Transcriptional Bursting in Gene Expression

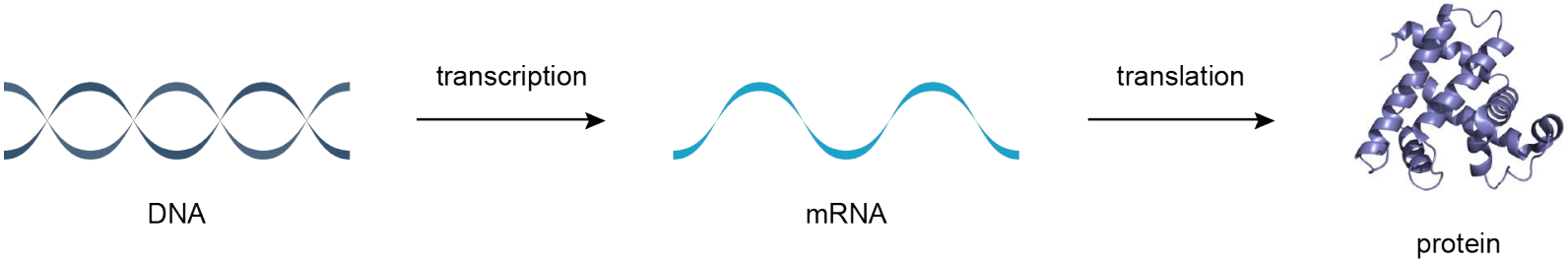
PH:549 Physics of Biological System

Aditya Sharma

Transcriptional Bursting in Gene Expression: Analytical Results for General Stochastic Models

by: Niraj Kumar[1], Abhyudai Singh[2], Rahul V. Kulkarni[1]*

Transcription



Even though the Central Dogma seems so straightforward, it is not. Here we only look on the structure of the process and sequential steps followed. But each of the individual steps in itself is very complex.
The process of gene expression in itself has inherent randomness which can be modelled.

Introduction and motive

It has been already known and well demonstrated that transcription in single cells is sporadic and the mRNA synthesis often occurs in bursts followed by periods of inactivity. This can give rise to high variability in gene expression and phenotypic variation in cells having same genetic makeup. Therefore there is a sufficient amount of motivation in quantifying various parameters like frequency and burst size for proper understanding.

What can we do to quantify? How to model them? And in what form can we obtain data?

1. Steady state measurements of mean and variance.
2. Obtain time lapse measurements.
3. Model general class of stochastic processes and match higher moments.

Modelling arrival process for mRNA creation

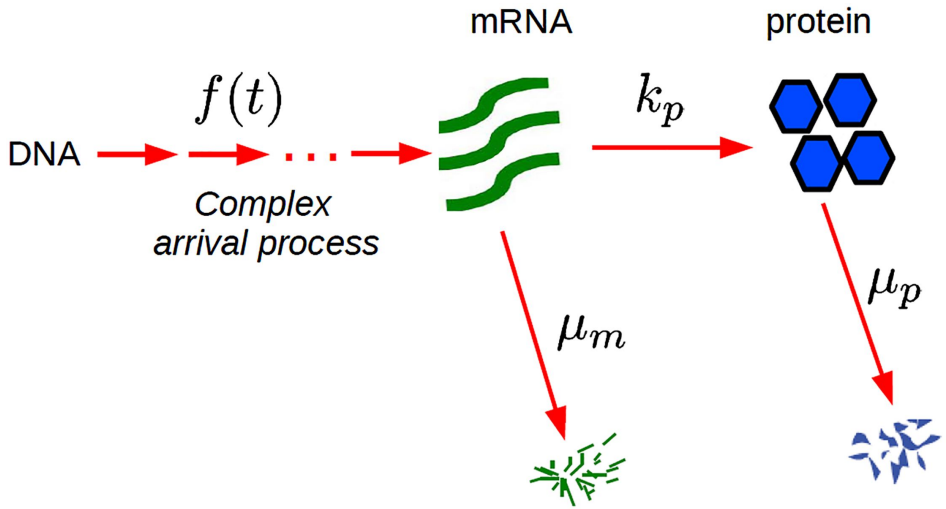
Terminology and assumptions

$f(t) \rightarrow$ General Arrival Time Distribution

$k_p \rightarrow$ Protein Production Rate

$\mu_p \rightarrow$ Protein Decay Rate

$\mu_m \rightarrow$ mRNA Decay Rate



Mathematical Description and Background

$$a^p(z) = \sum_{n=0}^{\infty} z^n p(n)$$

$a^p(z) \rightarrow$ Generating function of Protein burst distribution $p(n)$ by a single mRNA

$$A^p(z) = \sum_{n=0}^{\infty} z^n P(n)$$

$A^p(z) \rightarrow$ Generating function of Protein burst distribution $P(n)$ by a all the mRNAs in the burst

$$A^p(z) = A^m[a^p(z)].$$

$A^m(z) \rightarrow$ Generating function of mRNA burst distribution

$$a^p(z) = \frac{1}{[1 + \langle p_b \rangle (1-z)]}$$

$$\langle p_b \rangle = k_p/\mu_p$$

Mean Protein Burst Size

At least 1 mRNA

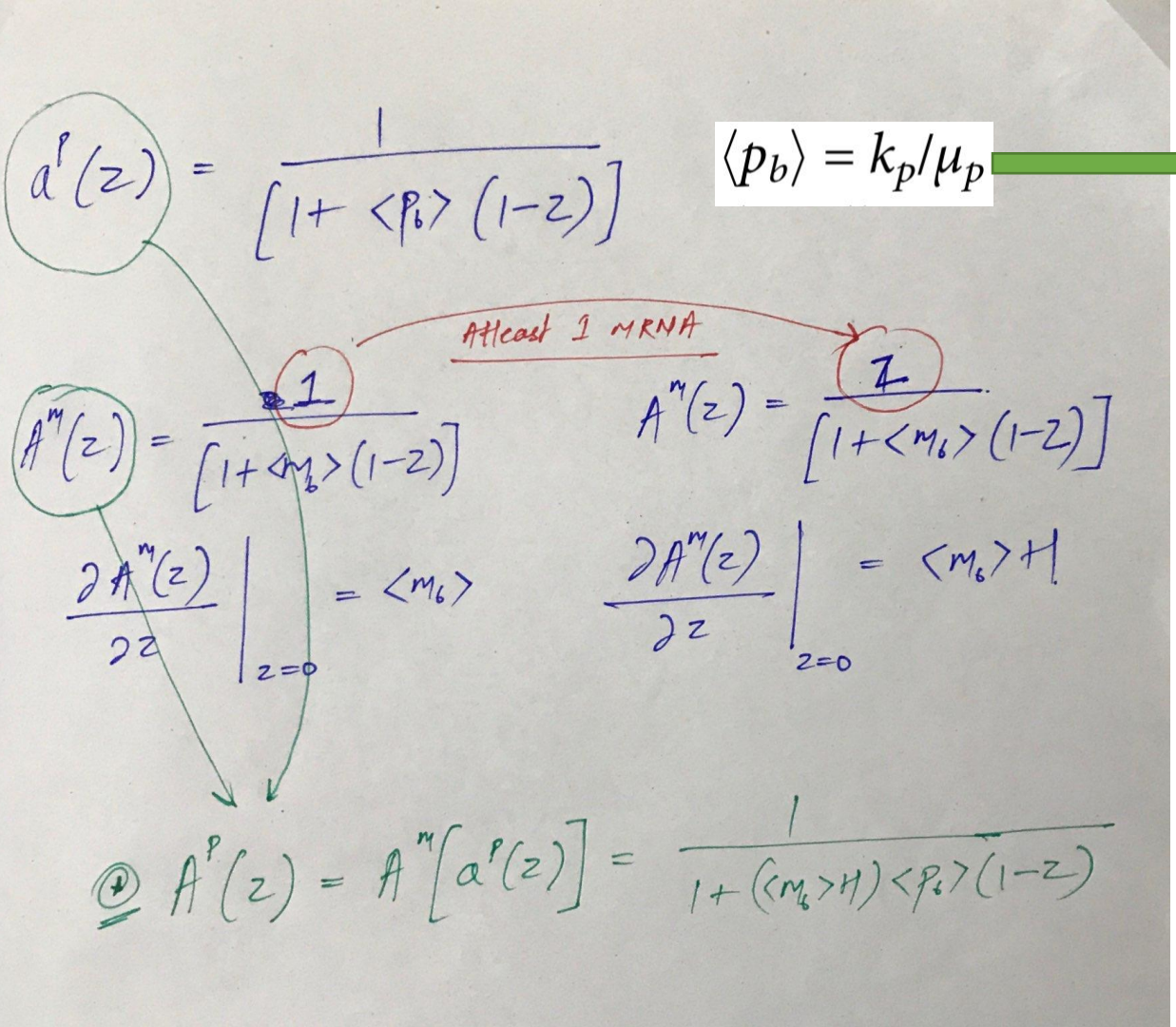
$$A^m(z) = \frac{1}{[1 + \langle m_b \rangle (1-z)]}$$

$$A^m(z) = \frac{z}{[1 + \langle m_b \rangle (1-z)]}$$

$$\left.\frac{\partial A^m(z)}{\partial z}\right|_{z=0} = \langle m_b \rangle$$

$$\left.\frac{\partial A^m(z)}{\partial z}\right|_{z=0} = \langle m_b \rangle + 1$$

Note that this can model Poisson when <mb> tends to 0 and Geometric when <mb> > 0


$a^p(z) = \frac{1}{[1+\langle p_b\rangle(1-z)]}$
$\langle p_b\rangle = k_p/\mu_p$
Mean Protein Burst Size
Atleast 1 mRNA
$A^m(z) = \frac{1}{[1+\langle m_b\rangle(1-z)]}$
$A^m(z) = \frac{1}{[1+\langle m_b\rangle(1-z)]}$
$\frac{\partial A^m(z)}{\partial z}\Big|_{z=0} = \langle m_b\rangle$
$\frac{\partial A^m(z)}{\partial z}\Big|_{z=0} = \langle m_b\rangle + 1$
$A^p(z) = A^m[a^p(z)] = \frac{1}{1+(\langle m_b\rangle+1)\langle p_b\rangle(1-z)}$

$$\langle m_s \rangle = \frac{k_b}{\mu_m} \langle m_b \rangle, \quad \langle p_s \rangle = \frac{k_b}{\mu_p} b,$$

$$\frac{b}{\langle m_b \rangle} = \frac{\mu_p}{\mu_m} \frac{\langle p_s \rangle}{\langle m_s \rangle}$$

Obtained by simple calculations

Experimental measurements of the first two moments of the steady-state distribution are sufficient to estimate the burst parameters, as has been done in multiple studies. However, when the arrival process is non-Poisson or if the burst distribution deviates from a geometric distribution, measurements of the first two steady-state moments are not sufficient for estimating the burst parameters.

Variance Term

$$\frac{\sigma^2_{m_s}}{\langle m_s \rangle^2} = \frac{1}{\langle m_s \rangle} + \frac{\mu_m}{k_b} + \frac{\mu_m}{2k_b} \left[K_g(\mu_m) - 1 + \frac{\sigma^2_{m_b}}{\langle m_b \rangle^2} - \left(1 + \frac{1}{\langle m_b \rangle} \right) \right],$$

$$\frac{\sigma^2_{p_s}}{\langle p_s \rangle^2} = \frac{1}{\langle p_s \rangle} + \frac{\mu_p}{k_b} + \frac{\mu_p}{2k_b} \left[K_g(\mu_p) - 1 + \frac{\sigma^2_{m_b}}{\langle m_b \rangle^2} - \left(1 - \frac{1}{\langle m_b \rangle} \right) + \left(\frac{\sigma^2_{p_b}}{\langle p_b \rangle^2} - \left(1 + \frac{1}{\langle p_b \rangle} \right) \right) \frac{1}{\langle m_b \rangle} \right]$$

n^{th} Moment $\Rightarrow \underline{\langle (N - \langle N \rangle)^n \rangle}$

For example: $\langle (N - \langle N \rangle)^2 \rangle = \langle N^2 \rangle - 2\langle N \rangle^2 + \langle N \rangle^2$

$$= \underline{\langle N^2 \rangle - \langle N \rangle^2}$$

$$\langle (N - \langle N \rangle)^3 \rangle = \langle N^3 \rangle - 3\langle N^2 \rangle \langle N \rangle + 3\langle N \rangle^3 + \langle N \rangle^3$$

$$= \underline{\langle N^3 \rangle - 3\langle N^2 \rangle \langle N \rangle + 4\langle N \rangle^3}$$

To calculate moments, we can modify generating functions :—

$$\langle N \rangle = \left. \frac{d A_p(z)}{d(z)} \right|_{z=1}$$

$$\langle N^2 \rangle = \left. \frac{d}{dz}\left(z \frac{d A_p(z)}{d(z)} \right) \right|_{z=1} = \left. \frac{d A_p(z)}{d(z)} \right|_{z=1} + z \left. \frac{d^2 A_p(z)}{dz} \right|_{z=1}$$

⋮

And so on.

Expression for the third Moment: skewness

$$\frac{\gamma_{m_s}\sigma_{m_s}^3}{m_s} = 1 + \langle m_s\rangle\langle m_b\rangle\mathcal{K}_1(\mu_m) + 2\langle m_b\rangle^2\mathcal{K}_2(\mu_m,\langle m_b\rangle) + (\sigma_{m_b}^2 + \langle m_b\rangle^2 - \langle m_b\rangle)\mathcal{K}_3(\mu_m,\langle m_b\rangle) + \frac{\langle m_b(m_b-1)(m_b-2)\rangle}{3\langle m_b\rangle},$$

$$\mathcal{K}_1(\mu_m) = K_g(2\mu_m) - K_g(\mu_m),$$

$$\mathcal{K}_2(\mu_m,\langle m_b\rangle) = \frac{K_g(\mu_m)-1}{4}\left(\frac{3}{\langle m_b\rangle} + K_g(2\mu_m) - 1\right),$$

$$\mathcal{K}_3(\mu_m,\langle m_b\rangle) = \frac{3}{2\langle m_b\rangle} + \frac{K_g(\mu_m)+K_g(2\mu_m)}{2} - 1.$$

Note that this is valid only for the burst limit $\mu_m \gg \mu_p$

This is a simplification made so that we avoid the details of the kinetic scheme for gene expressions.

$$\frac{\gamma_{p_s}\sigma_{p_s}^3}{p_s} = 1 + (A_1^p)^2\left[\frac{\langle p_s\rangle}{b}\mathcal{K}_1(\mu_p) + 2\mathcal{K}_2(\mu_p,A_1^p)\right] + A_2^p\mathcal{K}_3(\mu_p,A_1^p) + \frac{A_3^p}{3A_1^p},$$

$$A_1^p = \langle m_b\rangle\langle p_b\rangle,$$

$$A_2^p = \langle m_b\rangle(\sigma_{p_b}^2 - \langle p_b\rangle) + (\sigma_{m_b}^2 + \langle m_b\rangle^2)\langle p_b\rangle^2,$$

$$A_3^p = \langle p_b\rangle^3\langle m_b(m_b-1)(m_b-2)\rangle + 3\langle m_b(m_b-1)\rangle\langle p_b\rangle\langle p_b(p_b-1)\rangle + \langle m_b\rangle\langle p_b(p_b-1)(p_b-2)\rangle.$$

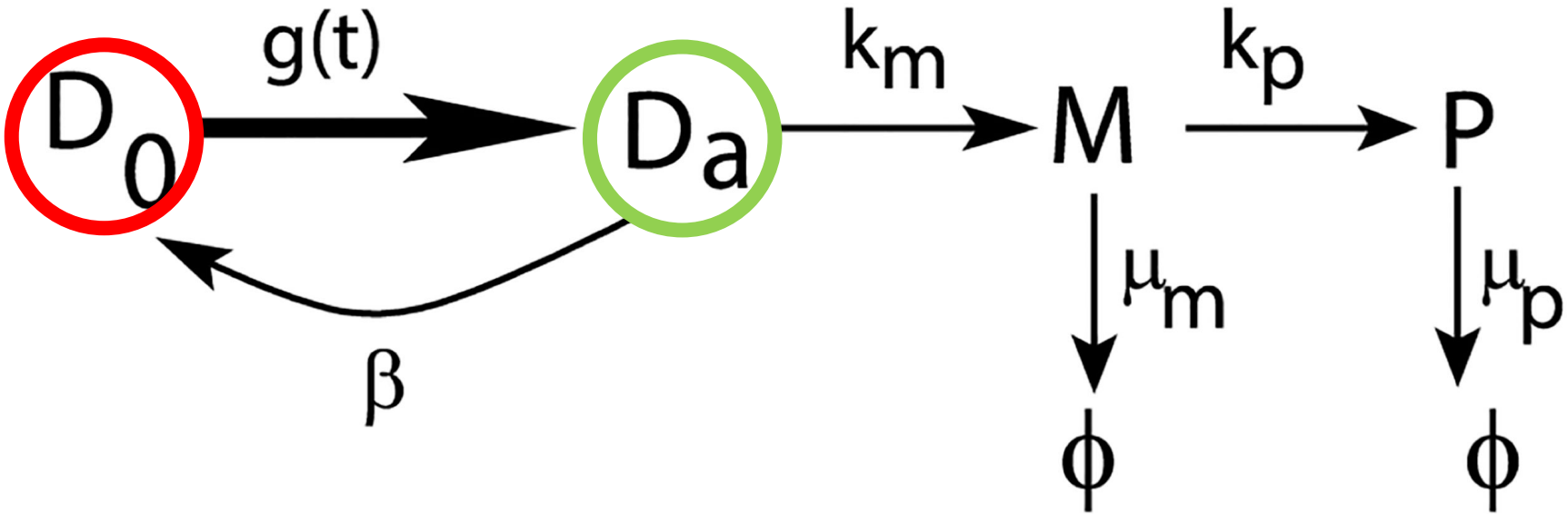
Problems with this approach

The results derived for the steady-state moments indicate that, if the burst arrival process is not a Poisson process, then it is no longer accurate to estimate burst parameters based on measurements of mean and variance only, as has been done in previous studies.

ON-OFF
State mechanism

Waiting time distribution

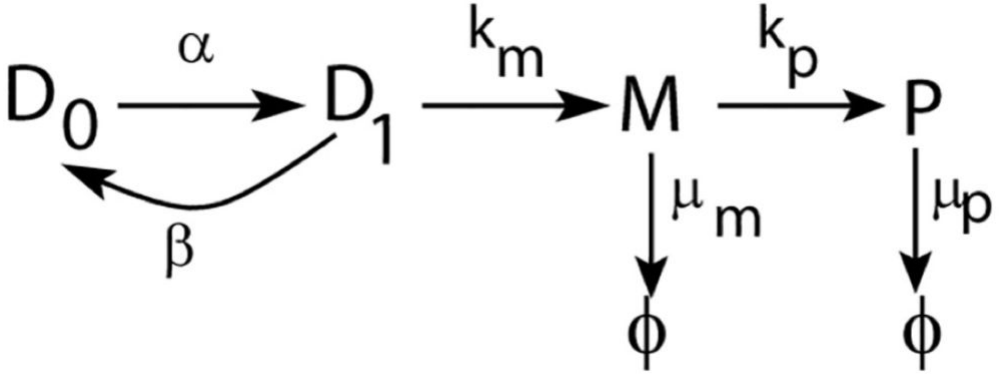
Burst arrival
Distribution

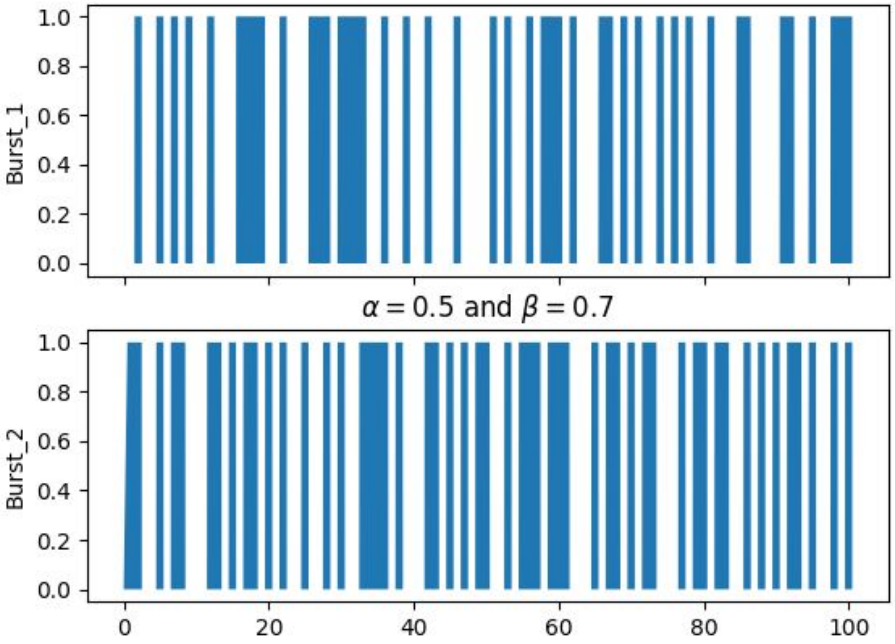


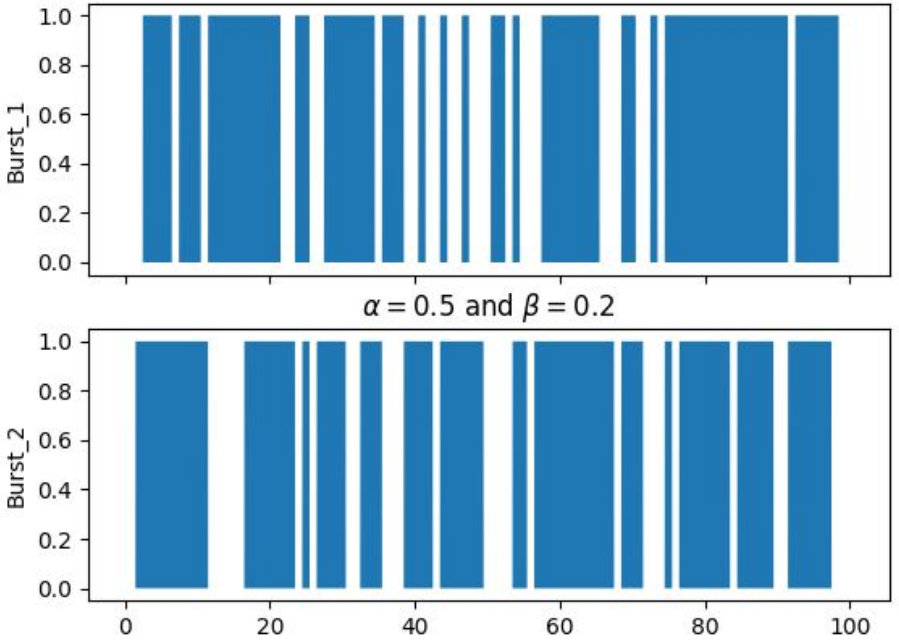
Iterative Procedure to follow:

1. Start with basic burst arriving model and use 3 moments to derive knowledge about waiting time distribution function.
2. If this consistently satisfies the other moments the you are done……
   Otherwise you need to modify burst arrival function.
3. Iterate again and again for more complex burst arrival functions.

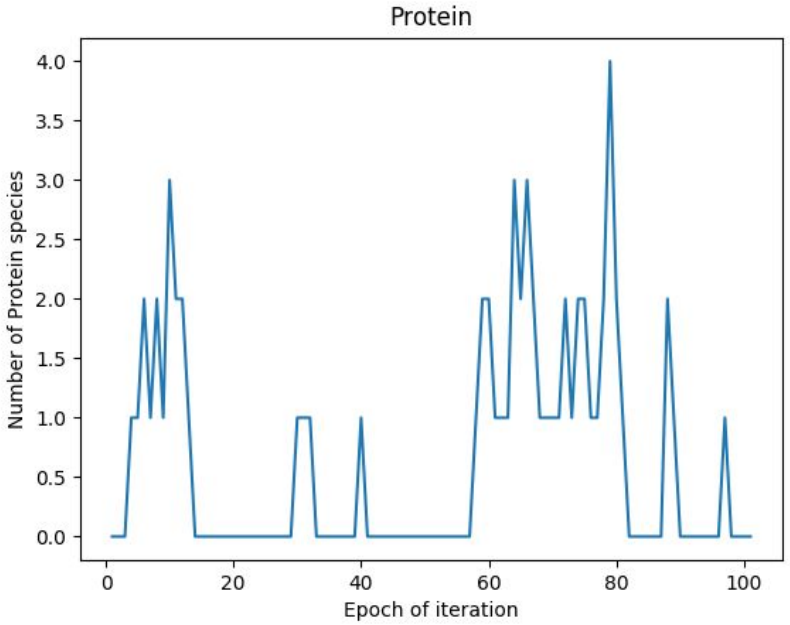
Some experimental simulations: (for a simplified version)



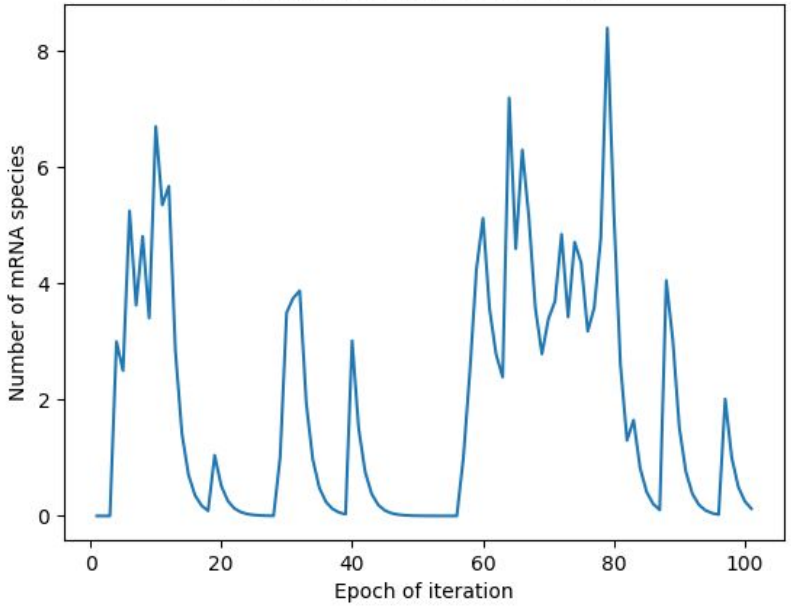


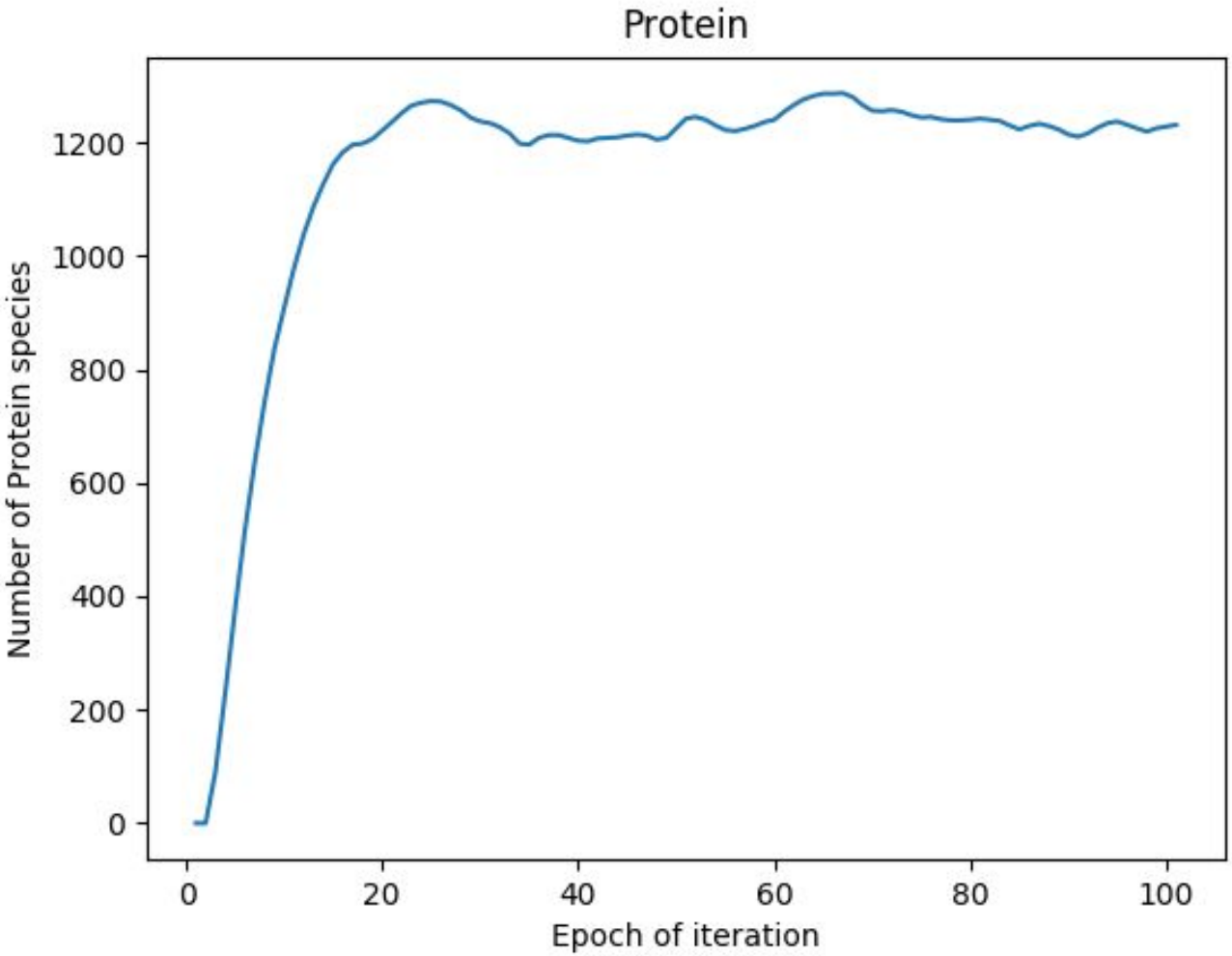


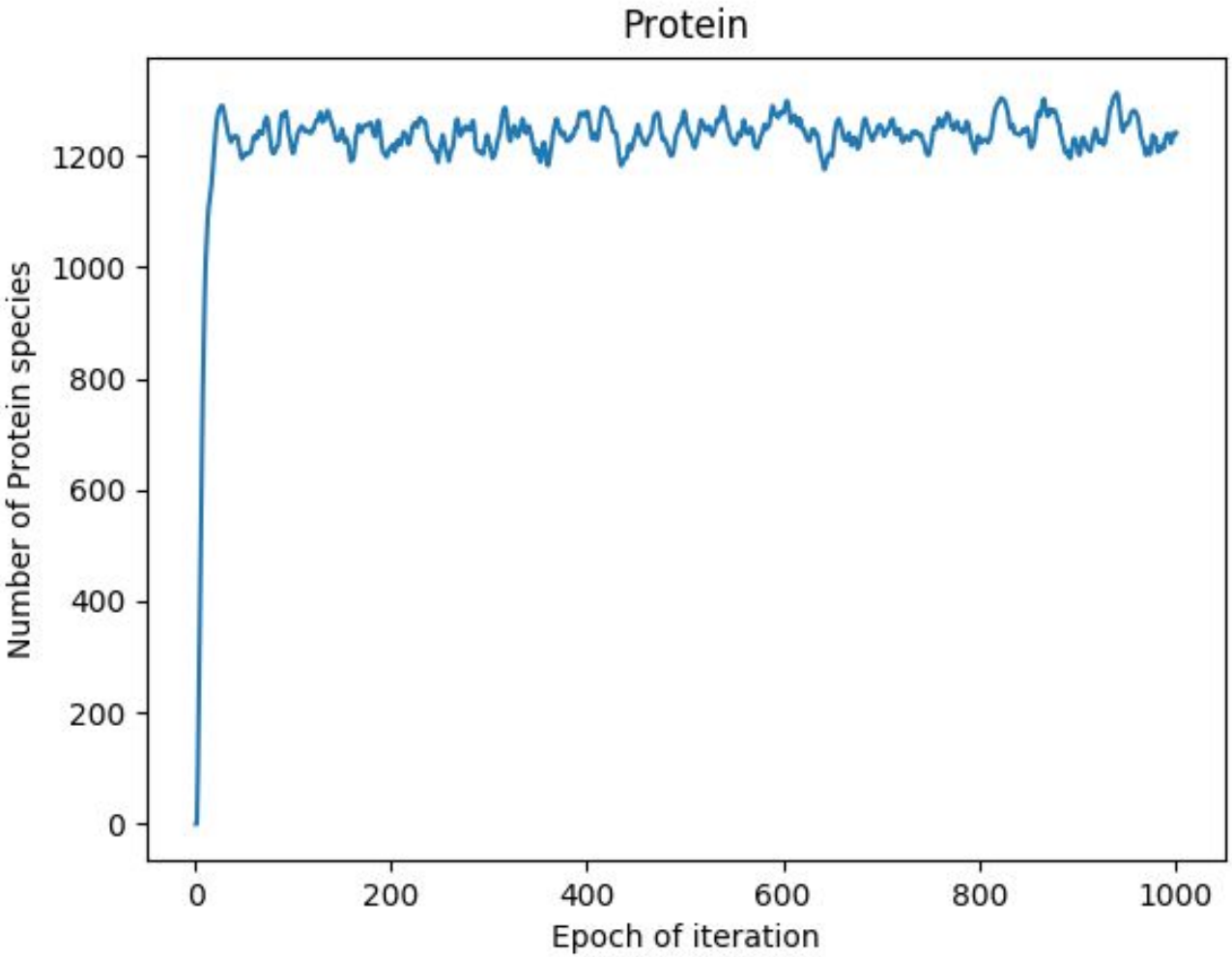
Effect of alpha and beta on the ON-OFF State distribution

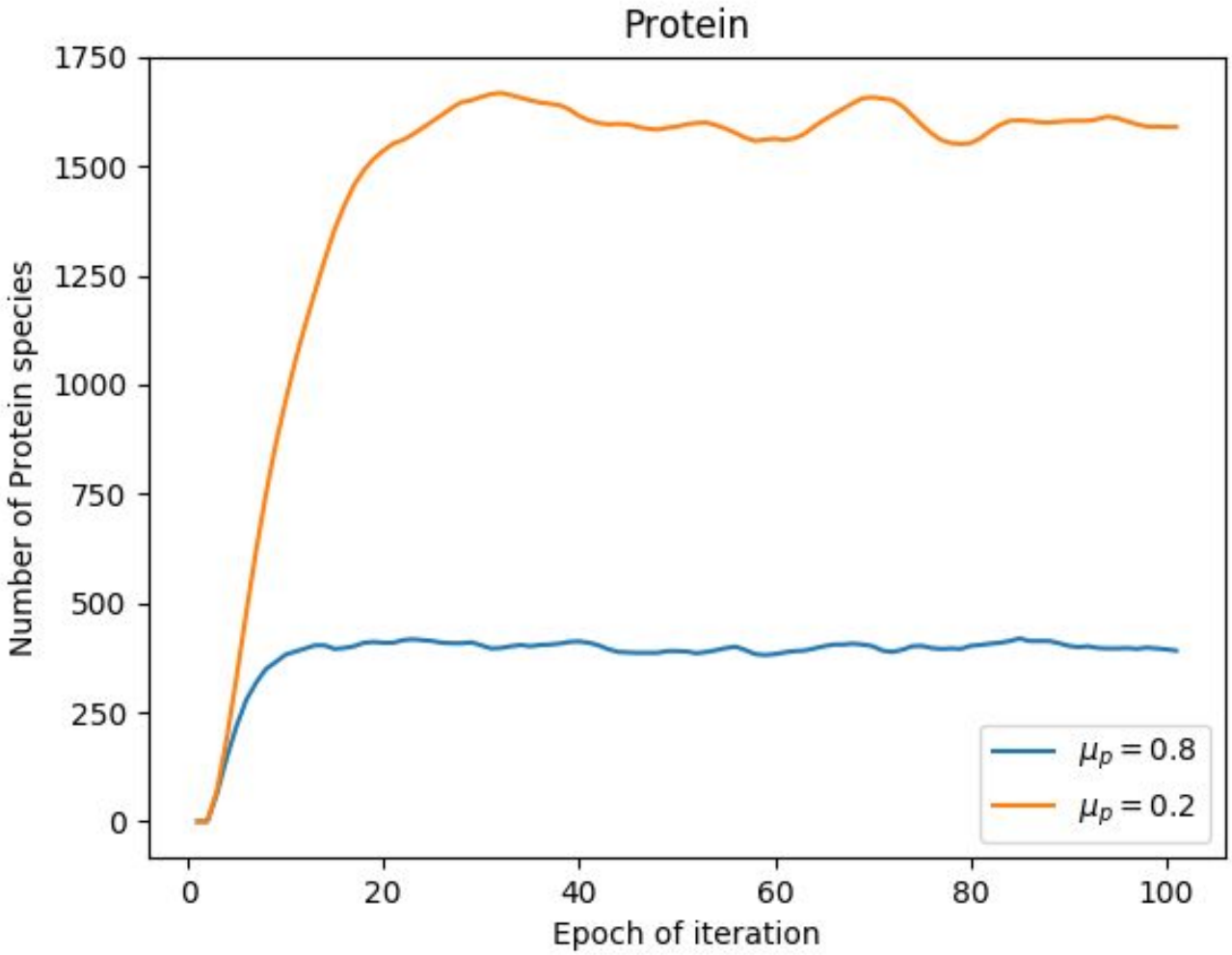


mRNA modelled as Geometric burst



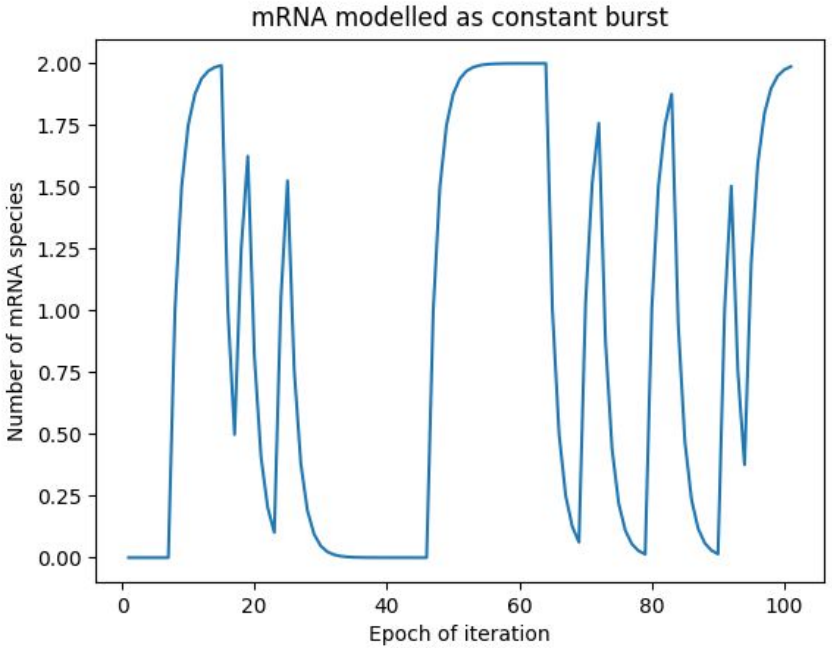

Protein
Number of Protein species
0
200
400
600
800
1000
1200
0
20
40
60
80
100
Epoch of iteration

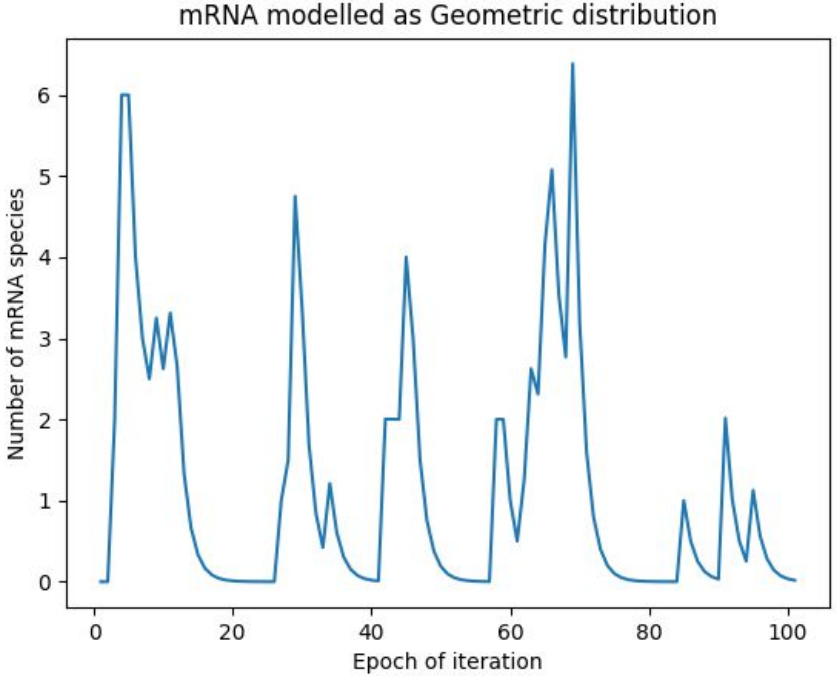

Protein
Number of Protein species
0
200
400
600
800
1000
1200
0
200
400
600
800
1000
Epoch of iteration



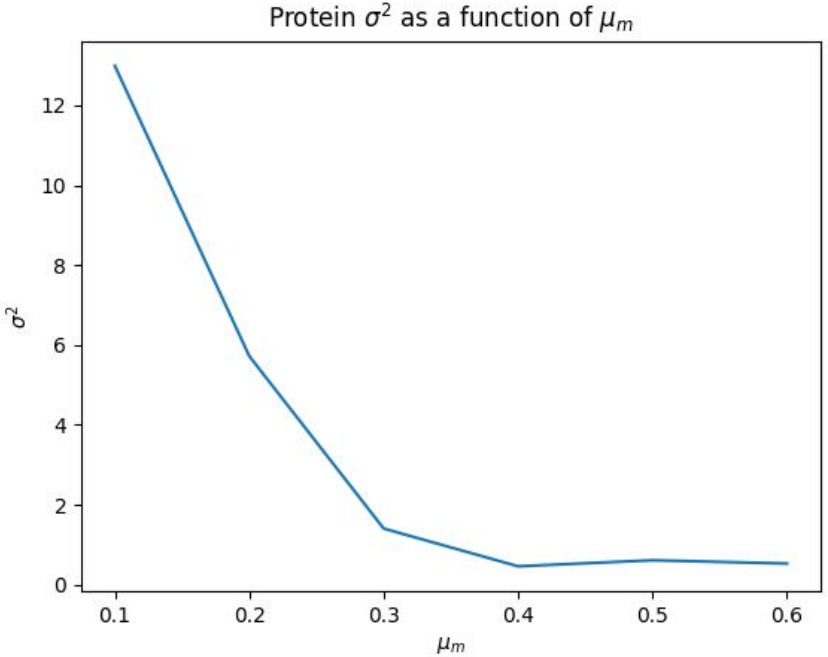
Effect of **μ**_p on the steady state Protein quantity

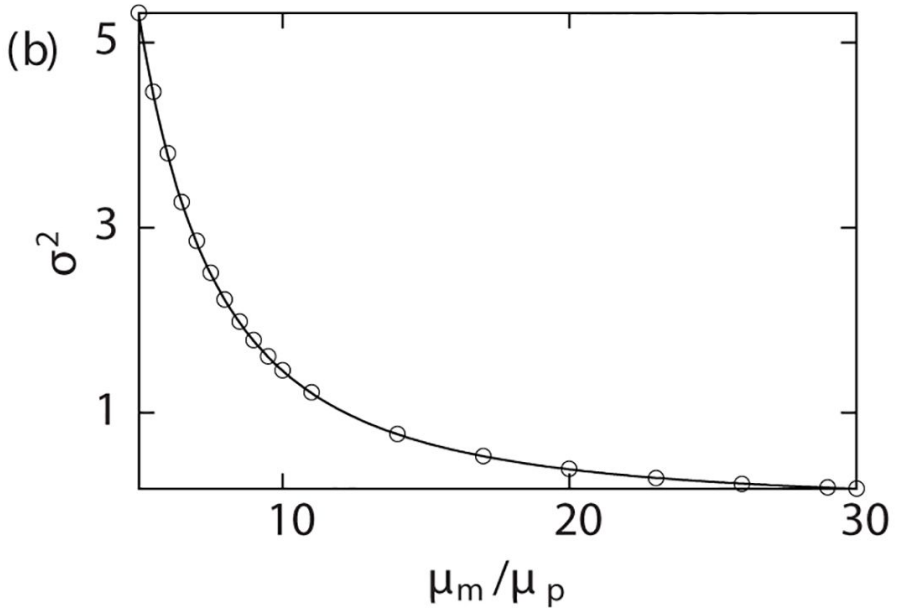
Modelling mRNA Burst as Geometric distribution





Second Moment Analysis





We see that both the graphs behave similarly.

Conclusion:

1. Time-waiting model is effective to a significant extent.
2. We can conclude upon the type of processes involved ar each step through some signatures that come in moments.

There are a lot of physical factors and biology which is involved which finds explanation in these models. However we see that these mathematical models are quite effective.

Thank you