5 - x_2)^2 + (x_6 - x_3)^2 \right]^{\frac{1}{2}} \quad (40)$$

Moreover, the abovementioned assumption allows us to rewrite the relation given in (23) as

$$S_2(\beta) = \frac{1}{\alpha_2^6} \int_V dx_1 dx_2 dx_3 \int_V dx_4 dx_5 dx_6 (e^{-\beta U_{1,2}} - 1) \quad (41)$$

where the dummy integration variables are changed to  $x_1, x_2, x_3$  for the first particle and  $x_4, x_5, x_6$  for the second particle from  $qs$  not to cause any confusion between the electric charges in coulombs.

As it is observed in (39) and (41), the potentials evaluated through bond and non-bond interactions appear as one factor of the exponential function of  $S_2(\beta)$ . Here, we have 6-tuple integrals and the structure inside this integral coming from the potentials is too complicated to evaluate these integrals either analytically or numerically. This urges us to bypass this complex structure of the considered integral evaluations through a new algorithm.

When we have identical and distinguishable atoms in our problem, then the total number of the atoms corresponds to  $N$  value appearing in the relation of free energy evaluation given in (38). We may assume that the pairwise interactions will occur through any potential for this case and this will result in a more simple model for evaluating the free energy whenever we are able to evaluate 6-tuple integrals appearing in (41). However, we are working on water molecules which have 2 hydrogen and 1 oxygen atoms. So, this structure is indistinguishable and includes different potentials that affects its free energy.

Here, we construct our algorithm first by looking at the total number of the atoms of water molecules and the possible pairwise interactions that will occur between their atoms. In this sense, to build the steps and the related relations of our proposed algorithm, we will discuss a 4 water molecules system and then extend the resulting relations to the general case. In this algorithm  $m$  will represent the total number of water molecules appearing in the mentioned general case. Our test system has 8 hydrogen and 4 oxygen atoms. There are O-H bond, H-O-H angle, Lennard-Jones and Coulomb potentials to be evaluated to get the final free energy value. For this purpose, we need to specify the pairwise interactions appearing between the mentioned atoms. These interactions are listed as follows:

1. The first pairwise interaction is between the oxygen atoms of distinguishable water molecules. There are 4 oxygen atoms. This means that we have totally  $m$  number of oxygen atoms that have pairwise interactions in the general case. This can be given as

$$N = m \quad (42)$$

where we have to evaluate the following Lennard-Jones and Coulomb potentials

$$U_{1,2} = \frac{a_{1,2}}{r_{1,2}^{12}} - \frac{b_{1,2}}{r_{1,2}^6} + \frac{q_1 q_2}{\epsilon r_{1,2}} \quad (43)$$

2. The second pairwise interaction is between any hydrogen atom and the other hydrogen atoms of the molecules that do not include that hydrogen atom. There are 7 atoms satisfying this case. Hence, in general, we can write the following relation

$$N = 2m - 1 \quad (44)$$

Similar with the previous case given in (43), Lennard-Jones and Coulomb potentials should be evaluated for this case.

3. The third pairwise interaction is between the hydrogen and oxygen atoms that have no bond with each other. That is, these atoms interact through Lennard-Jones and Coulomb potentials and the potential for this case can be evaluated by using (43). There are 7 atoms like that. So, we can formulate this case as

$$N = 2m - 1 \quad (45)$$

4. There is also a bond potential between the hydrogen and oxygen atoms. Each water molecule includes two hydrogen and one oxygen atoms interacting through bond potential

$$U_{1,2} = K_r (r_{1,2} - r_{eq})^2 \quad (46)$$

and this results in 3 atoms in interaction for all cases, that is,

$$N = 3 \quad (47)$$

5. The H-O-H angle potential is the next step in evaluating the free energy value. However, instead of taking this angle potential into consideration, we evaluate the bond energy between the hydrogen atoms of the same water molecule. Because, the total number of variables representing the H-O-H angle do not match the total number of integrands needed to be evaluated appearing in  $S_2(\beta)$  as it is seen in (41). Since there are 2 hydrogen atoms for all cases, we can write the following

$$N = 2 \quad (48)$$

To this end, the relation given in (46) is used for the potential evaluation process in this case.

The philosophy given up to here allows us to write the following relation for free energy evaluation of water molecule in gas phase.

$$A = A_1 + A_2 + A_3 + A_4 + A_5 \quad (49)$$

Here, to evaluate the  $A_1$ ,  $A_2$ ,  $A_3$ ,  $A_4$  and  $A_5$ , we use the steps 1,2,3,4 and 5 respectively. In this sense  $A_1$  can be written as

$$A_1 = -\frac{1}{\beta} \ln \left( \frac{1}{m!} \left[ \left( \frac{V}{\alpha_2^3} \right)^m + \frac{m(m-1)}{2} \tilde{V}^{m-2} \frac{1}{\alpha_2^6} \mathcal{K}_1 \right] \right) \quad (50)$$

where

$$\mathcal{K}_1 = \int_V dx_1 dx_2 dx_3 \int_V dx_4 dx_5 dx_6 \left( e^{-\beta \left[ \frac{a_{1,2}}{r_{1,2}^2} - \frac{b_{1,2}}{r_{1,2}^6} + \frac{q_1 q_2}{\epsilon r_{1,2}} \right]} - 1 \right) \quad (51)$$

Since the second and the third steps include the same potentials with the first step,  $A_2$  and  $A_3$  can be written as follows in terms of  $\mathcal{K}_1$

$$\begin{aligned} A_2 &= -\frac{1}{\beta} \ln \left( \frac{1}{(2m-1)!} \left[ \left( \frac{V}{\alpha_2^3} \right)^{2m-1} + \frac{(2m-1)(2m-2)}{2} \tilde{V}^{2m-3} \frac{1}{\alpha_2^6} \mathcal{K}_1 \right] \right) \\ A_3 &= -\frac{1}{\beta} \ln \left( \frac{1}{(2m-1)!} \left[ \left( \frac{V}{\alpha_2^3} \right)^{2m-1} + \frac{(2m-1)(2m-2)}{2} \tilde{V}^{2m-3} \frac{1}{\alpha_2^6} \mathcal{K}_1 \right] \right) \end{aligned} \quad (52)$$

The fourth and the fifth steps urge us to write the following relations for  $A_4$  and  $A_5$

$$\begin{aligned} A_4 &= -\frac{1}{\beta} \ln \left( \frac{1}{3!} \left[ \left( \frac{V}{\alpha_2^3} \right)^3 + 3\tilde{V} \frac{1}{\alpha_2^6} \mathcal{K}_2 \right] \right) \\ A_5 &= -\frac{1}{\beta} \ln \left( \frac{1}{2!} \left[ \left( \frac{V}{\alpha_2^3} \right)^2 + \frac{1}{\alpha_2^6} \mathcal{K}_2 \right] \right) \end{aligned} \quad (53)$$

where

$$\mathcal{K}_2 = \int_V dx_1 dx_2 dx_3 \int_V dx_4 dx_5 dx_6 \left( e^{-\beta K_r (r_{1,2} - r_{e_q})^2} - 1 \right) \quad (54)$$

It is clear that our new structure for evaluating the free energy is now more simple than the structure obtained in (36) since we have discrete integral structures to evaluate instead of a single one with a complicated integrand. However, we have still 6-tuple integrals in  $\mathcal{K}_1$  and  $\mathcal{K}_2$  to be evaluated for getting the resulting free energy value. At this point, it is time to apply the steps of the Fluctuation Free Integration method to overcome our multiple integral evaluation problem. For this purpose, the first step is to apply this method to  $\mathcal{K}_1$ . So we can start to calculate these integrals beginning from the last one. In this sense, we should take (35) into consideration. This results in the following relation

$$\mathcal{K}_1 \approx \int_V dx_1 dx_2 dx_3 \int dx_4 \int dx_5 \sum_{i_6=1}^n (\mathbf{e}_1^T \boldsymbol{\xi}_{i_6})^2 \left( e^{-\beta \gamma_1(x_1, \dots, x_5, \lambda_6^{(i_6)})} - 1 \right) \quad (55)$$

where

$$\begin{aligned} \gamma_1(x_1, \dots, x_5, \lambda_6^{(i_6)}) &\equiv \frac{a_{1,2}}{\left[ (x_4 - x_1)^2 + (x_5 - x_2)^2 + (\lambda_6^{(i_6)} - x_3)^2 \right]^6} \\ &- \frac{b_{1,2}}{\left[ (x_4 - x_1)^2 + (x_5 - x_2)^2 + (\lambda_6^{(i_6)} - x_3)^2 \right]^3} + \frac{q_1 q_2}{\epsilon \left[ (x_4 - x_1)^2 + (x_5 - x_2)^2 + (\lambda_6^{(i_6)} - x_3)^2 \right]^{\frac{1}{2}}} \end{aligned} \quad (56)$$

Tabelle 1: Free energy values for different numbers of water molecules.

# of Water Molecules	Free Energy Value (joule)
100	-394051.3185
500	-1947505.398
1000	-3881239.13
5000	-19264455.07
7000	-26930590.33
10000	-38412655.87

and  $\lambda_6^{(i_6)}$  and  $\boldsymbol{\xi}_{i_6}$  stand for the  $i_6$ -th eigenvalue and the corresponding eigenvector of the  $n \times n$  dimensional matrix representation of the 6-th independent variable respectively. The remaining integrals can also be calculated in the similar way depending on the relation given in (35) and the following final structure is obtained for  $\mathcal{K}_1$

$$\mathcal{K}_1 \approx \sum_{i_1=1}^n \cdots \sum_{i_6=1}^n \left[ \prod_{j=1}^6 \left( \mathbf{e}_1^T \boldsymbol{\xi}_{i_j} \right)^2 \right] \left( e^{-\beta \gamma_1(\lambda_1^{(i_1)}, \dots, \lambda_6^{(i_6)})} - 1 \right) \quad (57)$$

The next step is to determine  $\mathcal{K}_2$  appearing in  $A_4$  and  $A_5$ . If we use the relations given in (35) and (54), then we obtain the following relation in  $\mathcal{K}_2$  determination process

$$\mathcal{K}_2 \approx \int_V dx_1 dx_2 dx_3 \int dx_4 \int dx_5 \sum_{i_6=1}^n \left( \mathbf{e}_1^T \boldsymbol{\xi}_{i_6} \right)^2 \left( e^{-\beta K_r \gamma_2(x_1, \dots, x_5, \lambda_6^{(i_6)})} - 1 \right) \quad (58)$$

where

$$\gamma_2(x_1, \dots, x_5, \lambda_6^{(i_6)}) \equiv \left[ \left[ (x_4 - x_1)^2 + (x_5 - x_2)^2 + \left( \lambda_6^{(i_6)} - x_3 \right)^2 \right]^{\frac{1}{2}} - r_{eq} \right]^2 \quad (59)$$

The remaining integrals of  $\mathcal{K}_2$  can be evaluated as follows when we apply the Fluctuation Free Integration method

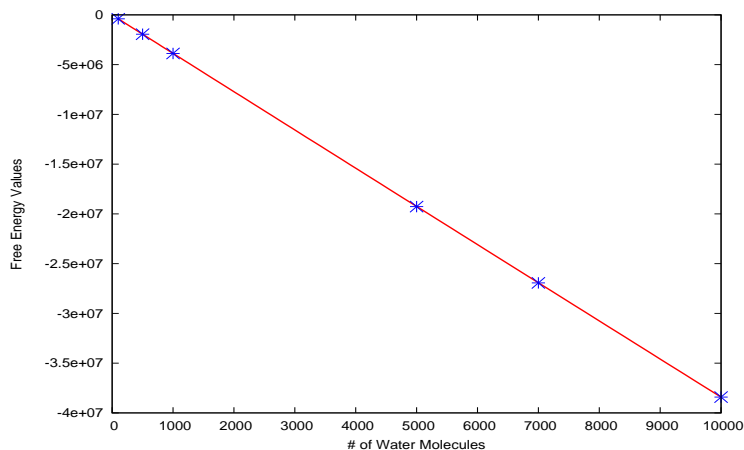
$$\mathcal{K}_2 \approx \sum_{i_1=1}^n \cdots \sum_{i_6=1}^n \left[ \prod_{j=1}^6 \left( \mathbf{e}_1^T \boldsymbol{\xi}_{i_j} \right)^2 \right] \left( e^{-\beta K_r \gamma_2(\lambda_1^{(i_1)}, \dots, \lambda_6^{(i_6)})} - 1 \right) \quad (60)$$

Finally, when we insert the result of (57) into (50) and (52) we obtain  $A_1$ ,  $A_2$  and  $A_3$ . To get  $A_4$  and  $A_5$ , we use (60) inside the relations given in (53). At the end, the results obtained for  $A_s$  are inserted into the relation (49) to determine the free energy value of the water molecule in gas phase.

## 5 Results and Discussion

We proposed a method to calculate the free energy of water molecules in gas phase by using a deterministic way. To achieve this, we developed a new algorithm by using two different methods from the literature. One of them is Mayer

Abbildung 1: Free energy values for different numbers of water molecules.



Cluster Expansion method. To benefit from this method, at first, we have rewritten the Helmholtz Free Energy relation in terms of the partition function. After that, we used the Mayer Cluster Expansion method to construct the power series expansion of the partition function. Under the binary interaction assumption, we composed an algorithm by clustering the water molecules. To this end, some equations were obtained which have six tuple integrals and we solved these equations by using the Fluctuation Free Integration method which is the second method coming from the literature.

The Free energy results obtained for different numbers of water molecules are given in Table 1. The evaluations were done under the following values  $T = 395.15$  and  $\hbar = 1.054571726 * 10^{-34} j.s$  where  $T$  and  $\hbar$  stands for temperature and reduced Planck constant respectively. Figure 1 shows the characteristic of the results given in Table 1 with respect to the different numbers of water molecules

## Acknowledgement

The first author is grateful to Professor Metin Demiralp for his valuable comments and to The Scientific & Technological Research Council of Turkey for its support.

## Literatur

- [1] C. Chipot, M.S. Shell, A. Pohorille, Free Energy Calculations – Theory and Applications, SpringerVerlag, Berlin, 2007.
- [2] M. Klimm, A. Bujotzek, M. Weber, Direct Reweighting Strategies in Conformation Dynamics, MATCH Commun. Math. Comput. Chem. **65**, 333–346 (2011).



- [3] J.E. Mayer, E. Montroll, Molecular distributions, *J. Chem. Phys.* **9** (1941) 2–16.
- [4] B. Tunga, M. Demiralp, Fluctuation Free Multivariate Integration Based Logarithmic HDMR in Multivariate Function Representation, *J. Math. Chem.* **49** (2011) 894–909 .
- [5] B. Tunga, M. Demiralp, Hybrid HDMR Method with an optimized hybridity parameter in multivariate function representation, *J. Math. Chem.* **50** (2012) 2223–2238.
- [6] J.E. Mayer, M.G. Mayer, *Statistical Mechanics*, John Wiley and Sons, 1940.
- [7] J.E. Mayer, Contribution to Statistical Mechanics, *J. Chem. Phys.* **10** (1942) 629–643.
- [8] M. Demiralp, A New Fluctuation Expansion Based Method for the Univariate Numerical Integration Under Gaussian Weights. WSEAS-2005 Proceedings, WSEAS 8-th International Conference on Applied Mathematics, 16-18 December, 2005, Tenerife, Spain, 68–73.
- [9] M. Demiralp, Convergence Issues in the Gaussian Weighted Multidimensional Fluctuation Expansion for the Univariate Numerical Integration. WSEAS Trans. Math. **4** (2005) 486–492.
- [10] M. Demiralp, Fluctuationlessness Theorem to Approximate Univariate Functions Matrix Representations. WSEAS Trans. Math. **8**(6) (2009) 258–267.
- [11] B. Tunga, M. Demiralp, Constancy Maximization Based Weight Optimization in High Dimensional Model Representation, *Numer. Algor.* **52** (2009) 435–459.
- [12] E. Gürvit, N.A. Baykara, M. Demiralp, Numerical Integration of Bivariate Functions Using Fluctuationlessness Theorem with a Trigonometric Basis Function to Deal with Highly Oscillatory Functions, In proceedings for the Applied Computing Conference (ACC'09 WSEAS International Conference), 28-30 September, 2009, Vouliagmeni, Greece, 400–405.
- [13] K. Hoffman, R. Kunze, *Linear Algebra 2ed.*, Prentice Hall Inc., 1971.