

The influence of nuclear compartmentalisation on stochastic dynamics of self-repressing gene expression

Marc Sturrock ^a, Shiyu Li ^b, and Vahid Shahrezaei ^b

^a Department of Physiology, Royal College of Surgeons in Ireland, Dublin, Ireland

^b Department of Mathematics, Imperial College London, London, United Kingdom

April 26, 2017

Gene expression is an inherently noisy process. This noise is generally thought to be deleterious as precise internal regulation of biochemical reactions is essential for cell growth and survival. Self-repression of gene expression, which is the simplest form of a negative feedback loop, is commonly believed to be employed by cellular systems to decrease the stochastic fluctuations in gene expression. When there is some delay in autoregulation, it is also believed that this system can generate oscillations. In eukaryotic cells, mRNAs that are synthesised in the nucleus must be exported to the cytoplasm to function in protein synthesis, whereas proteins must be transported into the nucleus from the cytoplasm to regulate the expression levels of genes. Nuclear transport thus plays a critical role in eukaryotic gene expression and regulation. Some recent studies have suggested that nuclear retention of mRNAs can control noise in mRNA expression. However, the effect of nuclear transport on protein noise and its interplay with negative feedback regulation is not completely understood. In this paper, we systematically compare four different simple models of gene expression. By using simulations and applying the linear noise approximation to the corresponding chemical master equations, we investigate the influence of nuclear import and export on noise in gene expression in a negative autoregulatory feedback loop. We first present results consistent with the literature, i.e., that negative feedback can effectively buffer the variability in protein levels, and nuclear retention can decrease mRNA noise levels. Interestingly we find that when negative feedback is combined with nuclear retention, an amplification in gene expression noise can be observed and is dependant on nuclear translocation rates. Finally, we investigate the effect of nuclear compartmentalisation on the ability of self-repressing genes to exhibit stochastic oscillatory dynamics.

Keywords:

Stochastic Gene Expression; Negative Feedback; Nuclear Compartmentalisation; Oscillations

Corresponding authors: marcsturrock@rcsi.ie, v.shahrezaei@imperial.ac.uk

1 Introduction

Gene expression is a complex, two-stage process, first the DNA of the gene is transcribed into messenger RNA (mRNA) by RNA polymerase (RNAP). The mRNA is subsequently translated into protein by ribosomes. In eukaryotic cells, mRNA synthesis is localised to the nucleus from which nascent mRNAs are exported to the cytoplasm for translation. In bacteria, which have no nucleus, translation occurs as soon as part of the mRNA is synthesised. The expression levels of many genes are tightly regulated by internal or environmental signals. Regulation is typically mediated by a special class of proteins called transcription factors, which either prevent the binding of RNAP to the promoter and thereby suppressing the gene expression or they promote the binding of RNAP to the promoter and thus promote gene expression. Eukaryotic transcription factors need to relocate from the cytosol to nucleus before they can interact with transcription binding sites on the DNA.

Gene expression and gene regulatory network dynamics are noisy due to intrinsic and extrinsic stochasticity (Elowitz et al., 2002; Swain et al., 2002; Blake et al., 2003; Raser and O'Shea, 2005; Shahrezaei and Swain, 2008b; Raj and van Oudenaarden, 2008). Intrinsic stochasticity arises from the probabilistic timing of the biochemical reactions, resulting in cell-to-cell variation in gene expression levels within clonal cell populations, even if the environment is completely homogeneous (Swain et al., 2002). Its effects are most pronounced when the number of biomolecules in the system is small (e.g., <100 since noise in a Poisson process is inversely proportional to the mean). Extrinsic stochasticity is generated from interactions of the system of interest with other stochastic systems in the cell or its environment, generating stochastic kinetic parameters, and it is responsible for cell-to-cell variation beyond what is expected from intrinsic variability generated by the reactions within the gene regulatory network.

Although, cells may take advantage of gene expression noise to enhance their fitness in fluctuating environments (Acar et al., 2008), noise is generally thought to be deleterious as precise internal regulation of biochemical reactions is essential for cell growth and survival. Various control mechanisms with different noise suppression abilities exist within gene networks and are employed by cells to attenuate stochastic fluctuations in gene expression level, and therefore help increase the stability of the cells to changing environments. One well studied, both experimentally and theoretically, is transcriptional feedback, where the protein expressed from a gene inhibits its own transcription (Becskei and Serrano, 2000). Negative autoregulation of gene expression is common in eukaryotes (Lee et al., 2002), occurring both transcriptionally and post-transcriptionally. Approximately 40% of known transcription factors in *Escherichia coli* are subject to negative transcriptional autoregulation (Thieffry et al., 1998; Rosenfeld et al., 2002). Negative feedback control is typically considered as an important means of repressing the noise level in gene expression by suppressing the variability of the protein level across cells (Savageau, 1974; Becskei and Serrano, 2000; Simpson et al., 2003; Swain, 2004; Boyer et al., 2005; Tao et al., 2007). However, introducing negative feedback without increasing transcription or translation rates can amplify the intrinsic noise level, as negative feedback causes a decrease in the average number of mRNAs and proteins, which results in an increase in the protein variability (Shahrezaei et al., 2008; Singh and Hespanha, 2009). It is argued that in fact negative feedback is more effective in reducing extrinsic noise (Shahrezaei et al., 2008; Singh and Hespanha, 2009). Recent work (Oyarzún et al., 2015) agrees with this and demonstrates that noise does not behave monotonically with respect to the repression strength nor to the promoter and repressor strength. Also, it is argued that there are physical limits on how much negative feedback can reduce noise (Zhang et al., 2009; Lestas et al., 2010), hence, the role of negative feedback mechanisms in suppression of fluctuations in gene expression is complex and can

depend on the molecular details of the regulation. Furthermore, a recent study by Albert and Roodman (2015) suggests that nuclear transport improves precise temporal patterning of negative feedback gene regulatory networks. This study used stochastic simulation with two different models of nuclear transport but it does not look specifically at noise in gene expression.

It is also known that negative feedback loops can create oscillations in mRNA and protein levels (Kobayashi and Kageyama, 2011; Geva-Zatorsky et al., 2010; Nelson et al., 2004). Negative feedback loops are commonly found in a diverse range of biological processes, including inflammation, meiosis, apoptosis and the heat shock response (Alberts et al., 2008; Lahav et al., 2004; Fall et al., 2002). The mathematical modelling of GRNs containing negative feedback loops has a long history with the earliest example coming from Goodwin, where a system of two ordinary differential equations (ODEs) was used to create a phenomenologically derived model of a self-repressing gene (Goodwin, 1965). Nonlinearity in the form of a Hill function was used to capture the feedback of protein on mRNA production and this has since been used by many other models. The work of Goodwin was continued and extended by Griffith who showed that such a system could not exhibit oscillatory dynamics without the introduction of another model species (Griffith, 1968). Mahaffy and co-workers developed this work further by introducing spatial structure and also delays accounting for transcription and translation (Mahaffy and Pao, 1984; Busenberg and Mahaffy, 1985; Mahaffy, 1988). Some years later, it was discovered that introducing delays to ODE models of negative feedback loops could produce sustained oscillatory dynamics (Tiana et al., 2002). Jensen et al. found the invocation of an unknown third species (as Griffith had done) could be avoided via the introduction of delay terms to a model of the Hes1 GRN (representing the processes of transcription and translation) (Jensen et al., 2003). The Hes1 GRN is a simple example of a GRN which possesses a single negative feedback loop and benefits from having been the subject of numerous biological experiments (Hirata et al., 2002; Kageyama et al., 2007; Kobayashi et al., 2009; Kobayashi and Kageyama, 2010, 2011). A delay differential equation (DDE) model of the Hes1 GRN was also studied in Monk (2003). The effect of low molecule numbers in Monk's DDE model of the Hes1 GRN was explored in Barrio et al. (2006). A spatio-temporal model of the Hes1 GRN, using a partial differential equation (PDE) approach, was presented in Sturrock et al. (2011) and extensions of this model were considered in Sturrock et al. (2012). A spatial stochastic model of the Hes1 GRN in embryonic stem cells was studied in Sturrock et al. (2013) – this model was derived using mass action kinetics and its mean behaviour was studied in Sturrock et al. (2014b). Wang et al. considered a spatially-homogeneous model of a self-repressing gene without nonlinearity in regulation in Wang et al. (2014), and were able to identify certain regions of the parameter space which yielded 'stochastic oscillations'.

Most proteins and RNAs are not able to freely move across the nuclear envelope in either direction (cytoplasm to nucleus nor nucleus to cytoplasm) because the nuclear membranes prevent the free passage of large molecules between the nucleus and the cytoplasm. The sole channels for proteins and RNAs through the nuclear membranes are provided by the nuclear pore complexes. RNA molecules that are synthesised in the nucleus must be efficiently exported to the cytoplasm to function in protein synthesis, and certain proteins such as transcription factors that are functional inside the nucleus, need to be transported into the nucleus from the cytoplasm. The traffic of proteins and RNAs through the nuclear pore complex, particularly the nuclear protein import and mRNAs export, plays a key role in regulation of gene expression in eukaryotic cells (Cooper and Hausman, 2004). But whether the selective traffic of proteins and RNAs can act to increase or decrease the stochastic fluctuations in eukaryotic gene expression have only recently been studied. A theoretical study has demonstrated that mRNA transport rate can regulate noise in gene expression (Singh and Bokes, 2012). More recently, Bahar Halpern et al. (2015);

Battich et al. (2015) have shown experimentally that nuclear retention of mRNA in single mammalian cells can reduce noise in mRNA levels.

Here, we study the interplay between nuclear transport and stochastic gene regulation in a self-repressing gene. We compare systematically the noise in gene expression in four simple models: the canonical model of stochastic gene expression, a model including stochastic gene expression with negative feedback, a stochastic gene expression model with nuclear compartmentalisation and finally a stochastic gene expression model with negative feedback and nuclear compartmentalisation. We find results consistent with previous studies for the role of negative feedback regulation and nuclear retention in suppression of noise in gene expression. But, we also find situations when combining the two systems can amplify noise in cytosolic protein levels. Finally, given the history of studying oscillatory dynamics in gene regulatory networks containing negative feedback loops, we also take the opportunity to investigate whether there exists a relationship between noise and oscillatory dynamics. Specifically, we explore the role of transport in producing stochastic oscillations as defined in Wang et al. (2014).

Model 1 Reactions	Model 1 Description	Model 2 Reactions	Model 2 Description
$gene_{on} \xrightleftharpoons[k_{off}]{k_{on}} gene_{off}$	Switching of gene between on and off states	$gene_{off} \xrightarrow{k_{on}} gene_{on}$	Switching on of gene
$gene_{on} \xrightarrow{k_1} m$	Transcription of mRNA	$gene_{on} + p \xrightarrow{k_{off}^p} gene_{off}$	Switching off of gene due to protein
$m \xrightarrow{k_2} m + p$	Translation of protein	$gene_{on} \xrightarrow{k_1} m$	Transcription of mRNA
$m \xrightarrow{d_2} \emptyset$	mRNA degradation	$m \xrightarrow{k_2} m + p$	Translation of protein
$p \xrightarrow{d_3} \emptyset$	protein degradation	$m \xrightarrow{d_2} \emptyset$	mRNA degradation
		$p \xrightarrow{d_3} \emptyset$	protein degradation
Model 3 Reactions	Model 3 Description	Model 4 Reactions	Model 4 Description
$gene_{on} \xrightleftharpoons[k_{off}]{k_{on}} gene_{off}$	Switching of gene between on and off states	$gene_{off} \xrightarrow{k_{on}} gene_{on}$	Switching on of gene
$gene_{on} \xrightarrow{k_1} m_1$	Transcription of nuclear mRNA	$gene_{on} + p_1 \xrightarrow{k_{off}^p} gene_{off}$	Switching off of gene due to cytoplasmic protein
$m_1 \xrightarrow{d_1} \emptyset$	Nuclear mRNA degradation	$gene_{on} \xrightarrow{k_1} m$	Transcription of mRNA
$m_1 \xrightarrow{T_1} m_2$	Translocation of nuclear mRNA to cytoplasm	$m_1 \xrightarrow{d_1} \emptyset$	Nuclear mRNA degradation
$m_2 \xrightarrow{k_2} m_2 + p_2$	Translation of cytoplasmic protein	$m_1 \xrightarrow{T_1} m_2$	Translocation of nuclear mRNA to cytoplasm
$m_2 \xrightarrow{d_2} \emptyset$	Cytoplasmic mRNA degradation	$m_2 \xrightarrow{k_2} m_2 + p_2$	Translation of cytoplasmic protein
$p_2 \xrightarrow{d_3} \emptyset$	Cytoplasmic protein degradation	$m_2 \xrightarrow{d_2} \emptyset$	Cytoplasmic mRNA degradation
$p_2 \xrightarrow{T_2} p_1$	Translocation of cytoplasmic protein to nucleus	$p_2 \xrightarrow{d_3} \emptyset$	Cytoplasmic protein degradation
$p_1 \xrightarrow{d_4} \emptyset$	Nuclear protein degradation	$p_2 \xrightarrow{T_2} p_1$	Translocation of cytoplasmic protein to nucleus
		$p_1 \xrightarrow{d_4} \emptyset$	Nuclear protein degradation

Table 1: Reactions of the 4 models considered and description of reactions.

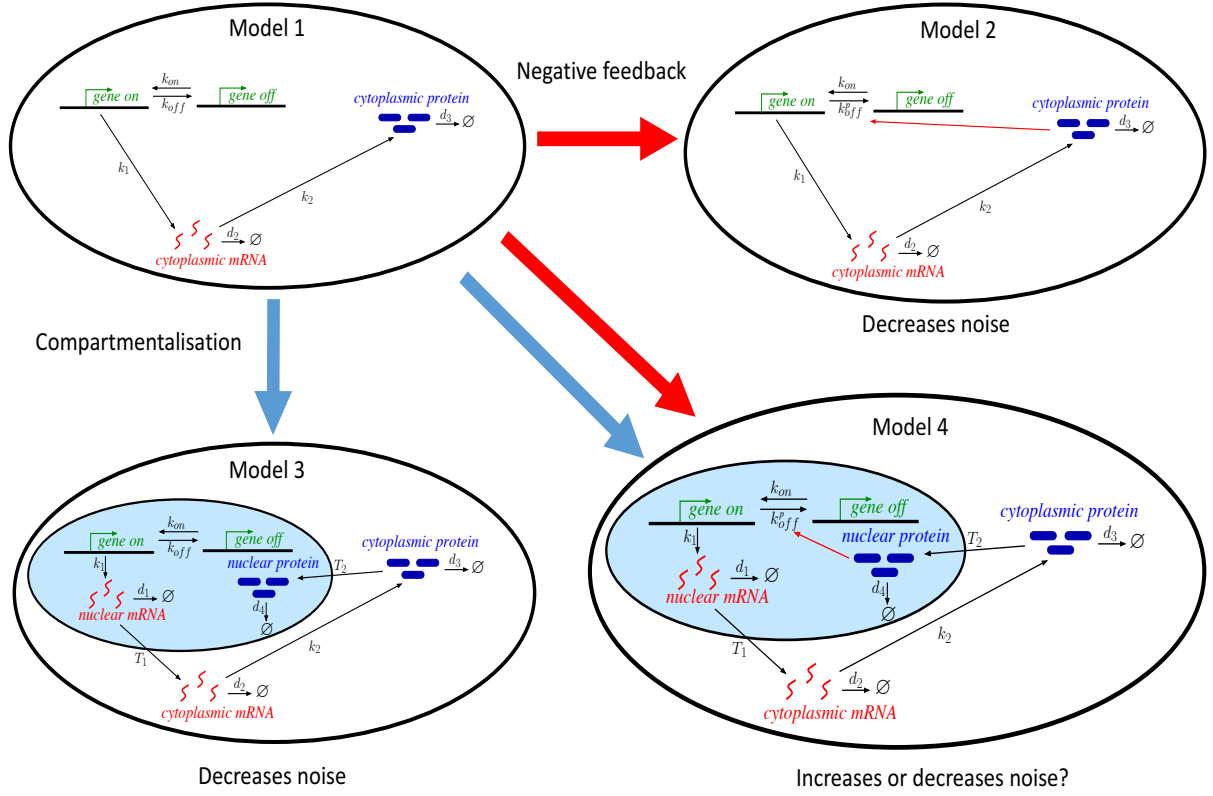


Figure 1: A schematic diagram showing the four different models considered in this study and illustrating the central question of this study, namely, how do compartmentalisation and negative feedback affect noise levels in gene expression? The top left diagram (model 1) depicts the well mixed gene expression model which represents a gene switching between an “on” and “off” state. When the gene is in the “on” state it can produce mRNA and this mRNA can subsequently be translated into protein. Both mRNA and protein are assumed to degrade linearly. By linking the abundance of protein to the rate which the gene switches off, we obtain a well-mixed negative feedback loop or a self-repressing gene (model 2). By adding a nuclear compartment to the model 1 and splitting mRNA and protein into nuclear and cytoplasmic species, we arrive at a compartmental model of gene expression (model 3). If we then link the abundance of nuclear protein to the rate at which the gene switches off we have a compartmental model of negative feedback (model 4).

2 Methods

2.1 Noise Measures

The most common quantitative measure of noise is the coefficient of variation (CV) squared, defined as

$$CV^2 = \frac{\sigma^2}{\langle X \rangle^2}$$

where σ^2 and $\langle X \rangle$ denote the variance and the mean steady-state number of molecule X . The coefficient of variation is a normalised measure of variability as it measures the extent of variability relative to the population mean.

Noise can also be quantified by the Fano factor, defined as

$$F = \frac{\sigma^2}{\langle X \rangle}.$$

The Fano factor of any Poisson process is 1, as the mean and the variance of any Poisson process are equal. Both the coefficient of variation squared and the Fano factor show the extent of variability in relation to the mean of the population.

Here, we are interested in comparing the noise levels between different models, therefore we define the following measure of relative noise between models M and N :

$$\log_{10} \frac{CV_M^2}{CV_N^2} \quad (1)$$

where CV_M^2 denotes the CV squared produced for mRNA or protein species in model M and CV_N^2 denotes the same for model N .

2.2 Chemical master equations

For any system of biochemical reactions, the lowest level of description of the stochastic dynamics is the chemical master equation (CME). The CME describes how the probability distribution of the molecular state of the system changes with time. In order to be clear and transparent about the model structures we consider and compare in this paper, we present the full chemical master equations for each model here. See Table 1 for an overview of the different model reactions.

Model 1

The first model captures gene expression in a well-mixed cell. This can be thought of as a model of gene expression in a prokaryotic cell. The model has three discrete variables, the number of free promoters (f), mRNAs (m) and proteins (n). F_0 represents the total number free promoters. The chemical master equation for this system can be written as

$$\begin{aligned} \frac{\partial P(m, n, f, t)}{\partial t} = & k_{off}(f+1)P(m, n, f+1, t) - k_{off}fP(m, n, f, t) \\ & + k_{on}(F_0 - f + 1)P(m, n, f-1, t) - k_{on}(F_0 - f)P(m, n, f, t) \\ & + k_1fP(m-1, n, f, t) - k_1fP(m, n, f, t) \\ & + k_2mP(m, n, n-1, f, t) - k_2mP(m, n, f, t) \\ & + d_1(m+1)P(m+1, n, f, t) - d_1mP(m, n, f, t) \\ & + d_2(m+1)P(m, m+1, n, f, t) - d_2mP(m, n, f, t) \\ & + d_3(n+1)P(m, n+1, n, f, t) - d_3nP(m, n, f, t) \\ & + d_4(n+1)P(m, n, n+1, f, t) - d_4nP(m, n, f, t). \end{aligned}$$

Model 2

The second model captures gene expression with an autoregulatory negative feedback loop in a well-mixed cell. This model can be attributed to the classical model of negative feedback first proposed by Goodwin (1965) and can be thought of as an autoregulatory feedback loop in a prokaryotic cell. The

model has three discrete variables, the number of free promoters (f), mRNAs (m) and proteins (n). F_0 represents the total number free promoters. The chemical master equation for this system can be written as

$$\begin{aligned} \frac{\partial P(m, n, f, t)}{\partial t} = & k_{off}^p n(f+1)P(m, n, f+1, t) - k_{off}^p n f P(m, n, f, t) \\ & + k_{on}(F_0 - f + 1)P(m, n, f-1, t) - k_{on}(F_0 - f)P(m, n, f, t) \\ & + k_1 f P(m-1, m, n, f, t) - k_1 f P(m, n, f, t) \\ & + k_2 m P(m, n, n-1, f, t) - k_2 m P(m, n, f, t) \\ & + d_1(m+1)P(m+1, m, n, f, t) - d_1 m P(m, n, f, t) \\ & + d_2(m+1)P(m, m+1, n, f, t) - d_2 m P(m, n, f, t) \\ & + d_3(n+1)P(m, n+1, n, f, t) - d_3 n P(m, n, f, t) \\ & + d_4(n+1)P(m, n, n+1, f, t) - d_4 n P(m, n, f, t). \end{aligned}$$

Model 3

The third model captures gene expression in a compartmentalised cell (i.e., one containing a nucleus). This is a simple model of gene expression in a eukaryotic cell. The model has five discrete variables, the number of free promoters (f), mRNAs in the nucleus (m_1), mRNAs in the cytoplasm (m_2), proteins in the nucleus (n_1) and proteins in the cytoplasm (n_2). F_0 represents the total number free promoters. The chemical master equation for this system can be written as

$$\begin{aligned} \frac{\partial P(m_1, m_2, n_1, n_2, f, t)}{\partial t} = & k_{off}(f+1)P(m_1, m_2, n_1, n_2, f+1, t) - k_{off} f P(m_1, m_2, n_1, n_2, f, t) \\ & + k_{on}(F_0 - f + 1)P(m_1, m_2, n_1, n_2, f-1, t) - k_{on}(F_0 - f)P(m_1, m_2, n_1, n_2, f, t) \\ & + k_1 f P(m_1 - 1, m_2, n_1, n_2, f, t) - k_1 f P(m_1, m_2, n_1, n_2, f, t) \\ & + k_2 m_2 P(m_1, m_2, n_1, n_2 - 1, f, t) - k_2 m_2 P(m_1, m_2, n_1, n_2, f, t) \\ & + d_1(m_1 + 1)P(m_1 + 1, m_2, n_1, n_2, f, t) - d_1 m_1 P(m_1, m_2, n_1, n_2, f, t) \\ & + d_2(m_2 + 1)P(m_1, m_2 + 1, n_1, n_2, f, t) - d_2 m_2 P(m_1, m_2, n_1, n_2, f, t) \\ & + d_3(n_1 + 1)P(m_1, m_2, n_1 + 1, n_2, f, t) - d_3 n_1 P(m_1, m_2, n_1, n_2, f, t) \\ & + d_4(n_2 + 1)P(m_1, m_2, n_1, n_2 + 1, f, t) - d_4 n_2 P(m_1, m_2, n_1, n_2, f, t) \\ & + T_1(m_1 + 1)P(m_1 + 1, m_2 - 1, n_1, n_2, f, t) - T_1 m_1 P(m_1, m_2, n_1, n_2, f, t) \\ & + T_2(n_2 + 1)P(m_1, m_2, n_1 - 1, n_2 + 1, f, t) - T_1 n_2 P(m_1, m_2, n_1, n_2, f, t). \end{aligned}$$

Model 4

The fourth model captures gene expression in a compartmentalised cell (i.e., one containing a nucleus) with a negative feedback. This model was explored before in Sturrock et al. (2013) and represents gene expression in a eukaryotic cell with negative feedback. The model has five discrete variables, the number of free promoters (f), mRNAs in the nucleus (m_1), mRNAs in the cytoplasm (m_2), proteins in the nucleus (n_1) and proteins in the cytoplasm (n_2). F_0 represents the total number free promoters. The

chemical master equation for this system can be written as

$$\begin{aligned}
\frac{\partial P(m_1, m_2, n_1, n_2, f, t)}{\partial t} = & k_{off}^p n_1 (f+1) P(m_1, m_2, n_1, n_2, f+1, t) - k_{off}^p f n_1 P(m_1, m_2, n_1, n_2, f, t) \\
& + k_{on} (F_0 - f + 1) P(m_1, m_2, n_1, n_2, f-1, t) - k_{on} (F_0 - f) P(m_1, m_2, n_1, n_2, f, t) \\
& + k_1 f P(m_1 - 1, m_2, n_1, n_2, f, t) - k_1 f P(m_1, m_2, n_1, n_2, f, t) \\
& + k_2 m_2 P(m_1, m_2, n_1, n_2 - 1, f, t) - k_2 m_2 P(m_1, m_2, n_1, n_2, f, t) \\
& + d_1 (m_1 + 1) P(m_1 + 1, m_2, n_1, n_2, f, t) - d_1 m_1 P(m_1, m_2, n_1, n_2, f, t) \\
& + d_2 (m_2 + 1) P(m_1, m_2 + 1, n_1, n_2, f, t) - d_2 m_2 P(m_1, m_2, n_1, n_2, f, t) \\
& + d_3 (n_1 + 1) P(m_1, m_2, n_1 + 1, n_2, f, t) - d_3 n_1 P(m_1, m_2, n_1, n_2, f, t) \\
& + d_4 (n_2 + 1) P(m_1, m_2, n_1, n_2 + 1, f, t) - d_4 n_2 P(m_1, m_2, n_1, n_2, f, t) \\
& + T_1 (m_1 + 1) P(m_1 + 1, m_2 - 1, n_1, n_2, f, t) - T_1 m_1 P(m_1, m_2, n_1, n_2, f, t) \\
& + T_2 (n_2 + 1) P(m_1, m_2, n_1 - 1, n_2 + 1, f, t) - T_1 n_2 P(m_1, m_2, n_1, n_2, f, t).
\end{aligned}$$

2.3 Linear noise approximation

In general, the full probability distribution from the CME can be analytically obtained only in a few simple cases (see e.g. Shahrezaei and Swain (2008a); Smith and Shahrezaei (2015)). However, to obtain results for noise only up to the second moment of the probability distributions are needed and there are different methods available for this. A common approach is the use of the linear noise approximation (LNA), which is obtained through van Kampen's Ω -expansion (Van Kampen and Reinhardt, 2007; Elf and Ehrenberg, 2003) of the CME. This approach is computationally much more efficient than using the stochastic simulation algorithm (SSA) of Gillespie (Gillespie, 1976), which produces statistically exact trajectories of the CME. The LNA allows us to explore large portions of the parameter space efficiently.

We consider the stochastic dynamics of a chemical system consisting of N molecular species and R chemical reactions. The CME for this system can be expressed as

$$\frac{\partial P(X, t)}{\partial t} = \Omega \sum_{j=1}^R \left(\prod_{i=1}^N E^{-S_{ij}} - 1 \right) f_j(X, \Omega) P(X, t) \quad (2)$$

where E denotes a step operator, S_{ij} is the stoichiometric matrix of the reactions, $f_j(X, \Omega)$ is the reaction propensity of the j th reaction and Ω is the system volume.

The copy number of chemical species i is written as the sum of its deterministic value and a random variable ξ_i

$$X_i = \Omega \phi_i + \Omega^{1/2} \xi_i \quad (3)$$

where ϕ_i is the macroscopic concentration of X_i defined from

$$\phi_i = \lim_{\Omega \rightarrow +\infty} X_i / \Omega. \quad (4)$$

Then the probability distribution $P(X, t)$ can be written as:

$$P(X, t) = P(\Omega \phi + \Omega^{1/2} \xi, t) = \prod (\xi, t). \quad (5)$$

By inserting $\prod(\xi, t)$ into equation (2), the CME can be approximated by:

$$\frac{\partial \prod(\xi, t)}{\partial t} = - \sum_{i,j} A_{ik} \frac{\partial (\xi_k \prod)}{\partial \xi_i} + \frac{1}{2} \sum_{i,j} \left(\prod [BB^T]_{ik} \frac{\partial^2 \prod}{\partial \xi_i \partial \xi_k} \right), \quad (6)$$

where

$$A_{ik} = \frac{\partial(S_i \cdot f)}{\partial \phi_k} \Big|_{\phi=\bar{\phi}}, \quad (7)$$

and

$$B = S \sqrt{\text{diag}(f(\bar{\phi}))}. \quad (8)$$

and $\bar{\phi}$ is the solution of

$$Sf(\bar{\phi}) = 0, \quad (9)$$

The steady state solution is

$$P(\xi) = ((2\pi)^{N/2} \sqrt{\text{Det}C})^{-1} \exp(-\xi^T C \xi / 2). \quad (10)$$

The stationary distribution for $P(X_i)$ is then taken to be the normal distribution

$$N(\text{mean} = \Omega \bar{\phi}_i, \sigma^2 = C_{ii}). \quad (11)$$

Which has zero mean vector and the covariance C is the solution of the (Lyapunov) matrix equation:

$$AC + CA^T + \Omega B B^T = 0. \quad (12)$$

We can then compute the Fano-factor for chemical species X_i by

$$F_i = \frac{\text{Var}(X_i)}{E(X_i)} = \frac{C_{ii}}{\Omega \bar{\phi}_i}. \quad (13)$$

We also check the accuracy of the LNA in comparison to SSA in Appendix A.

2.4 Parameter values and model comparison

The baseline parameter set is defined in Table 2. The values used are based on physiologically realistic parameters used for modelling similar gene regulatory network models (Sturrock et al., 2013; Monk, 2003; Terry et al., 2011). We neglect mRNA degradation in the nucleus as in Bahar Halpern et al. (2015); Battich et al. (2015). For simplicity, we do not assume mRNA is imported into the nucleus nor do we assume protein is exported from the nucleus. The parameters in our model, which are the most difficult to measure and for which there exists the least data to support their values are T_1 and T_2 . In order to shed light on the influence of these parameter values, we will explore ranges of these parameters in the results section.

In order to make a fair comparison between the four different models considered we match the mean levels of cytoplasmic mRNA and protein. To achieve this we vary the transcription (k_1) and translation (k_2) rates. We also match the mean transcriptional activity which we denote by g . In the models without negative feedback, the steady-state probability of the gene being active equals

$$\frac{k_{on}}{k_{on} + k_{off}},$$

hence we choose

$$k_{on} = \frac{k_{off}}{\frac{1}{g} - 1}$$

Description	Parameter	Value
Gene deactivation	k_{off}	0.05 min^{-1}
Gene deactivation due to protein	k_{off}^p	$0.05 \text{ min}^{-1} \text{ vol}^{-1}$
Nuclear mRNA degradation	d_1	0 min^{-1}
Cytoplasmic mRNA degradation	d_2	0.03 min^{-1}
Nuclear protein degradation	d_3	0.001 min^{-1}
Cytoplasmic protein degradation	d_4	0.001 min^{-1}
Cytoplasmic mRNA number	m_2	10
Cytoplasmic protein number	n_2	1000
Steady state gene occupancy	g	0.3
Transcription	k_1	-
Translation	k_2	-
Gene activation	k_{on}	-
mRNA export rate	T_1	$1 \times 10^{-6} \text{ to } 10 \text{ min}^{-1}$
protein import rate	T_2	$1 \times 10^{-6} \text{ to } 10 \text{ min}^{-1}$

Table 2: Description of parameters in the gene expression models and values used in results section unless otherwise stated. The upper partition shows values fixed to enable a fair model comparison. The lower partition shows the parameters which we vary.

whereas for the models with negative feedback the steady-state probability of the gene being active equals

$$\frac{k_{on}}{k_{on} + k_{off}^p \langle n \rangle_{ss}},$$

where $\langle n \rangle_{ss}$ denotes the steady-state average number of protein affecting the negative feedback (which is the nuclear levels of protein in the models with a nucleus). Hence, in models with negative feedback if k_{on} is equivalent to

$$k_{on} = \frac{k_{off}^p \langle n \rangle_{ss}}{\frac{1}{g} - 1},$$

then the probability of the gene being active is equal to g for the different models. By employing this matching scheme the differences in noise levels cannot be attributed to differences in the gene activity levels, nor the abundance of mRNA and protein but instead the differences in model structure. One should note as demonstrated previously by (Singh and Hespanha, 2009), the specific choice of adapting parameters of translation and transcription when negative feedback is introduced will affect the degree of suppression of molecular fluctuations. Furthermore, we note that as has been shown before, the specific choice of what parameters to vary to maintain protein levels could affect the outcome of our results (Singh and Hespanha, 2009). Alternatively, we could fix the transcription and translation rates and vary the mRNA and protein degradation rates. However, since these rate constants already take physiologically realistic values, we felt it better to fix them and vary parameters for which we have less information. For completeness, we have shown the dependence of the nuclear mRNA levels, nuclear protein levels, transcription rate and translation rate on transport parameters in Appendix B. We also study the effect explored including nuclear degradation of mRNA in Appendix D.

3 Results

We present here the 6 different possible unique model comparisons that can be made with the 4 models illustrated in Figure 1 and detailed in Table 1. We focus our comparison on cytoplasmic mRNA and protein noise levels and use the CV^2 metric, though we note our results remained the same if we use other measures of noise such as the Fano Factor. All calculations were performed using Julia using the “SymPy.jl” package and plotting was done using “PyPlot.jl” (Bezanson et al., 2014).

3.1 Negative feedback decreases protein noise but can increase mRNA noise

It is well established that negative feedback can reduce noise in prokaryotic gene expression. This has been shown both experimentally using synthetic biology techniques (Becskei and Serrano, 2000) and through the use of mathematical models (Singh et al., 2012). In Figure 2 we explore the relative noise of model 1 vs. model 2 with respect to varying the mean levels of mRNA (m) and protein (p) with all other parameters as in Table 2. Specifically, we visualise the output of equation (1), i.e., $\log_{10} \frac{CV_2^2}{CV_1^2}$ as computed using the LNA using a contour plot. Using this equation, if the output is negative then model 2 produces less noise than model 1 and vice versa if a positive output is produced while if the output is exactly zero then the two models produce equal noise. It can be seen that protein noise levels are always reduced in model 2 compared with model 1 but mRNA noise levels either are not significantly modified or can be even increased in a regime where mean mRNA levels exceed mean protein levels (for other parameter sets the noise in mRNA can be reduced in model 2 vs. model 1, for example see Appendix C). It is worth noting that such a regime is biologically unrealistic with protein expression levels typically 10 - 100 times that of mRNA (Greenbaum et al., 2003). Though protein noise is almost uniformly decreased regardless of the mean mRNA or protein value (see colour bar of Figure 2b), we find that protein noise reduction is at its most pronounced when protein levels are high whereas mRNA noise is approximately equal between the two different models for high mean protein levels. Hence, we find there is a decoupling between mRNA and protein noise, which is unexpected because protein noise is generally dominated by mRNA noise.

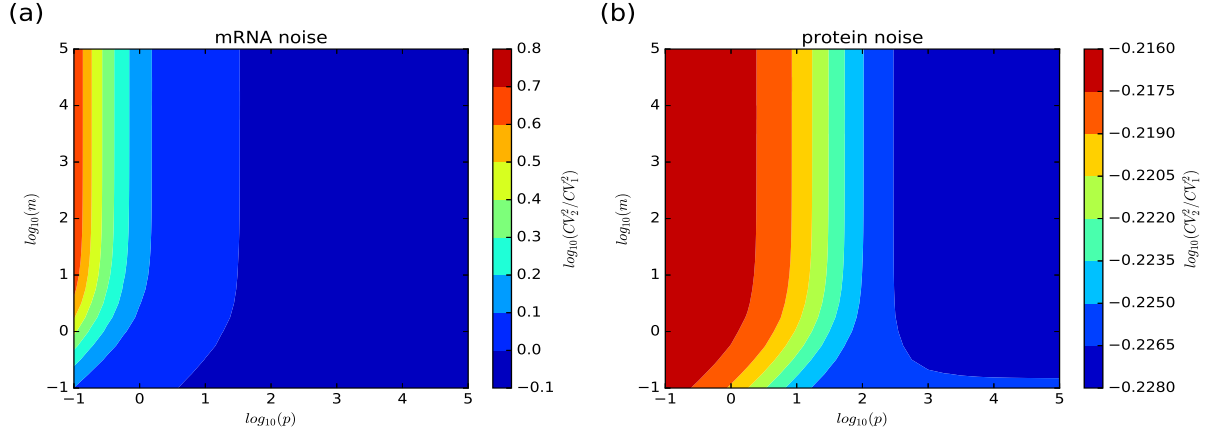


Figure 2: Comparison of relative noise produced by models 1 and 2 for different average numbers of mRNA and protein. Plot (a) shows relative mRNA noise and plot (b) shows relative protein noise where relative noise is computed using equation (1). Parameters as per Table 2.

3.2 Compartmentalisation decreases mRNA noise but can increase protein noise

Bahar Halpern et al. (2015); Battich et al. (2015) demonstrated that nuclear retention of mRNA can efficiently buffer cytoplasmic transcript levels from noise that emanates from transcriptional bursts. Their studies challenged the view that transcripts predominantly reside in the cytoplasm and revealed a role of the nucleus in dampening gene expression noise. In this section, we study the impact of the cell nucleus by comparing noise levels of model 3 and model 1 (again using equation (1) with other parameters as in Table 2) whilst varying the rate of nuclear export of mRNA and nuclear import of protein. We note that models 1 and 2 do not depend on transport parameters and hence are constant with respect to variation of T_1 and T_2 . Consistent with the findings of Bahar Halpern et al. (2015); Battich et al. (2015), we find that the addition of the nucleus reduces mRNA noise (see Figure 3(a)). However, by extending our study to the protein level we discover that protein noise can be increased if mRNA export and protein import are sufficiently fast ($> 1 \times 10^{-2} \text{min}^{-1}$). Moreover, protein noise can be reduced when compartmentalisation is introduced if mRNA export and protein import is slow ($< 1 \times 10^{-3} \text{min}^{-1}$). The level of mRNA noise is only dependent on T_1 (since there is no interaction between protein and mRNA in these models) while protein noise is sensitive to changes in both nuclear export of mRNA and nuclear import of protein.

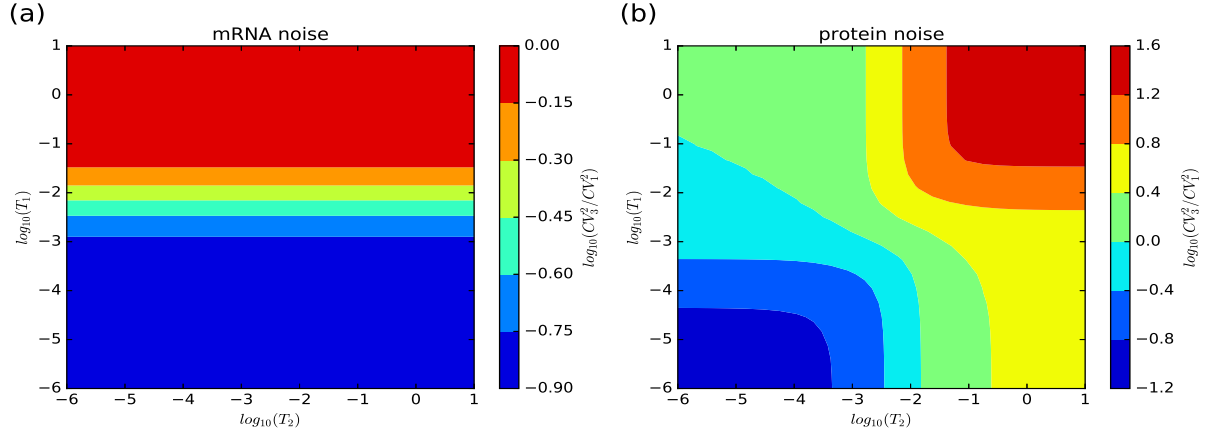


Figure 3: Comparison of relative noise produced by models 1 and 3 for different mRNA export and protein import rates. Plot (a) shows relative mRNA noise and plot (b) shows relative protein noise where relative noise is computed using equation (1). Parameters as per Table 2.

3.3 Negative feedback or compartmentalisation can be more effective in reducing protein noise depending on translocation rates

In the previous two sections, we have shown that negative feedback and compartmentalisation can effectively reduce both mRNA and protein noise levels. In this section, we compare models 2 and 3 to see whether negative feedback or compartmentalisation is more effective in reducing noise and display the results in Figure 4. As in the previous section, we vary both the mRNA nuclear export rate and protein nuclear import rate and display $\log_{10} \frac{CV_3^2}{CV_2^2}$ whilst keeping other parameters as in Table 2. We find that mRNA noise is always less in model 3 than model 2 and the two models produce equal noise in the fast nuclear export limit, consistent with the effects observed in the previous two figures. With respect to protein noise, we observe that it is greater in model 3 than model 2 if nuclear export of mRNA is fast and nuclear import of protein is fast while if translocation rates are slow then compartmentalisation is more effective in reducing noise than negative feedback. Hence, compartmentalisation can be more or less effective than negative feedback in reducing protein noise. As we noted previously, models 1 and 2 do not depend on transport parameters, therefore one can view Figure 4 as a rescaling of Figure 3.

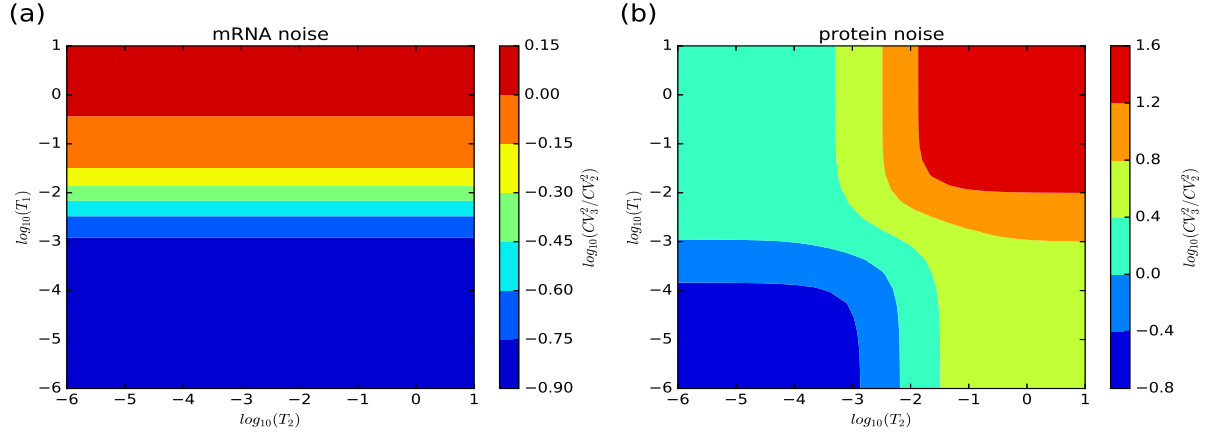


Figure 4: Comparison of relative noise produced by models 2 and 3 for different mRNA export and protein import rates. Plot (a) shows relative mRNA noise and plot (b) shows relative protein noise where relative noise is computed using equation (1). Parameters as per Table 2.

3.4 Compartmentalisation and negative feedback can increase or decrease noise levels depending on translocation rates

If both compartmentalisation and negative feedback can effectively decrease noise, it would seem plausible that combining the two would prove to be a robust method of reducing gene expression noise. In this section, we explore this idea by comparing model 4 with the other 3 models over a range of mRNA nuclear export rates and protein nuclear import rates. We present our results for this in Figure 5. Plots (a) and (b) show the relative noise of model 1 vs. model 4, plots (c) and (d) show the relative noise of model 2 vs. model 4 and plots (e) and (f) show the relative noise of model 3 vs. model 4. Strikingly, we find that the combination of both negative feedback and compartmentalisation can lead to an increase in both mRNA and protein noise compared to the other models. Specifically, combining slow protein import with fast mRNA export results in an increase in noise in both mRNA and protein while slow protein import combined with slow mRNA export is the most effective way to dampen noise. Furthermore, we find that the most effective way to increase protein noise in model 4 compared with models 1 or 2 is to make both mRNA export and protein import are fast. This suggests that eukaryotic cells can tune mRNA and protein noise levels by varying translocation rates. Additionally, from Figure 5 (b) and (d) we can see that there is an optimal import rate that minimises protein noise for mRNA export rates in the range 10^{-3} to 10^1 min^{-1} . Interestingly, experimentalists have estimated the nuclear export rates to be in the range 10^{-4} to 10^1 min^{-1} (Bahar Halpern et al., 2015), meaning that the import rate is crucial for determining whether or not noise is increased or decreased. By comparing the noise produced by model 3 vs. model 4 (plots (e) and (f)), we can see that the addition of negative feedback to a gene expression model with compartmentalisation can increase mRNA and protein noise if protein import is slow while if protein import is fast then the models produce approximately equal mRNA and protein noise levels.

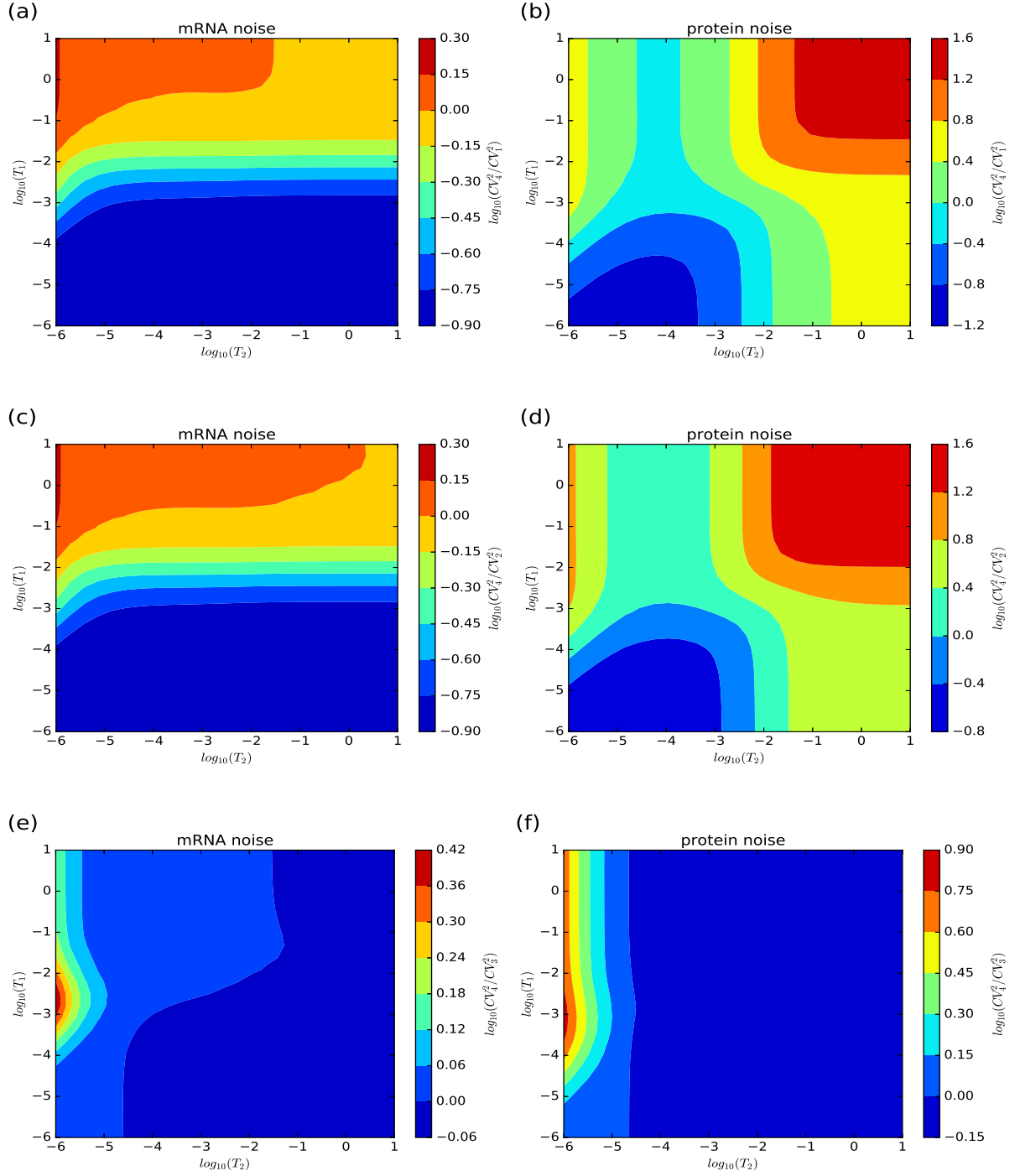


Figure 5: Comparison of relative noise produced by models 1 and 4 (plots (a) and (b)), 2 and 4 (plots (c) and (d)), and 3 and 4 (plots (e) and (f)) for different mRNA export and protein import rates. Plots (a), (c) and (e) show relative mRNA noise and plots (b), (d) and (f) show relative protein noise where relative noise is computed using equation (1). Parameters as per Table 2.

3.5 Fast decaying protein, noise and stochastic oscillations

While proteins are typically a lot more stable than mRNAs, there are some proteins that have fast active degradation. For example the Hes1 protein has a protein life time that is comparable to mRNA life time and this is a requirement for oscillatory dynamics at the protein level. In Appendix C (Figures 12 and 13), we have reproduced the noise characteristics of our models as explored above for the fast decaying protein case with $d_2 = d_3 = d_4 = 0.03 \text{ min}^{-1}$ (as is the case for the Hes1 gene regulatory network (Hirata et al., 2002)). Overall, we find the noise in the mRNA and fast decaying proteins are much more aligned. For example, it can be observed that the size of the region where model 3 decreases protein noise relative to model 1 is increased. This means for proteins with fast decay, nuclear retention is a much more effective way of reducing protein noise. Also, with fast protein decay, the region where protein noise is maximised in model 4 compared to the other models is now solely the slow protein import and fast mRNA export regime.

We next investigate the existence of oscillatory behaviour in models containing negative feedback with fast decaying protein, i.e. models 2 and 4 as it is known negative feedback with delay can produce oscillations as discussed in the introduction. We first attempted to look for peak frequencies in the power spectral density of long protein time series produced using the SSA but could not find any significant peaks. We also looked at the corresponding deterministic ODE system to see if oscillatory solutions could be found in the mean-field model but concluded that the introduction of a nonlinear Hill-function like term was necessary to observe such behaviour. Though we could not find the existence of deterministic oscillations, inspired by Wang et al. (2014), we investigated the existence of the stochastic analogue of such behaviour.

Wang et al. (2014) introduced a technique for assessing temporal regularity in protein spiking behaviour by dividing the state space into two regions I and II and studying the distribution of the times where the system leaves I to enter II. To avoid spurious transitions, events are only recorded where the protein level crosses successively the mean protein level $\langle P \rangle$ and $\langle P \rangle \pm 0.25\text{std}(P)$ where “std(P)” denotes the standard deviation of protein levels. Given the list of times where the system transits from low to high protein levels, we compute the probability of detecting n transitions within a time interval of fixed duration. We then compute the Fano Factor (as defined in section 2.1) associated with this probability distribution. Note that a Fano factor greater (less) than unity indicates super-Poissonian (respectively, sub-Poissonian) behaviour corresponding to a bunching (respectively, anti-bunching) of protein spikes. This spike anti-bunching can be viewed as a stochastic counterpart of deterministic oscillations.

In plots (a) and (b) of Figure 6, we show a sample time series produced by the SSA for model 2 and the corresponding probability distribution produced by computing 10000 of such time series. We show that model 2 does not exhibit subpoissonian protein spiking as indicated by the fact the Fano Factor is greater than 1, though under other parameter regimes it is shown that this model can exhibit stochastic oscillations, see Wang et al. (2014) for details. Plots (c) and (d) show the corresponding results for model 4 with $T_1 = 1 \text{ min}^{-1}$ and $T_2 = 1 \times 10^{-5} \text{ min}^{-1}$. The Fano Factor produced by model 4 is subpoissonian and hence we can say that transport alone can introduce regularity into protein spiking. Plot (e) shows the result of varying T_1 and T_2 on the Fano factor of transition numbers. We can see that the transport parameters can reduce the Fano factor below 1 but can also increase it significantly above 1. By comparing this figure with Figure 13 (b) from Appendix C, we can see that there is some overlap in terms of the areas of the (T_1, T_2) parameter space in which the relative noise of model 4 is increased and the Fano Factor of the transition numbers is below 1. In particular, the fast mRNA export coupled with slow protein import parameter regime results in both an increase in the relative noise of model 4 compared with

model 1 and subpoissonian protein spiking behaviour. To show this explicitly, Figure 6 (f), shows regions of the parameter space where subpoissonian behaviour in terms of number of transitions is observed and model 4 exceeds model 1 noise (highlighted with number 3), only subpoissonian behaviour in terms of number of transitions (highlighted with number 2), only noise in model 4 exceeds model 1 (highlighted with number 1) and neither subpoissonian or high relative noise in model 4 is observed (highlighted with number 0). We highlight the fact that in order to observe stochastic oscillations, the protein import does not have to be very slow but mRNA export must be fast.

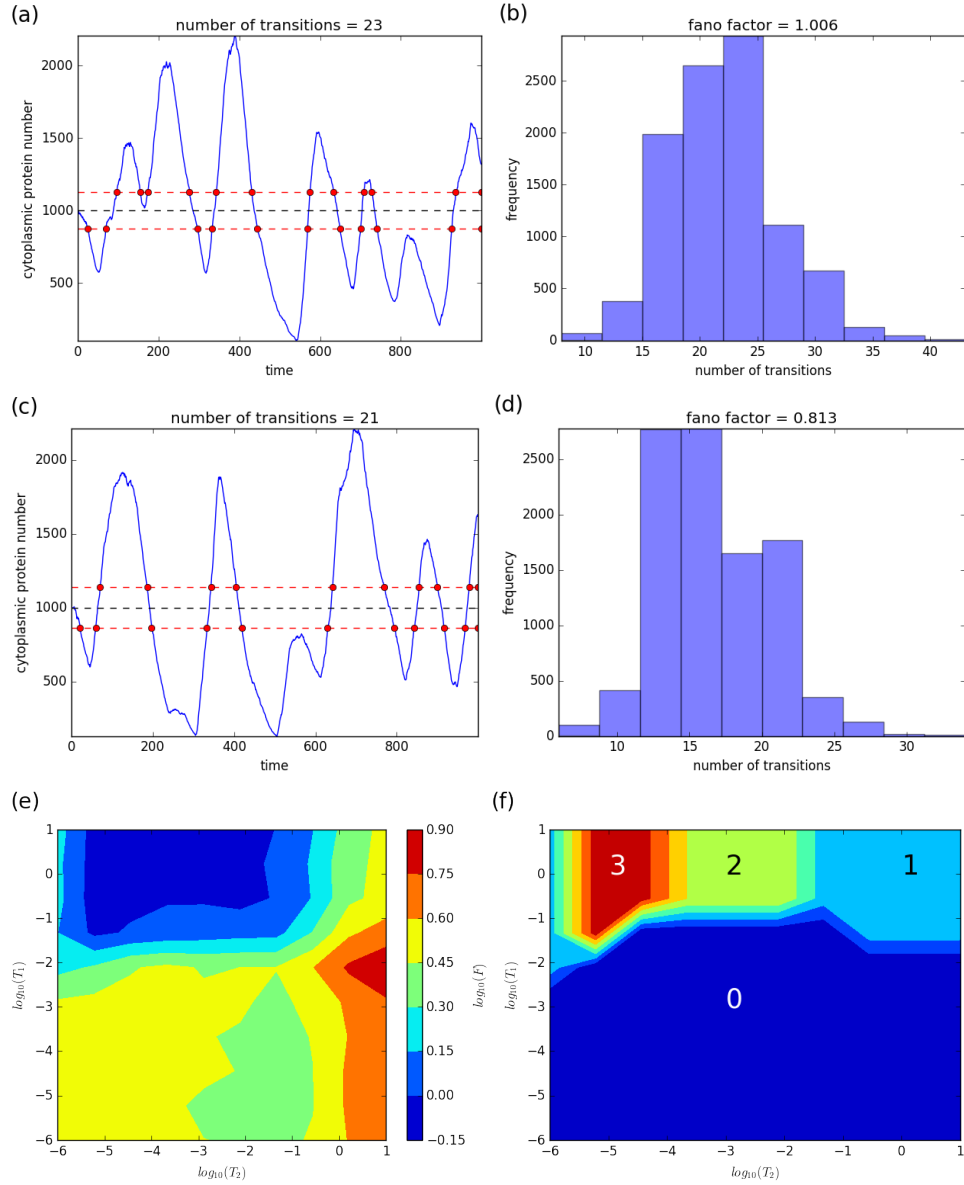


Figure 6: Plot (a) shows a cytoplasmic protein trajectory produced by the SSA for model 2 (blue solid line) along with the mean protein level, $\langle P \rangle$ (black dashed line) and $\langle P \rangle \pm 0.25\text{std}(P)$ (red dashed line). The red circles indicate the point of transitions. Plot (c) shows the corresponding plot for model 4 with $T_1 = 1\text{min}^{-1}$ and $T_2 = 1 \times 10^{-5}\text{min}^{-1}$. Plot (b) shows the probability distribution produced by recording the number of transitions from 10000 trajectories of the SSA for model 2 and plot (d) shows the corresponding plot for model 4. Plot (e) shows $\log_{10}$ of the Fano Factors produced from the distributions of number of transitions for different T_1 and T_2 values. Plot (f) shows the overlap between regions of the parameter space where noise in model 4 exceeds model 1 and spike anti-bunching is observed (3), only spike anti-bunching is observed (2), only model 4 noise exceeds model 1 (1) and neither spike anti-bunching or high noise in model 4 is observed (0). Parameters as per Table 2 with $d_2 = d_3 = d_4 = 0.03\text{min}^{-1}$.

3.6 Schizosaccharomyces pombe parameter study

Schizosaccharomyces pombe (S. Pombe), also called “fission yeast”, is a species of yeast which is used as a model organism in molecular and cell biology. We found values for mean mRNA levels (m), mean protein levels (p), mRNA decay rates (d_2) and protein decay rates (d_4) for 2390 different unique genes from experimental papers (Marguerat et al., 2012; Amorim et al., 2010; Christiano et al., 2014). These parameters and the correlations between them are displayed in Figure 7. It is clear from this figure that there exists a strong correlation between mean mRNA and protein levels and weak correlation between the other parameters.

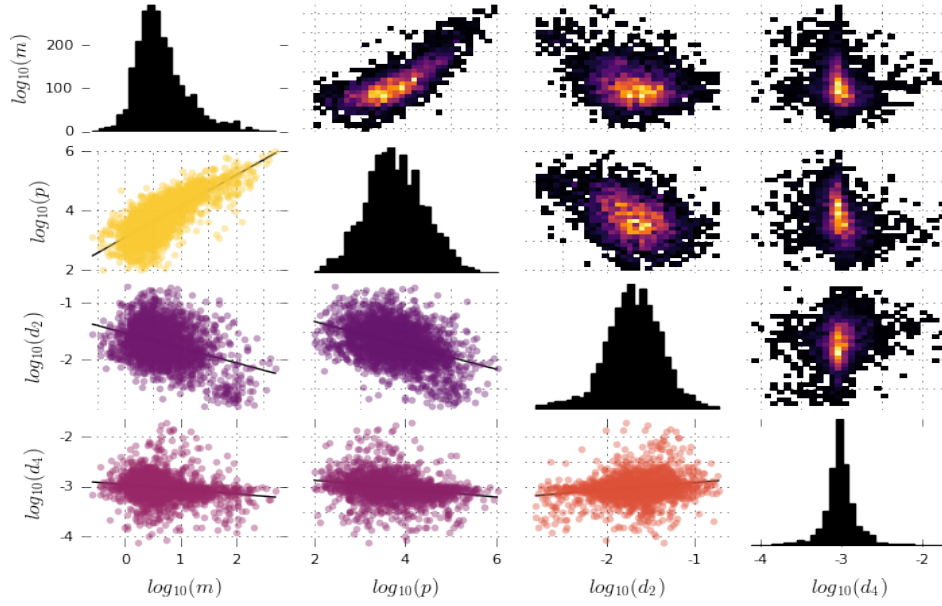


Figure 7: Correlation plot of 2390 unique S. Pombe parameter sets (Marguerat et al., 2012; Amorim et al., 2010; Christiano et al., 2014). Histograms are shown on diagonal panels and scatter plots comparing parameters are shown on off-diagonal panels.

We are interested in whether or not the results presented in sections 3.1 to 3.5 (the slow protein degradation case) and Appendix C (the fast protein degradation case) are sensitive to changes in parameters. To do this, we could perform a traditional parameter sensitivity analysis but this may lead us into unrealistic parameter regimes. Since we are particularly interested in realistic physiological parameters, we explore instead our noise comparison results using the above S. Pombe parameter sets. To do this, we compute the correlation distance between the model comparison results for slow and fast protein degradation cases and with those produced for the different S. Pombe parameter sets. I.e., we took the matrices used to create Figures 2-5 and Figures 12-13 and rearranged them into long vectors. We then created 2390 similar long vectors using different parameter values and computed the correlation distance (as implemented in the “Distances.jl” Julia package) between these vectors and those corresponding to the results presented in sections 3.1 to 3.5 and Appendix C. For two random variables X and Y the correlation distance is obtained by dividing their covariance distance by the product of their standard deviation

distances. I.e., the correlation distance is

$$\text{dCor}(X, Y) = \frac{\text{dCov}(X, Y)}{\sqrt{\text{dVar}(X) \text{dVar}(Y)}}. \quad (14)$$

If the value produced is close to zero, the results can be considered to be qualitatively similar. Results are summarised in the following Figure 8. We can observe negligible differences for the noise comparison results of model 3 vs. model 1 at the level of mRNA – which tells us that these results are particularly robust to changes in parameters. The other model comparisons have a similar structure, that is, the majority of parameter sets considered are close to the slow protein degradation case (distance less than 0.2) and a small fraction are close to the fast protein degradation case. We coloured the data points with the protein degradation rates which shows that the protein degradation rates dominate the results. However, it is possible to see a few genes for which the protein decay rate is relatively slow and the correlation distance exceeds 0.2.

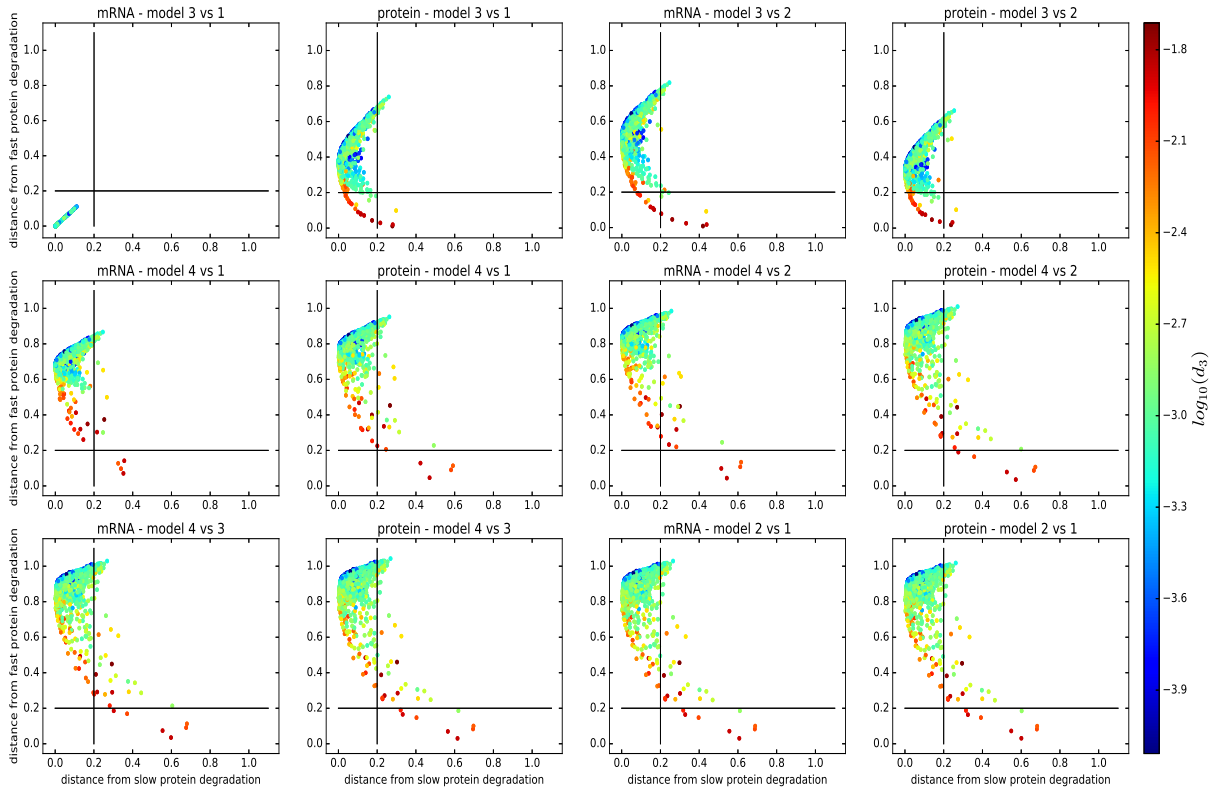


Figure 8: Distance correlation (computed using equation (14)) comparing noise comparison results for slow protein degradation case and fast protein degradation with 2390 unique *S. Pombe* parameter sets (Marguerat et al., 2012; Amorim et al., 2010; Christiano et al., 2014). Black lines highlight distances of 0.2 for slow and fast protein degradation cases. Colour bar shows the range of values found for the protein decay rate of the 2390 *S. Pombe* proteins.

4 Discussion

Gene regulation networks involving transcription factors are inherently spatial and stochastic. Mathematical models of gene circuits with negative feedback regulation have up until now been dominated by phenomenologically derived ODE and well-stirred stochastic models. While these well-stirred models may be appropriate for the study of prokaryotic cells, eukaryotic cells are more spatially segregated. In eukaryotic cells, mRNAs that are synthesized in the nucleus must be exported to the cytoplasm to function in protein synthesis, whereas proteins must be transported into the nucleus from the cytoplasm to regulate the expression levels of genes. This motivated us to compare and contrast stochastic gene expression models with and without negative feedback loops in both eukaryotic cells (where transport between the nucleus and cytoplasm is included) and prokaryotic cells. By studying four separate models of gene expression noise (see Figure 1) in a systematic manner, we were able to understand how the differences in model structure influence the noise produced.

We first wrote down the corresponding master equations describing the stochastic dynamics of each model (section 2.2). Since this master equation is neither analytically solvable nor computationally cheap to solve via stochastic simulation, we made use of a linear noise approximation. We verified the accuracy of this approximation by using long time simulations from the stochastic simulation algorithm for five different parameter regimes (see Appendix A). Using this linear noise approximation, we can solve the master equation cheaply and study a large portion of the translocation rate parameter space. We found that both the mRNA nuclear export rate and protein nuclear import rate had dramatic effects on the noise at both the mRNA and protein level.

It is known that both negative feedback and compartmentalisation can reduce gene expression noise (Savageau, 1974; Becskei and Serrano, 2000; Simpson et al., 2003; Swain, 2004; Boyer et al., 2005; Tao et al., 2007; Singh and Bokes, 2012; Bahar Halpern et al., 2015; Battich et al., 2015). The central question we sought to address in this paper was: what effect does the combination of both negative feedback and compartmentalisation have on gene expression noise? In the process of investigating the answer to this question we first found that negative feedback does not necessarily reduce mRNA noise. In particular, if the abundance of proteins is low and mRNA abundance is high, we can observe an increase in mRNA noise. Next we found that if mRNA export and protein import are fast, then compartmentalisation can increase protein noise. Hence, neither negative feedback nor compartmentalisation can be said to robustly reduce noise in every parameter regime. Finally, we found that if mRNA nuclear export is fast and protein nuclear import is slow, then negative feedback and compartmentalisation together can increase noise at both the mRNA and protein level. This suggests that negative feedback may be a less efficient method of reducing gene expression noise in eukaryotes compared to prokaryotes. Overall, our results suggest that eukaryotic cells have more control over noise levels than prokaryotic cells. By controlling the rate of influx and efflux of molecules from the nucleus, the nucleus can act as a noise tuner.

While we made every effort to choose our parameters to be physiologically realistic, we wanted to explore how sensitive our noise comparison results were to changes in parameter values. A parameter sensitivity analysis may stray into physiologically unrealistic parameter regimes, so instead we gathered 2390 realistic parameter sets corresponding to different *S. Pombe* genes. We then computed the corresponding noise comparison results for each of these parameter sets. In order to judge how far they were from the original results presented, we then used a correlation distance metric which allowed us to gauge how sensitive our results were to using these different values. We found that our results were, in general, in agreement with those presented in the main paper, and those parameter sets that were not in agreement

were instead close to the results presented in Appendix C.

In Appendix C we explored the noise properties of fast decaying proteins. We found compartmentalisation is more effective in reducing noise in fast decaying proteins. In the case of fast decaying protein, we also attempted to look for oscillations in our models with negative feedback but failed to find any significant peaks in the power spectral density. This confirms the notion that some nonlinearity in gene regulation is needed for observation of oscillatory behaviour. However, motivated by Wang et al. (2014), we investigated the existence of stochastic oscillations or spike anti-bunching. For our choice of parameter set, in the well mixed negative feedback model (model 2), we failed to find conditions necessary for stochastic oscillations but in certain translocation rate regimes, we found that the negative feedback model with nuclear compartmentalisation (model 4) could produce stochastic oscillations, suggesting that transportation processes can facilitate stochastic oscillatory behaviour.

These results are relevant to the case of the Hes1 gene regulatory network, a known eukaryotic self repressing transcription factor. It has been shown that noise in Hes1 expression can lead to heterogeneous differentiation responses (Kobayashi et al., 2009). It was argued by the authors that oscillatory dynamics contributed to this (Kobayashi and Kageyama, 2011), however, the experimental data shows time series consistent with that which we present in Figure 6 (a), i.e., highly noisy but not exhibiting stochastic oscillations. Hence, it may be the case that the Hes1 mRNA and protein molecules are in the fast mRNA export, slow protein import regime we have identified which produces highly noisy gene expression. To test this idea further, we found a more recent study of Hes1 by Phillips et al. (2016) which contains time series data for Hes1 luminescence levels for 15 cells. Using these time series, we applied the same criteria for determining whether or not stochastic oscillations are observed as we did in section 3.6 and the results are shown in Figure 9. We found that the data show superpoissonian transition times, consistent with the idea that Hes1 does not exhibit stochastic oscillations but instead is expressed in a highly noisy manner.

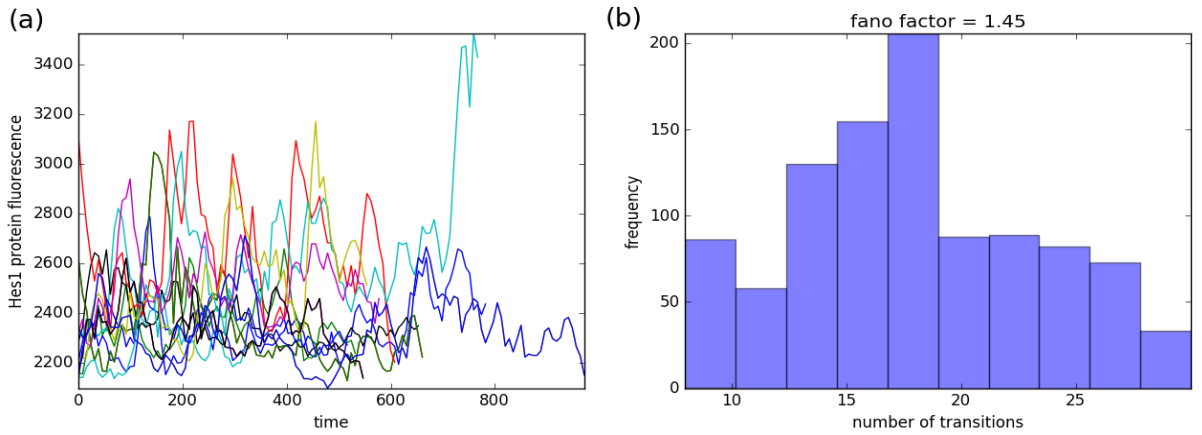


Figure 9: (a) 15 time series examples of single-cell experimental data showing Hes1 protein fluorescence levels from Phillips et al. (2016). (b) corresponding distribution of transition counts (as described in section 3.6) showing that the data is not consistent with the idea of stochastic oscillations.

In addition to studying naturally occurring negative feedbacks like Hes1, there has also been some interest in synthetically engineering negative feedback loops in mammalian cells (Tigges et al., 2009) with the aim of observing oscillatory dynamics. Our results suggests proper consideration of nuclear

transport is important for effective design of such synthetic networks. Also, the role played by nuclear compartmentalisation highlights the fact that design principles that have emerged from system and synthetic biology studies in prokaryotes may not be readily transferable to the eukaryotic systems.

Although we have incorporated some spatial structure of eukaryotic cells in our model, individual molecule free diffusion within the cells is ignored. Some studies (Cottrell et al., 2012) have demonstrated that diffusion can affect the complexity of fluctuations in biochemical networks, while other studies have mathematically predicted that diffusion of transcription factors can significantly enhance the noise in gene expression (van Zon et al., 2006), and Sturrock et al. (2014b) suggest that the spatially homogeneous approximations for bimolecular numbers do not agree with a one dimensional spatial model when the diffusion coefficients of molecules are slow. Therefore a further direction of our work is incorporating diffusive motion of mRNAs and proteins in our model, and studying how diffusion and molecular crowding affect the level of noise in eukaryotic gene expression, and exploring the difference between stochastic behaviours of the same biochemical network with compartments of different size. Also, the feedback mechanism presented in our model can be described as transcriptional auto-regulation. Within cells, however, there exist various types of feedback mechanisms, and their influence on noise can vary dramatically. Another further direction is therefore to include more possible feedback loops into our models of gene regulatory networks, and then study the effects of altering nuclear transport rates on noise in gene expression. In this study we have focused on intrinsic noise. Several studies have looked at the effect of feedback systems on extrinsic noise (Shahrezaei and Swain, 2008a; de Franciscis et al., 2016; Caravagna et al., 2013; Assaf et al., 2013). In particular, it has been shown that negative feedback is effective in reducing extrinsic noise as well as intrinsic noise. We predict that nuclear retention is effective in reducing extrinsic noise affecting transcription but not translation. However, a full investigation of the interplay between nuclear transport and extrinsic noise should be performed in a future study. Finally, it is known that genes involved in autoregulatory feedback loops often must form dimers before they can bind to their own promoters (such as Hes, Her). Though there has been some progress made in elucidating the role of dimerisation in autoregulatory feedback loops (Sturrock et al., 2014a), a systematic study of the impact of dimerisation on noise in such gene regulatory networks could prove fruitful.

Acknowledgements

VS would like to thank teaching replacement support from Leverhulme Grant RPG-2014-408.

Parameter set	T_1	T_2
1	2.5 min^{-1}	$3.00 \times 10^{-5} \text{ min}^{-1}$
2	$3.00 \times 10^{-5} \text{ min}^{-1}$	$3.00 \times 10^{-5} \text{ min}^{-1}$
3	2.5 min^{-1}	2.5 min^{-1}
4	$3.00 \times 10^{-5} \text{ min}^{-1}$	2.5 min^{-1}
5	$5.00 \times 10^{-3} \text{ min}^{-1}$	$5.00 \times 10^{-3} \text{ min}^{-1}$

Table 3: Different parameter sets used to check the accuracy of the LNA against the SSA

Appendix A: Checking accuracy of the linear noise approximation using the stochastic simulation algorithm

Although, closed form analytical expressions are possible to obtain in some cases, we use the linear noise equations to numerically compute the noise in our different models for different parameter values. The LNA can be computed much faster than the stochastic simulation algorithm. In Figure 10, we show how accurate it is for each model for different regions of the parameter set as defined in Table 3. We plot the result of computing the CV^2 from a long time trajectory computed by the SSA (20,000 minutes) using the Gillespie algorithm (Gillespie, 1976) along with the approximation from the LNA. The error bar represents the coefficient of variation produced by the cumulative CV^2 from the long time trajectory. The LNA appears to produce a good approximation of the SSA result in all cases.

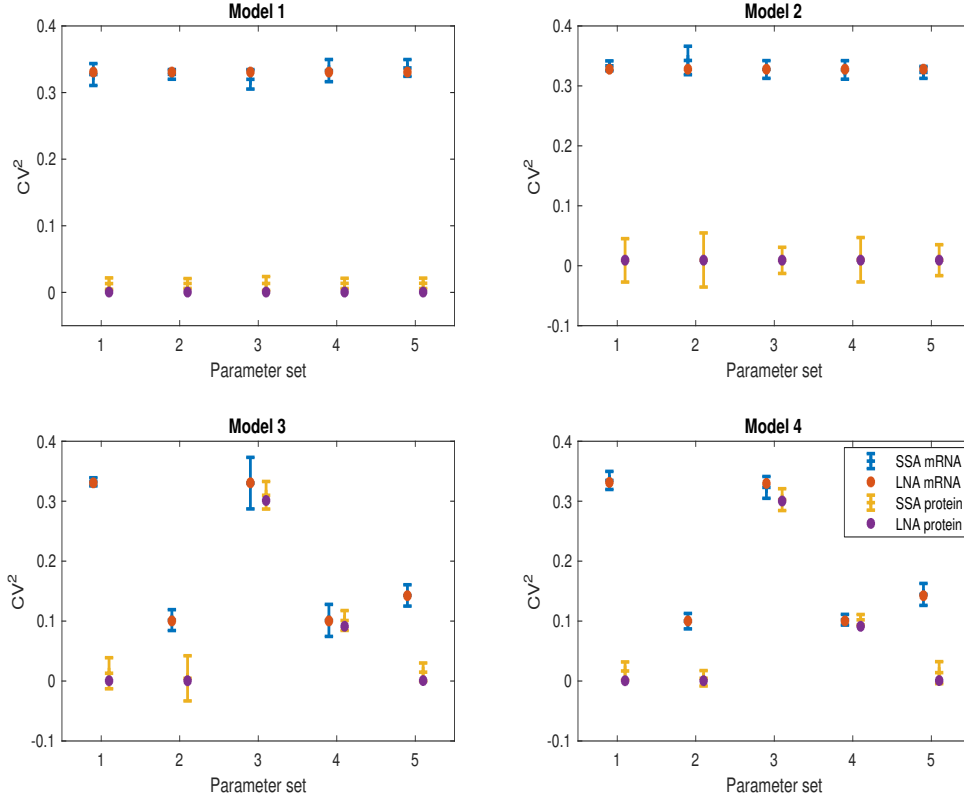


Figure 10: Plots showing comparison of LNA and SSA. The lines with error bars show the results from a long time simulation of the SSA (blue for mRNA and yellow for protein) and the filled circles show the corresponding approximation produced by the LNA (red for mRNA and purple for protein). The LNA results lie within the error bars for all cases. Parameters are as per Table 2 with T_1 and T_2 defined in Table 2.

Appendix B: Dependence of variable parameters on transport parameters

We show here how the abundance of nuclear mRNA and nuclear protein in the eukaryotic cell models are influenced by the variation in the transport parameters. As is shown in Figure 11, we consider very broad ranges which cover 7 orders of magnitude - this encapsulates experimental estimates of nuclear export rates which are in the range 10^{-4} to 10^1 min^{-1} (Bahar Halpern et al., 2015). Unsurprisingly, decreasing nuclear mRNA export (T_1) results in an increase in abundance of nuclear mRNA and increasing nuclear protein import (T_2) results in an increase in nuclear protein abundance. The transcription rate, k_1 , remains constant for different values of T_1 and T_2 . In fact it can be shown to be of the form $k_1 = \frac{d_2 m_2}{g}$, which equals 1 for the values in Table 2. If the nuclear degradation rate is non-zero, k_1 becomes dependent on T_1 , specifically in the form $k_1 = \frac{d_2 m_2 (T_1 + d_1)}{g T_1}$. Plot (d) shows how the translation rate k_2 increases with the protein nuclear import rate. As the nuclear import rate is increased, more protein must be created to keep the cytoplasmic number constant.

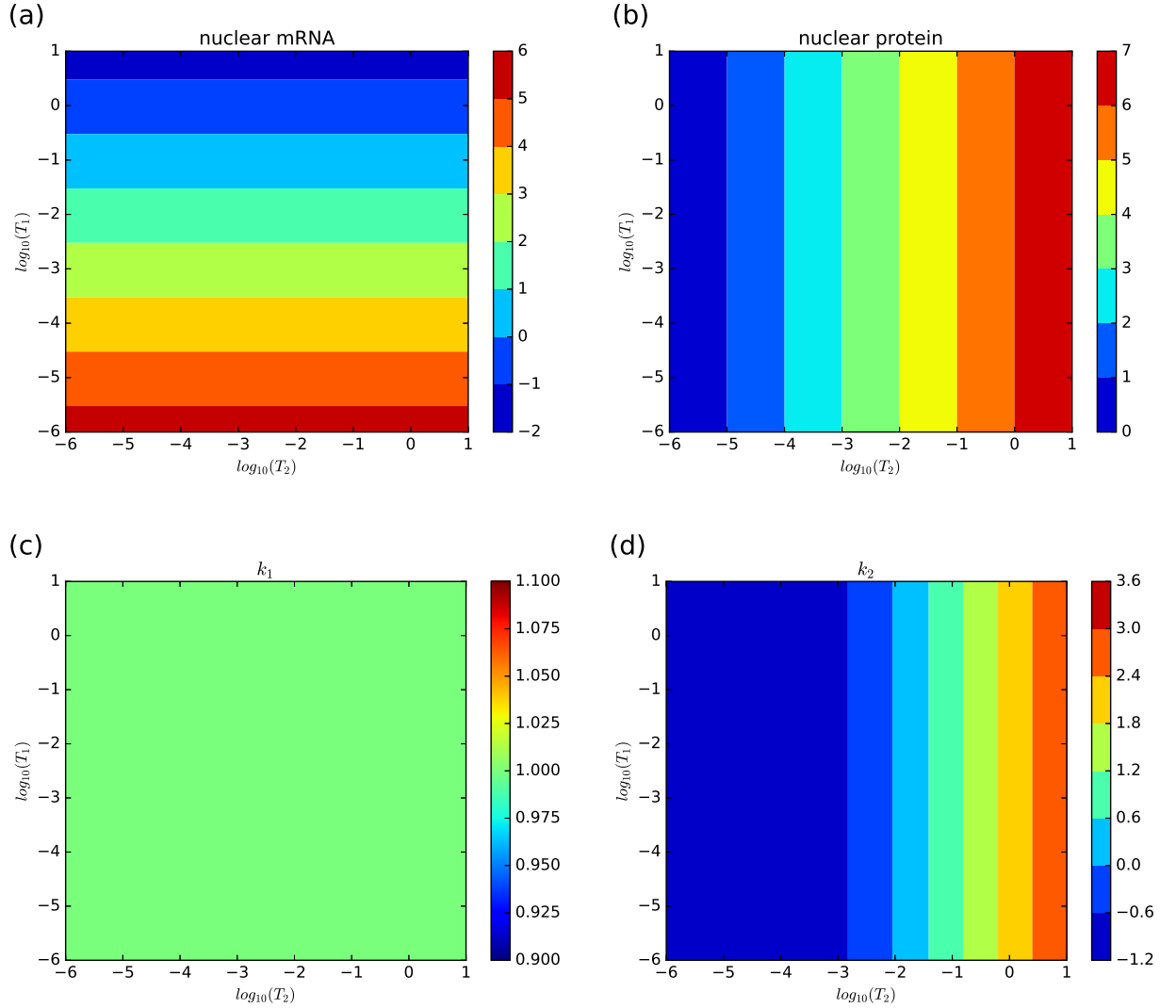


Figure 11: For different values of T_1 and T_2 , plot (a) shows the mean nuclear mRNA levels, plot (b) shows the corresponding mean protein levels, plot (c) shows the corresponding transcription rates, k_1 and plot (d) shows how the the translation rate k_2 varies. Parameters as per Table 2.

Appendix C: Relative noise plots for fast degradation regime

In this section, we show how the relative noise plots shown in the main paper differ for the case where $d_2 = d_3 = d_4 = 0.03 \text{ min}^{-1}$ as is the case for the Hes1 gene regulatory network (Hirata et al., 2002).

Appendix D: Sensitivity analysis of the nuclear degradation rate of mRNA

In the main paper and in the previous Appendix, we assumed that the nuclear degradation rate of mRNA (d_1) was zero in order to be consistent with some previous studies. Here, we explore how sensitive our results are to this assumption. To do this, we chose 500 values in the range of $(0, d_2)$ and computed

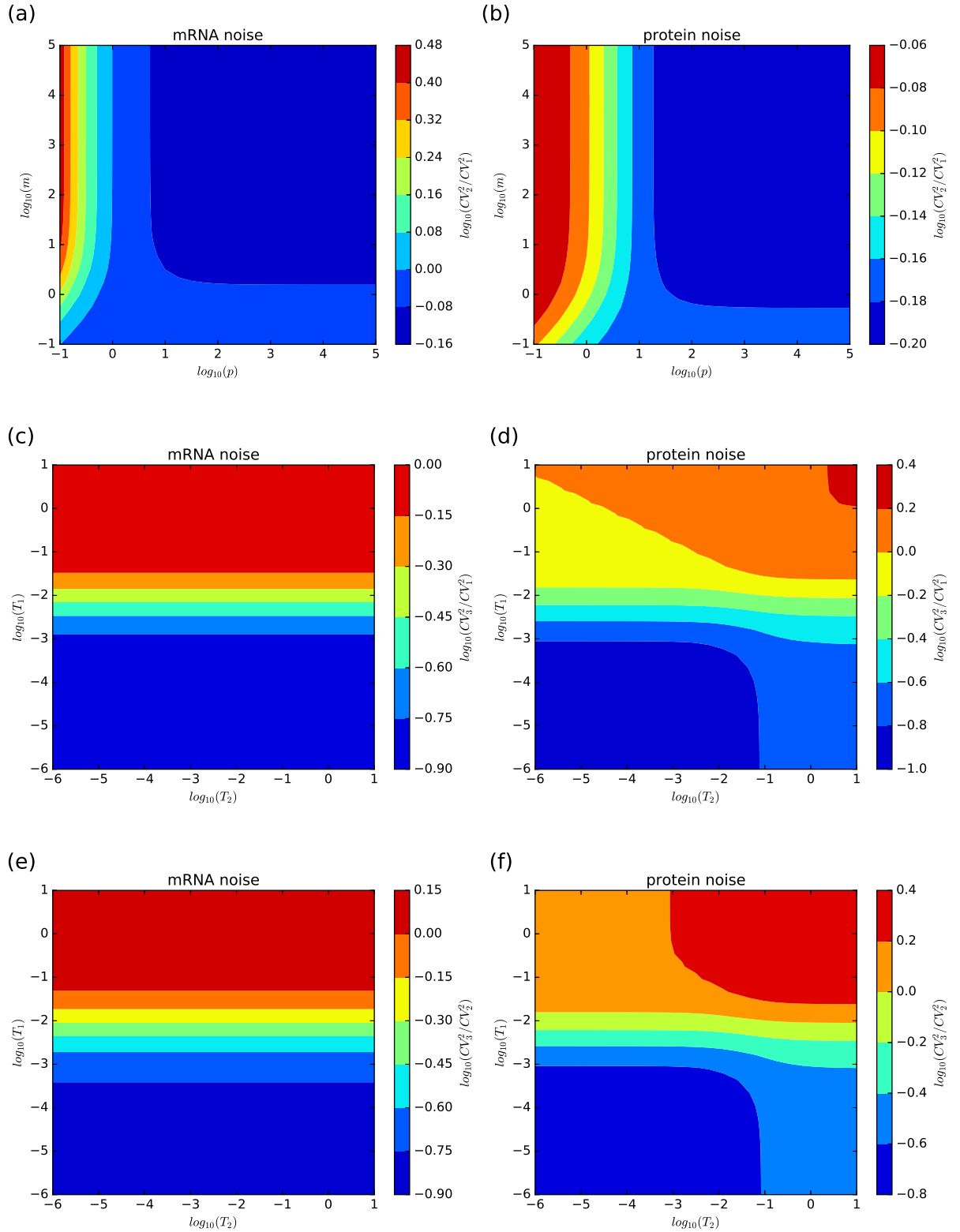


Figure 12: Comparison of relative noise produced by models 1 and 2 (plots (a) and (b)), 1 and 3 (plots (c) and (d)), and 2 and 3 (plots (e) and (f)) for different mRNA export and protein import rates. Plots (a), (c) and (e) show relative mRNA noise and plots (b), (d) and (f) show relative protein noise where relative noise is computed using equation (1). Parameters as per Table 2 with $d_2 = d_3 = d_4 \equiv 0.03 \text{ min}^{-1}$.

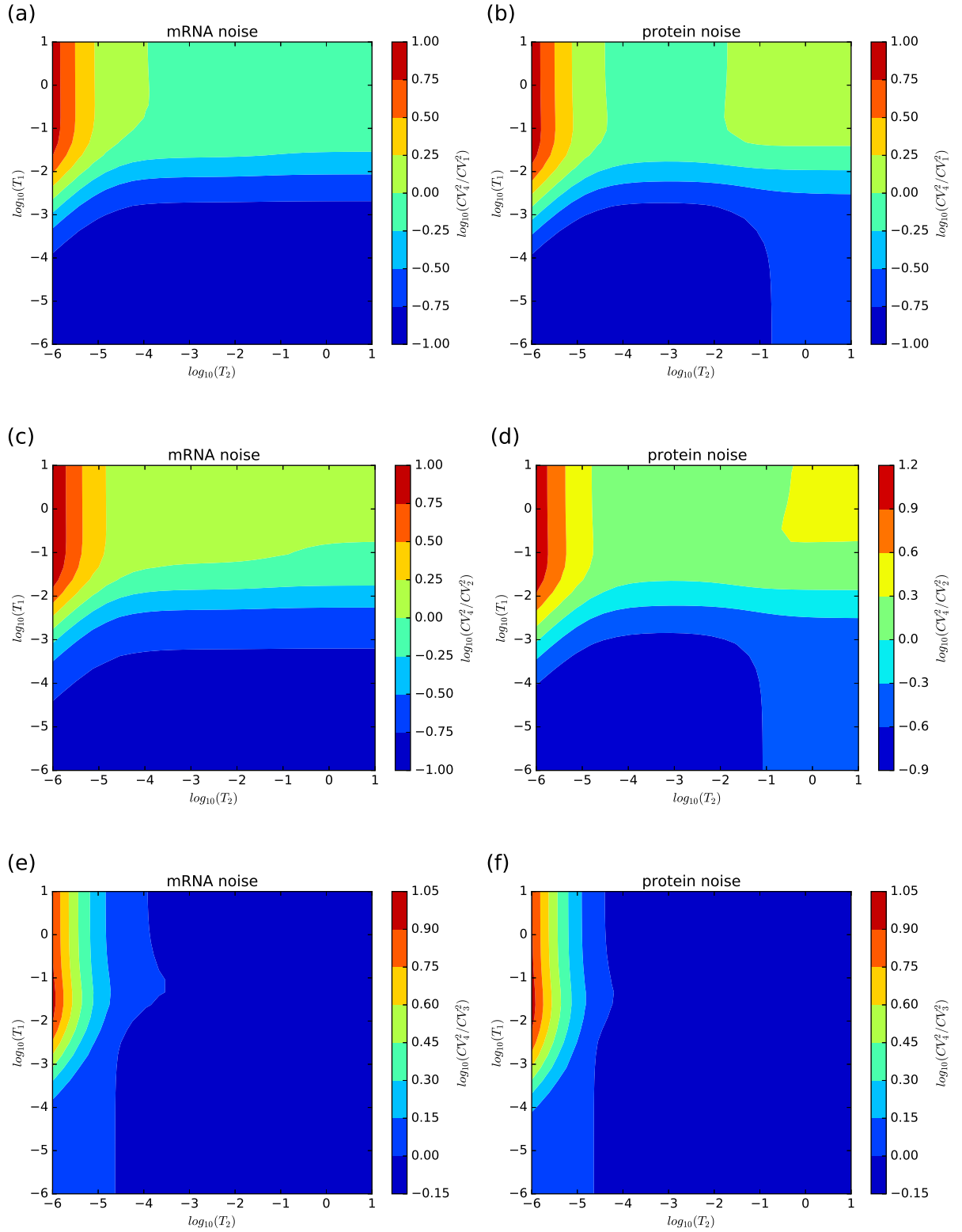


Figure 13: Comparison of relative noise produced by models 1 and 4 (plots (a) and (b)), 2 and 4 (plots (c) and (d)), and 3 and 4 (plots (e) and (f)) for different mRNA export and protein import rates. Plots (a), (c) and (e) show relative mRNA noise and plots (b), (d) and (f) show relative protein noise where relative noise is computed using equation (1). Parameters as per Table 2 with $d_2 = d_3 = d_4 = 0.03\text{min}^{-1}$.

the correlation distance between the model comparison results presented in the main paper with those produced for the different values of d_1 . As we did with the *S. Pombe* parameter study, we use the distance correlation metric and also check to see if including nuclear degradation affects the fast protein degradation results considered in the previous Appendix section. Results are summarised in the following Figure 14. As we found for the *S. Pombe* parameter study, the model 3 vs. model 1 noise comparison is particularly robust to changes in the nuclear mRNA degradation rate. We can also observe negligible differences until d_1 is close to the value of d_2 and we can also see that the slow protein degradation case (considered in the main paper) is less sensitive to changing the nuclear degradation rate of mRNA than the fast protein degradation case.

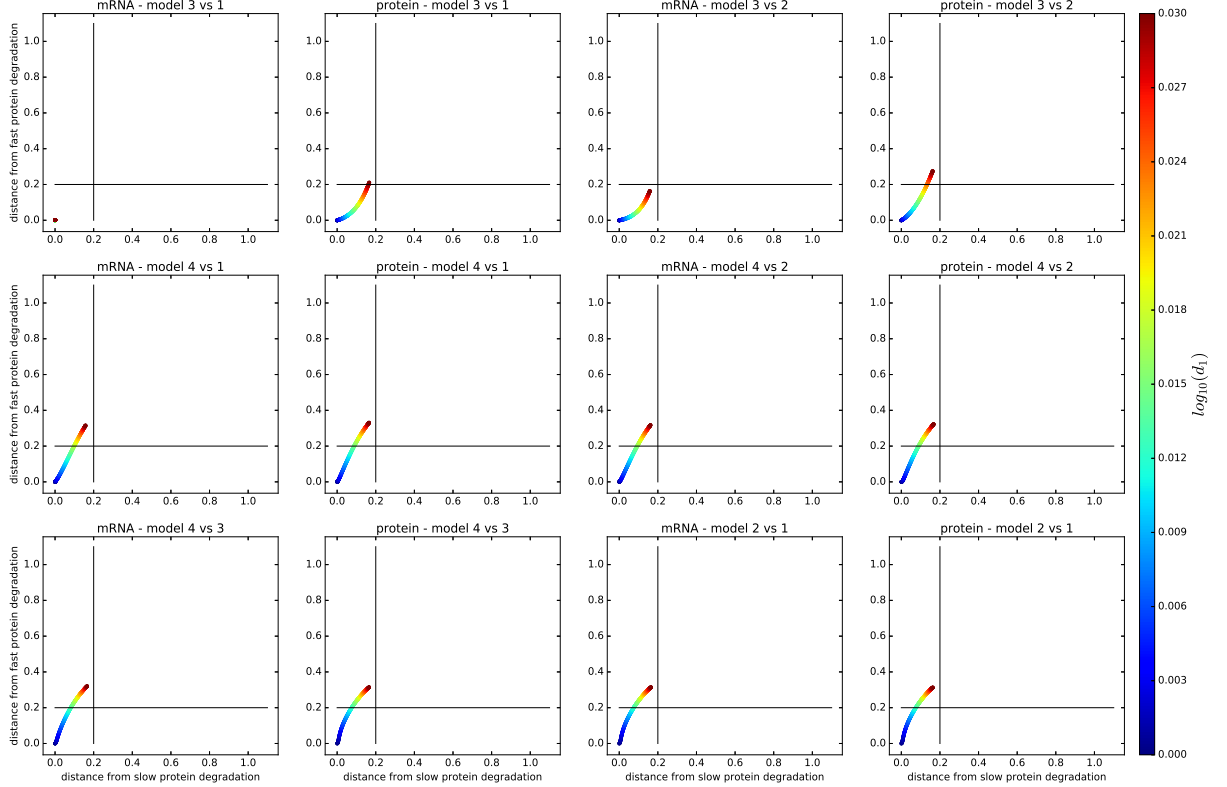


Figure 14: Distance correlation (computed using equation (14)) comparing noise comparison results for the slow protein degradation case and fast protein degradation case with noise comparison results produced using 500 nuclear degradation rates in the range $(0, d_2)$ for the same models. Black lines highlight distances of 0.2 for slow and fast protein degradation cases. Colour bar shows the range of values explored for the nuclear degradation rate (d_1).

References

- Acar, M., Mettetal, J. T., and van Oudenaarden, A. (2008). Stochastic switching as a survival strategy in fluctuating environments. *Nat. Genet.*, 40(4):471–475.
- Albert, J. and Rومان, M. (2015). Is the cell nucleus a necessary component in precise temporal patterning? *PloS one*, 10(7):e0134239.
- Alberts, B., Johnson, A., Lewis, J., Raff, M., Roberts, K., and Walter, P. (2008). *Molecular Biology of the Cell*. Garland Science, Taylor and Francis Group Ltd, Oxford, fifth edition.
- Amorim, M., Cotobal, C., Duncan, C., and Mata, J. (2010). Global coordination of transcriptional control and mrna decay during cellular differentiation. *Molecular systems biology*, 6.
- Assaf, M., Roberts, E., Schulten, Z., and Goldenfeld, N. (2013). Extrinsic noise driven phenotype switching in a self-regulating gene. *Physical Review Letters*, 111:058102.
- Bahar Halpern, K., Caspi, I., Lemze, D., Levy, M., Landen, S., Elinav, E., Ulitsky, I., and Itzkovitz, S. (2015). Nuclear Retention of mRNA in Mammalian Tissues. *Cell Rep.*, 13(12):2653–2662.
- Barrio, M., Burrage, K., Leier, A., and Tian, T. (2006). Oscillatory regulation of hes1: Discrete stochastic delay modelling and simulation. *PLoS ONE*, 2:e117.
- Battich, N., Stoeger, T., and Pelkmans, L. (2015). Control of Transcript Variability in Single Mammalian Cells. *Cell*, 163(7):1596–1610.
- Becskei, A. and Serrano, L. (2000). Engineering stability in gene networks by autoregulation. *Nature*, 405:590–593.
- Bezanson, J., Edelman, A., Karpinski, S., and Shah, V. B. (2014). Julia: A fresh approach to numerical computing. *arXiv preprint arXiv:1411.1607*.
- Blake, W., Kaern, M., Cantor, C., and Collins, J. J. (2003). Noise in eukaryotic gene expression. *Nature*, 422:633–637.
- Boyer, L., Lee, T., Cole, M., Johnstone, S., Levine, S., Zucker, J., Guenther, M., Kumar, R., Murray, H., Jenner, R., Gifford, D., Melton, D., Jaenisch, R., and Young, R. (2005). Core transcriptional regulatory circuitry in human embryonic stem cells. *Cell*, 122:947–956.
- Busenberg, S. and Mahaffy, J. M. (1985). Interaction of spatial diffusion and delays in models of genetic control by repression. *J. Math. Biol.*, 22:313–333.
- Caravagna, G., Mauri, G., and d’Onofrio, A. (2013). The interplay of intrinsic and extrinsic bounded noises in biomolecular networks. *PLoS One*, 8(2):e51174.
- Christiano, R., Nagaraj, N., Fröhlich, F., and Walther, T. C. (2014). Global proteome turnover analyses of the yeasts *s. cerevisiae* and *s. pombe*. *Cell reports*, 9(5):1959–1965.
- Cooper, G. M. and Hausman, R. E. (2004). *The Cell: A Molecular Approach*. ASM Press.
- Cottrell, D., Swain, P., and Tupper, P. (2012). Stochastic branching-diffusion models for gene expression. *PNAS*, 109:9699–9704.

- de Franciscis, S., Caravagna, G., Mauri, G., and d’Onofrio, A. (2016). Gene switching rate determines response to extrinsic perturbations in the self-activation transcriptional network motif. *Scientific reports*, 6:26980.
- Elf, J. and Ehrenberg, M. (2003). Fast evaluation of fluctuations in biochemical networks with the linear noise approximation. *Genome Res.*, 13(11):2475–2484.
- Elowitz, M., Levine, A., Siggia, E., and Swain, P. (2002). Stochastic gene expression in a single cell. *Science*, 297:1183–1186.
- Fall, C. P., Marland, E. S., Wagner, J. M., and Tyson, J. J. (2002). *Computational Cell Biology*. Springer, New York.
- Geva-Zatorsky, N., Dekel, E., Batchelor, E., Lahav, G., and Alon, U. (2010). Fourier analysis and systems identification of the p53 feedback loop. *Proc. Natl. Acad. Sci. USA*, 107:13550–13555.
- Gillespie, D. T. (1976). A general method for simulating the stochastic time evolution of coupled chemical reactions. *J. Comput. Phys.*, 22:403–434.
- Goodwin, B. C. (1965). Oscillatory behavior in enzymatic control processes. *Adv. Enzym. Regul.*, 3:425–428.
- Greenbaum, D., Colangelo, C., Williams, K., and Gernstein, M. (2003). Comparing protein abundance and mRNA expression levels on a genomic scale. *Genome Biol*, 4:117.
- Griffith, J. S. (1968). Mathematics of cellular control processes. i. negative feedback to one gene. *J. Theor. Biol.*, 20:202–208.
- Hirata, H., Yoshiura, S., Ohtsuka, T., Bessho, Y., Harada, T., K.Yoshikawa, and Kageyama, R. (2002). Oscillatory expression of the bhlh factor hes1 regulated by a negative feedback loop. *Science*, 298:840–843.
- Jensen, M. H., Sneppen, J., and Tiana, G. (2003). Sustained oscillations and time delays in gene expression of protein Hes1. *FEBS Lett.*, 541:176–177.
- Kageyama, R., Ohtsuka, T., and Kobayashi, T. (2007). The Hes1 gene family: repressors and oscillators that orchestrate embryogenesis. *Development*, 134:1243–1251.
- Kobayashi, T. and Kageyama, R. (2010). Hes1 regulates embryonic stem cell differentiation by suppressing Notch signaling. *Genes to Cells*, 15:689–698.
- Kobayashi, T. and Kageyama, R. (2011). Hes1 Oscillations Contribute to Heterogeneous Differentiation Responses in Embryonic Stem Cells. *Genes (Basel)*, 2:219–228.
- Kobayashi, T., Mizuno, H., Imayoshi, I., Furusawa, C., Shirahige, K., and Kageyama, R. (2009). The cyclic gene Hes1 contributes to diverse differentiation responses of embryonic stem cells. *Genes & Dev.*, 23:1870–1875.
- Lahav, G., Rosenfeld, N., Sigal, A., Geva-Zatorsky, N., Levine, A. J., Elowitz, M. B., and Alon, U. (2004). Dynamics of the p53-mdm2 feedback loop in individual cells. *Nat. Genet.*, 36:147–150.

- Lee, T., Rinaldi, N., Robert, F., Odom, D., Bar-Joseph, Z., Gerber, G., Hannett, N., Harbison, C., Thompson, C., Simon, I., Zeitlinger, J., Jennings, E., Murray, H., Gordon, B., Ren, B., Wyrick, J., Tagne, J., Volkert, T., Fraenkel, E., Gifford, D., and Young, R. (2002). Transcriptional Regulatory Networks in *Saccharomyces cerevisiae*. *Science*, 298:799–804.
- Lestas, I., Vinnicombe, G., and Paulsson, J. (2010). Fundamental limits on the suppression of molecular fluctuations. *Nature*, 467(7312):174–178.
- Mahaffy, J. M. (1988). Genetic control models with diffusion and delays. *Math. Biosci.*, 90:519–533.
- Mahaffy, J. M. and Pao, C. V. (1984). Models of genetic control by repression with time delays and spatial effects. *J. Math. Biol.*, 20:39–57.
- Marguerat, S., Schmidt, A., Codlin, S., Chen, W., Aebersold, R., and Bähler, J. (2012). Quantitative analysis of fission yeast transcriptomes and proteomes in proliferating and quiescent cells. *Cell*, 151(3):671–683.
- Monk, N. A. M. (2003). Oscillatory Expression of Hes1, p53, and NF- κ B driven by transcriptional time delays. *Curr. Biol.*, 13:1409–1413.
- Nelson, D. E., Ihekweba, A. E. C., Elliott, M., Johnson, J. R., Gibney, C. A., Foreman, B. E., Nelson, G., See, V., Horton, C. A., Spiller, D. G., Edwards, S. W., McDowell, H. P., Unitt, J. F., Sullivan, E., Grimley, R., Benson, N., Broomhead, D., Kell, D. B., and White, M. R. H. (2004). Oscillations in NF- κ B signaling control the dynamics of gene expression. *Science*, 306:704–708.
- Oyarzún, D., Lugagne, J., and Stan, G. (2015). Noise Propagation in Synthetic Gene Circuits for Metabolic Control. *ACS Synth. Biol.*, 4(2):116–125.
- Phillips, N. E., Manning, C. S., Pettini, T., Biga, V., Marinopoulou, E., Stanley, P., Boyd, J., Bagnall, J., Paszek, P., Spiller, D. G., et al. (2016). Stochasticity in the mir-9/hes1 oscillatory network can account for clonal heterogeneity in the timing of differentiation. *eLife*, 5:e16118.
- Raj, A. and van Oudenaarden, A. (2008). Nature, Nurture, or Chance: Stochastic Gene Expression and Its Consequences.
- Raser, J. M. and O’Shea, E. K. (2005). Noise in Gene Expression: Orgins, Consequences, and Control. *Science*, 309(5743):2010–2013.
- Rosenfeld, N., Elowitz, M., and Alon, U. (2002). Negative autoregulation speeds the response times of transcription networks. *J. Mol. Biol.*, 323:785–793.
- Savageau, M. A. (1974). Comparison of classical and autogenous systems of regulation in inducible operons. *Nature*, 252(5484):546–549.
- Shahrezaei, V., Ollivier, J. F., and Swain, P. S. (2008). Colored extrinsic fluctuations and stochastic gene expression. *Mol. Syst. Biol.*, 4:1–9.
- Shahrezaei, V. and Swain, P. S. (2008a). Analytical distributions for stochastic gene expression. *Proc. Natl. Acad. Sci. U. S. A.*, 105(45):17256–61.

- Shahrezaei, V. and Swain, P. S. (2008b). The stochastic nature of biochemical networks. *Curr. Opin. Biotechnol.*, 19:369–374.
- Simpson, M. L., Cox, C. D., and Saylor, G. S. (2003). Frequency domain analysis of noise in autoregulated gene circuits. *Proc. Natl. Acad. Sci. U. S. A.*, 100(8):4551–4556.
- Singh, A. and Bokes, P. (2012). Consequences of mRNA transport on stochastic variability in protein levels. *Biophys. J.*, 103(5):1087–1096.
- Singh, A. and Hespanha, J. P. (2009). Optimal feedback strength for noise suppression in autoregulatory gene networks. *Biophys. J.*, 96:4013–4023.
- Singh, A., Razooky, B. S., Dar, R. D., and Weinberger, L. S. (2012). Dynamics of protein noise can distinguish between alternate sources of gene-expression variability. *Mol. Syst. Biol.*, 8:1–9.
- Smith, S. and Shahrezaei, V. (2015). General transient solution of the one-step master equation in one dimension. *Phys. Rev. E*, 91(6):62119.
- Sturrock, M., Hellander, A., Aldakheel, S., Petzold, L., and Chaplain, M. A. J. (2014a). The Role of Dimerisation and Nuclear Transport in the Hes1 Gene Regulatory Network. *Bull. Math. Biol.*, 76:766–798.
- Sturrock, M., Hellander, A., Matzavinos, A., and Chaplain, M. A. J. (2013). Spatial stochastic modelling of the Hes1 gene regulatory network: intrinsic noise can explain heterogeneity in embryonic stem cell differentiation. *J. R. Soc. Interface*, 10(80):20120988.
- Sturrock, M., Murray, P. J., Matzavinos, A., and Chaplain, M. a. J. (2014b). Mean field analysis of a spatial stochastic model of a gene regulatory network. *J. Math. Biol.*, 71:921–959.
- Sturrock, M., Terry, A. J., Xirodimas, D. P., Thompson, A. M., and Chaplain, M. A. J. (2011). Spatio-temporal modelling of the Hes1 and p53-Mdm2 intracellular signalling pathways. *J. Theor. Biol.*, 273(1):15–31.
- Sturrock, M., Terry, A. J., Xirodimas, D. P., Thompson, A. M., and Chaplain, M. A. J. (2012). Influence of the nuclear membrane, active transport and cell shape on the hes1 and p53–mdm2 pathways: Insights from spatio-temporal modelling. *Bull. Math. Biol.*, 74:1531–1579.
- Swain, P., Elowitz, M., and Siggia, E. (2002). Intrinsic and extrinsic contributions to stochasticity in gene expression. *PNAS*, 99:12795–12800.
- Swain, P. S. (2004). Efficient attenuation of stochasticity in gene expression through post-transcriptional control. *J. Mol. Biol.*, 344(4):965–976.
- Tao, Y., Zheng, X., and Sun, Y. (2007). Effect of feedback regulation on stochastic gene expression. *J. Theor. Biol.*, 247:827–836.
- Terry, A. J., Sturrock, M., Dale, J. K., Maroto, M., and Chaplain, M. A. J. (2011). A spatio-temporal model of Notch signalling in the zebrafish segmentation clock: Conditions for synchronised oscillatory dynamics. *PLoS One*, 6:e16980.

- Thieffry, D., Huerta, A. M., Pérez-Rueda, E., and Collado-Vides, J. (1998). From specific gene regulation to genomic networks: a global analysis of transcriptional regulation in *Escherichia coli*. *BioEssays*, 20:433–440.
- Tiana, G., Jensen, M. H., and Sneppen, K. (2002). Time delay as a key to apoptosis induction in the p53 network. *Eur. Phys. J. B*, 29:135–140.
- Tigges, M., Marquez-Lago, T., Stelling, J., and Fussenegger, M. (2009). A tunable synthetic mammalian oscillator. *Nature*, 457(7227):309–312.
- Van Kampen, N. G. and Reinhardt, W. P. (2007). *Stochastic Processes in Physics and Chemistry*. AIP.
- van Zon, J. S., Morelli, M. J., Tvanase-Nicola, S., and ten Wolde, P. R. (2006). Diffusion of transcription factors can drastically enhance the noise in gene expression. *Biophys. J.*, 91(12):4350–4367.
- Wang, J., Lefranc, M., and Thommen, Q. (2014). Stochastic oscillations induced by intrinsic fluctuations in a self-repressing gene. *Biophys. J.*, 107(10):2403–2416.
- Zhang, J., Yuan, Z., and Zhou, T. (2009). Physical limits of feedback noise-suppression in biological networks. *Phys. Biol.*, 6(4):046009.