

2012

# Modeling spontaneous three-dimensional polymerization

Zhitong Zhao

*Michigan Technological University*

Copyright 2012 Zhitong Zhao

---

## Recommended Citation

Zhao, Zhitong, "Modeling spontaneous three-dimensional polymerization", Master's Thesis, Michigan Technological University, 2012.  
<http://digitalcommons.mtu.edu/etds/208>

Follow this and additional works at: <http://digitalcommons.mtu.edu/etds>

 Part of the [Mathematics Commons](#)

MODELING SPONTANEOUS THREE-DIMENSIONAL POLYMERIZATION

By

Zhitong Zhao

A THESIS

Submitted in partial fulfillment of the requirements for the degree of

MASTER OF SCIENCE

(Mathematical Sciences)

MICHIGAN TECHNOLOGICAL UNIVERSITY

2012

© 2012 Zhitong Zhao

This thesis, "Modeling Spontaneous Three-Dimensional Polymerization," is hereby approved in partial fulfillment of the requirements for the Degree of MASTER OF SCIENCE IN MATHEMATICAL SCIENCES.

Department of Mathematical Sciences

Signatures:

Thesis Advisor \_\_\_\_\_  
Dr. Iosif Pinelis

Committee Member \_\_\_\_\_  
Dr. Qiuying Sha

Committee Member \_\_\_\_\_  
Dr. Yue Li

Department Chair \_\_\_\_\_  
Dr. Mark S. Gockenbach

Date \_\_\_\_\_

To my adviser, teachers and friends

who didn't hesitate to criticize my work at every stage - without which I'd neither be who

I am nor would this work be what it is today.

# Contents

<b>List of Figures</b> . . . . .	<b>vi</b>
<b>List of Tables</b> . . . . .	<b>x</b>
<b>Acknowledgments</b> . . . . .	<b>xi</b>
<b>Abstract</b> . . . . .	<b>xii</b>
<b>1 Background</b> . . . . .	<b>1</b>
1.1 Introduction . . . . .	1
1.2 Dipoles . . . . .	4
1.2.1 Types Of Dipoles . . . . .	4
1.2.2 Dipole Moment . . . . .	4
1.2.2.1 Electric Dipole Moment . . . . .	4
1.2.2.2 Molecular Dipole Moment . . . . .	5
1.2.2.3 Bond Dipole Moment . . . . .	6
1.3 Spontaneous Polymerization . . . . .	7
1.4 Boltzmann Factor, Partition Function, And State Probabilities . . . . .	8

<b>2</b>	<b>Modeling</b>	<b>11</b>
2.1	Ideal 2D Model [1]	13
2.2	Ideal 3D Model	20
2.3	Physical 3D Model	26
2.4	Range-adjusted Physical 3D Model	32
<b>3</b>	<b>Results &amp; Conclusions</b>	<b>36</b>
3.1	Results	36
3.1.1	Physical 3D model with $t=600$	36
3.1.2	Physical 3D model with $t=1500$	41
3.1.3	Range-adjusted physical 3D model with $t=1500$	49
3.2	Conclusions	57
3.2.1	Dominant Parameters	57
3.2.2	The Role Of Time Limit	63
3.3	Future Plans	63
	<b>References</b>	<b>65</b>
<b>A</b>	<b>Matlab Code</b>	<b>67</b>
<b>B</b>	<b>Generalized Dipole Moment And Potential Energy Of Dipole</b>	<b>91</b>
<b>C</b>	<b>Units Transformation</b>	<b>94</b>
<b>D</b>	<b>Taylor Series Expansion Of The Configuration Factor <math>c_F</math></b>	<b>98</b>

<b>E Monotonic Kernel Fitting Function . . . . .</b>	<b>100</b>
--	------------

# List of Figures

1.1	Chemical bond examples. These figures are licensed under the Creative Commons Attribution-Share Alike license [2], [3]. . . . .	7
(a)	Covalent bond . . . . .	7
(b)	Ionic bond . . . . .	7
2.1	Four of adjacent $20 \times 20$ blocks, of which three are shown with oriented cells	14
2.2	Dipole distribution at the beginning of the polymerization process within a $20 \times 20$ block in 2D. . . . .	19
2.3	Dipole distribution at the presumed end of the polymerization process within a $20 \times 20$ block in 2D with different parameter settings. . . . .	20
2.4	Dipole distribution at the beginning of the polymerization process within a $10 \times 10 \times 10$ block in 3D. Note for colors of the small arrows: blue–right, cyan–left, white–up, black–down, red–front, green–back. . . . .	24
2.5	Dipole distribution at the presumed end of the polymerization process within a $10 \times 10 \times 10$ block in 3D when $\mathbf{a} = \mathbf{0.5}, \beta = \mathbf{1}$ , relatively small dipole size, with high temperature . . . . .	25



2.6	Dipole distribution at the presumed end of the polymerization process within a $10 \times 10 \times 10$ block in 3D when $\mathbf{a} = \mathbf{0.5}, \beta = \mathbf{2}$ , relatively small dipole size, with low temperature . . . . .	26
2.7	Dipole distribution at the presumed end of the polymerization process within a $10 \times 10 \times 10$ block in 3D when $\mathbf{a} = \mathbf{0.9}, \beta = \mathbf{1}$ , relatively large dipole size, with high temperature . . . . .	27
2.8	Dipole distribution at the presumed end of the polymerization process within a $10 \times 10 \times 10$ block in 3D when $\mathbf{a} = \mathbf{0.9}, \beta = \mathbf{2}$ , relatively large dipole size, with low temperature . . . . .	28
3.1	Average length of all chains in physical 3D model with $t=600$ . Count both polymers and independent monomers as chains. . . . .	37
3.2	Average length of nontrivial chains in physical 3D model with $t=600$ . Count only polymers as chains. . . . .	38
3.3	Number of all chains in physical 3D model with $t=600$ . Count both polymers and independent monomers as chains. . . . .	39
3.4	Number of nontrivial chains in physical 3D model with $t=600$ . Count only polymers as chains. . . . .	40
3.5	Average length of all chains in physical 3D model with $t=1500$ . Count both polymers and independent monomers as chains. . . . .	42
3.6	Number of all chains in physical 3D model with $t=1500$ . Count both polymers and independent monomers as chains. . . . .	43

3.7	Average length of nontrivial chains in physical 3D model with $t=1500$ . Count only polymers as chains. . . . .	44
3.8	Number of nontrivial chains in physical 3D model with $t=1500$ . Count only polymers as chains. . . . .	45
3.9	Comparison between two measurements on average length of chains in physical 3D model with $t=1500$ . . . . .	46
3.10	Comparison between two measurements on number of chains in physical 3D model with $t=1500$ . . . . .	47
3.11	Actual time of polymerization in physical 3D model with $t=1500$ . . . . .	48
3.12	Average length of all chains in range-adjusted physical 3D model with $t=1500$ . Count both polymers and independent monomers as chains. . . . .	50
3.13	Number of all chains in range-adjusted physical 3D model with $t=1500$ . Count both polymers and independent monomers as chains. . . . .	51
3.14	Average length of nontrivial chains in range-adjusted physical 3D model with $t=1500$ . Count only polymers as chains. . . . .	52
3.15	Number of nontrivial chains in range-adjusted physical 3D model with $t=1500$ . Count only polymers as chains. . . . .	53
3.16	Comparison between two measurements on average length of chains in range-adjusted physical 3D model with $t=1500$ . . . . .	54
3.17	Comparison between two measurements on number of chains in range-adjusted physical 3D model with $t=1500$ . . . . .	55

3.18 Actual time of polymerization in range-adjusted physical 3D model with t=1500. . . . .	56
3.19 Quadratic polynomial fitting using data in $5 \leq \gamma < 20$ in final 3D model. . .	59
3.20 Kernel function fitting using data in $5 \leq \gamma < 20$ in final 3D model. . . . .	60
3.21 Comparison between quadratic polynomial and kernel fitting functions when $5 \leq \gamma < 20$ in final 3D model. . . . .	62

# List of Tables

2.1	Full descriptive information of polymerization under different setting of parameters. NOD:total number of dipoles; NOA: number of all chains; ALOA: average length of all chains; NONT: number of nontrivial chains; ALONT: average length of nontrivial chains. . . . .	33
2.2	$\gamma$ interval and corresponding parameters based on physical criteria. . . . .	35

## **Acknowledgments**

I would like to thank all those who have helped me learn, understand and appreciate this subject. Without your help, this thesis can never be done. Particularly, I appreciate for all the guide and advice from my adviser during the last two years. I learned much more than I could have imagined. Not only in academic area, but also in the way of critical thinking. Besides, I would love to thank all my professors giving those useful statistics courses. Those courses provided me new perspectives in various disciplines, such as genetics and programming. Above all, I appreciate the opportunity I received at Michigan Technological University and really enjoy the life in the past two years.

## **Abstract**

For human beings, the origin of life has always been an interesting and mysterious matter, particularly how life arose from inorganic matter through natural processes. Polymerization is always involved in such processes. In this paper we built what we refer to as ideal and physical models to simulate spontaneous polymerization based on certain physical principles. As the modeling confirms, without taking external energy, small and simple inorganic molecules formed bigger and more complicated molecules, which are necessary ingredients of all living organisms. In our simulations, we utilized actual ranges of parameters according to their experimentally observed values. The results from the simulations led to a good agreement with the nature of polymerization. After sorting out through all the models that were built, we arrived at a final model that, it is hoped, can be used to simply and efficiently describe spontaneous polymerization using only three parameters: the dipole moment, the distance between molecules, and the temperature.

# Chapter 1

## Background

### 1.1 Introduction

Many studies have been done in the area of evolution since several centuries ago, but advanced research on the origin of life started only in the 20th century, resulting from developments in modern microbiology and biochemistry. Biopoiesis (or abiogenesis) is the study of how life may have arisen from inorganic matter through natural processes. Particularly, the term is typically related with the rise of life on Earth.

The biologist John Desmond Bernal, who coined the term biopoiesis, suggested that there were three well defined "stages" that could be used to describe the origin of life [4]:

† Stage 1: The development of biological monomers

† Stage 2: The development of biological polymers

† Stage 3: The evolution from molecules to cell

Most amino acids, typically essential components of life, can form through natural and inanimate chemical reactions, proven by the Miller-Urey experiment [5], [6], which was focused on simulating hypothetical conditions of early Earth in a laboratory, and testing for the occurrence of chemical origins of life. Similar procedures happened in other fundamental biochemicals, like nucleotides and saccharides. These biochemicals usually are ingredients of more complex molecules, such as proteins, polysaccharides, and nucleic acids (DNA, RNA). These three molecules are crucial to all life functions and can be found in all living organisms.

As Timothy Gowers's proposed in his Polymath project [7], a simple model should be constructed to describe abiogenesis but not necessarily evolution. He mentioned two ideas related to this topic, Cellular Automaton [8] and Self-organized Criticality [9]. Based on these two ideas, there are several properties the model should have:

1. The model should be dynamic and follow simple rules.
2. The outcome of the modeled process should not be too random nor too simple, and it should depend to a large extent on whether certain critical values of parameters are



exceeded.

3. The model should have a randomized initial status, but it also should have a certain level of robustness in the outcome.
4. The model should be able to produce macroscopic structures, which can be easily identified and classified based on some interesting criteria.
5. There should be interactions among those structures, and the interaction should lead to an optimal consequence.
6. Nontrivial outcomes should eventually emerge with high probability. However, there might be some preconditions, which will limit the probability. After all, not all planets in the universe have life.

Motivated by this, we built probabilistic models in both two dimensions and three dimensions to describe abiogenesis in Stage 2. Before describing our models, we first introduced some fundamental concepts related to our models.

## 1.2 Dipoles

### 1.2.1 Types Of Dipoles

In nature, there are mainly two kinds of dipoles, electric and magnetic ones. In this thesis, we only focus on topics related to electric dipoles. The definition of an electric dipole is: a pair of opposite but equal electric charges separated by a small distance.

### 1.2.2 Dipole Moment

#### 1.2.2.1 Electric Dipole Moment

The dipole moment is the product of magnitude of each charge and the distance of separation between the charges:

$$\mathbf{p} = q\mathbf{d}, \tag{1.1}$$

where  $q$  is the absolute value of the two charges and  $\mathbf{d}$  is the displacement vector from the negative charge to the positive charge.

Also, one has the superposition principle: the total dipole moment of a system can be obtained through vector addition.

The dipole moment is an important attribute of a dipole, since it provides an overall polarity measure, which will in turn affect the performance of dipole. The following discussion is limited to static phenomena; neither time-dependent nor dynamic polarization will be covered.

### **1.2.2.2 Molecular Dipole Moment**

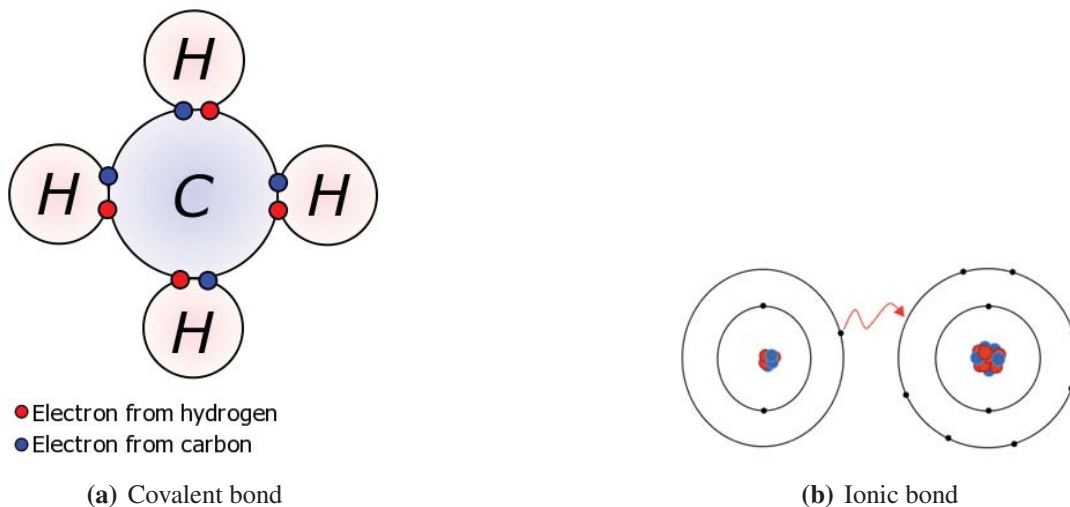
Many molecules have dipole moments because positive and negative charges are distributed nonuniformly over various atoms inside the molecules. Dipole moments exist in common molecules such as water, as well as in biomolecules such as proteins.

As mentioned in Section 1.2.2.1, we limit our discussion to static phenomena. Then our main concern is with permanent dipoles: there are two atoms in such a dipole, where one of them attracts more electrons, becoming more negative, and the other one becomes more positive, so that a stable dipole is formed. However, some molecules with dipolar groups inside do not show overall dipole moment at all because of cancellation of polarity. The highest molecular dipole moments are between 10 **D** and 11 **D** (measured in the units named *debye*, symbol: **D**) [10].

### 1.2.2.3 Bond Dipole Moment

Another important concept is the bond dipole moment, which is closely associated with the molecule dipole moment. A chemical bond usually forms when the balance between the forces of repulsion and electrostatic attractions is attained in the binding region. There are two kinds of chemical bond, ionic and covalent ones. A covalent bond can be recognized by the sharing of electrons between atoms. The prefix *co-* indicates the atoms both "own" the electrons. On the contrast, ionic bond is formed as one atom "takes away" electrons from the other. However, pure ionic bonding does not exist since the electrons can not be completely privately-owned in the binding region. Therefore, ionic and covalent binding are two extreme cases of reaching this electrostatic equilibrium. The bond dipole exists if a difference in polarity across the bond exists. Bond dipole moment can be calculated in the same way as electric dipole moment, which provides the measurement of the polarity of a chemical bond. Between the two extremes, ionic and covalent binds, a complete spectrum of bond dipole moment densities exists [11].

The total dipole moment of all covalent bonds in a molecule will be zero if the charge distribution is symmetric which will cancel the molecular polarity. Therefore dipole moments can take values from zero to the ionic extreme which approaches the product  $ner$  (where  $n$  is the number of electrons moved during the formation process of the ionic bond,  $e$  is the charge of electron,  $r$  is the bond length) [11]. Typical dipole moments for



**Figure 1.1:** Chemical bond examples. These figures are licensed under the Creative Commons Attribution-Share Alike license [2], [3].

simple diatomic molecules (a bond can be modeled as diatomic molecule) are in the interval of 0 to 11  $\mathbf{D}$  [10].

### 1.3 Spontaneous Polymerization

"Polymerization is a process of monomer molecules reacting together in a chemical reaction to form three-dimensional networks or polymer chains" [12]. In general, polymers that consist of repeated long chains or structures of the same monomer unit are referred to as homopolymers, whereas polymers that consist of at least two kinds of monomer molecules are referred to as copolymers. In this project, we only deal with homopolymers with chain-growth polymerization, which is a polymerization process where unsaturated monomer molecules add on to the beginning or end of a growing polymer chain one at

a time [12]. It involves the linking together of molecules incorporating chemical bonds, which means the properties of chemical bonds play a crucial role in the process of polymerization. There is another type of polymerization mechanism called step-growth polymerization, which involves reactions between functional groups of monomers, but it is not considered here.

## 1.4 Boltzmann Factor, Partition Function, And State Probabilities

For a point charge  $q$ , if the potential  $V$  is known at this point, then the potential energy this point charge has is

$$U = qV. \tag{1.2}$$

The potential  $V$  at a given position is generated by all charges in a system except the charge at the given position. Both the potential and potential energy in the above equation are scalar and additive

$$V = \sum_j V_j \text{ and } U = \sum_j U_j, \tag{1.3}$$

where  $V_j$  is the potential generated by  $j$ th charge at the given position; similarly for  $U_j$ .

For a dipole with charges of equal magnitude but opposite sign, the potential energy it has can be written as

$$U = (+q)V_+ + (-q)V_- \quad (1.4)$$

$V_+$  is the potential at  $+q$ ,  $V_-$  is the potential at  $-q$ .

In physics, the Boltzmann factor is defined as:

$$e^{-\frac{U}{k_B T}} = e^{-\beta U} \quad (1.5)$$

where

$$\beta = \frac{1}{k_B T},$$

$k_B$  is known as the Boltzmann constant,  $T$  is temperature. Then  $\beta$  is inversely proportional to temperature.

In a system with only discrete values of energy  $U_i$ , the partition function is

$$Z = \sum_i e^{-\frac{U_i}{k_B T}} = \sum_i e^{-\beta U_i}. \quad (1.6)$$

Systems in equilibrium at temperature  $T$  have probability  $P_i$  of occupying state  $i$  with

energy  $U_i$ , where

$$P_i = \frac{e^{-\beta U_i}}{Z} = \frac{e^{-\beta U_i}}{\sum_i e^{-\beta U_i}}. \quad (1.7)$$

Based on the above discussion, it is obvious that, if a dipole has a large potential energy, then it is unlikely to stay in state  $i$  for long. Instead, it will move to a position or orientation which have a lower potential energy. This follows from the law of conservation of energy, since the energy for movement is from the decrease of the potential energy. By (1.7), we can calculate the probability of a specified dipole's next movement given the locations and directions of the particular dipole and its neighbors.



# Chapter 2

## Modeling

As stated in Section 1.1, the models may reflect the two ideas, Cellular Automaton [8] and Self-organized Criticality [9].

Cellular automata are mathematical idealizations of a system. The time and space variables in this system are discrete. The space usually contains a regular uniform lattice, which can be extended repeatedly and infinitely, with a value of discrete variable at each site. The system's state can be identified by the values of the variables. A cellular automaton evolves in step of a discrete time variable. For example, moving from state  $i$  to state  $i + 1$  can be considered as one step. The variable value at one site is affected by the values of its neighbors, which are defined as all immediately adjacent sites. The updates of variable values at each site are determined by the values of their neighbors at the previous time step,

and based on a definite set of "local rules" [8].

Bak, Tang and Wiesenfeld's 1987 paper [9] demonstrated that in a self-organized system observed complexity will emerge regardless of details of the system: the emergence of critical behavior does not depend on the model parameters. Self-organized criticality in our topic generally implies the system will be developing spontaneously, and will eventually stabilize, regardless of the system's parameters.

According to the discussion above, our model will consist of a regular uniform lattice, in which there will be several blocks of the same size. The blocks of the lattice are assumed to be adjacent to one another so as to extend repeatedly and infinitely. At each site (or cell), there will be a dipole in it, with a few orientations, such as up, down, left, right in the case of a two-dimensional lattice. The local interaction, which only occurs within the immediately adjacent neighbors of a site, can be determined by (1.2) and (1.4).

Since potential energy is a scalar quantity, it is possible to sum up the energy of all interactions between the central dipole and its neighbors. Thereafter, we can predict the movement of a specified dipole using (1.7). To be more specific, there are two kinds of movements:

1. Orientation change: the dipole stays in the current site, only its orientation changes.
2. Position change: the dipole moves to an empty adjacent site from the previous one.

The orientation change is relatively simple to consider. We just need to calculate the potential energies of the particular dipole with all possible six orientations (in the 3D case) based on the configuration of its current neighbors. The position change is a little more complicated. The dipole will move to an empty adjacent site and leave its current site empty. Consequently, we need to calculate potential energies of this dipole with all possible orientations at every possible new site based on the distribution of its new neighbors.

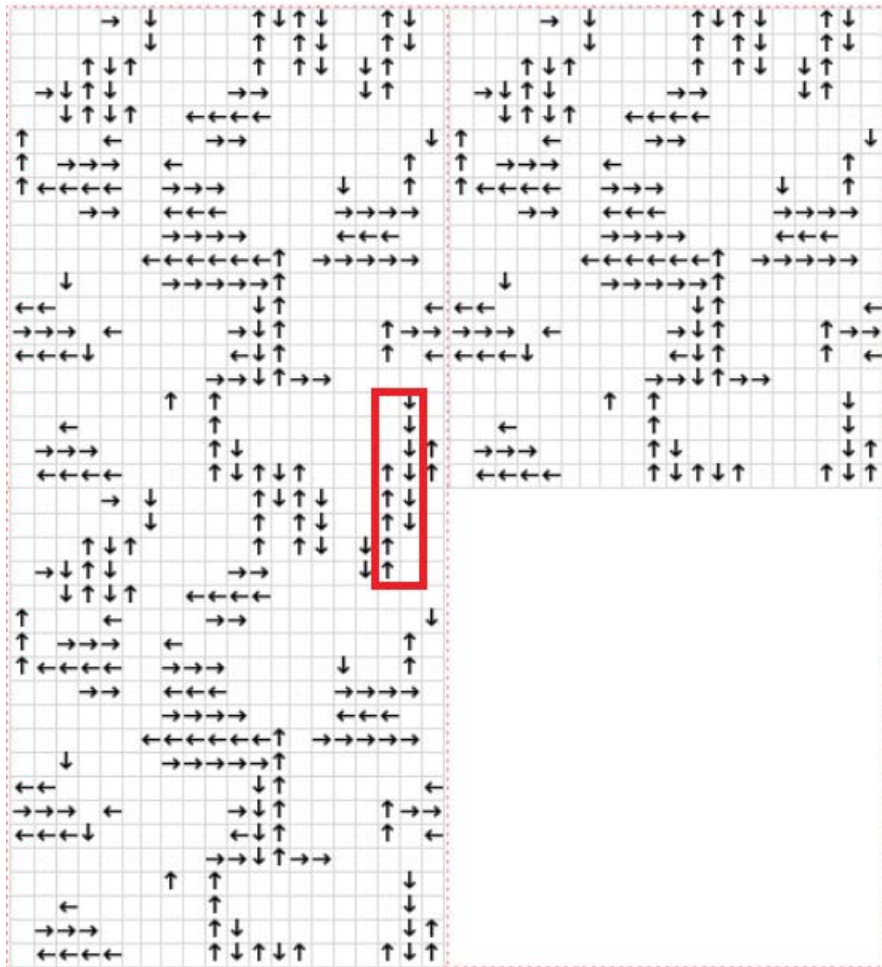
A dipole in a specified position with a specified orientation is in a certain state, and it has a corresponding potential energy level and a Boltzmann factor level. After listing all possible states the dipole could move into, we can estimate the probability of a given movement of this dipole.

Let us say that such a movement was "successful" if the dipole did change either its orientation or position. Each successful movement will also be referred to as a round for brevity.

## **2.1 Ideal 2D Model [1]**

In this case, each molecule is modeled as a small square in a lattice. It is polarized, that is, has a certain dipole moment, which may be oriented right, left, up, or down. The small squares form a grid, thus together constituting a big square (2D lattice). The big square

is one of the uniform blocks in the whole system, which is a periodic extension from the square to  $\mathbf{R}^2$ . This is illustrated in Figure 2.1.



**Figure 2.1:** Four of adjacent  $20 \times 20$  blocks, of which three are shown with oriented cells

In a two dimensional space, the potential of an electric charge  $q$  at distance  $r$  from it can be simplified as:

$$V = -q \ln r \tag{2.1}$$

if  $q$  is measured in an appropriate unit (otherwise, there may be an extra constant factor present).

Suppose  $a$  is the ratio of the length of the dipole  $d$  to the length of the square side  $l$ ,

$$a = \frac{d}{l},$$

so that  $0 \leq a < 1$ .

For simplicity, assume for now that every cell is a unit square with  $l = 1$ , and the charge is also unit:  $q = 1$ .

According to the dipoles' orientations and positions, there are several different potential energies of interaction between two dipoles:

$$1. \quad \begin{array}{|c|c|} \hline \rightarrow & \leftarrow \\ \hline \end{array} \quad \ln \frac{1}{1-a^2}$$

$$2. \quad \begin{array}{|c|} \hline \rightarrow \\ \hline \rightarrow \\ \hline \end{array} \quad \ln(1+a^2)$$

$$3. \quad \begin{array}{|c|} \hline \rightarrow \\ \hline \uparrow \\ \hline \end{array} \quad 0$$

$$4. \quad \begin{array}{|c|c|} \hline & \rightarrow \\ \hline \rightarrow & \\ \hline \end{array} \quad \ln \sqrt{1 + \frac{a^2}{4}}$$

$$5. \quad \begin{array}{|c|c|} \hline & \downarrow \\ \hline \rightarrow & \\ \hline \end{array} \quad \ln \frac{4+a^2}{4-a^2}$$

The energy of interaction of the above two head-to-head dipoles (Pattern 1) is

$$-\ln(1+a) - \ln(1-a) + \ln 1 + \ln 1 = \ln \frac{1}{1-a^2}$$

The energy of following patterns can be obtained similarly.

Remark 1: If any one of a pair of dipoles changes its orientation to the opposite one, the energy changes only in sign.

Remark 2: If any of the five above patterns is rotated, the energy does not change.

When  $0 < a < 1$ , one has the following inequalities:

$$\frac{1}{1-a^2} > 1+a^2 > \frac{4+a^2}{4-a^2} > \sqrt{1+\frac{a^4}{4}} > 1$$

This indicates that the energies of the five above cases follow the sequence

$$\text{pattern 1} > \text{pattern 2} > \text{pattern 5} > \text{pattern 4} > \text{pattern 3} ,$$

which implies that dipoles are most likely to be aligned as "head to tail" (opposite to pattern 1, which has lowest potential energy), since the system tends to have the lowest potential energy (at which state it is stable).

As stated at the beginning of Section 2.1, the cell size is set to be 1 by 1, with no unit. From (2.1), (1.4), the Boltzmann factor becomes

$$e^{-\beta U_i} = e^{-\beta f_i(a)}, \tag{2.2}$$

where  $f_i(a)$  is the potential energy of the given dipole in state  $i$ , which can be obtained by summing up the quantities in the corresponding patterns above. The expression of  $f_i(a)$  is determined by the configuration of the given dipole and its neighbors. Therefore, given the distribution of dipoles, we only need the values of  $a, T$  to determine the corresponding

probability. Then (1.7) can be written as

$$P_i = \frac{e^{-\beta f_i(a)}}{\sum_i e^{-\beta f_i(a)}}, \quad (2.3)$$

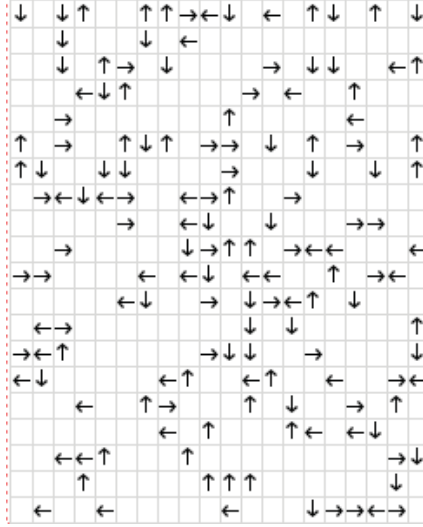
where the value of  $a$  reflects the relation between the dipole length and the molecule size,  $\beta$  is inversely proportional to the temperature, refer to

$$\beta = \frac{1}{k_B T}.$$

It is necessary to clarify the definition of the polymer length. The polymer length in this project means the number of dipoles in the polymer. Inside the red box in Figure 2.1, there are two polymers of length 5 and 6, respectively. If we only take the top left block into account, then the polymer in the 19th column oriented down, inside the red box will be cut into two parts, of lengths 2 and 4. So, we need to take the periodicity convention. In particular if a polymer takes an entire row or column of the 20 by 20 square, its length is taken to be 20, even though its "actual" length might be considered to be infinity.

Before polymerization, all the dipoles in the lattice are randomly distributed and oriented. The positions and orientations of dipoles are arbitrary with no order. There seldom will be polymers with length more than 3. Such a random initial distribution of dipoles is shown in Figure 2.2



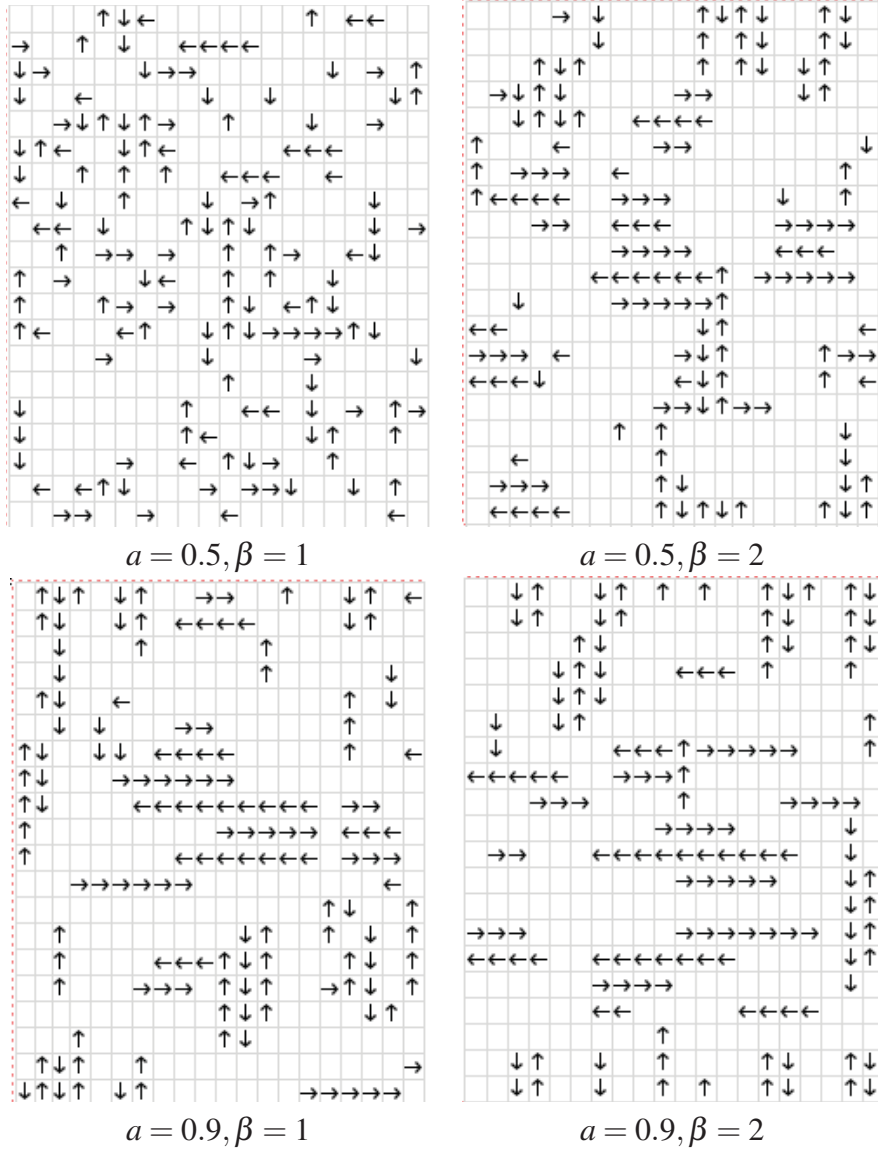


**Figure 2.2:** Dipole distribution at the beginning of the polymerization process within a  $20 \times 20$  block in 2D.

We simulated polymerization by setting qualitative parameters. Here are four cases:

- †  $a = 0.5, \beta = 1$ , relatively small dipole size, with high temperature.
- †  $a = 0.5, \beta = 2$ , relatively small dipole size, with low temperature.
- †  $a = 0.9, \beta = 1$ , relatively large dipole size, with high temperature.
- †  $a = 0.9, \beta = 2$ , relatively large dipole size, with low temperature.

After 1000 successful dipole movements in the polymerization process, many more polymers are seen. However, Figure 2.3 shows that, when  $a = 0.5$  and  $\beta = 1$  (small dipole size and high temperature) the degree of polymerization is not as high as in the other three cases during a certain time period.



**Figure 2.3:** Dipole distribution at the presumed end of the polymerization process within a  $20 \times 20$  block in 2D with different parameter settings.

## 2.2 Ideal 3D Model

Of course, the 3D case is much more realistic. In this case, each molecule is modeled as a small unit cube. Inside each cube there is a dipole, which may be oriented right, left, up,

down, front, or back. The small cubes together form a big cube (3D lattice). The big cube is one of the uniform blocks in the whole system, which is the periodic extension of the cube to  $\mathbf{R}^3$ .

The 3D case is in principle very similar to the 2D one. The main difference is the formula for the potential. Unlike (2.1), in 3D space the potential is defined as

$$V = q/r \tag{2.4}$$

if  $q$  and  $r$  are expressed in appropriate units of measurement; see Appendix C.

Suppose  $a$  is the ratio of the length of the dipole  $d$  to the length of the cube side  $l$ :

$$a = \frac{d}{l},$$

so that  $0 \leq a < 1$ .

For simplicity, assume for now that every cell is a unit cube with  $l = 1$ .

According to the dipoles' orientations and positions, there are several different potential energies of interaction between two dipoles:

$$1. \quad \begin{array}{|c|c|} \hline \rightarrow & \leftarrow \\ \hline \end{array} \quad q^2 \left( \frac{2}{1-a^2} - 2 \right)$$

$$2. \quad \begin{array}{|c|} \hline \rightarrow \\ \hline \rightarrow \\ \hline \end{array} \quad q^2 \left( 2 - \frac{2}{\sqrt{1+a^2}} \right)$$

2D Patterns:

$$3. \quad \begin{array}{|c|} \hline \rightarrow \\ \hline \uparrow \\ \hline \end{array} \quad 0$$

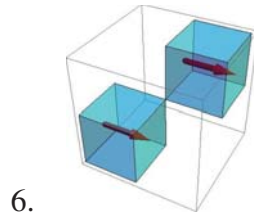
$$4. \quad \begin{array}{|c|c|} \hline & \rightarrow \\ \hline \rightarrow & \\ \hline \end{array} \quad q^2 \left( \sqrt{2} - \frac{\sqrt{1+(1-a)^2} + \sqrt{1+(1+a)^2}}{\sqrt{a^4+4}} \right)$$

$$5. \quad \begin{array}{|c|c|} \hline & \downarrow \\ \hline \rightarrow & \\ \hline \end{array} \quad q^2 \left( \frac{2\sqrt{2}}{2-\frac{a^2}{2}} - \frac{2}{\sqrt{2+\frac{a^2}{2}}} \right)$$

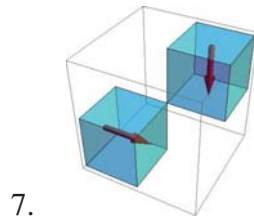
For the first pattern,

$$\frac{q^2}{1-a} + \frac{q^2}{1+a} - \frac{q^2}{1} - \frac{q^2}{1} = q^2 \left( \frac{2}{1-a^2} - 2 \right).$$

The energy of following patterns can be obtained similarly.

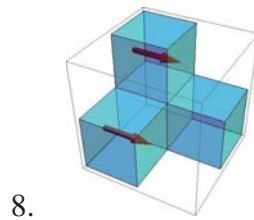


$$q^2 \left( \frac{2}{\sqrt{3}} - \frac{1}{\sqrt{3+a^2-2a}} - \frac{1}{\sqrt{3+a^2+2a}} \right)$$

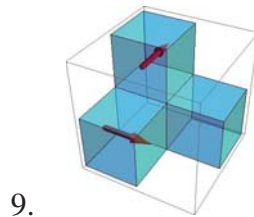


$$q^2 \left( \frac{1}{\sqrt{3+\frac{a^2}{2}+2a}} + \frac{1}{\sqrt{3+\frac{a^2}{2}-2a}} - \frac{2}{\sqrt{3+\frac{a^2}{2}}} \right)$$

3D Patterns:



$$q^2 \left( \sqrt{2} - \frac{2}{\sqrt{a^2+2}} \right)$$

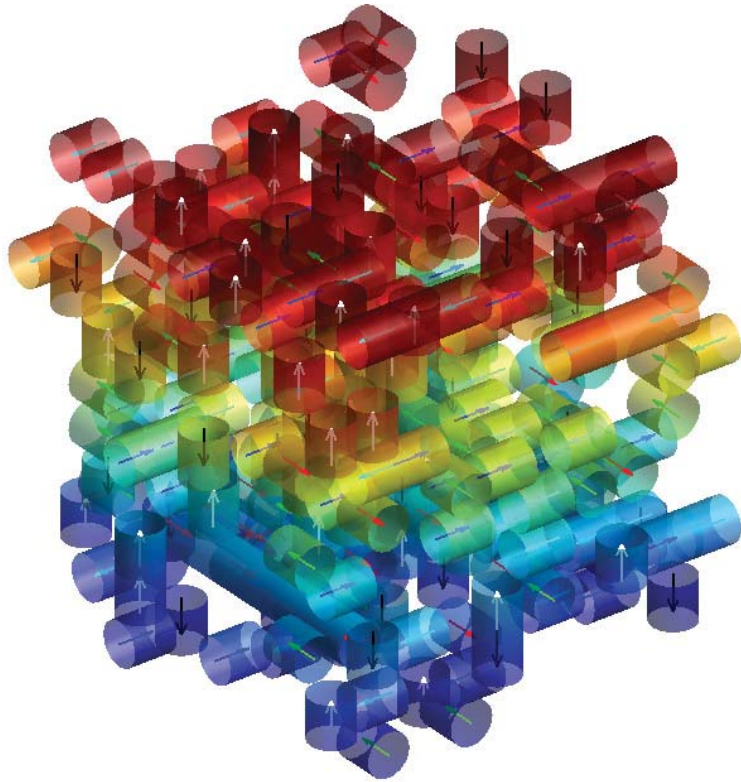


0

Remark 1: If any one of a pair of dipoles changes its orientation to the opposite one, the energy changes only in sign.

Remark 2: If any of the nine above patterns is rotated, the energy does not change.

In order to get a clear data visualization image, we set the block cube size to be 10 by 10 by 10, and the probability of any given cell being occupied by a dipole to be 0.05. In three

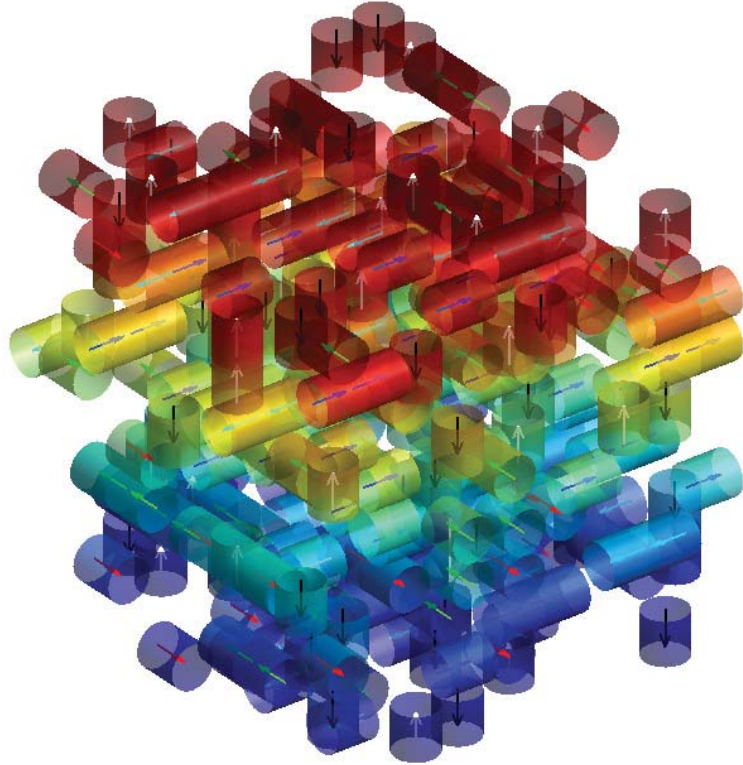


**Figure 2.4:** Dipole distribution at the beginning of the polymerization process within a  $10 \times 10 \times 10$  block in 3D. Note for colors of the small arrows: blue–right, cyan–left, white–up, black–down, red–front, green–back.

dimensions, formula (2.3) still applies. By setting appropriate values of  $a$  and  $\beta$ , qualitative understanding of polymerization can be obtained.

As shown in Figure 2.4, initially all the dipoles and their orientations in the cube are randomly distributed. In three dimensions, there are too many cells obscuring one another to get a complete visualization image, even though the cube is 10 by 10 by 10 cells (whereas the square was 20 by 20 cells).

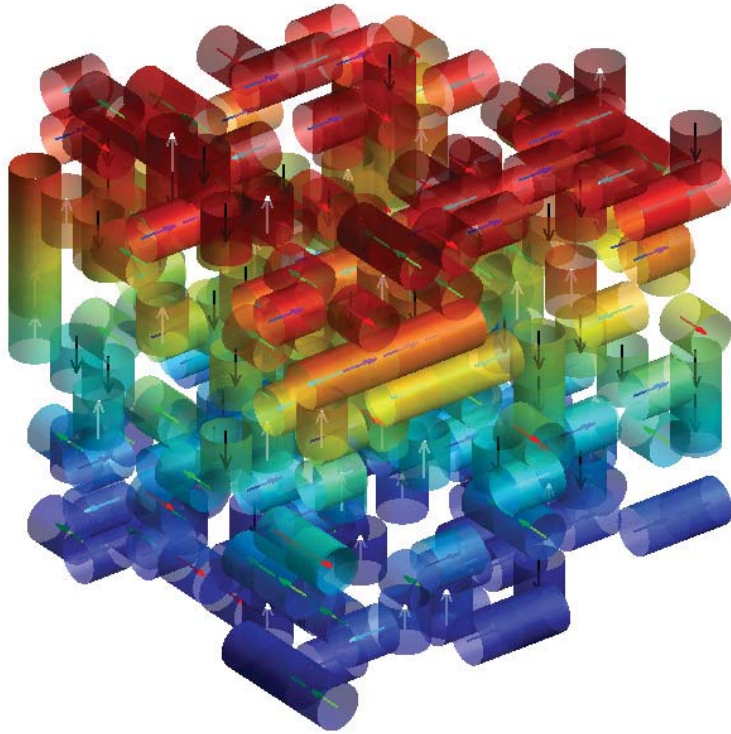
Use same  $a, \beta$  values as Section 2.1, the visualization images of dipoles after 600 successful



**Figure 2.5:** Dipole distribution at the presumed end of the polymerization process within a  $10 \times 10 \times 10$  block in 3D when  $\mathbf{a} = \mathbf{0.5}, \beta = \mathbf{1}$ , relatively small dipole size, with high temperature

movements are shown as in Figure 2.5 to Figure 2.8 .

From Figure 2.5 and Figure 2.6 , when the dipole length is relatively small ( $a = 0.5$ ) it is not easy to distinguish after-polymerization from prior-polymerization (Figure 2.4), nor to tell the difference between high temperature ( $\beta = 1$ ) and low temperature ( $\beta = 2$ ). However, there are significantly good degrees of polymerization when the dipole length is relatively large ( $a = 0.9$ ) (Figure 2.7 and Figure 2.8). The difference between high temperature ( $\beta = 1$ ) and low temperature ( $\beta = 2$ ) when  $a = 0.9$  is still hard to tell based only on these 3D images.



**Figure 2.6:** Dipole distribution at the presumed end of the polymerization process within a  $10 \times 10 \times 10$  block in 3D when  $\mathbf{a} = 0.5, \beta = 2$ , relatively small dipole size, with low temperature

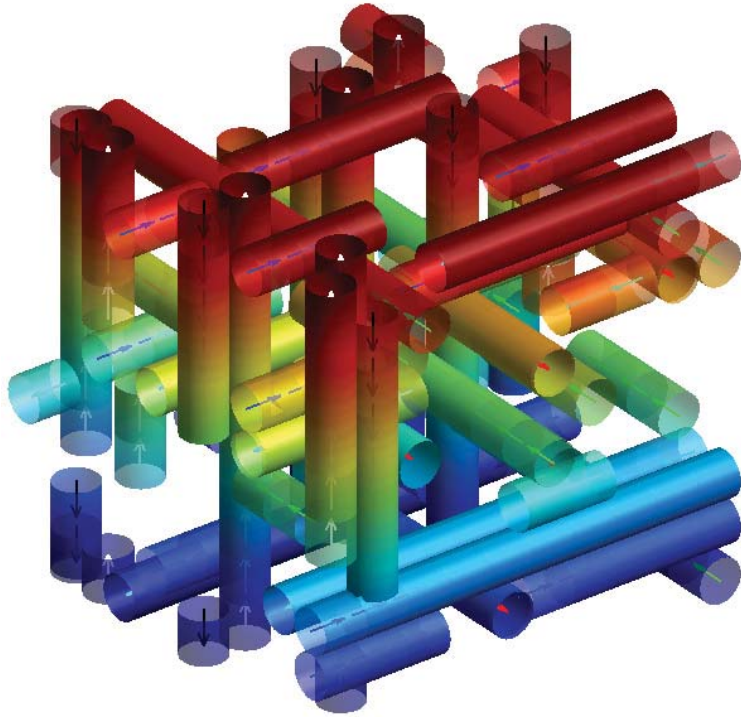
## 2.3 Physical 3D Model

So far, our main focus was to build ideal models describing the process of polymerization, with  $\beta$  value chosen arbitrarily. Now, we want to tie the model with real physical world.

There are three issues should be taken into consideration:

- a). Use real physical parameters to describe the model more accurately.
- b). Set restrictive conditions based on certain physical principles.





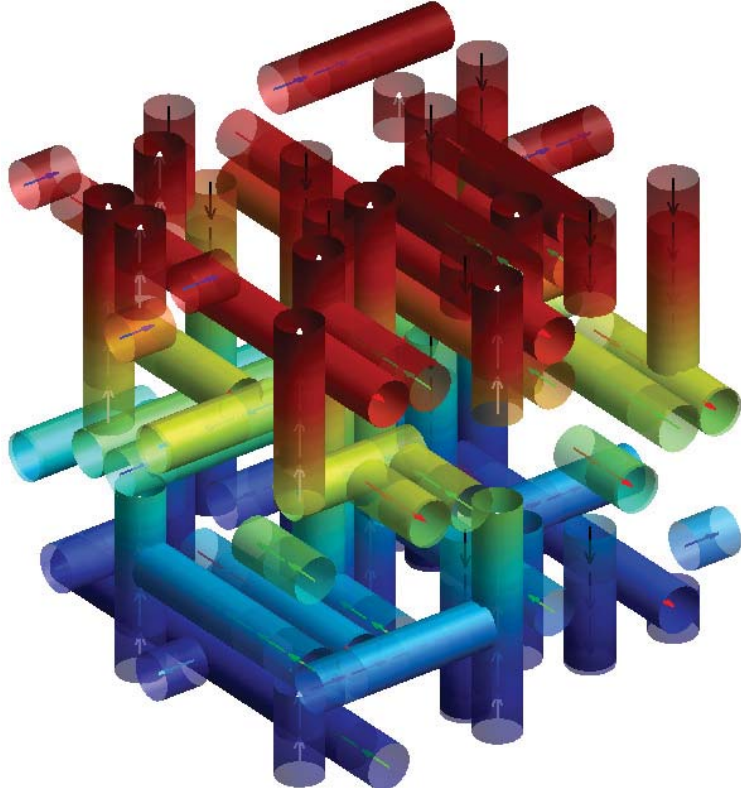
**Figure 2.7:** Dipole distribution at the presumed end of the polymerization process within a  $10 \times 10 \times 10$  block in 3D when  $\mathbf{a} = \mathbf{0.9}$ ,  $\beta = \mathbf{1}$ , relatively large dipole size, with high temperature

c). Need quantitative measurements to evaluate degree of polymerization besides graphics.

**Issue a):** We need to define variables in appropriate units.

†  $l$ : molecule size (or generally distance between two molecules), measured in  $\text{\AA}$ ,  $10^{-8}\text{cm}$ . for instance,  $l = 5$  means the molecule size is  $5 \text{\AA}$ . Then dipole length  $d = la$ .

†  $v$ : the number of elementary charge,  $|q| = ve$ ,  $v \in \mathbf{R}^+$ , where  $e$  is elementary charge



**Figure 2.8:** Dipole distribution at the presumed end of the polymerization process within a  $10 \times 10 \times 10$  block in 3D when  $\mathbf{a} = \mathbf{0.9}$ ,  $\beta = \mathbf{2}$ , relatively large dipole size, with low temperature

or charge of electron.

†  $M_d$ : "scale of dipole moment",  $M_d = vd = vla$ . This is transformed scale of typical dipole moment shown in Section 1.2.2. The relation between dipole moment and "scale of dipole moment" is  $M_d = \frac{\|\mathbf{p}\|}{e\text{\AA}}$ .

According to (2.4) and nine potential energy patterns in Section. 2.2, the potential energy of a given dipole in state  $i$  in the new physical model is

$$U_i = f_i(a) \cdot \frac{(v)^2}{l} \quad (2.5)$$

$$= f_i(a) \cdot \frac{(vla)^2}{a^2 l^3} \quad (2.6)$$

$$= g_i(a) \cdot \frac{M_d^2}{l^3}, \quad (2.7)$$

where  $f_i(a)$  is a function of configuration of the given dipole and its neighbor at state  $i$ , which can be obtained by summing up the quantities in the corresponding nine patterns in Section. 2.2. Set  $g_i(a) = \frac{f_i(a)}{a^2}$ . Further discussion about (2.7) is in Appendix B.

Based on (1.7) and (2.7), after coordinating various units among all parameters, Boltzmann factor can be written as

$$e^{-\beta U_i} = e^{-\beta \frac{M_d^2}{l^3} g_i(a)} \quad (2.8)$$

$$= e^{-\gamma g_i(a)}. \quad (2.9)$$

As shown in Appendix C, in our units system,  $\gamma = 557 \cdot \frac{300\text{K}}{T} \cdot \frac{M_d^2}{l^3}$ , see (C.11). 557 here is a constant without unit to coordinate the units between our units system and standard CGS system. Define  $\varepsilon = \frac{M_d}{l^{\frac{3}{2}}}$  as polarization coefficient.

If we substitute  $\beta$  and  $f_i(a)$  in (2.3) by  $\gamma$  and  $g_i(a)$ , then state probability is  $P_i = \frac{e^{-\gamma g_i(a)}}{\sum_i e^{-\gamma g_i(a)}}$ . Unlike  $\beta$ , which is only inversely associated with temperature,  $\gamma$  contains more information about a system, like the dipole moment, the distance between molecules and the temperature. This setting of parameters appears to be a good agreement with realistic phenomena.

**Issue b):** The following restrictive conditions should be set:

1.  $a < 0.5$ . In our model, dipole consists of two opposite point charges. To make sure the dipole is stable,  $a$  should be less than 0.5. Or the point charges might be attracted to another cell.
2.  $5 \leq l \leq 10$ . Molecules presented in organic synthesis usually have a dimension starting from a few Å. Our main concern is not those macromolecules (like DNA) which have already finished polymerization, but small molecules which are going to be involved in the process of polymerization. Therefore we set  $l$  be in the range of 5 Å to 10 Å.
3.  $0 < M_d \leq 2.1$ . As mentioned in Section 1.2.2.3, typical dipole moment for a chemical

bond is from 0 **D** to 11 **D**. However, the highest range 10 **D** to 11 **D** only exists in a highly ionic bond, which always contains a heavy metal atom that usually does not present in organic system. Then in our model the bond dipole moment is set to be in the range of 0 **D** to 10 **D**.  $M_d$  here has already been adjusted according to our units system, see Appendix C.

4.  $300 \leq T \leq 600$ . Set temperature to be in the range of 300K to 600K. Although 600K is rare in nature, it is still possible, like around volcanic vent.

Based on the above criteria, value of  $\gamma$  falls in the interval  $(0, 19.651]$ .  $\gamma$  can not be zero, which means no polarity and hence polymerization will not happen. The values of  $a$  are set to be in three levels: 0.1, 0.2, 0.3. When  $a$  takes values from these three levels,  $g_i(a)$  mentioned in (2.7) and (2.9) becomes a quantity  $c_F$  which only slightly depends on  $a$ , see Appendix D.

**Issue c):** Even the graphics will give a general idea about the degree of polymerization, it is not as informative as a quantitative measurement, such as a table describing numbers of polymer chains with different length and orientations, with which we can extract more information about degree of polymerization, see Table 2.1.

There are two reasonable way to define chains. First, all polymers and independent monomers can be considered as chains, we call this "all chains". Second, only polymers with length greater than or equal to two can be called chains, we call this "nontrivial

chains". Hence, there are two measurements for average length of chains and number of chains.

From Table. 2.1, NOA is identical to NONT, while ALOA is as same as ALONT. The reason is there is no independent monomer in all the six tables above, which leads to no difference between those two measurements. Meanwhile, it is obvious that when  $\gamma = 5$  the degree of polymerization is higher than  $\gamma = 13$  according to number of longer chains. regardless of  $a$  levels.

However, there are only two values of  $\gamma$  and corresponding information here. It does not provide enough information about the effect of  $\gamma$  on degree of polymerization. More data is needed to make reliable inference. Besides, Table 2.1 gives a good detailed view but is not ideal for comparisons. New measurement should be developed to monitor degree of polymerization among all  $\gamma$  values.

## **2.4 Range-adjusted Physical 3D Model**

In the previous section, we focused on several physical criteria on molecular level, like the distance between molecules, dipole moment. To be more accurate, we are taking the experimentally observed ranges of the parameters on atomic level into consideration. Due to our assumption, polymerization could only happen when dipole moments exist. Thereby

**Table 2.1**

Full descriptive information of polymerization under different setting of parameters. NOD:total number of dipoles; NOA: number of all chains; ALOA: average length of all chains; NONT: number of nontrivial chains; ALONT: average length of nontrivial chains.

	$\gamma = 5$							$\gamma = 13$						
	chain length	Right	Left	Up	Down	Front	Back	chain length	Right	Left	Up	Down	Front	Back
$a = 0.1$	1	0	0	0	0	0	0	1	0	0	0	0	0	0
	2	0	0	0	1	1	1	2	1	2	2	4	2	1
	3	0	1	2	3	1	1	3	3	2	1	1	1	2
	4	0	0	1	2	4	1	4	2	4	0	0	3	2
	5	0	0	2	3	0	2	5	2	1	1	1	1	1
	6	0	0	1	0	2	1	6	0	0	0	0	2	4
	7	0	0	0	0	1	0	7	0	0	1	1	1	0
	8	0	0	0	0	0	0	8	0	0	0	0	4	2
	9	0	0	0	0	0	0	9	0	0	0	0	0	0
	10	1	1	6	4	2	3	10	0	0	0	0	2	4
	NOD	10	13	86	74	60	55	NOD	29	31	19	23	95	101
NOA	1	2	12	13	11	9	NOA	8	9	5	7	16	16	
ALOA	10	6.5	7.17	5.69	5.45	6.11	ALOA	3.63	3.44	3.8	3.29	5.94	6.31	
NONT	1	2	12	13	11	9	NONT	8	9	5	7	16	16	
ALONT	10	6.5	7.17	5.69	5.45	6.11	ALONT	3.63	3.44	3.8	3.29	5.94	6.31	
$a = 0.2$	1	0	0	0	0	0	0	1	0	0	0	0	0	0
	2	0	1	0	0	1	2	2	0	0	2	0	4	2
	3	1	1	2	0	0	0	3	2	1	2	2	0	3
	4	0	0	0	2	1	0	4	3	4	5	3	3	2
	5	0	0	0	0	1	1	5	3	2	0	3	4	2
	6	0	0	2	2	2	2	6	0	0	1	0	1	2
	7	0	0	0	1	1	0	7	0	1	0	0	0	0
	8	0	0	0	0	1	2	8	0	0	1	3	0	0
	9	0	0	0	0	0	0	9	0	0	0	0	1	0
	10	3	2	3	2	4	3	10	1	0	1	1	0	0
	NOD	33	25	48	47	78	67	NOD	43	36	54	67	55	43
NOA	4	4	7	7	11	10	NOA	9	8	12	12	13	11	
ALOA	8.25	6.25	6.86	6.71	7.09	6.7	ALOA	4.78	4.5	4.5	5.58	4.23	3.91	
NONT	4	4	7	7	11	10	NONT	9	8	12	12	13	11	
ALONT	8.25	6.25	6.86	6.71	7.09	6.7	ALONT	4.78	4.5	4.5	5.58	4.23	3.91	
$a = 0.3$	1	0	0	0	0	0	0	1	0	0	0	0	0	0
	2	2	2	0	0	0	0	2	3	3	0	4	3	3
	3	2	1	0	1	0	0	3	1	3	1	1	2	0
	4	1	1	2	1	0	0	4	2	2	1	1	0	0
	5	0	0	1	0	2	0	5	0	1	1	1	0	2
	6	1	0	0	0	0	2	6	1	1	1	1	2	3
	7	1	3	1	2	1	1	7	1	0	1	2	0	0
	8	0	0	1	2	1	0	8	0	0	1	1	1	1
	9	0	0	0	0	0	0	9	0	0	1	0	0	0
	10	1	1	2	2	3	4	10	0	0	2	2	2	1
	NOD	37	42	48	57	55	59	NOD	30	34	62	68	52	52
NOA	8	8	7	8	7	7	NOA	8	10	9	13	10	10	
ALOA	4.63	5.25	6.86	7.13	7.86	8.43	ALOA	3.75	3.4	6.89	5.23	5.2	5.2	
NONT	8	8	7	8	7	7	NONT	8	10	9	13	10	10	
ALONT	4.63	5.25	6.86	7.13	7.86	8.43	ALONT	3.75	3.4	6.89	5.23	5.2	5.2	

we do not take covalent bond without polarity into account, but polar covalent bond and ionic bond, which two can be simply modeled as diatomic molecules. Typical amount of charges in such bond are in the range of 0 to  $3e$ . The smallest bond is 'H-H', the simplest hydrogen bond, with length  $0.74 \text{ \AA}$  [13].

In addition to the restrictive conditions mentioned in Section 2.3, the new model should satisfy the following criteria:

1.  $a < 0.5$ .
2.  $1 \leq l \leq 10$ .
3.  $la > 0.7$ .
4.  $0 < M_d = vla \leq 2.1$ .
5.  $0 < v \leq 3$ .
6.  $300 \leq T \leq 600$ .

In the range-adjusted physical model, set the shortest dipole length to be  $0.7 \text{ \AA}$ , the molecule size to be in the range  $1 \text{ \AA}$  to  $10 \text{ \AA}$  (wider than the one mentioned in Section 2.3),  $v$  to be any positive real number less than 3.

Section 2.3 explained the limitation of dipole distribution table. Then another measurement



**Table 2.2**

$\gamma$  interval and corresponding parameters based on physical criteria.

$a$	$\gamma_{min}$	$M_d$	$\gamma_{max}$	$l$	$M_d$	$T$
0.1	0 <sub>+</sub>	0 <sub>+</sub>	7.16	7	2.1	300
0.2	0 <sub>+</sub>	0 <sub>+</sub>	57.29	3.5	2.1	300
0.3	0 <sub>+</sub>	0 <sub>+</sub>	193.36	2.33	2.1	300

which can illustrate the changes related to various  $\gamma$  values is urgently needed. From dipole distribution tables, it is easy to extract total number of polymer chains for all six orientations and corresponding average length of chains. Of course, there are still two measurements for these two quantities, see Section 2.3 **Issue c**).

Table 2.2 shows the interval of  $\gamma$  and corresponding parameter values when  $\gamma$  reach its boundary. For example when  $a = 0.1$ ,  $\gamma$  is close at 0 when  $M_d$  is close to 0, and  $\gamma$  reach its maximum 7.16 when  $l = 7, M_d = 2.1, T = 300$ . Using  $\gamma$  intervals shown in Table.2.2, we are able to record all information which shows the effect of  $\gamma$  on the degree of polymerization.

The results are shown in the next chapter.

# Chapter 3

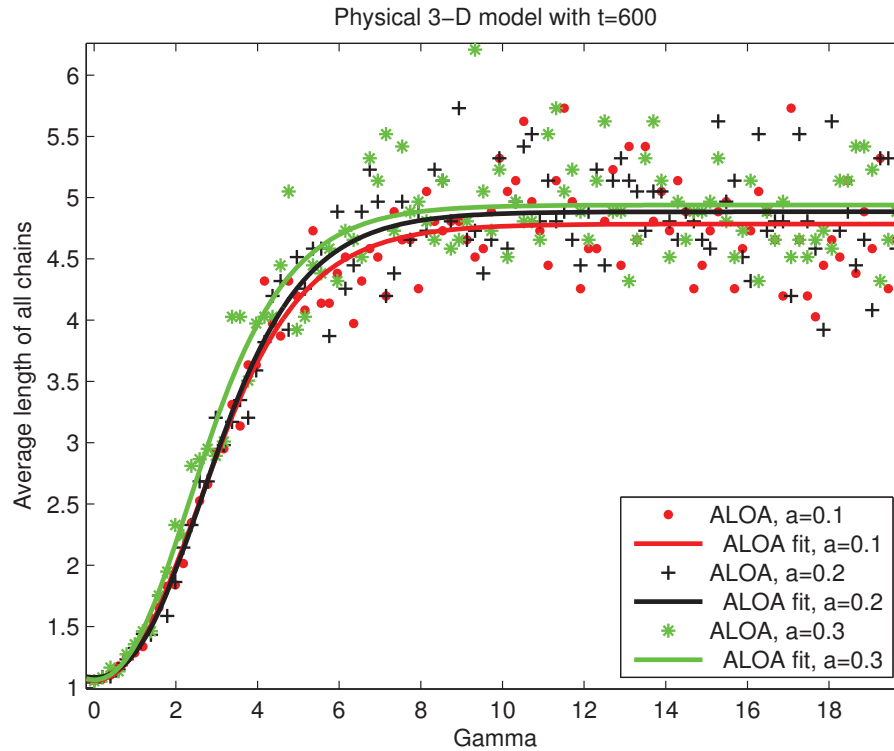
## Results & Conclusions

### 3.1 Results

Since in the 10 by 10 by 10 3D lattice, the occupation rate is set to be 0.05 for each orientation, which implies the total number of dipoles is approximately 300. The same initial 3D lattice were used through all simulations.

#### 3.1.1 Physical 3D model with $t=600$

Fix  $t$  at 600. It limits the total number of successful movements to 600. Averagely, 600 will allow each dipole moves twice.

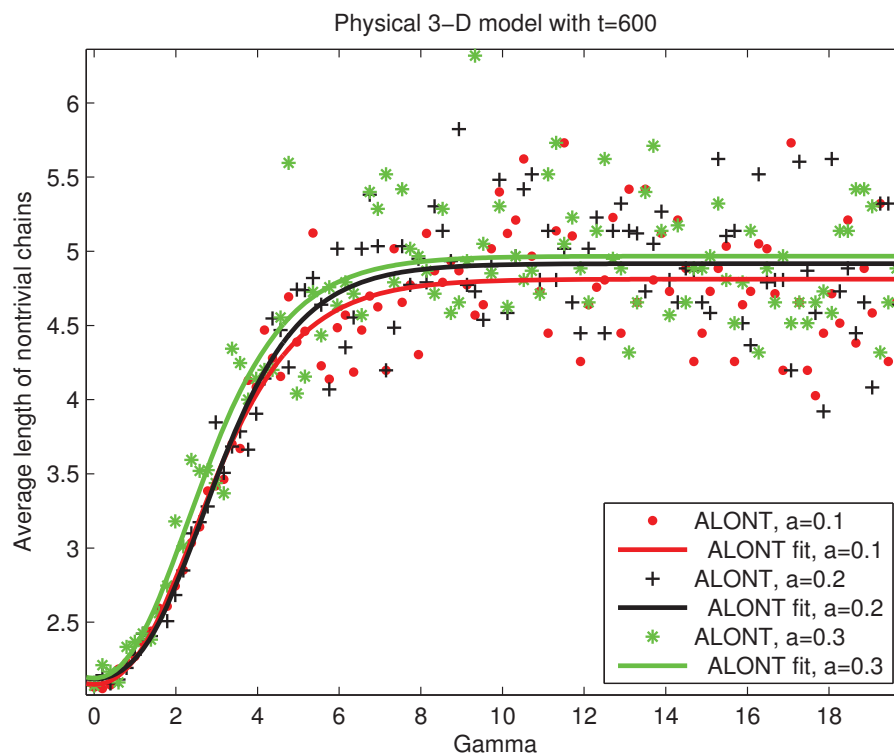


**Figure 3.1:** Average length of all chains in physical 3D model with  $t=600$ . Count both polymers and independent monomers as chains.

Figure 3.1 and Figure 3.2 show average length of chains in two measurements, while Figure 3.3 and Figure 3.4 present number of chains in two measurements.

None of traditional fitting functions, such as polynomial, works well for these data sets due to lack of monotony in the curve trend. Therefore we designed a monotonic kernel fitting function which concentrates on the curves of the scatter plots. The design of this kernel function can be found in Appendix. E. From Figure 3.1 to Figure 3.4, the goodness of fit is decent. The  $R^2$  value for all the fittings are above 0.85.

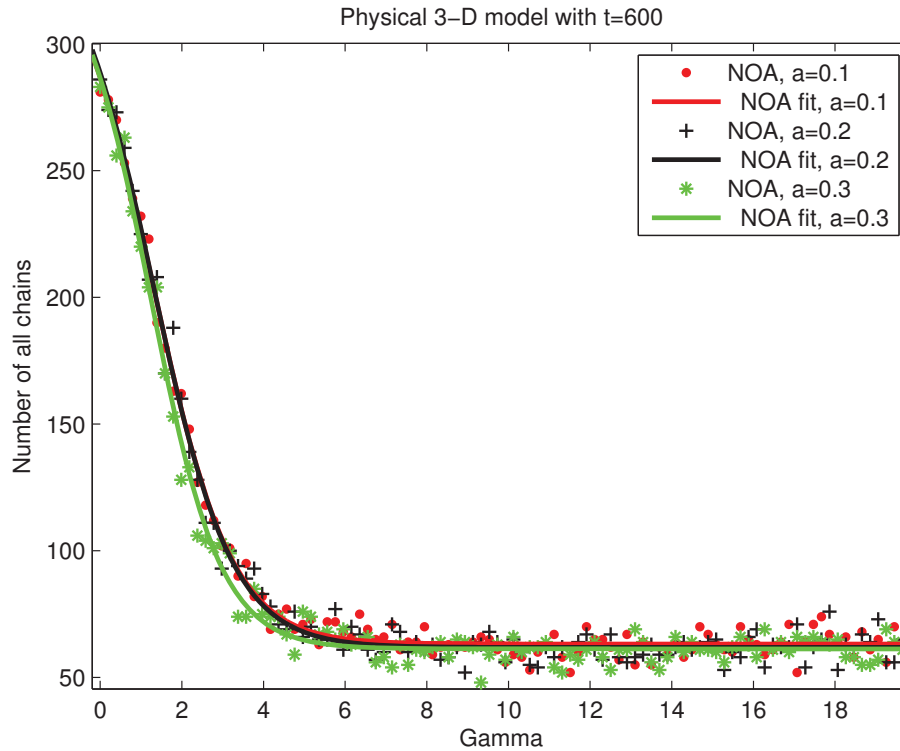
First, in Figure 3.4, the curve went up sharply when  $0 < \gamma < 2$ , decreased a little as  $\gamma$



**Figure 3.2:** Average length of nontrivial chains in physical 3D model with  $t=600$ . Count only polymers as chains.

approaching 5 and then became flat. Compared with Figure 3.3, it is clear that most monomers quickly form polymers with length two when  $\gamma$  goes from 0 to 2. As  $\gamma$  keeps increasing, longer chains are developed (polymers with length two become parts of longer chains) and hence the number of nontrivial chains will decrease.

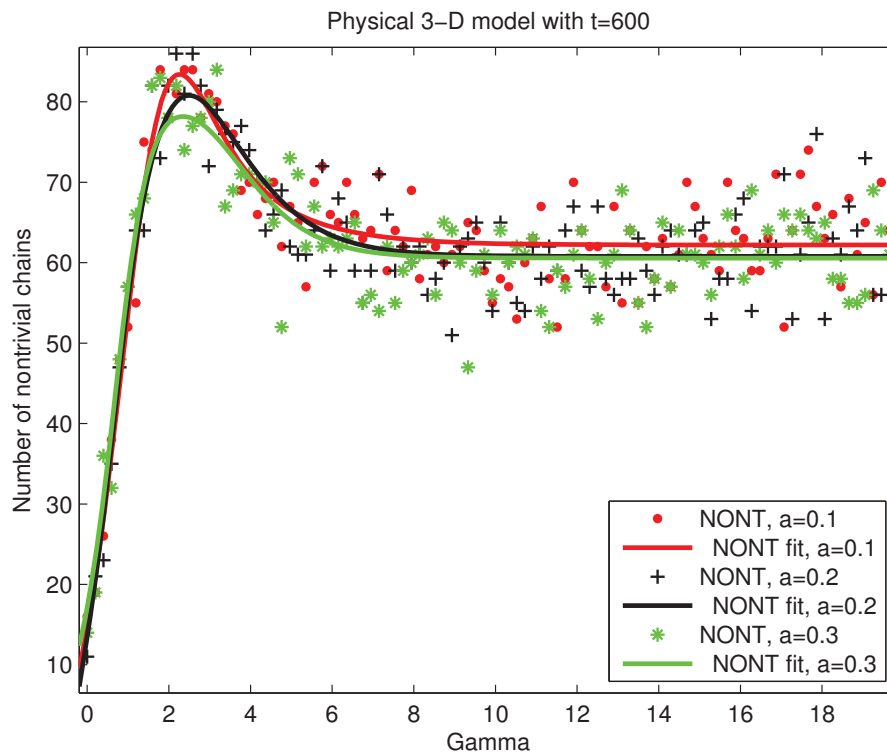
An important finding is that the differences in degree of polymerization among all  $a$  levels are quite slight. Those fitting curves are really close to each other. The finding coincides with the discussion in Appendix D, which proves that the potential energy of dipoles mainly depend on the configuration of dipoles instead of  $a$  levels and hence  $a$  has weak influence



**Figure 3.3:** Number of all chains in physical 3D model with  $t=600$ . Count both polymers and independent monomers as chains.

on the degree of polymerization. With higher value of  $a$ , the average length of chains will be just a little bigger and the number of chains will be a litter smaller.

Another finding is that even we used two measures to describe average length and number of chains, the differences between these two measurements vanished after reaching certain points. For example, when  $\gamma > 6$  the differences are tiny (this finding can be seen clearly in Figure 3.9 and Figure 3.10 in the next section). The differences only exist significantly when  $\gamma$  is small (less than 2). With the degree of polymerization becoming higher, there are much less independent monomers in the 3D lattice, thereby the two measures turn to be identical, refer to Table 2.1.



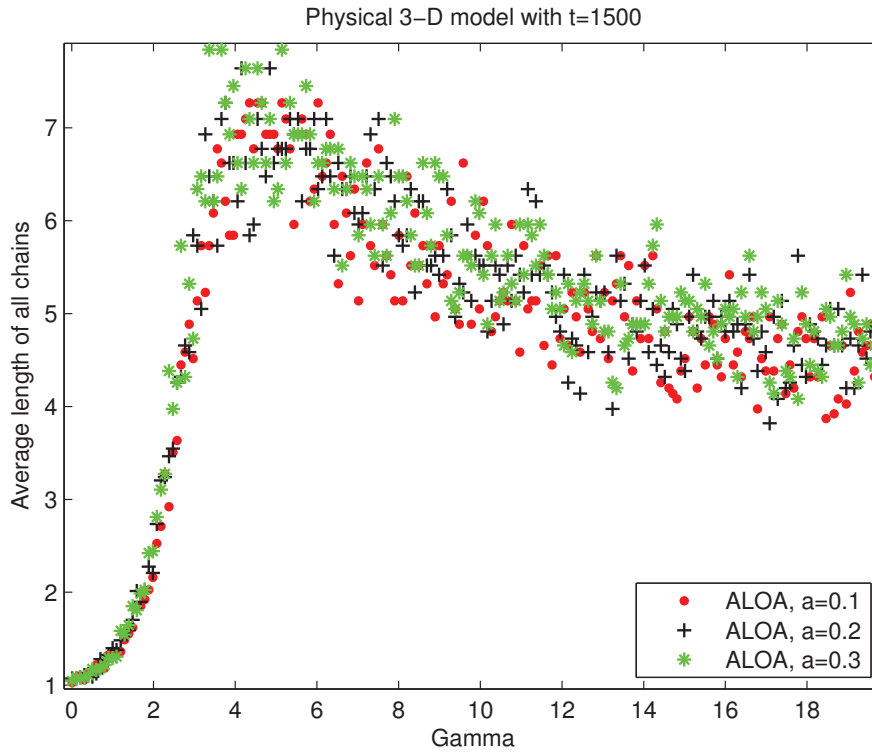
**Figure 3.4:** Number of nontrivial chains in physical 3D model with  $t=600$ . Count only polymers as chains.

Although the kernel fitting function works perfectly here, it requires different combinations of the kernel functions designed for the particular patterns. From Figure 3.1 to Figure 3.4, we had already used three different forms of fitting functions. Besides, the forms of combinations of kernel fitting functions are too complicated to interpret the meaning of the parameters. In the following sections, there are several patterns that need extra effort to design the fitting function. But the main reason we abandon curve fitting in the following sections is that not all the points from the data sets are useful (this will be discussed later). Curve fitting based on those unnecessary points will be misleading. Therefore we only use scatter plots in the coming sections to show the trend of data.

### 3.1.2 Physical 3D model with $t=1500$

In previous section, we set maximal successful movement  $t = 600$  for all cases. When  $t = 600$ , Matlab code is still able to return outcomes. However, when  $t$  is increasing, like  $t = 1000$ , higher  $\gamma$  value will lead to an endless calculation. The reason for the dead loop is the definition of successful move, which is the discrete time variable  $t$ . We defined that only if the dipole changed its position or orientation, it will be counted as a successful move. But some systems might reach a stable status, in which dipoles will seldom change its position or orientation any more. In this situation, a computer will run for a quite long time to simulate those unimportant movements. Thereby, a stop rule is necessary. The rule can be defined as: if any randomly chosen dipole remains its position and orientation after a hypothetical move, this will add 1 to "stable status indicator"; when this indicator accumulates consecutively to 300, then stable status reaches. Consequently, the time variable, which recorded the number of successful moves, provides important information about the process of polymerization. Define the number of successful moves as actual time,  $t_a$ .  $t$  is the upper bound of  $t_a$ .

Figure 3.5 to Figure 3.8 show the two quantities, average length and number of chains, measured under different definitions of chains. Quantities using the definition of "all chains" are plotted by solid markers, as quantities using the definition of "nontrivial chains" are plotted by hollow markers. Figure 3.9 and Figure 3.10 make comparisons between



**Figure 3.5:** Average length of all chains in physical 3D model with  $t=1500$ . Count both polymers and independent monomers as chains.

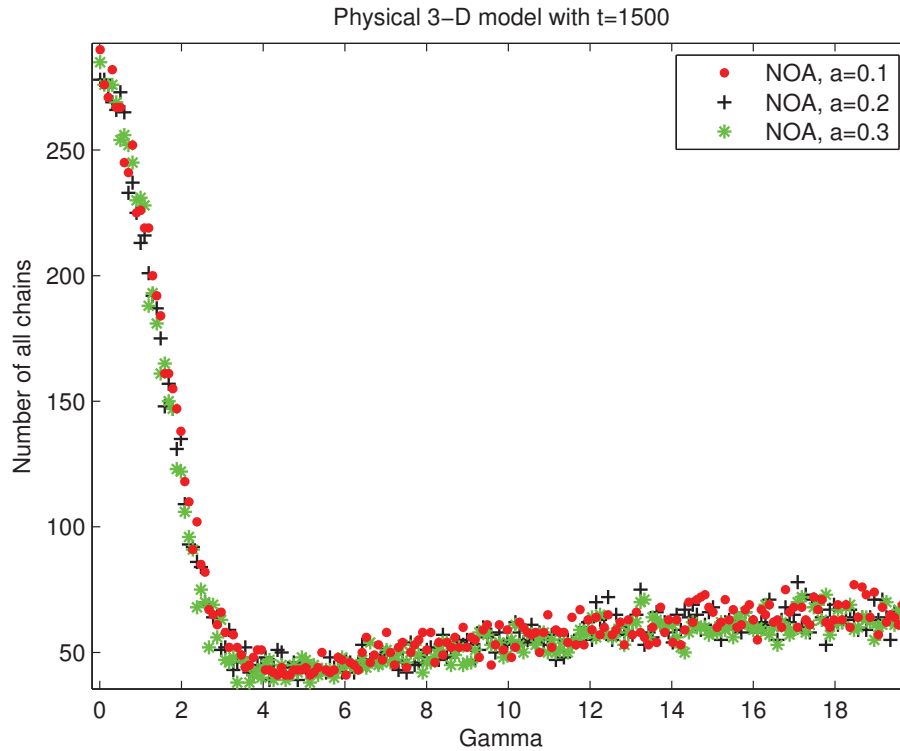
those two measure methods for average length and number of chains respectively. It can be observed the curves under the two measures begun to merge around  $\gamma = 2$  and overlap after  $\gamma = 4$ , when the scatter plot show that those hollow markers are filled by those solid markers.

Compared with plots in Section 3.1.1, there are two main differences in this section:

† Curve patterns.

† Vertical values of critical points.

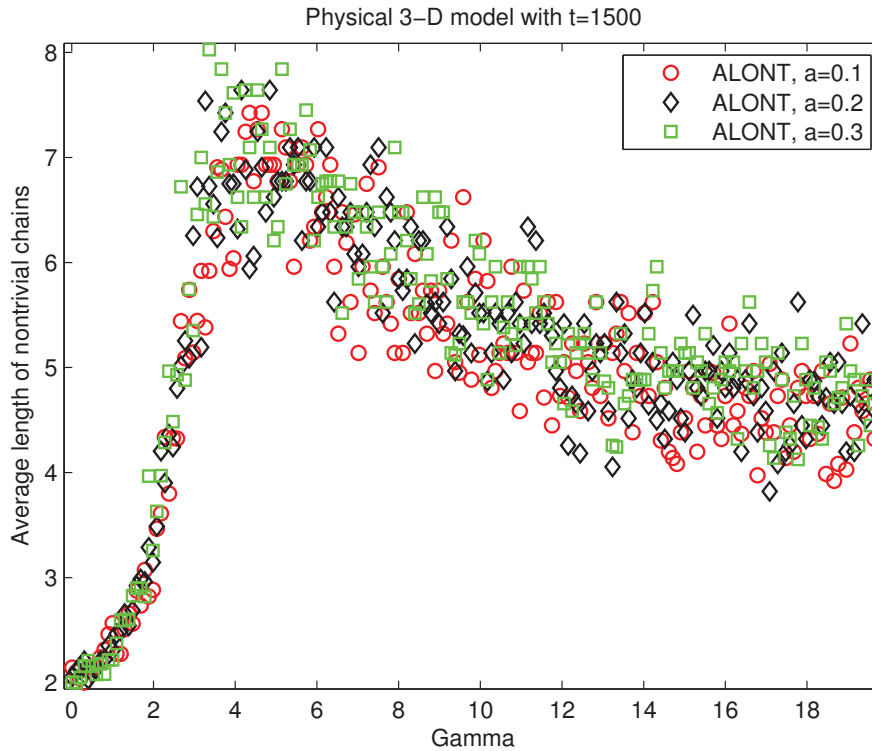




**Figure 3.6:** Number of all chains in physical 3D model with  $t=1500$ . Count both polymers and independent monomers as chains.

In Section 3.1.1, for average length of chains, the curves first went up and then became flat in both measures (Figure 3.1, Figure 3.2), while the curves of number of chains in both measures became flat when  $\gamma > 4$  (Figure 3.3, Figure 3.4). In current section, the curves of average length of chains first rapidly reached their peaks and then decreased gradually (Figure 3.9). While the curves of number of chains went down to their bottoms around  $\gamma = 4$  and then increased gradually (Figure 3.10).

Another obvious difference is about the vertical critical value. In Figure 3.1 and Figure 3.2, the values of maximal average length are around 5. In Figure 3.9, the value of maximal average length is above 7. Since the total number of dipoles is fixed, increase in average

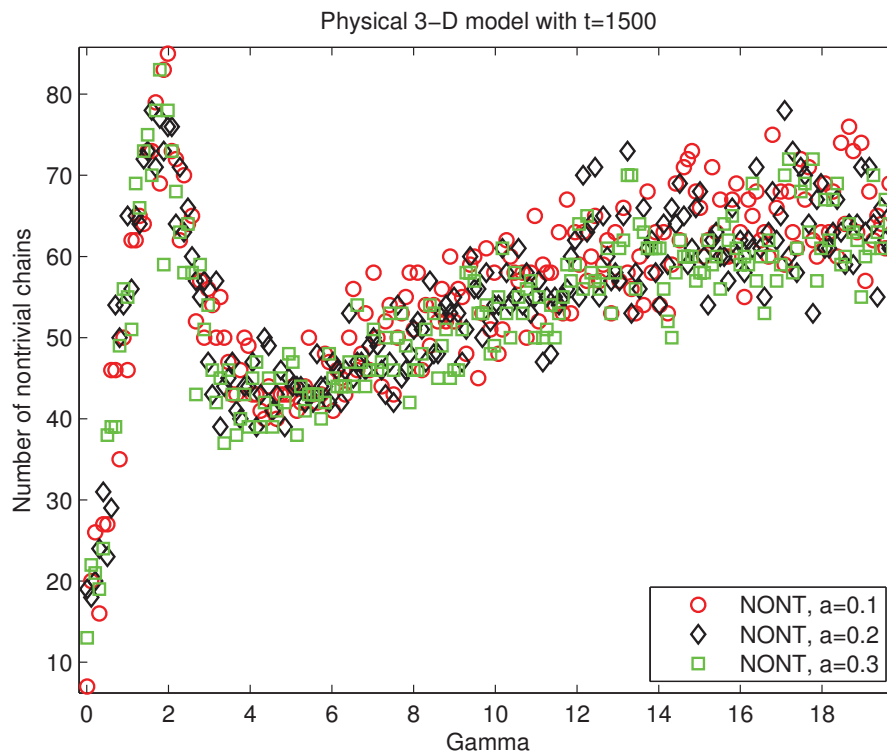


**Figure 3.7:** Average length of nontrivial chains in physical 3D model with  $t=1500$ . Count only polymers as chains.

length means decrease in number of chains.

The reason for these disparities is the time variable, which is the successful movements of dipoles. In Section 3.1.1,  $t$  is fixed at 600, in every system dipoles made totally 600 successful movements. On the contrast, not all system in this section fulfilled 1500 movements, Figure 3.11 illustrates this. When  $\gamma < 4$ , the system made 1500 successful movements, which means the system was still unstable. After  $\gamma = 4$ ,  $t$  started to drop dramatically, which indicates the system became stable.

After comparing Figure 3.11 with all others from Figure 3.1 to Figure 3.4, we found that the

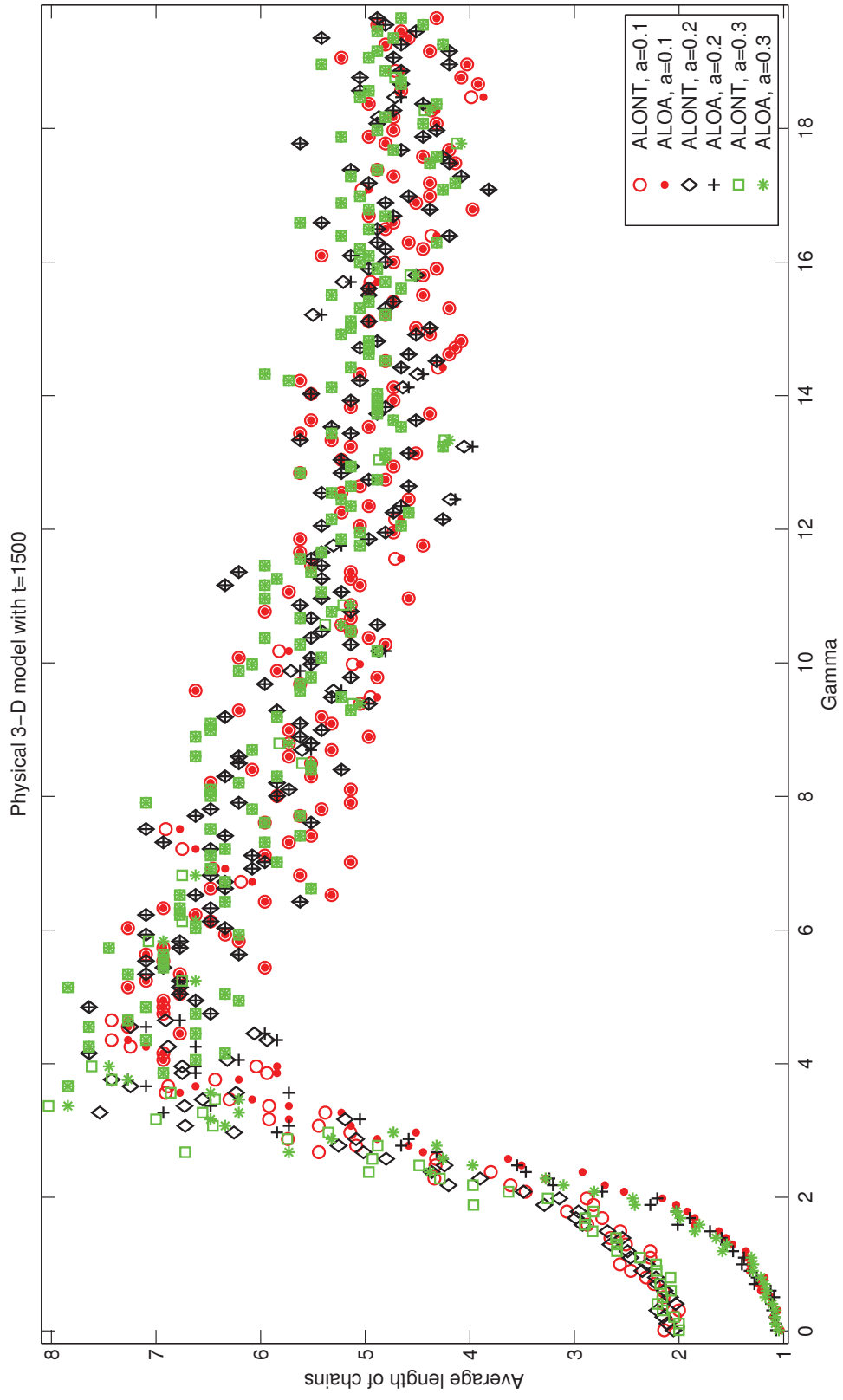


**Figure 3.8:** Number of nontrivial chains in physical 3D model with t=1500. Count only polymers as chains.

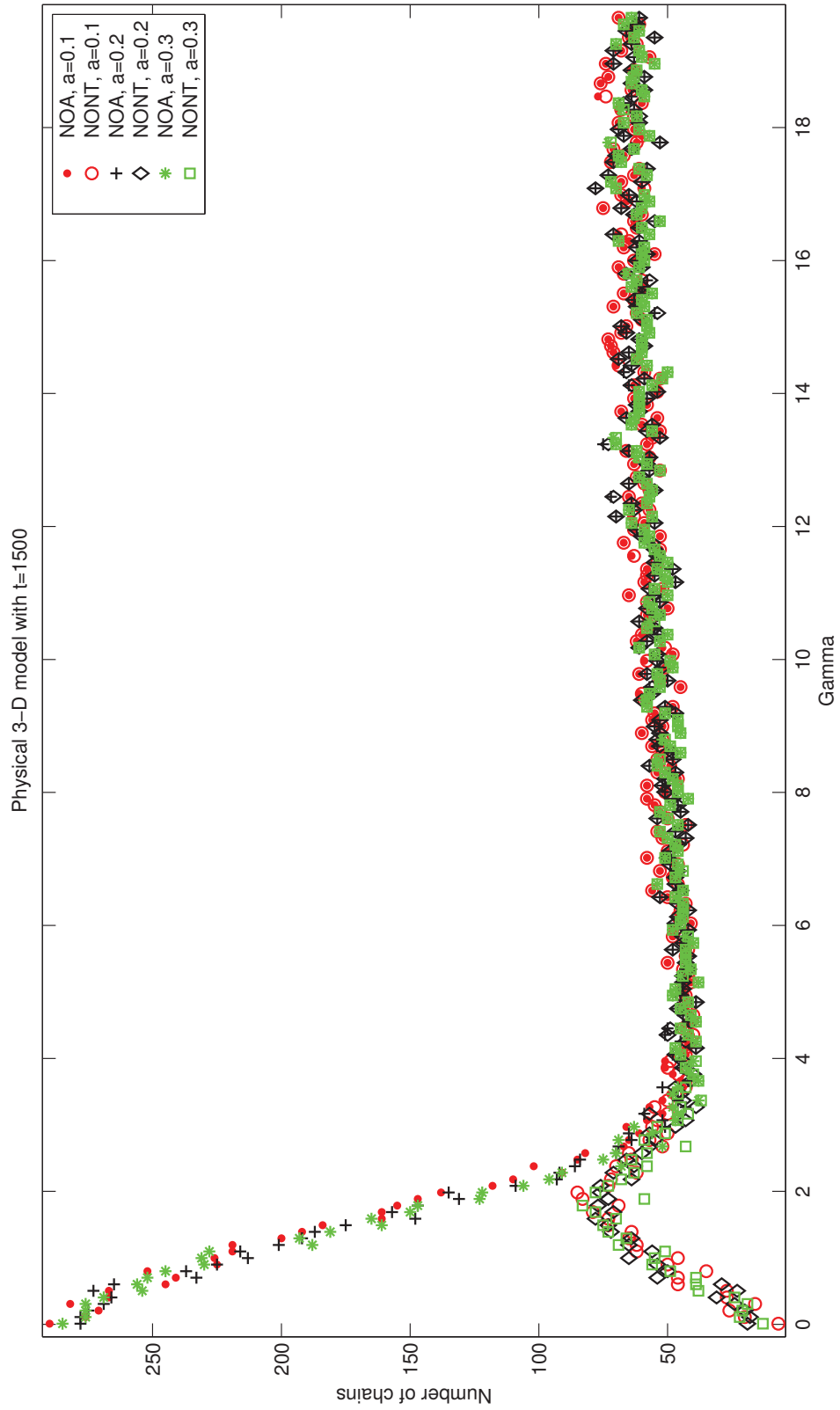
inflections (change in curve trend) happened when the system was getting stable (around  $\gamma = 4$ ).

Before the system becomes stable, there still is potential for monomers to form longer polymer chains. As a result, we should exclude unstable systems in our comparisons.

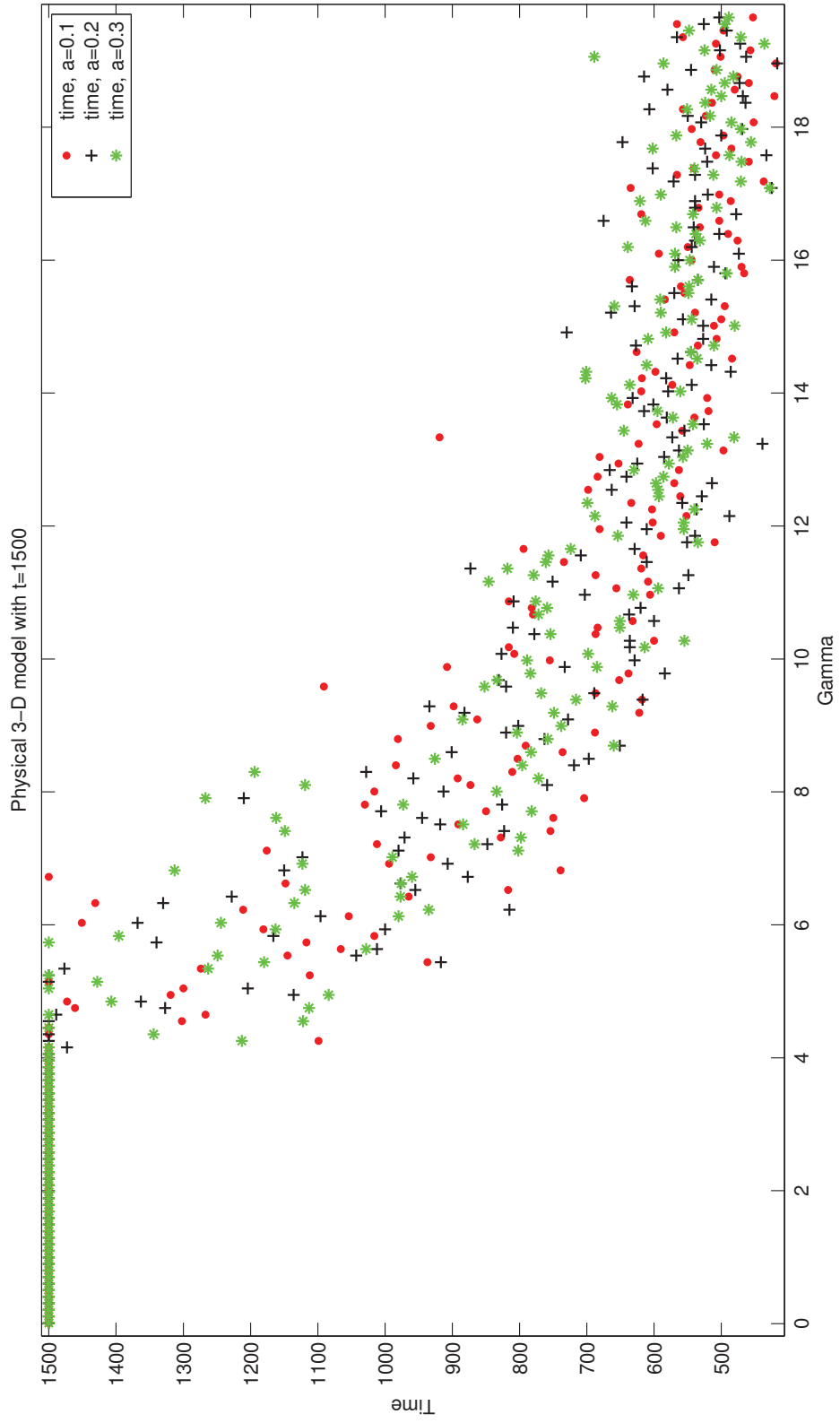
However, the unfinished process of polymerization did provide some valuable information. It explicitly illustrated that if given enough time, systems with lower  $\gamma$  value have potential to form longer chains. This will be discussed in Section 3.2.



**Figure 3.9:** Comparison between two measurements on average length of chains in physical 3D model with  $t=1500$ .



**Figure 3.10:** Comparison between two measurements on number of chains in physical 3D model with  $t=1500$

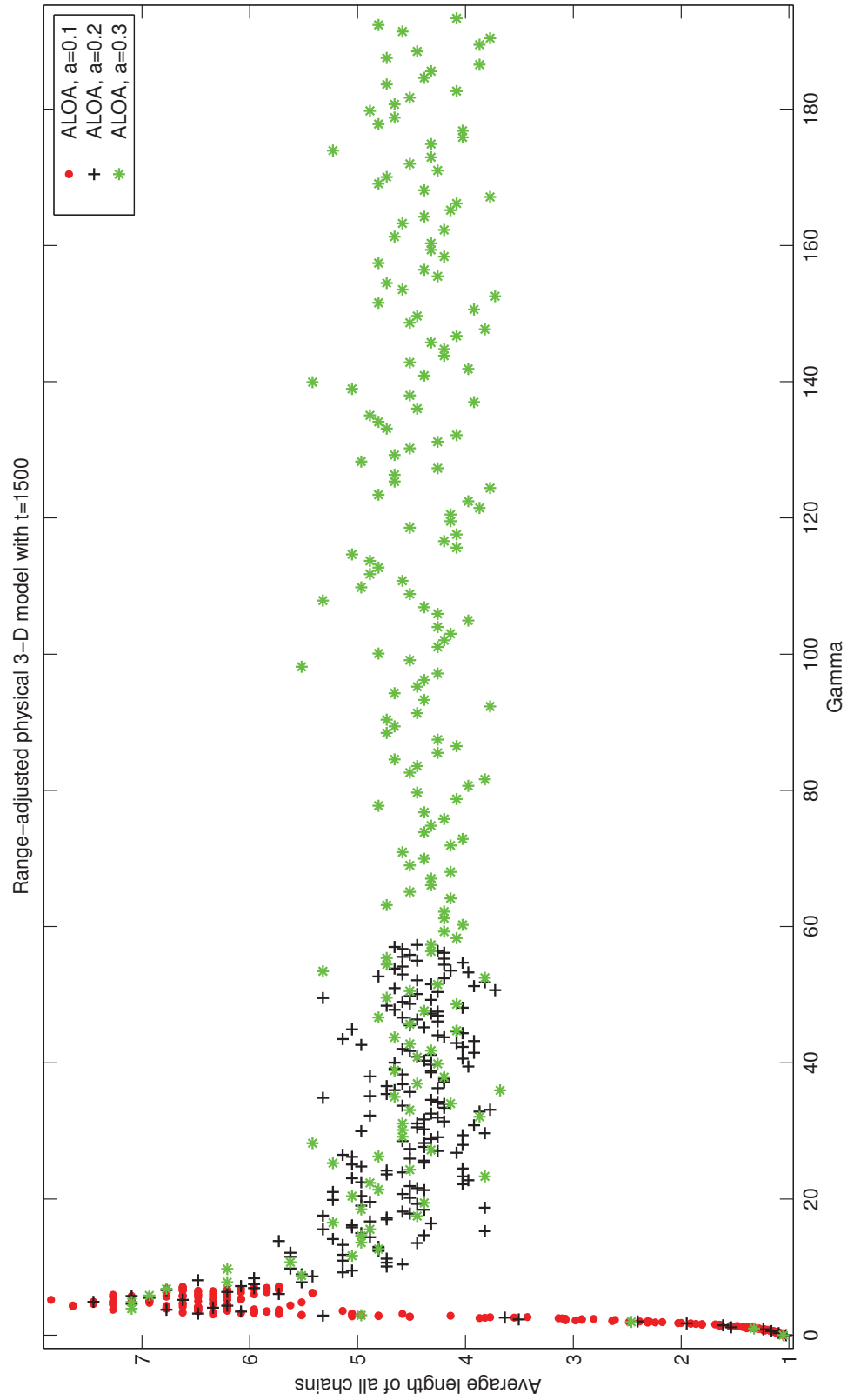


**Figure 3.11:** Actual time of polymerization in physical 3D model with  $t=1500$ .

### 3.1.3 Range-adjusted physical 3D model with $t=1500$

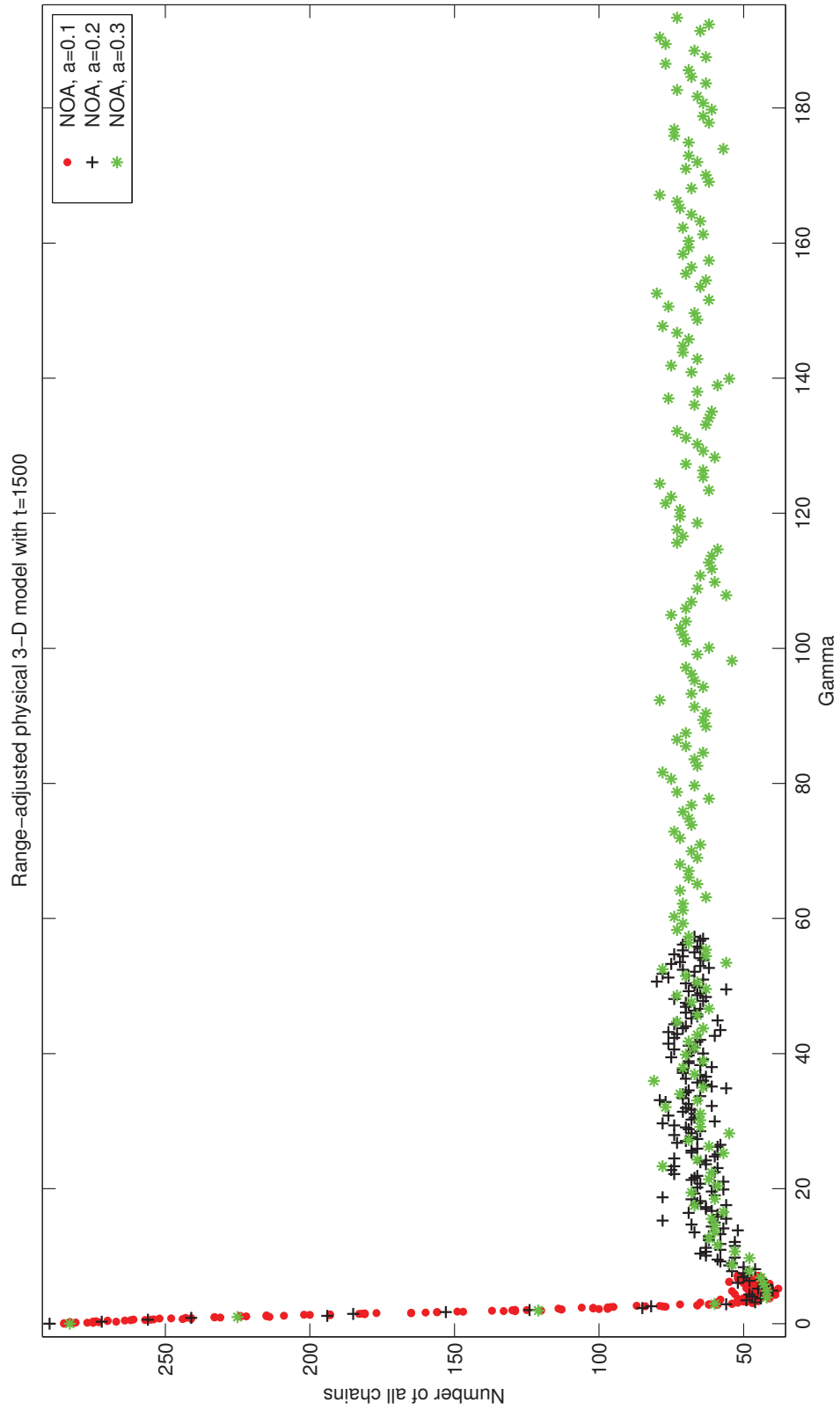
Due to those restrictions in Section 2.4,  $\gamma$  has different intervals for each level of  $a$ , see Table 2.2. The interval of  $\gamma$  when  $a = 0.3$  is much wider than  $a = 0.1$ . The scatter plot patterns of each  $a$  level show that their vertical values are close to each other and somehow overlapped at same  $\gamma$ . Horizontally, the patterns of  $a = 0.2$  is an extension of patterns of  $a = 0.1$ ; the patterns of  $a = 0.3$  is an extension of patterns of  $a = 0.2$ . See Figure 3.12 to Figure 3.17.

An interesting finding is that when  $\gamma > 20$ , all the curves become flat. When  $0 < \gamma < 20$ , all the scatter plot patterns are similar to previous ones in Section 3.1.2, including average length of chains, number of chains, and time.

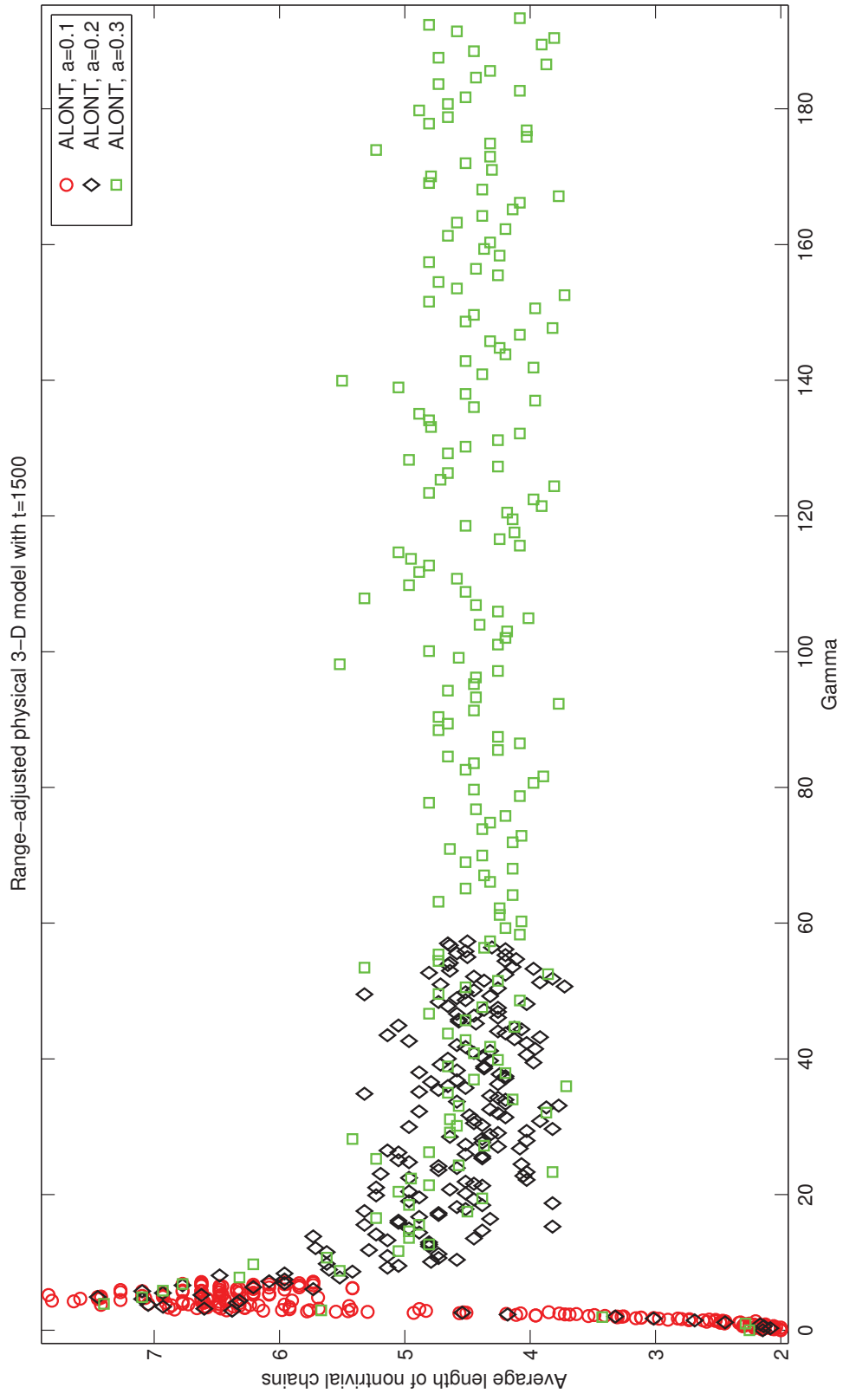


**Figure 3.12:** Average length of all chains in range-adjusted physical 3D model with  $t=1500$ . Count both polymers and independent monomers as chains.

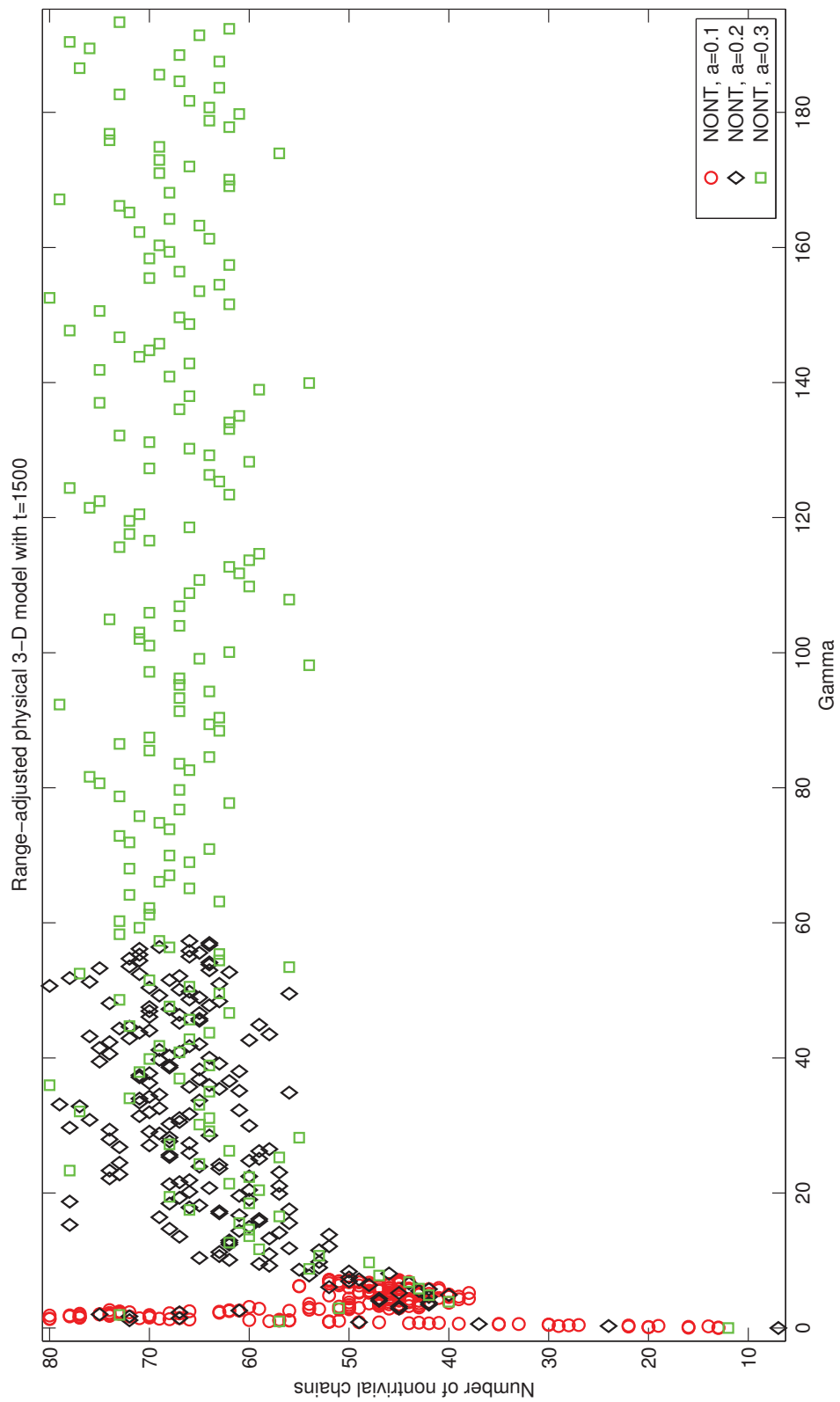




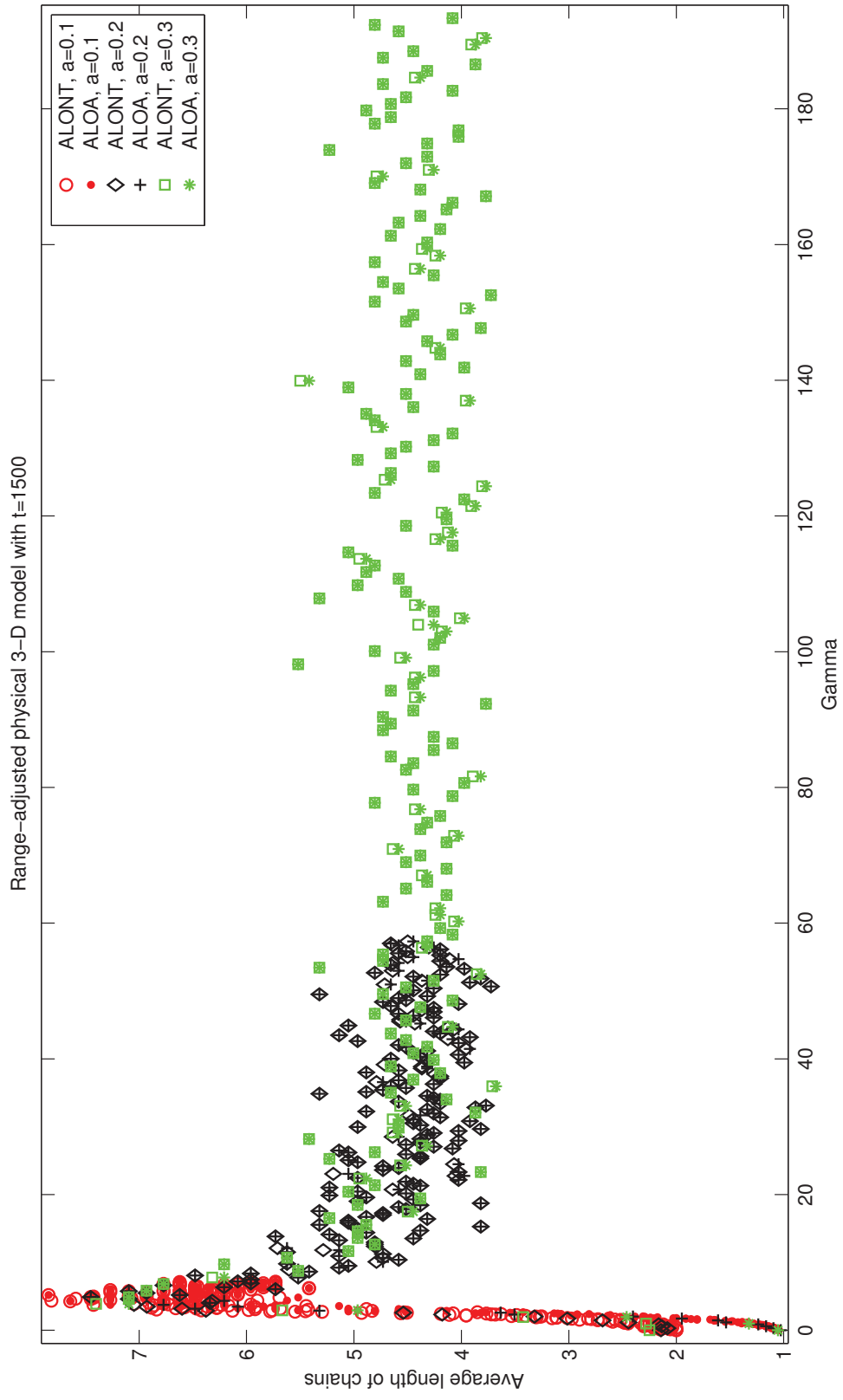
**Figure 3.13:** Number of all chains in range-adjusted physical 3D model with  $t=1500$ . Count both polymers and independent monomers as chains.



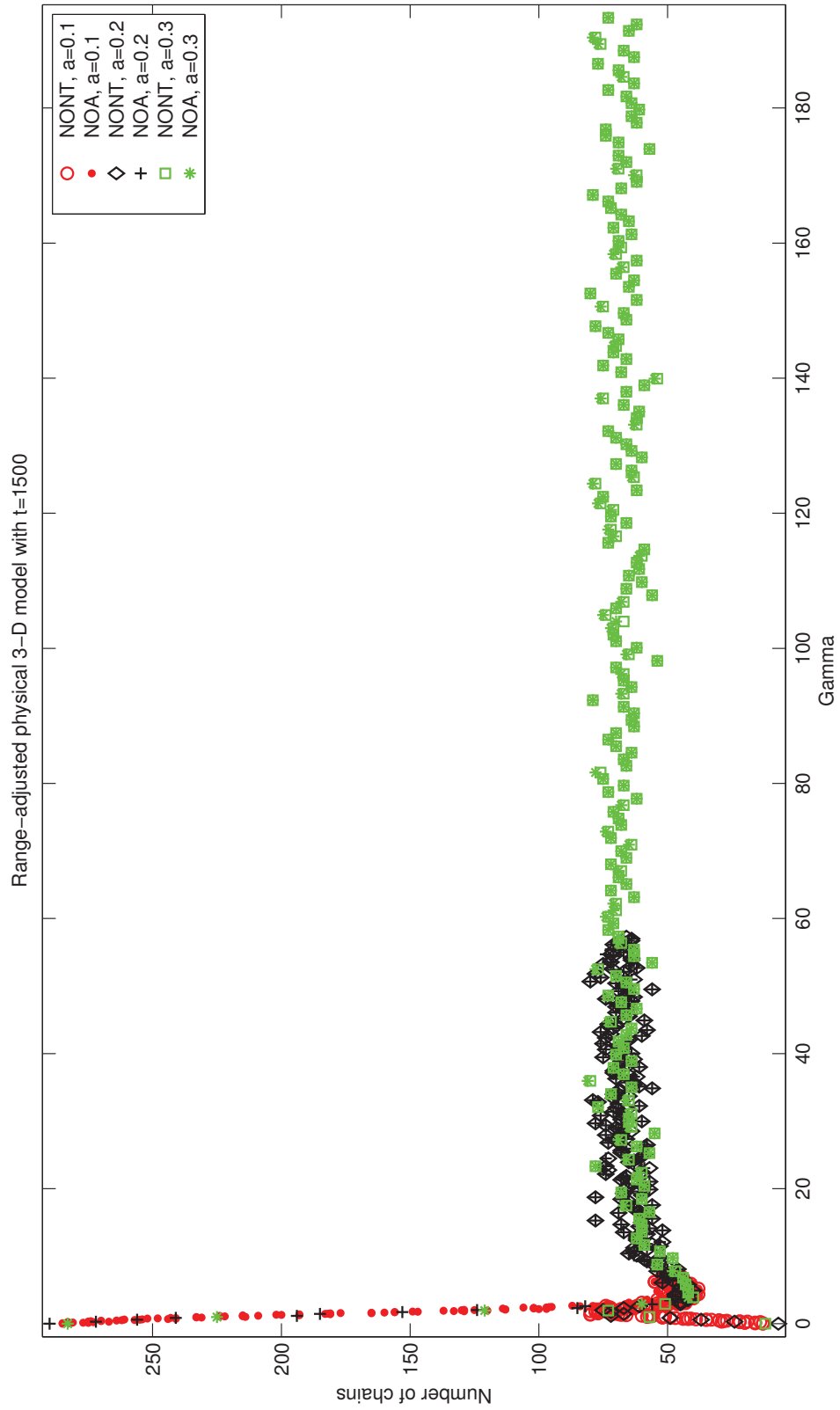
**Figure 3.14:** Average length of nontrivial chains in range-adjusted physical 3D model with  $t=1500$ . Count only polymers as chains.



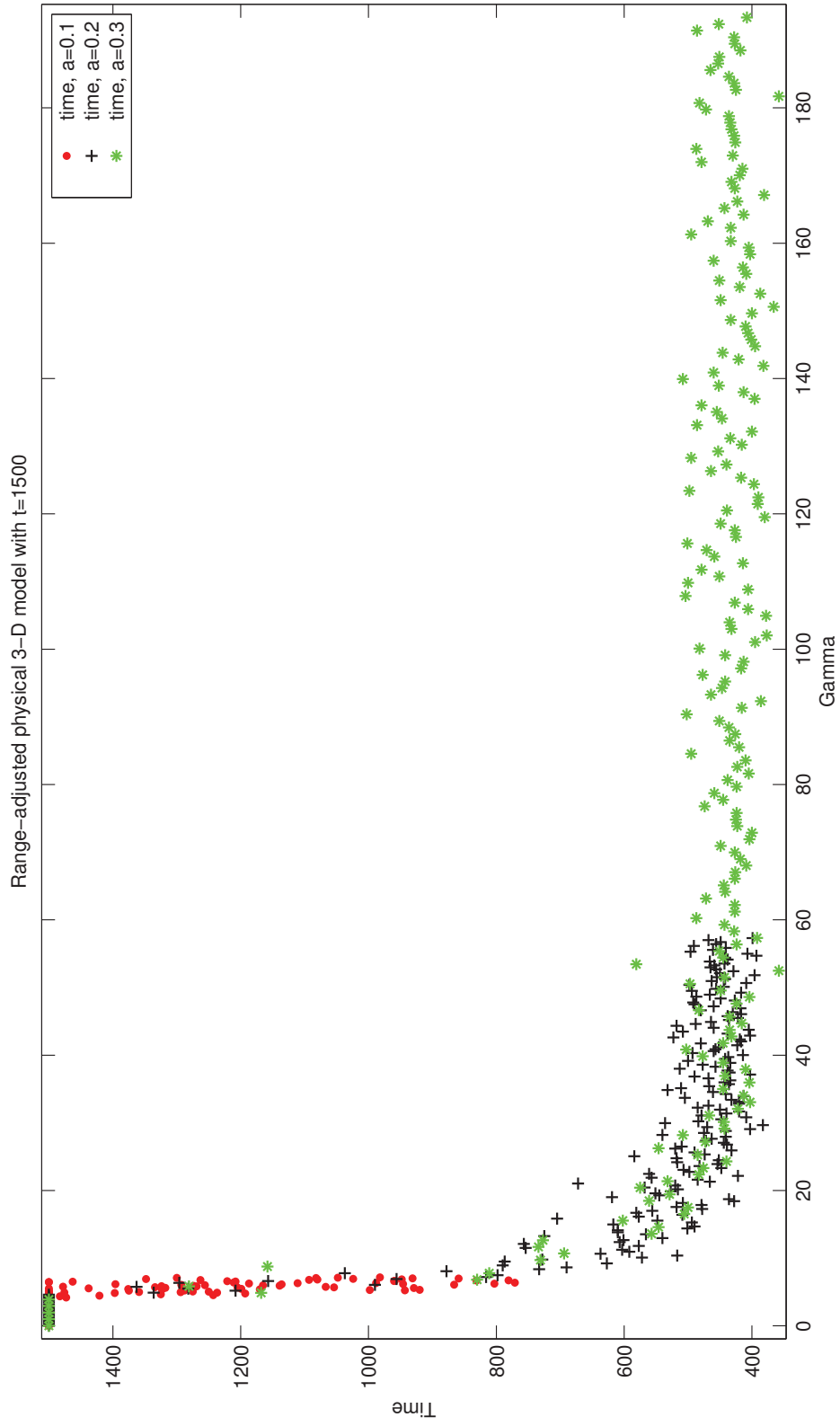
**Figure 3.15:** Number of nontrivial chains in range-adjusted physical 3D model with  $t=1500$ . Count only polymers as chains.



**Figure 3.16:** Comparison between two measurements on average length of chains in range-adjusted physical 3D model with  $t=1500$ .



**Figure 3.17:** Comparison between two measurements on number of chains in range-adjusted physical 3D model with  $t=1500$ .



**Figure 3.18:** Actual time of polymerization in range-adjusted physical 3D model with  $t=1500$ .

## 3.2 Conclusions

### 3.2.1 Dominant Parameters

Based discussions in Section 3.1, we can divide  $\gamma$  values into several small intervals when  $t = 1500$ :

† (0,2]: System is unstable.  $t_a$  stays at 1500. Lot of monomers rapidly form polymers with length two. Average length of chains in two measurements are both growing, number of nontrivial chains is growing as well while number of all chains is falling.

† (2,4]: System is still unstable.  $t_a$  stays at 1500. Plots under two measurements start to merge and overlap. Polymers with length two become part of longer chains. Average length of chains in two measurements keep increasing, number of chains in two measurements are decreasing.

† (4,20]: System turns to be stable.  $t_a$  drops dramatically. Plots under two measurements are identical. Average length of chains decreases gradually while number of chains rise slowly.

† 20+: System remains stable.  $t_a$  stays around 450. Two measurements are identical. Both average length (around 4.5) and number of chains (around 65) stay unchanged.

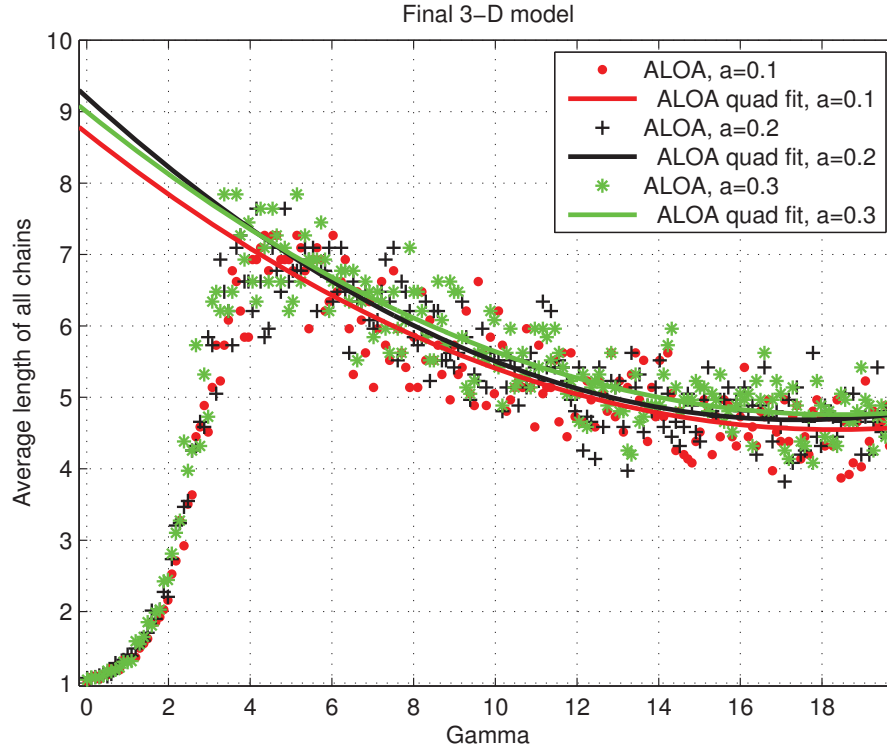
Since we are only interested in stable systems, and when  $\gamma > 20$  no obvious changes happen, then our main focus should be  $5 \leq \gamma < 20$  while  $t = 1500$  (since when  $4 \leq \gamma < 5$  there still are some unstable systems, we exclude this interval). The physical 3D model can be simply used for this purpose if we exclude data when  $\gamma < 5$ .

Because two measurements are identical in this interval, it does not matter which one we use. The total number of dipoles is fixed, so if we know average length of chains then the number of chains can be easily calculated. The scatter plot pattern implies that curve fitting will perform well. We used two approaches to fit the curves, quadratic polynomial fitting (Figure 3.19) and kernel function fitting (Figure 3.20).

From Figure 3.19 and Figure 3.20, the difference among all three  $a$  levels are slight when  $5 \leq \gamma < 20$ . Thereby, in our final model, the only dominant parameter will be  $\gamma$ . Both quadratic polynomial fitting and kernel function fitting give  $R^2$  value around 0.75, which shows decent goodness of fit. Although these two fitting curves are almost same when  $5 \leq \gamma < 20$ , see Figure 3.21, there are two crucial differences between them:

†  $\gamma < 5$ : The quadratic polynomial fitting curves are close to each other among all three  $a$  levels. The fitting functions have similar coefficients. Therefore we can combine three fitting curves to one using mean of corresponding coefficients. This can not be done under kernel fitting functions due to disparity in curve patterns when  $\gamma < 5$ . And the disparity leads to big differences in the corresponding coefficients of the fitting





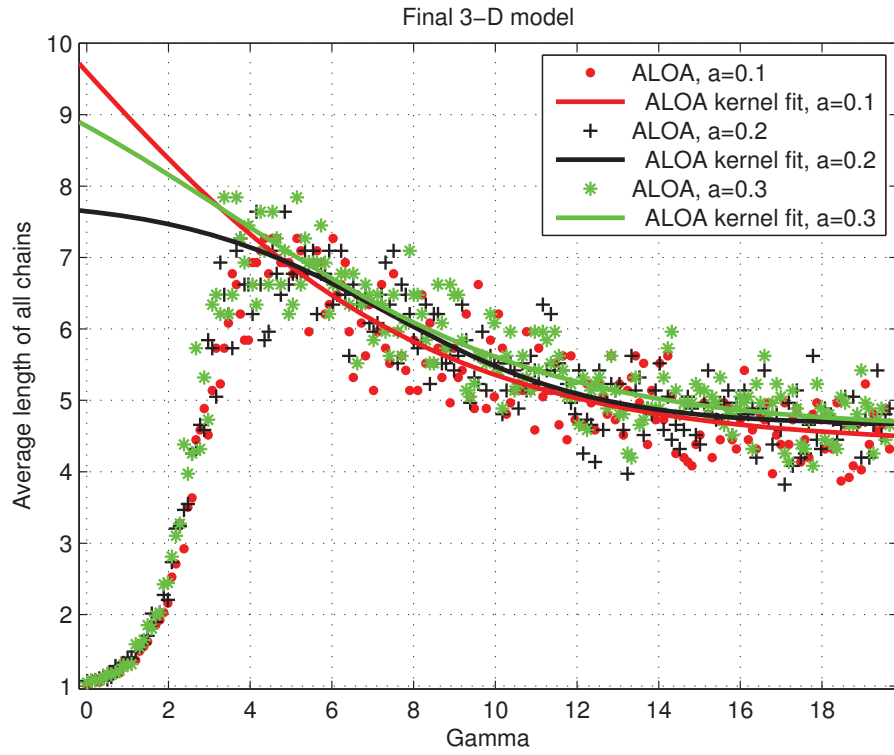
**Figure 3.19:** Quadratic polynomial fitting using data in  $5 \leq \gamma < 20$  in final 3D model.

functions.

†  $\gamma > 20$ : The quadratic polynomial fitting curves will go up dramatically, which does not coincide with our findings in Figure 3.16. On the contrast, the kernel fitting functions will go flat in this interval.

Using mean of estimated coefficients among three  $a$  levels in the quadratic polynomial fitting functions, we get the relation between  $\gamma$  and average length of all chains:

$$ALO A = 0.01083\gamma^2 - 0.4147\gamma + 8.706 \tag{3.1}$$



**Figure 3.20:** Kernel function fitting using data in  $5 \leq \gamma < 20$  in final 3D model.

Recall

$$\gamma = 557 \varepsilon^2 \frac{300K}{T} \tag{3.2}$$

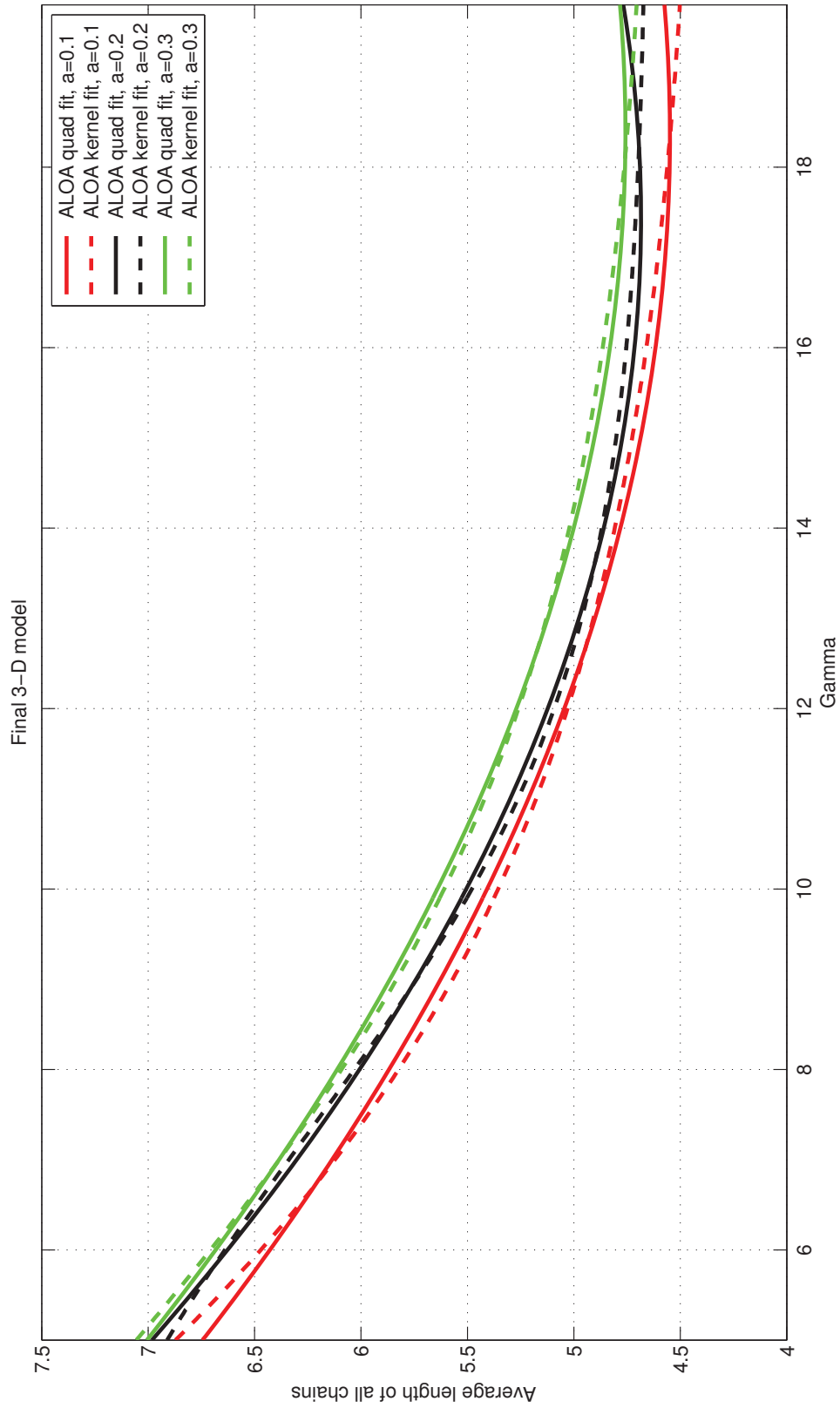
$$= 557 \frac{M_d^2}{l^3} \frac{300K}{T}. \tag{3.3}$$

Substituted  $\gamma$  in 3.1 by 3.3, we get

$$ALOA = 0.01083\left(557\frac{M_d^2}{l^3}\frac{300K}{T}\right)^2 - 0.4147\left(557\frac{M_d^2}{l^3}\frac{300K}{T}\right) + 8.706. \quad (3.4)$$

It seems that quadratic fitting is a good choice to predict the average length of all chains in stable systems when  $\gamma < 5$ . But based on reality, kernel fitting function is more appropriate because it is monotonic. If we could have eliminated the disparity in curves of kernel fitting functions when  $\gamma < 5$ , kernel fitting function would be undoubtedly the best choice.

Through all the above discussions, we draw the conclusion that the degree of polymerization can be determined by only three parameters, dipole moment, distance between molecules, and circumstance temperature. The parameter  $a$  is not important when  $a$  is small.



**Figure 3.21:** Comparison between quadratic polynomial and kernel fitting functions when  $5 \leq \gamma < 20$  in final 3D model.

### 3.2.2 The Role Of Time Limit

Time limit has no effect on the degree of polymerization of systems that reach stable before arriving at the time limit. However, time limit greatly determines the performance of polymerization in systems which will not be stable until the time limit. For instance, this difference can be found in Figure 3.1 ( $t = 600$ ) and Figure 3.5 ( $t = 1500$ ). From Figure 3.19, the quadratic fitting curve when  $\gamma < 5$  seems to be a plausible way to predict average length of chains in stable systems. However, this should be verified by more simulated data sets.

### 3.3 Future Plans

With recent development in monitoring molecular charge distribution accurately [14], our model provides a possible approach to predict the performance of polymerization. However, to further improve the availability of our model, the following plans can be applied:

- † The most possible and reasonable plan is to apply larger limit value of successful movements  $t$ . As mentioned in Section 3.2.2, using much larger limit value of  $t$  will extract more information about stable system with small  $\gamma$ . It will enable us to test

the fitting curves in Figure 3.19 and Figure 3.20 by using new data, and then improve our kernel fitting functions if possible.

- † The second plan is to increase the block size. In our 3D models, we used 10 by 10 by 10 cube as a block. In reality, the polymer chain can consist of hundreds monomers.
- † The last plan could be a better generalized model. In all of our models, orientations of dipoles are discrete. It is possible to model it using continuous orientation variable in 3D.

The above plans are all about improving our models. Another aspect is to verify our models. Since there are no data about the origin of life, it is impossible to verify our models through that way. However, there are a lot of chemical experiments about polymerization have been done. Is is plausible to use polymerization data to verify out models even though the molecules in that processes are larger than simple and inorganic molecules.

# References

- [1] I. Pinelis, "Spontaneous Two-Dimensional Polymerization," unpublished.
- [2] DynaBlast. (2006, Jan 28), *Covalently bonded hydrogen and carbon* [Online].  
Available: <http://en.wikipedia.org/wiki/File:Covalent.svg>.
- [3] EliseEtc. (2012, Feb 5), *Diagram of electron transfer between Li and F* [Online].  
Available: [http://en.wikipedia.org/wiki/File:Ionic\\_bonding.svg](http://en.wikipedia.org/wiki/File:Ionic_bonding.svg).
- [4] J. D. Bernal, "The Physical Basis of Life," *Proc. of the Physical Society*, vol. 62, no. 9, p. 537, 1949.
- [5] S. L. Miller, "Production of Amino Acids Under Possible Primitive Earth Conditions," *Science*, vol. 117, no. 3046, p. 528, 1953.
- [6] S. L. Miller, "Organic Compound Synthesis on the Primitive Earth," *Science*, vol. 130, no. 3370, p. 245, 1959.
- [7] T. Gowers. (2009, Nov), *Polymath and the origin of life* [Online]. Available:  
<http://gowers.wordpress.com/2009/11/07/polymath-and-the-origin-of-life>.

- [8] S. Wolfram, “Statistical Mechanics of Cellular Automata,” *Reviews of Modern Physics*, vol. 55, no. 3, pp. 601–644, 1983.
- [9] P. Bak *et al.*, “Self-organized criticality: An explanation of the 1/f noise,” *Physical Review Letters*, vol. 59, no. 4, pp. 381–384, 1987.
- [10] G. Barrow, *Physical chemistry*. Columbus, OH: McGraw Hill, 2nd ed., 1966.
- [11] R. F. Bader, *An introduction to the electronic structure of atoms and molecules*. Hamilton, ON, Canada: Clarke, Irwin, 1970.
- [12] R. Young, *Introduction to Polymers*. Boca Raton, FL: Chapman and Hall, 2nd ed., 1991.
- [13] G. N. Lewis, “The Atom and the Molecule,” *Journal of the American Chemical Society*, vol. 38, p. 772, 1916.
- [14] F. Mohn *et al.*, “Imaging the charge distribution within a single molecule,” *Nature Nanotechnology*, vol. 7, pp. 227–231, 2012.



# Appendix A

## Matlab Code

```
function n=countF(M)

% Front(orientation of center cell in 3*3*3 matrix), count the
% number to calculate the potential-----
% a1*n1+a2*n2+a3*n3+a4*n4+a5*n5+a6*n6+a7*n7

n1=(M(2,2,1)==-3)+(M(2,2,3)==-3)-(M(2,2,1)==3)-(M(2,2,3)==3);

n2=(M(2,1,2)==3)+(M(2,3,2)==3)+(M(1,2,2)==3)+(M(3,2,2)==3)...
    -(M(2,1,2)==-3)-(M(2,3,2)==-3)-(M(1,2,2)==-3)-(M(3,2,2)==-3);

n3=(M(1,2,1)==3)+(M(3,2,1)==3)+(M(1,2,3)==3)+(M(3,2,3)==3)+...
    (M(2,1,1)==3)+(M(2,3,1)==3)+(M(2,1,3)==3)+(M(2,3,3)==3)-...
    (M(1,2,1)==-3)-(M(3,2,1)==-3)-(M(1,2,3)==-3)-(M(3,2,3)==-3)-...
```

$$(M(2, 1, 1) == -3) - (M(2, 3, 1) == -3) - (M(2, 1, 3) == -3) - (M(2, 3, 3) == -3) ;$$

$$\begin{aligned} n4 = & (M(1, 2, 1) == -2) + (M(3, 2, 1) == 2) + (M(1, 2, 3) == 2) + (M(3, 2, 3) == -2) + \dots \\ & (M(2, 1, 1) == 1) + (M(2, 3, 1) == -1) + (M(2, 1, 3) == -1) + (M(2, 3, 3) == 1) - \dots \\ & (M(1, 2, 1) == 2) - (M(3, 2, 1) == -2) - (M(1, 2, 3) == -2) - (M(3, 2, 3) == 2) - \dots \\ & (M(2, 1, 1) == -1) - (M(2, 3, 1) == 1) - (M(2, 1, 3) == 1) - (M(2, 3, 3) == -1) ; \end{aligned}$$

$$\begin{aligned} n5 = & (M(1, 1, 2) == 3) + (M(1, 3, 2) == 3) + (M(3, 1, 2) == 3) + (M(3, 3, 2) == 3) - \dots \\ & (M(1, 1, 2) == -3) - (M(1, 3, 2) == -3) - (M(3, 1, 2) == -3) - (M(3, 3, 2) == -3) ; \end{aligned}$$

$$\begin{aligned} n6 = & (M(1, 1, 1) == 3) + (M(1, 3, 1) == 3) + (M(3, 1, 1) == 3) + (M(3, 3, 1) == 3) + \dots \\ & (M(1, 1, 3) == 3) + (M(1, 3, 3) == 3) + (M(3, 1, 3) == 3) + (M(3, 3, 3) == 3) - \dots \\ & (M(1, 1, 1) == -3) - (M(1, 3, 1) == -3) - (M(3, 1, 1) == -3) - (M(3, 3, 1) == -3) - \dots \\ & (M(1, 1, 3) == -3) - (M(1, 3, 3) == -3) - (M(3, 1, 3) == -3) - (M(3, 3, 3) == -3) ; \end{aligned}$$

$$\begin{aligned} n7 = & (M(1, 1, 1) == -2) + (M(1, 3, 1) == -2) + (M(3, 1, 1) == 2) + (M(3, 3, 1) == 2) + \dots \\ & (M(1, 1, 3) == 2) + (M(1, 3, 3) == 2) + (M(3, 1, 3) == -2) + (M(3, 3, 3) == -2) - \dots \\ & (M(1, 1, 1) == 2) - (M(1, 3, 1) == 2) - (M(3, 1, 1) == -2) - (M(3, 3, 1) == -2) - \dots \\ & (M(1, 1, 3) == -2) - (M(1, 3, 3) == -2) - (M(3, 1, 3) == 2) - (M(3, 3, 3) == 2) + \dots \\ & (M(1, 1, 1) == 1) + (M(1, 3, 1) == -1) + (M(3, 1, 1) == 1) + (M(3, 3, 1) == -1) + \dots \\ & (M(1, 1, 3) == -1) + (M(1, 3, 3) == 1) + (M(3, 1, 3) == -1) + (M(3, 3, 3) == 1) - \dots \\ & (M(1, 1, 1) == -1) - (M(1, 3, 1) == 1) - (M(3, 1, 1) == -1) - (M(3, 3, 1) == 1) - \dots \\ & (M(1, 1, 3) == 1) - (M(1, 3, 3) == -1) - (M(3, 1, 3) == 1) - (M(3, 3, 3) == -1) ; \end{aligned}$$

$$n = [n1, n2, n3, n4, n5, n6, n7] ;$$

function n=countR(M)

```

% Right(orientation of center cell in 3*3*3 matrix), count the
% number to calculate the potential----
% a1*n1+a2*n2+a3*n3+a4*n4+a5*n5+a6*n6+a7*n7

n1=(M(2,1,2)==-1)+(M(2,3,2)==-1)-(M(2,1,2)==1)-(M(2,3,2)==1);

n2=(M(2,2,1)==1)+(M(1,2,2)==1)+(M(2,2,3)==1)+(M(3,2,2)==1)...
    -(M(2,2,1)==-1)-(M(1,2,2)==-1)-(M(2,2,3)==-1)-(M(3,2,2)==-1);

n3=(M(2,1,1)==1)+(M(2,3,1)==1)+(M(2,1,3)==1)+(M(2,3,3)==1)+...
    (M(1,1,2)==1)+(M(1,3,2)==1)+(M(3,1,2)==1)+(M(3,3,2)==1)-...
    (M(2,1,1)==-1)-(M(2,3,1)==-1)-(M(2,1,3)==-1)-(M(2,3,3)==-1)-...
    (M(1,1,2)==-1)-(M(1,3,2)==-1)-(M(3,1,2)==-1)-(M(3,3,2)==-1);

n4=(M(2,1,1)==3)+(M(2,3,1)==-3)+(M(2,1,3)==-3)+(M(2,3,3)==3)+...
    (M(1,1,2)==2)+(M(1,3,2)==-2)+(M(3,1,2)==-2)+(M(3,3,2)==2)-...
    (M(2,1,1)==-3)-(M(2,3,1)==3)-(M(2,1,3)==3)-(M(2,3,3)==-3)-...
    (M(1,1,2)==-2)-(M(1,3,2)==2)-(M(3,1,2)==2)-(M(3,3,2)==-2);

n5=(M(1,2,1)==1)+(M(3,2,1)==1)+(M(1,2,3)==1)+(M(3,2,3)==1)-...
    (M(1,2,1)==-1)-(M(3,2,1)==-1)-(M(1,2,3)==-1)-(M(3,2,3)==-1);

n6=(M(1,1,1)==1)+(M(1,3,1)==1)+(M(3,1,1)==1)+(M(3,3,1)==1)+...
    (M(1,1,3)==1)+(M(1,3,3)==1)+(M(3,1,3)==1)+(M(3,3,3)==1)-...
    (M(1,1,1)==-1)-(M(1,3,1)==-1)-(M(3,1,1)==-1)-(M(3,3,1)==-1)-...
    (M(1,1,3)==-1)-(M(1,3,3)==-1)-(M(3,1,3)==-1)-(M(3,3,3)==-1);

```

```

n7=(M(1,1,1)==2)+(M(1,3,1)==-2)+(M(3,1,1)==-2)+(M(3,3,1)==2)+...
    (M(1,1,3)==2)+(M(1,3,3)==-2)+(M(3,1,3)==-2)+(M(3,3,3)==2)-...
    (M(1,1,1)==-2)-(M(1,3,1)==2)-(M(3,1,1)==2)-(M(3,3,1)==-2)-...
    (M(1,1,3)==-2)-(M(1,3,3)==2)-(M(3,1,3)==2)-(M(3,3,3)==-2)+...
    (M(1,1,1)==3)+(M(1,3,1)==-3)+(M(3,1,1)==3)+(M(3,3,1)==-3)+...
    (M(1,1,3)==-3)+(M(1,3,3)==3)+(M(3,1,3)==-3)+(M(3,3,3)==3)-...
    (M(1,1,1)==-3)-(M(1,3,1)==3)-(M(3,1,1)==-3)-(M(3,3,1)==3)-...
    (M(1,1,3)==3)-(M(1,3,3)==-3)-(M(3,1,3)==3)-(M(3,3,3)==-3);
n=[n1,n2,n3,n4,n5,n6,n7];

```

```

function n=countU(M)
% Up(orientation of center cell in 3*3*3 matrix), count the
% number to calculate the potential----
% a1*n1+a2*n2+a3*n3+a4*n4+a5*n5+a6*n6+a7*n7

n1=(M(1,2,2)==-2)+(M(3,2,2)==-2)-(M(1,2,2)==2)-(M(3,2,2)==2);

n2=(M(2,2,1)==2)+(M(2,1,2)==2)+(M(2,3,2)==2)+(M(2,2,3)==2)...
    -(M(2,2,1)==-2)-(M(2,1,2)==-2)-(M(2,3,2)==-2)-(M(2,2,3)==-2);

n3=(M(1,2,1)==2)+(M(3,2,1)==2)+(M(1,2,3)==2)+(M(3,2,3)==2)+...
    (M(1,1,2)==2)+(M(1,3,2)==2)+(M(3,1,2)==2)+(M(3,3,2)==2)-...
    (M(1,2,1)==-2)-(M(3,2,1)==-2)-(M(1,2,3)==-2)-(M(3,2,3)==-2)-...
    (M(1,1,2)==-2)-(M(1,3,2)==-2)-(M(3,1,2)==-2)-(M(3,3,2)==-2);

```

```

n4=(M(1,2,1)==-3)+(M(3,2,1)==3)+(M(1,2,3)==3)+(M(3,2,3)==-3)+...
    (M(1,1,2)==1)+(M(1,3,2)==-1)+(M(3,1,2)==-1)+(M(3,3,2)==1)-...
    (M(1,2,1)==3)-(M(3,2,1)==-3)-(M(1,2,3)==-3)-(M(3,2,3)==3)-...
    (M(1,1,2)==-1)-(M(1,3,2)==1)-(M(3,1,2)==1)-(M(3,3,2)==-1);

n5=(M(2,1,1)==2)+(M(2,3,1)==2)+(M(2,1,3)==2)+(M(2,3,3)==2)-...
    (M(2,1,1)==-2)-(M(2,3,1)==-2)-(M(2,1,3)==-2)-(M(2,3,3)==-2);

n6=(M(1,1,1)==2)+(M(1,3,1)==2)+(M(3,1,1)==2)+(M(3,3,1)==2)+...
    (M(1,1,3)==2)+(M(1,3,3)==2)+(M(3,1,3)==2)+(M(3,3,3)==2)-...
    (M(1,1,1)==-2)-(M(1,3,1)==-2)-(M(3,1,1)==-2)-(M(3,3,1)==-2)-...
    (M(1,1,3)==-2)-(M(1,3,3)==-2)-(M(3,1,3)==-2)-(M(3,3,3)==-2);

n7=(M(1,1,1)==1)+(M(1,3,1)==-1)+(M(3,1,1)==-1)+(M(3,3,1)==1)+...
    (M(1,1,3)==1)+(M(1,3,3)==-1)+(M(3,1,3)==-1)+(M(3,3,3)==1)-...
    (M(1,1,1)==-1)-(M(1,3,1)==1)-(M(3,1,1)==1)-(M(3,3,1)==-1)-...
    (M(1,1,3)==-1)-(M(1,3,3)==1)-(M(3,1,3)==1)-(M(3,3,3)==-1)+...
    (M(1,1,1)==-3)+(M(1,3,1)==-3)+(M(3,1,1)==3)+(M(3,3,1)==3)+...
    (M(1,1,3)==3)+(M(1,3,3)==3)+(M(3,1,3)==-3)+(M(3,3,3)==-3)-...
    (M(1,1,1)==3)-(M(1,3,1)==3)-(M(3,1,1)==-3)-(M(3,3,1)==-3)-...
    (M(1,1,3)==-3)-(M(1,3,3)==-3)-(M(3,1,3)==3)-(M(3,3,3)==3);

n=[n1,n2,n3,n4,n5,n6,n7];

```

```

function w=weight(a,b,M)
% checked
a1=2/(1-a^2)-2;

```

```

a2=2-2/sqrt(1+a^2);

a3=sqrt(2)-(sqrt(1+(1-a)^2)+sqrt(1+(1+a)^2))/sqrt(a^4+4);

a4=2*sqrt(2)/(2-a^2/2)-2/sqrt(2+a^2/2);

a5=sqrt(2)-2/sqrt(2+a^2);

a6=2/sqrt(3)-1/sqrt(3-2*a+a^2)-1/sqrt(3+2*a+a^2);

a7=1/sqrt(3+a^2/2+2*a)+1/sqrt(3+a^2/2-2*a)-2/sqrt(3+a^2/2);

% a1=-2/(1-a^2)+2;

% a2=-2+2/sqrt(1+a^2);

% a3=-sqrt(2)+(sqrt(1+(1-a)^2)+sqrt(1+(1+a)^2))/sqrt(a^4+4);

% a4=-2*sqrt(2)/(2-a^2/2)+2/sqrt(2+a^2/2);

% a5=-sqrt(2)+2/sqrt(2+a^2);

% a6=-2/sqrt(3)+1/sqrt(3-2*a+a^2)+1/sqrt(3+2*a+a^2);

% a7=-1/sqrt(3+a^2/2+2*a)-1/sqrt(3+a^2/2-2*a)+2/sqrt(3+a^2/2);

e=[a1,a2,a3,a4,a5,a6,a7];

wcR=exp(-b*e*countR(M)');

wcL=1/wcR;

wcU=exp(-b*e*countU(M)');

wcD=1/wcU;

wcF=exp(-b*e*countF(M)');

wcB=1/wcF;

w=[wcR,wcL,wcU,wcD,wcF,wcB];

function subM=submat(i,j,k,M)

%pick the submatrix from the big Matrix

m=size(M,1);

```

```

subM=M([1+mod(i-2,m),1+mod(i-1,m),1+mod(i,m)],...
        [1+mod(j-2,m),1+mod(j-1,m),1+mod(j,m)],...
        [1+mod(k-2,m),1+mod(k-1,m),1+mod(k,m)]);

function [Mnew,N]=upd(a,b,n,M)

%update function. a is the length of dipole.

% b is the Boltzman constant. n

%is times of loops. M is initial matrix.

N=1;

t=0;

dirs=[1,-1,2,-2,3,-3];

m=size(M,1);

while N<=n

    if t>300 break;end

    % really really really important!!initialize matrix

    WM=[];pos=[];

    IJK=ceil(m*rand(1,3));

    i0=IJK(1);j0=IJK(2);k0=IJK(3);

    if M(i0,j0,k0)==0

        continue

    end

    org=M(i0,j0,k0); %original direction%

    M(i0,j0,k0)=0;

    subM=submat(i0,j0,k0,M);

    i1=[1+mod(i0-2,m),1+mod(i0-1,m),1+mod(i0,m)];

    j1=[1+mod(j0-2,m),1+mod(j0-1,m),1+mod(j0,m)];

```

```

k1=[1+mod(k0-2,m),1+mod(k0-1,m),1+mod(k0,m)];

%find 0's and record their positions
[x,y]=find(subM==0);
z=ceil(y/3);
y=1+mod(y-1,3);
pos=[i1(x)',j1(y)',k1(z)'];
n0=length(x);
%weight matrix for submatrix
for i=1:n0
    WM(i,:)=weight(a,b,submat(pos(i,1),
        pos(i,2),pos(i,3),M));
end

u=sum(sum(WM));
v=u*rand(1);
WMT=WM';
WMR=WMT(:);
n1=1+sum(cumsum(WMR)<v);
np=ceil(n1/6);           %number to locate position
nd=1+mod(n1-1,6);      %number to locate direction
%if direction in the chosen cell doesn't change
%, then not count%
if org==dirs(nd) && pos(np,1)==i0
    && pos(np,2)==j0 && pos(np,3)==k0
        M(i0,j0,k0)=org;

```



```

        t=t+1;continue;

    end

    M(pos(np,1),pos(np,2),pos(np,3))=dirs(nd);

    N=N+1;

    t=0;

end

Mnew=M;

N=N-1;

%generate a m*m*m matrix

function M=gen(p,m)

dirs=[0 1 -1 2 -2 3 -3];

%directions, index for the following, index search

R=rand(m,m,m);

%use logic judge get numbers indicate times

%when they are ture, then could get

%the corresponding interval,use 1+..

%since index start from 1, 0 cant work.

result=1+(R>=1-6*p)+(R>=1-5*p)+(R>=1-4*p)

+(R>=1-3*p)+(R>=1-2*p)+(R>=1-p);

M=dirs(result); %map resuslt(1 2 3 4 5 6 7) to dirs

%count the length of chains

function C=lofc(A)

m=size(A,1);

```

```

C=zeros(m+5,6);

while sum(sum(sum(abs(A))))~=0

    IJK=ceil(m*rand(1,3));

    i0=IJK(1);j0=IJK(2);k0=IJK(3);

    n=0;

    if A(i0,j0,k0)==0 continue;end

    switch A(i0,j0,k0)

        case 1

            i=i0;j=j0;k=k0;

            %if A(i,1+mod(j-2,m),k)==1 continue;end

            while A(i,1+mod(j-2,m),k)==1

                n=n+1;

                A(i,1+mod(j-2,m),k)=0;

                j=1+mod(j-2,m);

            end

            i=i0;j=j0;k=k0;

            while A(i,j,k)==1

                n=n+1;

                A(i,j,k)=0;

                j=1+mod(j,m);

            end

            C(n,1)=C(n,1)+1;

        case -1

            %if A(i,1+mod(j,m),k)==-1 continue;end

            i=i0;j=j0;k=k0;

```

```

while A(i,1+mod(j,m),k)==-1
    n=n+1;
    A(i,1+mod(j,m),k)=0;
    j=1+mod(j,m);
end
i=i0;j=j0;k=k0;
while A(i,j,k)==-1
    n=n+1;
    A(i,j,k)=0;
    j=1+mod(j-2,m);
end
C(n,2)=C(n,2)+1;
case 2
%if A(1+mod(i,m),j,k)==2 continue;end
i=i0;j=j0;k=k0;
while A(1+mod(i,m),j,k)==2
    n=n+1;
    A(1+mod(i,m),j,k)=0;
    i=1+mod(i,m);
end
i=i0;j=j0;k=k0;
while A(i,j,k)==2
    n=n+1;
    A(i,j,k)=0;
    i=1+mod(i-2,m);
end

```

```

C(n,3)=C(n,3)+1;
case -2
%if A(1+mod(i-2,m),j,k)==-2 continue;end
i=i0;j=j0;k=k0;
while A(1+mod(i-2,m),j,k)==-2
    n=n+1;
    A(1+mod(i-2,m),j,k)=0;
    i=1+mod(i-2,m);
end
i=i0;j=j0;k=k0;
while A(i,j,k)==-2
    n=n+1;
    A(i,j,k)=0;
    i=1+mod(i,m);
end
C(n,4)=C(n,4)+1;
case 3
%if A(i,j,1+mod(k,m))==3 continue;end
i=i0;j=j0;k=k0;
while A(i,j,1+mod(k,m))==3
    n=n+1;
    A(i,j,1+mod(k,m))=0;
    k=1+mod(k,m);
end
i=i0;j=j0;k=k0;
while A(i,j,k)==3

```

```

        n=n+1;

        A(i, j, k)=0;

        k=1+mod(k-2, m);

    end

    C(n, 5)=C(n, 5)+1;

case -3

    %if A(i, j, 1+mod(k-2, m))==-3 continue;end

    i=i0; j=j0; k=k0;

    while A(i, j, 1+mod(k-2, m))==-3

        n=n+1;

        A(i, j, 1+mod(k-2, m))=0;

        k=1+mod(k-2, m);

    end

    i=i0; j=j0; k=k0;

    while A(i, j, k)==-3

        n=n+1;

        A(i, j, k)=0;

        k=1+mod(k, m);

    end

    C(n, 6)=C(n, 6)+1;

end

end

C(m+2, :)=sum(C(2:m, :));

C(m+4, :)=sum(C(1:m, :));

for i=1:6

    for j=1:m

```

```

        C(m+1,i)=C(m+1,i)+j*C(j,i);
    end

    C(m+3,i)=(C(m+1,i)-C(1,i))/C(m+2,i);

    C(m+5,i)=C(m+1,i)/C(m+4,i);

end

C(isnan(C)==1) = 0;

```

```

function [C,time,beta]=cnepl(B,n)

d=size(B,1)+5;

C=zeros(d,6,n);

time=zeros(1,n);

beta=linspace(0.01,7.1614,n);

for i=1:n

    [A,t]=upd(0.1,100*beta(i),1500,B);

    C(:, :, i)=lofc(A);

    time(i)=t;

    %m=length(C);

    %D(i,1)=beta(i);

    %number of chains%

    %D(i,2)=sum(C(m-1,:));

    %average length of chains%

    %D(i,3)=C(m-1,:)*C(m,:)/D(i,2);

end

```

```

function [C,time,beta]=cnepl2(B,n)

```

```

d=size(B,1)+5;

C=zeros(d,6,n);

time=zeros(1,n);

beta=linspace(0.01,57.2914,n);

for i=1:n

    [A,t]=upd(0.2,25*beta(i),1500,B);

    C(:, :, i)=lofc(A);

    time(i)=t;

    %m=length(C);

    %D(i,1)=beta(i);

    %number of chains%

    %D(i,2)=sum(C(m-1, :));

    %average length of chains%

    %D(i,3)=C(m-1, :)*C(m, :)' / D(i,2);

end

function [C,time,beta]=cnep3(B,n)

d=size(B,1)+5;

C=zeros(d,6,n);

time=zeros(1,n);

beta=linspace(0.01,193.359,n);

for i=1:n

    [A,t]=upd(0.3,100/9*beta(i),1500,B);

    C(:, :, i)=lofc(A);

    time(i)=t;

    %m=length(C);

```

```

%D(i,1)=beta(i);

%number of chains%

%D(i,2)=sum(C(m-1,:));

%average length of chains%

%D(i,3)=C(m-1,:)*C(m,:)/D(i,2);

end

function [C1,t1,C2,t2,C3,t3,beta]=cnepall(B,n)

C1=zeros(size(B,1)+5,6,n);C2=zeros(size(B,1)+5,6,n);

C3=zeros(size(B,1)+5,6,n);

beta=linspace(0.01,19.651,n);

t1=zeros(1,n);

t2=zeros(1,n);

t3=zeros(1,n);

%beta=linspace(0.01,40,n);

for i=1:n

    [A1,ti1]=upd(0.1,100*beta(i),1500,B);

    [A2,ti2]=upd(0.2,25*beta(i),1500,B);

    [A3,ti3]=upd(0.3,100/9*beta(i),1500,B);

    C1(:, :, i)=lofc(A1);C2(:, :, i)=lofc(A2);C3(:, :, i)=lofc(A3);

    t1(i)=ti1;t2(i)=ti2;t3(i)=ti3;

    %m=length(C1);

    %D1(i,1)=beta(i);D2(i,1)=beta(i);D3(i,1)=beta(i);

    %number of chains%

    %D1(i,2)=sum(C1(m-1,:));

    %D2(i,2)=sum(C2(m-1,:));

```



```

%D3(i,2)=sum(C3(m-1,:));

%average length of chains%

%D1(i,3)=C1(m-1,:)*C1(m,:)/D1(i,2);

%D2(i,3)=C2(m-1,:)*C2(m,:)/D2(i,2);

%D3(i,3)=C3(m-1,:)*C3(m,:)/D3(i,2);

end

```

```

function h=matu(B)

A=zeros(size(B)+2);

n=length(A);

A(2:n-1,2:n-1,2:n-1)=B;

Ur=zeros(size(A));

Ul=zeros(size(A));

Vu=zeros(size(A));

Vd=zeros(size(A));

Wf=zeros(size(A));

Wb=zeros(size(A));

```

```

Ur(A==1)=1;

Ul(A==-1)=-1;

Vu(A==2)=-1;

Vd(A==-2)=1;

Wf(A==3)=-1;

Wb(A==-3)=1;

```

```

% U(A==0)=0;V(A==0)=0;W(A==0)=0;
% U(A==1)=1;V(A==1)=0;W(A==1)=0;
% U(A==-1)=-1;V(A==-1)=0;W(A==-1)=0;
% U(A==2)=0;V(A==2)=1;W(A==2)=0;
% U(A==-2)=0;V(A==-2)=-1;W(A==-2)=0;
% U(A==3)=0;V(A==3)=0;W(A==3)=1;
% U(A==-3)=0;V(A==-3)=0;W(A==-3)=-1;

T=zeros(size(A));

% [x,y,z]=meshgrid(0.5:1:n-0.5,0.5:1:n-0.5,-0.5:-1:-n+0.5);

[xr,yr,zr]=meshgrid(0:1:n-1,0.5:1:n-0.5,0.5:1:n-0.5);
[xl,yl,zl]=meshgrid(1:1:n,0.5:1:n-0.5,0.5:1:n-0.5);

[xu,yu,zu]=meshgrid(0.5:1:n-0.5,1:1:n,0.5:1:n-0.5);
[xd,yd,zd]=meshgrid(0.5:1:n-0.5,0:1:n-1,0.5:1:n-0.5);

[xf,yf,zf]=meshgrid(0.5:1:n-0.5,0.5:1:n-0.5,1:1:n);
[xb,yb,zb]=meshgrid(0.5:1:n-0.5,0.5:1:n-0.5,0:1:n-1);

% [x,y,z]=meshgrid(0,0:1:n,0:1:n);
% [sx sy sz] = meshgrid(0:1:10,0:1:10,0:-1:-10);

```

```

% verts=stream3(x,y,z,T,V,T,x,y,z);

% streamline(verts)

% % view(3)

% % alpha(0.6);

% % axis tight

% % shading interp;

% % camlight; lighting gouraud

% hold on;

% verts=stream3(sx,sy,sz,T,V,T,sx,sy,sz);

% streamtube(verts,1);

% alpha(0.7);

% % streamtube(x,y,z,T,V,T,x,y,z);

% view(3)

% axis tight

% shading interp;

% camlight; lighting gouraud

% hold on;

quiver3(xr,yr,zr,Ur,T,T,'b','LineWidth',1.5);

hold on;

quiver3(xl,yl,zl,Ul,T,T,'c','LineWidth',1.5);

hold on;

% coneplot(Ur,T,T,xr,yr,zr);

```

```

% hold on;

%

% coneplot(Ul,T,T,xl,yl,zl);

% hold on;

verts_r=stream3(xr,yr,zr,Ur,T,T,xr,yr,zr);

streamtube(verts_r,1)

alpha(0.6)

hold on;

verts_l=stream3(xl,yl,zl,Ul,T,T,xl,yl,zl);

streamtube(verts_l,1)

alpha(0.6)

hold on;

quiver3(xu,yu,zu,T,Vu,T,'r','LineWidth',1.5);

hold on;

quiver3(xd,yd,zd,T,Vd,T,'g','LineWidth',1.5);

hold on;

verts_u=stream3(xu,yu,zu,T,Vu,T,xu,yu,zu);

streamtube(verts_u,1)

alpha(0.6)

hold on;

verts_d=stream3(xd,yd,zd,T,Vd,T,xd,yd,zd);

```

```

streamtube(vertsd,1)

alpha(0.6)

hold on;

quiver3(xf,yf,zf,T,T,Wf,'k','LineWidth',1.5);

hold on;

quiver3(xb,yb,zb,T,T,Wb,'w','LineWidth',1.5);

hold on;

vertsf=stream3(xf,yf,zf,T,T,Wf,xf,yf,zf);

streamtube(vertsf,1)

alpha(0.6)

hold on;

vertsb=stream3(xb,yb,zb,T,T,Wb,xb,yb,zb);

streamtube(vertsb,1)

alpha(0.6)

hold off;

view(3)

shading interp;

camlight; lighting gouraud

%*****%

% vertsl=stream3(xl,yl,zl,Ul,T,T,x,y,z);

```

```

% streamline(verts1)

% hold off;

% axis tight

function h=mat3dv(A)

U=zeros(size(A));

V=zeros(size(A));

W=zeros(size(A));

n=length(A);

U(A==0)=0;V(A==0)=0;W(A==0)=0;

U(A==1)=1;V(A==1)=0;W(A==1)=0;

U(A==-1)=-1;V(A==-1)=0;W(A==-1)=0;

U(A==2)=0;V(A==2)=1;W(A==2)=0;

U(A==-2)=0;V(A==-2)=-1;W(A==-2)=0;

U(A==3)=0;V(A==3)=0;W(A==3)=1;

U(A==-3)=0;V(A==-3)=0;W(A==-3)=-1;

% X=ones(size(A));

% Y=ones(size(A));

% Z=ones(size(A));

T=zeros(size(A));

% [x,y,z]=meshgrid(0.5:1:19.5,0.5:1:19.5,-0.5:-1:-19.5);

% quiver3(x,y,z,U,T,T,'r');

% grid minor;

% view(30,40);

% camproj perspective;

```

```

[x,y,z]=meshgrid(0.5:1:n-0.5,0.5:1:n-0.5,-0.5:-1:-n+0.5);
% for i=1:4
% subplot(2,2,i);

% i=1;
% quiver3(x(:, :, 1+5*(i-1):5*i),y(:, :, 1+5*(i-1):5*i),
z(:, :, 1+5*(i-1):5*i),T(:, :, 1+5*(i-1):5*i),
T(:, :, 1+5*(i-1):5*i),W(:, :, 1+5*(i-1):5*i), 'y');
% hold on;
% quiver3(x(:, :, 1+5*(i-1):5*i),y(:, :, 1+5*(i-1):5*i),
z(:, :, 1+5*(i-1):5*i),U(:, :, 1+5*(i-1):5*i),
T(:, :, 1+5*(i-1):5*i),T(:, :, 1+5*(i-1):5*i), 'b', 'LineWidth', 2);
% hold on;
% quiver3(x(:, :, 1+5*(i-1):5*i),y(:, :, 1+5*(i-1):5*i),
z(:, :, 1+5*(i-1):5*i),T(:, :, 1+5*(i-1):5*i),
V(:, :, 1+5*(i-1):5*i),T(:, :, 1+5*(i-1):5*i), 'r', 'LineWidth', 2);
% hold off;
% % end
% view(30,40);
% camproj perspective;

quiver3(x,y,z,T,T,W, 'k', 'LineWidth', 2);
hold on;

```

```
quiver3(x,y,z,U,T,T,'b','LineWidth',2);

hold on;

quiver3(x,y,z,T,V,T,'r','LineWidth',2);

hold off;

% end

view(30,40);

camproj perspective;

%

% daspect([1,1,1]);

% h=coneplot(X,Y,Z,U,V,W,'nointerp',0.4);

% set(h,'FaceColor','blue','EdgeColor','none')

% view(30,40);

% camproj perspective;
```



## Appendix B

# Generalized Dipole Moment And Potential Energy Of Dipole

The dipole moment is the sum of the products of the amount of each charge and the distance between their centroid. The centroid of charges in a system is similar to the center of mass of a system [11]. Dipole moment can be described as

$$\mathbf{p} = \sum_{i=1}^n q_i \mathbf{r}_i, \quad (\text{B.1})$$

where  $q_i$  is the quantity of  $i$ th charge with signs,  $\mathbf{r}_i$  is the distance vector from system centroid to the  $i$ th charge. The simplest example of this is a dipole with equal magnitude but opposite sign charges separated by some distance.

With the definition of electric field  $\mathbf{E}$ , Coulomb's law for the electric field at a point  $\mathbf{r}$  due to a point charge  $q$  is

$$\mathbf{E} = \frac{q\mathbf{r}}{r^3}. \quad (\text{B.2})$$

Similarly, the electric field at a point  $\mathbf{r}$  due to a number of point charges  $q_n$ , located at positions  $\mathbf{r}_n$ , is given by

$$\mathbf{E}(\mathbf{r}) = \sum_{i=1}^n \frac{q_i(\mathbf{r} - \mathbf{r}_i)}{|\mathbf{r} - \mathbf{r}_i|^3}. \quad (\text{B.3})$$

The energy of dipole in an external electric field can be written as

$$U_p = -\mathbf{p} \cdot \mathbf{E}. \quad (\text{B.4})$$

An object with polarity is influenced by a torque  $\tau$  when presented in an external electric field. The torque will align the dipole with the field, and result in an orientation of lower potential energy. For a spatially uniform electric field  $\mathbf{E}$ , the torque is given by

$$\tau = \mathbf{p} \times \mathbf{E}. \quad (\text{B.5})$$

In general, the torque can be calculated if we know the dipole moment and the distribution

of electric field around the dipole.

The potential energy of two dipoles  $\mathbf{p}$  and  $\mathbf{p}'$ , a distance  $\mathbf{r}$  apart, is given by

$$U_{\mathbf{p}\mathbf{p}'} = -\mathbf{p} \cdot \mathbf{E}_{\mathbf{p}'} = \frac{\mathbf{p} \cdot \mathbf{p}' - 3(\mathbf{p} \cdot \hat{\mathbf{r}})(\mathbf{p}' \cdot \hat{\mathbf{r}})}{r^3}, \quad (\text{B.6})$$

where  $\hat{\mathbf{r}} = \frac{\mathbf{r}}{|\mathbf{r}|}$ ,  $\mathbf{E}_{\mathbf{p}'}$  is electric field generated by  $\mathbf{p}'$ .

Since each dipole moment in uniform lattice has same mode, therefore  $\|\mathbf{p}\| = \|\mathbf{p}'\|$ , according to (B.6),

$$U_{\mathbf{p}\mathbf{p}'} = -\mathbf{p} \cdot \mathbf{E}_{\mathbf{p}'} = -\frac{\mathbf{p} \cdot \mathbf{p}' - 3(\mathbf{p} \cdot \hat{\mathbf{r}})(\mathbf{p}' \cdot \hat{\mathbf{r}})}{r^3} \quad (\text{B.7})$$

$$= -\|\mathbf{p}\|^2 \frac{\cos \theta_{\mathbf{p}\mathbf{p}'} - 3 \cos \theta_{\mathbf{p}\mathbf{r}} \cos \theta_{\mathbf{p}'\mathbf{r}}}{r^3} \quad (\text{B.8})$$

$$= \frac{\|\mathbf{p}\|^2}{r^3} f, \quad (\text{B.9})$$

where scale function  $f$  describes the position association among  $\mathbf{p}, \mathbf{p}', \mathbf{r}$ .

# Appendix C

## Units Transformation

In Section 2.2, we mentioned if  $q$  and  $r$  are in appropriate units, the potential energy can be written as (2.4),  $V = q/r$ . The unit system is centimeter-gram-second (CGS) system of units, which only uses centimeter, gram, second as three base units, all the other units can be derived from these three base units.

We shall use the following two units in CGS system:

$$\dagger \text{ Unit of charge: statC} := 1 \text{ statC} = 1 \text{ g}^{\frac{1}{2}} \cdot \text{cm}^{\frac{3}{2}} \cdot \text{s}^{-1}$$

$$\dagger \text{ Unit of energy: erg} := 1 \text{ erg} = 1 \text{ g} \cdot \text{cm}^2 \cdot \text{s}^{-2}$$

For elementary charge  $e$ :

$$1e = 4.80320425 \times 10^{-10} \text{statC}.$$

For angstroms  $\text{\AA}$ :

$$1 \text{\AA} = 10^{-8} \text{ cm}$$

For Boltzmann constant  $k_B$ :

$$k_B = 1.3806488 \times 10^{-16} \text{erg} \cdot \text{K}^{-1}.$$

For the unit of dipole moment  $\mathbf{D}$ :

$$\begin{aligned} 1 \mathbf{D} &= 10^{-18} \text{ statC} \cdot \text{cm} \\ &= 10^{-10} \text{ statC} \cdot \text{\AA} \\ &\approx 0.2082 \text{ e}\text{\AA}. \end{aligned}$$

The potential energy of two point elementary charges separated by one angstrom in (1.2)

can be adjusted in the following form

$$U = \frac{e^2}{\text{\AA}} \quad (\text{C.1})$$

$$= \frac{(4.80320425 \times 10^{-10} \text{statC})^2}{10^{-8} \text{cm}} \quad (\text{C.2})$$

$$= 2.30708 \times 10^{-11} \text{erg}. \quad (\text{C.3})$$

Then inside Boltzmann factor (1.5),

$$\beta U = \frac{U}{k_B T} \quad (\text{C.4})$$

$$= \frac{U}{k_B \cdot 300\text{K}} \cdot \frac{300\text{K}}{T} \quad (\text{C.5})$$

$$= \frac{2.30708 \times 10^{-11} \text{erg}}{1.3806488 \times 10^{-16} \text{erg} \cdot \text{K}^{-1} \cdot 300\text{K}} \cdot \frac{300\text{K}}{T} \quad (\text{C.6})$$

$$= 557 \cdot \frac{300\text{K}}{T} \quad (\text{C.7})$$

K is unit 'Kelvin' for temperature  $T$ .

Since the above equations are derived for two point elementary charges separated by one

angstrom. If the charge  $q$  and distance  $r$  become  $q = ve$  and  $r = l$  in our units, then

$$\beta U = 557 \cdot \frac{300\text{K}}{T} \cdot \frac{v^2}{l} \quad (\text{C.8})$$

Furthermore, if energy  $U$  is specified in state  $i$  in our 3-D physical model,

$$\beta U_i = 557 \cdot \frac{300\text{K}}{T} \cdot \frac{v^2}{l} \cdot f_i(a) \quad (\text{C.9})$$

$$= 557 \cdot \frac{300\text{K}}{T} \cdot \frac{v^2 l^2 a^2}{l^3} \cdot \frac{f_i(a)}{a^2} \quad (\text{C.10})$$

$$= 557 \cdot \frac{300\text{K}}{T} \cdot \frac{M_d^2}{l^3} \cdot g_i(a). \quad (\text{C.11})$$

## **Appendix D**

# **Taylor Series Expansion Of The Configuration Factor $c_F$**

Using Taylor series to expand expressions in Section 2.2 around 0 to the second power:



Pattern	Initial Expression	Taylor Series Expansion
1	$\frac{2}{1-a^2} - 2$	$2a^2 + O(a)^3$
2	$2 - \frac{2}{\sqrt{1+a^2}}$	$a^2 + O(a)^3$
3	0	0
4	$\sqrt{2} - \frac{\sqrt{1+(1-a)^2} + \sqrt{1+(1+a)^2}}{\sqrt{a^4+4}}$	$-\frac{a^2}{4\sqrt{2}} + O(a)^3$
5	$\frac{2\sqrt{2}}{2 - \frac{a^2}{2}} - \frac{2}{\sqrt{2 + \frac{a^2}{2}}}$	$\frac{3a^2}{4\sqrt{2}} + O(a)^3$
6	$\frac{2}{\sqrt{3}} - \frac{1}{\sqrt{3+a^2-2a}} - \frac{1}{\sqrt{3+a^2+2a}}$	$O(a)^3$
7	$\frac{1}{\sqrt{3 + \frac{a^2}{2} + 2a}} + \frac{1}{\sqrt{3 + \frac{a^2}{2} - 2a}} - \frac{2}{\sqrt{3 + \frac{a^2}{2}}}$	$\frac{a^2}{3\sqrt{3}} + O(a)^3$
8	$\sqrt{2} - \frac{2}{\sqrt{a^2+2}}$	$\frac{a^2}{2\sqrt{2}} + O(a)^3$
9	0	0

From all the above extensions, when  $a$  is close to 0, all nine patterns can be expressed in term of  $c_f \cdot a^2$ , where  $c_f$  is a constant specified by a particular pattern. Then  $g(a)$  mentioned in (2.7) only very slightly depends on value of  $a$ ,  $g(a) = \sum c_f + O(a)^3 = c_F + O(a)^3$ . Consequently, given same configuration, the difference of potential energies  $U$  among all three  $a$  levels ( $a = 0.1, a = 0.2, a = 0.3$ ) are tiny.

# Appendix E

## Monotonic Kernel Fitting Function

As mentioned in Section 3.1, we need a fitting function to capture the monotonic curve trends in various intervals. This can be fulfilled through setting an always non-negative function as the derivative of the fitting function. And the non-negative function will be our kernel function  $k(x)$ . Support the curve fitting function is  $C(x)$  and its derivative is  $C'(x)$ .

A good choice of kernel function is

$$k(x) = \frac{1}{\cosh x^2}.$$

$k(x)$  can be scaled and shifted to  $l(x)$ :

$$l(x) = k\left(\frac{x-a}{b}\right)$$

For Figure 3.1 and Figure 3.2,  $C'(x)$  should be always non-negative, and it should be close 0 around  $x = 0$  and this requires  $C'(x)$  in the form

$$\begin{aligned}
 C'(x) &= xl(x) \\
 &= x \left[ ck \left( \frac{x-a}{b} \right) \right]. \\
 C(x) &= \int C'(x) dx = \int x \left[ ck \left( \frac{x-a}{b} \right) \right] dx \\
 &= -b_1^2 c_1 \log(\cosh \frac{x-a}{b}) + bcx \tanh \frac{x-a}{b} + c_0,
 \end{aligned}$$

where  $b, c$  should be greater than 0,  $c_0$  is the initial value in integration.

For Figure 3.3,  $C'(x)$  should be always non-positive.

$$\begin{aligned}
 C'(x) &= l(x) \\
 C(x) &= \int C'(x) dx = \int ck \left( \frac{x-a}{b} \right) dx \\
 &= bc \tanh \frac{x-a}{b} + c_0,
 \end{aligned}$$

where  $b, c$  should be less than 0,  $c_0$  is the initial value in integration.

For Figure 3.4, it gets more complicated. It's not monotonic anymore. However, we can cut the whole interval into two parts at some point  $\phi$  around  $\gamma = 2$ . In those two parts, the trends are still monotonic, increasing and decreasing respectively. Only one kernel will not be enough to get a good fit. Then  $C'(x)$  should be written as

$$\begin{aligned}
C'(x) &= (\phi - x) [l_1(x) + l_2(x)] \\
&= (\phi - x) \left[ c_1 k\left(\frac{x-a_1}{b_1}\right) + c_2 k\left(\frac{x-a_2}{b_2}\right) \right], \\
C(x) &= \int C'(x) dx \\
&= b_1^2 c_1 \log(\cosh \frac{x-a_1}{b_1}) + b_2^2 c_2 \log(\cosh \frac{x-a_2}{b_2}) \\
&\quad + b_1 c_1 (\phi - x) \tanh \frac{x-a_1}{b_1} + b_2 c_2 (\phi - x) \tanh \frac{x-a_2}{b_2} + c_0,
\end{aligned}$$

where  $b_1, b_2, c_1, c_2$  are greater than 0 and  $t$  is around 2.  $c_0$  is the initial value in integration.

So  $C'(x)$  will be positive when  $x < \phi$  and negative after that.