A Stochastic Model for a General Bio-sensor

Rachael Fulton, Martina Marba

October 7, 2007

1 Introduction

The stochastic fluctuation in gene expression arises in one of two ways intrinsic and extrinsic noise. Here, intrinsic noise is the focus of interest: Cells are intrinsically noisy biochemical reactors; low reactant numbers can lead to significant statistical fluctuations in molecule numbers and reaction rates.[6] It has been brought to light that 'gene expression has a stochastic component due to the single molecule nature of the gene and the small number of copies of individual DNA-binding proteins in the cell'[7]. 'Noise is seen as being detrimental and volatile in different systems of interest. However, living systems are inherently noisy and are optimised to function in the presence of stochastic fluctuations'[6]. In this case the system of interest is a single cell or bacteria of a bacterial whole cell bio-sensor, noise is investigated on a cellular level using a stochastic modelling algorithm called Gillespie algorithm. 'A stochastic model is a tool for estimating probability distributions of potential outcomes by allowing random variation in one or more inputs over time'[1].It can give a better prediction of the behaviour of a system, especially on a single cellular level. It takes into account a number of parameters that contribute to the model in a random manner rather than assuming everything can be predicted deterministically.[2.1]

2 Aims

The aim of the stochastic approach to modelling the general bio-sensor was generally to provide, as accurately as possible, a prediction of the behaviour

of the system on a single cell level. This was achieved in a number of different ways:

- Find a simple model of the system that still accurately represents the behaviour of the full model but needs less data.
- Investigate the output and measure the noise with different values of signal.
- locate the parameters that the system is most sensitive to and investigate the effect these have on the value at which P4 saturates and the effect these have on the noise in the output.
- Explore the effect of positive feedback on the output of P4 and investigate the noise with and without positive feedback.
- Explore the effect of leakiness of Phz on the output of P4 and investigate the noise with and without leakiness.
- Consider the deterministic approach and run the stochastic simulation for a large number of cells to investigate whether it is equivalent to the deterministic approach.

3 Background Theory

3.1 Gillespie Algorithm

'The Gillespie algorithm allows a discrete and stochastic simulation of the system with few reactants because every reaction is individually simulated'. The aim of this algorithm is to draw two random numbers at each time step, one is to determine the next reaction index (choose which reaction will occur) and the second is to determine the time until the next reaction. It works using the number of molecules rather than the concentration. A concentration can be converted into the number of molecules by multiplying it by the volume of the system and Avogadro's constant. 'The quantity characterising each reaction is the probability $a_\mu(t)dt$ that, given the state of the system at time t, reaction μ will occur in volume V in the interval (t,t+dt). $a_\mu(t)$ is the product of two parts, the reaction rate c_μ and the number of possible reactions μ in volume V. This is called the propensity function for each reaction'.[3-5]

3.1.1 Implementation of the Algorithm

The implementation of the algorithm is as follows:[3][5]

Step 1: Initialise the number of molecules in the system, reactions constants and random number generators.

Step 2: Find the time τ after t at which the next reaction will take place, by drawing a random number from an exponential probability density function. This is known as the Monte Carlo Step.

Step 3: Choose at random the reaction which will occur at time t+ τ. Draw a random number from a uniform distribution between 0 and 1.

Step 4: Go back to step one and reiterate this process for as long as you want to follow the system.

3.2 Measuring noise

'Intrinsic noise is determined by the structure, reaction rates and species concentrations of the underlying biochemical networks'[6]. The relative size of noise can be measured in a number of ways. In this report, only two were investigated, the Fano factor and the coefficient of variation are used. The Fano factor is defined as:

$$Fano = \frac{\sigma^2}{\mu}, \tag{1}$$

where σ^2 is the variance and μ is the mean of the output. The Fano factor of an arbitrary stochastic system reveals deviations from Poissonian behaviour and it is a sensitive measure of noise. [6] There is also the coefficient of variation, which is defined as:

$$coeffvar = \frac{\sigma}{\mu}, \tag{2}$$

where σ is the standard deviation and, again, μ is the mean. This coefficient is a measure of dispersion of a probability distribution and is often used as another measure of noise[7]. The two are obviously conected via the ratio.

4 The Model

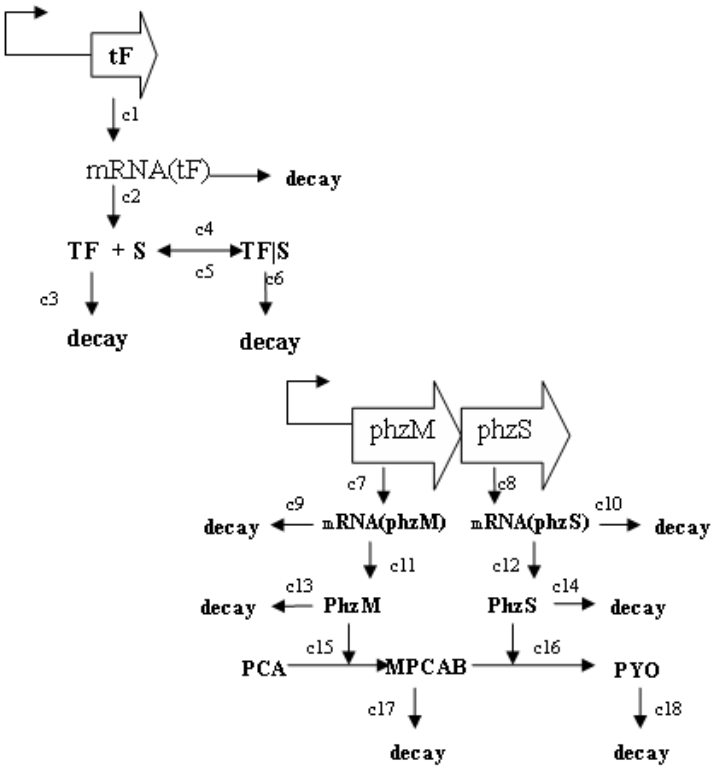
Instead of solving a full set of ODE's, Gillespie algorithm was used to get a probability distribution of the output over all possible states for each time t.

4.1 The Specific System

The specific system studied by the Glasgow IGEM team was used as a basis for the general model of a bio-sensor that is capable to detect a pollutant. The experemental team constructed two systems in parallel, one to detect xylene and another to detect salicylate (both referred to in the model as signal). 'In these systems, the pollutant forms a complex with a constitutively expressed transcription factor (TF), which in turn drives the expression of proteins allowing the bacteria to degrade the pollutant. By replacing the genes that encode these proteins with a reporter gene, an easily detectable output signal, such as pH or colour change is produced, thus creating a bio-sensor'[9]. The two systems the experimental team constructed were as folllows:the first.

4.2 Full Model

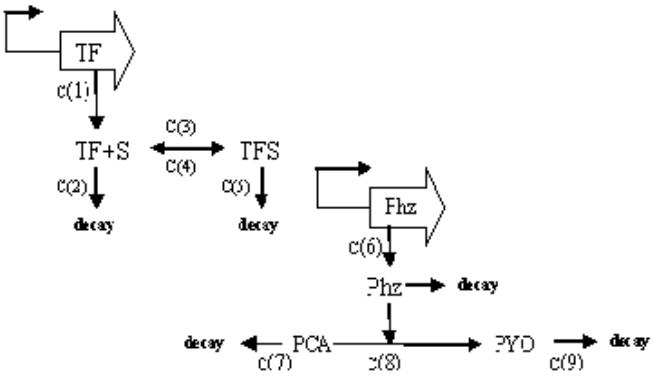
Initially a full model for the specific biobrick system the Glasgow IGEM team was studying,was derived. This could be used for the stochastic approach and ODE's were derived for the deterministic approach. The full model included each of the mRNA binding sites as well as each of the transcription and translation reactions. The full model can be illustrated on figure 1. This model could have been extended further, for example by considering each step binding and unbinding. However, this would make the model completely unusable as there are no available rate constants for most of the stages. As is seen from the diagram below, there is already a great deal of rate constants needed ($c(1) - c(18)$).



Due to the lack of available data and information, the majority of the rate constants are very unlikely to be found. Instead the model was simplified to a point at which it was still a realistic approximation of the complete biological system.

4.3 Simplified Model

The simplified model for a general bio-sensor is as follows. With TF being the transcription factor, S being the signal, TFS the complex of the two, and P3 and P4 the two proteins formed in the cascade. In the specific case studied P4 is the measurable output Pyocyanin. Below is a diagram illustrating the simplified model and a table of rate constants:



reaction	rate constant	propensity function
$\phi \to TF$	$\alpha = c(1)$	$a(1) = c(1)$
$TF \to \phi$	$\delta_{TF} = c(2)$	$a(2) = c(2) * X(1)$
$TF + S \to TFS$	$K1 * S = c(3)$	$a(3) = c(3) * X(1)$
$TFS \to TF + S$	$Km1 = c(4)$	$a(4) = c(4) * X(2)$
$TFS \to \phi$	$\delta_{TFS} = c(5)$	$a(5) = c(5) * X(2)$
$\phi \to P3$	$\frac{\beta * TFS}{\gamma + TFS} = c(6)$	$a(6) = c(6)$
$P3 \to \phi$	$\delta_{P3} = c(7)$	$a(7) = c(7) * X(3)$
$P3 \to P4$	$\alpha_2 = c(8)$	$a(8) = c(8) * X(3)$
$P4 \to \phi$	$\delta_{P4} = c(9)$	$a(9) = c(9) * X(4)$

5 Method

5.1 Simulating the Bio-sensor

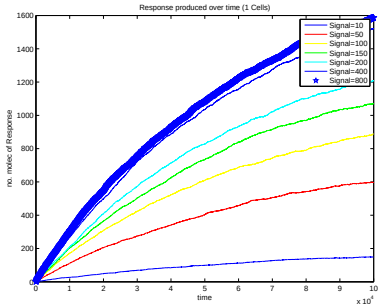
A Matlab code was written using Gillespie algorithm implementing the above model for a general bio-sensor (code(1.1)). However, problems occurred for the specific system studied by the team. It was found that the binding and unbinding reactions of TFS are much faster than the other reactions for example $c(3) = 10^6 * S$ and $c(4) = 10^6 * 4$, whereas $c(8) = 1.3$. This increases τ, the time until next reaction, which in turn makes the code run agonisingly slow for practical purposes. The reactions are interconnected and not easy to separate. In order to get round this problem, a method called slow scale

stochastic simulation algorithm was used. The code(1.1) was amended so as to to calculate the fast reactions seperatelycode(1.2), so that the propensity functions could be changed. The and the code then skips the fast reactions and simulates only the slow reactions code(1.3) and was used as follows:

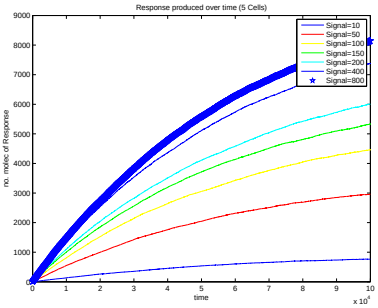
- Code(1.2) was simulated over a very small time interval of 0.0005s for each signal, in order to obtain a sampling interval for TF which consisted of $mean +/- std$.
- TF was sampled from the obtained interval and an equation to compute TFS was obtained and implemented in code (1.3) $TFS = \frac{K1*TF*S}{Km1}$.
- The final four reactions were simulated using code (1.3) over the required time interval (100000s) for a given number of simulations and a given number of cells.
- Plots of P4 output over time and P4 output with increasing signal could then be obtained and investigated. This was all implemented in code(1.3)
- The noise present in the output of P4 could also be studied and is implemented in code(1.3).

Below are graphs and tables of noise calculated for output of P4 over time for different signals. Each graph represents a different number of cells being simulated (1,5,10 and 100), averaged over 10 runs. 1 cell 5 cells 10 cells

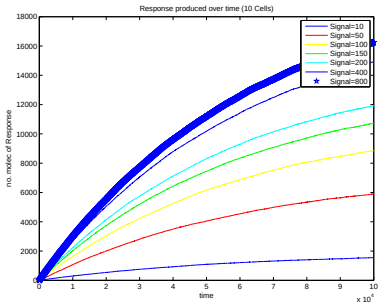
1 cell



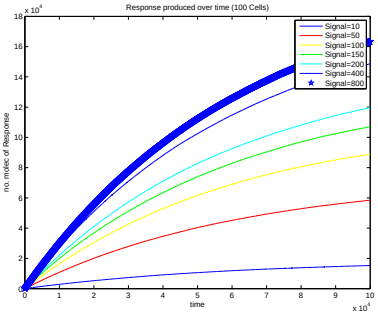
5 cells



10 cells



100 cells



Signal(μM)	10	50	100	150	200	400	800
Fano	19.3	74.2	113.9	133.6	153.9	188.8	196.8
coeff var	0.4467	0.4452	0.4475	0.4499	0.4503	0.4469	0.4435

Signal(μM)	10	50	100	150	200	400	800
Fano	183	734	1114	1352	1525	1882	2038
coeff var	0.4556	0.4407	0.4491	0.4502	0.4460	0.4478	0.4476

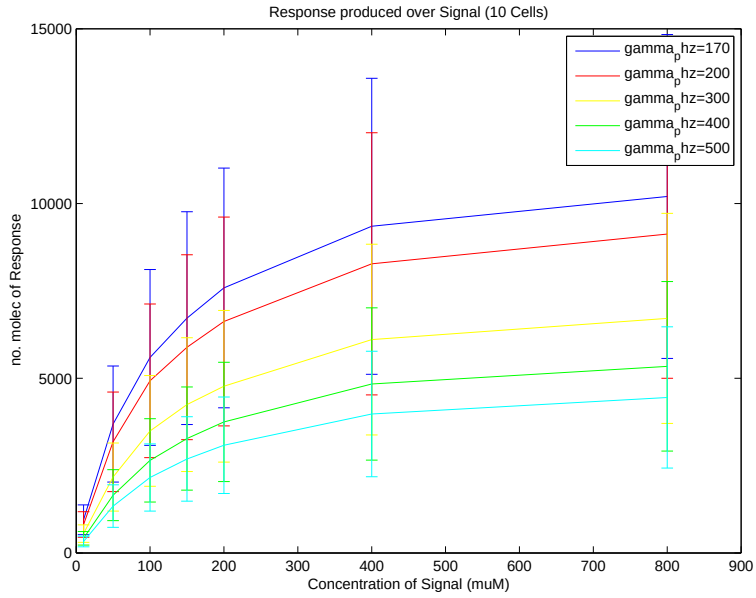
Signal(μM)	10	50	100	150	200	400	800
Fano	183	734	1114	1352	1525	1882	2038
coeff var	0.4511	0.4475	0.4462	0.4478	0.4481	0.4445	0.4475

100 cells

Signal(μM)	10	50	100	150	200	400	1927
Fano	1927	7323	11158	13426	15023	18661	20507
coeff var	0.4475	0.4455	0.4460	0.4465	0.4460	0.4426	0.4469

As was expected, increasing signal gives a higher output of P4 as is shown in each of the above graphs. Inspecting the fano factor shows that as the output of P4 increases so does the noise in the system. However, the fano factor did not behave as was expected when increasing the number of cells. The noise in the system is expected to decrease. To investigate this further the coefficient of variation was calculated and, as the number of cells increses, the noise does decrease. This is particulary evident when inspecting the values of noise for 10 and 100 cells as for each signal, the coefficient of variation is lower for 100 cells.

The goal was to construct a biosensor that will yeild a graded response to signal. Due to the extreme difficulties experienced by the Glasgow IGEM wetab in obtaining data, simulations were compared to that in a paper that investigated a similar system by Willardson et al[10]. In this paper a graded response of the luminescent output is measured over different concentrations of signal.The main parameter that affects this was found to be γ. Using the plot from the Willardson et al paper [10], where the output reached its half maximum at around S=200 μM so γ was fixed at 170μM. If γ is smaller than this then the output will saturate very quickly and at very low signals. Different values of γ were investigated ranging from 170 to 500. The simulation was run for 1,5 and 10 cells each for 10 runs. A graph for pyocyanin output over signal for each of the 5 values of γ (10 cells) is given below with standard deviation intervals of output at each value of signal. The noise in the system was calculated over signal for each of the values of gamma for 1,5 and 10 cells, however only the tables for $\gamma = 170$ and $\gamma = 500$ for 10 cells have been inculded below so as to compare the results:



$\gamma_{phz} = 170$

Signal(μM)	10	50	100	150	200	400	800
Fano	183	734	1114	1352	1525	1882	2038
coeff var	0.4511	0.4475	0.4462	0.4478	0.4481	0.4445	0.4475

$\gamma_{phz} = 500$

Signal(μM)	10	50	100	150	200	400	800
Fano	63	277	430	536	619	802	895
coeff var	0.4398	0.4526	0.4430	0.4497	0.4507	0.4490	0.4482

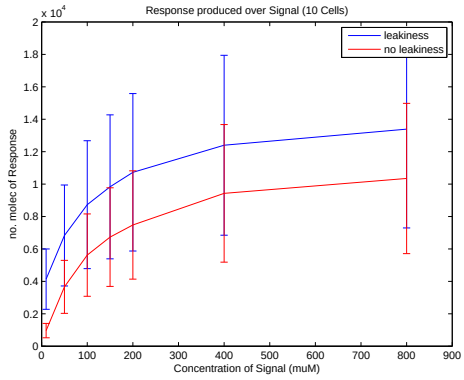
As is evident from the graph a higher value of γ gives a more graded response of P4, as was expected from observations made in the Willardson et al paper [10]. When γ increases, investigating the Fano factor shows that the noise in the output decreases, so the output of the protein is more stable.

5.2 leakiness

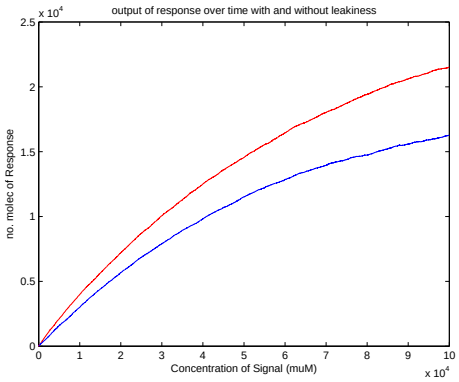
Leakiness of the promoter is unregulated expression of the gene. This means that even without any signal or bound transcription factor, there is still mRNA regulation of the gene and the outpt P4 will still be produced. This

can be modelled by changing $c(6)$ to: $\frac{\beta * TFS}{\gamma + TFS} + \alpha$. So, if $\alpha > 0$ leakiness occurs and if $\alpha = 0$ it doesn't.If the value of α is comparable to the value of the original c(6), the signal itself will have very little effect and the sensor would be useless. Below are two graphs illustrating the behaviour of the system with and without leakiness. The first is a graded response of output to signal the lower line illustrates without leakiness and the higher line illustrates with. The second is output of pyocyanin over time the bottom line illustrates without leakiness and the top with. The simulations represented in these graphs were run for 10 cells and averaged over 10 runs. All of the parameter constants and initial concentrations of proteins were the same in all simulations, only the α was changed.

leakiness/ no leakiness pyo vs signal



leakiness/ no leakiness pyo vs time



No Leakiness, 10 cells, $\gamma_{phz} = 170$

Signal(μM)	10	50	100	150	200	400	800
Fano	202	718	1133	1354	1471	1884	2046
coeff var	0.4511	0.4422	0.4483	0.4479	0.4428	0.4465	0.4441

Leakiness, 10 cells, $\gamma_{phz} = 170$

Signal(μM)	10	50	100	150	200	400	800
Fano	827	1399	1759	1978	2168	2451	2737
coeff var	0.4465	0.4519	0.4482	0.4478	0.4489	0.4440	0.4515

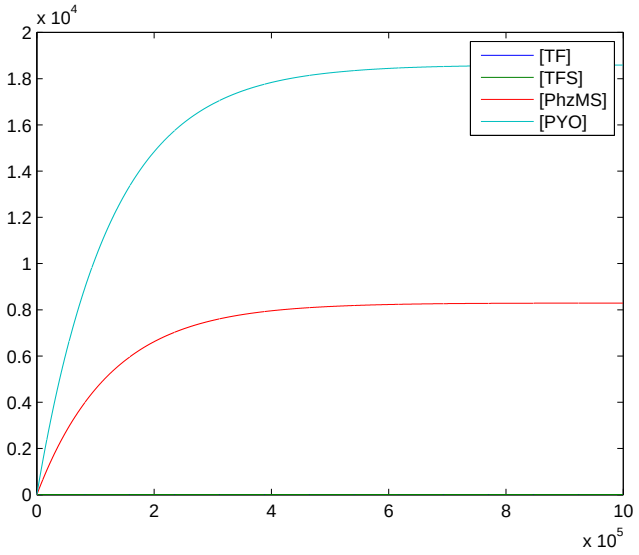
After examination of the graphs it is apparent that the promoter with leakiness provides a higher output of P4, this could give a false indication of the level of Signal detected.

5.3 positive feedback

will finish later Positive feedback is an interaction that increases or amplifies the response of the system in which it is incorporated. Although increasing the output of the final protein P4, positive feedback makes the system very noisy. Because the rate constants were so high the code(1.3) was used with feeback put as true and the results were as follows: **will put a picture of positive feedback here***** ***waiting on data hopefuly will get it today***

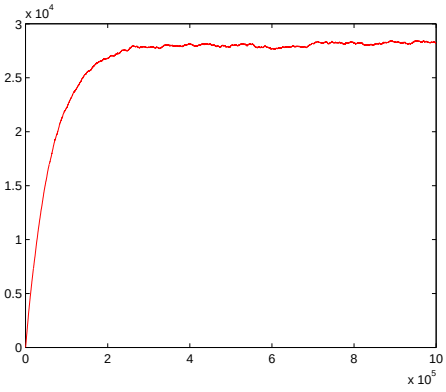
5.4 stochastic vs deterministic

Below is a graph from the deterministic approach, an equivalent system was modelled stochastically for 10 and 100 cells averaged over 10 runs. Below is the graph obtained from the deterministic ODE approach. The the line of interest in this case is the top line which represents the output of P4 or Pyocyanin.

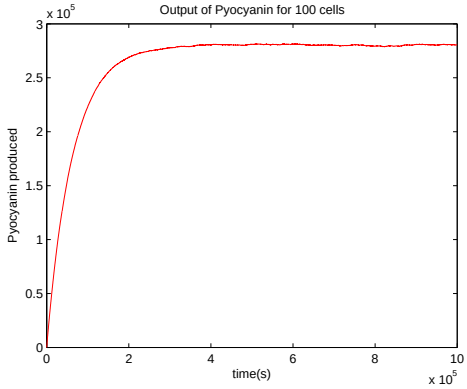


The two graphs below represent the stochastic approach for 10 and 100 cells. In both approaches all of the rate constants and initial concentrations are the same to ensure the two approaches are comparable.

leakiness/ stochastic 10 cells



stochastic 100 cells



Stochastic modelling may give a more accurate description of the behaviour in the system for a single cell or even 10 cells as it models the noise that occurs. However, as is evident from the above graphs, for a very large number of cells the two approaches are equivalent due to the fact that as the number of cells increases the noise cancells out.

5.5 Conclusions and reccomendations

Using the results from the different systems modelled certain reccomendations can be made for others building a biosensor. Changing gamma can give a graded response to signal. However, if $\gamma < 170$ the output of P4 saturates at very lower values of signal. If at all possible, leakiness should be completely avoided. However, due to the unpredicable nature of biological systems, this cannot always be the case. If leakiness does occur, it should be as low a level as possible. Positive feedback does give a higher output of the response P4, i does not however, give a graded output of P4 at different signals. The noise is also higher when positive feedback is implemented.