Simulation studies on Origin of Life

By

SHILADITYA SWARNAKAR(11MS002)

SUPERVISOR: DR. SUPRATIM SENGUPTA

Department of Physics
INDIAN INSTITUTE OF SCIENCE EDUCATION AND RESEARCH, KOLKATA

A thesis submitted in partial fulfillment of the requirements
for the award of the degree of Master of Science.

APRIL 20, 2017

Abstract

This article uses mathematical modeling and computer simulations to study the origin-of-life.A simplified model(Two Company's Model) of the origin of life is described that maps in a straightforward fashion to a more complex model. It address the conditions under which a living state can emerge from a nonliving state, and find that a spatially extended system with moderate diffusion can greatly facilitate the emergence of the living state. Most importantly, in this regime, the probability of emergence scales with the size of the system, so that the emergence of life somewhere is virtually guaranteed for sufficiently large systems.

Dedication and acknowledgements

I am grateful to Dr. Supratim Sengupta for constant help and for the references.

AUTHOR'S DECLARATION

I declare that the work in this thesis was carried out by me in accordance with the requirements of the University's Regulations and Code of Practice for the degree of Masters in Science and it has not formed the basis of award of any other degree anywhere.

SIGNED: .....DATE: .....

SUPERVISOR'S SIGNATURE: ..... DATE: .....

TABLE OF CONTENTS

List of Figures

CHAPTER 1

INTRODUCTION

In probability theory, the Gillespie algorithm generates a statistically correct trajectory of a stochastic equation. It was created by Joseph L. Doob.We use this method to model origin of life in a spatially localized stochastic transition. Life as we know it relies on the existence of self-replicating biopolymers. Modern cells use DNA, RNA, and proteins, where DNA helps in information storage and proteins acts as catalysts.It is usually thought that the interdependent system of DNA, RNA and proteins that sustains today,Äôs cells is too complex to have arisen all at once.A simpler biopolymer system that could have existed in early organisms is the RNA world hypothesis.The discovery of ribozymes supported a hypothesis, known as the RNA World Hypothesis, that earlier forms of life may have relied solely on RNA to store genetic information and to catalyze chemical reactions.This hypothesis was proposed independently by Carl Woese, Francis Crick and Leslie Orgel in the 1960s – decades before the discovery of ribozymes – and soon after the double-helical structure of DNA was determined.[5] According to the RNA World Hypothesis, life later evolved to use DNA and proteins due to RNA's relative instability and poorer catalytic properties, and gradually, ribozymes became increasingly phased out. The presence of biopolymer catalysts is an essential feature that distinguishes living systems from non-living chemical systems.It is thought that initially from the pool of chemicals monomers were formed.These monomers resulted in the formation of activated monomers and finally polymers. The polymers of certain length in aqueous medium can fold and act as enzyme to catalyze its formation as well as formation of other polymers and hence start autocatalysis. A key feature of living systems is that they are auto catalytic,i.e. they can reproduce and make more of their own components.It is biopolymers that allow organisms to do this.The nonliving and living states are two dynamically stable states of the same chemical system and the origin of life is a stochastic transition between these two states that is initiated by concentration fluctuations involving

relatively small numbers of molecules in a localized spatial region.The central idea of this article is shown schematically in Fig. 1.1.(P.C Google image) We suppose that there is some local region

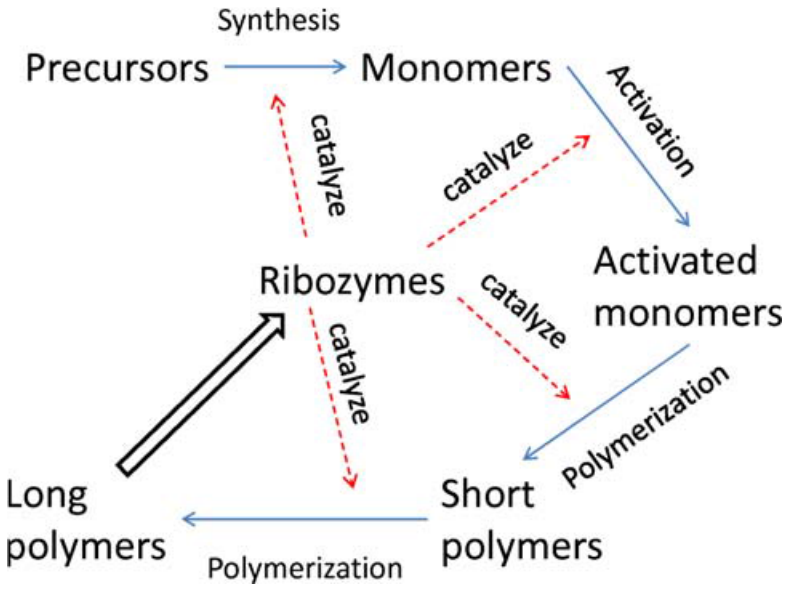

Figure 1.1:

Schematic diagram of an autocatalytic RNA polymerization system in which long polymers act as ribozymes that catalyze the different reactions of the system(P.C Google image)

in which the chemical reactions that give rise to life can begin to occur.The net concentration of polymers in the local region is determined by a balance between the rates of formation and escape.

CHAPTER 2

A DETOUR: GILLESPIE'S ALGORITHM

2.1 Gillespie's Method

2.1.1 Why do we need Gillespie's algorithm?

There are many ways to model biochemical system depending on size details required etc. One method is writing down a set of coupled differential equation where each equation describes a number of reactions.The variables are time dependent concentration of participating molecules and parameters are reaction constants.Concentrations are defined for large numbers of molecules. Moreover when the number of molecules is large any two reactions can take place at the same time. The system of ordinary differential equations for concentrations thus represents a collections of reactions occurring simultaneously all through the reaction volume.This method break down when the numbers of molecules become small, and reactions now occur in some random order rather than simultaneously. One then needs to adopt a new language for the description of the system: probabilities for the state of the system defined by the number of molecules of each type at a given time, replace the differentiable concentrations.[1] These probabilities evolve in time as such or such reaction takes place randomly among all possible reactions. Gillespie's algorithm is a way of implementing consistently this probabilistic description of a biological system. Gillespie algorithm allows a discrete and stochastic simulation of a system with few reactants because every reaction is explicitly simulated.

2.1.2 The algorithm

The crux of the algorithm is the drawing of two random numbers at each time step, one to determine after how much time the next reaction take place, the second one to choose which one of the reactions will occur.Below is a summary of the steps to run the algorithm:

• ***Initialization*** **:** Initialize the number of molecules in the system, reaction constants, and random number generators.

• ***MonteCarloStep*** **:** Generate random numbers to determine the next reaction to occur as well as the time interval. The probability of a given reaction to be chosen is proportional to the number of substrate molecules.

• ***Update*** **:** Increase the time step by the randomly generated time in Step 2. Update the molecule count based on the reaction that occurred.

• ***Iterate*** **:** Go back to Step 2 unless the number of reactants is zero or the simulation time has been exceeded.

2.1.3 A simple model to understand the algorithm: Reversible binding of A and B to form AB dimer

Consider a system of molecules of two types: A and B.[2] In the system A and B reversibly bind together to form AB dimers. So there are two reactions. The first is where one molecule of A reacts reversibly with one B molecule to form an AB dimer, and the second is where an AB dimer dissociates into an A and a B molecule. The reaction rate constant for a given single A molecule reacting with a given single B molecule is k_D, and the reaction rate for an AB dimer breaking up is $k_A B$. So, for example if at time t there is one molecule of each type then the rate of dimer formation is k_D, while if there are n_A molecules of type A and n_B molecules of type B, the rate of dimer formation is $k_D n_A n_B$. If there are $n_A n_B$ dimers then the rate of dimer dissociation is $k_A n_{AB}$. The total reaction rate is R_{tot}. $R_{tot} = k_D n_A n_B + k_B n_A B$

$$\text{A+B} \xrightarrow{k_a} \text{AB}$$
$$\text{AB} \xrightarrow{k_b} \text{A+B}$$

In the algorithm, we advance forward in time in two steps: calculating the time to the next reaction, and determining which of the possible reactions the next reaction is. Reactions are assumed to be completely random, so if the reaction rate at a time t is R_{tot}, then the time, δt, until the next reaction occurs is a random number drawn from exponential distribution function with mean$\frac{1}{R_{tot}}$. Thus, we advance time from t to t + δt. The probability that this reaction is an A molecule binding to a B molecule is simply the fraction of total rate due to this type of reaction, i.e.,the probability that reaction is $P(A + B \rightarrow AB) = \frac{k_D n_A n_B}{R_{tot}}$

The probability that the next reaction is an AB dimer dissociating is just 1 minus that. So with these two probabilities we either form a dimer by reducing $n_A$and n_B by one, and increase n_{AB} by one, or we dissociate a dimer and increase $n_A$and n_B by one and decrease n_{AB} by one.

Now we have both advanced time to t + δt, and performed a single reaction. The Gillespie algorithm just repeats these two steps as many times as needed to simulate the system for however long we want. With $n_A = n_B = 10$ and $n_{AB} = 0$ at t=0 ,and $k_D = 2$ and $k_B = 1$ we perform the simulation in MATLAB shown in Fig. 2.1 We see the number of molecules stabilize after

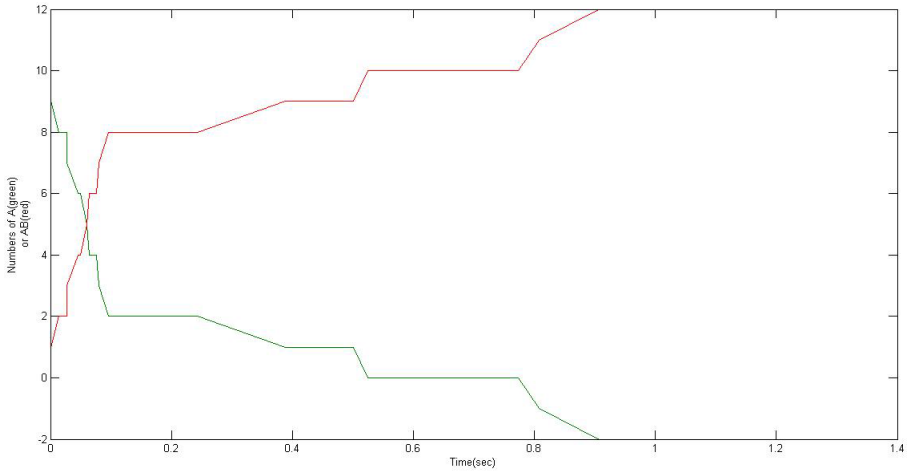

Figure 2.1:

The plot shows how number of A molecules and number of AB dimer changes with time

sometime.

CHAPTER 3

RNA POLYMERIZATION MODEL

3.1 RNA polymerization model

3.1.1 Mathematical model

In the RNA system, we suppose that precursor food molecules are available in the environment at concentrations F_1 and F_2. We suppose that monomers, denoted by A, can be synthesized from F_1, with rate constant s.These monomers can react with F_2 to produce activated monomers, A^*, with rate constant a. RNA polymers of length n are denoted An. An activated monomer can react with a polymer to extend its length by 1 with rate constant r. All molecules can escape from the system at a rate u.P is the total polymer concentration of all length.

- $F_1 \rightarrow$ A Rate constant s
- A+$F_2 \rightarrow A^*$ Rate constant a
- A^*+$A_n \rightarrow A_{n+1}$ Rate constant r
- All molecules can leave the system at the rate u

Under this condition we can model the system with the following differential equation.[3]

$$\frac{dA}{dt} = sF_1 - aF_2A - rAA^* - uA \tag{3.1}$$

$$\frac{dA_n}{dt} = rA^*(A_{n-1} - A_n) - uA_n \tag{3.2}$$

$$\frac{dA^*}{dt} = aF_2A - rA^*(A+P) - uA^* \tag{3.3}$$

3.1.2 Model Analysis

We are interested in the steady-state concentrations when the rates of change are all zero in 3.1 to 3.3 . From 3.2, we know that $A_n/A_{n-1} = z$ where

$$z = rA^*/(u + rA) \tag{3.4}$$

It follows that $A_n = Az^{n-1}$ and hence

$$P = rAA^*/u \tag{3.5}$$

From 3.1 we can write

$$A^* = \frac{sF_1 - aF_2A - uA}{rA} \tag{3.6}$$

and from 3.5 and 3.6 we can write

$$P = \frac{sF_1 - aF_2A - uA}{u} \tag{3.7}$$

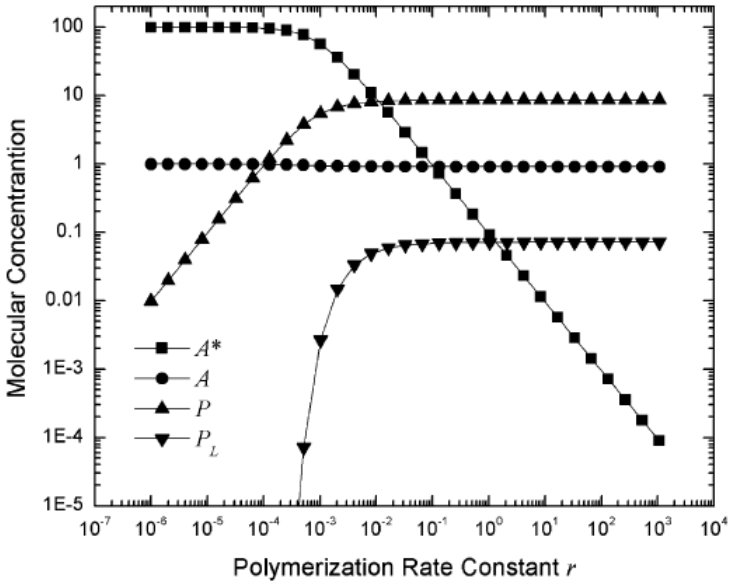

Figure 3.1:

The plot shows how variation of A,A^*,P_L and P with r

We are interested in the way the behavior of the system depends on the polymerization rate.Fig 3.1([3]) shows the effect of varying r over many orders of magnitude while fixing all other parameters with the following values:F1 = F2 = 1, a = 1, s = 1, u = 0.01. These values are arbitrary but they are sufficient to illustrate the qualitative behavior of the model. The limiting

regimes of small and large r can be easily understood. If r is small, it can be seen that A and A^* are roughly constant, and $A^* >> A$. In 3.1,synthesis term $sF_1$is balanced by three terms that reduce A.The most important of these is the term for the activation reaction, aF_2A. If synthesis is balanced only by the activation reaction, then A and sF_1/aF_2 (this is 1.0 with the parameters chosen). In 3.3, A^* is produced by the activation reaction, and in the small r regime, the most important term limiting A^* is the escape term. Therefore,$A*$ and aF_2A/u and sF_1/u (this is 100 with the parameters chosen). $A^* >> A$ in this example because we have supposed that most of the A undergoes the activation reaction before it escapes. We know from 3.7 that $P = rAA^*/u$;hence, in this regime, $Pandr(sF_1)^2/(u^2aF_2)$, which is 10^4r with the parameters chosen. P is small in this regime because r is small, and P is proportional to r, as can be seen from the left side of Fig. 3.1.

We will now consider the large r regime.[3] It can be seen that A is only very slightly lower for large r than it is for small r. Although the polymerization term rAA* in 3.1 is larger when r is larger, it turns out that A^* is small in this regime, so the most important term limiting A is the activation reaction, as it was for small r. Therefore it is still approximately true that A and sF_1/aF_2. If r is large, then the total polymer concentration P is much larger than A. The most important negative term in 3.2 is $rA*P$.Rearranging we obtain A^* $(uaF_2)^{0.5}/r$.which is 0.1/r with these parameters.Thus A^* is small and varies inversely with r when r is large, as can be seen on the right of Fig. 3.1. The total polymer concentration is P $sF_1/(uaF_2)^{0.5}$ This is independent of r for large r, and is large compared to A, as we assumed above. If r is large, almost all monomers get incorporated into polymers. Therefore, increasing r even further does not increase the polymer concentration once this limit is reached. The distribution of polymer lengths is exponentially decreasing, even when r is large, because z is always less than 1. Nevertheless, when r is large, the distribution tails off more slowly; hence, there is an appreciable concentration of long polymers. P_L reaches a fairly large constant value in this regime, as shown on the right of Fig. 3.1. If we have catalytic feedback

$$r = r_0 + kP_L$$

where P_L is the term due to ribosyme catalysis. It is clear that the spontaneous rates of polymerization,synthesis, and activation are important parameters in our model because these will control the concentrations of polymers in the dead state. If these rates are too low then long polymers will not arise in sufficient frequency to initiate the transition. In particular, if there is no spontaneous polymerization (r0 = 0, and r = kP_L), then both the living and the dead states still exist, but the dead state has zero polymer concentration, so there is no way of getting out of the dead state even in a finite volume. It is therefore important that conditions in the local region enable these reactions to occur to some degree, even if the spontaneous rates are low. It is worth pointing out, however, that once the system shifts to the living state, the spontaneous polymerization rate becomes unimportant because $kP_L >> r0$. This means that once life originates, it is not restricted to the favorable local region in which it arose. The reaction system could spread

to a region where r0 = 0,even though it could not arise in such a region.We can see the phase diagram which is similar to the phase diagram of the new model proposed: replicator model.

CHAPTER 4

REPLICATOR MODEL

4.1 A simple model for origin of life

4.1.1 Model description

Now we introduce the simplest model we can have that has the same phase diagram as RNA Polymerization model. The simplified model will then be useful to investigate the dynamic properties of the stochastic transition to life.We consider a single type of replicating molecule whose concentration is ϕ.We have the following scenario:

- Spontaneous replication rates by random polymerization is s. This process does't depend on concentration.
- Existing replicators may be copied with a rate constant r,representing a process of non-living template-directed synthesis. This process is proportional to the current replication concentration, process is proportional to the current replication ϕ.
- Replicators can act both as enzyme(polymerase) and template at the same time and can be replicated by a process called catalytic replication which depends both on the concentration of the template and the polymerase.(both are ϕ)
- The increase in replicator concentration is limited by finite resources. The simplest way to model this is to assume a finite carrying capacity of the system, corresponding to a concentration $\phi = 1$, and to multiply all the growth rates by a factor $1 - \phi$.
- Replicator can die at the rate u.

The differential equation describing this model is[5]:

$$\frac{d\phi}{dt} = (s + r\phi + k\phi^2)(1-\phi) - u\phi \tag{4.1}$$

We try to plot the phase diagram of the model (plot of s Vs k)(Fig 4.1)

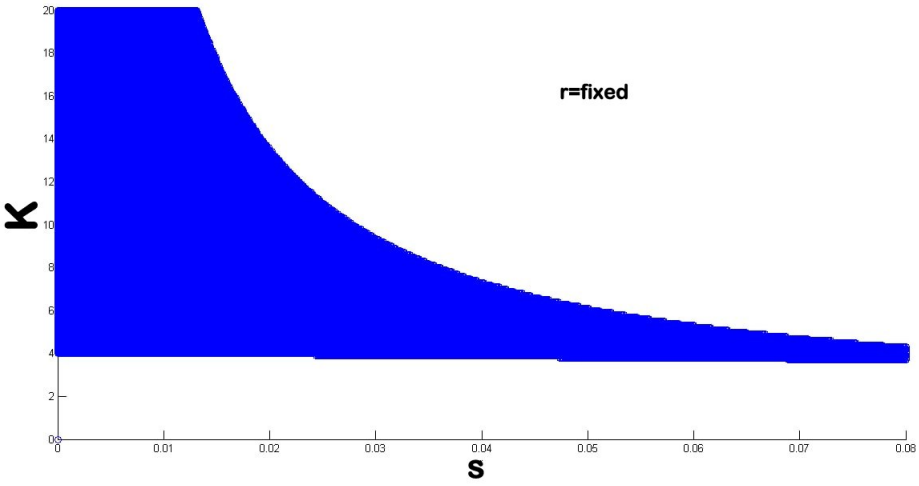

Figure 4.1:

The plot shows phase diagram of replicator model with r=0 and u=1

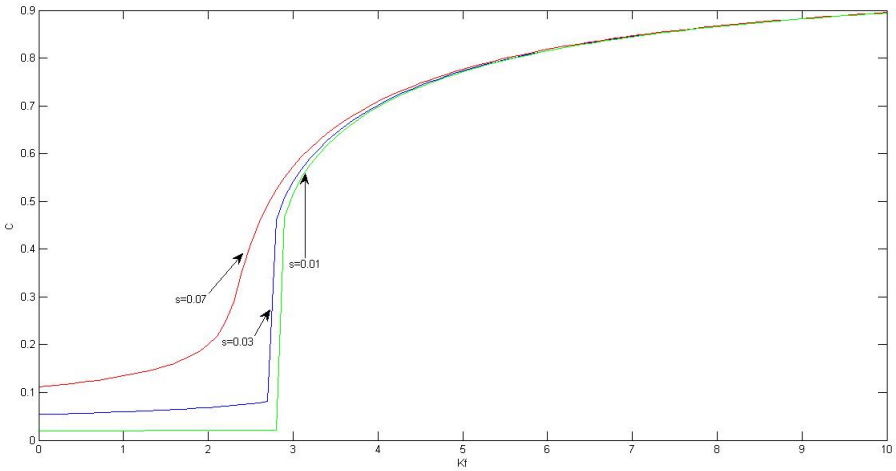

Figure 4.2:

The plot shows the variation of k with concentration(ϕ) for three different s

The region above the blue region has only one real root and is called the living state.The region below the blue region also has one root and is the dead state and the blue region if the

bistable state which has three real root : two stable root separated by one unstable root.The dead state can be identified from the concentration of the replicators.[4] The real root for which the concentration of replicators is minimum is the dead state and the real root for which replicator concentration is maximum is the living state.The lower stable state is a dead state, controlled by the balance between spontaneous generation and death while the upper state is controlled by catalytic term k. By varying r we can see the phase boundaries to vary but the shape remains same.

In Fig. 4.2 we see variation of replicator concentration with catalytic constant k.With s=0.01 and s= 0.03 an S-shaped curve.The upper and lower branches of this curve are stable states that we call living and dead,respectively.In the living state,the concentration of the catalyst is high,and the major mode of production for the catalyst is auto-catalytic(k reaction).In the dead state,the concentration of the catalyst is low and the catalysts are primarily formed chemically(s and r reactions).The middle branch of theS is unstable,and this solution does not arise dynamically. Fig. 4.1 also shows an example with s=0.07,where there is a single stable solution at all values of k.The solution changes smoothly from a dead to a living state as k increases.The boundary between living and dead is not sharp in this case.The point of inflection on this curve is where the catalyst concentration increases most rapidly with k. This is a convenient way of defining the boundary.

CHAPTER 5

TWO'S COMPANY MODEL

5.1 A spatial version of the replicator model

5.1.1 Model Description

In this model[5] we try to show spatial fluctuation and limited diffusion rate can lead to stochastic transition to life.We consider a 2D square lattice of L*L sites. The number of molecules on any one site may be n = 0, 1, 2 or 3 only. The carrying capacity of the whole lattice is therefore $3L^2$. If there are N molecules on the lattice, the mean concentration relative to the carrying capacity is $\phi = N/2L^2$. There are 3-n vacancies on a site with n molecules. Vacancies(3-n) are treated as resources.Here we can have the following process :

- Spontaneous replication at the rate $s(3-n)$
- Linear growth which depends on the number of molecules and molecules are added by this process at the rate $rn(3-n)/2$
- Catalytic growth.The rate of catalytic replication is defined as k/2 times the number of ways of picking a replicator-template pair times the number of resources and is $kn(n-1)(3-n)/2$
- The death rate of a molecule on a site with n molecules is un (i.e.u per molecule).
- Hopping of molecules between sites is implemented using either local or a global hopping rules in the following way. Each molecule attempts to hop at rate h. A destination site is chosen for the molecule. In the case of local hopping rules, the destination is chosen at random from the 8 neighboring sites of the original site. In the case of global hopping rules, the destination site is chosen at random from all the other sites on the lattice. The molecule

hops successfully to the destination site if it finds a vacancy there, i.e. the hop is successful with probability 1-n/3.

5.1.2 Observations from two's company model

We at first look at the processes possible for different N and we have:

- n=0 : Spontaneous and death
- n=1 : Spontaneous, linear and death
- n=2 : Spontaneous,linear,catalytic and death.
- n=3 : Death

So we see catalytic replication is only possible when n=2 while for all other n its zero. Hence the name two's company three is a crowd.

5.1.3 Algorithm and simulations

The code was written for the two company model by using Gillespie's method. At a time at each lattice point we can have spontaneous replication,linear growth,catalytic replication,death and hopping.We run this for a large number of times and observe the result.Simulations are initiated in the dead state(with parameter s=0.02,u=1,k=9,r=0) and followed until a transition to the living state occurs. The concentration in the dead state should be close to the lower stable solution of 4.1. In order to initiate the stochastic simulation in the dead state, each lattice site is seeded with molecular numbers sampled from a binomial distribution with average concentration equal to the lower stable root of the replicator equation(Fig. 5.1).

The system remains in the dead state for a long time until a localized patch of high concentration arises that is sufficiently large to be stable in the living state. The red dot are the molecules that were synthesized catalytically.Initially we see umber of red dots are very small hence it the dead state(Fig. 5.1) and later it becomes living(Fig. 5.2).The clusters of red dots appear and disappear until the a living patch becomes stable and makes the entire lattice living.We performed many simulations of this model in order to measure the way the time taken for the transition depends on the parameters.

5.1.4 How does limited diffusion and spatial fluctuation leads to formation of life?

We define T_{sys} = T_{reg} + T_{spead}, where T_{reg} is the time until a local transition occurs in any one region of the lattice, and T_{spread} is the time for the living state to spread from one region across the rest of the lattice, and T_{sys} is the time required for the full lattice to reach the living state.We measured the $Tsys$ by calculating the time required for the system to go to the living state(i.e the

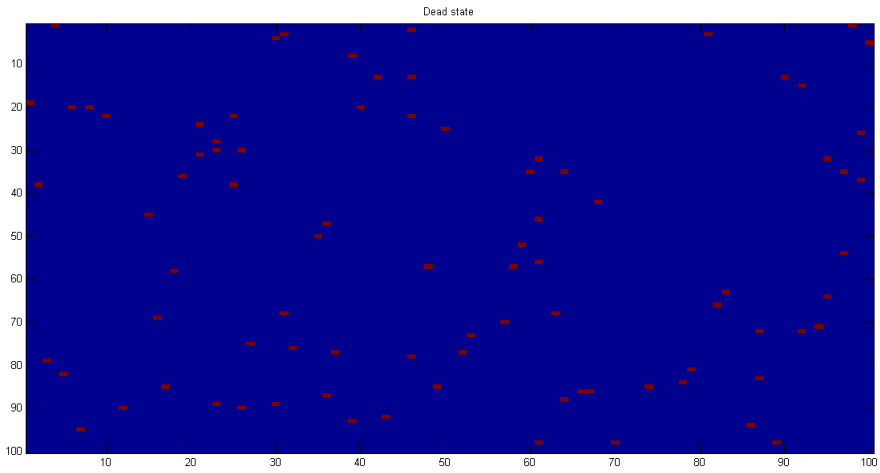

Figure 5.1:

The dead state

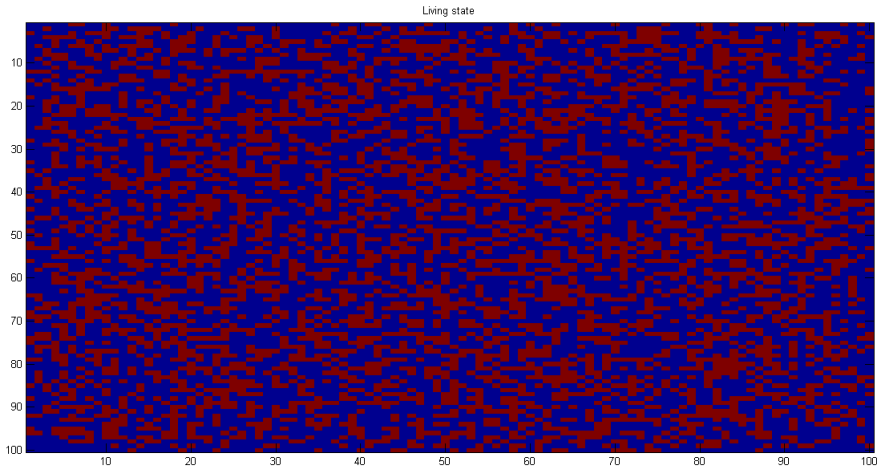

Figure 5.2:

The living state

second stable solution of the replicator equation) and in order to ensure the system is in living state the time was averaged.

Fig. 5.3 shows the average $Tsys$ as h is varied with a fixed system size.We see that local and global hopping converge at large h which shows that the system is well mixed.When h is small the transition time with local hopping is shorter than that with global hopping which shows that the local concentration fluctuations that arise in the local hopping model make the transition

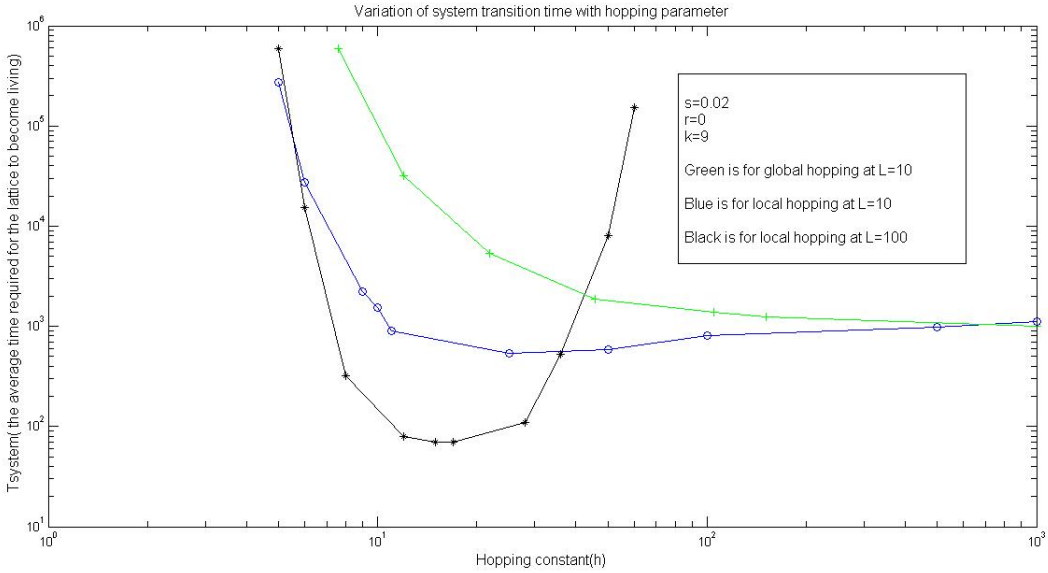

Figure 5.3:

variation of T_{sys} wih hopping parameter h

much easier. It should be remembered that the spontaneous generation rate s is very small, so the concentration of molecules in the dead state is very small. For the catalytic process to occur, it requires two molecules on the same site. If h is too small, molecules do not encounter one another frequently.If molecules do encounter one another and a replication occurs, there will now be three molecules on one site and no further replication occurs. It is necessary for one of these molecules to hop away before a second replication can occur; hence, h should not be too small. It can be seen that,for a particular h for local hopping transition is fastest. This effect is not very strong in the small lattice (L = 10), but is very pronounced in the larger lattice (L = 100) also shown in Fig. 5.3.Finally we have shown that limited diffusion and spatial fluctuation results in formation of faster.

5.2 Discussion and conclusion

Different types of models can be simulated using the Gillespie's method. We have used the method extensively in the article.We have first discussed the RNA polymerization model. We proposed a bunch of differential equation and analyzed it with parameter variation.The key point is that there is a dead (or non-living) state in which P_L is very small and $r \approx r_0$, so that polymerization occurs at the spontaneous rate, and there is a living state in which $kP_L >> r_0$, so that polymerization is auto catalytic and occurs at a much higher rate than the spontaneous rate. We are most interested in the region where both states are stable. The bistable region occurs

if r_0 is fairly small and k is fairly large. If k is too small, only the dead state is stable, and if k is too large, only the living state is stable. Furthermore, if r_0 is large then there is only one stable state for all values of k. Then we discuss the simplest model which generates the same phase diagram as the RNA polymerization model which is the replicator model. Finally in Two's company model we have seen how a lattice becomes living at for a certain hopping constant(diffusion parameter).It is therefore important to have a diffusion rate that is not too large in order for the transition to the living state to occur at an appreciable rate in the first place. In the article we haven't considered the role of complimentary strand and parasites. Their role can also be included and a similar study can be made.(formation of the coupled differential equation and their spatial version). Parasite polymer were an important part of the early environment so their role needs to be studied extensively.

APPENDIX A

APPENDIX A

Code-1 (Gillespie's algorithm for reversible binding of A and B to form AB dimer)

```
Input the reactants and products n_a, n_b,n_ab

Input the rate of reactions k_a and k_b respectively and store them in a array r as r(1) and r(2)

Set initial t=0

while (t less than a choosen value)
{
        r_tot=(k_a*n_a*n_b)+(k_b*n_ab)
        set r(1)=r(1)*n_a*n_b and r(2)=r(2)*n_ab
        delta_t=choose a random number from exponential distribution
        t=t+delta_t
        p=r/r_tot
        s=sum of array p
        q=cumulative summation of (p/s)
        ran=choose a random number between 0 and 1
        If (ran >0 and ran<q(1))
        {
                Reaction 1 is choosen. Perform the reaction i.e decrease n_a,n_b by 1 and
                increase n_ab by 1
        }
        Else
        {
                Reaction 2 is choosen. Perform the reaction i.e decrease n_ab by 1 and increase
                n_a and n_b by 1

        }

 }

Plot n_a,n_b and n_ab Vs t
```

Code-2 (Phase space plot of k Vs s)

```
Set the rates of  linear growth and death rate as r and u respectively.

 The spontaneous rate is s and the catalytic rate is k

 loop ( vary k and s)
{
        Finding roots of the cubic equation with coefficients k,r-k,s-r+u,-s
        If( number of real roots are 3 )
        {
                Store the values of the k and s in an array 'y'
        }
}

Plot k Vs s and color the values of k and s which is in the array y
```

Code-3 (Phase space plot of k Vs concentration of replicases(C))

```
Set the rates of  spontaneous rate,linear growth,death rate as s,r , u respectively.

The catalytic rate is k

 loop ( vary k )
{
        Finding roots of the cubic equation with coefficients k,r-k,s-r+u,-s
        If( the root is not imaginary and not negetive )
        {
                Store the values of the k and the root in an array
        }
}

Plot k vs the the root of the equation
```

Code-4 (Two's Company Model(Local Hopping))

Set the rates of spontaneous,linear,catalytic growth rate and death rate as s,r,k,u respectively and set t=0

Find roots of the replicator equation set to 0 with the given value i.e root of the cubic polynomial with coefficient k,r-k,s-r+u, -s. Store the root as phi1 and phi2 where phi1 is smallest and phi2 largest read positive root.

Choose the smaller positive real root as dead state(phi1) and larger positive real root as living state.(phi2) and choose the lattice size L

Declare a array such that R(i,j,1),R(i,j,2),R(i,j,3),R(i,j,4) stores the total number of particles,particles generated by spontaneous process, particles generated by linear growth and particles generated by catalytic process in the given (i,j) position of the lattice respectively.

R(i,j,2) is seeded with molecular numbers sampled from a binomial distribution with average concentration of the dead state(phi1)

R(i,j,1)= R(i,j,2)+ R(i,j,3)+ R(i,j,4)

Calculate concentration of the lattice s2=sum of R(:,:,1) and phi=s2/(3*L*L)

```
While (phi less than phi2)
{
        Loop( i running from 1 to L and j running from 1 to L )
        {
                n=R(i,j,1)
                sgr=s*(3-n)     // spontaneous
                lgr=r*n*(3-n)/2 //linear
                rgr=k*n*(n-1)*(3-n)/2  //catalytic
                dr=u*n   //death rate
                 hr=h*n   //hopping rate
```

```
Normalize srg,lrg,rgr,dr,hr and store them in itself
p1(1)=sgr
p1(2)=lgr
p1(3)=rgr
p1(4)=dr
p1(5)=hr

s1=sgr+lgr+rgr+dr+hr
p1=p1/s1
q1=cumulative summation of p1
z1=1
Choose a random number 'ran'
while(q1(z1)<ran)
{
        z1=z1+1;
}

Z1 gives us the reaction chosen by Gillespie algorithm and we perform the reactions
 if(z1==1)
{
R(i,j,2)=R(i,j,2)+1
R(i,j,1)=R(i,j,1)+1
}
if(z1==2)
{
R(i,j,3)=R(i,j,3)+1
R(i,j,1)=R(i,j,1)+1
}

if(z1==3)
{



  R(i,j,4)=R(i,j,4)+1
  R(i,j,1)=R(i,j,1)+1
  }
 if(z1==4)
  {
  sam=[2 3 4] //in case of death choosing randomly which particle to delete

    while ()
  {
      x=choose a number randomly from  the array 'sam'

      if((R(i,j,x) not equal to 0)
       {
         R(i,j,x)=R(i,j,x)-1
         R(i,j,1)=R(i,j,1)-1

         break;
       }

    }

}

  If(z==5)
  {
          sam=[2 3 4]
```

```
while 1
{
        x=randomly choose one element from 'sam'
        if((R(i,j,x) not equal to 0))
        {
        break;
       }
}

If( (i,j) point is not in the boundary)
{
    hopp=[R(i-1,j-1,1) R(i-1,j,1) R(i-1,j+1,1) R(i,j-1,1) R(i,j+1,1)
    R(i+1,j-1,1) R(i+1,j,1) R(i+1,j+1,1)]
    hopp=1-hopp/3                    //probability of hopping
    p2=[i-1 i-1 i-1 i i i+1 i+1 i+1]
    p3=[j-1 j j+1 j-1 j+1 j-1 j j+1]


     select an element from hopp depending on the probability given by the
     by content of hopp ( the same way we did for Gillespie algorithm)

     if( R(p2(hoppto),p3(hoppto),1)~=3)
     {
        R(i,j,1)=R(i,j,1)-1;
        R(i,j,x)=R(i,j,x)-1;
        R(p2(hoppto),p3(hoppto),1)=R(p2(hoppto),p3(hoppto),1)+1;
        R(p2(hoppto),p3(hoppto),x)=R(p2(hoppto),p3(hoppto),x)+1;
      }


}
```

```
                Else
                {
                        When the (i,j) is in the boundary we do the same thing just the 'hopp'
                        Array changes and corresponding p1 and p3
                        For i=1,j=1   hopp=[R(1,2,1) R(2,1,1) R(2,2,1)]
                                      p2=[1,2,2]
                                      p3=[2,1,2]
                        For i=1 j=L   hopp=[R(1,L-1,1) R(2,L-1,1) R(2,L,1)]
                                      p2=[1,2,2]
                                      p3=[L-1,L-1,L]
                        For i=L j=1   hopp=[R(L-1,1,1) R(L-1,2,1) R(L,2,1)]
                                      p2=[L-1,L-1,2]
                                      p3=[1,2,2]
                        For i=L j=L   hopp=[R(L-1,L-1,1) R(L-1,L,1) R(L,L-1,1)]
                                      p2=[L-1,L-1,L]
                                      p3=[L-1,L,L-1]



                }

        }

}

delta_t= choose a random number from exponential distribution with mean
1/(srg+lgr+rgr+dr+hr)(Gillespie's algorithm)

t=t+delta_t

|

    Calculate concentration of the lattice s2=sum of R(:,:,1) and phi=s2/(3*L*L)

}
```

After completion we plotted the R with different colors i.e the particles generated by spontaneous, catalytic, linear growth with different colors on the lattice. The time for various 'h' values are noted and plotted for the Tsys Vs h plot.

Code 5 (Two's company model(Global hopping)

Set the rates of spontaneous,linear,catalytic growth rate and death rate as s,r,k,u respectively and set t=0

Find roots of the replicator equation set to 0 with the given value i.e root of the cubic polynomial with coefficient k,r-k,s-r+u, -s. Store the root as phi1 and phi2 where phi1 is smallest and phi2 largest read positive root.

Choose the smaller positive real root as dead state(phi1) and larger positive real root as living state.(phi2) and choose the lattice size L

Declare a array such that R(i,j,1),R(i,j,2),R(i,j,3),R(i,j,4) stores the total number of particles,particles generated by spontaneous process, particles generated by linear growth and particles generated by catalytic process in the given (i,j) position of the lattice respectively.

R(i,j,2) is seeded with molecular numbers sampled from a binomial distribution with average concentration of the dead state(phi1)

R(i,j,1)= R(i,j,2)+ R(i,j,3)+ R(i,j,4)

Calculate concentration of the lattice s2=sum of R(:,:,1) and phi=s2/(3*L*L)

```
While (phi less than phi2)
{
        Loop( i running from 1 to L and j running from 1 to L )
        {
                n=R(i,j,1)
                sgr=s*(3-n)     // spontaneous
                lgr=r*n*(3-n)/2 //linear
                rgr=k*n*(n-1)*(3-n)/2  //catalytic
                dr=u*n   //death rate
                 hr=h*n   //hopping rate
```

```
Normalize srg,lrg,rgr,dr,hr and store them in itself
p1(1)=sgr
p1(2)=lgr
p1(3)=rgr
p1(4)=dr
p1(5)=hr

s1=sgr+lgr+rgr+dr+hr
p1=p1/s1
q1=cumulative summation of p1
z1=1
Choose a random number 'ran'
while(q1(z1)<ran)
{
        z1=z1+1;
}

Z1 gives us the reaction chosen by Gillespie algorithm and we perform the reactions
 if(z1==1)
{
R(i,j,2)=R(i,j,2)+1
R(i,j,1)=R(i,j,1)+1
 }
if(z1==2)
{
R(i,j,3)=R(i,j,3)+1
R(i,j,1)=R(i,j,1)+1
 }

if(z1==3)
{
R(i,j,4)=R(i,j,4)+1
R(i,j,1)=R(i,j,1)+1
```

```
 }
if(z1==4)
 {
 sam=[2 3 4] //in case of death choosing randomly which particle to delete

   while ()
 {
     x=choose a number randomly from  the array 'sam'

     if((R(i,j,x) not equal to 0)
      {
        R(i,j,x)=R(i,j,x)-1
        R(i,j,1)=R(i,j,1)-1

        break;
      }

  }

}

  If(z==5)
  {
          sam=[2 3 4]

          while 1
          {
                  x=randomly choose one element from 'sam'
                  if((R(i,j,x) not equal to 0))
                  {
                  break
                 }
          }
```

```
I1=1,j1=1
Loop( i1>1 and i1<L and j1>1 and j1<L)
{
        If( (i1,j1) not equal to (i,j))
        {
        Store R(i1,j1,1) in 'hopp'

        Put the element i1 to p2 and j1 to p3
        and keep doing that one after another.
        }
}

hopp=1-hopp/3                //probability of hopping


 select an element from hopp depending on the probability given by the
 by content of hopp ( the same way we did for Gillespie algorithm)

 if( R(p2(hoppto),p3(hoppto),1)~=3)
 {
    R(i,j,1)=R(i,j,1)-1;
    R(i,j,x)=R(i,j,x)-1;
    R(p2(hoppto),p3(hoppto),1)=R(p2(hoppto),p3(hoppto),1)+1;
    R(p2(hoppto),p3(hoppto),x)=R(p2(hoppto),p3(hoppto),x)+1;
  }
```

```
}
```

```
        }

        delta_t= choose a random number from exponential distribution with mean
        1/(srg+lgr+rgr+dr+hr)(Gillespie's algorithm)

        t=t+delta_t

        Calculate concentration of the lattice s2=sum of R(:,:,1) and phi=s2/(3*L*L)

}
```

After completion we plotted the R with different colors i.e the particles generated by spontaneous, catalytic, linear growth with different colors on the lattice. The time for various 'h' values are noted and plotted for the Tsys Vs h plot.

BIBLIOGRAPHY

[1] F. HAYOT AND C. JAYAPRAKASH, *A tutorial on cellular stochasticity and gillespie's algorithm (draft)*, (2006).

[2] D. KARIG, *Introduction to stochastic simulation with the gillespie method*, (2005).

[3] W. M AND H. PG, *Origin of self-replicating biopolymers: autocatalytic feedback can jump-start the rna world.*, J Mol Evol, 69 (2009), pp. 541–544.

[4] J. A. SHAY, C. HUYNH, AND P. G. HIGGS, *The origin and spread of a cooperative replicase in a prebiotic chemical system*, Journal of Theoritical Biology, 364 (2015), pp. 249–259.

[5] WU AND HIGGS, *The origin of life is a spatially localized stochastic transition.*, Biology Direct, 7 (2012), p. 42.