

# Probing Inter-Cell Variability using Bulk Measurements

Harrison Steel<sup>†</sup> and Antonis Papachristodoulou<sup>†</sup>

**Abstract**—The measurement of noise is critical when assessing the design and function of synthetic biological systems. Cell-to-cell variability can be quantified experimentally using single-cell measurement techniques such as flow cytometry and fluorescent microscopy. However, these approaches are costly and impractical for high-throughput parallelised experiments, which are frequently conducted using plate-reader devices. In this paper we describe reporter systems that allow estimation of the cell-to-cell variability in a biological system’s output using only measurements of a cell culture’s bulk properties. We analyse one potential implementation of such a system which is based upon a fluorescent protein FRET reporter pair, finding that with typical parameters from the literature it is able to reliably estimate variability. We also briefly describe an alternate implementation based upon an activating sRNA circuit. The feasible region of parameter values for which the reporter system can function is assessed, and the dependence of its performance on both extrinsic and intrinsic noise is investigated. Experimental realisation of these constructs can yield novel reporter systems that allow measurement of a synthetic gene circuit’s output, as well as the intra-population variability of this output, at little added cost.

A major challenge in the design of reliable synthetic biological systems is assessing and regulating the impact of variability and noise upon their behaviour[1]. In a particular system, noise sources can be broadly classified as arising from either intrinsic (due to the stochastic nature of the system’s constituent biochemical processes) or extrinsic (due to fluctuations in the concentration and behaviour of cellular components with which it interacts) sources[2], [3]. Extrinsic noise that impacts the behaviour of synthetic circuits can be caused by both short- and long-term phenomena: It can be introduced by temporary fluctuations in cellular machinery abundance[4] (such as as ribosome sequestration, which impacts the translation rate of synthetic circuits[5], [6], or uneven distribution of proteins during cell division[7]), as well as epigenetic differences in proteome and cell state that are passed on as cells divide[8], [9], [10]. Total gene expression noise is thus a function of both cell-wide fluctuations, as well as variability in gene-specific regulation[11].

In many cases noise can negatively influence the function of synthetic biological circuits[12], [13]. Fluctuations in the concentration of individual elements of a synthetic network can propagate throughout the system, impacting the behaviour of other components[4], [14]. This has motivated the design and implementation of synthetic systems that

reduce variability[15], for example via inclusion of feedback control architectures[16], [17], [18]. At the same time, in certain circumstances noise can be beneficial: In natural systems variability between cells provides an evolutionary advantage in changing environments[19], [20], and can enhance information transfer in genetic networks[21]. Similarly, synthetic systems can be designed that benefit from the stochasticity of gene expression[22], making this factor essential in some cases for accurate modelling of their behaviour[23], [24].

When synthetic biological designs are realised experimentally, assessment of their noise properties and performance variability is therefore a critical part of their characterisation. Fluorescent reporter proteins are frequently used as an output for such systems due to the ease with which they are measured[25]. Experiments are often repeated many times (under theoretically identical conditions) using plate-reader hardware[26] to observe variation in their behaviour, which is then quantified in terms of the variability in a mean experimental outcome (for example, a cell culture’s bulk fluorescence output). However, measurement of only a cell culture’s mean fluorescence can disguise important but often unnoticed behaviours[27], such as bi-modal responses[28]. To accurately characterise many systems, measurement of intra-population cell-to-cell variability is thus required[27], for which techniques such as fluorescence microscopy and flow cytometry (which can measure individual fluorescence levels of a large number of cells) are employed extensively[29]. Recently, improvements in the capabilities of these single-cell measurement technologies have facilitated high-throughput studies of cell-to-cell variability[30]. However, single-cell measurement techniques remain time- and equipment-intensive, making their use unfavourable if large numbers of samples must be measured at regular time intervals, as is the case when parallelised experiments are used to test synthetic circuits over a range of input parameter combinations (often done using a plate reader).

This experimental challenge motivates the aim of this paper: To design a synthetic biological reporter circuit that allows characterisation of cell-to-cell variability using only measurements of a cell culture’s mean behaviour. Such a reporter could be used in any synthetic biological system with a fluorescent protein output to give an estimate of cell-to-cell variability at little added cost. This would be of great utility for many high-throughput noise-measurement experiments (which currently require flow cytometry) as it

<sup>†</sup>: Authors with the Department of Engineering Science, University of Oxford, Oxford, OX1 3PJ, UK. e-mail: ({harrison.steel, antonis}@eng.ox.ac.uk). H. Steel is supported by the General Sir John Monash Foundation. A. Papachristodoulou is supported in part by EPSRC project EP/M002454/1.

would allow them to be performed efficiently and in parallel using a plate-reader device. We achieve noise estimation via a biological implementation of a multiplication function that can be used to measure a variable's mean value, as well as its mean squared value, at a population level. It is then possible to estimate the cell-to-cell standard deviation of a system's output, which previously required measurements at a single-cell level. Thus, though our system does not return the complete shape of the expression distribution (as does flow cytometry), for cases in which the distribution mean and width are the primary parameters of interest (as in many past studies[13], [15], [17], [18]) it can provide an efficient experimental alternative.

We begin the paper with a description of the simple statistics behind this estimation procedure, showing how measurement of only bulk culture parameters can allow quantification of cell-to-cell variability. In the case of log-normally distributed behaviour (as is often found in biological systems[31]), we demonstrate that this provides a direct estimate of the width of the system's output distribution. We then describe a fluorescent-protein FRET implementation for such a circuit (an alternate implementation using an activating sRNA circuit is described in Supplementary Section 2), and discuss expected parameter values for its operation. The FRET system's variance-estimating performance is then analysed, and its operational parameter range and sensitivity to both intrinsic and extrinsic noise is assessed. These results are discussed with reference to potential experimental conditions in which our reporter system might be employed, and potential sources for error (as well as alternate implementations which might minimise these) are outlined.

## NOISE DISTRIBUTIONS AND QUANTIFICATION

We consider a measurable random biological variable of interest,  $X$ , which might (for example) correspond to the fluorescence output from a single cell due to production of a fluorescent reporter protein within that cell. The *expectation* of  $X$ ,  $E[X]$ , is its mean value over a large number of cells. In this context  $E[X]$  would (approximately) correspond to the total fluorescence of a cell colony divided by the number of cells of which it is comprised, hence giving the average fluorescence per cell. Many factors are not considered in this approximation, such as the attenuation of emitted fluorescence as it passes through experimental media (thereby making cells further from a detector device appear slightly dimmer), however these secondary effects will be ignored in our current analysis. The variance of a random variable  $X$  is given by[32]:

$$\text{Var}[X] = E[X^2] - E[X]^2, \quad (1)$$

where  $E[X^2]$  is the expected (arithmetic mean) value of the variable  $X$  squared, also referred to as the second *moment* of the random variable  $X$ . In the current context,  $X^2$  would correspond to the squared fluorescence value of a single cell.

Values of  $X$  for individual cells will vary within a population, and when a large number of cells are measured individually (i.e. at a single-cell level) the approximate distribution of  $X$  can be ascertained. For many variables of interest in biological systems, such as the level of mRNA or protein produced, this distribution is observed to be approximately log-normal during steady growth[33] (though it may be better described by a normal distribution for cells in stationary phase[34]). It has been shown that a log-normal distribution can emerge in such systems due to the inherent complexity of the noisy biochemical processes involved[31]. A log-normal distribution for a given variable will be denoted  $\text{Lognormal}(\mu, \sigma^2)$ , where  $\mu$  and  $\sigma$  are the mean and standard deviation respectively of the variable's natural logarithm. The  $n$ th moment of a log-normal distribution can be calculated analytically[35]:

$$E[X^n] = e^{n\mu + \frac{1}{2}n^2\sigma^2}, \quad (2)$$

where  $n$  is a positive integer. In reality we will not be able to measure the absolute value of  $E[X^2]$ , but rather a value "calculated" by our reporter system that is proportional to its value. We denote the measured second moment  $E_m[X^2] = \gamma_m E[X^2]$ , where  $\gamma_m$  reflects this proportionality and is termed the 2nd moment gain.

Experimentally, variation between cells in the value of a parameter  $X$  is often quantified in terms of the width of the distribution when the variable  $X$  is plot on a log scale (i.e. the parameter  $\sigma$ ). We can derive an expression for this variable for the log-normal distribution in terms of measurable quantities, giving:

$$\sigma = \sqrt{\ln\left(\frac{E_m[X^2]}{E_m[X]^2}\right) - \ln(\gamma_m)}, \quad (3)$$

where  $E_m[X]$  is the measured first moment (the mean). The square-root-log dependence of  $\sigma$  on  $E_m[X^2]$  is worth noting, as it means small changes in  $\sigma$  will be measurable as large changes in  $E_m[X^2]$ . Even if we do not estimate/measure the parameter  $\gamma_m$ , we note that  $\sigma$  is an increasing function of  $E_m[X^2]/E_m[X]^2$  (when the argument is real), meaning that as the quantity  $E_m[X^2]/E_m[X]^2$  grows, so does  $\sigma$ . If the underlying distribution is not log-normal and/or unknown we can estimate the population standard deviation using (1), though again estimation of  $\gamma_m$  is necessary.

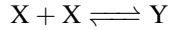
Noise strength (fano factor) is a common parameter of interest when characterising the variability of biological systems, and can be used to discriminate between different mechanisms contributing to noise in protein abundance[36], [37]. This parameter, defined as the ratio of the expression variance to its mean is constant for a perfectly poissonian process[38]. However, for most biological process (for which a log-normal distribution is anticipated) the coefficient of variation provides a better measure of gene-expression noise[37]. The (arithmetic) Coefficient of Variation ( $CV$ ) is defined as the ratio of the arithmetic standard-deviation to the

mean ( $SD[X]/E[X]$  where  $SD[X] = \sqrt{Var[X]}$ ) due to the linear relationship observed between these parameters[39]. We therefore have:

$$CV = \sqrt{e^{\sigma^2} - 1} \approx \sigma \quad (4)$$

where the last approximation uses the truncated Taylor-series  $e^x \approx 1 + x$  which is valid for small sigma ( $\sigma \lesssim 0.8$ ). This highlights the appropriateness of the coefficient of variation for assessing noise in biological systems exhibiting log-normal behaviour: It is proportional to the width of the expression distribution when plot on a log scale.

We have thus outlined a method for determining the level of variability in a parameter  $X$ , as is often measured via flow cytometry, but here only using measurements of bulk parameters. To build an estimator of  $E_m[X^2]$  we must create a system with fundamental structure:



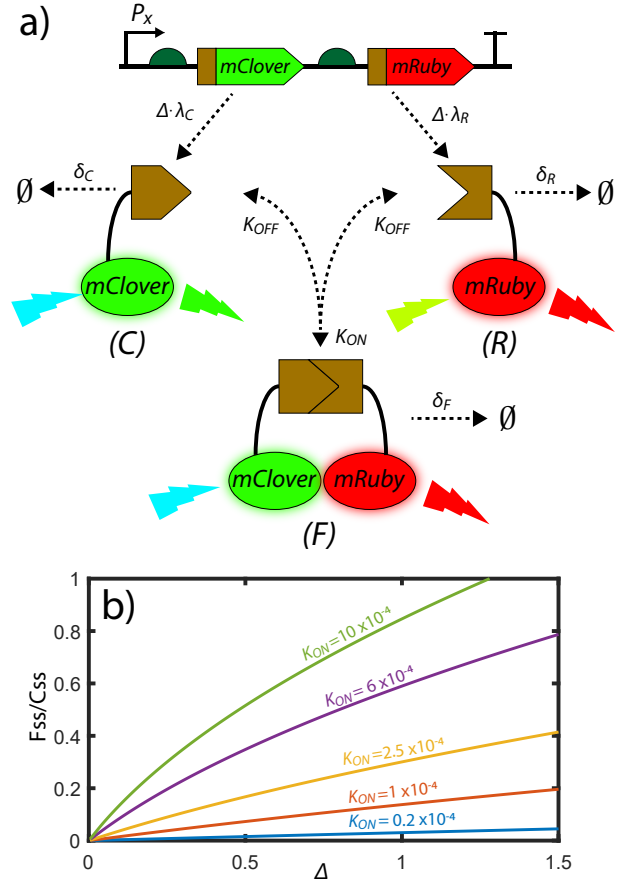
where we can both measure  $X$ , and the variable  $Y \propto X^2$ . We now describe and analyse a possible FRET implementation for such a variance estimating circuit. A second potential implementation, based upon an activating sRNA circuit, is described in Supplementary Section 2.

#### A BIOLOGICAL IMPLEMENTATION: FRET REPORTER PAIR

One approach to measuring a variable  $X$  in terms of its mean ( $E_m[X]$ ) and second moment ( $E_m[X^2]$ ) is to use a dimerising FP-FRET (Fluorescent Protein - Fluorescence Resonance Energy Transfer) protein pair that can approximate a multiplication function (Fig. 1a). In this case we have chosen the mClover3 ( $C$ ) and mRuby3 ( $R$ ) fluorescent proteins[40], but many other possibilities exist[41]. Both fluorescent fusion proteins are expressed in tandem from the promoter  $P_x$ , for which we aim to measure the variability in transcription. This variability may arise from any upstream processes (e.g. gene regulatory networks) interacting with  $P_x$ .

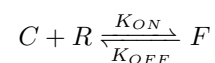
When spatially separated, the fluorescent proteins have their usual excitation/emission behaviour. However, when they are brought into close proximity ( $\sim 10\text{nm}$  or less[41]) it is possible to excite the complex at the lower excitation wavelength of mClover3 (termed the *donor*), and measure a change in fluorescence in the emission spectrum of mRuby3 (the *acceptor*). This occurs via non-radiative energy transfer through long-range dipole to dipole interactions when the fluorophores are proximal[42]. A diversity of approaches to measuring FRET exist[41], though we will concentrate on ratiometric measurements of FRET intensity which may be performed using a plate reader (or fluorescence microscope or flow cytometer)[43].

To co-localise the two fluorescent proteins such that FRET can occur we fuse them to complementary protein binding domains, as has been done in a number of FRET-based



**Fig. 1: A FRET Reporter system for analysis of gene expression variability.** **a)** Structure of the system, in which the concentration of a FRET-active dimer ( $F$ ) is approximately proportional to the square of the concentration of its unbound components  $C$  and  $R$ .  $C$  and  $R$  each consist of a fusion of a fluorescent protein (mClover3 and mRuby3 respectively) with a protein binding domain (brown). Dashed arrows represent reactions and their directions (with rate parameters as defined in the text), and degradation/dilution reactions ( $\delta$ ) represent removal of species to the null state  $\emptyset$ . The DNA operon (top) consists of a promoter ( $P_x$ ), Ribosome Binding Sites (RBS, dark-green), genes (labelled), and a transcriptional terminator. **b)** The dependence of the ratio  $F_{SS}/C_{SS}$  (as defined by (6) and (7)) on the value of  $\Delta$ . Colours denote different values of  $K_{ON}$  in units of  $s^{-1}$ . For an ideal multiplication operation  $F_{SS}/C_{SS}$  would be linear in  $\Delta$ , which is best satisfied for small values of  $K_{ON}$ .

protein localisation studies[44], [45]. Each fluorescent protein can therefore exist in isolation, or they can dimerise to form a complex  $F$  (Fig. 1a). The binding strength of the complementary domains represents the major parameter for tuning in our system, which could initially be investigated experimentally using a pair with ligand-dependent binding[45], [46]. Their binding properties will determine the forward and backward rates ( $K_{ON}$  and  $K_{OFF}$ ) for the reaction:



If we assume the concentrations of both  $C$  and  $R$  are proportional to the gene expression variability that we aim to analyse (the parameter  $\Delta$  introduced at  $P_x$ ), then this system provides multiplicative ability when the concentrations of  $C$  and  $R$  are greater than  $F$ . Following the terminology of the previous section the concentration (and hence fluorescence) of either mClover or mRuby is a proxy for the variable  $X$  which we aim to measure, and the concentration of  $F$  represents the variable  $Y$ . We can model this system using a system of differential equations of the form:

$$\dot{C} = \Delta\lambda_C - \delta_C C - K_{ON}CR + K_{OFF}F \quad (5a)$$

$$\dot{R} = \Delta\lambda_R - \delta_R R - K_{ON}CR + K_{OFF}F \quad (5b)$$

$$\dot{F} = K_{ON}CR - K_{OFF}F - \delta_F F \quad (5c)$$

where  $C$ ,  $R$  are the concentrations of the mClover and mRuby fusion proteins respectively, and  $F$  is the concentration of the bound FRET complex.  $\Delta$  is the random scaling factor introduced by  $P_x$  which we aim to measure,  $\lambda_{C,R}$  are lumped transcription/translation rates (we model these in a single step for simplicity), and  $\delta_{C,R,F}$  are degradation rates for the corresponding species. By setting the time derivatives to zero it is possible to solve the system of equations (5) analytically to find the steady-state (ss) concentrations of each complex, giving:

$$C_{ss}, R_{ss} = \frac{\Delta\lambda_{C,R} - \delta_F F_{ss}}{\delta_{C,R}}. \quad (6)$$

and

$$F_{ss} = a(b + \Delta c - (\Delta^2 d + 2bc\Delta + b^2)^{0.5}) \quad (7)$$

where:

$$a = 1/(2K_{ON}\delta_F^2) \quad (8a)$$

$$b = \delta_C\delta_R(K_{OFF} + \delta_F) \quad (8b)$$

$$c = K_{ON}\delta_F(\lambda_C + \lambda_R) \quad (8c)$$

$$d = K_{ON}^2\delta_F^2(\lambda_C - \lambda_R)^2 \quad (8d)$$

For an ideal multiplication operation to occur, we aim to tune system parameters such that  $F_{ss} \propto \Delta^2$ , which would then mean (approximately) that  $F_{ss} \propto C_{ss}^2, R_{ss}^2$ . Observing that the constant term ( $F_{ss}|_{\Delta=0}$ ) and first derivative ( $\frac{\partial F_{ss}}{\partial \Delta}|_{\Delta=0}$ ) are zero, to achieve  $F_{ss} \propto \Delta^2$  we thus desire:

$$\frac{\partial^2 F_{ss}}{\partial \Delta^2} = \frac{ab^2(c^2 - d)}{(\Delta^2 d + 2bc\Delta + b^2)^{1.5}} \sim \text{Constant}, \quad (9)$$

for all  $\Delta$ . This criteria is best satisfied if  $b$  is large and  $c$  and  $d$  are small. Here  $b$  can be made large by setting the protein degradation terms to be fast (i.e.  $\delta_{C,R,F}$  are large). However, tuning these parameters can be challenging (and may introduce noise): A more effective approach is to focus on  $c$  and  $d$  which can be made small by reducing  $K_{ON}$  (and in the case of  $d$  setting  $\lambda_C = \lambda_R$ ).  $K_{ON}$  is the rate at which the fusion proteins  $C$  and  $R$  dimerise, which can be reduced by making the binding of their attached protein domains weak. By minimising  $F_{ss}$  this constraint also achieves  $C, R \propto \Delta$  in (6). This follows from an intuitive

analysis of the opposite scenario, in which a large  $F_{ss}$  means most protein is in its dimerised form, and thus  $F_{ss}$  will be proportional to whichever of  $C_{ss}$  or  $R_{ss}$  is smaller.

These results can be rearranged to give:

$$F_{ss} = \frac{K_{ON}C_{ss}}{K_{OFF} + \delta_F} \cdot \frac{\Delta(\lambda_R - \lambda_C) + \delta_C C_{ss}}{\delta_R} \quad (10)$$

which if  $\lambda_R = \lambda_C$  and  $\delta_R = \delta_C$  (i.e. both fusion proteins are expressed and degraded at equal rates), simplifies to:

$$F_{ss} = \frac{K_{ON}}{K_{OFF} + \delta_F} C_{ss}^2 \approx \gamma_m C_{ss}^2 \quad (11)$$

giving a straight-forward estimate of  $\gamma_m$  for our simulated system. In the case where  $\delta_F \ll K_{OFF}$  we have  $\gamma_m \approx 1/K_d$ , where  $K_d = K_{OFF}/K_{ON}$  is the dissociation constant of the dimerisation process.

## PARAMETER SELECTION

We express all parameters values in terms of molecules per cell to aid stochastic simulations: Assuming a cell volume of  $0.6 \mu\text{m}^3$  a 1 nM concentration corresponds to approximately 1 molecule per cell. The lumped transcription/translation rates ( $\lambda_{C,R}$ ) can be expressed as a product of the constituent processes,  $\lambda_{C,R} = \alpha_m \beta_{C,R}/\delta_m$ , where  $\alpha_m$  is the transcription rate of the (common) mRNA,  $\beta_{C,R}$  are the translation rates of the two fusion proteins, and  $\delta_m$  is the mRNA degradation rate. We set the transcription rate from the promoter as  $\alpha_m = 0.05 \text{ s}^{-1}$  to correspond with typical levels from a mid-strength promoter in *E. coli*[47], [48], and the translation initiation rates are set as  $\beta_{C,R} = 0.05 \text{ s}^{-1} \text{ mRNA}^{-1}$  as is measured for mid-strength RBSs[47], [49]. The mRNA degradation rate is set as  $\delta_m = 4.1 \times 10^{-3} \text{ s}^{-1}$  to represent a typical mRNA half-life of 2.8 minutes for *E. coli* growing in exponential phase[50]. Combining these parameters we have  $\lambda_{C,R} = 0.61 \text{ s}^{-1}$ . The protein degradation rate is set as  $\delta_{C,R,F} = 3.9 \times 10^{-4} \text{ s}^{-1}$  to correspond to dilution during growth with a 30 minute doubling time[48]. These parameter values give a typical equilibrium protein abundance of  $\lambda_{C,R}/\delta_{C,R,F} \sim 1500 \times \Delta \text{ cell}^{-1}$ , and correspond closely to those assumed in other studies of noise in *E. coli*[51].

As described in the previous section, the values of  $K_{ON}$  and  $K_{OFF}$  are the most important tuning dial for achieving the system's desired multiplicative behaviour. Generally the dimerisation timescales observed in FP-FRET experiments are substantially faster than cellular processes that they aim to measure, such as phosphorylation[52]. We therefore set  $K_{OFF} = 1 \text{ s}^{-1}$  so that our system reflects the typical fast response times of  $\sim 10 \text{ s}$  for FRET dimerisation[53]. To guide our selection of  $K_{ON}$  values we consider a typical dissociation constant for protein-protein binding in a FP-FRET system,  $K_d = 4.4 \mu\text{M}$  [52]. This value is very similar to that measured for the same protein binding domains when they are not fused to fluorescent markers[52], which (assuming this consistency transfers to other systems)

would simplify the task of selecting appropriate binding domains from the literature to implement our system. Using these values for  $K_{OFF}$  and  $K_d$  (and converting to units of molecules per cell) we have  $K_{ON} = 2.4 \times 10^{-4} \text{ s}^{-1}$ , which will be used to guide the range of  $K_{ON}$  values investigated. For typical experiments in which our variability reporter might be employed[17], [54], [55] we find values of  $\sigma$  in the range  $0.5 \sim 1.5$ , though we will focus analysis on the lower end of this range (as this is where disruption of our system's performance by noise and parameter variation will be greatest).

An important consideration in the FRET implementation's design is that its output remains measurable: If very small  $K_{ON}$  values are chosen (to best approximate a multiplication function), it may be challenging to measure the small amount of FRET signal against a high level of background fluorescent proteins that are not undergoing FRET[56]. In an *in vitro* study Gao *et al.* found that the FRET ratio was measurable with very high signal-to-noise for monomers at a concentration of  $\sim 1 \mu\text{M}$  which (for their measured dimerisation  $K_d$ ) corresponded to  $\sim 15\%$  of the proteins being in their FRET-active dimerised state[52]. Further, Potzkei *et al.* demonstrated that accurate FRET measurements are possible in living *E. coli* when only a small fraction of a sensor was in its FRET-active state[57]. Due to overlap of fluorescent protein spectra, measurements include excitation and emission cross-talk which is exhibited as a baseline fluorescence level in the emission band of the acceptor fluorescent protein. Hoppe *et al.* demonstrated one approach that uses measured fluorescence values to calculate the relative concentrations of monomers and a FRET-active dimer in the presence of this cross-talk[58] (and a range of other methods exist[56]). Furthermore, they demonstrated that FRET signals are distinguishable even when the concentration of the FRET-active dimer is  $<5\%$  of that of one of its constituent components[58]. Therefore, in the following sections we analyse our circuit's behaviour in terms of the relative concentrations of its components ( $E_m[F]$ ,  $E_m[C]$ ), assuming that these can be inferred from accurate measurements of fluorescence (enabled by characterisation of the particular FRET pair's excitation/emission interaction[41], [58]).

#### DETERMINISTIC ANALYSIS

In this section we analyse the system's steady-state output (given by (6) and (7)) using the parameter values discussed above, and with the value of  $K_{ON}$  varying between simulations. A similar analysis is performed for the alternate activating sRNA implementation in Supplementary Section 2. In Fig. 1b the ratio  $F_{ss}/C_{ss}$  is plot for varying  $K_{ON}$ ; for ideal multiplicative behaviour we desire a linear relation, which is best achieved for low values of  $K_{ON}$ . As discussed in the previous section this parameter's value represents an experimental trade-off: Smaller  $K_{ON}$  values uniformly reduce the concentration of  $F$  (and hence the

measured FRET dynamic range), but mean that the system better approximates a multiplication function.

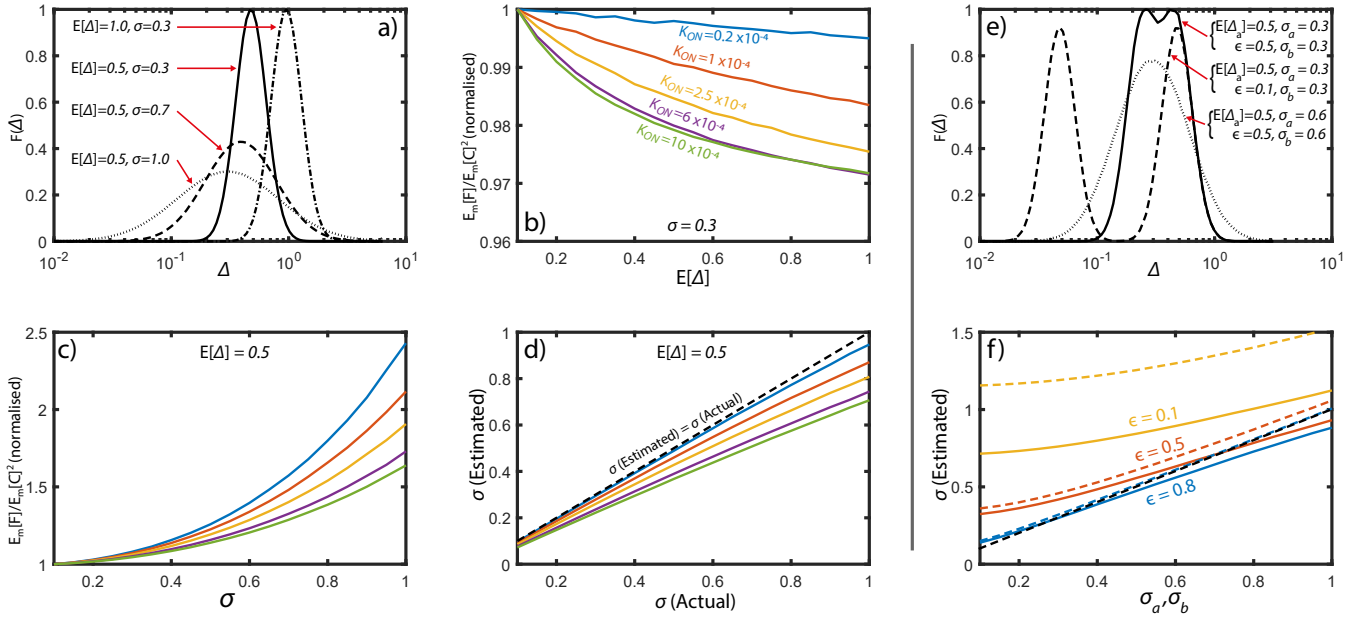
We now examine the system's ability to retrieve parameters of the underlying probability distribution of  $\Delta$  when only measurements of the circuit's mean outputs ( $E_m[F]$ ,  $E_m[C]$ , and  $E_m[R]$ ) are available. The values for  $\Delta$  are sampled from a log-normal distribution ( $\Delta \in \text{Lognormal}(\mu, \sigma^2)$ ). We vary the value of  $\sigma$ , which is proportional to the distribution's width when plot on log-scale, and  $E[\Delta]$ , the expected value of  $\Delta$  when sampled from this distribution.  $E[\Delta]$  therefore corresponds to the distribution's arithmetic mean, as would be measured using a plate reader. Fig. 2a provides examples of log-normal distributions with varying parameters: Changing  $E[\Delta]$  with  $\sigma$  fixed shifts the distribution's peak, but maintains its width, whilst changing  $\sigma$  with  $E[\Delta]$  fixed flattens the distribution, and shifts its peak (since  $\mu = \ln(E[\Delta]) - \sigma^2/2$  depends on both  $E[\Delta]$  and  $\sigma$ ).

In Fig. 2b,c we calculate the system's mean output over  $10^6$  log-normally distributed  $\Delta$  values, and examine the experimentally measurable ratio of population means  $E_m[F]/E_m[C]^2$  (which appears in (3)) as distribution parameters ( $E[\Delta]$ ,  $\sigma$ ) vary. As  $E[\Delta]$  varies this ratio would ideally remain constant if  $C \propto \Delta$  and  $F \propto \Delta^2$ , which is what is observed in Fig. 2b with the best agreement occurring for small  $K_{ON}$ . In Fig. 2c this ratio is plot for varying  $\sigma$ , demonstrating the system's sensitivity to increasing variability in  $\Delta$ . At this point we have not assumed any fundamental properties of the log-normal distribution (though it has been used to generate random samples), and as such the ratio  $E_m[F]/E_m[C]^2$  can be used to assess relative changes in noise (the variability of  $\Delta$ ) in a given system, even if the noise distribution deviates from log-normal (though the shape of the curves in Fig. 2b,c may change). In Fig. 2d we have used this measurable ratio (and the approximation  $\gamma_m = 1/K_d$  from (11)) to directly estimate the parameter  $\sigma$  using (3), demonstrating that the best agreement is found for low  $K_{ON}$ . In Fig. S1 we repeat this analysis when  $\lambda_C \neq \lambda_R$ , which gives a similar outcome.

A useful application of our variance-estimating system would be to identify when biological processes exhibit a bi-modal output, which is obscured by traditional measurements of only a cell culture's mean fluorescence[27], [38]. We therefore analyse our system's behaviour when  $\Delta$  is sampled from a sum ( $\Delta_s$ ) of two log-normal distributions ( $\Delta_a$  and  $\Delta_b$ ) with equal weighting:

$$\Delta_s = \Delta_a + \Delta_b \sim \text{Lognormal}(\mu_a, \sigma_a^2) + \text{Lognormal}(\mu_b, \sigma_b^2) \quad (12)$$

Examples of such distributions are plot in Fig. 2e, where we have set  $\sigma_a = \sigma_b$  and we define  $\mu_b$  such that  $E[\Delta_b] = \epsilon E[\Delta_a]$ . In a standard plate-reader fluorescence measurement a decrease in  $\epsilon$  (i.e. the distribution becoming increasingly bi-modal) would appear as a decrease in mean expression (Fig.



**Fig. 2: Deterministic Analysis of FRET reporter system.** **a)** Log-normal distributions for varying  $E[\Delta]$  and  $\sigma$ . **b)** Normalised (by value at  $E[\Delta] = 0.1$ ) value of the system output  $E_m[F]/E_m[C]^2$  (which is calculated as an average of system output for  $10^6$  values of log-normally distributed  $\Delta$ ) as the distribution's mean ( $E[\Delta]$ ) is varied for a fixed value of  $\sigma$ . An ideal system would be invariant under changes in  $E[\Delta]$ , which is best achieved for small  $K_{ON}$  (colours denote different values of  $K_{ON}$  in units of  $s^{-1}$  and are maintained in panels c,d). **c)** System outputs simulated similarly to panel b, but with varying  $\sigma$  and fixed  $E[\Delta]$ . An ideal system would show a strong dependence on  $\sigma$ , making experimental measurement of changes in this ratio easier, which is best achieved for small  $K_{ON}$ . **d)** Comparison of the value of  $\sigma$  estimated using (3) (and the approximation  $\gamma_m = K_{ON}/K_{OFF}$  from (11)) with the underlying distribution's actual value. Best estimation accuracy is achieved for small  $K_{ON}$ . **e)** Bi-modal log-normal distributions described in (12) with varying  $\sigma$  and  $\epsilon$ . **f)** Estimated value of  $\sigma$  (solid lines) and “true” value of  $\sigma$  (dashed lines, the coefficient of variation of the bi-modal distribution) for bi-modal distributions with  $K_{ON} = 1 \times 10^{-4} s^{-1}$ . Black dashed line indicates  $\sigma(\text{Estimated}) = \sigma_a, \sigma_b$ .

S2), which does not give clear indication of an increased variation due to bi-modality. However, in Fig. 2f bi-modality is exhibited as a substantial increase in the estimated value of  $\sigma$ , though our system's ability to estimate the “true” value for  $\sigma$  deteriorates as the underlying distribution deviates from uni-modal log-normal. Though bi-modality substantially shifts the estimated  $\sigma$ , it is not distinguishable from the case in which there is a very wide uni-modal distribution with equal  $\sigma$ .

#### PARAMETER DEPENDENCE AND NOISE SENSITIVITY

The extent to which the deterministic behaviour of the variance-estimating system will be achieved in an experimental realisation will depend upon its performance in a noisy cellular environment. In this section we further analyse the FRET reporter implementation, demonstrating that its functionality is maintained under variations in parameter values, and in the presence of both extrinsic and intrinsic noise.

To assess the range of gene expression levels over which a system with a given  $K_{ON}$  value can operate we calculated the accuracy of  $\sigma$  estimation as  $E[\Delta]$  varied over two orders of magnitude (Fig. 3a). This range corresponds

to mean protein abundances per cell varying between approximately 150 and 15,000 (see *Parameter Selection* section). Estimation accuracy declines (approximately) proportionally to  $\log(E[\Delta])$ , with the tradeoff between accuracy and measurability (i.e. the FRET signal being large compared to the background fluorescence) being governed by  $K_{ON}$ . Thus, by decreasing this value variance estimation will be more accurate at high protein abundances (high  $E[\Delta]$ ), but it will be increasingly challenging to measure the FRET signal when protein abundance is low.

Fig. 3b compares estimated values of  $\sigma$  as the parameter  $\gamma_m$  used in (3) varies. In practice a calibration experiment would be performed using flow-cytometry and a plate reader to give simultaneous measurements of  $\sigma$ ,  $E_m[F]$ , and  $E_m[C]$  from which  $\gamma_m$  can be estimated using (3). The value of  $\gamma_m$  would then (ideally) carry over when different upstream circuits (i.e. inputs at  $P_x$ ) are connected to the FRET reporter system. In Fig. 3b we simulate such a calibration experiment to estimate  $\gamma_m^*$  when  $\sigma = 0.5$  (and hence for  $\gamma_m^*$  we have  $\sigma(\text{Estimated}) = \sigma(\text{Actual})$  at this point). We observe that even when this parameter is over- or under-estimated, it does not have a major impact on the estimated value of  $\sigma$  over the range  $\sigma = 0.5 \sim 1.5$ .

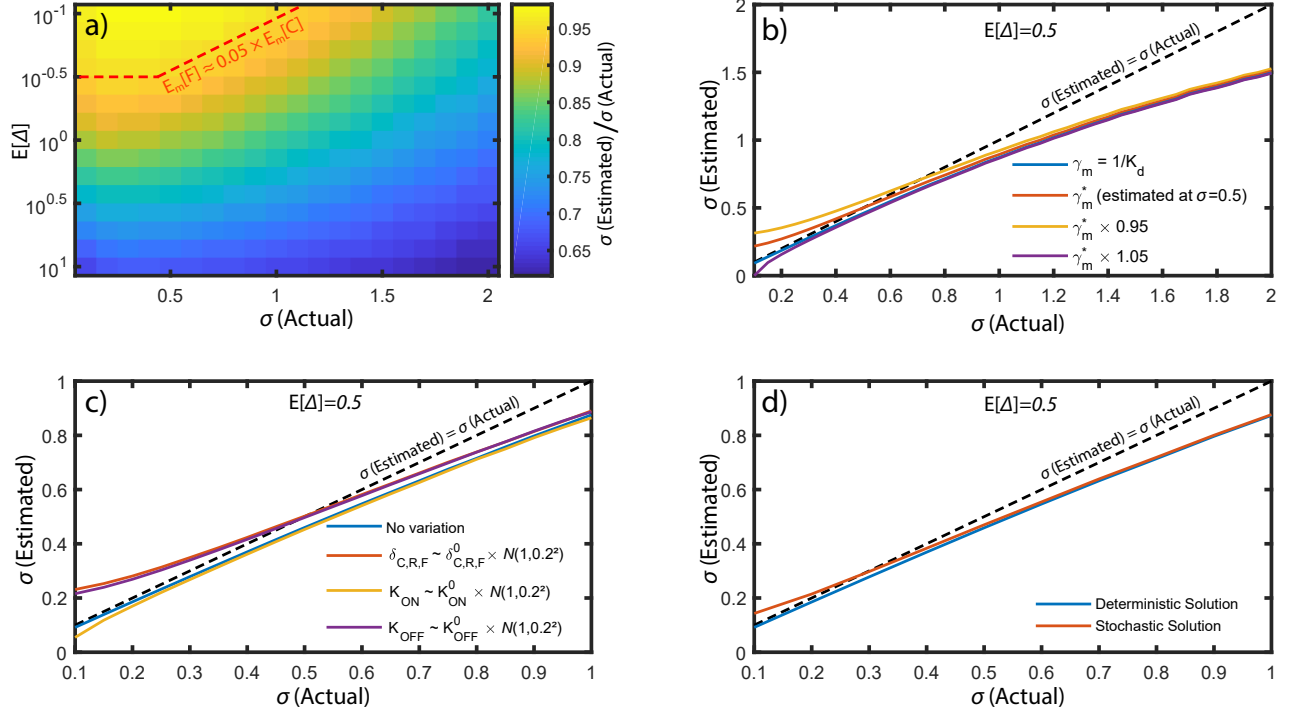


Fig. 3: **Parameter sensitivity and noise analysis with  $K_{ON} = 1 \times 10^{-4} \text{ s}^{-1}$ .** **a)** Accuracy of  $\sigma$  estimation over an experimentally-relevant range of  $\sigma$  and  $E[\Delta]$  values. The area above the dashed red line indicates the parameter regime in which the concentration of the FRET-active complex  $F$  is less than 5% of  $C$ . **b)** Dependence of calculated  $\sigma$  upon the estimated value of  $\gamma_m^*$  at  $\sigma = 0.5$ . **c)** Estimated values of  $\sigma$  when system parameters are sampled from a normal distribution. Note that we scale each of  $\delta_{C,R,F}$  by the same randomly sampled value, since each is (predominantly) determined by the shared factor of cellular growth rate. **d)** Comparison of estimated values of  $\sigma$  for the deterministic steady state solution (e.g. (6) and (7)) and stochastic simulations (described in Supplementary Section 1).

When system parameters are varied individually, we observe that for a given change in parameter value a similar magnitude of variation is observed in the quantity  $E_m[F]/E_m[C]^2$  (Fig. S3). In practice isolated parameter variation is unlikely: It would require a coordinated shift in the mean value of a single parameter (independently of all others) across all cells in a population. More likely are consistent shifts across a group of parameters[51]; for example,  $\lambda_C$  and  $\lambda_R$  might decrease by a similar amount (e.g. due to ribosome sequestration), or  $\delta_{C,R,F}$  might vary together (due to a change in mean cellular growth rate). In both these cases we find the opposing actions cancel each other out, returning the system to its previous output level.

In order to estimate the variability of an upstream process we ultimately require minimal noise to be introduced within the variance-estimating system itself. Since our system's purpose is to assess the noise in the upstream regulatory architecture / gene circuit that interacts with  $P_x$ , the variability we aim to measure is that which occurs “upstream” of the point at which transcription is initiated from  $P_x$ . We are thereby defining factors such as the varying copy number of the plasmid holding  $P_x$  (should our system be

implemented on a plasmid) as being extrinsic to the system's internally introduced variability. Much of the variability in output protein levels arising within our system will be caused by variation at the transcription and translation steps[31]. We can model noise introduced due to extrinsic sources during this process by defining a new parameter  $\Delta_*$ :

$$\Delta_* = \Delta \times \Delta_2 \quad (13)$$

where  $\Delta$  is the variability at  $P_x$  as defined previously, and  $\Delta_2$  is the noise introduced during the gene expression process within our variance-estimating system itself. Due to the large number of secondary biological factors that combine to determine the rate of transcription and translation we anticipate that  $\Delta_2$  will also be distributed log-normally[31], and we denote its underlying parameters  $\mu_2$  and  $\sigma_2$ . Because  $\Delta_*$  is the product of two log-normal distributions it will be similarly distributed, with parameters  $\mu_* = \mu + \mu_2$  and  $\sigma_*^2 = \sigma^2 + \sigma_2^2 + 2\sigma\sigma_2\rho$ , where  $\rho$  is the correlation between the random variables  $\Delta$  and  $\Delta_2$ . For un-correlated noise ( $\rho = 0$ ),  $\sigma_* = \sigma\sqrt{1 + \sigma_2^2/\sigma^2}$ , and thus the influence of a fixed  $\sigma_2$  decreases rapidly as  $\sigma$  grows. For perfectly correlated parameters ( $\rho = 1$ ) we have  $\sigma_* = \sigma + \sigma_2$  which represents the “worst case” scenario: For any  $|\rho| < 1$  we have an estimation error of  $|\sigma_* - \sigma| < |\sigma_2|$ .



Past studies have considered extrinsic noise to be the component of total noise that is correlated between expression of different proteins within a cell[2] (implying a value  $\rho = 1$ ). However, with parameter  $\Delta_2$  we aim to model variability arising due to sources that are external to our circuit’s stochastic biochemical reactions, but which may differ between the expression of the different proteins in our system[31]. For example, differing coding sequences for the two fusion proteins would mean that the (extrinsic) variability in the concentration of charged tRNA species may impact their translation rates to different degrees. Thus, we expect a large (though not unity) value for  $\rho$ , since most of this noise arises due to shared parameters determining rates of transcription and translation[31]. That aside, we proceed with the worst case scenario in which the variability estimated by our system ( $\sigma_*$  measured from  $\Delta_*$ ) differs from the “true” variability ( $\sigma$  from the random parameter  $\Delta$ ) by an amount  $\sigma_2$  due to the impact of extrinsic noise on the variance-estimating system itself.

The important consideration in this case is therefore the difference between  $\sigma_2$  and  $\sigma$ . We can estimate this from past studies, in which extrinsic noise (measured by  $CV$ , which by (4) is approximately equal to  $\sigma$  in the low  $\sigma$  region) was observed to increase by a factor of  $\sim 5$  when a previously constitutively expressed fluorescence marker was put under the control of the inducible LacI transcription factor[2]. Therefore, considering that circuits that are substantially more complex and noisy than a LacI inducible promoter might generally be analysed by our proposed variance estimating system, extrinsic noise is only likely to have a minor influence on its estimating capability. Furthermore, by replacing the promoter  $P_x$  with a constitutive promoter known to provide low-noise transcription it may be possible to estimate  $\sigma_2$  (though this parameter would show some level of context-dependence), and directly subtract its value from the measured  $\sigma_*$  to find  $\sigma$ .

Although the transcription and translation processes are likely to be a more substantial source of noise, we also analysed the impact of extrinsic variability in  $\delta_{C,R,V}$  and  $K_{ON,OFF}$  by sampling each of these values from a normal distribution (with relatively large standard deviation) and averaging the system’s performance over a large number of trials (Fig. 3c). We observe that the influence of parameter noise decreases as the level of variability that we are attempting to measure grows, because in this case it is not correlated with variability in  $\Delta$ .

When biomolecular systems operate at low concentration scales the fact that interactions occur in discrete increments may substantially alter a system’s behaviour[59]. To examine the influence of intrinsic noise introduced by stochasticity in our system we developed a stochastic model of its behaviour, described fully in Supplementary Section 1. We separate the bulk transcription/translation process (characterised by rates  $\lambda_{C,R}$ ) considered in the analytical

analysis of our system into its individual steps, in order to explicitly simulate mRNA concentration dynamics. We anticipate large fluctuations in mRNA concentration due to its lower abundance, whereas for the higher concentration proteins the stochastic model results should converge towards the equilibrium levels predicted by analytical analysis of the differential equations[59]. The stochastic model was simulated for  $10^5$  trials over a range of values for  $\sigma$ . The estimated value of  $\sigma$  in the presence of stochasticity is plot in Fig. 3d and compared to that achieved by the analytical solution: We observe that the impact of intrinsic noise decreases rapidly as  $\sigma$  grows, which is anticipated since the influence of stochasticity within our system is not correlated with the variability in  $\Delta$ . Examples of the results from individual stochastic simulations are presented in Figs. S4-S8, illustrating how the influence of stochastic noise on the probability distribution of each species’ abundance decreases as  $\sigma$  grows.

## DISCUSSION

We have demonstrated that a biological reporter system which returns a variable’s mean value ( $X$ ), as well as an output proportional to this value squared ( $Y \propto X^2$ ), can provide an estimation of cell-to-cell variability from only bulk measurements of a cell culture’s fluorescence. Though this does not return a complete distribution of reporter expression levels, for experiments in which distribution mean and width are of primary interest our system can eliminate the need for expensive and time-consuming single-cell measurements. It thereby prevents important behavioural properties (such as increased variability introduced by bi-modality) from being obscured as they are by traditional mean-only measurements[27], [38]. One potential implementation for such a system was described, based upon a FP-FRET reporter pair that reversibly dimerises to yield a measurable change in fluorescence. A second potential implementation, based upon an activating sRNA circuit architecture, is outlined in Supplementary Section 2. We find that (for typical parameter values from the literature) the FRET system is able to accurately estimate the variability of gene expression from a promoter  $P_x$  from which transcription is controlled by an upstream genetic circuit. We find that both intrinsic and extrinsic noise have minimal influence on our system’s functionality, except for the case in which the level of variability it aims to measure is very small. There are, however, a number of constraints that limit its operational range which will now be discussed.

The region for which accurate  $\sigma$  estimation is achieved is constrained to roughly two orders of magnitude of expression levels ( $E[\Delta]$ ) from  $P_x$  (Fig. 3a), a dynamic range which is likely sufficient for analysis of most synthetic biological systems. This range can be shifted by tuning of  $K_{ON}$ ; reduced  $K_{ON}$  shifts the system’s ideal performance range to higher  $E[\Delta]$  values. For a given value of  $K_{ON}$  the feasible range is bounded below by the limitations of FRET



fluorescence measurement sensitivity, as discussed in the *Parameter Selection* section. In Fig. 3a we took this bound to be the point at which  $E_m[F]$  was 5% of  $E_m[C]$ , though in practice this bound will depend on the fluorescent protein FRET pair chosen, media conditions (i.e. background fluorescence), and the particular measurement hardware employed. The upper bound on the system's operation is constrained by its departure from true multiplicative behaviour as  $E[\Delta]$  grows, which reduces the accuracy of  $\sigma$  estimation. The difference between the estimated and actual value of  $\sigma$  (as observed in Figs. 2d and 3) could be substantially reduced by performing calibration experiments over a range of  $\sigma$  values so that a more accurate expression linking  $\sigma$ ,  $E_m[F]$ , and  $E_m[C]$  could be derived and fit to this particular implementation; this would avoid the assumption that the system attains perfectly multiplicative behaviour in its output  $E_m[F] \propto E_m[C]^2$  as is made implicitly when using (3).

Direct measurement of  $\sigma$  (which approximates  $CV$  when  $\sigma \lesssim 0.8$ ) requires the underlying distribution to be approximately log-normal, as is found ubiquitously in intra-cellular reaction dynamics[39]. This distribution arises due to the large number of noise sources which are applied multiplicatively to determine the outcome of biological processes, and thus only when one particular (non log-normal) noise source dominates a system's output is departure from log-normal behaviour anticipated[31]. We investigated this possibility for the case of bi-modality (Figs. 2e,f), demonstrating that though our design was not able to directly estimate the coefficient of variation for the bi-modal system, it was able to effectively rank them (i.e. gave higher outputs for distributions with greater variation). This highlights the usefulness of the measured parameter  $E_m[F]/E_m[C]^2$ , which increases in the presence of noise (even when not log-normal distributed) and allows the relative variability of two systems to be compared (if not estimated absolutely) without knowledge of the parameter  $\gamma_m$ .

In Fig. 2d  $\gamma_m$  was estimated using the approximation in (11), which holds for  $\lambda_C = \lambda_R$ . This equality is not necessary for our system to function (Fig. S1), though it could be achieved in practice: Both proteins are expressed from the same mRNA and are highly stable (so have approximately equal  $\delta_{C,R}$ , these being primarily due to dilution), and so their relative levels could be tuned by adjusting their RBSs (the RBS for  $C$  would likely need to be slightly stronger than that for  $R$  to account for transcriptional read-through of  $R$ ). Regardless of whether  $\lambda_C \approx \lambda_R$  is achieved,  $\gamma_m$  can be measured with flow cytometry and a plate-reader experiment, from which measured  $\sigma$  and mean fluorescence levels can be used to calculate its value using (3) (this process is simulated in Fig. 3b). Alternatively, if the approximation in (11) is satisfied then  $\gamma_m$  could be measured by expressing one fluorescent reporter from an inducible promoter and performing a titration experiment to

determine  $K_d$ . Because  $\gamma_m$  is defined by parameters of the FRET reporter system itself, its value should remain fixed when different circuits are placed upstream of  $P_x$ , unless these were to actively interfere with the reporter binding processes (changing  $K_{ON}$  and  $K_{OFF}$ ).

Stochastic simulation of our system demonstrated that its sensitivity to intrinsic noise is generally small, as anticipated for biological systems where concentrations of interacting components are large[51]. The influence of extrinsic noise was found to have substantial impact only when the variability introduced by  $P_x$  is of approximately the same level as that introduced within our system. Since our system expresses stable proteins, and noise introduced by transcriptional regulation of  $P_x$  is not considered as occurring "within" the system (this is included in the parameter  $\Delta$  which we aim to measure), it represents what are the least noisy gene expression conditions obtainable without feedback[2]. Based on similar systems[2] we would anticipate the system's total internal noise level to be  $\sigma \approx 0.1 \sim 0.2$ , which is largely determined by transcriptional bursting[60]. The constraint (that a typical upstream regulatory system contributes significantly more noise than the expression of fluorescent reporters) is satisfied for many biological systems: Elowitz *et al.*[2] found that introducing an inducible transcription factor upstream of a fluorescent reporter increased extrinsic noise by a factor of  $\sim 5$ . Furthermore, studies of synthetic biological circuits for which our reporter might be employed[17], [54], [55] typically yield values of  $\sigma$  in the range  $0.5 \sim 1.5$ , which coincides with the range considered in our simulations (Fig. 3a,b), and is above the  $\sigma$  levels at which intrinsic or extrinsic variability were found to impact our circuit (Figs. 3c,d). We thus conclude that our reporter system's functionality will be minimally disturbed by noise, with the exception of the case in which it is used to study a feedback regulatory system that is able to reduce variability of its output (i.e. whatever interacts with  $P_x$ ) to below the variability internal to our system.

As is the case for all synthetic biological systems, a range of secondary factors could disturb the behaviour of this (or any similar) variance estimating system. The dynamic range of the FRET complex is dependent upon the protein binding domains chosen and the linkers to their fluorescent markers[41], and can be improved by utilising "sticky" fluorescent protein FRET pairs[61]. For a stable cell-wide equilibrium (dependent on  $K_{ON}$  and  $K_{OFF}$ ) to develop between proteins in their unbound ( $C, R$ ) and FRET-active bound ( $F$ ) state, it is important that each is well mixed through the cell. This could be violated if reporter proteins displayed a strong tendency to form inclusion bodies, or to be localised near a single transcription site[62] (though this would be unlikely if the system was introduced on a multi-copy plasmid). Transient interference with our reporter system by other cellular processes could also degrade its behaviour (for example,

if the protein-protein binding domain used on  $C$  and  $R$  bound strongly to a native protein). Similarly, particular parameter values could be perturbed by changes in context (for example  $K_{ON}$  impacted by internal cellular crowding, or the efficiency of FRET emission varying with PH[58]), though such context-dependencies are challenges for all synthetic biological systems[63].

It is possible and likely that alternate implementations could be designed that are simpler, and provide better performance, than those proposed in this study. To estimate the variability of the parameter  $\Delta$  the only fixed design requirement is that a system has states  $X$  and  $Y \propto X^2$ , both of which are measurable. Recently “sticky” FRET pairs have been engineered by introducing weak hydrophobic interactions between the fluorescent proteins, thereby ensuring their proximity (hence boosting FRET efficiency) when they are brought together by the process that they aim to measure[61]. Similar protein engineering techniques could potentially be undertaken to further increase dimerisation affinity (but avoid homo-dimerisation), thereby eliminating the need for inclusion of protein binding domains. If variance estimation could be achieved using a single protein, then the noise introduced within the system would be reduced even further. Such a system could be conceivably built using a fluorescent reporter ( $X$ ) that could form a homodimer, with the dimerised state ( $Y$ ) being distinguishable via bulk measurement (for example, using homo-FRET[64], [65]) from the monomeric state. Zhao *et al.* developed such a system, in which homodimerisation of a cpYFP construct results in a change in its excitation spectrum[66].

## CONCLUSION

Mean-only measurements of biological systems frequently obscure complex behaviours, making single-cell measurements of variability critical to their accurate characterisation. In this paper we have demonstrated that it is possible to estimate both the mean and cell-to-cell variability of a biological system’s output using only measurements of a cell culture’s bulk fluorescence. This is achieved by designing synthetic biological systems with measurable outputs that are proportional to the rate of transcription, and to this rate squared. We proposed a potential fluorescent protein FRET implementation of such a system, and demonstrated that it achieves this functionality over experimentally-relevant parameter ranges. We analysed the impact of both extrinsic and intrinsic noise, finding that they were only significant if the upstream system being measured had very low levels of variability. The results in this paper thus illustrate the feasibility of a variance-estimating reporter system and, along with consideration of experimental factors, could be used to guide the *in vivo* implementation of such a system in future work.

## REFERENCES

- [1] M. Khammash, “An engineering viewpoint on biological robustness.” *BMC biology*, vol. 14, no. 1, p. 22, 2016.
- [2] M. Elowitz, A. Levine, E. Siggia, and P. Swain, “Stochastic gene expression in a single cell,” *Science*, vol. 297, no. 5584, pp. 1183–6, 2002.
- [3] P. S. Swain, M. B. Elowitz, and E. D. Siggia, “Intrinsic and extrinsic contributions to stochasticity in gene expression.” *Proceedings of the National Academy of Sciences of the United States of America*, vol. 99, no. 20, pp. 12 795–800, 2002.
- [4] D. A. Oyarzún, J. B. Lugagne, and G. B. V. Stan, “Noise propagation in synthetic gene circuits for metabolic control,” *ACS Synthetic Biology*, vol. 4, no. 2, pp. 116–125, 2015.
- [5] J. L. Snoep, L. P. Yomano, H. V. Westerhoff, and L. O. Ingram, “Protein burden in *Zymomonas mobilis*: Negative flux and growth control due to overproduction of glycolytic enzymes,” *Microbiology*, vol. 141, no. 9, pp. 2329–2337, 1995.
- [6] Y. Qian and D. Del Vecchio, “Mitigation of ribosome competition through distributed sRNA feedback,” *Proc. of IEEE Conference on Decision and Control*, pp. 1–30, 2016.
- [7] M. Soltani, C. A. Vargas-Garcia, D. Antunes, and A. Singh, “Inter-cellular Variability in Protein Levels from Stochastic Expression and Noisy Cell Cycle Processes,” *PLoS Computational Biology*, vol. 12, no. 8, pp. 1–23, 2016.
- [8] J.-W. Veening, W. K. Smits, and O. P. Kuipers, “Bistability, Epigenetics, and Bet-Hedging in Bacteria,” *Annual Review of Microbiology*, vol. 62, no. 1, pp. 193–210, 2008.
- [9] L. Robert, G. Paul, Y. Chen, F. Taddei, D. Baigl, and A. B. Lindner, “Pre-dispositions and epigenetic inheritance in the *Escherichia coli* lactose operon bistable switch,” *Molecular Systems Biology*, vol. 6, no. 357, 2010.
- [10] B. Snijder and L. Pelkmans, “Origins of regulated cell-to-cell variability,” *Nature Reviews Molecular Cell Biology*, vol. 12, no. 2, pp. 119–125, 2011.
- [11] A. Sanchez and I. Golding, “Genetic Determinants and Cellular Constraints in Noisy Gene Expression,” *Science*, no. December, pp. 1188–1193, 2013.
- [12] Y. Dublanche, K. Michalodimitrakis, N. Kummerer, M. Foglierini, and L. Serrano, “Noise in transcription negative feedback loops: simulation and experimental analysis,” *Mol Syst Biol*, vol. 2, p. 41, 2006.
- [13] K. F. Murphy, R. M. Adams, X. Wang, and J. J. Collins, “Tuning and controlling gene expression noise in synthetic gene networks,” *Nucleic Acids Research*, vol. 38, no. 8, pp. 2712–2726, 2010.
- [14] M. Thattai and A. van Oudenaarden, “Intrinsic noise in gene regulatory networks,” *Proceedings of the National Academy of Sciences*, vol. 98, no. 15, pp. 8614–8619, 2001.
- [15] A. Becskei and L. Serrano, “Engineering stability in gene networks by autoregulation,” *Nature*, vol. 405, no. June, pp. 590–593, 2000.
- [16] D. Nevozhay, R. M. Adams, K. F. Murphy, K. Josic, and G. Balázsi, “Negative autoregulation linearizes the dose-response and suppresses the heterogeneity of gene expression,” *PNAS*, vol. 106, no. 13, pp. 5123–8, 2009.
- [17] T. Folliard, H. Steel, T. P. Prescott, G. Wadhams, L. J. Rothschild, and A. Papachristodoulou, “A synthetic recombinase-based feedback loop results in robust expression,” *ACS Synthetic Biology*, 2017.
- [18] Y. Boada, A. Vignoni, and J. Picó, “Engineered Control of Genetic Variability Reveals Interplay among Quorum Sensing, Feedback Regulation, and Biochemical Noise,” *ACS Synthetic Biology*, vol. 6, no. 10, pp. 1903–1912, 2017.
- [19] G. Balazsi, A. Van Oudenaarden, and J. J. Collins, “Cellular decision making and biological noise: From microbes to mammals,” *Cell*, vol. 144, no. 6, pp. 910–925, 2011.
- [20] L. T. Macneil and A. J. M. Walhout, “Gene regulatory networks and the role of robustness and stochasticity in the control of gene expression,” *Genome Research*, pp. 645–657, 2011.
- [21] G. Rodrigo and J. F. Poyatos, “Genetic Redundancies Enhance Information Transfer in Noisy Regulatory Circuits,” *PLoS Computational Biology*, vol. 12, no. 10, pp. 1–20, 2016.
- [22] C. Briat, A. Gupta, and M. Khammash, “Antithetic Integral Feedback Ensures Robust Perfect Adaptation in Noisy Biomolecular Networks,” *Cell Systems*, vol. 2, no. 1, pp. 15–26, 2016.
- [23] T. Tian and K. Burrage, “Stochastic models for regulatory networks of the genetic toggle switch,” *Proceedings of the National Academy of Sciences of the United States of America*, vol. 103, no. 22, pp. 8372–8377, 2006.
- [24] N. Politi, L. Pasotti, S. Zucca, and P. Magni, “Modelling the effects of cell-to-cell variability on the output of interconnected gene networks

- in bacterial populations,” *BMC Systems Biology*, vol. 9, no. Suppl 3, p. S6, 2015.
- [25] F. Delvigne, H. Pêcheux, and C. Tarayre, “Fluorescent Reporter Libraries as Useful Tools for Optimizing Microbial Cell Factories: A Review of the Current Methods and Applications,” *Frontiers in Bioengineering and Biotechnology*, vol. 3, no. September, pp. 1–8, 2015.
  - [26] M. Chavez, J. Ho, and C. Tan, “Reproducibility of High-Throughput Plate-Reader Experiments in Synthetic Biology,” *ACS Synthetic Biology*, vol. 6, no. 2, pp. 375–380, 2017.
  - [27] B. F. Brehm-stecher and E. A. Johnson, “Single-Cell Microbiology: Tools, Technologies, and Applications Single-,” *Microbiology and Molecular Biology Reviews*, vol. 68, no. 3, pp. 538–559, 2004.
  - [28] D. A. Siegle and J. C. Hu, “Gene expression from plasmids containing the araBAD promoter at subsaturating inducer concentrations represents mixed populations,” *Proceedings of the National Academy of Sciences*, vol. 94, no. 15, pp. 8168–8172, 1997.
  - [29] B. P. Tracy, S. M. Gaida, and E. T. Papoutsakis, “Flow cytometry for bacteria: Enabling metabolic engineering, synthetic biology and the elucidation of complex phenotypes,” *Current Opinion in Biotechnology*, vol. 21, no. 1, pp. 85–99, 2010.
  - [30] O. K. Silander, N. Nikolic, A. Zaslaver, A. Bren, I. Kikoin, U. Alon, and M. Ackermann, “A genome-wide analysis of promoter-mediated phenotypic noise in *Escherichia coli*,” *PLoS Genetics*, vol. 8, no. 1, 2012.
  - [31] J. Beal, “Biochemical complexity drives log-normal variation in genetic expression,” *Engineering Biology*, vol. 1, no. 1, pp. 55–60, 2017.
  - [32] P. Billingsley, *Probability and Measure - Third Edition*. Wiley, 1995.
  - [33] M. Bengtsson, A. Ståhlberg, P. Rorsman, and M. Kubista, “Gene expression profiling in single cells from the pancreatic islets of Langerhans reveals lognormal distribution of mRNA levels,” *Genome Research*, vol. 15, pp. 1388–1392, 2005.
  - [34] B. Banerjee, S. Balasubramanian, G. Ananthakrishna, T. V. Ramakrishnan, and G. V. Shivashankar, “Tracking Operator State Fluctuations in Gene Expression in Single Cells,” *Biophysical Journal*, vol. 86, no. 5, pp. 3052–3059, 2004.
  - [35] N. L. Johnson, S. Kotz, and N. Balakrishnan, *Continuous univariate distributions. Vol. 1. Second edition. Wiley Series in Probability and Mathematical Statistics: Applied Probability and Statistics*. Wiley, 1994.
  - [36] R. Guantes, B. Cayrol, F. Busi, and V. Arluison, “Positive regulatory dynamics by a small noncoding RNA: speeding up responses under temperature stress,” *Molecular BioSystems*, vol. 8, no. 6, p. 1707, 2012.
  - [37] M. Kaern, T. C. Elston, W. J. Blake, and J. J. Collins, “Stochasticity in gene expression: From theories to phenotypes,” *Nature Reviews Genetics*, vol. 6, no. 6, pp. 451–464, 2005.
  - [38] E. M. Ozbudak, M. Thattai, I. Kurtser, A. D. Grossman, and A. van Oudenaarden, “Regulation of noise in the expression of a single gene,” *Nature Genetics*, vol. 31, no. 1, pp. 69–73, 2002.
  - [39] C. Furusawa, T. Suzuki, A. Kashiwagi, T. Yomo, and K. Kaneko, “Ubiquity of log-normal distributions in intra-cellular reaction dynamics,” *Biophysics*, vol. 1, pp. 25–31, 2005.
  - [40] B. T. Bajar, E. S. Wang, A. J. Lam, B. B. Kim, C. L. Jacobs, E. S. Howe, M. W. Davidson, M. Z. Lin, and J. Chu, “Improving brightness and photostability of green and red fluorescent proteins for live cell imaging and FRET reporting,” *Scientific Reports*, vol. 6, no. 1, p. 20889, 2016.
  - [41] B. T. Bajar, E. S. Wang, S. Zhang, M. Z. Lin, and J. Chu, “A guide to fluorescent protein FRET pairs,” *Sensors*, vol. 16, no. 9, pp. 1–24, 2016.
  - [42] T. Forster, “Zwischenmolekulare energiewanderung und fluoreszenz,” *Annalen der physik*, vol. 437, no. 1-2, pp. 55–75, 1948.
  - [43] B. Hochreiter, A. P. Garcia, and J. A. Schmid, “Fluorescent Proteins as Genetically Encoded FRET Biosensors in Life Sciences,” *Sensors*, vol. 15, pp. 26281–26314, 2015.
  - [44] C. Zhang, Z. H. Wei, and B. C. Ye, “Imaging and tracing of intracellular metabolites utilizing genetically encoded fluorescent biosensors,” *Biotechnology Journal*, vol. 8, no. 11, pp. 1280–1291, 2013.
  - [45] V. Stein and K. Alexandrov, “Synthetic protein switches: Design principles and applications,” *Trends in Biotechnology*, vol. 33, no. 2, pp. 101–110, 2015.
  - [46] E. M. W. M. Van Dongen, L. M. Dekkers, K. Spijker, E. W. Meijer, L. W. J. Klomp, and M. Merkx, “Ratiometric fluorescent sensor proteins with subnanomolar affinity for Zn(II) based on copper chaperone domains,” *Journal of the American Chemical Society*, vol. 128, no. 33, pp. 10754–10762, 2006.
  - [47] D. Kennell and H. Riezman, “Transcription and translation initiation frequencies of the *Escherichia coli* lac operon,” *Journal of Molecular Biology*, vol. 114, no. 1, pp. 1–21, 1977.
  - [48] S. T. Liang, M. Bipatnath, Y. C. Xu, S. L. Chen, P. Dennis, M. Ehrenberg, and H. Bremer, “Activities of constitutive promoters in *Escherichia coli*,” *Journal of Molecular Biology*, vol. 292, no. 1, pp. 19–37, 1999.
  - [49] A. Pai and L. You, “Optimal tuning of bacterial sensing potential,” *Molecular Systems Biology*, vol. 5, no. 286, pp. 1–11, 2009.
  - [50] H. Chen, K. Shiroguchi, H. Ge, and X. S. Xie, “Genome-wide study of mRNA degradation and transcript elongation in *Escherichia coli*,” *Molecular Systems Biology*, vol. 11, no. 1, pp. 781–781, 2015.
  - [51] V. Shahrezaei, J. F. Ollivier, and P. S. Swain, “Colored extrinsic fluctuations and stochastic gene expression,” *Molecular Systems Biology*, vol. 4, no. 196, pp. 1–9, 2008.
  - [52] R. Gao, Y. Tao, and A. M. Stock, “System-level mapping of *Escherichia coli* response regulator dimerization with FRET hybrids,” *Molecular Microbiology*, vol. 69, no. 6, pp. 1358–1372, 2008.
  - [53] L. Sanford and A. Palmer, *Recent Advances in Development of Genetically Encoded Fluorescent Sensors*, 1st ed. Elsevier Inc., 2017, vol. 589.
  - [54] T. Folliard, B. Mertins, H. Steel, T. Prescott, T. Newport, C. Jones, G. Wadhams, T. Bayer, J. Armitage, A. Papachristodoulou, and L. Rothschild, “Ribo-attenuators: novel elements for reliable and modular riboswitch engineering,” *Scientific Reports*, 2017.
  - [55] A. A. K. Nielsen, B. S. Der, J. Shin, P. Vaidyanathan, V. Paralanov, E. A. Strychalski, D. Ross, D. Densmore, and C. A. Voigt, “Genetic circuit design automation,” *Science*, vol. 352, no. 6281, p. aac7341, 2016.
  - [56] D. W. Piston and G. J. Kremers, “Fluorescent protein FRET: the good, the bad and the ugly,” *Trends in Biochemical Sciences*, vol. 32, no. 9, pp. 407–414, 2007.
  - [57] J. Potzkei, M. Kunze, T. Drepper, T. Gensch, K.-e. Jaeger, and J. Büchs, “Real-time determination of intracellular oxygen in bacteria using a genetically encoded FRET-based biosensor,” *BMC Biology*, vol. 10, no. 1, p. 28, 2012.
  - [58] A. Hoppe, K. Christensen, and J. A. Swanson, “Fluorescence Resonance Energy Transfer-Based Stoichiometry in Living Cells,” *Biophysical Journal*, vol. 83, no. 6, pp. 3652–3664, 2002.
  - [59] B. P. Ingalls, *Mathematical Modelling in Systems Biology : An Introduction*. MIT Press, 2014.
  - [60] R. D. Dar, B. S. Razoosky, L. S. Weinberger, C. D. Cox, and M. L. Simpson, “The low noise limit in gene expression,” *PLoS ONE*, vol. 10, no. 10, pp. 1–21, 2015.
  - [61] L. H. Lindenburg, M. Malisaukas, T. Sips, L. V. Oppen, S. P. W. Wijnands, S. F. J. V. D. Graaf, and M. Merkx, “Quantifying Stickiness: Thermodynamic Characterization of Intramolecular Domain Interactions To Guide the Design of Förster Resonance Energy Transfer Sensors,” *Biochemistry*, vol. 53, pp. 6370–6381, 2014.
  - [62] Á. Goñi-Moreno, I. Benedetti, J. Kim, and V. De Lorenzo, “Deconvolution of Gene Expression Noise into Spatial Dynamics of Transcription Factor-Promoter Interplay,” *ACS Synthetic Biology*, vol. 6, no. 7, pp. 1359–1369, 2017.
  - [63] S. Cardinale and A. P. Arkin, “Contextualizing context for synthetic biology - identifying causes of failure of synthetic biological systems,” *Biotechnology Journal*, vol. 7, no. 7, pp. 856–866, 2012.
  - [64] P. Sharma, R. Varma, R. C. Sarasij, K. Gousset, G. Krishnamoorthy, M. Rao, S. Mayor, I. K. Gousset, G. Krishnamoorthy, M. Rao, and S. Mayor, “Nanoscale organization of multiple GPI-anchored proteins in living cell membranes,” *Cell*, vol. 116, no. 4, pp. 577–89, 2004.
  - [65] A. N. Bader, E. G. Hofman, J. Voortman, P. M. van Bergen En Henegouwen, and H. C. Gerritsen, “Homo-FRET imaging enables quantification of protein cluster sizes with subcellular resolution,” *Biophysical Journal*, vol. 97, no. 9, pp. 2613–2622, 2009.
  - [66] Y. Zhao, Q. Hu, F. Cheng, N. Su, A. Wang, Y. Zou, H. Hu, X. Chen, H. M. Zhou, X. Huang, K. Yang, Q. Zhu, X. Wang, J. Yi, L. Zhu, X. Qian, L. Chen, Y. Tang, J. Loscalzo, and Y. Yang, “SoNar, a Highly Responsive NAD<sup>+</sup>/NADH Sensor, Allows High-Throughput Metabolic Screening of Anti-tumor Agents,” *Cell Metabolism*, vol. 21, no. 5, pp. 777–789, 2015.

#### ACKNOWLEDGEMENTS

H. S. is supported by the General Sir John Monash Foundation. A.P. is supported in part by EPSRC project EP/M002454/1.

#### AUTHOR CONTRIBUTIONS STATEMENT

H.S. Conceived the idea and performed simulations. H.S. and A.P. wrote the manuscript.

#### COMPETING FINANCIAL INTERESTS

The authors declare no competing financial interests.