

# Identifying sources of variation and the flow of information in biochemical networks

Clive G. Bowsher<sup>a,1</sup> and Peter S. Swain<sup>b,1</sup>

<sup>a</sup>School of Mathematics, University of Bristol, Bristol BS8 1TW, United Kingdom; and <sup>b</sup>SynthSys—Synthetic and Systems Biology, University of Edinburgh, Edinburgh EH9 3JD, United Kingdom

Edited by Ken A. Dill, Stony Brook University, Stony Brook, NY, and approved March 6, 2012 (received for review November 24, 2011)

**To understand how cells control and exploit biochemical fluctuations, we must identify the sources of stochasticity, quantify their effects, and distinguish informative variation from confounding “noise.” We present an analysis that allows fluctuations of biochemical networks to be decomposed into multiple components, gives conditions for the design of experimental reporters to measure all components, and provides a technique to predict the magnitude of these components from models. Further, we identify a particular component of variation that can be used to quantify the efficacy of information flow through a biochemical network. By applying our approach to osmosensing in yeast, we can predict the probability of the different osmotic conditions experienced by wild-type yeast and show that the majority of variation can be informational if we include variation generated in response to the cellular environment. Our results are fundamental to quantifying sources of variation and thus are a means to understand biological “design.”**

analysis of variance | internal history | gene expression | signal transduction | intrinsic and extrinsic noise

Cells must make decisions in fluctuating environments using stochastic biochemistry. Such effects create variation between isogenic cells, which despite sometimes being disadvantageous for individuals may be advantageous for populations (1). Although the random occurrence and timing of chemical reactions are the primary intracellular source, we do not know how much different biochemical processes contribute to the observed heterogeneity (2). It is neither clear how fluctuations in one cellular process will affect variation in another nor how an experimental assay could be designed to quantify this effect. Further, we cannot distinguish variation that is extraneous “noise” from that generated by the flow of information within and between biochemical networks. We will show that a general technique to decompose fluctuations into their constituent parts provides a solution to these problems.

Previous work divided variation in gene expression in isogenic populations into two components (3, 4): intrinsic and extrinsic variation. Both components necessarily include a variety of biochemical processes yet dissecting the effects of these processes has previously not been possible. Intrinsic variation should be understood as the average “variability” in gene expression between two copies of the same gene under identical intracellular conditions (4); extrinsic variation is the additional variation generated by interaction with other stochastic systems in the cell and the cell’s environment. Single-cell experiments established that stochasticity generated during gene expression can be substantial in both bacteria (3, 5) and eukaryotes (6, 7), but did not identify the biochemical processes that generate this variation, regardless of whether the variation is intrinsic or extrinsic.

## Decomposing Variation in Biochemical Systems

Consider a fluctuating molecular species in a biochemical system and let the random variable  $Z$  be the number of molecules of that species, for example a transcription factor in a gene regulatory network or the number of active molecules of a protein in a signaling network. Suppose we are interested in how variation in  $Z$  is determined by three stochastic variables, labeled  $Y_1$ ,  $Y_2$ , and  $Y_3$  (Fig. 1).

Each  $Y$  could be, for example, the number of molecules of another biochemical species, a property of the intra- or extracellular environment, a characteristic of cell morphology (8), a reaction rate that depends on the concentration of a cellular component such as ATP, or even the number of times a particular type of reaction has occurred. We emphasize that the  $Y$  variables are the stochastic variables whose effects are of interest: They are not all possible sources of stochasticity.

We wish to determine how fluctuations in  $Y_1$ ,  $Y_2$ , and  $Y_3$  affect fluctuations in  $Z$ , the output of the network. Intuitively, we can measure the contribution of, say,  $Y_1$  to  $Z$  by comparing the size of fluctuations in  $Z$  when  $Y_1$  is free to fluctuate with the size of these fluctuations when  $Y_1$  is “fixed” in some way. Mathematically, we can fix  $Y_1$  by conditioning probabilities on the history of  $Y_1$ : the value of  $Y_1$  at the present time and at all previous times. By using histories, we capture the influence of the past behavior of the system on its current behavior (9, 10). For example, fluctuations in protein numbers depend on the history of mRNA levels because proteins typically do not finish responding to a change in the level of mRNA before mRNA levels change again. If  $Y_1^{\mathcal{H}}$  then is the history of  $Y_1$  up to time  $t$ , the expected contribution of fluctuations in  $Y_1$  to the variation of  $Z$  at time  $t$  is

$$E\{V[Z(t)] - V[Z(t)|Y_1^{\mathcal{H}}]\}, \quad [1]$$

where  $E$  denotes expectation (here, taken over all possible histories of  $Y_1$ ) and  $V$  denotes variance. The notation  $Z(t)|Y_1^{\mathcal{H}}$  is read as  $Z(t)$  conditioned on, or given, the history at time  $t$  of the stochastic variable  $Y$ . We use  $E$ , as for example in  $E[Z(t)|Y_1^{\mathcal{H}}]$ , to denote averaging over all random variables except those given in the conditioning. Therefore,  $E[Z(t)|Y_1^{\mathcal{H}}]$  is itself a random variable: it is a function of the random variables generating  $Y_1^{\mathcal{H}}$  (we give a summary of the properties of conditional expectations in the *SI Text*). Eq. 1 can be shown to be equal to  $V\{E[Z(t)|Y_1^{\mathcal{H}}]\}$ .

**General Decomposition of Variation.** To determine the effects of fluctuations in multiple  $Y$  variables on the variation in  $Z$ , we must successively condition on groups of  $Y$  variables. We prove (*Appendix*) that

$$\begin{aligned} V[Z(t)] = & \overbrace{E\{V[Z(t)|(Y_1, Y_2, Y_3)^{\mathcal{H}}]\}}^{\text{from sources other than } Y_1, Y_2, Y_3} \\ & + \overbrace{E\{V[E[Z(t)|(Y_1, Y_2, Y_3)^{\mathcal{H}}]|(Y_1, Y_2)^{\mathcal{H}}]\}}^{\text{from } Y_3} \\ & + \overbrace{E\{V[E[Z(t)|(Y_1, Y_2)^{\mathcal{H}}]|Y_1^{\mathcal{H}}]\}}^{\text{from } Y_2} + \overbrace{V\{E[Z(t)|Y_1^{\mathcal{H}}]\}}^{\text{from } Y_1}. \quad [2] \end{aligned}$$

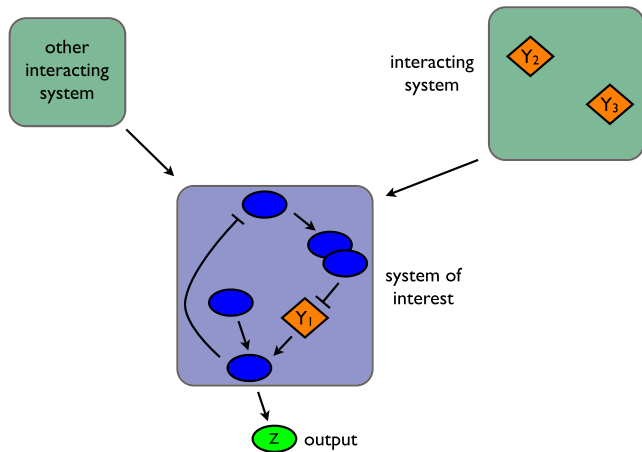
Author contributions: C.G.B. and P.S.S. designed research, performed research, analyzed data, and wrote the paper.

The authors declare no conflict of interest.

This article is a PNAS Direct Submission.

<sup>1</sup>To whom correspondence may be addressed. E-mail: c.bowsher@bristol.ac.uk or peter.swain@ed.ac.uk.

This article contains supporting information online at [www.pnas.org/lookup/suppl/doi:10.1073/pnas.1119407109/-DCSupplemental](http://www.pnas.org/lookup/suppl/doi:10.1073/pnas.1119407109/-DCSupplemental).



**Fig. 1.** Decomposing fluctuations in the output of a biochemical system. In this example, we consider how fluctuations in three variables, denoted  $Y_i$ , affect fluctuations in  $Z$ , the system's output.  $Y_1$  is a biochemical species within the system being studied;  $Y_2$  and  $Y_3$  are variables in other stochastic systems that interact with the system of interest, but whose dynamics are principally generated independently of that system. We show  $Y_2$  and  $Y_3$  in the same system, but they need not be.

Each of the middle terms can be understood as a mean difference between two conditional variances. For example, the third term, generated by fluctuations in  $Y_2$ , is equal to the average difference in the variance of  $Z$  when  $Y_1$  is fixed and when  $Y_1$  and  $Y_2$  are fixed:  $E\{V[Z(t)|Y_1^{\mathcal{H}}] - V[Z(t)|(Y_1, Y_2)^{\mathcal{H}}] \}$ . This term therefore quantifies the additional variation generated by fluctuations in  $Y_2$  when the history of  $Y_1$  is given. It requires first averaging  $Z(t)$  over all other stochastic variables given a particular history of  $Y_1$  and  $Y_2$  (the resulting expectation is a function of  $Y_1^{\mathcal{H}}$  and  $Y_2^{\mathcal{H}}$ ), then the calculation of the variance of this conditional expectation given the particular history of  $Y_1$  (the variance is taken over all  $Y_2^{\mathcal{H}}$ ), and finally the average of the resulting conditional variance over all  $Y_1^{\mathcal{H}}$ .

Our decomposition of variance holds at any point in time for any stochastic dynamic system and can be generalized to any number of groups of  $Y$  variables (Appendix, Eq. 13). It is unchanged if some of the stochastic  $Y$  variables are constant over time, with conditioning on their histories then being equivalent to conditioning on the values of the variables. As indicated, Eq. 2 has one term for each of the  $Y$  variables. This term becomes zero if the corresponding  $Y$  variable is deterministic or if the history of that  $Y$  variable is independent of  $Z(t)$  given the history of the other  $Y$  variables involved in the conditioning of that term. The remaining variance in  $Z$  after conditioning on the joint history of all the  $Y$  variables is measured by the first term. It requires the variance of  $Z(t)$  given the history of  $Y_1, Y_2$  and  $Y_3$  (the variance is taken over all other stochastic variables in the system) and then the average of this variance over all possible histories of  $Y_1, Y_2$ , and  $Y_3$ .

Eq. 2 is only unique given a particular choice of the conditioning: we could, for example, have decided to condition on the history of  $Y_2$  in the last term. Mathematically, there is no unique way to decompose the variance if we wish to determine the effects of fluctuations of more than one  $Y$  variable on variation in  $Z$ . Rather the decomposition is sequential, with all  $Y$  variables appearing in multiple components. Biological systems, however, often suggest a natural order of conditioning that follows their flow of control. The hierarchy described by the central dogma is one example. Nevertheless, freedom in choosing the order of conditioning can be an advantage because different decompositions give different insights and may be more or less amenable to experimental analysis.

**Defining Transcriptional and Translational Variation.** We can identify the contribution of a subsystem to downstream fluctuations by conditioning on the history of that subsystem. As an example, to investigate whether transcription or translation is more “noisy” during gene expression, we should condition on the history of the mRNA levels of the gene of interest,  $M^{\mathcal{H}}$ , and on the history of all stochastic processes extrinsic to gene expression. Examples of such processes include the synthesis and turnover of ribosomes and the rate of cellular growth. We will denote the collection of extrinsic processes as  $Y_e$ . If  $Z$  is the number of molecules of the expressed protein, then

$$V[Z(t)] = \overbrace{E\{V[Z(t)|(M, Y_e)^{\mathcal{H}}]\}}^{\text{translational}} + \overbrace{E\{V[E[Z(t)|(M, Y_e)^{\mathcal{H}}]|Y_e^{\mathcal{H}}]\}}^{\text{transcriptional}} + \overbrace{V\{E[Z(t)|Y_e^{\mathcal{H}}]\}}^{\text{from extrinsic effects}} \quad [3]$$

Variation from translation is the extra variation generated on average once the history of fluctuations in mRNA and  $Y_e$  is given (the first term in Eq. 3); variation from transcription is the extra variation generated by fluctuating levels of mRNA given the history of fluctuations in  $Y_e$ .

### Measuring the Components of Variation

**Using Conjugate Reporters.** Our general decomposition of variance for a biochemical system (Appendix, Eq. 13) allows us also to determine exact conditions that reporter systems should satisfy in order to quantify the different sources of variation. Such conditions establish a basis for experimental design. First, we need a reporter for the biochemical output of interest,  $Z$ . This reporter could be a fluorescently tagged protein for a genetic network or the nucleo-cytoplasmic ratio for a fluorescently tagged transcription factor that translocates in response to a signaling input. Second, given a variable of interest,  $Y$ , we use a conjugate reporter,  $Z'$ , that by definition must obey two conditions:

- $Z'(t)$  is conditionally independent of  $Z(t)$  given the history of  $Y$ ;
- $Z'(t)$  and  $Z(t)$  have the same conditional means and the same conditional variances given the history of  $Y$ .

The first, conditional independence, implies that given the history of fluctuations in  $Y$ , no other fluctuations result in the level of the first reporter correlating (in the most general sense) with the level of the second (10, 11). The second condition says that  $E[Z'(t)|Y^{\mathcal{H}}] = E[Z(t)|Y^{\mathcal{H}}]$  and  $V[Z'(t)|Y^{\mathcal{H}}] = V[Z(t)|Y^{\mathcal{H}}]$ , and in practice often means that the reporter system generating  $Z'$  is as close a copy as possible of the relevant subsystem generating  $Z$ .

The reporter system is thus designed so that it is only fluctuations in  $Y$  that “cause” a covariance between the level of  $Z$  and the level of  $Z'$ . Consequently, measuring this covariance quantifies the effects of fluctuations in  $Y$  on the variance of  $Z$ . Mathematically,  $\text{Cov}[Z(t), Z'(t)] = V\{E[Z(t)|Y^{\mathcal{H}}]\}$ , from the conjugacy of the reporters and the law of total covariance. Depending on the choice of  $Y$ , this covariance gives either the last term or the sum of the last terms in Eq. 2. The mean squared difference between the reporters,  $E[(Z(t) - Z'(t))^2]/2$ , complements the covariance measurement and gives the sum of the remaining terms in Eq. 2 (Appendix).

We need a reporter for each component of Eq. 2 and so four in total: a reporter for  $Z$  from which we can measure  $V[Z]$ , the variance of  $Z$  across a population of isogenic cells; a reporter conjugate to  $Z$  given the history of  $Y_1$  whose covariance with  $Z$ , again across an isogenic population, gives the last term of Eq. 2; a reporter conjugate to  $Z$  given the history of  $Y_1$  and  $Y_2$  whose

covariance with  $Z$  gives the sum of the last two terms of Eq. 2; and a reporter conjugate to  $Z$  given the history of  $Y_1$ ,  $Y_2$ , and  $Y_3$  whose covariance with  $Z$  gives the sum of the last three terms (Appendix). These reporters can in principle be constructed in the same cell or, if simultaneously distinguishing four reporters is technically challenging, in pairs in different cells (with the reporter for  $Z$  and one of its conjugate reporters comprising a pair).

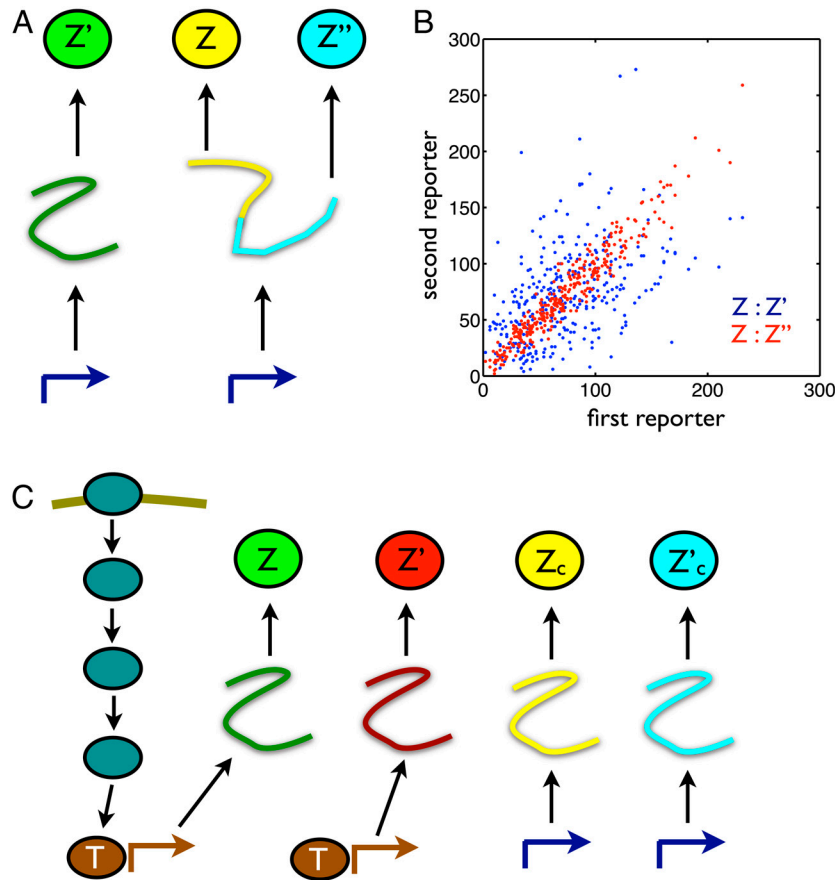
**Measuring Transcriptional and Translational Variation.** Returning to the example of measuring transcriptional and translational contributions to variation in gene expression (Eq. 3), we can construct the appropriate conjugate reporters by having a reporter for the level of the protein and a bicistronic mRNA coding for two other reporters of the protein—each with a distinct fluorescent tag—in the same cell (Fig. 2A). Only fluctuations in mRNA levels and extrinsic variables generate correlations between the two bicistronic reporters  $Z$  and  $Z''$  (12), and their covariance therefore equals the sum of the last two terms in Eq. 3 provided the conditions for conjugate reporters are met. Their mean squared difference (halved) then measures translational variation, the first term in Eq. 3. We should therefore construct the mRNA for all reporters to have identical rates of transcription, translation, and turnover. The reporters  $Z$  and  $Z'$  are condition-

ally independent given the history of the extrinsic fluctuations, and their covariance measures the last term in Eq. 3 (Fig. 2B).

Such bicistronic mRNAs have been constructed in *Escherichia coli*, but for distinguishable fluorescent proteins tagged to two different rather than identical proteins (13). We can show that these measurements give an upper bound on the translational variance: for CheY and CheZ from *E. coli*'s chemotaxis network, we show that the average translational variance for the two proteins, normalized by the product of the means of their fluorescence, is less than  $0.2^2$  (SI Text). As we will show later (Eq. 9 and SI Text), transcriptional variation usually dominates translational variation for typical parameters appropriate for *E. coli*.

### Identifying Sources of Variation in Cell Signaling

Much gene expression is initiated by signaling networks (14), and we will study examples of such expression to illustrate how to apply our decomposition (Fig. 2C). The variation observed may be determined predominantly by stochasticity in upstream signal transduction rather than by gene expression itself (15) and will not only be a consequence of biochemical noise but also a signature of information flowing through the network (16, 17). Fluctuations in gene expression can carry information on environmental changes because new rates of transcription are often caused by such changes (18). By having a general decomposition of



**Fig. 2.** Designs of conjugate reporters to measure the effects of different cellular subsystems on variation in output. (A) To distinguish transcriptional from translational effects, three reporters are needed including a bicistronic mRNA with two independent ribosome binding sites. (B) Simulated results for the reporters in A assuming that extrinsic fluctuations affect only the rate of transcription, which fluctuates between three different levels (reactions and parameter values are given in SI Text). Blue dots show  $Z$  plotted against  $Z'$ : The average spread along the  $Z = Z'$  diagonal equals the sum of  $V[Z]$  and the extrinsic variance; the average spread perpendicular to the diagonal equals the sum of the transcriptional and translational variation (SI Text). Red dots show  $Z$  plotted against  $Z''$ : the average spread along the diagonal equals the sum of  $V[Z]$ , extrinsic, and transcriptional variation; the average spread perpendicular to the diagonal equals translational variation. For the parameters chosen (SI Text), the translational noise (coefficient of variation) is 0.12; the transcriptional noise is 0.39; and the extrinsic noise is 0.41. These numbers agree with Eqs. 9 through 11 to two decimal places. (C) Four reporters are needed to distinguish transductional variation from variation generated by gene expression. Here, a signaling network activates a transcription factor,  $T$ , in response to extracellular inputs. To measure variation in the output  $Z$  arising from gene expression, we require two conjugate reporters,  $Z$  and  $Z'$ , whose expression is controlled by this transcription factor. To find a bound on transductional variation, we use two further conjugate and constitutively expressed reporters,  $Z_c$  and  $Z'_c$ .

variance, we can distinguish both signaling from expression effects and information flow from noise.

**Informational Variation.** We begin by considering variation generated in response to the extracellular environment. We will condition on the extracellular input signals,  $X$ , sensed by the signaling network. This input could be a collection of ligands or, for example, light intensity for a circadian system. In nature, different levels of input would correspond to different states of the environment. If  $Z$  is the number of protein molecules expressed from a downstream gene at a particular time, then

$$V[Z] = E\{V[Z|X]\} + V\{E[Z|X]\}. \quad [4]$$

We write Eq. 4 assuming that the biochemical system responds sufficiently quickly to reach steady state before the input changes (otherwise the system would still be responding to a previous environmental change when the next change happens). Therefore,  $X$  in Eq. 4 is not a history but takes a unique, constant value for each state of the environment. If  $X$  is dynamic, then Eq. 4 can be rewritten replacing  $X$  with its history  $X^{\infty}$ .

The variance on the left-hand side of Eq. 4,  $V[Z]$ , is generated by both variation in the extracellular input  $X$  and by variation in the output  $Z$  given a particular value of that input. If we were to measure cellular variation with the cell in its natural environment (and could not simultaneously detect the value of  $X$ ), we would see variation in the output because of  $X$  taking different values for different measurements of the output. In the laboratory, however, we can measure the distribution of output holding the level of  $X$  fixed. For example, with a fluorescent reporter for  $Z$ , we can use either microscopy (15) or fluorescence activated cell sorting (19) to measure the mean,  $E[Z|X]$ , and variance,  $V[Z|X]$ , of the response conditional on the level of the input. We can then determine  $V[Z]$  as follows: by calculating the variance of  $E[Z|X]$  for a given distribution of the input, we find  $V\{E[Z|X]\}$ , the last component in Eq. 4; by calculating the expectation of  $V[Z|X]$  over the distribution of the input, we find  $E\{V[Z|X]\}$ , the first component. We will discuss subsequently different choices of the input distribution.

We interpret the last term of Eq. 4 as the “informational” component of the variance of  $Z$ . When the means,  $E[Z|X]$ , are not the same for any two values of  $X$ , the size of this component compared to  $V[Z]$  typically indicates how difficult it is for the cell to decide the state of the environment  $X$  from the output of the sensing network. Because  $E[Z|X]$  gives the mean output for a given input, the magnitude of the variance of  $E[Z|X]$  reflects the extent to which the system can respond to a change of input. The first term of Eq. 4 describes stochasticity generated by the system during the process of sensing. The larger this term relative to  $V[Z]$ , typically the greater the deterioration in transfer of information about the environmental state to the output.

Unambiguous identification of the environmental state usually becomes easier as the fraction of the output variance that is informational increases because the distributions of output for each environmental state then typically overlap less. From information theory (20), identifying the environmental state is undermined by such overlap because different environmental states can then generate the same output (although some states are typically more likely to generate that output than others). Increased transductional variation, for example, usually leads to broader and overlapping output distributions. If the mean output when  $X$  is in state  $s$  is  $\mu_s$  and the variance of the output is then  $\sigma_s^2$ , we define the informational fraction of output variance as

$$i_Z = \frac{V\{E[Z|X]\}}{V[Z]} = \left(1 + \frac{E[\sigma_s^2]}{V[\mu_s]}\right)^{-1}, \quad [5]$$

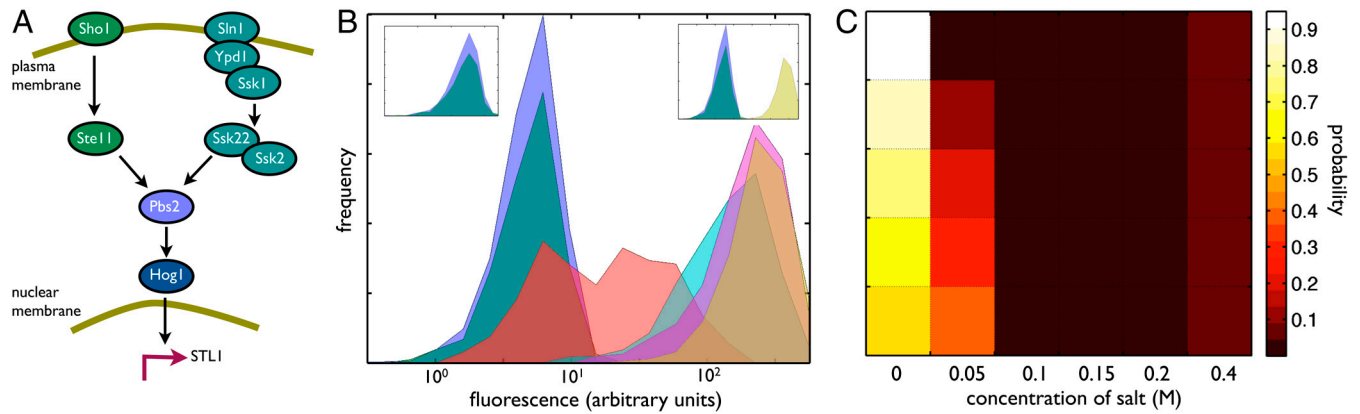
where the expectation and variance are over the probability distribution for the states of  $X$ . As the informational fraction tends

to its maximum value of one, then  $E[\sigma_s^2] \ll V[\mu_s]$ : the average “width” of a conditional output distribution typically becomes much less than the average distance between the means of two of these output distributions (SI Text). Heuristically, each output distribution is less likely to overlap with another and the system should become more efficient at transmitting information.

We can make these ideas more precise by relating the informational fraction to mutual information, the standard measure of information transfer (20). When the joint distribution of the input and output is Gaussian, the mutual information between  $X$  and  $Z$  equals  $\frac{1}{2} \ln[1 - i_Z]^{-1}$  (SI Text): an increasing function of  $i_Z$ . As the informational fraction increases so too must the mutual information between  $X$  and  $Z$ , and  $Z$  becomes more informative about the environmental state. We consider  $V[\mu_s]/E[\sigma_s^2]$  as a generalized signal-to-noise ratio that increases whenever  $i_Z$  increases. Indeed, this signal-to-noise ratio simplifies to give the one familiar from information theory for the jointly Gaussian case (21). The information capacity is the maximum possible mutual information over all potential input distributions (20). In general, we can show that the capacity is bounded below by the same increasing function,  $\frac{1}{2} \ln[1 - i_Z]^{-1}$ , evaluated using a suitable choice of the distribution of  $X$  (SI Text), and provided the conditional means  $\mu_s$  are not exactly the same for different states of  $X$ . Further, under mild conditions on the joint distribution of the input and output we can prove that the mutual information between  $X$  and  $Z$  becomes maximal as the informational fraction becomes close to one (SI Text). Such a convergence of  $i_Z$  occurs, for example, in the limit of large numbers of molecules. Nevertheless, the informational fraction is not an exact proxy for mutual information because mutual information, unlike  $i_Z$ , accounts for the entire joint distribution of input and output. As a consequence, though, reliably estimating the mutual information can require substantially more data than reliably estimating the informational fraction.

**Using Informational Variation to Analyze Osmosensing in Budding Yeast.** We can apply these ideas to measurements of the transcriptional response of budding yeast to hyperosmotic stress (Fig. 3A). Pelet et al. made single-cell measurements of a Yellow Fluorescent Protein (YFP) expressed from the promoter of STL1 under different conditions of osmotic stress by using various concentrations of extracellular salt (Fig. 3B) (19). STL1 is the typical reporter for the response to a hyperosmotic change (22). Using this data and Eq. 5, we can determine the informational fraction for any particular probability distribution of extracellular salt (the input  $X$ ). The distribution of osmotic stress encountered by yeast in the wild is unknown. We can, however, find distributions of salt concentration that give high informational fractions, an example of “inverse ecology” (23–26).

From the Pelet et al. data for the STL1 promoter, we find the highest informational fraction (of around 0.8) for distributions that have a high probability of low osmotic stress, zero probability for intermediate stresses, and a low probability for high stress (Fig. 3C). Intermediate levels of stress have zero probability in these distributions because the corresponding expression from STL1 covers a broad range and extends into the levels of expression typical of both high and low stress (Fig. 3B). If expression from STL1 is “read” by the cell to determine the level of osmotic stress, then this readout is most unambiguous for distributions of salt concentrations that have these properties (Fig. 3B, Insets). If we believe that the network has been selected to have an ability to distinguish different environmental states (27), then we can go further and suggest that the natural distribution of osmotic stress experienced by yeast is of the same form. The osmosensing network then “expects” frequent low level osmotic stress interspersed with rare, high level stress. An alternative distribution with frequent high and infrequent low level stress has a substantially lower informational fraction (of around 0.15) and does not



**Fig. 3.** Determining informational variation for osmosensing in budding yeast allows us to predict the probability of the different osmotic conditions experienced by yeast. (A) Hyperosmotic stress is sensed by two pathways in budding yeast, which activate the MAP kinase kinase kinases Ste11 and Ssk2/22 (22). Both these kinases activate the MAP kinase kinase Pbs2, which in turn activates the MAP kinase Hog1. Activated Hog1 translocates from the cytosol to the nucleus and initiates new gene expression. (B) Histograms of fluorescence data from a YFP reporter expressed from the promoter for STL1 and measured by Pelet et al. (19). Fluorescence levels typically increase with increasing extracellular salt: Blue corresponds to zero extracellular salt; dark green to 0.05 M salt; red to 0.1 M; cyan to 0.15 M; magenta to 0.2 M; and brown to 0.4 M. Approximately 1,000 data points were measured for each concentration (19) and are shown using 20 bins for the fluorescence level (calculated in log-space). The left inset shows the same histograms but weighted by the probability of the different salt concentrations for an input distribution that has a low informational fraction of output variance; the right inset is analogous but for an input distribution that has a high informational fraction of output variance. (C) The five probability distributions for extracellular salt that give the five highest informational fractions (each approximately equal to 0.8 because of the high degree of overlap of the fluorescence distributions for zero and 0.05 M salt). Each distribution is read horizontally. We calculated the informational fraction for all possible probability distributions of the six concentrations of extracellular salt that were chosen experimentally. The informational fraction decreases continuously from around 0.8 to zero. A uniform probability distribution of salt gives an informational fraction of approximately 0.6.

lead to distinct responses from STL1 in different environmental states. Low levels of expression are now ambiguous about the environmental state. The joint probability of a low level of expression and a particular state of the environment is a product of the probability of the state and the probability of low expression given that state. These joint (and the posterior) probabilities are now similar for the states of low and high stress when expression is low. In the high stress environment, low levels of expression are quite rare (lying in the long left-hand tail of the output distribution in Fig. 3B), but the high stress environment itself occurs with a high probability. Similarly, although the low stress environment occurs rarely, when it does so, low levels of expression are highly probable.

We emphasize that this example is intended to be illustrative, and there are numerous caveats (17): considering expression of several genes as the output could change the predicted input distribution; expression of YFP is itself a “noisy” measure of levels of protein; and inputs in yeast’s natural environment other than osmotic shock may affect the output response.

**Quantifying Components of Variation in Yeast’s Osmosensing.** We can decompose the first, noninformational term in Eq. 4 into a component generated by gene expression and a component generated by fluctuations in processes extrinsic to the expression of  $Z$ . Most experiments are carried out for given levels of input  $X$ , and therefore we will consider the decomposition of  $V[Z|X]$ . Any decomposition of  $V[Z|X]$  can be inserted into Eq. 4 to give the decomposition of  $V[Z]$ . The transcription factors,  $T$ , activated by upstream signaling (Fig. 3A) are extrinsic to gene expression, and we will define all other extrinsic processes by  $Y_{eT}$ . The decomposition of  $V[Z|X]$  is then (Appendix, Eq. 15)

$$V[Z|X] = \underbrace{E\{V[Z|(Y_{eT}, T)^{\otimes}, X]|X\}}_{\text{from gene expression}} + \underbrace{V\{E[Z|(Y_{eT}, T)^{\otimes}, X]|X\}}_{\text{from extrinsic effects}}. \quad [6]$$

To measure these components, we introduce a reporter  $Z'$  conjugate to  $Z$  given both  $X$  and the joint history of the transcription factors and the other extrinsic variables. Measuring the covariance of  $Z$  with this reporter for a given  $X$  will determine the extrinsic variance for that level of input; measuring half of the mean squared difference between  $Z$  and  $Z'$  will determine the intrinsic variation arising from gene expression. We can construct the appropriate conjugate reporter using a second copy of the promoter and gene for  $Z$  but with the corresponding protein marked with a different fluorescent tag (Fig. 2C).

Combining Eqs. 4 and 6, we can show that informational variation can be the dominant source of heterogeneity. In their study of osmosensing, Pelet et al. expressed a Cyan Fluorescent Protein from a second copy of the STL1 promoter in the same cells (19). This reporter is conjugate to the original YFP reporter (measuring output  $Z$ ) given the salt concentration,  $X$ , and the history of the extrinsic variables. Its covariance with  $Z$  for a given  $X$  therefore gives the extrinsic variation in  $Z$  for that level of  $X$ . Using this data (SI Text) and defining the extrinsic fraction of variance as the ratio of the extrinsic variance to  $V[Z|X]$ , we can show that the extrinsic fraction is typically about 45% for each fixed concentration of salt (noting that  $V\{E[Z|X]\}$  is then zero). When we allow for a probability distribution over osmotic conditions that generates a high informational fraction for expression of STL1, we find that the extrinsic fraction becomes about 90%, of which almost 90% is generated by informational variation. In contrast, salt distributions that generate a low, positive informational fraction (such as Fig. 3B, Left Inset) give an extrinsic fraction of around 45%, with almost no contribution from informational variation. Our results thus imply that it is important to include the effects of the natural distributions of environmental states when considering cellular heterogeneity.

**Distinguishing Variation due to Gene Expression from Variation due to Upstream Signaling.** We can further decompose the variation of  $Z$  for a given  $X$  to distinguish the variation generated by the signaling network from that generated by gene expression:

$$\begin{aligned}
V[Z|X] = & \overbrace{E\{V[Z|(Y_{eT}, T)^{\mathcal{Z}}, X]|X\}}^{\text{from gene expression}} \\
& + \overbrace{E\{V[E[Z|(Y_{eT}, T)^{\mathcal{Z}}, X]|Y_{eT}, X]|X\}}^{\text{from signal transduction}} \\
& + \overbrace{V\{E[Z|Y_{eT}, X]|X\}}^{\text{from other extrinsic effects}}, \quad [7]
\end{aligned}$$

where the last two terms sum to give the last term of Eq. 6. Transductional variation at a given  $X$  is therefore the extra variation generating by fluctuating levels of the transcription factor given the history of the other extrinsic processes and the level of  $X$ . To directly measure the third component of Eq. 7, which arises from those processes extrinsic to gene expression other than  $T$ , would require a reporter conjugate to  $Z$  given the input and the history of only those extrinsic processes. It is difficult to see how to construct such a reporter without having to create a copy of the upstream signaling network and have transduction in that copy insulated from transduction in the original network.

We can, however, find a lower bound on this extrinsic component, and consequently an upper bound on the transductional component, by introducing a reporter that is conditionally independent of  $Z$  given the history of all extrinsic variables except the transcription factor  $T$  (and given the input  $X$ ). An example is a constitutively expressed reporter, which is conditionally independent of  $Z$  given the input  $X$  and the history of  $Y_{eT}$ . This reporter, denoted  $Z_c$ , need not have the same conditional mean as  $Z$  (Fig. 2C). Introducing a further reporter conjugate to  $Z_c$  given  $X$  and  $Y_{eT}$ , we can prove that (SI Text)

$$\overbrace{V[E[Z|Y_{eT}, X]|X]}^{\text{from other extrinsic effects}} \geq \frac{\text{Cov}[Z, Z_c|X]}{\text{Cov}[Z_c, Z_c|X]} \cdot \text{Cov}[Z, Z_c|X], \quad [8]$$

where  $Z'_c$  is the conjugate reporter to  $Z_c$  and Cov denotes covariance. This inequality reflects the intuition that an observed covariance between expression of genes from different networks, here  $Z$  and  $Z_c$ , is determined by the strength of the extrinsic fluctuations common to both (15, 28). The prefactor corrects for the different conditional means (given  $Y_{eT}$  and  $X$ ) of the two processes. Eq. 8 becomes an equality if transduction through the signaling network makes this conditional mean of  $Z$  a linear function of the conditional mean of  $Z_c$  (SI Text). Eq. 8 implies an upper bound on the variation generated by signal transduction because we have already measured the sum of the last two terms of Eq. 7, the extrinsic variance, using  $\text{Cov}[Z, Z'|X]$ . To measure a component of Eq. 7 and to find bounds on the others, four reporters are therefore needed in one cell or three pairs of reporters in different cells:  $Z$  and  $Z'$ ;  $Z$  and  $Z_c$ ; and  $Z_c$  and  $Z'_c$  (Fig. 2C).

**Decomposing Variation in Yeast's Pheromone Response.** Such reporters have already been constructed by Colman-Lerner et al., who studied the pheromone pathway in budding yeast (15). In the presence of extracellular pheromone, this pathway activates a MAP kinase cascade and a transcription factor, Ste12 (29). Colman-Lerner et al. analyzed variation by equating the output of the pathway to the product of the downstream gene expression measured per unit of the upstream signaling response and the response from upstream signaling itself. They constructed, among others, three strains: one expressing two fluorescent proteins from two copies of the pheromone-responsive promoter PRM1 (equivalent to  $Z$  and  $Z'$ ); one expressing two fluorescent proteins from two copies of the promoter for actin (equivalent to  $Z_c$  and  $Z'_c$ ); and another expressing fluorescent proteins from the promoters of PRM1 and actin (equivalent to  $Z$  and  $Z_c$ ). We can therefore reanalyze their data using Eq. 7 (SI Text). We find that

the fraction of variance in  $Z$  generated by gene expression is low and approximately 0.1; that the fraction generated by processes extrinsic to gene expression (other than fluctuations in the transcription factor Ste12) is greater than approximately 0.5 (from Eq. 8); and that the fraction generated by signal transduction is therefore less than approximately 0.4 for cells exposed to 1.25 nM of pheromone. We can conclude that signal transduction is a less substantial source of variation than other processes extrinsic to gene expression for these data.

### Analyzing Variation in Models of Biochemical Systems

Given that we now have a general method for decomposing variance, can we make predictions of how the components of variance will behave as properties of the system change? Indeed, our method of conjugate reporters can also be used to compute the components of the variance for a given model. We can either use the chemical master equation to model both reporters and to calculate their covariance (30) or use Monte Carlo simulations to numerically estimate the covariances of the reporters (31).

**Calculating Transcriptional and Translational Variation.** For example, we calculated the three components of Eq. 3 for a standard model of gene expression (32) and with the assumption that the extrinsic variables  $Y_e$  are dominated by fluctuations in the rate of transcription (33). At steady state, with  $Z$  being the number of proteins, we find that the translational and the transcriptional components are (SI Text)

$$E\{V[Z(t)|(M, Y_e)^{\mathcal{Z}}]\} = E[Z], \quad [9]$$

and

$$E\{V[E[Z(t)|(M, Y_e)^{\mathcal{Z}}]|Y_e^{\mathcal{Z}}]\} = \frac{\tau_m}{\tau_m + \tau_z} \frac{E[Z]^2}{E[M]}, \quad [10]$$

where  $\tau_m$  denotes the lifetime of the mRNA and  $\tau_z$  denotes the lifetime of the protein. If  $\tau_e$  is the lifetime of fluctuations in the rate of transcription, then the extrinsic component is

$$V\{E[Z(t)|Y_e^{\mathcal{Z}}]\} = \frac{\tau_e(\tau_z\tau_e + \tau_m\tau_e + \tau_m\tau_z)}{(\tau_m + \tau_z)(\tau_e + \tau_m)(\tau_e + \tau_z)} E[Z]^2 \eta_e^2, \quad [11]$$

where  $\eta_e$  is the coefficient of variation of the fluctuations in the rate of transcription (their standard deviation over their mean). For this model, the translational component is equal to the mean number of proteins and the transcriptional component is determined partly by the coefficient of variation of fluctuations in mRNA levels (the square of which is the reciprocal of  $E[M]$ ) scaled by a function of the ratio of the mRNA to protein lifetimes.

**Intrinsic and Extrinsic Variation.** Throughout, we condition on the histories of stochastic variables (9–11, 34), and such conditioning is needed for a general definition of intrinsic and extrinsic variation because it allows for extrinsic variables that fluctuate on any timescale. If we let  $Y_e$  denote all extrinsic variables, then by conditioning on the history of  $Y_e$  we have

$$V[Z(t)] = \overbrace{E\{V[Z(t)|Y_e^{\mathcal{Z}}]\}}^{\text{intrinsic variation}} + \overbrace{V\{E[Z(t)|Y_e^{\mathcal{Z}}]\}}^{\text{extrinsic variation}}, \quad [12]$$

for expression of a protein  $Z$ . This decomposition is mathematically analogous to that given originally (4), except that the conditioning is on the history of the extrinsic variables rather than on just a single value of each extrinsic variable. The original conditioning is only valid if the extrinsic variables change substantially more slowly than the dynamics of the system of interest (34, 35). This change in definition alters neither the interpretation of

intrinsic variation as the average variability in gene expression between two identical copies of the gene subject to the same extrinsic fluctuations (empirically, the mean squared difference in levels of expression from the two genes) nor the measurements needed for experimental assays (34). These measurements are the covariance between a reporter for  $Z$  and another reporter conjugate given the history of the extrinsic variables (3), as required by our general decomposition.

Intrinsic variation is often heuristically understood as the “variation generated inherently” by the biochemical reactions that comprise the system (36, 37). However, intrinsic variation also depends on the environment in which the system is embedded, that is on the extrinsic processes themselves (4, 34, 35). This dependence is made explicit in Eq. 12:  $V[Z(t)|Y_e^{\mathcal{H}}]$  is unambiguously the variance due to processes that are not extrinsic, but its magnitude depends on the realized history of  $Y_e$ . We must therefore take its expectation over all such histories to obtain a useful numerical measure. We can also write the intrinsic variation as equal to the expectation of  $\{Z(t) - E[Z(t)|Y_e^{\mathcal{H}}]\}^2$ , that is the average squared deviation of  $Z(t)$  away from its mean given the history of the extrinsic variables.

An implication is that attempting to validate a model by comparing its variance components with experimental measurements can be problematic because the model must include the extrinsic processes appropriately. Detailed models of these processes, however, may not always be necessary. For example, a model specifying a mean, variance, and autocorrelation function for an extrinsic process could potentially give informative comparisons with data. Importantly, once a model has been selected and parameterized statistically, we can use conjugate reporters to evaluate its different components of variation in silico, some of which may not be easily measured experimentally. As an illustration, we considered constitutive gene expression for four different models each of which has a single source of extrinsic fluctuations, in either degradation or synthesis of mRNA or protein (*SI Text*). Calculating the intrinsic variation using the master equation augmented with a conjugate reporter, we found that extrinsic fluctuations in some processes increased the measurement of intrinsic noise in protein levels (the variability between two conjugate reporters normalized by the product of their means), and extrinsic fluctuations in others decreased it, relative to a model with no extrinsic fluctuations. Therefore, predictions from a model with only intrinsic fluctuations cannot even act as an upper or lower bound on the intrinsic noise measured experimentally. Once a model has been validated statistically, perhaps using time-series data, then results such as Eqs. 9–11 can be used to understand and predict the importance of different sources of variation. For example, Eq. 9 predicts that transcriptional variation usually dominates translational variation for typical parameters appropriate for *E. coli* (*SI Text*).

Finally, we point out that we have chosen an “operational” definition of intrinsic variation—the average variability in gene expression between two copies of the same gene under identical cellular conditions—in part because of the ease of constructing reporters by copying the promoter and other control regions of a gene of interest. If desired, this definition can be made more inclusive: the average variability of the output between two copies of the same subsystem under identical cellular conditions. Such a definition, for example, could include fluctuations in signal transduction in the example of Fig. 2C as “intrinsic,” with the first component in Eq. 4 then constituting the intrinsic variation.

## Discussion

A challenge when investigating variation in any biochemical network is the influence of the wider stochastic system in which the network is embedded. The general decomposition of variance we have introduced allows the variance to be decomposed into as many components as there are potential sources of variation and

provides exact mathematical expressions for each component. Further, we show that all components can be measured using reporters that are conjugate to the reporter for the original system. For each component, we give two conditions that these reporters must satisfy.

Through its use of conjugate reporters, our approach provides both a means to compute the magnitude of different components of variance for mathematical models and a framework for thinking about experimental approaches to quantifying sources of variation. Although we have described experimental realizations of conjugate reporters for two different types of decomposition (Fig. 2), the design of conjugate reporters to measure some components of variation may be challenging. For example, fluctuations in the numbers of mitochondria have been proposed to generate much of the variation in the rates of transcription seen in mammalian cells (33). To measure the contribution of such fluctuations to variation in gene expression, our decomposition requires a reporter conjugate to the output given only the history of the levels of mitochondria. Currently, constructing such a reporter is difficult, but this prediction in itself is important. Knowing the exact conditions required for a reporter is the first step in experimental design. If those conditions are challenging to establish, the experimenter can consider different decompositions of variance or alternative approaches. Using our techniques for quantifying sources of variation to analyze models whose parameters have been fitted to experimental data, for example, provides such an alternative approach. More generally, advances in synthetic biology that aim to create biochemistry “orthogonal” to the endogenous biochemistry may make suitably conjugate reporters commonplace (38).

Using our decomposition, we can identify a component of the variance determined by environmental signals or inputs—the informational variance—and as such can disentangle fluctuations that carry information from those that disrupt it. For given input distributions, measuring the fraction of the variance in the output that is informational allows quantitative comparison of the efficacy of different biochemical networks for information transfer and so addresses why one network architecture might be selected over another (39). The informational component can also be used for “inverse ecology”: to determine the distribution of inputs from which the network is best able to unambiguously transduce an input signal given the network’s structure (Fig. 3). Our analysis of data for osmosensing in budding yeast implies that the majority of variation can be generated by the response of the network to extracellular signals (we find that such environmental stochasticity can generate 80% of the variation in the response of the osmosensing network in yeast).

Stochasticity is now believed to pervade molecular and cellular biology, but the principal biochemical processes that generate stochasticity are mostly unknown. Our general decomposition of variance, each component of which can be evaluated with suitably constructed conjugate reporters, provides a means to quantify the effect of fluctuations in one biochemical process on the variation in the constituents of another. It provides a measure of the contribution of information flow to biochemical variation: a contribution that can be the dominant source of variation, at least for some distributions of input. Our approach thus provides a mathematical foundation for studies investigating the biological role of stochasticity and variation in cellular decision-making.

## Appendix

### General Decomposition of Variance for Stochastic Dynamic Systems.

We denote a stochastic dynamic system with  $n$  variables in all by  $\mathbf{Y} = \{Y_{1t}, \dots, Y_{n-1,t}, Z_t\}_{t \geq 0}$  and write, for example,  $Y_{1t}$  in place of  $Y_1(t)$ . We will use  $\mathcal{H}_{jt}$  to denote the history at time  $t$  of  $\mathbf{Y}_j$ , where  $\mathbf{Y}_j$  is some subset of the  $n$  variables of the full system. (The subscript  $j$  will index the different subsets or collections of variables; the subsets may overlap.) We note for those familiar

with stochastic process theory that  $\mathcal{H}_{jt} = \sigma(\mathbf{Y}_{js}; s \leq t)$ . Intuitively, knowledge of  $\mathcal{H}_{jt}$  is equivalent to knowing the trajectory of all of the variables in the vector  $\mathbf{Y}_j$  for times  $s$  up to and including time  $t$ .

Our general theorem for decomposition of variance is as follows. Let  $\mathbf{Y}$  be a stochastic dynamic system with a variable of interest  $Z_t$  whose variance exists (is finite). Suppose we have the histories  $\mathcal{H}_{1t}, \mathcal{H}_{2t}, \dots, \mathcal{H}_{kt}$ ,  $k \geq 2$ , each one for a different collection of variables of the system. Then,

$$V[Z_t] = E\{V[Z_t|\mathcal{H}_{1t}, \mathcal{H}_{2t}, \dots, \mathcal{H}_{kt}]\} + \sum_{j=2}^k E\{V[E[Z_t|\mathcal{H}_{1t}, \dots, \mathcal{H}_{jt}]]|\mathcal{H}_{1t}, \dots, \mathcal{H}_{j-1,t}\} + V\{E[Z_t|\mathcal{H}_{1t}]\}. \quad [13]$$

**Proof:** We now derive Eq. 13 (omitting the subscripts  $t$  in order to give the proof in its most general form where each  $\mathcal{H}_j$  is any  $\sigma$  field). To see that Eq. 14 below holds, notice that terms cancel to give the identity  $V[Z] = E\{V[Z]\}$ :

$$V[Z] = E\{V[Z|\mathcal{H}_1, \mathcal{H}_2, \dots, \mathcal{H}_k]\} + \sum_{j=2}^k E\{-V[Z|\mathcal{H}_1, \dots, \mathcal{H}_j] + V[Z|\mathcal{H}_1, \dots, \mathcal{H}_{j-1}]\} + E\{-V[Z|\mathcal{H}_1] + V[Z]\}. \quad [14]$$

For any two  $\sigma$  fields  $\mathcal{G}_1, \mathcal{G}_2$ , a conditional version of the law of total variance is seen to hold:

$$V[Z|\mathcal{G}_1] = V[E[Z|\mathcal{G}_1, \mathcal{G}_2]|\mathcal{G}_1] + E[V[Z|\mathcal{G}_1, \mathcal{G}_2]|\mathcal{G}_1], \quad [15]$$

and hence on taking (unconditional) expectations of both sides and rearranging,

$$E\{V[Z|\mathcal{G}_1]\} - E\{V[Z|\mathcal{G}_1, \mathcal{G}_2]\} = E\{V[E[Z|\mathcal{G}_1, \mathcal{G}_2]|\mathcal{G}_1]\},$$

because  $E\{E[V[Z|\mathcal{G}_1, \mathcal{G}_2]|\mathcal{G}_1]\} = E\{V[Z|\mathcal{G}_1, \mathcal{G}_2]\}$ . Therefore, the summation in Eq. 14 simplifies to give  $\sum_{j=2}^k E\{V[E[Z|\mathcal{H}_1, \dots, \mathcal{H}_j]|\mathcal{H}_1, \dots, \mathcal{H}_{j-1}]\}$ , [by setting  $\mathcal{G}_1 = \sigma(\mathcal{H}_1, \dots, \mathcal{H}_{j-1})$  and  $\mathcal{G}_2 = \mathcal{H}_j$  for  $j = 2, \dots, k$ ]. Using the law of total variance to simplify the last two terms of Eq. 14 completes the proof. When  $k = 1$  Eq. 13 simplifies to give a two-way decomposition: a measure theoretic version of the law of total variance. The two-way decomposition applied to the internal history  $(\mathcal{H}_{1t}, \mathcal{H}_{2t}, \dots, \mathcal{H}_{jt})$  is  $V[Z_t] = E\{V[Z_t|\mathcal{H}_{1t}, \mathcal{H}_{2t}, \dots, \mathcal{H}_{jt}]\} + V\{E[Z_t|\mathcal{H}_{1t}, \mathcal{H}_{2t}, \dots, \mathcal{H}_{jt}]\}$ . It then follows from Eq. 13 that for any  $j \geq 2$ ,

$$V\{E[Z_t|\mathcal{H}_{1t}, \mathcal{H}_{2t}, \dots, \mathcal{H}_{jt}]\} = V\{E[Z_t|\mathcal{H}_{1t}]\} + \sum_{i=2}^j E\{V[E[Z_t|\mathcal{H}_{1t}, \dots, \mathcal{H}_{it}]]|\mathcal{H}_{1t}, \dots, \mathcal{H}_{i-1,t}\}. \quad [16]$$

**Conjugate Reporters Identify all Components of the Variance Decomposition.** We define a reporter  $Z'_t$  to be first-moment conjugate to  $Z_t$  for the history  $\mathcal{H}_t$  if (i)  $Z_t$  and  $Z'_t$  are conditionally independent

given  $\mathcal{H}_t$  and (ii)  $E[Z_t|\mathcal{H}_t] = E[Z'_t|\mathcal{H}_t]$ . (It is implicit in our definition that neither the history  $\mathcal{H}_t$  nor the process  $Z_t$  is “affected” by introducing the reporter  $Z'_t$  into the system.) The following derivation shows that the measurement of variance components using the (contemporaneous) covariance of conjugate reporters at time  $t$  remains valid despite the “dependence” of  $Z_t$  on the histories of the relevant fluctuations as well as on the value of those variables at time  $t$  [see also Hilfinger and Paulson (34)].

We wish to measure each of the variance components in the decomposition given by Eq. 13. These measurements require a reporter  $Z'_t$  that is first-moment conjugate to  $Z_t$  for the history  $(\mathcal{H}_{1t}, \dots, \mathcal{H}_{jt})$  for each  $j = 1, \dots, k$ , which gives  $(k + 1)$  reporters in all including the original one for  $Z_t$ . Measurement of each of their covariances with  $Z_t$ , denoted  $\text{Cov}[Z_t, Z'_{jt}]$ , identifies all of the  $(k + 1)$  terms in the decomposition.

**Proof:** That  $\text{Cov}[Z_t, Z'_{jt}] = V\{E[Z_t|\mathcal{H}_{1t}, \dots, \mathcal{H}_{jt}]\}$  follows directly from the law of total covariance after noting that condition (i) for the conjugacy of the reporter (conditional independence) implies that the conditional covariance  $\text{Cov}[Z_t, Z'_{jt}|\mathcal{H}_{1t}, \dots, \mathcal{H}_{jt}]$  equals zero. Alternatively, the property may be derived as follows. Condition (ii) for the conjugacy of the reporter implies that  $E[Z_t] = E[Z'_{jt}]$ , and thus

$$\begin{aligned} \text{Cov}[Z_t, Z'_{jt}] &= E\{E[Z_t Z'_{jt}|\mathcal{H}_{1t}, \dots, \mathcal{H}_{jt}]\} - E[Z_t]^2 \\ &= E\{E[Z_t|\mathcal{H}_{1t}, \dots, \mathcal{H}_{jt}]^2\} \\ &\quad - \{E[E[Z_t|\mathcal{H}_{1t}, \dots, \mathcal{H}_{jt}]]\}^2 \\ &= V\{E[Z_t|\mathcal{H}_{1t}, \dots, \mathcal{H}_{jt}]\}, \end{aligned}$$

where we have again used the conditional independence of condition (i). It follows from Eq. 16 that  $\text{Cov}[Z_t, Z'_{jt}]$  identifies the sum of the last  $j$  terms in the decomposition in Eq. 13. Therefore,  $\text{Cov}[Z_t, Z'_{jt}] - \text{Cov}[Z_t, Z'_{j-1,t}]$  equals the  $j$ th term from the end of the equation, for  $j = 1, \dots, k$ . The first term of Eq. 13 is obtained as  $V[Z_t] - \text{Cov}[Z_t, Z'_{kt}]$ .

Furthermore, suppose that, for  $j = 1, \dots, k$ ,  $Z'_t$  has the same second conditional moment as  $Z_t$ :  $E[Z'_t{}^2|\mathcal{H}_{1t}, \dots, \mathcal{H}_{jt}] = E[Z_t{}^2|\mathcal{H}_{1t}, \dots, \mathcal{H}_{jt}]$ . We then say that  $Z'_t$  is conjugate to  $Z_t$  for the history  $\mathcal{H}_{1t}, \dots, \mathcal{H}_{jt}$ . The variance now decomposes as

$$V[Z_t] = \frac{1}{2} E[(Z_t - Z'_t)^2] + \text{Cov}[Z_t, Z'_t], \quad [17]$$

because  $E[(Z_t - Z'_t)^2] = E\{E[(Z_t - Z'_t)^2|\mathcal{H}_{1t}, \dots, \mathcal{H}_{jt}]\} = 2E\{E[Z_t^2|\mathcal{H}_{1t}, \dots, \mathcal{H}_{jt}] - E[Z_t|\mathcal{H}_{1t}, \dots, \mathcal{H}_{jt}]^2\} = 2E\{V[Z_t|\mathcal{H}_{1t}, \dots, \mathcal{H}_{jt}]\}$ . Eq. 17 formally justifies the empirical measure of intrinsic noise proposed by Elowitz et al. (3).

**ACKNOWLEDGMENTS.** We thank Tannie Liverpool, Andrew Millar, Ted Perkins, Guido Sanguinetti, and Margaritis Voliotis for critical comments on an earlier version of the manuscript and particularly Alejandro Colman-Lerner, Serge Pelet, and Matthias Peter for providing a copy of their data. C.G.B. acknowledges support under a Medical Research Council–Engineering and Physical Sciences Research Council funded Fellowship in Bioinformatics. P.S.S. is supported by a Scottish Universities Life Sciences Alliance chair in Systems Biology.

1. Eldar A, Elowitz MB (2010) Functional roles for noise in genetic circuits. *Nature* 467:167–173.
2. Raj A, van Oudenaarden A (2008) Nature, nurture, or chance: Stochastic gene expression and its consequences. *Cell* 135:216–226.
3. Elowitz MB, Levine AJ, Siggia ED, Swain PS (2002) Stochastic gene expression in a single cell. *Science* 297:1183–1186.
4. Swain PS, Elowitz MB, Siggia ED (2002) Intrinsic and extrinsic contributions to stochasticity in gene expression. *Proc Natl Acad Sci USA* 99:12795–12800.

5. Ozbudak EM, Thattai M, Kurtser I, Grossman AD, van Oudenaarden A (2002) Regulation of noise in the expression of a single gene. *Nat Genet* 31:69–73.
6. Blake WJ, Kaern M, Cantor CR, Collins JJ (2003) Noise in eukaryotic gene expression. *Nature* 422:633–637.
7. Raser JM, O’Shea EK (2004) Control of stochasticity in eukaryotic gene expression. *Science* 304:1811–1814.
8. Rinott R, Jaimovich A, Friedman N (2011) Exploring transcription regulation through cell-to-cell variability. *Proc Natl Acad Sci USA* 108:6329–6334.



9. Rausenberger J, Kollmann M (2008) Quantifying origins of cell-to-cell variations in gene expression. *Biophys J* 95:4523–4528.
10. Bowsher CG (2011) Information processing by biochemical networks: A dynamic approach. *J R Soc Interface* 8:186–200.
11. Bowsher CG (2010) Stochastic kinetic models: Dynamic independence, modularity and graphs. *Ann Stat* 38:2242–2281.
12. Swain PS (2004) Efficient attenuation of stochasticity in gene expression through post-transcriptional control. *J Mol Biol* 344:965–976.
13. Kollmann M, Lovdok L, Bartholomé K, Timmer J, Sourjik V (2005) Design principles of a bacterial signalling network. *Nature* 438:504–507.
14. Brivanlou AH, Darnell JE (2002) Signal transduction and the control of gene expression. *Science* 295:813–818.
15. Colman-Lerner A, et al. (2005) Regulated cell-to-cell variation in a cell-fate decision system. *Nature* 437:699–706.
16. Perkins TJ, Swain PS (2009) Strategies for cellular decision-making. *Mol Syst Biol* 5:326.
17. Cheong R, Rhee A, Wang CJ, Nemenman I, Levchenko A (2011) Information transduction capacity of noisy biochemical signaling networks. *Science* 334:354–358.
18. Tkacik G, Callan CG, Bialek W (2008) Information flow and optimization in transcriptional regulation. *Proc Natl Acad Sci USA* 105:12265–12270.
19. Pelet S, et al. (2011) Transient activation of the HOG MAPK pathway regulates bimodal gene expression. *Science* 332:732–735.
20. Shannon CE, Weaver W (1999) *The Mathematical Theory of Communication* (University of Illinois Press, Urbana, IL).
21. Tostevin F, Wolde PRT (2010) Mutual information in time-varying biochemical systems. *Phys Rev E Stat Nonlin Soft Matter Phys* 81:061917.
22. Hohmann S (2002) Osmotic stress signaling and osmoadaptation in yeasts. *Microbiol Mol Biol Rev* 66:300–372.
23. Dekel E, Mangan S, Alon U (2005) Environmental selection of the feed-forward loop circuit in gene-regulation networks. *Phys Biol* 2:81–88.
24. Libby E, Perkins TJ, Swain PS (2007) Noisy information processing through transcriptional regulation. *Proc Natl Acad Sci USA* 104:7151–7156.
25. Kalisky T, Dekel E, Alon U (2007) Cost-benefit theory and optimal design of gene regulation functions. *Phys Biol* 4:229–245.
26. Tkacik G, Walczak AM, Bialek W (2009) Optimizing information flow in small genetic networks. *Phys Rev E Stat Nonlin Soft Matter Phys* 80:031920.
27. Mitchell A, et al. (2009) Adaptive prediction of environmental changes by microorganisms. *Nature* 460:220–224.
28. Pedraza JM, van Oudenaarden A (2005) Noise propagation in gene networks. *Science* 307:1965–1969.
29. Bardwell L (2005) A walk-through of the yeast mating pheromone response pathway. *Peptides* 26:339–350.
30. Van Kampen NG (1981) *Stochastic Processes in Physics and Chemistry* (North-Holland, Amsterdam).
31. Gillespie DT (1977) Exact stochastic simulation of coupled chemical reactions. *J Phys Chem* 81:2340–2361.
32. Thattai M, van Oudenaarden A (2001) Intrinsic noise in gene regulatory networks. *Proc Natl Acad Sci USA* 98:8614–8619.
33. Neves RPD, et al. (2010) Connecting variability in global transcription rate to mitochondrial variability. *PLoS Biol* 8:e1000560.
34. Hilfinger A, Paulsson J (2011) Separating intrinsic from extrinsic fluctuations in dynamic biological systems. *Proc Natl Acad Sci USA* 108:12167–12172.
35. Shahrezaei V, Ollivier JF, Swain PS (2008) Colored extrinsic fluctuations and stochastic gene expression. *Mol Syst Biol* 4:196.
36. Kaern M, Elston TC, Blake WJ, Collins JJ (2005) Stochasticity in gene expression: From theories to phenotypes. *Nat Rev Genet* 6:451–464.
37. Shahrezaei V, Swain PS (2008) The stochastic nature of biochemical networks. *Curr Opin Biotechnol* 19:369–374.
38. An W, Chin JW (2009) Synthesis of orthogonal transcription-translation networks. *Proc Natl Acad Sci USA* 106:8477–8482.
39. Cağatay T, Turcotte M, Elowitz MB, Garcia-Ojalvo J, Süel GM (2009) Architecture-dependent noise discriminates functionally analogous differentiation circuits. *Cell* 139:512–522.

# Supporting Information

Bowsher and Swain 10.1073/pnas.1119407109

## SI Text

### A Brief Primer on Conditional Expectations

Conditional expectations are not commonly used outside of probability and statistics. We present here a short introduction and list of their properties.

For two random variables  $X$  and  $Y$ , the conditional expectation  $E[X|Y]$  is itself a random variable because it is a function of the random variable  $Y$ . For continuous random variables,  $E[X|Y]$  is defined as

$$E[X|Y = y] = \int dx x p(x|y), \quad [\text{S1}]$$

where  $p(x|y)$  is the conditional probability density of  $X$  given  $Y$ . It satisfies  $p(x|y) = p(x, y)/p(y)$ , by Bayes's rule. The conditional expectation has the following properties for any three random variables  $X$ ,  $Y$  and  $Z$ :

i. If  $X$  and  $Y$  are independent, then

$$E[X|Y] = E[X]. \quad [\text{S2}]$$

ii. If  $X$  and  $Y$  are conditionally independent given  $Z$ , then

$$E[X|Y, Z] = E[X|Z]. \quad [\text{S3}]$$

iii. For real constants  $a$  and  $b$

$$E[aX + bY|Z] = aE[X|Z] + bE[Y|Z]. \quad [\text{S4}]$$

iv. If knowing the random variable  $Z$  implies that  $X$  is known, then

$$E[XY|Z] = XE[Y|Z] \quad [\text{S5}]$$

and so  $E[g(Z)|Z] = g(Z)$  for any (measurable) function  $g(Z)$ .

v. If knowing  $Z$  implies that  $X$  is known, then

$$E[Y|X] = E[E[Y|Z]|X]. \quad [\text{S6}]$$

vi. For any  $Y$ ,

$$E[E[X|Y]] = E[X]. \quad [\text{S7}]$$

The conditional expectation  $E[X|Y]$  is, in the sense of minimizing the mean squared error, the best approximation to  $X$ . For any real-valued function  $f(Y)$ , it can be shown that

$$E[(X - f(Y))^2] \geq E[(X - E[X|Y])^2]. \quad [\text{S8}]$$

### Useful Additional Properties of Conditionally Independent Reporters

Suppose that  $Z'_t$  is conditionally independent of  $Z_t$  given  $(X, \mathcal{H}_t)$ , where  $X$  is a time invariant random variable (for example, one controlled in an experiment or set by the environment) and  $\mathcal{H}_t$  is some history. Then

$$\begin{aligned} \text{Cov}[Z_t, Z'_t|X] &= E\{E[Z_t Z'_t|X, \mathcal{H}_t]|X\} - E\{E[Z_t|X, \mathcal{H}_t]|X\} \\ &\quad \cdot E\{E[Z'_t|X, \mathcal{H}_t]|X\} \\ &= \text{Cov}[E[Z_t|X, \mathcal{H}_t], E[Z'_t|X, \mathcal{H}_t]|X], \end{aligned}$$

[S9]

where we have used the conditional independence of  $Z_t$  and  $Z'_t$ . If  $Z'_t$  is also first-moment conjugate to  $Z_t$  for  $(X, \mathcal{H}_t)$ , then we shall show that the covariance of the reporters conditional on  $X$  identifies the second component of the decomposition of conditional variance,  $V[Z_t|X] = E[V[Z_t|X, \mathcal{H}_t]|X] + V[E[Z_t|X, \mathcal{H}_t]|X]$ . When the two reporters are first-moment conjugate,  $E[Z_t|X, \mathcal{H}_t] = E[Z'_t|X, \mathcal{H}_t]$ , which implies that  $E[Z_t|X] = E[Z'_t|X]$ , and therefore

$$\begin{aligned} \text{Cov}[Z_t, Z'_t|X] &= \text{Cov}[E[Z_t|X, \mathcal{H}_t], E[Z'_t|X, \mathcal{H}_t]|X] \\ &= V\{E[Z_t|X, \mathcal{H}_t]|X\}. \end{aligned} \quad [\text{S10}]$$

Consequently, the average conditional covariance (averaging with respect to the distribution of  $X$ ) gives  $E\{\text{Cov}[Z_t, Z'_t|X]\} = E\{V[E[Z_t|X, \mathcal{H}_t]|X]\}$ .

### Transcriptional and Translational Variance: Reaction Network and Parameter Values Used in Simulations for Fig. 2B

We used the *Facile* compiler (1) and the *EasyStoch* simulator (2), which encodes the Gibson–Bruck (3) version of the Gillespie algorithm (4). We specify the model and the parameters used to generate the data underlying Fig. 2B in the main paper in the format employed by *Facile* (see Table S1). Comments are marked with a hash and the initial numbers of molecules are denoted with N. Any chemical species not specified initially has zero molecules.

For convenience, we simultaneously simulate three reporters (the original system of interest, a copy, and a bistronic reporter). To model extrinsic fluctuations in  $v_0$  the rate of transcription, we use ‘dummy’ chemical species, S1, S2, and S3, to control the propensity of transcription. Only one of these species exists at any given time and transitions between the three forms of S generate transitions in the value of  $v_0$ . We denote the protein reporter equivalent to  $Z$  in Fig. 2A as B, the reporter equivalent to  $Z'$  as A, and the reporter equivalent to  $Z''$  as C.

### Interpreting Scatter Plots of Measurements of Reporters

Plotting single-cell measurements of one reporter,  $Z$ , against measurements for a reporter conjugate to  $Z$  given some history  $Y^{\mathcal{H}}$  (denoted  $Z'$ ) gives a scatter plot where the extents of the scatter of points parallel and perpendicular to the  $Z = Z'$  diagonal measure different components of the variance. A typical example is shown in Fig. S1. Each point represents measurements of a reporter and its conjugate in a single cell and has coordinates  $(Z, Z')$ . Note that each reporter has the same mean value from the conditions of conjugacy. For each point, we can define  $d_{\perp}$ , which measures the distance from the point to the  $Z = Z'$  diagonal, and  $d_{\parallel}$ , which measures the distance along the diagonal that the point lies from the point corresponding to the mean value (Fig. S1).

We can show that the mean value of  $d_{\perp}^2$ , the spread of the points perpendicular to the  $Z = Z'$  diagonal, satisfies

$$E[d_{\perp}^2] = \frac{1}{2}E[(Z - Z')^2] \quad \text{[S11]}$$

because the point of intersection (red dot in Fig. S1) is  $((Z + Z')/2, (Z + Z')/2)$ . For any point  $(Z, Z')$ ,  $d_{\perp}^2$  is then

$$d_{\perp}^2 = \left(Z' - \frac{Z + Z'}{2}\right)^2 + \left(Z - \frac{Z + Z'}{2}\right)^2, \quad \text{[S12]}$$

giving Eq. S11 taking expectations. The right-hand side of Eq. S11 corresponds generally to a sum of terms in the decomposition of variance, with the particular sum being determined by the choice of the conditioning used to select the conjugate reporter.

Similarly, the mean value of  $d_{\parallel}^2$ , the spread of the points along the diagonal, satisfies

$$E[d_{\parallel}^2] = \text{Cov}[Z, Z'] + \frac{1}{2}E[(Z - Z')^2] + \text{Cov}[Z, Z']. \quad \text{[S13]}$$

If  $E[(Z - Z')^2]/2$  corresponds to a sum of terms in the decomposition of variance, then  $\text{Cov}[Z, Z']$  corresponds to the sum of the remaining terms. By definition,

$$\begin{aligned} d_{\parallel}^2 &= \left(\frac{Z + Z'}{2} - E[Z]\right)^2 + \left(\frac{Z + Z'}{2} - E[Z]\right)^2 \\ &= \frac{1}{2}\left(Z - E[Z] + Z' - E[Z]\right)^2, \end{aligned} \quad \text{[S14]}$$

implying that

$$E[d_{\parallel}^2] = V[Z] + \text{Cov}[Z, Z'] \quad \text{[S15]}$$

and giving Eq. S13 because  $V[Z] = E[(Z - Z')^2]/2 + \text{Cov}[Z, Z']$  (Eq. 17 in the main text).

### Translational Variation: Analyzing the Data of Kollmann et al.

Using similar arguments to those given in the *Appendix* section of the main text, we can also show that

$$\begin{aligned} E[(Z - Z_c)^2] &= E\{V[Z|Y^{\mathcal{H}}]\} + E\{V[Z_c|Y^{\mathcal{H}}]\} \\ &\quad + E\{(E[Z|Y^{\mathcal{H}}] - E[Z_c|Y^{\mathcal{H}}])^2\} \end{aligned} \quad \text{[S16]}$$

if  $Z$  and  $Z_c$  are conditionally independent given some history  $Y^{\mathcal{H}}$ . Consequently,

$$E[(Z - Z_c)^2] \geq E\{V[Z|Y^{\mathcal{H}}]\} + E\{V[Z_c|Y^{\mathcal{H}}]\}, \quad \text{[S17]}$$

where both terms on the right-hand side can be measured using conjugate reporters with equal second conditional moments:  $E\{V[Z|Y^{\mathcal{H}}]\}$  is equal to  $E[(Z - Z')^2]/2$  if  $Z'$  is a reporter conjugate to  $Z$  given  $Y^{\mathcal{H}}$ , and  $E\{V[Z_c|Y^{\mathcal{H}}]\}$  is equal to  $E[(Z_c - Z'_c)^2]/2$  if  $Z'_c$  is a reporter conjugate to  $Z_c$  given  $Y^{\mathcal{H}}$ .

Kollmann et al. (5) measured gene expression of the chemotaxis proteins CheY, tagged with YFP, and CheZ, tagged with CFP, with both proteins expressed from the same mRNA. CheY-YFP and CheZ-CFP should be conditionally independent given the joint history of the levels of the bicistronic mRNA,  $M$ , and the stochastic variables extrinsic to gene expression,  $Y_e$ . We can therefore use Eq. S17 with CheY-YFP denoted by

$Z$ , CheZ-CFP denoted by  $Z_c$ , and  $Y^{\mathcal{H}}$  being the joint history of  $M$  and  $Y_e$ . Kollmann et al. found that

$$E[(Z - Z_c)^2] \simeq 2 \times 0.2^2 E[Z]E[Z_c], \quad \text{[S18]}$$

where  $Z$  and  $Z_c$  are measured in fluorescence units (5). Consequently,

$$\begin{aligned} &\frac{1}{2}\left[E\{V[Z|(M, Y_e)^{\mathcal{H}}]\} + E\{V[Z_c|(M, Y_e)^{\mathcal{H}}]\}\right] \\ &\leq 0.2^2 E[Z]E[Z_c], \end{aligned} \quad \text{[S19]}$$

and therefore the average translational variance for the two proteins, normalized by the product of their mean fluorescences, is less than  $0.2^2$ .

### Finding Bounds on Components of the Variance Decomposition

Suppose that the reporter  $Z_t''$  is conditionally independent of the reporter  $Z_t$  given  $(X, \mathcal{H}_t)$ , where  $X$  is again a time invariant random variable. To find a lower bound on  $V\{E[Z_t|X, \mathcal{H}_t]\}$ , we begin with a conditional form of the Cauchy-Schwarz inequality:

$$\text{Cov}[W, W''|X]^2 \leq V[W|X] \cdot V[W''|X], \quad \text{[S20]}$$

for arbitrary random variables  $W$  and  $W''$ . From Eq. S9,

$$\text{Cov}[Z_t, Z_t''|X] = \text{Cov}[E[Z_t|X, \mathcal{H}_t], E[Z_t''|X, \mathcal{H}_t]|X],$$

and therefore the Cauchy-Schwarz inequality directly implies that

$$V\{E[Z_t|X, \mathcal{H}_t]|X\} \geq \frac{\text{Cov}[Z_t, Z_t''|X]^2}{V\{E[Z_t''|X, \mathcal{H}_t]|X\}}, \quad \text{[S21]}$$

where the denominator  $V\{E[Z_t''|X, \mathcal{H}_t]|X\}$  can itself be measured by the covariance (conditional on  $X$ ) between  $Z_t''$  and a reporter conjugate to  $Z_t''$  for the conditioning  $(X, \mathcal{H}_t)$ . The lower bound in Eq. S21 becomes an equality when  $E[Z_t|X, \mathcal{H}_t]$  is a linear function of  $E[Z_t''|X, \mathcal{H}_t]$ .

### Distinguishing Variation due to Gene Expression from Variation due to Signal Transduction: Analyzing the Data of Colman-Lerner et al.

Colman-Lerner et al. used the promoter for PRM1 driving YFP to quantify the response of budding yeast to pheromone (6). From Eqs. 7 and 8, we can write an inequality for the variation generated by signal transduction:

$$\begin{aligned} &\overbrace{E\{V[E[Z|(Y_{eT}, T)^{\mathcal{H}}, X]|Y_{eT}, X]|X\}}^{\text{from signal transduction given } X} \\ &\leq \text{Cov}[Z, Z'|X] - \frac{\text{Cov}[Z, Z_c|X]^2}{\text{Cov}[Z_c, Z'_c|X]}, \end{aligned} \quad \text{[S22]}$$

where  $Z$  is a reporter for the output of the system;  $Z'$  is a reporter conjugate to  $Z$  given the history of all extrinsic variables;  $Z_c$  is a reporter for a constitutively expressed gene; and  $Z'_c$  is a reporter conjugate to  $Z_c$  given the history of extrinsic variables (Fig. 2C). Alejandro Colman-Lerner kindly provided: average fluorescence measurements (the total fluorescence in individual cells divided by the area of the cells) of a strain expressing both YFP and CFP from two copies of the promoter for PRM1 across a population of

172 cells (corresponding to  $Z$  and  $Z'$  in Eq. S22); average fluorescence measurements of a strain expressing both YFP and CFP from two copies of the promoter for ACT1 (actin) across a population of 292 cells (corresponding to  $Z_c$  and  $Z'_c$  in Eq. S22); and average fluorescence measurements of a strain expressing CFP from the promoter of ACT1 and YFP from the promoter of PRM1 across a population of 233 cells.

To adjust for the different brightness of CFP and YFP, we corrected the measurements of CFP in the strain expressing CFP and YFP from the promoter of PRM1 to have the same median as the measurements of YFP. We corrected the measurements of CFP in the strain expressing CFP and YFP from the promoter of ACT1 and the measurements of CFP in the strain expressing CFP from ACT1 and YFP from the promoter of PRM1 to have the same median as measurements of YFP in the strain expressing CFP and YFP from the promoter for ACT1.

Both the CFP and YFP measurements should also be corrected for cellular autofluorescence, although autofluorescence is less of a problem for YFP because it is brighter. We were unable to correct the data for autofluorescence, and numerical values should be interpreted with this caveat.

### The Informational Fraction of Variance

**Some Intuition.** For an output  $Z$  and input  $X$ , let  $E[Z|X] = \mu(X)$  and  $V[Z|X] = \sigma^2(X)$ . Then,

$$V[Z] = V\{E[Z|X]\} + E\{V[Z|X]\} = V[\mu(X)] + E[\sigma^2(X)], \quad \text{[S23]}$$

and the informational fraction of the output variance is

$$\begin{aligned} \iota_{Z|X} &= \frac{V\{E[Z|X]\}}{V[Z]} = \frac{V[\mu(X)]}{V[\mu(X)] + E[\sigma^2(X)]} \\ &= \left(1 + \frac{E[\sigma^2(X)]}{V[\mu(X)]}\right)^{-1}. \end{aligned} \quad \text{[S24]}$$

In the main text, we denote  $\iota_{Z|X}$  by  $\iota_Z$ .

Imagine drawing two independent realizations of the input  $X$  from its distribution, denoted by  $X_1, X_2$ . Write the corresponding expected outputs conditional on the realized inputs as  $\mu_1$  and  $\mu_2$ , where  $\mu_i = \mu(X_i)$ . Then, the typical distance between the two conditional means obtained is

$$\frac{1}{2}E[(\mu_1 - \mu_2)^2] = V[\mu_i], \quad \text{[S25]}$$

because  $E[\mu_1\mu_2] = E[\mu_1]^2$ . The expected conditional variance for each draw is simply  $E[\sigma_i^2]$ . Therefore, as the informational fraction tends to one,

$$\frac{1}{2}E[(\mu_1 - \mu_2)^2] \gg E[\sigma^2(X)], \quad \text{[S26]}$$

and the typical distance between the means of a pair of conditional distributions for the output  $Z$  becomes much larger than the expected variability or “width” of those distributions: The conditional output distributions typically overlap less. Heuristically, each output distribution is less likely to overlap with another and the system should become more efficient at transmitting information. We make these ideas more precise by providing formal connections between the informational fraction and information theory below.

**Input and Output with a Jointly Gaussian Distribution and a General Upper Bound on the Conditional Entropy of the Output.** Consider the

input and output  $(X, Z)$  to be a continuous random vector, with the support of  $Z$  equal to  $(-\infty, \infty)$ . Let  $z$  be the rescaled output with variance equal to 1,  $z = Z/V[Z]^{1/2}$ . The rescaling affects neither the informational fraction, nor the mutual information of input and output (7). Note that  $1 - \iota_{Z|X} = E\{V[z|X]\}$  and that the entropy of a Gaussian distribution with variance  $v$  is equal to  $\frac{1}{2}\ln(2\pi ev)$ . Now  $V[z|X = x] \geq \frac{1}{2\pi e}\exp\{2h(z|X = x)\}$  because the Gaussian distribution has the maximum entropy for a given variance. It follows after taking the expectation of both sides of the inequality and applying Jensen’s inequality that

$$1 - \iota_{Z|X} = E\{V[z|X]\} \geq \frac{1}{2\pi e}\exp\{2h(z|X)\},$$

where  $h(z|X) = E_x[h(z|X = x)]$ . Therefore, an upper bound for the conditional entropy of the rescaled output is given by

$$h(z|X) \leq \frac{1}{2}\ln\{2\pi e[1 - \iota_{Z|X}]\}. \quad \text{[S27]}$$

The upper bound decreases as  $\iota_{Z|X}$  increases, placing an upper limit on how uncertain the output can be given the state of the input.

When the signaling mechanism obeys “Gaussian statistics,” or more precisely the conditional distribution of output given input,  $p(z|X)$ , is Gaussian with variance not depending on  $X$ , it is seen that  $V[z|X] = \frac{1}{2\pi e}\exp\{2h(z|X)\}$  and therefore Eq. S27 holds with equality in this case. If the input  $X$  is also normally distributed then  $(X, Z)$  has a bivariate normal distribution and  $z$  is therefore normally distributed with variance equal to 1. The mutual information  $I(X; Z) = I(X; z)$ . It follows directly that

$$\begin{aligned} I(X; Z) &= h(z) - h(z|X) = \frac{1}{2}\ln\{2\pi e\} - \frac{1}{2}\ln\{2\pi e[1 - \iota_{Z|X}]\} \\ &= -\frac{1}{2}\ln\{1 - \iota_{Z|X}\}, \end{aligned} \quad \text{[S28]}$$

which is familiar on recognizing that  $\iota_{Z|X} = \text{Corr}[Z, X]^2$ , because  $E[Z|X]$  is a linear function of  $X$  (see Eq. S31) for a bivariate normal distribution. Throughout, we define the correlation of any two random variables  $T, U$  to be  $\text{Corr}[T, U] = \text{Cov}[T, U]/V[T]^{1/2}V[U]^{1/2}$ .

**A Lower Bound on Information Capacity Is Set by  $\iota_{Z|X}$ .** We will now prove that when the conditional means  $\mu(X)$  are different for all values of  $X$ , the information capacity  $C$  of the biochemical mechanism satisfies the lower bound given by

$$C = \sup_{p(X)} I(X; Z) \geq \frac{1}{2}\ln[1 - \iota_{Z|\tilde{X}}]^{-1}, \quad \text{[S29]}$$

where the supremum (“maximum”) is taken over a set of possible input distributions,  $\mathcal{S}$ . The informational fraction is evaluated for an input distribution corresponding to a Gaussian distribution for  $\mu(X)$ . The higher the informational fraction, the larger the lower bound on the capacity.

Consider the mechanism

$$W \xrightarrow{f} X \xrightarrow{p(Z|X)} Z, \quad \text{[S30]}$$

where  $p(Z|X)$  represents the biochemical transduction,  $Z \in (-\infty, \infty)$ ,  $W \in (-\infty, \infty)$ , and the function  $f$  is a continuously differentiable, one-to-one mapping. Where necessary, we transform the biophysical output (which is often positive), for example by taking its logarithm, so that  $Z$  is real-valued. For any distribution  $p(W)$ , we have that  $I(W; Z) = I(X; Z)$  (7).

We will need the following result. For any two random variables  $T, U$ , the informational fraction satisfies the equality

$$i_{U|T} = \text{Corr}(U, E[U|T])^2, \quad \text{[S31]}$$

from the definition of correlation and because

$$\begin{aligned} \text{Cov}(U, E[U|T]) &= E\{E[(U - E[U])(E[U|T] - E[U])|T]\} \\ &= E\{E[U|T]^2 - 2E[U]E[U|T] + E[U]^2\} \\ &= V\{E[U|T]\}. \end{aligned}$$

Therefore,  $i_{Z|W} = \text{Corr}(Z, E[Z|W])^2$ . Furthermore, if  $E[Z|W]$  is a linear (affine) function of  $W$ , then  $\text{Corr}(Z, E[Z|W]) = \text{Corr}(Z, W)$  and  $i_{Z|W} = \text{Corr}(Z, W)^2$  for any distribution  $p(W)$ .

Notice that for the mechanism in Eq. S30, the random variables  $E[Z|X]$  and  $E[Z|W]$  are equal because  $f$  is an invertible function and conditioning on  $X$  is therefore equivalent to conditioning on  $W$  (mathematically, the conditioning sigma field  $\sigma(X) = \sigma(W)$ ). It follows immediately that the corresponding informational fractions are equal:

$$i_{Z|W} = i_{Z|X}, \quad \text{[S32]}$$

where we have used  $V[Z|X] = V[Z|W]$ .

The essential insights in the proof of the lower bound in Eq. S29 are to approach the problem via the “augmented” mechanism in Eq. S30 and to then notice that a certain choice of the function  $f$  will result in  $E[Z|W]$  being a linear function of  $W$ . This choice is useful because it is known how to bound the mutual information from below when the input is Gaussian, using the squared correlation of input and output. As we have seen, when the conditional expectation is a linear function of  $W$ ,  $i_{Z|W} = \text{Corr}(Z, W)^2$ . The choice of  $f$  is to set

$$X = \mu^{-1}(W) \quad \text{where } \mu(x) = E[Z|X = x].$$

The inverse biochemical “response” function,  $\mu^{-1}$ , is expected to be smooth (continuously differentiable) for biophysically reasonable response functions  $\mu(x)$ . When  $f$  is set equal to  $\mu^{-1}$  in Eq. S30,

$$E[Z|W = w] = E[Z|X = \mu^{-1}(w)] = \mu(\mu^{-1}(w)) = w,$$

or more concisely  $E[Z|W] = E[Z|X] = W$ , which is the linearity in  $W$  we set out to achieve.

It now follows, recalling Eq. S32, that  $\text{Corr}(Z, W)^2 = i_{Z|X}$  for any distribution  $p(W)$  and the implied input distribution  $p(X)$ . Let  $\tilde{W}$  denote the artificial input when that random variable has a Gaussian distribution. Then,  $I(\tilde{W}; Z) \geq \frac{1}{2} \ln[1 - \text{Corr}(Z, \tilde{W})^2]^{-1}$  (8). We have now shown that

$$\begin{aligned} C &\geq I(\tilde{X}; Z) = I(\tilde{W}; Z) \geq \frac{1}{2} \ln[1 - \text{Corr}(Z, \tilde{W})^2]^{-1} \\ &= \frac{1}{2} \ln[1 - i_{Z|\tilde{X}}]^{-1}, \end{aligned}$$

where  $\tilde{X} = \mu^{-1}(\tilde{W})$  has the distribution implied by the Gaussian distribution of  $\tilde{W}$ .

In biology, because natural input distributions have not been widely measured, the set  $\mathcal{S}$  of possible input distributions  $p(X)$  must be specified by the investigator, and a range of choices for  $\mathcal{S}$  may be entertained. To implement the capacity bound in Eq. S29, one can range over choices for the mean and variance of the

Gaussian  $\tilde{W}$ , excluding those choices that imply a distribution  $p(X)$  that one wants to omit from  $\mathcal{S}$ . Armed with the function  $\sigma^2(X) = V[Z|X]$ , both  $p(\tilde{X})$  and the informational fraction  $i_{Z|\tilde{X}}$  can be computed by Monte Carlo sampling from  $p(\tilde{W})$ , using the relation  $\tilde{X} = \mu^{-1}(\tilde{W})$ . To implement the capacity bound, the other input distributions in  $\mathcal{S}$  need not be specified. Finally, the informational fraction  $i_{Z|\tilde{X}}$  should be maximized over the set of distributions  $p(\tilde{X})$  given by the allowed means and variances for  $\tilde{W}$ .

As a simple illustration, consider the Gaussian noise channel of information theory given by  $Z = gX + \eta_{Z|X}$ , where  $g$  is a constant and  $\eta_{Z|X}$  is normally distributed with zero mean and a constant variance  $\sigma_{Z|X}^2$  that is not dependent on  $X$ . Let  $\mathcal{S}$  consist of input distributions satisfying  $E[X] = 0$  and  $V[X] \leq \sigma^2$ . Because  $W = gX$  here, we set  $E[\tilde{W}] = 0$  and  $V[\tilde{W}] \leq g^2\sigma^2$ . The informational fraction  $i_{Z|\tilde{X}} = g^2V[\tilde{X}]/(g^2V[\tilde{X}] + \sigma_{Z|X}^2)$ , which is maximized by setting  $V[\tilde{X}] = \sigma^2$ , which implies  $V[\tilde{W}] = g^2\sigma^2$ . The corresponding, maximized lower bound for the capacity given by Eq. S29 is then equal to  $\frac{1}{2} \ln[1 - i_{Z|\tilde{X}}]^{-1} = \frac{1}{2} \ln[1 + (g^2\sigma^2/\sigma_{Z|X}^2)]$ , which is exactly equal to the capacity of the Gaussian noise channel with input “power” constraint  $\sigma^2$ . Our lower bound on the information capacity is a tight one for the Gaussian channel.

**Information Transfer When  $i_{Z|X}$  Is Large.** Consider now a setting in which the biochemical mechanism and the input distribution can vary as  $n \rightarrow \infty$ , where  $n$  labels the sequence of mechanisms and input distributions. Suppose that  $i_{Z_n|X_n}$ , the informational fraction for  $Z_n$ , converges to its maximum value of 1 and that the unconditional distribution  $p(z_n)$  of the rescaled output of the signaling mechanism does not become ever less uncertain as  $n \rightarrow \infty$ . By uncertainty in the continuous case (or differential entropy), we mean the logarithm of the effective volume of the smallest set that contains most of the probability (9). We will give a concrete biochemical example of such asymptotic behavior below.

More precisely, suppose  $i_{Z_n|X_n} \rightarrow 1$  and that  $h(z_n) \not\rightarrow -\infty$  (or, equivalently,  $h(z_n)$  is bounded below by a constant for all  $n$ ), where  $z_n = Z_n/V[Z_n]^{1/2}$  as before. Then, Eq. S27 implies that  $-h(z_n|X_n) \rightarrow +\infty$  and hence

$$I(X_n; Z_n) = I(X_n; z_n) \rightarrow \infty \quad \text{as } i_{Z_n|X_n} \rightarrow 1. \quad \text{[S33]}$$

Biophysically reasonable transduction mechanisms are expected to give rise to unconditional distributions for the rescaled output,  $p(z_n)$ , that reflect the uncertainty of the input rather than having differential entropy that is unbounded below. If the input distribution varies as the limit is taken, we assume its uncertainty (differential entropy) does not become ever less as  $n \rightarrow \infty$ .

As an example of such asymptotics, suppose we hold the input distribution constant for simplicity and consider the linear noise approximation (LNA) of output at time  $t$ , for which (10)

$$Z_{t,n} = \Omega_n \phi(t, X) + \Omega_n^{1/2} \xi(t, X), \quad \text{[S34]}$$

where  $\Omega_n$  is the system size,  $\phi(t, X)$  is the deterministic solution for output concentration at time  $t$ , and the random variable  $\xi(t, X)$  can be shown in the case of the LNA not to depend on  $\Omega_n$  (11). Notice that Eq. S34 makes no assumption of Gaussianity. Let  $\tilde{Z}_{t,n} := Z_{t,n}/\Omega_n$  denote the output concentration. We can see that, as the system size  $\Omega_n \rightarrow \infty$ , then  $i_{Z_{t,n}|X} \rightarrow 1$  because

$$\frac{E\{V[\tilde{Z}_{t,n}|X]\}}{V\{E[\tilde{Z}_{t,n}|X]\}} = \frac{E\{V[\xi(t, X)|X]\}}{\Omega_n V[\phi(t, X)] + \Omega_n^{1/2} \text{Cov}\{\phi(t, X), E[\xi(t, X)|X]\} + V\{E[\xi(t, X)|X]\}} = \mathcal{O}(\Omega_n^{-1}).$$

Furthermore, it follows from Eq. S34 that  $\tilde{Z}_{t,n} \rightarrow \phi(t, X)$ : The output concentration converges (almost surely) to the deterministic solution, which is a function of  $X$ . Let  $\tilde{z}_{t,n} = \tilde{Z}_{t,n} V[\tilde{Z}_{t,n}]^{-1/2}$  be the rescaled output with a variance of 1. The differential entropy of the rescaled output  $h(\tilde{z}_{t,n}) \rightarrow h(\phi(t, X)) - \frac{1}{2} \ln V\{\phi(t, X)\}$ , under suitable regularity conditions. We make the mild assumption that the distribution of the continuous input is such that  $|h(\phi(t, X))| < \infty$  and  $V\{\phi(t, X)\} < \infty$ . It then follows from Eq. S27 and the above argument that  $I(X; Z_{t,n}) = I(X; \tilde{z}_{t,n}) \rightarrow \infty$  as  $\Omega_n \rightarrow \infty$ . Information transfer becomes perfect in the limit of large system size. Because the LNA tells us about moments but not distributions, it is not clear how to prove this property without making use of the informational fraction and the implied properties when the informational fraction tends to its maximal value of 1.

### Determining the Informational Fraction for Osmosensing in Budding Yeast

Pelet et al. (12) used YFP to report gene expression from the STL1 promoter for six different concentrations of extracellular salt. They recorded fluorescence levels from approximately 1,000 cells for each concentration of salt.

Letting  $P_i$  denoting the probability of the environment having a salt concentration equal to  $S_i$ , then the informational fraction for a given probability distribution of extracellular salt is

$$\begin{aligned} \text{informational fraction} &= \frac{V\{E[Z|S]\}}{V[Z]} \\ &= \frac{\sum_i P_i E[Z|S_i]^2 - (\sum_i P_i E[Z|S_i])^2}{\sum_i P_i E[Z^2|S_i] - (\sum_i P_i E[Z|S_i])^2}, \end{aligned} \quad [\text{S35}]$$

where we have used  $V[W] = E[W^2] - E[W]^2$ . Therefore, if  $y_{i,j}$  is the average fluorescence level of YFP in the  $j$ th cell (the total fluorescence in that cell divided by the area of the cell) when the salt concentration is  $S_i$  and there are  $N_i$  such cells, then our empirical measure of the informational fraction is given by

$$\frac{\sum_i P_i \left( \frac{1}{N_i} \sum_{j=1}^{N_i} y_{i,j} \right)^2 - \left[ \sum_i P_i \frac{1}{N_i} \sum_{j=1}^{N_i} y_{i,j} \right]^2}{\sum_i P_i \frac{1}{N_i} \sum_{j=1}^{N_i} y_{i,j}^2 - \left[ \sum_i P_i \frac{1}{N_i} \sum_{j=1}^{N_i} y_{i,j} \right]^2}. \quad [\text{S36}]$$

We exhaustively searched the possible probability distributions over the six different concentrations of extracellular salt and determined the probability distributions that have high informational fractions. We discretized  $P_i$  (to have 21 equally spaced values, between and including the values zero and one) and looped through all possible values of  $P_i$  for each  $i$ , calculating the informational fraction only when  $\sum_i P_i = 1$ .

For two reporters that are conjugate given the history of the stochastic variables extrinsic to gene expression, then the total extrinsic fraction for a particular concentration of salt is defined as the ratio of the covariance of the two reporters to the variance of the output  $Z$ :

$$\begin{aligned} \text{total extrinsic fraction} &= \frac{\text{Cov}[Z, Z'|S_i]}{V[Z]} \\ &= \frac{\sum_i P_i E[ZZ'|S_i] - (\sum_i P_i E[Z|S_i])(\sum_i P_i E[Z'|S_i])}{\sum_i P_i E[Z^2|S_i] - (\sum_i P_i E[Z|S_i])^2} \end{aligned} \quad [\text{S37}]$$

for the experiments of Pelet *et al.* If  $c_{i,j}$  is the average fluorescence measured from the CFP reporter in the  $j$ th cell when the concentration of salt is  $S_i$ , then our empirical measure of the total extrinsic fraction is given by

$$\frac{\sum_i P_i \frac{1}{N_i} \sum_{j=1}^{N_i} y_{i,j} c_{i,j} - \left( \sum_i P_i \frac{1}{N_i} \sum_{j=1}^{N_i} y_{i,j} \right) \left( \sum_i P_i \frac{1}{N_i} \sum_{j=1}^{N_i} c_{i,j} \right)}{\sum_i P_i \frac{1}{N_i} \sum_{j=1}^{N_i} y_{i,j}^2 - \left[ \sum_i P_i \frac{1}{N_i} \sum_{j=1}^{N_i} y_{i,j} \right]^2}. \quad [\text{S38}]$$

The two fluorescent proteins, CFP and YFP, have different brightness, and we multiply each  $c_{i,j}$  by a correction factor so that the median of the YFP measurements is equal to the median of the CFP measurements for each concentration of salt.

Both the CFP and YFP measurements should also be corrected for cellular autofluorescence, although autofluorescence is less of a problem for YFP because it is brighter. We were unable to correct the data for autofluorescence, and numerical values should be interpreted with this caveat.

### Calculation of the Variance Components from the Chemical Master Equation

The conjugate reporter method allows us to find analytical expressions for the components of the variance. Consider gene expression with one extrinsic variable (Fig. S24). In a mathematical model, this extrinsic variable corresponds to a fluctuating propensity for a particular reaction (2, 13), and we will begin with a fluctuating propensity for transcription. If this propensity has three states reflecting, for example, three states of the extracellular environment (Fig. S2B), then we can define  $\kappa_{01}$  to be the probability per unit time of transitioning from the state 0 to state 1 (conditional on being in state 0);  $\kappa_{10}$  to be the probability per unit time of transitioning back;  $\kappa_{12}$  to be the probability per unit time of transitioning from state 1 to state 2; and  $\kappa_{21}$  to be the probability per unit time of transitioning back. With all the  $\kappa_{ij}$  identical, we used this model to generate the data for Fig. 2B.

Initially, we will consider the decomposition of the variance into intrinsic and extrinsic components,

$$V[Z(t)] = E\{V[Z(t)|v_{0,t}^{\mathcal{H}}]\} + V\{E[Z(t)|v_{0,t}^{\mathcal{H}}]\}, \quad [\text{S39}]$$

and therefore require reporters that are conditionally independent given the history of  $v_0$ , the propensity for the transcriptional reaction, and that have the same means conditional on  $v_{0,t}^{\mathcal{H}}$ . An identical copy of the system exposed to the same fluctuations in  $v_0$  satisfies both these conditions (Fig. S24). We will denote the number of mRNAs from each copy of the system as  $m_1$  and  $m_2$  and the number of proteins from each copy as  $n_1$  and  $n_2$ . The probability of having  $m_1$  mRNAs and  $n_1$  proteins from the first copy and  $m_2$  mRNAs and  $n_2$  proteins from the second is  $P(m_1, n_1, m_2, n_2, v_0^{(i)}, t)$ , with  $i$  denoting the state of the extrinsic variable. For brevity, we will write  $P^{(i)}$  for  $P(m_1, n_1, m_2, n_2, v_0^{(i)}, t)$  and only explicitly write (with subscripts) the number of molecules when these differ from either  $m_1, n_1, m_2$ , or  $n_2$ . The corresponding master equations for the dual reporter systems are then (see Fig. S24 for definitions of the parameters)

$$\begin{aligned} \frac{\partial P^{(i)}}{\partial t} = & v_0^{(i)} [P_{m_1-1}^{(i)} - P^{(i)}] + d_0 [(m_1 + 1)P_{m_1+1}^{(i)} - m_1 P^{(i)}] \\ & + d_1 [(n_1 + 1)P_{n_1+1}^{(i)} - n_1 P^{(i)}] + v_1 m_1 [P_{n_1-1}^{(i)} - P^{(i)}] \\ & + v_0^{(i)} [P_{m_2-1}^{(i)} - P^{(i)}] + d_0 [(m_2 + 1)P_{m_2+1}^{(i)} - m_2 P^{(i)}] \\ & + d_1 [(n_2 + 1)P_{n_2+1}^{(i)} - n_2 P^{(i)}] + v_1 m_2 [P_{n_2-1}^{(i)} - P^{(i)}] \\ & + \begin{cases} \kappa_{10} P^{(1)} - \kappa_{01} P^{(0)} & \text{if } i = 0 \\ \kappa_{01} P^{(0)} - (\kappa_{10} + \kappa_{12}) P^{(1)} + \kappa_{21} P^{(2)} & \text{if } i = 1, \\ \kappa_{12} P^{(1)} - \kappa_{21} P^{(2)} & \text{if } i = 2 \end{cases} \end{aligned} \quad \text{[S40]}$$

where there is one equation for each state of the extrinsic variable (here,  $v_0^{(i)}$ ).

We can solve Eq. S40 exactly for the moments of the probability distribution  $P^{(i)}$ . We will use  $\mathbf{s}$  to represent the vector of numbers of species,  $\mathbf{s} = [m_1, n_1, m_2, n_2]$ , and, for brevity, we will use angled brackets to denote expectations:

$$\langle f(\mathbf{s}) \rangle_i = \sum_{\mathbf{s}} P(\mathbf{s}, v_0^{(i)}) f(\mathbf{s}) = \sum_{\mathbf{s}} P^{(i)}(\mathbf{s}) f(\mathbf{s}) \quad \text{[S41]}$$

for any function  $f(\mathbf{s})$  and where the expectation is taken with the particular value of  $v_0$  fixed. By multiplying Eq. S40 by either  $m_1$  or  $m_2$  and summing over all states described by  $P^{(i)}$  (over all values of  $m_1, m_2, n_1$ , and  $n_2$ ), we find that the mean mRNA for either copy then obeys

$$\begin{aligned} \frac{\partial \langle m \rangle_i}{\partial t} = & v_0^{(i)} \sum_{\mathbf{s}} P^{(i)} - d_0 \langle m \rangle_i \\ & + \begin{cases} \kappa_{10} \langle m \rangle_1 - \kappa_{01} \langle m \rangle_0 & \text{if } i = 0 \\ \kappa_{01} \langle m \rangle_0 - (\kappa_{10} + \kappa_{12}) \langle m \rangle_1 + \kappa_{21} \langle m \rangle_2 & \text{if } i = 1, \\ \kappa_{12} \langle m \rangle_1 - \kappa_{21} \langle m \rangle_2 & \text{if } i = 2 \end{cases} \end{aligned} \quad \text{[S42]}$$

where the sum over  $P^{(i)}$  is over all states of the system for a given  $v_0$ . Performing this summation in Eq. S40 gives

$$\begin{aligned} \frac{\partial}{\partial t} \sum_{\mathbf{s}} P^{(0)} &= \kappa_{10} \sum_{\mathbf{s}} P^{(1)} - \kappa_{01} \sum_{\mathbf{s}} P^{(0)}, \\ \frac{\partial}{\partial t} \sum_{\mathbf{s}} P^{(1)} &= \kappa_{01} \sum_{\mathbf{s}} P^{(0)} - (\kappa_{10} + \kappa_{12}) \sum_{\mathbf{s}} P^{(1)} + \kappa_{21} \sum_{\mathbf{s}} P^{(2)}, \\ \frac{\partial}{\partial t} \sum_{\mathbf{s}} P^{(2)} &= \kappa_{12} \sum_{\mathbf{s}} P^{(1)} - \kappa_{21} \sum_{\mathbf{s}} P^{(2)}, \end{aligned} \quad \text{[S43]}$$

and so

$$\begin{aligned} \sum_{\mathbf{s}} P^{(0)} &= \frac{\kappa_{10} \kappa_{21}}{\kappa_{10} \kappa_{21} + \kappa_{01} \kappa_{12} + \kappa_{01} \kappa_{21}}, \\ \sum_{\mathbf{s}} P^{(1)} &= \frac{\kappa_{01} \kappa_{21}}{\kappa_{10} \kappa_{21} + \kappa_{01} \kappa_{12} + \kappa_{01} \kappa_{21}}, \\ \sum_{\mathbf{s}} P^{(2)} &= \frac{\kappa_{01} \kappa_{12}}{\kappa_{10} \kappa_{21} + \kappa_{01} \kappa_{12} + \kappa_{01} \kappa_{21}} \end{aligned} \quad \text{[S44]}$$

at steady state. The mean protein for either copy satisfies

$$\begin{aligned} \frac{\partial \langle n \rangle_i}{\partial t} = & v_1 \langle m \rangle_i - d_1 \langle n \rangle_i \\ & + \begin{cases} \kappa_{10} \langle n \rangle_1 - \kappa_{01} \langle n \rangle_0 & \text{if } i = 0 \\ \kappa_{01} \langle n \rangle_0 - (\kappa_{10} + \kappa_{12}) \langle n \rangle_1 + \kappa_{21} \langle n \rangle_2 & \text{if } i = 1. \\ \kappa_{12} \langle n \rangle_1 - \kappa_{21} \langle n \rangle_2 & \text{if } i = 2 \end{cases} \end{aligned} \quad \text{[S45]}$$

The simultaneous equations, Eq. S42 and Eq. S45 together with Eq. S42, can be solved at steady state, straightforwardly with computer algebra packages such as Mathematica (Wolfram Research).

Similarly, by multiplying Eq. S40 by, for example,  $m_1^2$  and averaging, we can find equations for the second moments:

$$\begin{aligned} \frac{\partial \langle m^2 \rangle_i}{\partial t} = & 2v_0^{(i)} \langle m \rangle_i + v_0^{(i)} \sum_{\mathbf{s}} P^{(i)} + d_0 \langle m \rangle_i - 2d_0 \langle m^2 \rangle_i \\ & + \begin{cases} \kappa_{10} \langle m^2 \rangle_1 - \kappa_{01} \langle m^2 \rangle_0 & \text{if } i = 0 \\ \kappa_{01} \langle m^2 \rangle_0 - (\kappa_{10} + \kappa_{12}) \langle m^2 \rangle_1 + \kappa_{21} \langle m^2 \rangle_2 & \text{if } i = 1 \\ \kappa_{12} \langle m^2 \rangle_1 - \kappa_{21} \langle m^2 \rangle_2 & \text{if } i = 2 \end{cases} \end{aligned} \quad \text{[S46]}$$

for the mean square number of molecules of mRNA;

$$\begin{aligned} \frac{\partial \langle n^2 \rangle_i}{\partial t} = & d_1 \langle n \rangle_i + 2v_1 \langle mn \rangle_i + v_1 \langle m \rangle_i - 2d_1 \langle n^2 \rangle_i \\ & + \begin{cases} \kappa_{10} \langle n^2 \rangle_1 - \kappa_{01} \langle n^2 \rangle_0 & \text{if } i = 0 \\ \kappa_{01} \langle n^2 \rangle_0 - (\kappa_{10} + \kappa_{12}) \langle n^2 \rangle_1 + \kappa_{21} \langle n^2 \rangle_2 & \text{if } i = 1 \\ \kappa_{12} \langle n^2 \rangle_1 - \kappa_{21} \langle n^2 \rangle_2 & \text{if } i = 2 \end{cases} \end{aligned} \quad \text{[S47]}$$

for the mean square number of molecules of protein; and

$$\begin{aligned} \frac{\partial \langle mn \rangle_i}{\partial t} = & v_0^{(i)} \langle n \rangle_i + v_1 \langle m^2 \rangle_i - (d_0 + d_1) \langle mn \rangle_i \\ & + \begin{cases} \kappa_{10} \langle mn \rangle_1 - \kappa_{01} \langle mn \rangle_0 & \text{if } i = 0 \\ \kappa_{01} \langle mn \rangle_0 - (\kappa_{10} + \kappa_{12}) \langle mn \rangle_1 + \kappa_{21} \langle mn \rangle_2 & \text{if } i = 1 \\ \kappa_{12} \langle mn \rangle_1 - \kappa_{21} \langle mn \rangle_2 & \text{if } i = 2 \end{cases} \end{aligned} \quad \text{[S48]}$$

for the mean product of mRNA and protein numbers. We solve Eqs. S46, S47, and S48 at steady state simultaneously using the solutions of Eqs. S42 and S45 and so compute the stationary second moments.

Finally, we need the covariance between the two reporters,  $\langle n_1 n_2 \rangle$ , to determine the extrinsic variance. From Eq. S40, we find three sets of coupled equations:

$$\frac{\partial \langle n_1 n_2 \rangle_i}{\partial t} = 2v_1 \langle m_1 n_2 \rangle_i - 2d_1 \langle n_1 n_2 \rangle_i + \begin{cases} \kappa_{10} \langle n_1 n_2 \rangle_1 - \kappa_{01} \langle n_1 n_2 \rangle_0 & \text{if } i = 0 \\ \kappa_{01} \langle n_1 n_2 \rangle_0 - (\kappa_{10} + \kappa_{12}) \langle n_1 n_2 \rangle_1 + \kappa_{21} \langle n_1 n_2 \rangle_2 & \text{if } i = 1 \\ \kappa_{12} \langle n_1 n_2 \rangle_1 - \kappa_{21} \langle n_1 n_2 \rangle_2 & \text{if } i = 2 \end{cases} \quad \text{[S49]}$$

to determine the covariance between the proteins;

$$\frac{\partial \langle m_1 n_2 \rangle_i}{\partial t} = v_0^{(i)} \langle n \rangle_i + v_1 \langle m_1 m_2 \rangle_i - (d_0 + d_1) \langle m_1 n_2 \rangle_i + \begin{cases} \kappa_{10} \langle m_1 n_2 \rangle_1 - \kappa_{01} \langle m_1 n_2 \rangle_0 & \text{if } i = 0 \\ \kappa_{01} \langle m_1 n_2 \rangle_0 - (\kappa_{10} + \kappa_{12}) \langle m_1 n_2 \rangle_1 + \kappa_{21} \langle m_1 n_2 \rangle_2 & \text{if } i = 1 \\ \kappa_{12} \langle m_1 n_2 \rangle_1 - \kappa_{21} \langle m_1 n_2 \rangle_2 & \text{if } i = 2 \end{cases} \quad [\text{S50}]$$

to determine the covariance of the mRNA of one copy of the system with the protein of another ( $\langle m_1 n_2 \rangle = \langle m_2 n_1 \rangle$  from symmetry); and

$$\frac{\partial \langle m_1 m_2 \rangle_i}{\partial t} = 2v_0^{(i)} \langle m \rangle_i - 2d_0 \langle m_1 m_2 \rangle_i + \begin{cases} \kappa_{10} \langle m_1 m_2 \rangle_1 - \kappa_{01} \langle m_1 m_2 \rangle_0 & \text{if } i = 0 \\ \kappa_{01} \langle m_1 m_2 \rangle_0 - (\kappa_{10} + \kappa_{12}) \langle m_1 m_2 \rangle_1 + \kappa_{21} \langle m_1 m_2 \rangle_2 & \text{if } i = 1 \\ \kappa_{12} \langle m_1 m_2 \rangle_1 - \kappa_{21} \langle m_1 m_2 \rangle_2 & \text{if } i = 2 \end{cases} \quad [\text{S51}]$$

to determine the covariance of the mRNAs from the two copies of the systems. We solve Eqs. S49, S50, and S51 at steady state. To find the final moments, we sum the moments calculated for each state of the extrinsic variable because

$$\langle f(\mathbf{s}) \rangle = \sum_{\mathbf{s}, i} P(\mathbf{s}, v_0^{(i)}) f(\mathbf{s}) = \sum_{\mathbf{s}, i} P^{(i)} f(\mathbf{s}) = \sum_i \langle f(\mathbf{s}) \rangle_i \quad [\text{S52}]$$

for any function  $f(\mathbf{s})$ .

All these equations can be straightforwardly modified to study extrinsic fluctuations in a different kinetic rate. For example, if the translation rate fluctuates then we replace  $v_0^{(i)}$  by  $v_0$  and the translation rate  $v_1$  by the appropriate  $v_1^{(i)}$  in all the equations. To have two or more rates fluctuating (2), we can either extend the number of states of  $P^{(i)}$  if the extrinsic fluctuations are uncorrelated or have more than one parameter changing with state  $i$  if the extrinsic fluctuations are correlated. Our analytical results verify the behavior found previously through simulation for particular values of parameters (2).

**Extrinsic Fluctuations in Transcription Need Not Change the Form of the Intrinsic Noise.** Consider extrinsic fluctuations in  $v_0$ , so that  $v_0$  has three states:  $v_0^{(i)}$ , where  $i$  runs from 0 to 2. From Eq. S44,

$$\langle v_0 \rangle = \frac{\kappa_{10} \kappa_{21} v_0^{(0)} + \kappa_{01} \kappa_{21} v_0^{(1)} + \kappa_{01} \kappa_{12} v_0^{(2)}}{\kappa_{01} \kappa_{12} + \kappa_{01} \kappa_{21} + \kappa_{10} \kappa_{21}}, \quad [\text{S53}]$$

and we find that

$$\langle m \rangle = \frac{\langle v_0 \rangle}{d_0}; \quad \langle n \rangle = \frac{v_1}{d_1} \langle m \rangle. \quad [\text{S54}]$$

To compare with previous work (14–19), we will give our results in terms of the coefficient of variation,  $\eta$  (the standard deviation of a variable divided by its mean). For the intrinsic noise, we have that

$$\eta_{\text{int}}^2 = \frac{1}{\langle n \rangle} + \frac{d_1}{d_0 + d_1} \frac{1}{\langle m \rangle}, \quad [\text{S55}]$$

which has the same form for the system when no extrinsic fluctuations are present (16, 19) (Fig. S3). If we assume that  $v_0^{(0)} = v_0(1 - \epsilon)$ ,  $v_0^{(1)} = v_0$ , and  $v_0^{(2)} = v_0(1 + \epsilon)$  for a constant  $v_0$  and  $\epsilon$  and that  $\kappa_{01} = \kappa_{10} = \kappa_{12} = \kappa_{21} = \kappa$ , then the extrinsic noise equals

$$\eta_{\text{ext}}^2 = \frac{d_0 d_1 (d_0 + d_1 + \kappa)}{(d_0 + d_1)(d_0 + \kappa)(d_1 + \kappa)} \eta_{v_0}^2 \quad [\text{S56}]$$

and is proportional to the square of the noise in  $v_0$ ,  $\eta_{v_0}^2$ , as expected (20). We note that  $\eta_{v_0}^2 = \frac{2\epsilon^2}{3}$  with this choice of  $v_0^{(i)}$ .

**Extrinsic Fluctuations in Translation Can Increase the Intrinsic Noise.** We can proceed similarly with extrinsic fluctuations in the

translation rate. If we let  $v_1^{(0)} = v_1(1 - \epsilon)$ ,  $v_1^{(1)} = v_1$ , and  $v_1^{(2)} = v_1(1 + \epsilon)$  and  $\kappa_{01} = \kappa_{10} = \kappa_{12} = \kappa_{21} = \kappa$ , then

$$\langle m \rangle = \frac{v_0}{d_0}; \quad \langle n \rangle = \frac{v_1}{d_1} \langle m \rangle \quad [\text{S57}]$$

and

$$\eta_{\text{int}}^2 = \frac{1}{\langle n \rangle} + \frac{d_1}{d_0 + d_1} \frac{1}{\langle m \rangle} \left[ 1 + \frac{d_0 + d_1}{d_0 + d_1 + \kappa} \eta_{v_1}^2 \right] \quad [\text{S58}]$$

with

$$\eta_{\text{ext}}^2 = \frac{d_1}{d_1 + \kappa} \eta_{v_1}^2. \quad [\text{S59}]$$

We see that the intrinsic noise is larger than the intrinsic noise of an equivalent system with no extrinsic fluctuations ( $\eta_{v_1} = 0$  and Eq. S55) because of the factor in square brackets in Eq. S58 (Fig. S3). This factor depends on both the magnitude and lifetime of the fluctuations in  $v_1$ , as well as the lifetime of both mRNA and protein molecules.

**Extrinsic Fluctuations in the Degradation of mRNA Can Increase the Intrinsic Noise.** Having extrinsic fluctuations in the degradation rates of either mRNA or protein gives more complex behaviors because such fluctuations directly affect the lifetime of fluctuations in proteins (2). Assuming, as before, that  $d_0^{(0)} = d_0(1 - \epsilon)$ ,  $d_0^{(1)} = d_0$ , and  $d_0^{(2)} = d_0(1 + \epsilon)$ , that  $\kappa_{01} = \kappa_{10} = \kappa_{12} = \kappa_{21} = \kappa$ , and that  $\eta_{d_0}^2 < 1$ , then

$$\langle m \rangle = \frac{v_0}{d_0} \left[ 1 + \frac{d_0}{d_0 + \kappa} \eta_{d_0}^2 + \dots \right]; \quad \langle n \rangle = \frac{v_1}{d_1} \langle m \rangle, \quad [\text{S60}]$$

where we have omitted terms of order  $\eta_{d_0}^4$  and higher. The intrinsic noise is approximately

$$\eta_{\text{int}}^2 \approx \frac{1}{\langle n \rangle} + \frac{d_1}{d_0 + d_1} \frac{1}{\langle m \rangle} \left[ 1 + \frac{d_0^2 (2d_0 + d_1 + \kappa)}{(d_0 + d_1)(d_0 + \kappa)(d_0 + d_1 + \kappa)} \eta_{d_0}^2 \right], \quad [\text{S61}]$$

and the extrinsic noise is approximately

$$\eta_{\text{ext}}^2 \approx \frac{d_0 d_1 (d_0 + d_1 + \kappa)}{(d_0 + d_1)(d_0 + \kappa)(d_1 + \kappa)} \eta_{d_0}^2, \quad [\text{S62}]$$

where higher order corrections in  $\eta_{d_0}$  have been omitted. The intrinsic noise is therefore larger than the intrinsic noise of an equivalent system with no extrinsic fluctuations ( $\eta_{d_0} = 0$ ) providing  $\eta_{d_0}$  is sufficiently small (Fig. S3), and the mean number of proteins has increased above the value predicted by purely deterministic dynamics.



**Extrinsic Fluctuations in the Degradation of Protein Can Decrease the Intrinsic Noise.** Assuming again that  $d_1^{(0)} = d_1(1 - \epsilon)$ ,  $d_1^{(1)} = d_1$ , and  $d_1^{(2)} = d_1(1 + \epsilon)$ , that  $\kappa_{01} = \kappa_{10} = \kappa_{12} = \kappa_{21} = \kappa$ , and that  $\eta_{d_1}^2 < 1$ , then

$$\langle m \rangle = \frac{v_0}{d_0}; \quad \langle n \rangle \simeq \frac{v_1}{d_1} \langle m \rangle \left[ 1 + \frac{d_1}{d_1 + \kappa} \eta_{d_1}^2 \right], \quad [\text{S63}]$$

where we ignore terms of order  $\eta_{d_1}^4$  and higher. The intrinsic noise is

$$\eta_{\text{int}}^2 \simeq \frac{1}{\langle n \rangle} + \frac{d_1}{d_0 + d_1} \frac{1}{\langle m \rangle} \times \left[ 1 - \frac{d_1^2(2d_0(d_0 + d_1) - (d_1 + \kappa)(2d_1 + \kappa))}{(d_0 + d_1)(d_1 + \kappa)(d_0 + d_1 + \kappa)(2d_1 + \kappa)} \right] \eta_{d_1}^2, \quad [\text{S64}]$$

and

$$\eta_{\text{ext}}^2 \simeq \frac{d_1}{d_1 + \kappa} \eta_{d_1}^2, \quad [\text{S65}]$$

where we have omitted higher order terms in  $\eta_{d_1}$ . We expect  $d_0 > d_1$  (19) and  $\kappa \geq d_1$  (21). The intrinsic noise is therefore typically smaller than the intrinsic noise of an equivalent system with no extrinsic fluctuations (Fig. S3), and the mean number of proteins has increased above the value predicted by deterministic dynamics.

**Decomposing the Intrinsic Noise.** In Eq. 3 of the main paper, we decompose the intrinsic noise into transcriptional and translational components. We further argue that a bicistronic reporter correctly measures the translational component and when combined with the original reporter for the system will allow all three components of the variance to be measured. We apply these ideas to calculate the transcriptional and translational components of the intrinsic noise when there are extrinsic fluctuations in the propensity for transcription. The master equation for a bicistronic reporter (Fig. S2C) is

$$\begin{aligned} \frac{\partial P^{(i)}}{\partial t} = & v_0^{(i)} [P_{m-1}^{(i)} - P^{(i)}] + d_0 [(m+1)P_{m+1}^{(i)} - mP^{(i)}] \\ & + d_1 [(n_1+1)P_{n_1+1}^{(i)} - n_1P^{(i)}] + v_1 m [P_{n_1-1} - P^{(i)}] \\ & + d_1 [(n_2+1)P_{n_2+1}^{(i)} - n_2P^{(i)}] + v_1 m [P_{n_2-1} - P^{(i)}] \\ & + \begin{cases} \kappa_{10}P^{(1)} - \kappa_{01}P^{(0)} & \text{if } i = 0 \\ \kappa_{01}P^{(0)} - (\kappa_{10} + \kappa_{12})P^{(1)} + \kappa_{21}P^{(2)} & \text{if } i = 1, \\ \kappa_{12}P^{(1)} - \kappa_{21}P^{(2)} & \text{if } i = 2 \end{cases} \end{aligned} \quad [\text{S66}]$$

- Siso-Nadal F, Ollivier JF, Swain PS (2007) Facile: A command-line network compiler for systems biology. *BMC Syst Biol* 1:36.
- Shahrezaei V, Ollivier JF, Swain PS (2008) Colored extrinsic fluctuations and stochastic gene expression. *Mol Syst Biol* 4:196.
- Gibson M, Bruck J (2000) Efficient exact stochastic simulation of chemical systems with many species and many channels. *J Phys Chem A* 104:1876–1889.
- Gillespie DT (1977) Exact stochastic simulation of coupled chemical reactions. *J Phys Chem* 81:2340–2361.
- Kollmann M, Lovdok L, Bartholomé K, Timmer J, Sourjik V (2005) Design principles of a bacterial signalling network. *Nature* 438:504–507.
- Colman-Lerner A, Gordon A, Serra E, Chin T, Resnekov O, et al. (2005) Regulated cell-to-cell variation in a cell-fate decision system. *Nature* 437:699–706.
- Kraskov A, Stögbauer H, Grassberger P (2004) Estimating mutual information. *Phys Rev E Stat Nonlin Soft Matter Phys* 69:066138.
- Mitra PP, Stark JB (2001) Nonlinear limits to the information capacity of optical fibre communications. *Nature* 411:1027–1030.
- Cover TM, Thomas JA (2006) *Elements of Information Theory* (Wiley-Interscience, Hoboken, NJ).
- Van Kampen NG (1981) *Stochastic Processes in Physics and Chemistry* (North-Holland, Amsterdam).

where  $m$  is the number of molecules of the mRNA and we assume three different states of the extrinsic variable (here  $v_0$ ). We can solve Eq. S66 for its moments following the approach used for Eq. S40. The equations undergo only minor changes with some straightforward replacements (for example,  $\langle m_1^{(i)} m_2^{(i)} \rangle$  becomes  $\langle [m^{(i)}]^2 \rangle$ ). We find that the transcriptional and translational components of the intrinsic noise are

$$\eta_{\text{transc}}^2 = \frac{d_1}{d_0 + d_1} \frac{1}{\langle m \rangle}; \quad \eta_{\text{transl}}^2 = \frac{1}{\langle n \rangle}, \quad [\text{S67}]$$

showing that our theoretical definitions (Eq. 3) give a natural decomposition.

Eq. S67 implies that transcriptional variation is often greater than translational variation. Typical lifetimes of mRNA in *Escherichia coli* are several minutes, but protein numbers are often mostly reduced through dilution. Assuming a cell cycle of 50 min (22) and an average lifetime of an mRNA of 3 min (23), then  $d_1/(d_0 + d_1)$  is approximately 0.06, and so  $\eta_{\text{transc}}^2/\eta_{\text{transl}}^2 \simeq 0.06 \frac{\langle n \rangle}{\langle m \rangle}$ . Consequently, transcriptional variation is greater than translational variation if  $\langle n \rangle$  is approximately greater than 18 times  $\langle m \rangle$ , which is not uncommon: The average number of proteins per mRNA is around 540 (24).

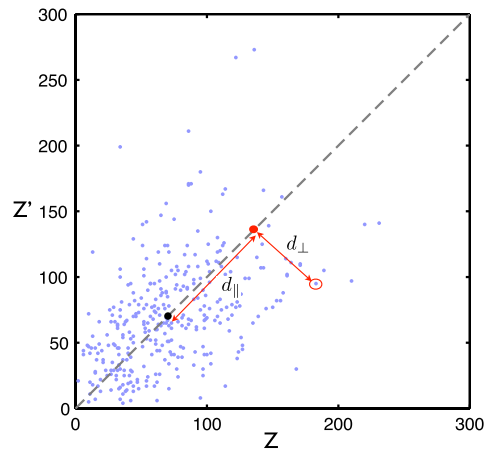
### Verifying Conditional Independences

To use conjugate reporters to determine the components of the variance of a given model, we must check that the appropriate conditional independences are satisfied. Suppose we wish to verify that two reporters  $Z$  and  $Z'$  are conditionally independent given the history  $Y^{\mathcal{H}}$ . Suppose further that the future dynamics of these  $Y$  variables can depend on their own histories, but (given those histories) are independent of the history of all other variables in the model. Then, informally,  $Z$  and  $Z'$  are conditionally independent given  $Y^{\mathcal{H}}$  if we can first simulate the realization of  $Y$  (to time  $t$ ), and then simulate two subsystems independently (or “separately”) given that history of  $Y$  to obtain  $Z_t$  and  $Z'_t$ .

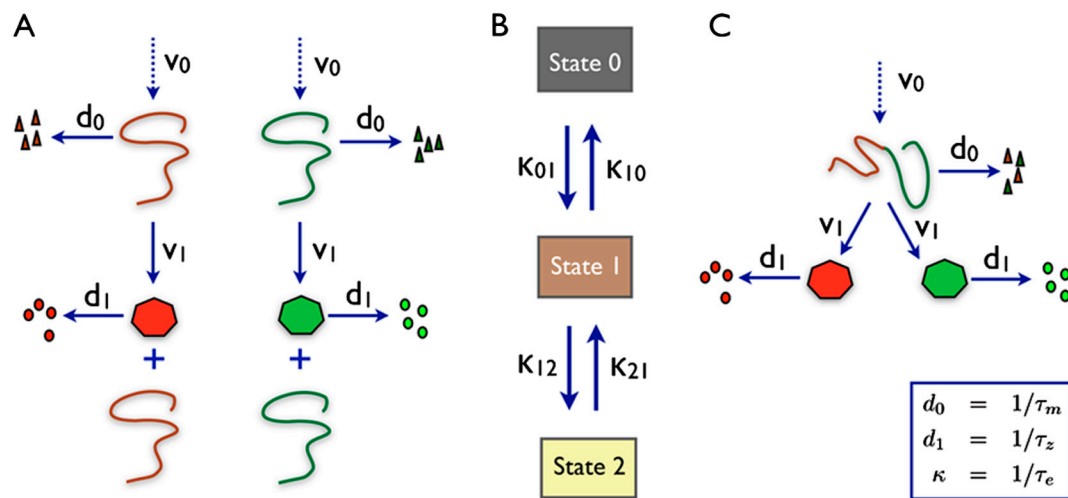
One of us (C.G.B.) has developed the necessary mathematical theory for verification of conditional independence properties in stochastic kinetic models (chemical master equations) in general (25, 26). An algorithm, MIDIA, that applies this theory to test conditional independences has been implemented in R and is freely available (27).

- Grima R (2010) An effective rate equation approach to reaction kinetics in small volumes: theory and application to biochemical reactions in nonequilibrium steady-state conditions. *J Chem Phys* 133:035101.
- Pelet S, et al. (2011) Transient activation of the HOG MAPK pathway regulates bimodal gene expression. *Science* 332:732–735.
- Paulsson J (2004) Summing up the noise in gene networks. *Nature* 427:415–418.
- Peccoud J, Ycart B (1995) Markovian modeling of gene-product synthesis. *Theor Popul Biol* 48:222–234.
- Kepler TB, Elston TC (2001) Stochasticity in transcriptional regulation: Origins, consequences, and mathematical representations. *Biophys J* 81:3116–3136.
- Thattai M, van Oudenaarden A (2001) Intrinsic noise in gene regulatory networks. *Proc Natl Acad Sci USA* 98:8614–8619.
- Friedman N, Cai L, Xie XS (2006) Linking stochastic dynamics to population distribution: An analytical framework of gene expression. *Phys Rev Lett* 97:168302.
- Jahnke T, Huisinga W (2007) Solving the chemical master equation for monomolecular reaction systems analytically. *J Math Biol* 54:1–26.
- Shahrezaei V, Swain PS (2008) Analytical distributions for stochastic gene expression. *Proc Natl Acad Sci USA* 105:17256–17261.
- Swain PS, Elowitz MB, Siggia ED (2002) Intrinsic and extrinsic contributions to stochasticity in gene expression. *Proc Natl Acad Sci USA* 99:12795–12800.
- Rosenfeld N, Young JW, Alon U, Swain PS, Elowitz MB (2005) Gene regulation at the single-cell level. *Science* 307:1962–1965.

22. Phillips R, Kondev J, Theriot J (2009) *Physical Biology of the Cell* (Garland Science, New York).
23. Bernstein JA, Lin PH, Cohen SN, Lin-Chao S (2004) Global analysis of Escherichia coli RNA degradosome function using DNA microarrays. *Proc Natl Acad Sci USA* 101:2758–2763.
24. Lu P, Vogel C, Wang R, Yao X, Marcotte EM (2007) Absolute protein expression profiling estimates the relative contributions of transcriptional and translational regulation. *Nat Biotechnol* 25:117–124.
25. Bowsler CG (2010) Stochastic kinetic models: Dynamic independence, modularity and graphs. *Ann Stat* 38:2242–2281.
26. Bowsler CG (2011) Information processing by biochemical networks: A dynamic approach. *J R Soc Interface* 8:186–200.
27. Bowsler CG (2011) Automated analysis of information processing, kinetic independence and modular architecture in biochemical networks using MIDIA. *Bioinformatics* 27:584–586.



**Fig. S1.** Typical plot of single-cell measurements of a reporter versus measurements of its conjugate reporter. These data are simulated and are given in numbers of proteins per cell. A typical measurement is highlighted by a red circle, and  $d_{\perp}$  and  $d_{\parallel}$  are shown for this measurement. The diagonal  $Z = Z'$  is shown by dashes. The mean  $(E[Z], E[Z'])$  lies on this diagonal and is shown as a black dot. The point of intersection with the diagonal of the line from  $(Z, Z')$  perpendicular to the diagonal is shown as a red dot. This line has a gradient of  $-1$ , and the point of intersection is  $((Z + Z')/2, (Z + Z')/2)$ .



**Fig. S2.** Reactions for models of gene expression. (A) Conjugate reporters given the history of all stochastic processes extrinsic to gene expression. Here,  $v_0$  is the probability of transcription per unit time;  $v_1$  is the probability of translation per unit time per molecule;  $d_0$  is the degradation rate of mRNA per unit time per molecule; and  $d_1$  is the degradation rate of protein per unit time per molecule. (B) The local environment is modeled as a Markov chain. It transitions between three states generating extrinsic fluctuations in a parameter that correspondingly transitions between three values. (C) A bicistronic reporter for measuring the translational component of variation in gene expression. The inset shows the correspondence between the notation here and that in the main text. We simulated this model (with all  $\kappa_{ij}$  identical and equal to  $\kappa$ ) to generate the data for Fig. 2B.



**Table S1. Reaction network and parameter values used in simulations for Fig. 2B**

---

```
variable eta = 0.5
# Z' reporter
DA + S2 → S2 + DA + MA; v02 = 0.01
DA + S1 → S1 + DA + MA ; v01= v02*(1-eta)
DA + S3 → S3 + DA + MA ; v03 = v02*(1+eta)
MA → MA + A ; v1 = 0.2
MA → null; d0 = 0.0167
A → null; d1 = 0.0017
# Z reporter
# transcription
DB + S2 → S2 + DB + MB; v02
DB + S1 → S1 + DB + MB; v01
DB + S3 → S3 + DB + MB; v03
# translation
MB → MB + B; v1
# degradation
MB → null; d0
B → null; d1
# Z" (bicistronic) reporter
MB → MB + C; v1
C → null; d1
# state for fluctuations in v0
S1 → S2; k12 = d1/30
S2 → S1; k21 = k12
S2 → S3; k23 = k12
S3 → S2; k32 = k12
INIT
DA = 1 N;
DB = 1 N;
S2 = 1 N;
```

---