

Molecular Dynamics and the Water Molecule

An Introduction to Molecular Dynamics for Physical Chemistry Students

Enrique Saiz and M. Pilar Tarazona

Departamento de Química Física, Universidad de Alcalá, 28871 Alcalá de Henares, Spain

In spite of the importance of molecular dynamics (MD) procedures as research tools for simulating complex systems such as fluids, polymer chains, and chemical reactions (1, 2), the topic is generally not discussed in physical chemistry textbooks; and although there are excellent books (3) devoted to the topic, they are aimed at graduate students or researchers more than at students. One reason for this is that, to understand the method and its possibilities, one must have ample knowledge of the system (a polymer chain, for instance) that is to be studied. This problem can be surmounted by applying the molecular dynamic method to simple systems that are adequate for introducing students to this computer simulation procedure, although they do not have research interest because they are very well known.

With this aim, we developed a simple general approach to molecular dynamics that can be used in physical chemistry courses. It can be included in various parts of the physical chemistry program, such as those devoted to liquids, macromolecules, or transport phenomena. If PCs are available students can do the calculations by themselves, but this is not necessary. The lesson starts with some general notions about molecular dynamics and continues by applying molecular dynamics first to an isolated atom, then to a couple of interacting atoms, and finally to a very simple and very well-known molecule, the water molecule. Of course the results obtained for this molecule are already known, but it is precisely this fact that permits students to get an insight into MD procedures and possible applications.

A molecular dynamic calculation consists of analyzing the evolution of the sample with time. In the case of a molecule, the atoms that form it are continuously moving, the bonds are vibrating, the angles are bending, the whole molecule is rotating, etc. If we were able to follow the exact position of every atom as a function of time—for instance, by taking periodic pictures of the molecule after each time increment δ —we could build a table with the coordinates of each atom in the molecule as a function of time and use it to calculate properties that depend on the conformation of the molecule—that is, on the relative positions of the atoms. For instance, to calculate the average distance between two atoms of the molecule, one must employ the coordinates contained in the table to compute the distance and then perform the time average of all values thus obtained. The same procedure could be used for any other bond length, interatomic distance, valence or rotational angle, etc.

Certainly this example seems to be quite easy. A real MD calculation is a bit more complicated than simply writing down the atomic coordinates for the molecule, since there are two important differences between an actual MD analysis and this trivial example: the size of the table and the procedure for computing the atomic coordinates.

The table of coordinates in an actual MD calculation may be very large. It normally contains several thousand lines obtained by computing coordinates with very small time increments and trying to cover a relatively large time interval. In addition, every atom in the molecule or particle in the system produces one column with three coordinates.

Thus, the table becomes too large to be handled. However, this inconvenience only means that the whole analysis should be performed by a computer, either during the actual MD calculation or, alternatively, by saving the table on disk for subsequent analysis employing appropriate software.

As for the computation of the atomic coordinates, since it is obvious that we can not photograph the molecule, the classical trajectories of the atoms are calculated by integrating Newton's equation of motion

$$\mathbf{F} = m\mathbf{a} \quad (1)$$

for the system under study.

System Formed by One Isolated Particle

Let us assume, for the first example, that our system is formed by just one particle having a mass, m , such as the one represented in Figure 1, which is isolated from the outside and is not subject to any kind of interaction. A physical picture of such a system will be a single atom moving in free vacuum. This particle will have a certain kinetic energy, E_k , which will be determined by the temperature (having $kT/2$ for each translational degree of freedom) and which, in turn, will govern the velocity, \mathbf{v} , at which this particle is moving, according to the equation:

$$E_k = \frac{3}{2} kT = \frac{1}{2} m\mathbf{v}^2 \quad (2)$$

If the temperature of the system is absolute zero ($T = 0$ K), then $\mathbf{v} = 0$, and the particle would be motionless so that the coordinates will not change with time, and we have finished the MD analysis. However, let us assume that this is not the case. Instead we suppose that the temperature is $T > 0$ and consequently the particle is moving with a certain velocity $\mathbf{v} \neq 0$. But if there are no interactions, there will also be no forces and no accelerations. Therefore \mathbf{v} will remain constant, which means that the movement of the particle will be uniform. We could define an arbitrary coordinate system with the origin at the initial position of the particle and the x axis along the direction of $\hat{\mathbf{v}}$. Then the position and velocity of the particle at initial time t would be

$$\left. \begin{aligned} x(t) = y(t) = z(t) &= 0 \\ v_x(t) &= v \neq 0 \\ v_y(t) = v_z(t) &= 0 \end{aligned} \right\} \quad (3)$$

Since the movement is uniform, it is trivial to integrate the equation of motion to compute the position of the particle as function of time:

$$\left. \begin{aligned} x(t + \delta) &= x(t) + v\delta \\ y(t + \delta) &= y(t) = 0 \\ z(t + \delta) &= z(t) = 0 \end{aligned} \right\} \quad (4)$$

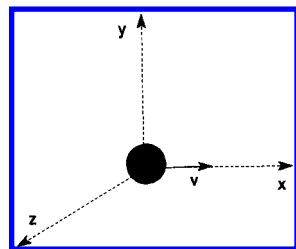


Figure 1. A single particle subject to no interactions and moving with a thermal velocity \mathbf{v} . A cartesian frame is defined with the origin located at the initial position of the particle and the x axis having the direction of \mathbf{v} .

so that the table of coordinates will look like Table 1. Certainly it is not difficult to obtain this even if we do not have big computers. However, the negative aspect is that the system is too simple to be of interest.

System Formed by Two Interacting Particles

Let us complicate the system by assuming that now we have two particles, which are not linked by any kind of bond and which, at the initial time t , are separated by a distance r , as indicated in Figure 2. The system is isolated and does not interact with the outside. The kinetic energy is

$$E_k = \frac{6}{2} kT = \frac{1}{2} (mv_1^2 + mv_2^2) \quad (5)$$

We have assumed that both particles have the same mass m ; otherwise we should write the equation with m_1 and m_2 .

The difference with the previous example is that now, even if the system is isolated, each particle will interact with the other, producing a potential energy that could be represented by a function such as that of Lennard-Jones:

$$E_p = \frac{a}{r^{12}} - \frac{b}{r^6} \quad (6)$$

The effect of this potential is that the two particles repel each other when they are close and attract each other when they are far apart. This interaction can be described as a force given by the gradient of the potential that will determine the acceleration to which both particles are subjected.

$$\mathbf{F}_r = -\nabla E = \frac{dE}{d\mathbf{r}} = \left(\frac{12a}{r^{14}} - \frac{6b}{r^8} \right) \mathbf{r} = m\mathbf{a} = m \left(\frac{d^2 \mathbf{r}}{dt^2} \right) \quad (7)$$

This vectorial equation represents a set of six differential equations, one for each coordinate and each particle within the system. Thus we can select an arbitrary reference system—for instance, taking the origin at the initial position of one of the particles and the x axis along its ini-

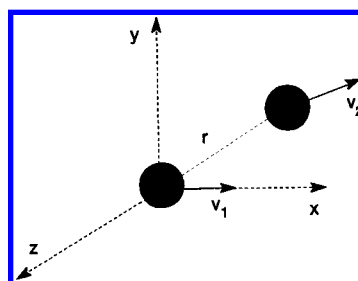


Figure 2. A system formed by two particles that are not bonded but interact according to a Lennard-Jones potential. The coordinate system is defined with the origin at the initial position of one particle and the x axis along the initial velocity of the same particle.

tial velocity as it is represented in Figure 2. Then, the force acting over the first particle will be

$$\mathbf{F}_1 = (F_x, F_y, F_z) \quad (8)$$

while the force over the second particle will be of the same magnitude but with the opposite direction.

Let us assume that we know the initial positions and velocities of both particles. In practice, these initial values are arbitrarily assigned, since the final results are quite insensitive to the starting point. Then, we can compute the forces acting over the particles and the accelerations that these forces produce, and integrate the equation of motion (eq 7) to obtain their position as a function of time.

However, the equation is too complicated to allow an analytical solution. The problem is that the position of the particles determines the forces and therefore the accelerations, which will then modify the velocity and the next position of the particles. Thus the movement is neither uniform, nor uniformly accelerated, nor any other kind of simple motion and therefore numerical procedures are required to integrate its equation.

The procedure most frequently used for this kind of calculation employs the same approach used for the numerical integration of any general function $f(x)$: namely, to replace the actual function by a stair function obtained by taking the values $f(x_0 + i\delta)$ at regularly spaced positions of x , with δ being the interval of integration and i representing an integer number. Let us assume that the actual acceleration changes with time, as indicated by the solid curve on Figure 3. The interval δ is divided into halves and the value of the acceleration at time $t + i\delta$ is assumed to be constant from $t + (i - 1/2)\delta$ to $t + (i + 1/2)\delta$. Thus, the actual function \mathbf{a} is replaced by the stair indicated in the figure. Since the acceleration remains constant between two consecutive half intervals, the velocity will change linearly, with a slope of $\mathbf{a}(t + i\delta)$, from $\mathbf{v}[t + (i - 1/2)\delta]$ to $\mathbf{v}[t + (i + 1/2)\delta]$ and this last value could be used as an average for the interval $t + i\delta$, $t + (i + 1)\delta$, so that the movement in this interval will be uniform and the position will change linearly with a slope of $\mathbf{v}[t + (i + 1/2)\delta]$ from $\mathbf{r}(t + i\delta)$ to $\mathbf{r}[t + (i + 1)\delta]$.

Writing down these variations of \mathbf{v} and \mathbf{r} , one obtains the leapfrog algorithm:

$$\begin{aligned} \mathbf{v}[t + (i + 1/2)\delta] &= \mathbf{v}[t + (i - 1/2)\delta] + \mathbf{a}(t + i\delta)\delta \\ \mathbf{r}[t + (i + 1)\delta] &= \mathbf{r}(t + i\delta) + \mathbf{v}[t + (i + 1/2)\delta]\delta \end{aligned} \quad (9)$$

Thus at the first step ($i = 0$), the initial positions $\mathbf{r}(t)$ are used to compute the forces $\mathbf{F}(t)$ and accelerations $\mathbf{a}(t)$. These, together with the initial velocities that are assigned to the time $t - \delta/2$, are employed to compute the new velocities at half-step $\mathbf{v}(t + \delta/2)$ with the first equation of the al-

Table 1. Coordinates of an Isolated Particle as a Function of Time^a

Time	x	y	z
t	0	0	0
$t + \delta$	δv	0	0
\vdots	\vdots	\vdots	\vdots
$t + n\delta$	$n\delta v$	0	0

^aThe reference frame was arbitrarily centered at the initial position of the particle with the x axis having the direction of the velocity, \mathbf{v} .

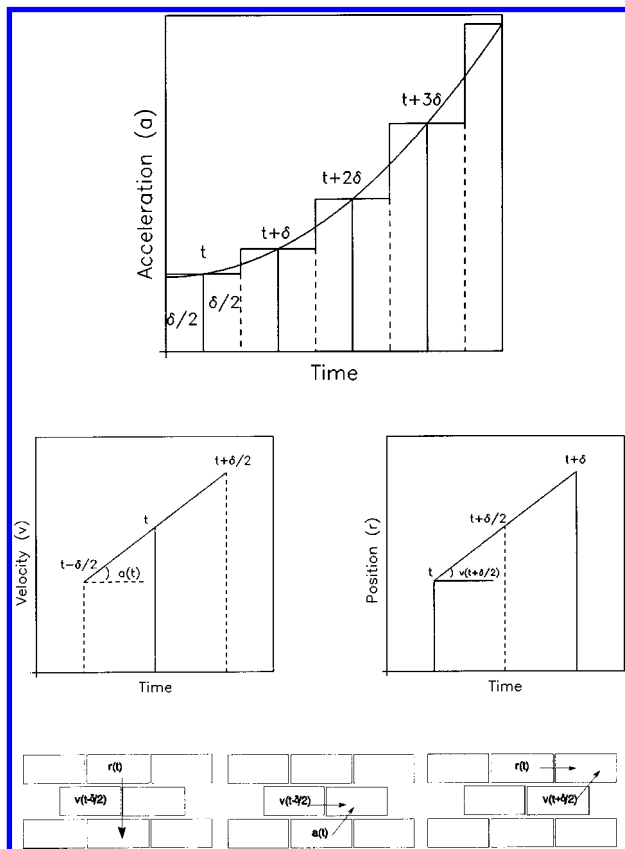


Figure 3. Leapfrog algorithm for integrating the equation of motion. Actual acceleration is replaced by a stair function obtained by dividing the integration interval δ into halves and assuming that the value $a(t)$ remains constant from $t - \delta/2$ to $t + \delta/2$. If the acceleration stays constant, the velocity changes linearly, with slope $a(t)$, and its value can be taken as average for the interval $t, t + \delta$ so that the movement will be uniform in that interval and the position will change linearly with slope $v(t + \delta/2)$. The algorithm is named for its way of advancing along time.

gorithm. The new velocities are then used, together with the positions $\mathbf{r}(t)$, to advance the coordinates to $\mathbf{r}(t + \delta)$ with the second equation and the whole procedure is repeated, computing positions as a function of time for $i = 1$ to $i = n - 1$. This way of advancing along time, jumping between full steps (δ) and half steps ($\delta/2$), which is illustrated in the last diagram of Figure 3, is responsible for the name of this procedure: *the leapfrog algorithm*.

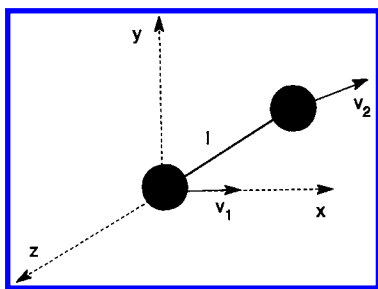


Figure 4. A diatomic molecule with two identical atoms separated by a distance l . See legend for Figure 2.

This procedure has the limitation of requiring small values for the time increment δ in order to obtain a good approximation between the actual acceleration and the stair function used to replace it. Of course, the best choice for the value of δ depends on the system to be studied, but it is common to use values of $\delta \approx 1$ fs (10^{-15} s) and to perform several millions of integration cycles in order to reach total times in the range of nanoseconds (10^{-9} s).

Since the velocities appear in the equations of the algorithm, it is easy to scale them up every certain number of steps to keep the temperature of the system constant for calculations performed on the canonical ensemble (constant particle number N , volume V , and temperature T). It is also possible to scale the velocities to keep the total energy (kinetic plus potential) constant in order to perform calculations within the microcanonical ensemble (fixed values of particle number N , volume V , and total energy E). This uncomplicated way of scaling the velocities makes this algorithm the most frequently used approach for MD calculations.

Diatomic Molecule

If our system is a diatomic molecule, such as that represented in Figure 4, instead of two independent particles as we have considered up to now, the situation would be very similar. In fact, the kinetic energy would be given by eq 5, the only difference being that the interaction among particles will no longer be a Lennard-Jones potential (eq 6) but a bond potential such as

$$E_{\text{bond}} = \frac{k}{2} (l - l_0)^2 \quad \mathbf{F}_l = -\frac{dE}{dl} = k(l - l_0) \quad (10)$$

where k represents the force constant of the bond and l_0 is the equilibrium bond length.

Thus the interaction is different, but the general procedure is the same: the positions are used to compute forces and accelerations, which are incorporated into the leapfrog algorithm to obtain positions as a function of time.

Water Molecule

The analysis of a more complicated molecule, for instance water (represented in Fig. 5), would also be quite similar. One noteworthy difference is that in polyatomic molecules we used to be more interested in the relative movements of the atoms with respect to each other than in displacements or rotations of the whole molecule. For this reason, the atomic velocities are scaled every certain number of steps to remove global rotations and translations of the whole molecule. This procedure eliminates six degrees of freedom (assuming that the molecule is not linear, in which

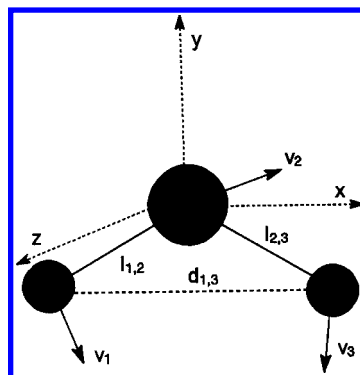


Figure 5. The water molecule. The oxygen atom is numbered as atom 2 and its initial position is taken as the origin for an arbitrary coordinate system.

case we would be removing only five). The relationship between temperature and atomic velocities would then be

$$E_T = E_k - E_{\text{glob. rot.}} - E_{\text{glob. trans}} = \frac{3N-6}{2} kT = \frac{1}{2} \sum_{j=1}^N m_j v_j^2 \quad (11)$$

where the sum expands over all N atoms of the molecule ($j = 1$ to N), and the atomic velocities v_j give zero values for the total linear and angular momenta of the whole molecule.

There are two kind of interactions among these atoms. We have first the bond forces, which have two contributions from bonds 1-2 and 2-3. Both bonds are of the O-H type and therefore they will have the same values of k and l_0 constants, although the instantaneous lengths may differ. This energy would then be

$$E_{\text{bonds}} = E_{12} + E_{23} = \frac{k}{2} \left\{ (l_{12} - l_0)^2 + (l_{23} - l_0)^2 \right\} \quad (12)$$

On the other hand, the bond angle 1-2-3 (the valence angle H-O-H) will also contribute to the potential energy of the molecule, as represented by the function

$$E_{\text{angles}} = \frac{k}{2} (\theta_{123} - \theta_0)^2 \quad (13)$$

with θ_0 and k being, respectively, the unstressed value of the angle and the force constant opposing its deformation.

Since there are two contributions to the potential energy, there will also be two contributions to the interatomic forces (which should be obtained by differentiation of the energies) and to the accelerations. The calculation follows the same pattern as in the previous examples, although it is a little bit more complicated.

In fact, it may be instructive to perform the calculation for this simple molecule. The results obtained are summarized in Figure 6, which represents the variation of the O-H bond length, the distance between the two H atoms, and the H-O-H bond angle as a function of time. The parameters used for this calculation were $T = 300$ K, time increment $\delta = 1$ fs, $n = 5 \times 10^5$ integration cycles to cover a total time of 0.5 ns, $l_0 = 0.95$ Å, $k = 1007.5$ kcal/mol Å² for the bond stretching energy (eq 12) and $\theta_0 = 110^\circ$, $k = 0.02$ kcal/mol (°)² for the energy produced by deformation of the valence angle (eq 13). The whole calculation required four minutes on a PC-486.

Before examining the results shown in Figure 6, it may be worth recalling a few things that we know, at least on the qualitative level, about the water molecule. For instance, we know that the atoms are supposed to move with the bonds stretching and compressing around a well-determined value $\langle l \rangle$ of the bond length, which should be very close to the unstressed value, l_0 , used as a parameter in the definition of the energy function. Something similar could be said about the bond angle for which $\langle \theta \rangle \approx \theta_0$. Finally, the H...H distance should also oscillate, because of modifications in O-H bond length and H-O-H bond angle, around an equilibrium value $d_0 = 1.556$ Å obtained when $l = l_0 = 0.95$ Å and $\theta = \theta_0 = 110^\circ$. However, it is important to realize that d_0 is not an energetic parameter, since these two H atoms do not interact with each other. This is exactly the behavior represented by Figure 6, in which l oscillates approximately between 0.90 and 1.0 Å with an average value $\langle l \rangle = 0.9501$ Å; d is in the range 1.45–1.65 Å with $\langle d \rangle = 1.556$ Å; and θ moves from 104 to 116° with an average of $\langle \theta \rangle = 110.01^\circ$.

But Figure 6 also indicates that the amplitude of the oscillations increases in the order $l < d < \theta$. The bond length changes are within 0.05 Å of the bond's average value, so that the maximum change in d is on the order of 0.1 Å and

in θ it amounts to ca. 6°. The amplitude of these oscillations can also be represented by computing the standard deviation of the values obtained across time: that is, the differences between the average of the square and the square of the average—for instance $\langle l^2 \rangle - \langle l \rangle^2$. These standard deviations amount to 3.2×10^{-4} Å² for the bond length, 4.4×10^{-3} Å² for the H...H distance, and 18.5 (°)² for the bond angle. The length of one O-H bond changes only by stretching or compressing that particular bond, whereas the H...H distance is modified by changing the length of either of the two O-H bonds or by modifying the H-O-H bond angle. Thus it is reasonable to predict that the d distance will produce larger oscillations than the individual O-H bonds. For the valence angle, the amplitude of the oscillation is due to the small value of the force constant for its deformation as compared with the values of force constants opposing bond elongation. Thus, according to eq 13, and with the values of the parameters indicated above, the energy required to deform θ by 11° is 1.2 kcal/mol, which, following eq 12, would produce a deformation in the bond length l of only ca. 0.05 Å. Of course, the amplitude of any of these deformations depends on temperature, so that performing the calculation at higher temperatures will produce larger oscillations in all these values.

Finally, we can also find in Figure 6 something that we know from spectroscopy: that the frequencies of the stretching vibrational modes of water (3642 and 3756 cm⁻¹) are higher than the frequency of the bending mode (1595 cm⁻¹). Thus, the oscillations for the O-H distance are much faster

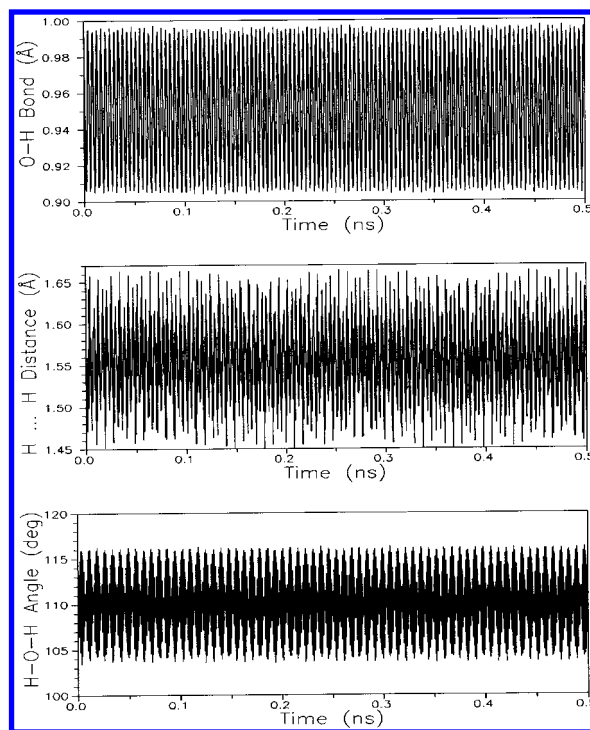


Figure 6. An MD calculation for the water molecule. Values of O-H bond length, distance between the two H atoms, and H-O-H bond angle are represented as functions of time. The calculation was performed at 300 K employing a time increment $\delta = 1$ fs and repeating the cycle of integration 5×10^5 times to cover a total time of 0.5 ns. Values used for the energetic parameters were $l_0 = 0.95$ Å, $k = 1007.5$ kcal/mol Å² for bond stretching and $\theta_0 = 110^\circ$, $k = 0.02$ kcal/mol (°)² for deformation of the valence angle.

than those of the H–O–H angle. Inspection of Figure 6 shows that l changes with time much faster than θ does. For instance, the separation between two consecutive peaks is roughly 4.3×10^{-3} and 9×10^{-3} ns respectively in the top and bottom panels of Figure 6.

It seems then that the MD calculations can produce results that agree with basic ideas about simple molecules such as water. Of course, if all the capabilities of this procedure were restricted to study such small molecules, its application would be very rare. However, the system studied can be much more complicated—a polymer, a liquid, etc.—and as far as the MD calculations are concerned, the procedure is the same as that described in the examples above. The only differences are that energy functions may be more

complicated and the number of atoms within the molecule or particles in the system may be larger.

Acknowledgment

This work was supported by the DGICYT through grant PB94-0364.

Literature Cited

1. *Computational Modeling of Polymers*; Bicerano, J., Ed.; Dekker: New York, 1992.
2. Allen, M. P.; Tildesley, D. J. *Computer Simulation of Liquids*; Clarendon: Oxford, 1987.
3. Haile, J. M. *Molecular Dynamics Simulation*; Wiley: New York, 1992.