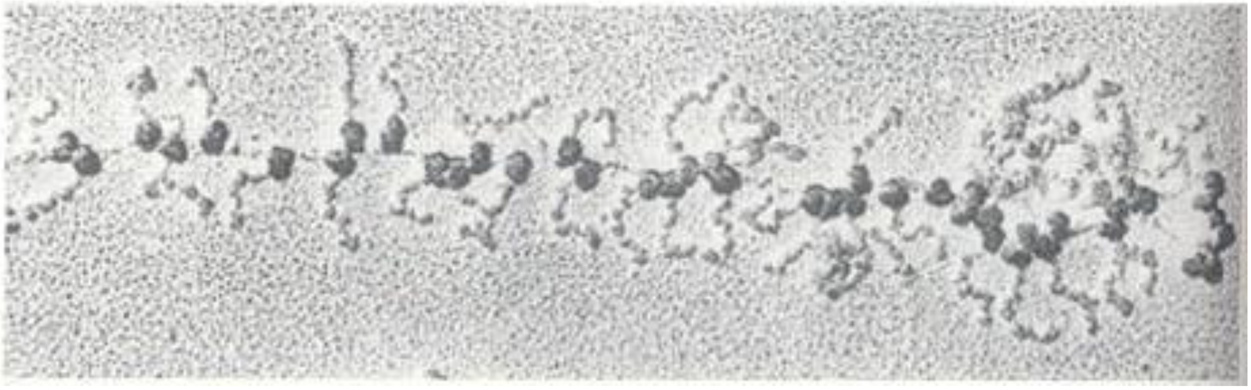
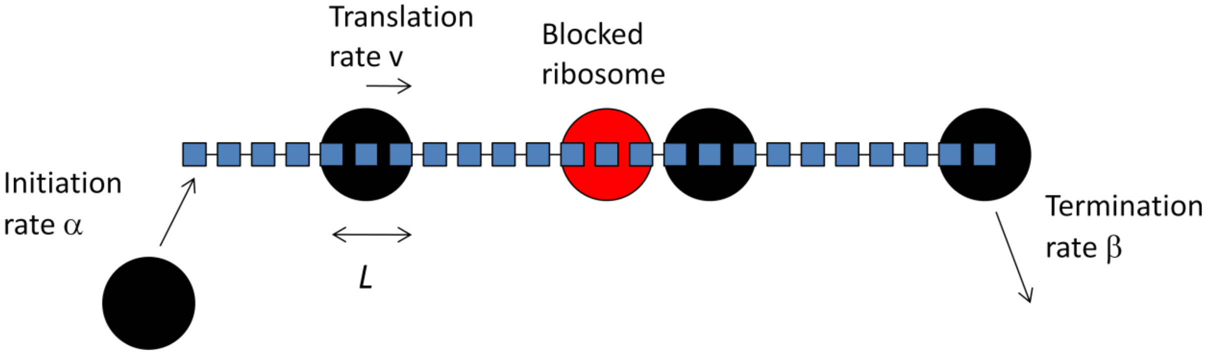
Project 5 – Ribosome Dynamics and the TASEP Model



Protein synthesis occurs when ribosomes move along mRNAs, one codon at a time, and translate the mRNA sequence into an amino acid sequence with the help of tRNAs. This EM picture shows many ribosomes moving along the same mRNA from left to right. The protein emerging from each ribosome is also visible (bigger on the right). If you are unfamiliar with how protein synthesis works, you could look up the background on Wikipedia or elsewhere - https://en.wikipedia.org/wiki/Protein_biosynthesis

When there are many ribosomes on the same mRNA, each one can only move if there is a space in front of it. Traffic jams occur, just like with cars on a highway. To model ribosome dynamics we will use a well-studied model in statistical physics which is known as the TASEP - totally asymmetric exclusion process. Mathematical details of this model have been carefully worked out (Shaw *et al.* 2003), and detailed applications of the model to translation of real genes have been made (Zia *et al.* 2011; Riba *et al.* 2019).



For our TASEP model of ribosome dynamics we will assume that ribosomes bind at one end of the mRNA (codon 1) and move along one codon at a time until they reach the other end (codon 300). Each ribosome has a size that covers L codons, so the distance of closest approach of two ribosomes is L. We will draw a ribosome as a circle of diameter L. In fact L = 11 codons for *E. coli* ribosomes.

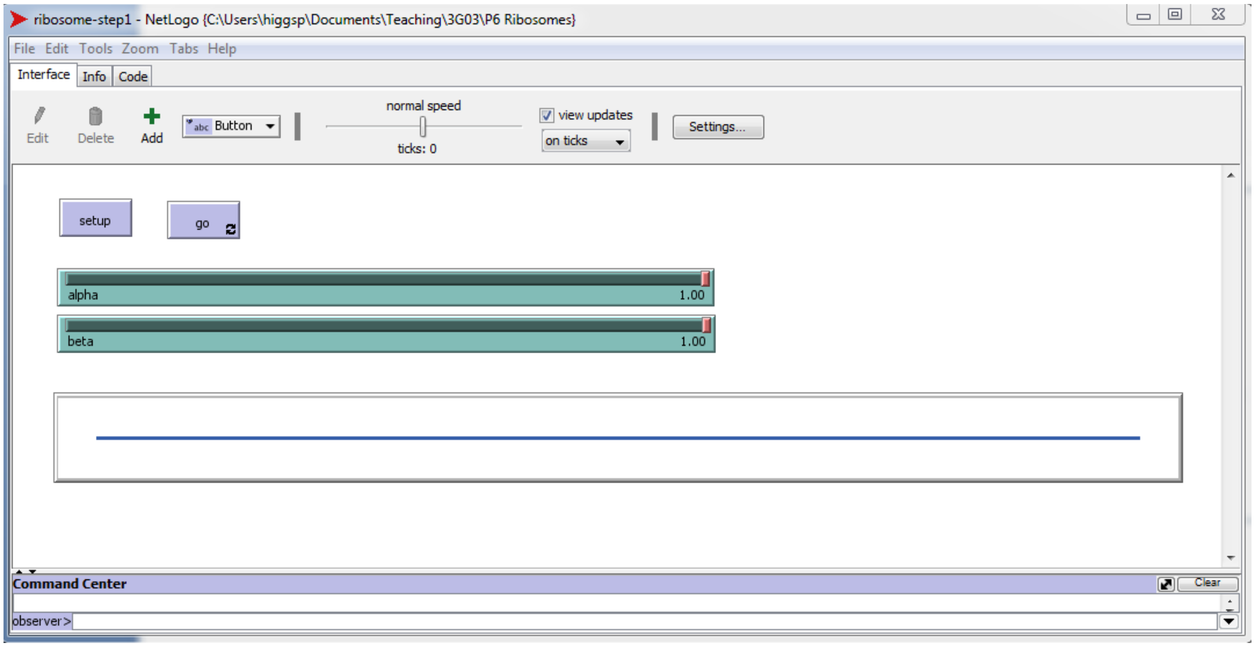
The binding of a ribosome at the first codon is called initiation of translation. In our model, this occurs at a rate α, but this can only occur if there is no previous ribosome blocking the first *L* codons. The detaching of a ribosome from the last codon is called termination. In our model, this occurs at a rate β.

The speed of translation is the rate of addition of one amino acid to the growing protein and the movement of a ribosome forward by one codon. This speed depends on the concentrations of tRNAs and is different for each codon in the genetic code. *E. coli* ribosomes can translate around 20 codons per second. In our model, however, we will suppose that all codons are the same, and we will set the speed per codon to be $v = 1$. The rates of initiation and termination, α and β, are measured relative to this. A ribosome can only move forward to the next site if it is not blocked by the one in front.

Step 1 Setup.

- Start a new Netlogo program from scratch. There is no initial program for this exercise. Make the world window of custom size going from -11 to 310 in the x direction and -11 to 11 in the y direction.
- Add a setup and a go button.
- Add a slider for the rates α and β going from 0 to 1. The speed per codon is always $v = 1$, so there is no need to have a slider for v.
- When the setup button is run, set all the sites to be white. Then set a line of sites representing the mRNA to be blue. These have coordinates pycor = 0 and 1 ≤ pxcor ≤ 300. This represents an mRNA of 300 codons.

It should look something like this:



alpha 1.00

beta 1.00

Step 2 Binding routine.

Define some global variables:

L (the ribosome size, i.e. the number of codons that is covered by one ribosome)

dt (the time step)

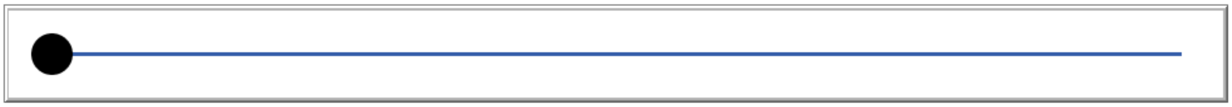
In the setup routine, let dt = 0.1, and L = 11 (which is correct for E. coli ribosomes).

Let patches-own a variable called speed. Set speed = 1 for all the codons except the last one (patches with pxcor from 1 to 299). For the last codon (pxcor = 300), set speed = β.

Write a routine called 'to bind' that creates a ribosome on patch 1. In this routine, create one turtle, and set its initial variables:

```
set xcor 1
set ycor 0
set shape "circle"
set color black
set size L
set heading 90 (that means it will move to the right when it moves)
```

Call the bind routine at the end of the setup routine, so that you begin with one ribosome on the first site. It should look like this:



Step 3 Dynamics

Now add the dynamics of the ribosomes. The structure of the code in the 'to go' routine will look like this

```
to go
  ask turtles [
    if xcor < 300 [
      ;code for moving ribosomes forward one codon
    ]
    if xcor = 300 [
      ;code for removing the ribosome from the last site
    ]
  ]
  ;code for binding a new ribosome at the beginning
  tick
end
```

You will find it useful to use `any?` - for example:

```
      if (any? turtles with [...this property...]) [...do this...]
      if (not any? turtles with [...this property...]) [...do this...]
```

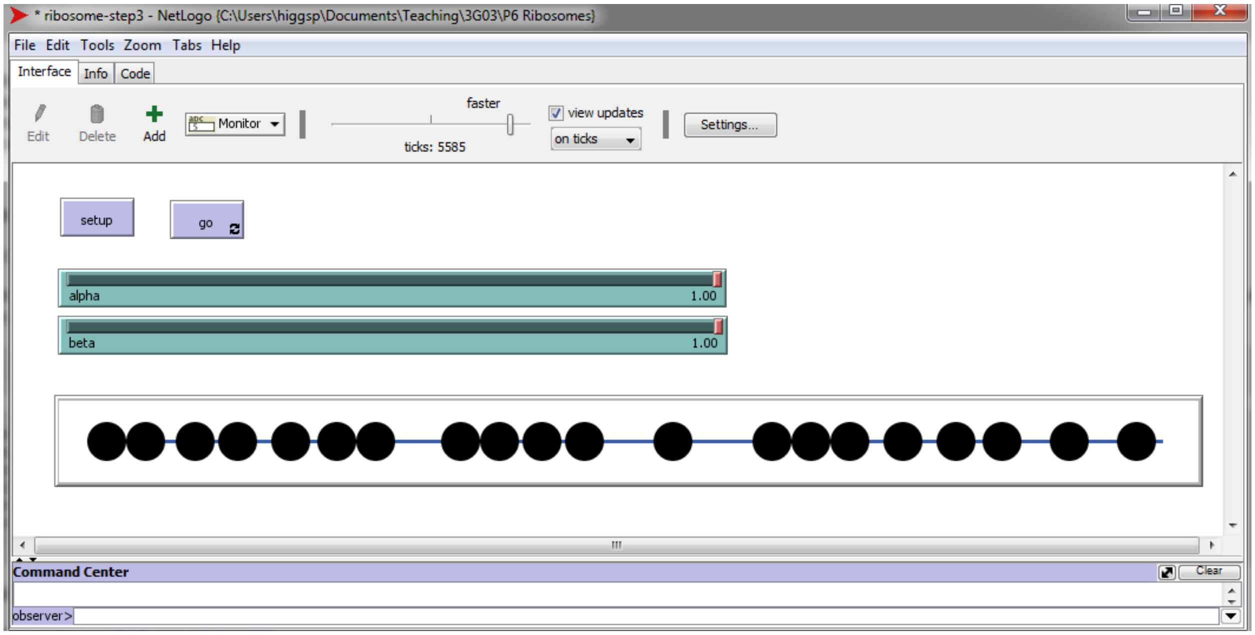
Look up `any?` in the Netlogo help pages.

The speed per codon is $v = 1$; therefore, the probability of moving forward one codon in one time step is δt. But this can only happen if the turtle is not blocked by the one in front. A turtle with `xcor = x` can only move if there is not any turtle with `xcor = x + L`. Remember that the ribosomes are not allowed to get closer than L, so there should never be any other ribosomes with `x < xcor < x + L.`

If there is a ribosome on the last site, its probability of detaching from the end is βδt. There is nothing that can block the last one. If it does detach then it disappears from the model, *i.e.* the turtle dies.

The binding probability of a new ribosome at site 1 is $\alpha\delta t$. This can only happen if the first *L* codons are free, *i.e.* if there are not any turtles with `xcor ≤ L`, then bind a new ribosome with probability $\alpha\delta t$. Call the bind routine that you already wrote.

Now when you click go you should see the ribosomes moving like this:



* ribosome-step3 - NetLogo {C:\Users\higgsp\Documents\Teaching\3G03\P6 Ribosomes}

ticks: 5585

☑ view updates

on ticks

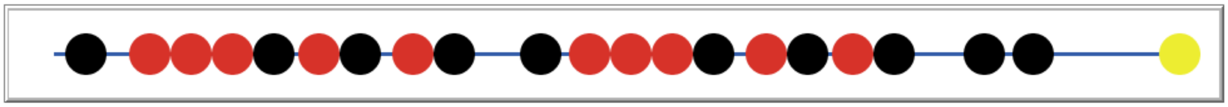
alpha 1.00

beta 1.00

Step 4 Colouring the ribosomes.

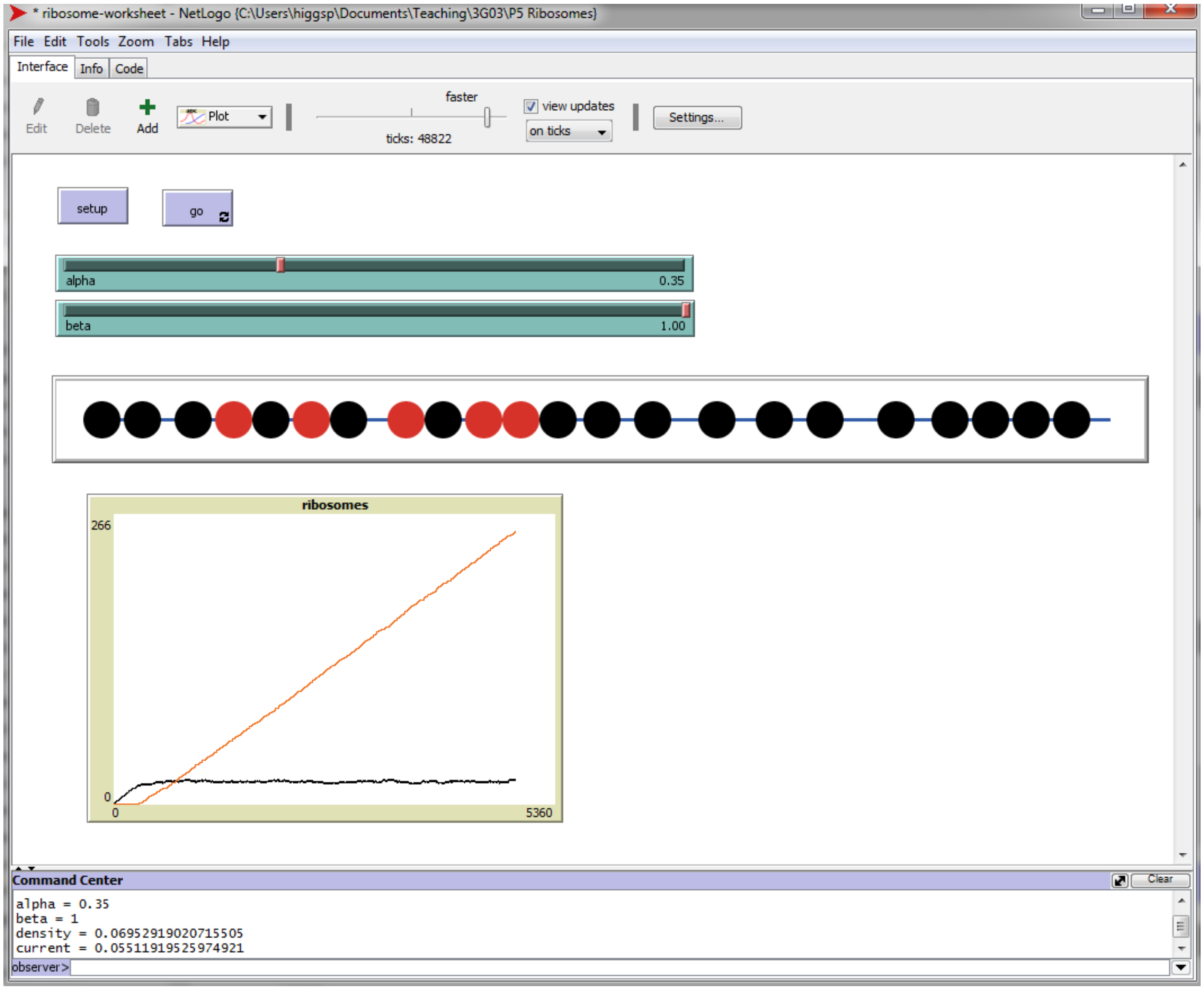
At the end of the go routine before the tick, set the colour of the ribosomes to indicate whether they are blocked.

- black if not blocked
- red if blocked
- yellow if on the last site



Step 5 Measuring the density and current of ribosomes

Add a variable `nbound` that counts the number of ribosomes currently bound to the mRNA. This is simply 'count turtles'. Add a variable `ncomplete` that counts the number of ribosomes that have completed synthesis. This can be set to 0 in the setup, and can be incremented by 1 each time a ribosome leaves the last codon. Add a plot that shows `nbound` and `ncomplete` as a function of time. It should look something like this.



alpha 0.35

beta 1.00

```
alpha = 0.35
beta = 1
density = 0.06952919020715505
current = 0.05511919525974921
observer>
```

The variables that we want to measure from this model are the density of ribosomes on the mRNA and the current of ribosomes along the mRNA. The density at any given moment is the number of ribsomes per unit length of mRNA: `nbound /300` . We want to measure the average of the density over a long time. The current is the number of ribosomes per unit time moving past any point on the mRNA. The current is the same at every point when averaged over time. The easiest way to measure the current is to count the number of ribosomes that leave the end of the mRNA and to divide by the time taken.

As we start with only one ribosome, the initial part of the simulation is not yet at a steady state. We want to measure the average current and density in the steady state, therefore we need to ignore the beginning of the simulation. We will start measuring when `ncomplete` = 50 and stop measuring when `ncomplete` = 250. Thus there will be 200 ribosomes that reach completion during the measurement period. To do this, define two variables `tsteps` and `density-sum` and set them both to 0 in the setup routine. Then add the following piece of code to the end of the go routine.

```
if ncomplete >= 50 [
  set tsteps tsteps + 1
  set density-sum density-sum + nbound / 300
]

if ncomplete = 250 [
  set density density-sum / tsteps
  set current 200 / (dt * tsteps)
  print " "
  type "alpha = " print alpha
  type "beta = " print beta
  type "density = " print density
  type "current = " print current
  stop
]
```

This says that we start counting when `ncomplete` reaches 50. `tsteps` counts the number of time steps since that point, and `density-sum` sums the densities at all times since that point. When we reach 250, the average density over the measurement period is `density-sum/tsteps,` and the average current during the measurement is `200 / (tsteps * dt)`. After this we print out the important variables and stop the simulation.

Results From The Program

Case 1: Set $\beta = 1$. Measure the current J and the density ρ for many values of α in the range 0 to 0.5. Plot graphs of J and ρ as a function of α across this range. The theory of Shaw et al. (2003) described below predicts that there is a critical value of α at which a transition between two types of behaviour occurs. The critical value is

$$\alpha_c = \frac{1}{1 + \sqrt{L}}$$

As, we have $L = 11$, α_c= 0.232. Therefore, include this value of α in your measurements. It is expected that the current increases with α for low values of α, but does not increase any further once $\alpha > \alpha_c$.

Case 2: Set β = 0.1. Measure J and ρ for many values of α in the range 0 to 0.5. Plot graphs of J and ρ as a function of α across this range. In this case, the theory predicts that there is a critical value of α = β at which a different type of transition occurs. Therefore, include the value of α = β = 0.1 in your measurements.

Comparison with theory

Shaw et al (2003) have shown that there are three possible phases of this model according to whether initiation, translation, or termination is rate limiting. When initiation is rate limiting, we are in the low density phase. The ribosomes are spread out because there is a long time between initiation of successive ribosomes. In the low density phase the current and density are:

$$J_{ld} = \frac{\alpha(1-\alpha)}{1+(L-1)\alpha}$$

$$\rho_{ld} = \frac{\alpha}{1+(L-1)\alpha}$$

You can look at the Shaw paper for interest, but don't worry if you don't understand the details. This is a fairly complex piece of statistical physics. Note that in the Shaw paper, the density is defined as the fraction of codons that are covered by ribosomes, whereas we have defined it as the number of ribosomes per unit length. These definitions differ by a factor of L. Therefore the equations in the paper look slightly different from the ones in this sheet.

If the initiation rate increases, we reach a point where translation becomes rate limiting, which is called the maximum current phase. At this point there is a traffic jam caused by the high density of traffic (like rush hour). The transition from low density to maximum current phase occurs at the critical value $\alpha = \alpha_c$ given above. For $\alpha > \alpha_c$ the current is constant and does not increase any more, even if α is increased. The current and density in the maximum current phase are:

$$J_{mc} = \frac{1}{\left(1+\sqrt{L}\right)^2}$$

$$\rho_{mc} = \frac{1}{\sqrt{L}(1+\sqrt{L})}$$

When termination is rate limiting, we are in the high density phase. There is a traffic jam with a high density caused by the single slow point (like road works). A transition from low density phase to high density phase occurs when $\alpha = \beta$, in the case where $\beta < \alpha_c$. The current and density in the high density phase are:

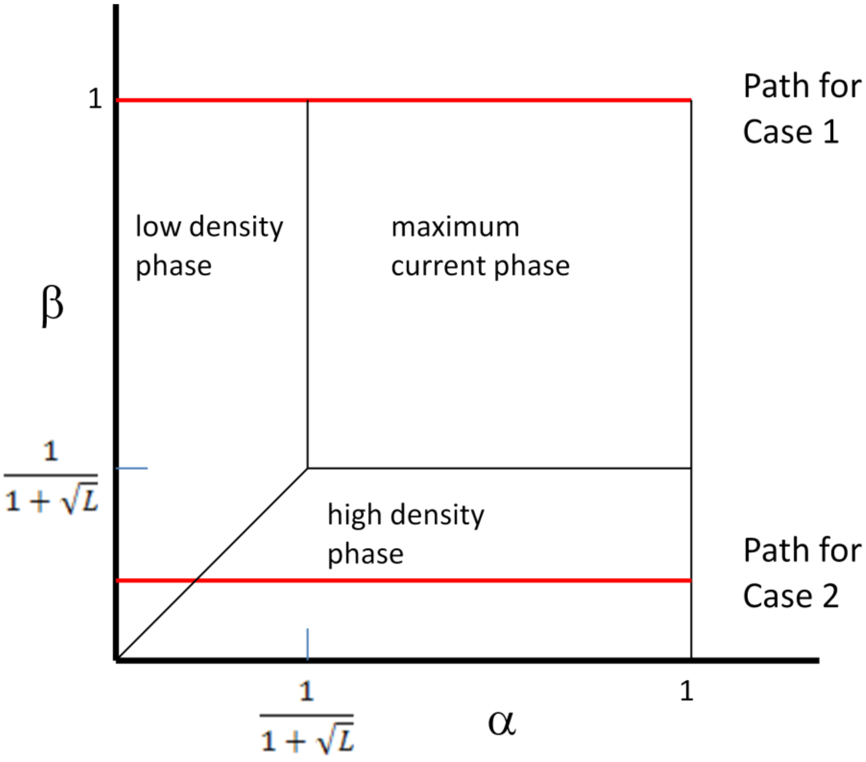
$$J_{hd} = \frac{\beta(1-\beta)}{1+(L-1)\beta}$$

$$\rho_{hd} = \frac{1-\beta}{L}$$

Plot the theory curves for J and ρ as a function of α on the same graphs as the measured values.

- For Case 1, use the low density formulae for $\alpha < \alpha_c$, and the maximum current formulae for $\alpha > \alpha_c$.

- For Case 2, use the low density formulae for $\alpha \le \beta$, and the high density formulae for $\alpha > \beta$. There is a jump in the density at $\alpha = \beta$ in this case (as shown in Fig 7 of Shaw et al).



The phase diagram for the TASEP model is shown here. The red lines show the paths in the phase diagram that we used for cases 1 and 2.

For your Report:
(a) **40 marks -** When you submit the report for this assignment, include a working version of your netlogo program showing that it successfully carries out all the steps above.
(b) **20 marks -** Include the graphs that plot the theory curves on top of the simulation results. Comment on whether the theory fits the simulation well or not. If there are differences, is it because of statistical error or because there are systematic differences between the measurement and the theory? If you want to reduce the statistical error, you could run the program for longer to get a better time average, or you could run it several times for the same parameter and average the results.
(c) **20 marks -** In the β = 0.1 case, you will probably find that the density increases fairly smoothly with α, and that the predicted jump in the curve at α = 0.1 is not very obvious. The reason for this is that there are not very many ribosomes on the mRNA at the same time. The max possible number is 300/11 = 27. Phase transitions are always rounded out when the number of particles is small. To consider a case where the theory is more convincing, we need more particles. We could either increase the length of the mRNA or decrease the size of the ribosomes.

Neither of these is biologically realistic, but we are checking the physics theory here, so don't worry about it. It will be comforting to have a good physics theory next time you are stuck on the QEW.

Repeat the analysis of the β = 0.1 case with the same length 300, but smaller particles of size L = 3, in which case the maximum possible number of particles is 100. You may find that the system takes quite a long time to come to a steady state. Measuring the average between 50 and 250 completed particles is probably not very good, because the system is still not in a steady state after 50 have completed. Change the program so that you start the measurement after nstart particles and stop after nstop particles, where nstop = nstart + nsample. How big does nstart have to be to ensure that you are in a steady state? How big does nsample have to be to ensure that you have a good average during the measurement period? In your report, say which values you tried and how you decided which values to use for your measurement. Show the graph of simulation and theory for this case. You will probably find that the transition occurs much more obviously in this case, and that the theoretical prediction is more convincing.

References

Riba, A., Di Nanni, N., Mittal, N., Arhne, E., Schmidt, A. and Zavolan, M. (2019) Protein synthesis rates and ribosome occupancies reveal determinants of translation elongation rates. *PNAS* 116: 15023-15032.

Shaw, L.B., Zia, R.K.P. and Lee, K.H. (2003) Totally asymmetric exclusion process with extended objects: A model for protein synthesis. *Physical Review E* 68, 021910

Zia, R.K.P., Dong, J.J. and Schmittmann, B. (2011) Modeling translation in protein synthesis with TASEP: a tutorial and recent developments. *J. Stat. Phys.* 144: 405-428.