Error correction, information and stochastic control (optimisation) in biological systems.
Supervisor: Burroughs (Mathematics). Second supervisor: McAinsh (WMS).

Background. Biological processes continually restructure themselves, synthesising new proteins (as existing proteins are turned over or functional requirements change) and building mechanical/dynamic structures (as a cell responds to its environment, moves or divides). These processes have to be 'fit for purpose', so proteins must have the correct amino acid sequence otherwise their function is compromised, and mechanical structures must have structural integrity and meet the requisite functional requirements. These processes thus have quality control processes and error correction mechanisms. How error correction works is beginning to be understood, but mathematical models are needed to understand the principles.

One of the key principles is information. There is a balance between speed of performing a process and its accuracy, as the adage goes "Less haste, more speed" – performing a process faster leads to more errors. This is well established in protein synthesis which is a multi-step process [1,2]. The ability to identify the sequence correctly takes time and energy, thus performing this process faster means that the information (what the sequence is) is less accurate. Information generation is thus key to error correction.

For illustration, consider a simple system of two states, labelled 1,2, with transfer rates $sk_+ : 1 \to 2$, $sk_- : 2 \to 1$. Let the state at time t be X_t. At equilibrium $P(X = 1) = k_-/(k_+ + k_-)$, so if $k_- > k_+$ state 1 is preferred. The factor s determines the speed of exchange.

Now consider having a finite time T such that you want to be in state 1 at time T, starting the process with $P(X_0 = 1) = 1/2$, equilprobable. So if exchange is slow, or off ($s = 0$), $P(X_T = 1) = 1/2$, whilst if exchange is fast it approaches the equilibrium value. So we need s high, in particular $sk_+T >> 1$, achieving stationarity and the highest probbaility of being in state 1.

To increase $P(X_T = 1)$, we need information as to what state we are in. Generating that information requires time, so if we have been in the current state for time τ we will know the state to accuracy $\alpha(\tau)$, $P(S = k|X = k) = \alpha(\tau)$ where S is the inferred state. By changing the transfer rates dependent on the information gained (we make $k_+ = k_-$ here)

$$\text{rate } f(\tau)sk : \ 1 \to 2 : \ \text{rate } g(\tau)sk : \ 2 \to 1,$$

such that f, g are monotonically decreasing, increasing respectively, ($f(0) = g(0) = 1/2$), we can increase $P(X_T = 1)$. For instance $f = 1 - \alpha, g = \alpha$. The task is to find the optimal s given T, k, f, g, a small s will enable accurate state inference but switches slowly, so $P(X_T = 1) = 1/2$, the initial state. Rapid switching gives no information, so $P(X_T = 1) = 1/2$ again. This can be improved upon by making s time dependent ($s(t)$), i.e. start with rapid switching then decrease the switch rate so that the time stayed in the current states increase thereby improving accuracy of inferred state.

The project. The project will lay the mathematical foundations of decison making and control of error correction in biological processes. There are two systems that are ideal for establishing a mathematical framework, translation (protein synthesis) [1,2] and cell division (chromosome congression) [3]. The former is a multi-step process, essentially aborting a synthesis step if the information suggests an error early in the process. The latter is a mechanical process where chromosomes are captured by two spindle poles, and force sensors are used to infer the state (information generation) and make a decision as to whether the attachments are correct. Thus, a spatial simulation will be constructed; mapping to a non-spatial system may be possible through determining the event distributions - the above assumes first order kinetics whereas binding will be higher order (not Exponentially distributed). This fundamentally changes the optimisation problem.

For these systems you will build simple stochastic models, simulate the systems and explore parameter space to understand the constraints. If time permits you will express the problem as a stochastic control problem (eg for the time dependence $s(t)$ above) and use sequencial Monte Carlo techniques to solve it [4].

Desirable skills An understanding of Markov processes is required, whilst understanding the basics of stochastic differential equations would be helpful (for cell division example). Programming will likely be in MatLab but other languages acceptable if preferred, e.g. C++, python.

Opportunities for a PhD. This mini-project can lead to a PhD with Burroughs and McAinsh, and a suitable external partner. For instance,

- This could develop into a PhD focussed on building a general theory. The main techniques would then be stochastic control theory and Markov decision processes. A suitable external would be in Kent University, a protein translation expert.
- This could develop into a PhD focused on just the cell division component, developing the models to be more realistic and testing the predictions using tracked chromosome data from the latest advanced microscopes; we already have suitable data and generating more on a BBSRC grant. A suitable external is a microscopy company.

Other possible directions can be discussed.

References:

[1] Dynamic basis of fidelity and speed in translation: Coordinated multistep mechanisms of elongation and termination. Prabhakar A., Choi J., Wang J., Petrov A, Puglisi JD. 2017. Protein Science 26:1352-62.

[2] Mathematical and Computational Modelling of Ribosomal Movement and Protein Synthesis: an overview. Tobias von der Haar. *Computational and Structural Biotechnology* 2014, 1,e201204002.

[3] Building an integrated model of chromosome congression. P. Auckland and A.D. McAinsh. 2015, *J Cell Sci.* 128:3363-74

[4] Sequential Monte Carlo for Model Predictive Control. N. Kantas, J.M. Maciejowski, A. Lecchini-Visintini. 2008, *Nonlinear Model Predictive Control*, Eds. Magni et al., *LNCIS* 384, pp 263