ICES REPORT 13-26

August 2013

Atomistic and Coarse-Grained Modeling With Application to Hexane

by

Brian DeVincentis. (In collaboration with Kathryn Farrell and J. T. Oden)

The Institute for Computational Engineering and Sciences
The University of Texas at Austin
Austin, Texas 78712

Reference: Brian DeVincentis. (In collaboration with Kathryn Farrell and J. T. Oden), Atomistic and Coarse-Grained Modeling With Application to Hexane, ICES REPORT 13-26, The Institute for Computational Engineering and Sciences, The University of Texas at Austin, August 2013.

Atomistic and Coarse-Grained Modeling With Application to Hexane

Brian DeVincentis

August 16, 2013

1 Introduction

As engineers and scientists study smaller systems (such as biological cells or nanoscale computer transistors) that begin to approach the length scale of an atom, it becomes necessary to understand the behavior of atoms and molecules in detail. Determining the precise interactions between atoms requires the use of quantum mechanics. Unfortunately, this requires great computational power and is not feasible for systems with a large number of atoms. Therefore, over the past few decades researchers have developed simplified models describing the interactions between atoms which attempt to describe the forces atoms exert on one another. The models are not perfectly accurate but still have the ability to predict many phenomena at the atomic scale. For example, they can be used to calculate the density of water at a given temperature, to calculate the thermal conductivity of a material, and even simulate the folding of proteins used in the study of biological systems.

Unfortunately, when systems approach dimensions of 10 nanometers, the number of atoms makes the problem too computationally expensive to simulate with today's computers. As a result, engineers and scientists are actively researching methods to develop coarse-grained models of atoms and molecules. These coarse-grained models are simplifications of their atomistic counterparts. The idea is to replace groups of atoms and molecules with beads which act like atoms themselves. Interactions between atoms contained within the bead are ignored and, instead, only interactions between beads are considered. The result is a less accurate but more efficient model for simulating atoms and molecules.

The purpose of this document is to provide an overview of atomistic models and one method of coarse-grained modeling. These ideas are then applied to the hexane molecule.

2 Atomistic/Molecular Models

A molecular model is defined by a set of atoms, the forces exerted between them, their positions, and their velocities. Typically, and for the purposes of this document, atoms are considered point particles, although sometimes they are given size and shape. Each atom must have a defined chemical identity which affects the forces that atoms exert on one another. Although some molecular models predict and simulate bond formation and destruction, the models considered here require the bonds to be predefined.

The forces between atoms are determined by force fields, also called empirical potentials. Force fields define the functional form of the forces between atoms as well as the parameters for these functions dependent on the type of atoms involved in an interaction. Three well-developed and commonly used force fields are AMBER (Assisted Model Building with Energy Refinement), CHARMM

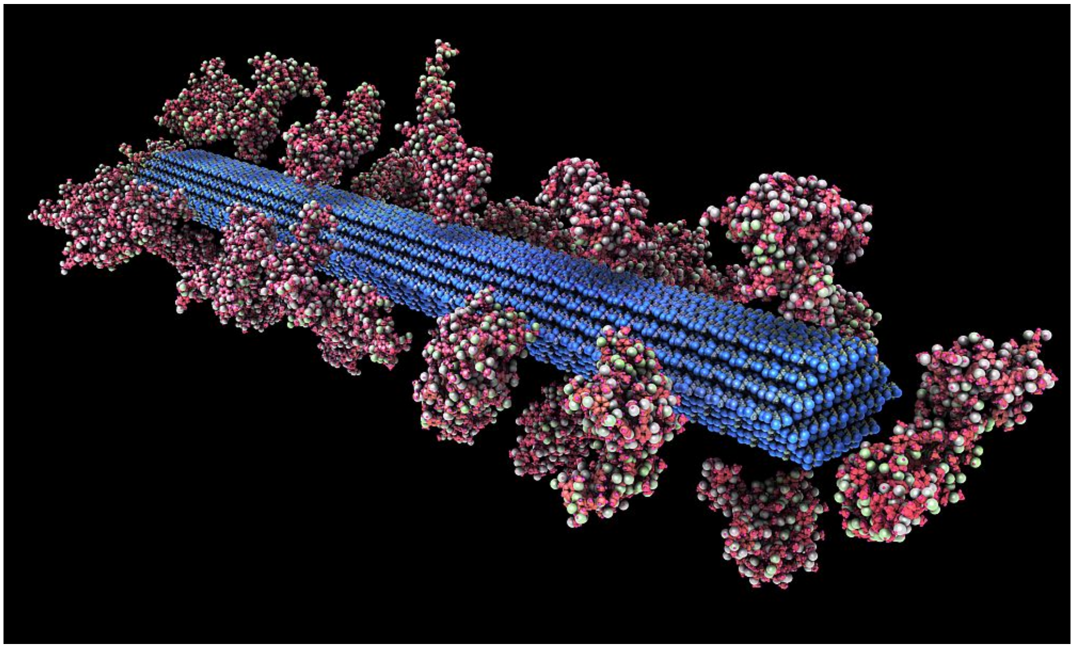


Figure 1: Example Molecular Simulation: Lignin Reprecipitation onto Cellulose [5]

(Chemistry at HARvard Macromolecular Mechanics), and OPLS (Optimized Potentials for Liquid Simulations). For the hexane application in this document, OPLS is used.

There are two categories of methods that can be used to perform a molecular simulation. The first is molecular dynamics which moves each atom according to the net force exerted on it by other atoms. This is done using a time stepping integration method: at each time step the net force on each atom is calculated, then each atom is moved based on that net force. The second method is Monte Carlo simulation which randomly moves atoms according to a probability distribution. Instead of determining the trajectory of each atom over time, it produces states of the system that are statistically likely.

Typically, a well-tested molecular dynamics code is used to perform a simulation. The codes AMBER, CHARMM, and OPLS implement the force fields for which they are named. GROMACS (GROningen MAchine for Chemical Simulations) and LAMMPS (Large-scale Atomic/Molecular Massively Parallel Simulator) are other commonly used codes that can use a variety of force fields. LAMMPS was used for the atomistic and coarse-grained simulations of hexane in this document.

The purpose of the simulations is not to determine the trajectory of each atom, but rather the properties that the system exhibits as a whole. For this purpose, the concepts of statistical mechanics, discussed in Section 3, are applied to the results to calculate macroscopic thermodynamic properties.

Returning to the previously mentioned force field, there are five types of interactions in atomistic models: non-bonded, bonded, angular, dihedral, and improper. Each is described below. Note that in this document the formulas are given for the potential energy of an interaction. The force is equal to the derivative of the potential energy with respect to position.

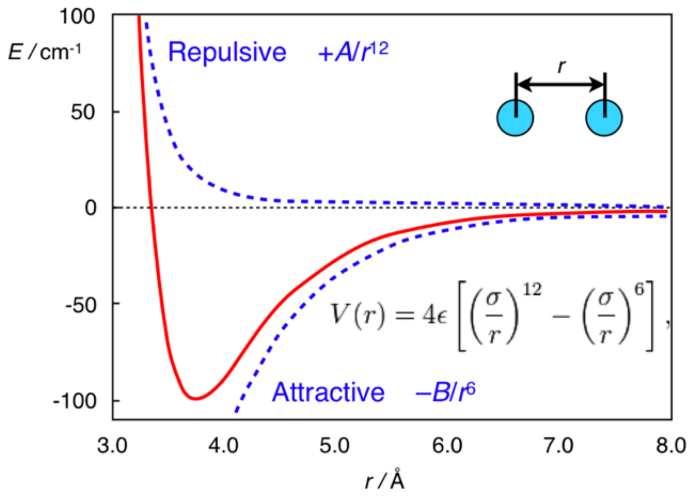


Figure 2: Lennard-Jones Potential [4]

2.1 Non-Bonded

Non-bonded interactions describe Van der Waals and Coulombic interactions. Typically, the Van der Waals interactions are modeled using the Lennard-Jones potential,

$$E = 4\epsilon\left[\left(\frac{\sigma}{r}\right)^{12} - \left(\frac{\sigma}{r}\right)^{6}\right], \tag{1}$$

where r is the separation between two atoms, ϵ determines the magnitude of the potential, and σ is the separation at which the potential energy is 0. See Figure 2 for a graph of (1). If $r < \sigma$ then as r decreases, the force becomes increasingly repulsive. If $r > \sigma$ then as r increases, the force becomes less repulsive until it reaches the point at which the force is 0, called the equilibrium separation. At any separation greater than the equilibrium, the force is attractive.

The Coulombic potential is defined as,

$$E = \frac{Cq_1q_2}{r}, \tag{2}$$

where q_1 and q_2 are the charges of the atoms, r is their separation distance, and C is a constant.

2.2 Bonded

In the vast majority of force fields, including OPLS, covalent bonds are modeled as harmonic springs. The energy potential is defined by,

$$E = K_r(r - r_0)^2, \tag{3}$$

where K_r is the spring constant, r is the separation between atoms, and r_0 is the equilibrium separation. The magnitude of the force between the two particles is equal to the derivative of the energy with respect to the separation. That is,

$$|F| = |2K_r(r - r_0)|. \tag{4}$$

As the separation r increases or decreases from the equilibrium distance, r_0, the force attracting or repelling the particles increases linearly. As a result, the bonded interaction attempts to keep the atoms at the equilibrium separation.

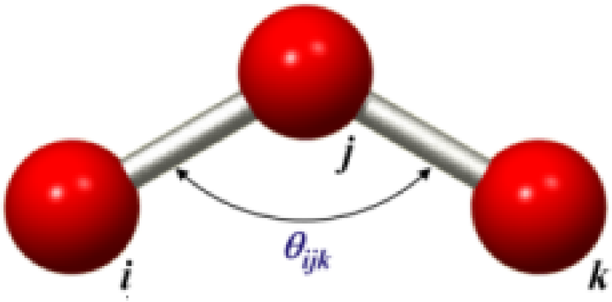


Figure 3: Diagram of an angle [2]

2.3 Angular

Angular interactions exert forces on a set of three atoms in which a center atom is bonded to two other atoms (see Figure 3). In OPLS, an angle is modeled as a harmonic spring and its energy potential is defined by,

$$E = K_\theta(\theta - \theta_0)^2, \tag{5}$$

where K_θ is a constant, θ is the angle, and θ_0 is the equilibrium angle. Note that this potential has the same functional form as the potential for bonded interactions.

2.4 Dihedral

A dihedral angle is formed by a group of four atoms bonded covalently in series (see Figure 4). The dihedral angle describes the angle between the plane formed by the first three atoms and the plane formed by the last three atoms. The dihedral potential describes the energy required to twist the center bond which causes the dihedral angle to change. In OPLS, its energy potential is defined by,

$$E = \sum_i \frac{V_{1,i}}{2}(1 + \cos\phi_i) + \frac{V_{2,i}}{2}(1 - \cos 2\phi_i) + \frac{V_{3,i}}{2}(1 + \cos 3\phi_i) + \frac{V_{4,i}}{2}(1 - \cos 4\phi_i), \tag{6}$$

where the V's determine the magnitude of the potential and ϕ is the dihedral angle.

2.5 Improper

An improper angle is formed when one atom is directly bonded to three other atoms (see Figure 5). The potential attempts to keep the four atoms at a defined equilibrium angle. It is included only in some force fields such as CHARMM but not OPLS.

3 Statistical Mechanics for Molecular Systems

The force fields described in Section 2 allow for the simulation of atomistic and molecular systems. But the following questions arise: What do the results of a molecular simulation mean? How can the simulation of atoms be useful for the analysis of a much larger system? Statistical mechanics answers these questions by relating macroscopic systems to microscopic systems through the concepts of probability and statistics. It allows for the interpretation of the results of a molecular simulation in order to make conclusions about a macroscopic system.

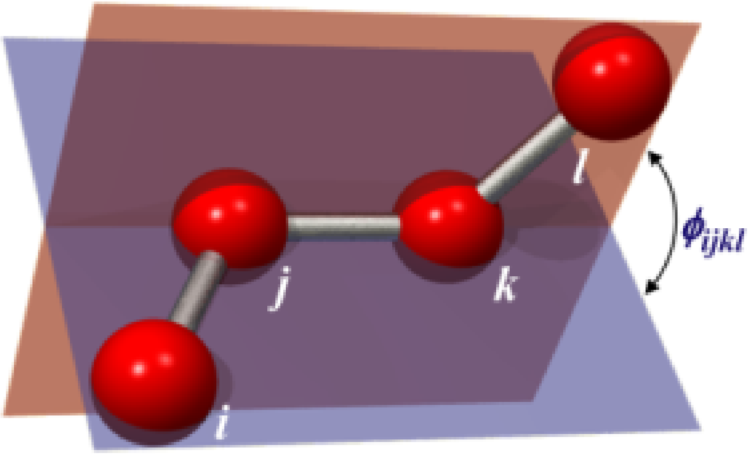


Figure 4: Diagram of a dihedral angle [2]

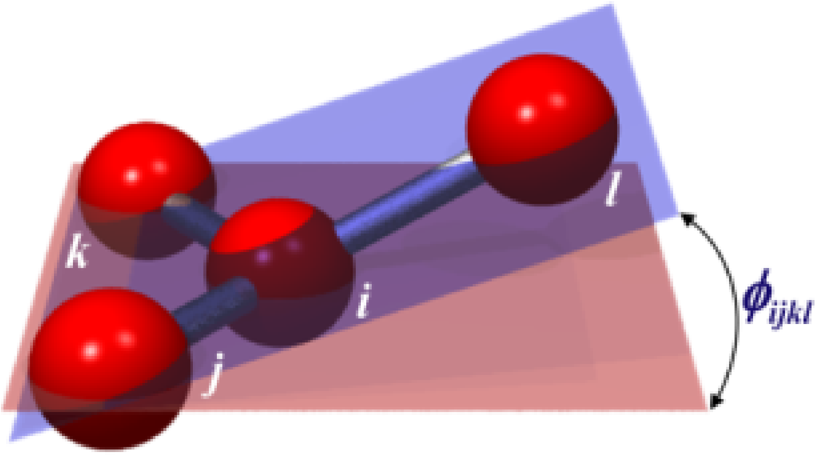


Figure 5: Diagram of an improper angle [2]

3.1 A Quick Side Note: Hamiltonian Mechanics

For molecular systems, Hamiltonian mechanics is a more convenient, though equivalent, way to describe a mechanical system than Newtonian mechanics. For a molecular system of n particles, the Hamiltonian, H, is defined as the total energy [8],

$$H(\boldsymbol{q}, \boldsymbol{p}) = \sum_{i=1}^{n} \frac{p_i^2}{2m_i} + U(\boldsymbol{q}), \tag{7}$$

where the first term is the kinetic energy dependent solely on the vector $\boldsymbol{p}$ of momenta, and the second term is the potential energy dependent solely on the vector $\boldsymbol{q}$ of positions. The potential energy function, U, is the sum of the potentials defined by the force field. The dynamics of a Hamiltonian system are determined by the equations [8],

$$\dot{q}_i = \frac{\partial H}{\partial p_i} \qquad \text{and} \qquad \dot{p}_i = -\frac{\partial H}{\partial q_i}. \tag{8}$$

3.2 Ensembles

The ensemble is a theoretical concept that helps to establish relations between macroscopic and microscopic systems. An ensemble the set of possible states of a microscopic system that are all consistent with a certain set of thermodynamic conditions. An ensemble may require all its possible states to be at the same temperature, pressure, or other conditions. These thermodynamic conditions can then be associated with a macroscopic system.

An important idea is that of an ensemble's probability distribution. The distribution describes the probability that a micro-system will be in a certain state out of all the possible states that the system can take on. Recall that a molecular system's state is defined by the positions, $\boldsymbol{q}$, and momenta, $\boldsymbol{p}$, of all the atoms. The distribution defines how likely it is for the molecular system to have a certain $\boldsymbol{q}$ and $\boldsymbol{p}$ out of all the possible $\boldsymbol{q}$'s and $\boldsymbol{p}$'s. Different types of ensembles have different forms of probability distributions. See the sections on microcanonical and canonical ensembles below for their probability distributions.

At the heart of statistical mechanics is the ergodic hypothesis, which relates an average over a macro-system to an average over the corresponding micro-system. For a macro-system, the time average is the average value of a property, defined by the phase function $F(\boldsymbol{q}, \boldsymbol{p})$, calculated at every macro-state during a dynamic trajectory of the system [8],

$$\bar{F} = \lim_{T \to \infty} \frac{1}{T} \int_0^T F\big(\boldsymbol{q}(\tau), \boldsymbol{p}(\tau)\big) d\tau, \tag{9}$$

where T is the time period. Calculating the time average of the macro-system is not generally computationally feasible. For a micro-system, the ensemble average is the average value of a property, again defined by the phase function $F(\boldsymbol{q}, \boldsymbol{p})$, over all the possible micro-states weighted by the probability distribution for that ensemble [8],

$$\langle F \rangle = \int F(\boldsymbol{q}, \boldsymbol{p}) \rho(\boldsymbol{q}, \boldsymbol{p}) d\boldsymbol{q}^N d\boldsymbol{p}^N, \tag{10}$$

where ρ is the probability distribution and N is the number of particles times the number of coordinate axes. As an example, for a system with n particles, $q^N = (q_1, q_2, q_3, ..., q_{3n-2}, q_{3n-1}, q_{3n})$.

The ergodic hypothesis states that the time average of the macro-system is equal to the ensemble average of the micro-system if the averages are taken over a long period of time and a large number of states respectively. The ergodic hypothesis makes it possible to approximate macroscopic properties from a microscopic system by calculating the ensemble average of the corresponding properties of the microscopic system.

3.3 NVE (Microcanonical) Ensemble

The NVE ensemble, also called the microcanonical ensemble, describes a set of states of a molecular system that all have the same number of particles N, occupy the same volume V, and have the same total energy E. One of the principles of statistical mechanics is that a micro-system has an equal probability of being in any of the possible states of a microcanonical ensemble. This makes the calculation of an ensemble average very simple for a microcanonical ensemble, it is just the average value of a property at the observed states. This is a fundamental postulate of statistical mechanics.

3.4 NVT (Canonical) Ensemble

The NVT ensemble, also called the canonical or Gibbs ensemble, describes a set of states of a molecular system that have the same number of particles N, occupy the same volume V, and are at the same temperature T. In the canonical ensemble, energy is not constant. Instead it maintains a constant temperature. To do this, the Hamiltonian, given in (7), must be modified. The modified equation is called a Nosé-Hoover equation.

The canonical ensemble has a more complicated probability distribution, called the Boltzmann distribution. The probability of a particular state defined as [8],

$$\rho(\boldsymbol{q},\boldsymbol{p}) = \frac{e^{-\beta H(\boldsymbol{q},\boldsymbol{p})}}{\int e^{-\beta H(\boldsymbol{q},\boldsymbol{p})} d\boldsymbol{q} d\boldsymbol{p}}, \quad \beta = \frac{1}{k_B T}, \tag{11}$$

where H is the Hamiltonian (total energy), k_B is the Boltzmann constant, and T is the temperature. Note that for high energies the probability is very low, which makes sense intuitively since a system naturally tries to reach is lowest energy state.

3.5 NPH and NPT Ensembles

The NPH ensemble describes a system in which the number of particles N, the pressure P, and the enthalpy H are held constant. The NPT ensemble differs in that the enthalpy is allowed to fluctuate, but the temperature T must remain constant. As with the canonical ensemble, the energy is not constant. Therefore, they also require modified versions of the Hamiltonian.

3.6 Calculation of Macroscopic Properties

Using the principles of statistical mechanics it is possible to calculate macroscopic properties, such as temperature and density, based on the results of a molecular simulation.

Recall the ensemble average (10). This integral may be approximated by sampling points in $\boldsymbol{q}$–$\boldsymbol{p}$ space, called the phase space. If the points are sampled uniformly then [7, 3],

$$\langle F \rangle = \int F(\boldsymbol{q},\boldsymbol{p})\rho(\boldsymbol{q},\boldsymbol{p}) d\boldsymbol{q}^N d\boldsymbol{p}^N \approx \frac{1}{L}\sum_{i=1}^{L} F(\boldsymbol{q}_i,\boldsymbol{p}_i)\rho(\boldsymbol{q}_i,\boldsymbol{p}_i), \tag{12}$$

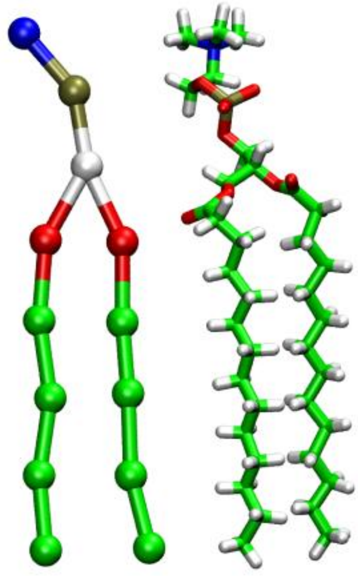


Figure 6: Example of Coarse-Graining [6]

where L is the number of samples. However, if the samples are generated according to anther distribution g, this approximation becomes,

$$\langle F \rangle = \int F(\boldsymbol{q},\boldsymbol{p})\rho(\boldsymbol{q},\boldsymbol{p}) d\boldsymbol{q}^N d\boldsymbol{p}^N \approx \frac{1}{L}\sum_{i=1}^{L} \frac{F(\boldsymbol{q}_i,\boldsymbol{p}_i)\rho(\boldsymbol{q}_i,\boldsymbol{p}_i)}{g(\boldsymbol{q}_i,\boldsymbol{p}_i)}. \tag{13}$$

Details for this derivation are given in [7].

The conditions for simulating a molecular system governed by any of the ensembles described above can be enforced in molecular dynamics codes, including LAMMPS. This is accomplished by generating positions and velocities sampled according to the proper ensemble distribution function, as enforced by the Nosé-Hoover-style equations of motion [1]. Therefore, the generating distribution g is actually the ensemble distribution ρ. The approximation of the macroscopic values in (13) then becomes,

$$\langle F \rangle = \frac{1}{L}\sum_{i=1}^{L} \frac{F(\boldsymbol{q}_i,\boldsymbol{p}_i)\rho(\boldsymbol{q}_i,\boldsymbol{p}_i)}{\rho(\boldsymbol{q}_i,\boldsymbol{p}_i)} = \frac{1}{L}\sum_{i=1}^{L} F(\boldsymbol{q}_i,\boldsymbol{p}_i). \tag{14}$$

In conclusion, a thermodynamic property's ensemble average can be approximated by averaging its values calculated at each step of the molecular simulation if the code is sampling from the proper ensemble distribution function.

4 Coarse-Grained Models

A coarse-grained model is a simplification of an atomistic model, created for the purpose of improving the efficiency of the simulation while simultaneously maintaining properties of interest. To define a coarse-grained model, atoms are grouped into beads and the interactions between these beads are assigned.

The process of grouping atoms into beads is called mapping. It is convenient to replace groups of atoms that appear commonly throughout a system with one type of bead. The position of the bead can be determined in various ways. A common choice is to place it at the center of mass of the group of atoms it replaces. Another option is to place the bead at the position of one of the atoms in the group.

After the mapping is defined, the interaction potentials between the beads must be chosen and their parameters defined. Coarse-grained models are simulated in the same way as atomistic models. The same potentials, dynamic equations, and molecular dynamics codes can be used for coarse-grained models. For the purposes of this research, it is assumed that the coarse-grained potential energy assumes the same functional form as the OPLS atomistic empirical potential discussed in Section 2. The process by which the parameters are determined is discussed in the next two sections.

5 Model Inference

Model inference is the process of determining a theoretical model's parameters based on collected data. The theoretical model is fit to the data by optimizing the parameters' values. Now consider the application of model inference to coarse-grained modeling. The data is generated from the atomistic model, the theoretical model is the coarse-grained model, and the parameters are those in the force field functions.

The most common method of model inference is the method of least squared error. It determines the parameters to be those values which minimize the error between the data and the theoretical model. The method of force matching uses least squared error to calibrate the parameters of a coarse-grained model. Force matching attempts to apply the same forces between beads as the net force between the atoms in one bead and the atoms is another. [9]

Another method of model inference is called Bayesian inference, which applies Bayes' Theorem to parameter optimization. One of the advantages of Bayesian inference is that instead of considering the parameters as single values, they are considered random variables which each have a probability distribution. In addition to optimizing the parameters, this approach allows for a better understanding of a model's uncertainty. The next section describes Bayesian inference in more detail.

5.1 Bayesian Inference

Bayesian inference employs Bayes' Theorem given by the equation,

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)}, \tag{15}$$

where A and B are events. $P(A|B)$ is a function that determines the probability of event A occurring given that B has already occurred. This is called a conditional probability. Similarly, $P(B|A)$ is the probability of B given A. The remaining terms, $P(A)$ and $P(B)$ are the probabilities of events A and B occurring respectively. The theorem is derived from the definition of a conditional probability.

From a simplistic point of view, Bayes' Theorem is just a relationship between probabilities. If three are known, then the fourth can be found. However, when used in Bayesian probability the equation takes on a specific meaning. It is thought of as updating the probability of A based on the occurrence of B. Its purpose is to define the change in the probability of A if B occurs.

When this idea of Bayesian probability is applied to model inference, Bayes' Theorem becomes a useful tool for determining a theoretical model's parameters when data is observed. Bayes' Theorem can be restated as,

$$\pi(\theta|D) = \frac{\pi(D|\theta)\pi(\theta)}{\pi(D)}, \tag{16}$$

where quantitative probabilities, P, have been replaced by probability density functions, π. θ is the set of model parameters and D is the set of observed data. The prior, $\pi(\theta)$, represents the previous information regarding the probability distribution for each of the parameters. If nothing is known, then the prior is a uniform distribution. The likelihood, $\pi(D|\theta)$, is the probability of reproducing the observational data given a set of model parameters. As the error between the data and the model decreases, the likelihood increases. Finally, the posterior, $\pi(\theta|D)$, is the probability of a set of values for the model's parameters after including the data. The result of Bayesian inference is not the optimal values for each of the parameters, but instead the optimal probability density of each parameter. Furthermore, it can be used to include new data into the existing model by simply replacing the old prior with the latest posterior and repeating the calculation.

6 Comparison of Models

In the process of creating a model, the modeler will be required to make choices. In the case of coarse-graining, the modeler can choose the mapping method and how to represent the interactions between beads. The various options result in multiple candidate models. Ultimately, these models must be compared in order choose the best. Bayesian statistics provides a method for making such a comparison. It allows for a quantitative description of the relative plausibilities of models. The plausibility of a model is the probability that it will most accurately predict a future outcome, compared to a set of models. The denominator of Bayes' Theorem, $\pi(D)$, is called the evidence, and in the case of Bayesian inference represents the probability of observing the data for a given model. The evidences of candidate models can be compared to determine the plausibility of each model.

A higher form of Bayes' Theorem can be stated for the purpose of calculating model plausibility,

$$\pi(M_i|D,\mathcal{M}) = \frac{\pi(D|M_i)\pi(M_i|\mathcal{M})}{\pi(D|\mathcal{M})}, \tag{17}$$

where M_i is a candidate model and $\mathcal{M}$ is the set of all candidate models. The sum of the plausibilities of the all the candidate models must be one,

$$\sum_i \pi(M_i|D,\mathcal{M}) = 1. \tag{18}$$

The prior plausibility, $\pi(M_i|\mathcal{M})$, is equal to one over the number of candidate models if no previous information is known. The evidence, $\pi(D|M_i)$, is found from Bayesian inference. The model plausibility is given by $\pi(M_i|D,\mathcal{M})$. And $\pi(D|\mathcal{M})$ is a normalizing constant satisfying (18). The model with a plausibility closest to one is considered the best model.

7 Hexane

Hexane is a molecule with a chain of 6 carbon atoms surrounded by 14 hydrogen atoms (see Figure 7). For this document, its potentials were modeled using the OPLS force field and simulated using the molecular dynamics code LAMMPS. In the coarse-grained models, hexane was simplified from 20 atoms and 19 bonds to 3 beads and 2 bonds (see Figure 8). Each bead replaces two carbon

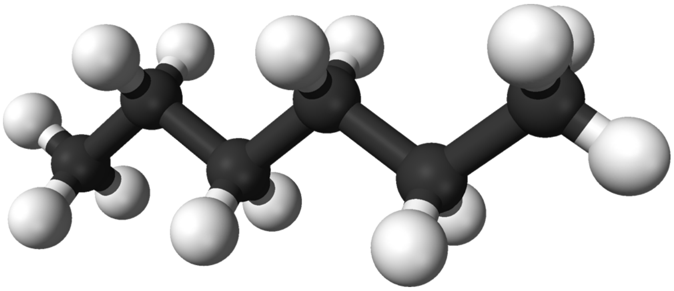


Figure 7: Hexane Molecule, black spheres are carbon atoms and white spheres are hydrogen atoms.

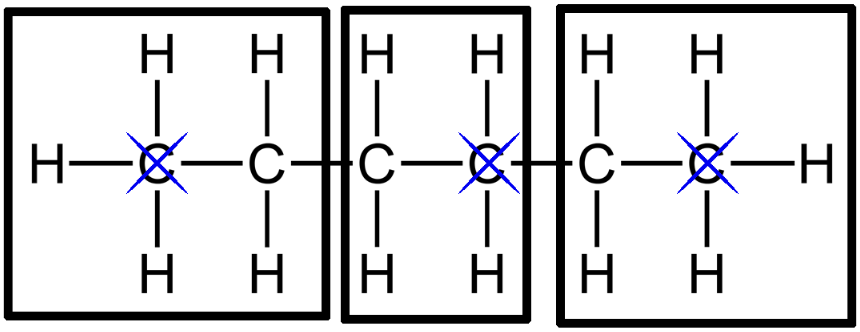


Figure 8: Hexane Coarse-Grained Beads, position of beads marked by blue x's

atoms and their associated hydrogen atoms. There are two types of beads since the outer beads have the same structure. The outer beads were placed on the atoms with the extra hydrogen and the middle bead was placed on the fourth carbon atom from the left.

The purpose of the atomistic model was to provide data for the Bayesian inference of the coarse-grained models. The data was generated by simulating hexane in a canonical ensemble at a temperature of approximately 300K. Data was collected periodically during the simulation. The total potential energy of hexane was recorded as well as the positions of the carbon atoms that determine the positions of the beads. Ten data points were recorded over 1000 ensemble time steps. Each time step is 1 femtosecond.

The Bayesian inference code, QUESO (Quantification of Uncertainty for Estimation, Simulation and Optimization), was used to perform the coarse-grained model inference. The code required the parameters' priors and the definition of the likelihood. The priors were defined as uniform distributions within a given range. The likelihood function calculated the probability of observing the data for a given set of model parameters. First, it calculated the error between the data and the coarse-grained model's prediction, where the error was defined as the difference between the predictions of the potential energy. Then the likelihood was calculated as the probability of this error occurring according to an assumed probability distribution. Both Gaussian and log normal distributions were used, and the results of both are presented.

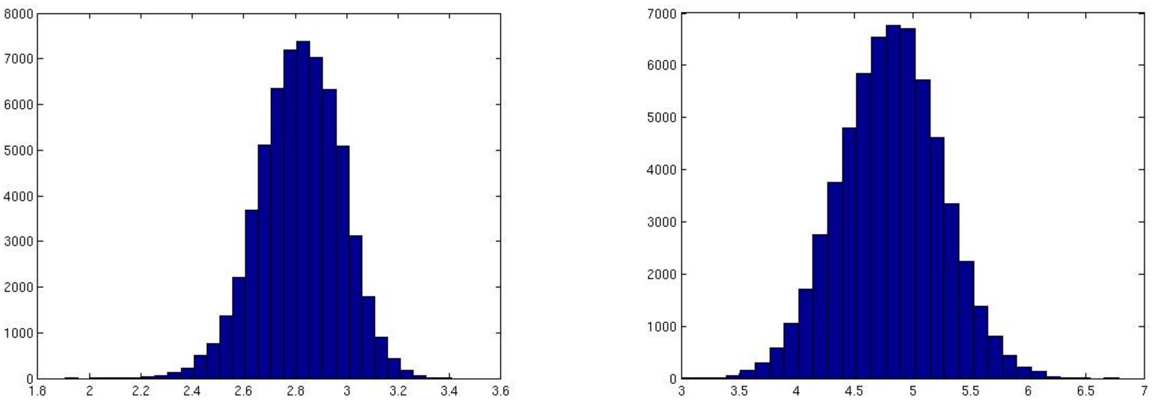
8 Results

This section presents the results of the coarse-graining process described above for three candidate hexane models: bonds only, angles only, and both bonds and angles. The histograms in this section show the probability distribution for each model parameter. The x-axis is the parameter's value. The y-axis is related to the number of Markov Chain Monte Carlo samples. For each model, both a Gaussian and log normal likelihood error probability distribution were used, and the results of both are presented separately. A variance of 0.1 was used.

8.1 Bond

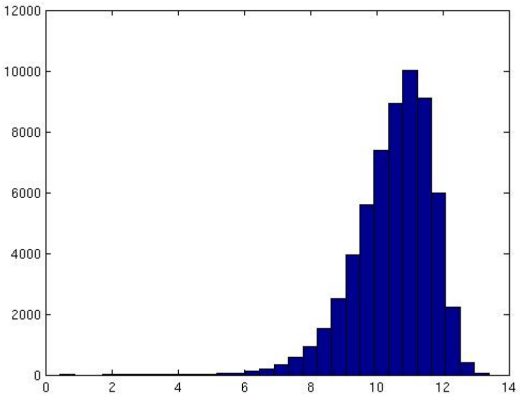
The bond and angle model's potential energy is the sum of the potential energy in the bonds between the beads and a constant. The two bonds are of the same type (they both connect an outer bead to the middle bead), therefore, the parameters K_r and r_0 are the same for both bonds.

$$U_{CG} = \sum_{bonds} K_r(r - r_0)^2 + C \tag{19}$$



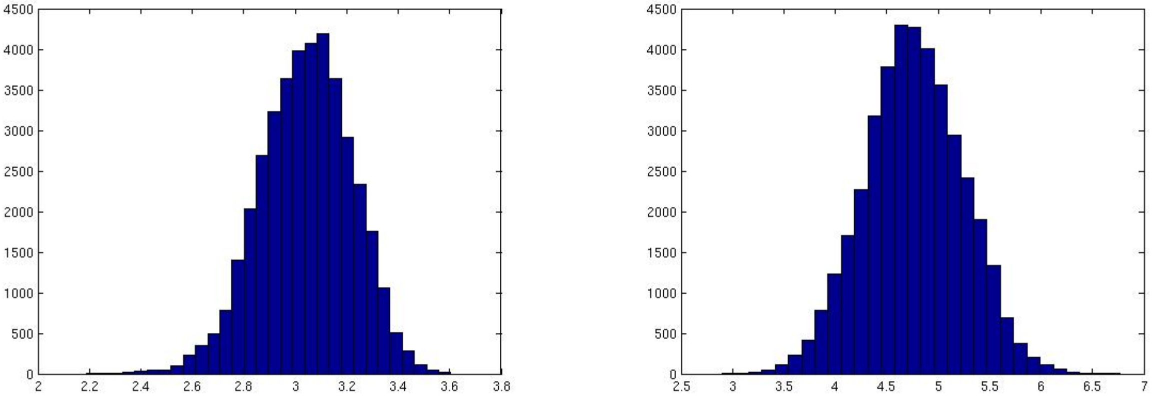
(a) r_0

(b) K_r

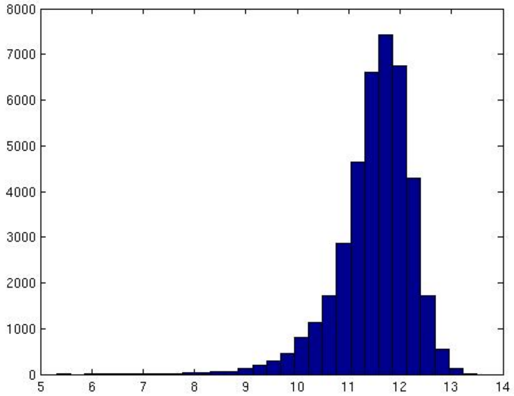


(c) C

Figure 9: Bond Model Parameters (Gaussian Likelihood)



(a) r_0 (b) K_r



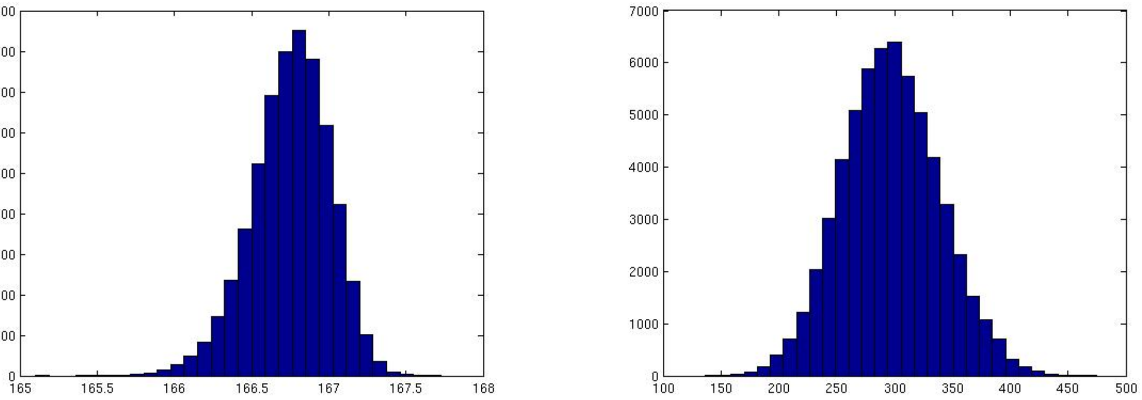
(c) C

Figure 10: Bond Model Parameters (Log Normal Likelihood)

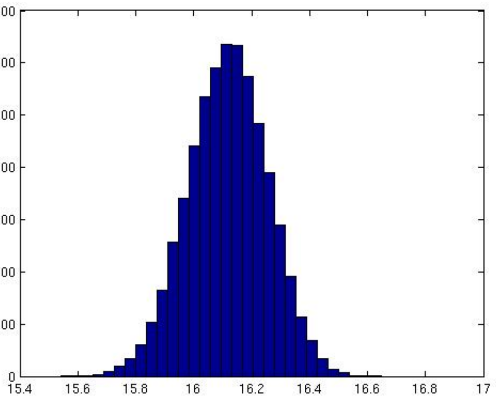
8.2 Angle

The angle model's potential energy is the sum of the potential energy in the angle between the three beads and a constant.

$$U_{CG} = K_\theta(\theta - \theta_0)^2 + C \tag{20}$$

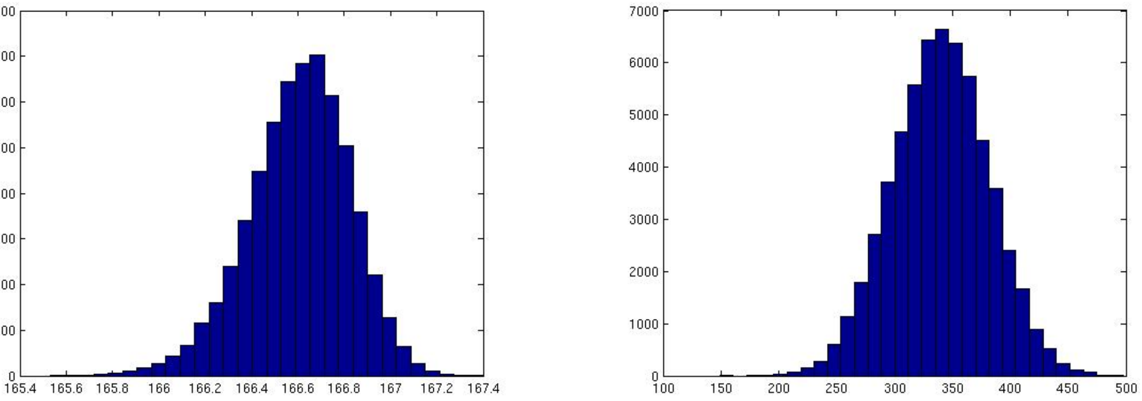


(a) θ_0 (b) K_θ



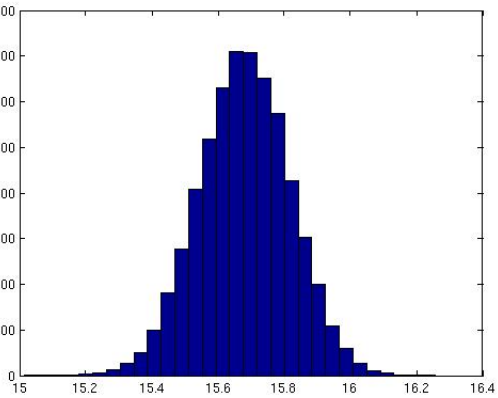
(c) C

Figure 11: Angle Model Parameters (Gaussian Likelihood)



(a) θ_0

(b) K_θ



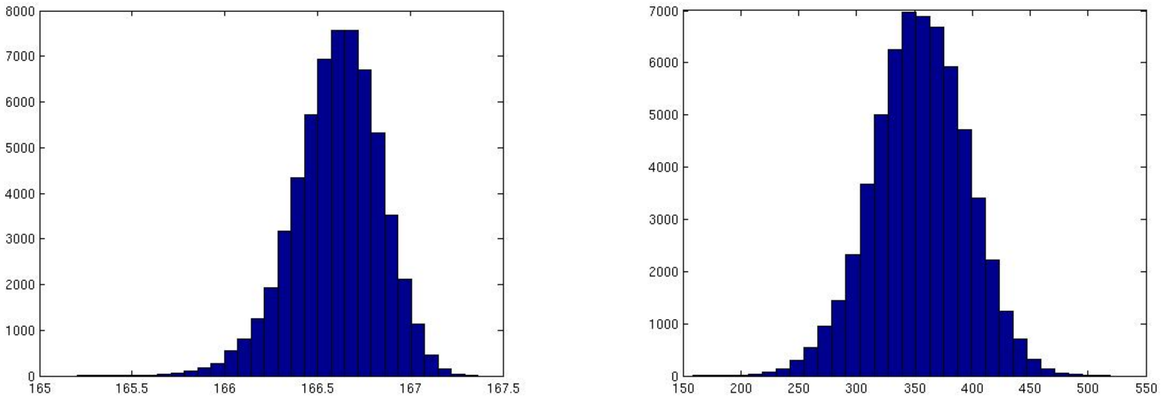
(c) C

Figure 12: Angle Model Parameters (Log Normal Likelihood)

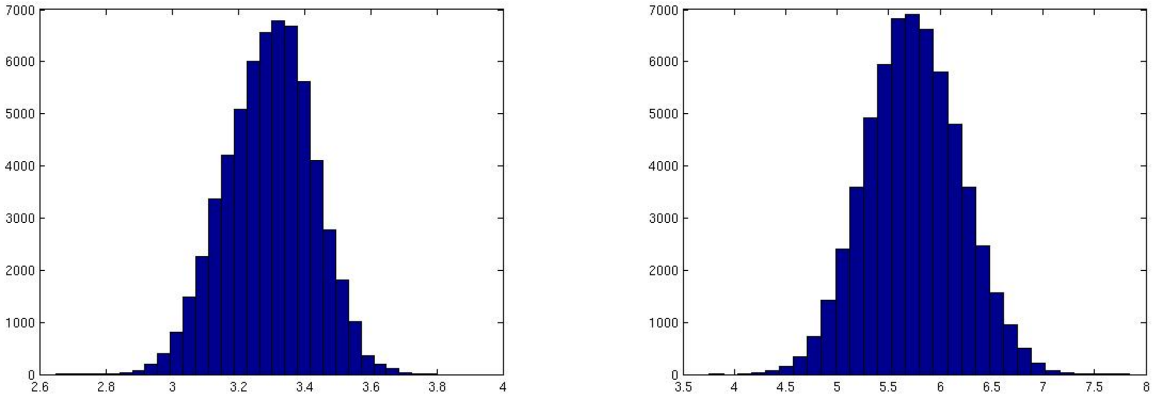
8.3 Bond and Angle

The bond model's potential energy is the sum of the potential energy in the bonds, the potential energy in the angle, and a constant.

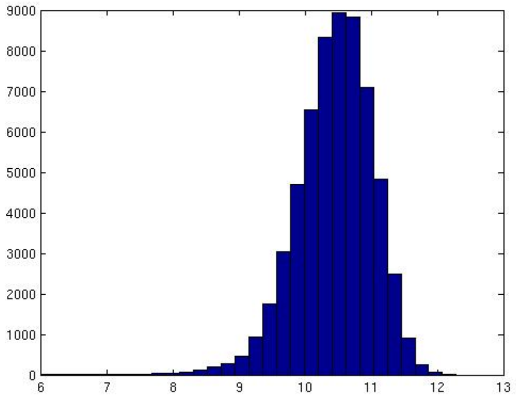
$$U_{CG} = \sum_{bonds} K_r(r - r_0)^2 + K_\theta(\theta - \theta_0)^2 + C \tag{21}$$



(a) θ_0 (b) K_θ

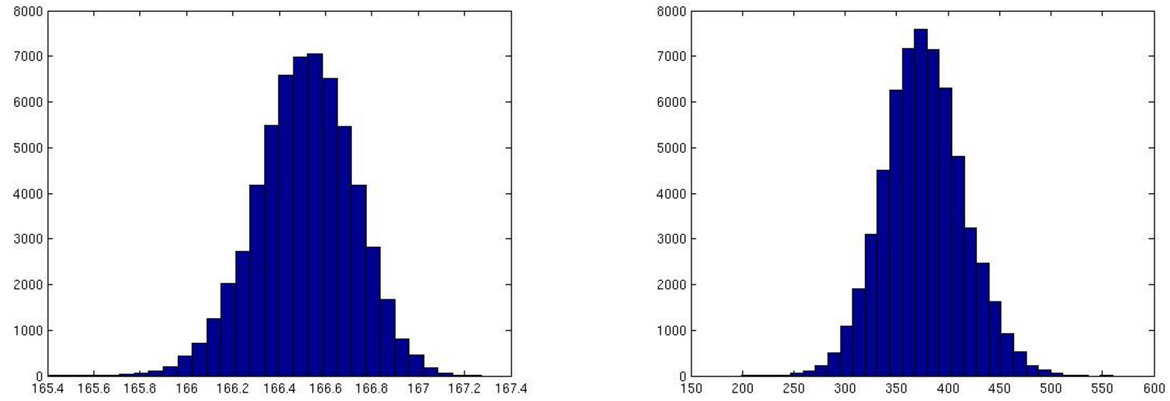


(c) r_0 (d) K_r



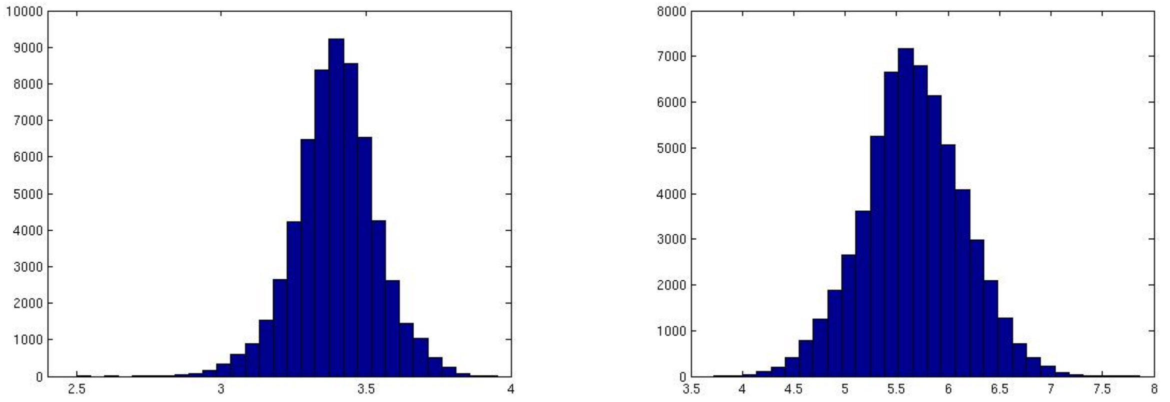
(e) C

Figure 13: Bond and Angle Model Parameters (Gaussian Likelihood)



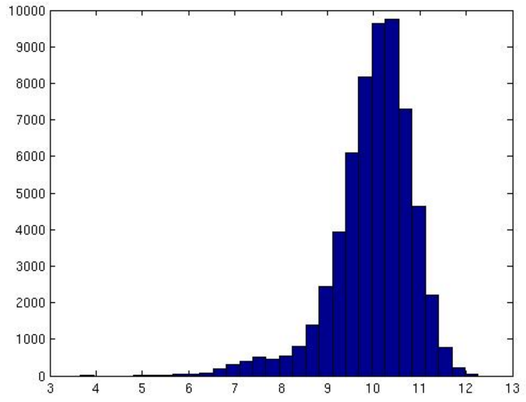
(a) θ_0

(b) K_θ



(c) r_0

(d) K_r



(e) C

Figure 14: Bond and Angle Model Parameters (Log Normal Likelihood)

Model	**C**	θ_0	K_θ	r_0	K_r	**Log Evidence**	**Evidence**	**Plausibility**
Bond	11.0			2.83	4.90	-30.3	6.93E-14	3.03E-04
Angle	16.2	166.8	300			-84.5	2.01E-37	8.78E-28
Bond Angle	10.5	166.7	360	3.35	5.75	-22.2	2.28E-10	0.9997

Figure 15: Results for Models Using Gaussian Likelihood

Model	**C**	θ_0	K_θ	r_0	K_r	**Log Evidence**	**Evidence**	**Plausibility**
Bond	11.8			3.10	4.75	-57.9	7.15E-26	3.40E-09
Angle	15.7	166.7	350			-66.9	8.82E-30	4.19E-13
Bond Angle	10.3	166.6	375	3.45	5.75	-38.4	2.10E-17	0.999999997

Figure 16: Results for Models Using Log Normal Likelihood

9 Comparison of Models

Using the concepts of model comparison and plausibility, described in Section 6, the plausibility of each of the three models was calculated. The plausibility calculation relied on the evidence values which were determined by QUESO.

From the tables in Figures 15 and 16, it is clear that the bond and angle model is a much better model than the only bonds or only angle models.

10 Likelihood Error Probability Distribution

Two likelihood error probability distributions were used in this document: Gaussian and log normal. A log normal distribution is one in which the log of a random variable is assumed to be normally distributed rather than the random variable itself. Overall the two produced similar results with minor differences in some values.

All the results presented used a variance of 0.1. Small changes to the variance has little or no effect on the resulting parameter values, although it does have a significant effect on the evidences since it changes the likelihood error probability distribution. Decreasing the variance results in narrower parameter probability distributions, but smaller evidence values. Increasing the variance results in wider probability distributions. If the increase is large it may change the shape of the distribution – QUESO may find multiple optimal values for some parameters.

11 Constant Parameter in Models

The constant potential energy parameter has little or no effect on this Bayesian inference process since the three models presented effectively have a constant term that is a function of the other parameters. However, it does have an effect on the values of the other parameters. Without the constant, some of the parameters are on different orders of magnitude than one would expect in an atomistic system. For example, the equilibrium distance, r_0, is around 0.3 while an atomistic value for r_0 would normally be around 2.0. When the constant is added, the parameters have values that more closely resemble an atomistic system.

This constant represents the potential energy between the atoms inside the beads. Remember that a coarse-grained model only keeps track of the potentials between beads. When there is no constant, the potentials between beads must account for all the potential energy of the molecule. Using LAMMPS I calculated the potential energy between the atoms in each bead for one data point. I found the total potential energy for the three beads to be 12.4 which is similar to the values for C in Figures 15 and 16.

It is important for the parameters in coarse-grained systems to be similar those in atomistic systems because the coarse-grained system is treated identically an atomic system when it is simulated. In this document, the quality of the coarse-grained model was only measured by how well it predicts the potential energy. But it is also important for the coarse-grained model to predict the kinetic energy, temperature, and positions of the beads correctly.

To determine if the constant parameter improved the coarse-grained models, the bond and angle model was simulated in a canonical ensemble with and without the constant parameter. Then the bond lengths between the beads were compared with the atomistic system. The result: atomistic 2.53, with constant 2.52, without constant 1.11. This clearly shows that the constant improves the prediction of the positions of the beads. (Note this was done before I found an error in the mapping function, but I think the results would be similar if recalculated.)

12 Conclusion

After an introduction to atomistic and coarse-grained modeling, this document presents the results of modeling hexane as three candidate coarse-grained models: only bonded, only angular, and both bonded and angular interactions. Each model's parameters were determined using Bayesian inference to minimize the error in the prediction of the potential energy. The three models were compared by calculating each model's plausibility. The both bonded and angular interactions model was found to be a much better model than the other two. It was also found that the addition of a constant to each of the models improved the prediction of the beads' positions. The coarse-graining process was performed using both Gaussian and log normal likelihood functions, which had very similar results.

References

[1] LAMMPS Manual. Available: http://http://lammps.sandia.gov/doc/Manual.html.

[2] PumMa. Available: http://www.pumma.nl/index.php/Theory/Theory.

[3] Wei Cai. "Importance Sampling in Monte Carlo Simulation of Rare Transition Events", 2005. Available: http://micro.stanford.edu/~caiwei/LLNL-CCMS-2005-Summer/Lecture_Notes/impsample_lecture_all.pdf.

[4] Cambridge MedChem Consulting. Molecular interactions, 2012. Available: http://www.cambridgemedchemconsulting.com/resources/molecular_interactions.html.

[5] Susan Gregurick. Available: http://ascr-discovery.science.doe.gov/shared/exa_biol_fig1.html.

[6] Temple University Institute for Computational Molecular Science. Available: https://icms.cst.temple.edu/software.html.

[7] S. Lehti and V. Kanmaki. "Computing Methods in High Energy Physics". Lecture Notes, 2010. Institute of Physics, University of Helsinki.

[8] J. T. Oden. *An Introduction to Mathematical Modeling.* Wiley, 2011.

[9] M. Scott Shell. “Coarse graining and multiscale techniques”. Lecture Notes, 2012. Department of Chemical Engineering, University of California Santa Barbara.