Identifying sources of variation in biochemical networks

Peter Swain
Biological Sciences
University of Edinburgh

SULSA
Scottish Universities Life Sciences Alliance

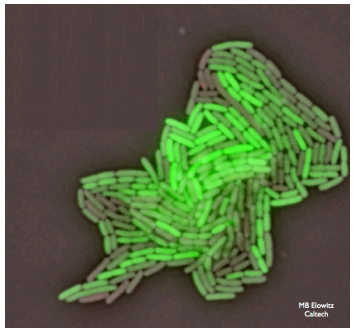
SynthSys
SYNTHETIC & SYSTEMS BIOLOGY

Substantial stochasticity has been measured in the biochemistry of many organisms:

Humans

Variability and memory of protein levels in human cells

Alex Sigal[1]*, Ron Milo[1]*†, Ariel Cohen[1]*, Naama Geva-Zatorsky[1], Yael Klein[1], Yuvalal Liron[1], Nitzan Rosenfeld[1], Tamar Danon[1], Natalie Perzov[1] & Uri Alon[1]



Yeast

Control of Stochasticity in Eukaryotic Gene Expression

Jonathan M. Raser and Erin K. O'Shea*

Slime moulds

Transcriptional Pulsing of a Developmental Gene

Jonathan R. Chubb,[1,2,*] Tatjana Trcek,[1] Shailesh M. Shenoy,[1] and Robert H. Singer[1]

Bacteria

Stochastic Gene Expression in a Single Cell

Michael B. Elowitz,[1,2*] Arnold J. Levine,[1] Eric D. Siggia,[2] Peter S. Swain[2]

Some questions

1. How should I expect fluctuations in an other system, or in several different systems, to affect fluctuations in my system of interest?

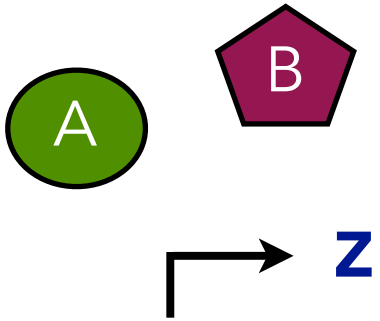
2. How can I measure the effects of such fluctuations?

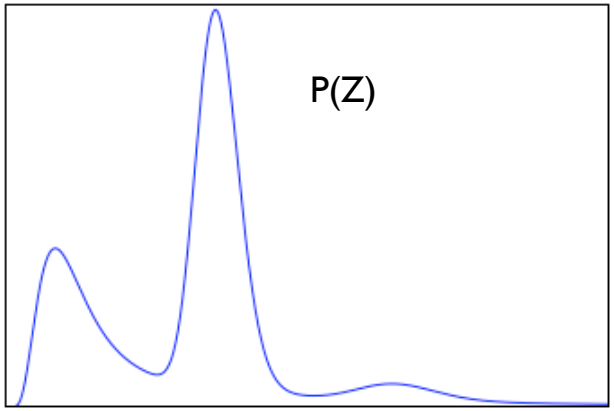
3. How can I distinguish fluctuations generated by information transfer from those generated by `noise'?

4. How do I relate such measurements to models?

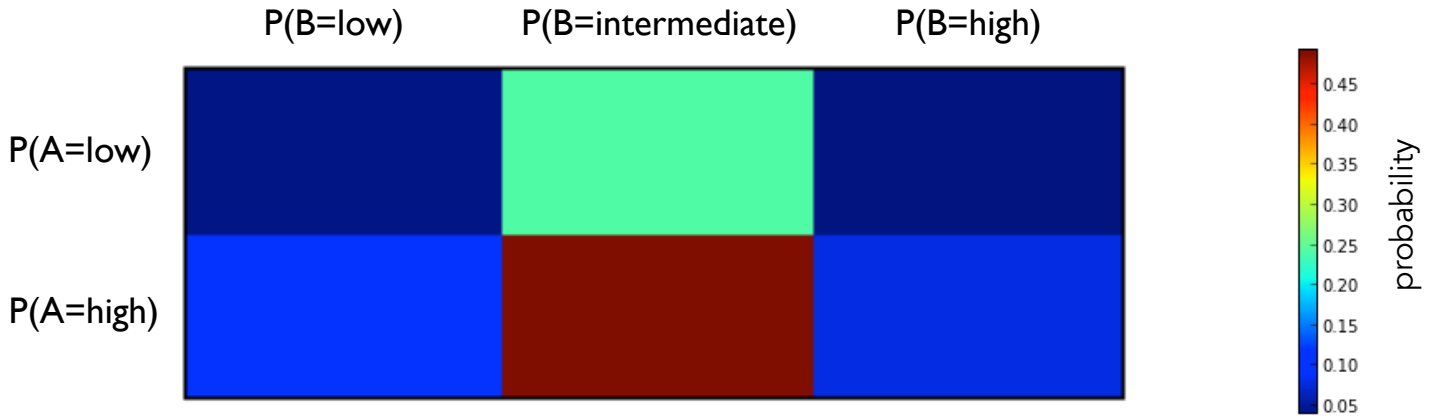
Variation is generated by fluctuations in levels of cellular components.

Consider a gene Z controlled by two transcription factors and a collection of single cell measurements.

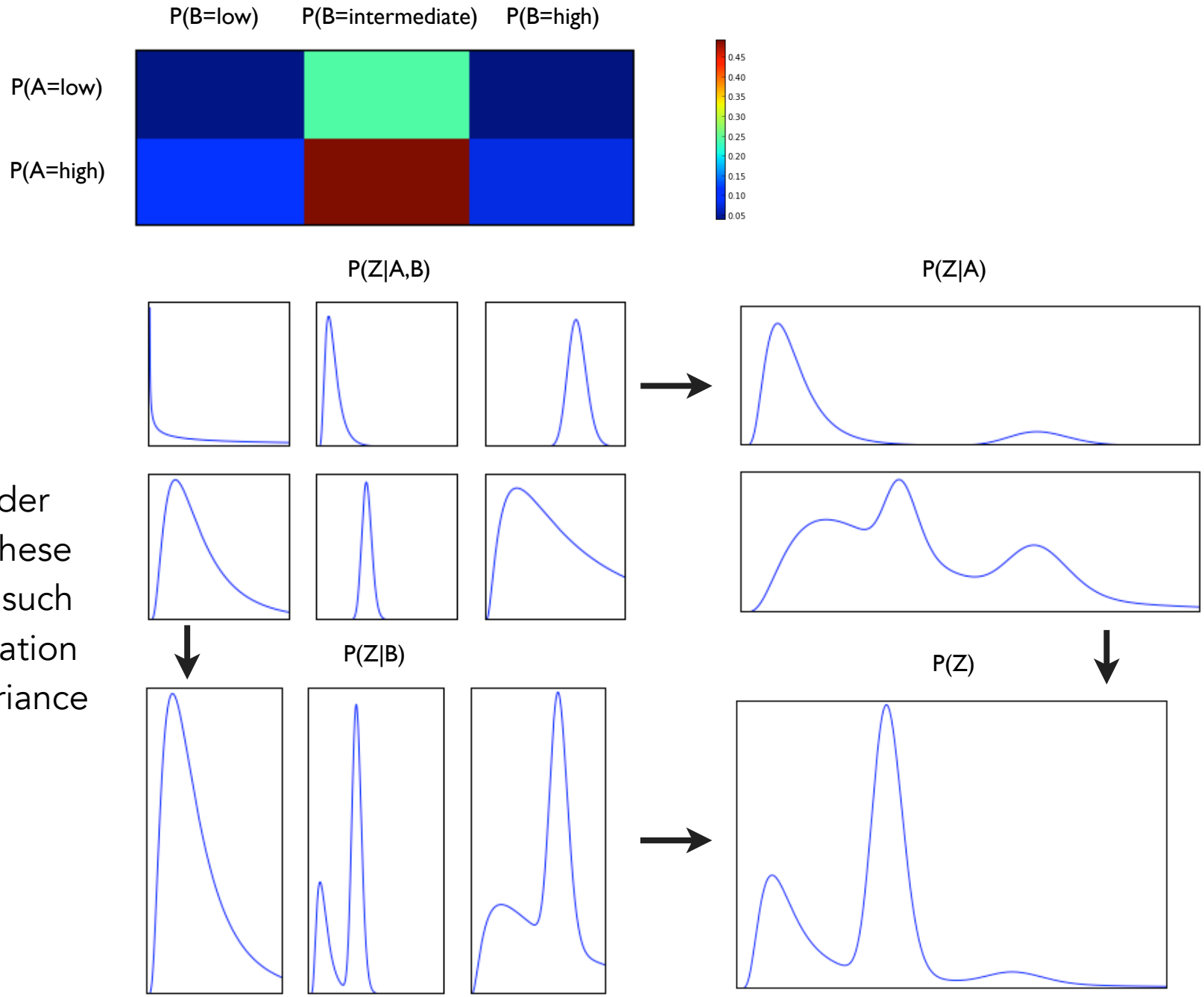




Each cell can have one of two levels of A and one of three levels of B.



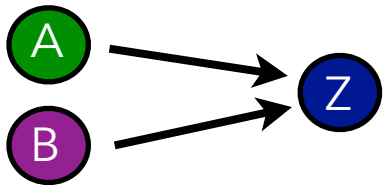
We can consider different components of the distribution of Z.



We can consider moments of these distributions, such as the expectation E[Z|A] and variance V[Z|A,B].

A mathematical aside: conditional expectations

Consider an output Z that is itself stochastic and depends on two stochastic variables A and B



then one conditional expectation of Z is

$$E[Z|B=b] = \int dzda\, z\, P(Z=z, A=a|B=b)$$
$$= e_Z(B)$$

which is a itself a random variable with probabilities given by $P(B=b)$.

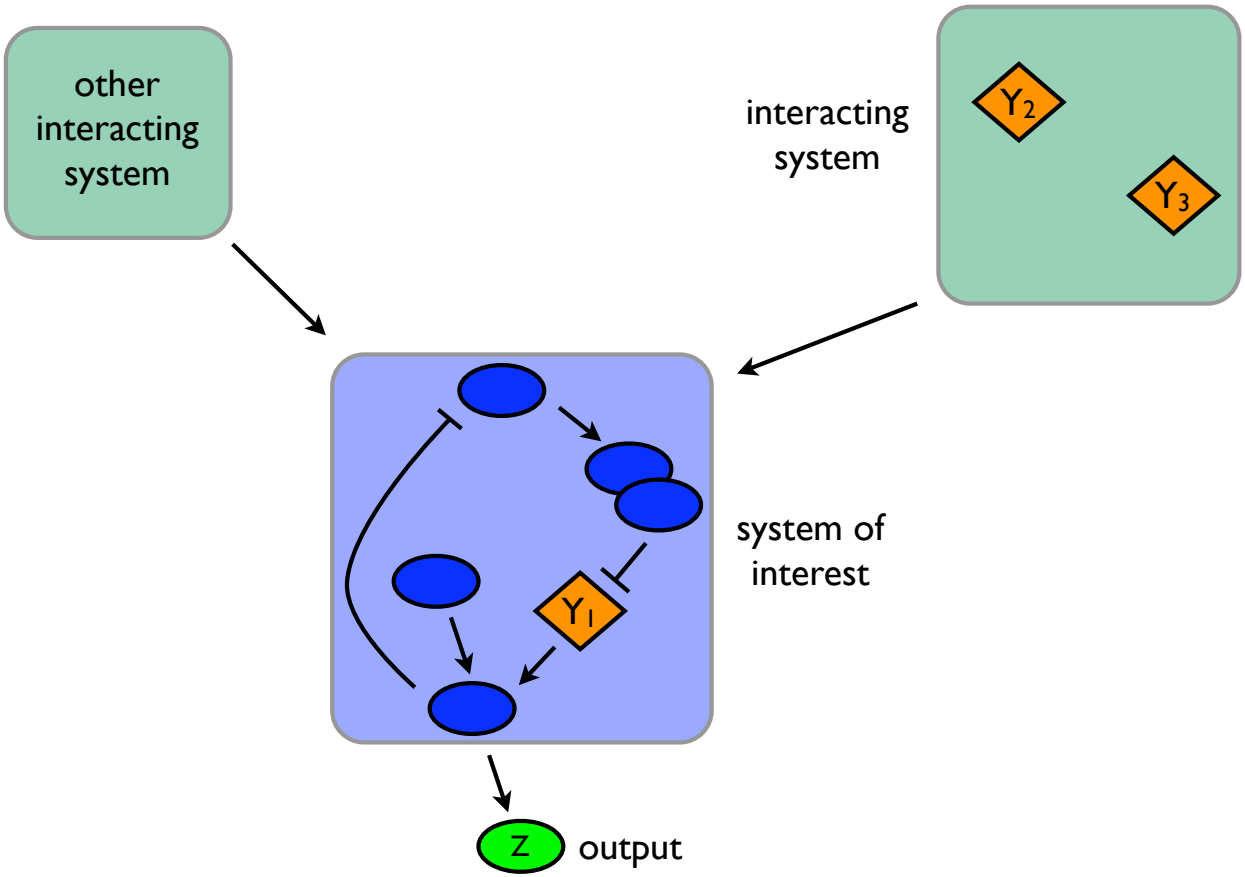
The conditional variance of Z is defined as

$$V[Z|B] = E[Z^2|B] - E[Z|B]^2$$

expectations are always taken over all variables except those for which there is explicit conditioning

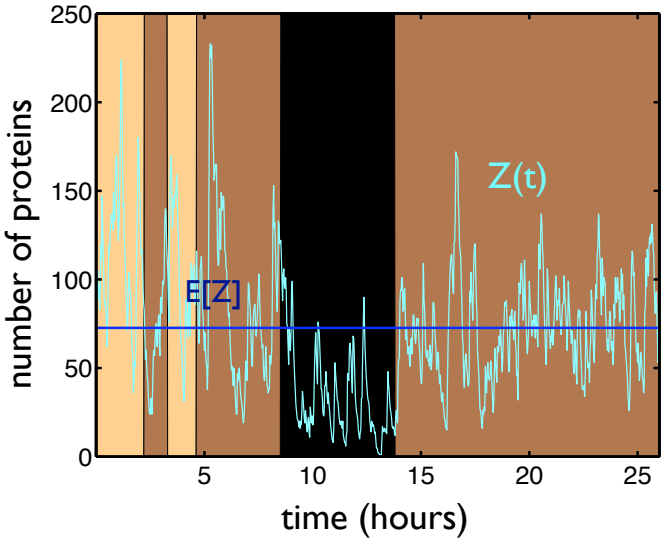
In our decomposition, the stochastic variables of interest can be in the system under study or in other interacting systems.

We can consider the stochasticity generated in the output Z(t) by fluctuations in the stochastic variables: $Y_1(t)$, $Y_2(t)$, and $Y_3(t)$.

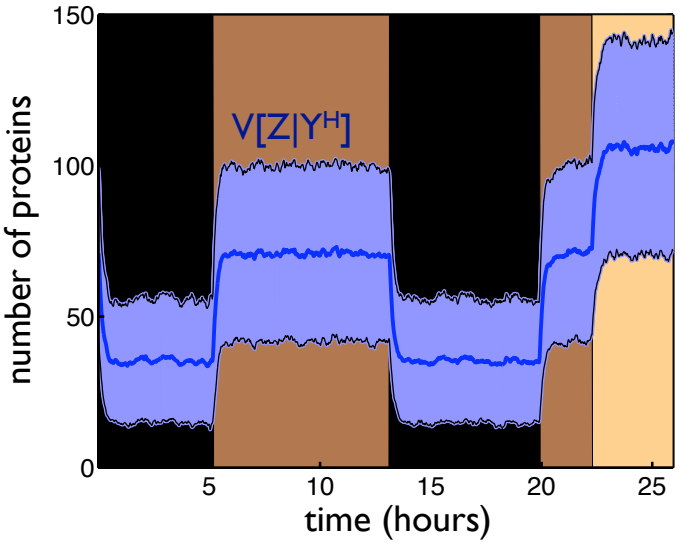


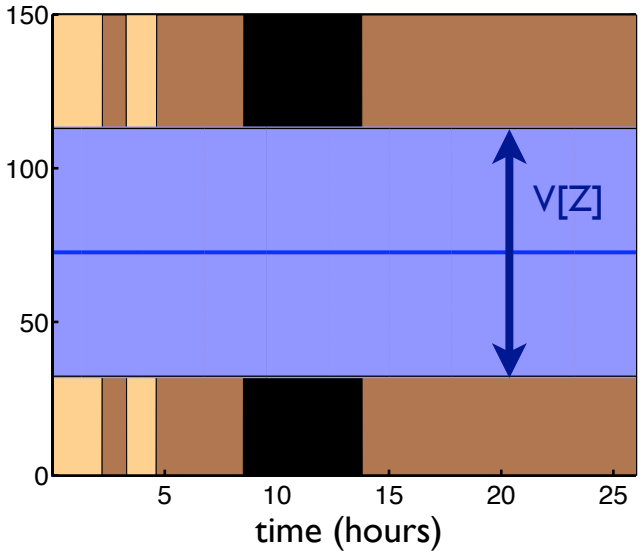
Using conditioning of probabilities, we can mathematically "fix" Y and examine the fluctuations in Z when Y is fixed.

For example, let Z(t) be the number of proteins expressed from a gene and let Y be the number of active transcription factors, which can be either **low**, **medium**, or **high**.

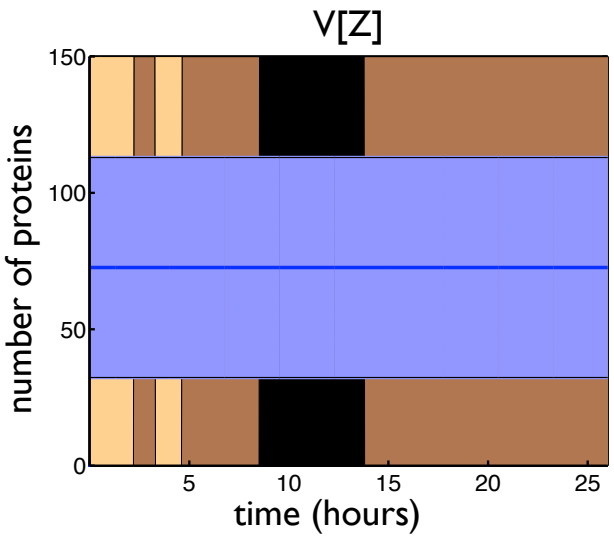


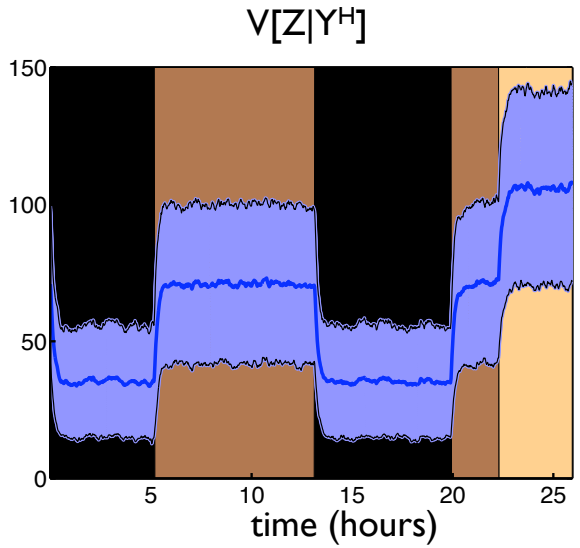
The variance of Z conditional on the history of Y is always smaller than V[Z].





To characterize the variation generated by fluctuations of Y, we compare fluctuations in the output when Y is fixed and when Y fluctuates.





The stochasticity contributed by Y to Z is the mean difference between the variance of Z and the variance of Z conditioned on Y :

$$E\Big\{V[Z] - V[Z|Y^{\mathcal{H}}]\Big\} = V\Big\{E[Z|Y^{\mathcal{H}}]\Big\}$$

where we condition on the entire history of Y, $Y^{\mathcal{H}}$, because all biochemical networks have memory.

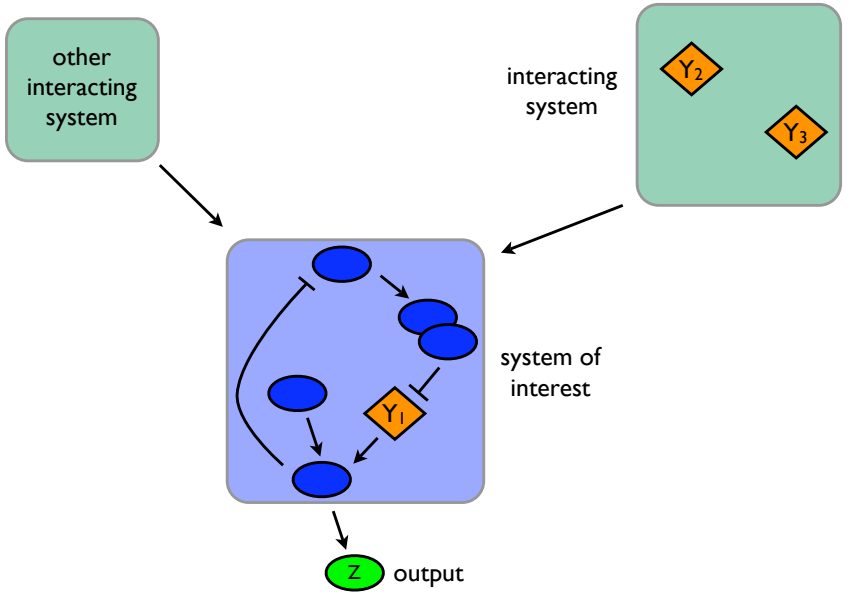
the trajectory of Y: the value of Y at the present time and at all previous times

We can prove that the variance of Z partitions into one term for each of the Y variables and a term for any other stochastic sources.

$$V[Z(t)] = \overbrace{E\Big\{V[Z(t)|(Y_1,Y_2,Y_3)^{\mathcal{H}}]\Big\}}^{\text{from sources other than the } Y\text{s}} + \overbrace{E\Big\{V\Big[E[Z(t)|(Y_1,Y_2,Y_3)^{\mathcal{H}}]\Big|(Y_1,Y_2)^{\mathcal{H}}\Big]\Big\}}^{\text{from } Y_3}$$
$$+ \overbrace{E\Big\{V\Big[E[Z(t)|(Y_1,Y_2)^{\mathcal{H}}]\Big|Y_1^{\mathcal{H}}\Big]\Big\}}^{\text{from } Y_2} + \overbrace{V\Big\{E[Z(t)|Y_1^{\mathcal{H}}]\Big\}}^{\text{from } Y_1}$$

$$E\Big\{V[Z(t)|Y_1^{\mathcal{H}}] - V[Z(t)|(Y_1,Y_2)^{\mathcal{H}}]\Big\}$$

where $\mathcal{H}$ denotes history



Unpacking terms in the decomposition

The term $E[V[E[Z|Y_1,Y_2]|Y_1]]$ is the additional variance in Z generated by Y_2 when Y_1 is given.

$$E\Big\{V\Big[E[Z|Y_1,Y_2]\Big|Y_1\Big]\Big\}$$

first expectation calculated over all stochastic variables except Y_1 and Y_2

$$E\Big\{V\Big[e_Z(Y_1,Y_2)\Big|Y_1\Big]\Big\}$$

variance calculated over Y_2

$$E\Big\{\text{variance of } e_Z \text{ for fluctuating } Y_2 \text{ (given } Y_1)\Big\}$$

last expectation calculated over Y_1

More generally, we have a decomposition that determines the effects of *n* different sources of stochasticity on the output Z:

$$V[Z(t)] = \overbrace{E\{V[Z(t)|\mathbf{Y}_n^{\mathcal{H}}]\}}^{\text{from sources other than } \mathbf{Y}_n} + \sum_{i=2}^{n} \overbrace{E\{V\left[E[Z(t)|\mathbf{Y}_i^{\mathcal{H}}]\middle|\mathbf{Y}_{i-1}^{\mathcal{H}}\right]\}}^{\text{from } Y_i \text{ given } \mathbf{Y}_{i-1}} + \overbrace{V\{E[Z(t)|\mathbf{Y}_1^{\mathcal{H}}]\}}^{\text{from } Y_1}$$

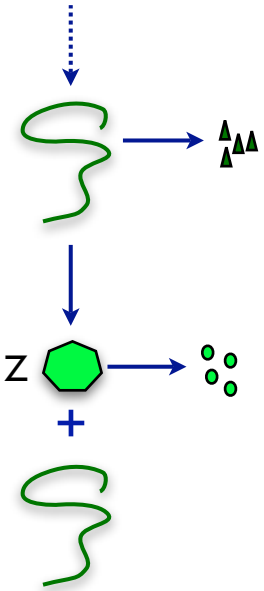
$$E\left[\, V[Z(t)|\mathbf{Y}_{i-1}^{\mathcal{H}}] - V[Z(t)|\mathbf{Y}_i^{\mathcal{H}}] \,\right]$$

with $\mathbf{Y}_i^{\mathcal{H}}$ being the history of the collection of stochastic variables Y_1, ..., Y_i

The decomposition is only unique given a choice of the conditioning.

Example: Decomposing variation of protein expression into two components gives intrinsic and extrinsic variation.

$$V[Z(t)] = \overbrace{E\left\{V[Z(t)|Y_e^{\mathcal{H}}]\right\}}^{\text{intrinsic variation}} + \overbrace{V\left\{E[Z(t)|Y_e^{\mathcal{H}}]\right\}}^{\text{extrinsic variation}}$$



We condition on Y_e^H , all processes extrinsic to gene expression. Examples include the number of free RNA polymerases, the number of free ribosomes, the number of free exosomes, and the number of free proteasomes.

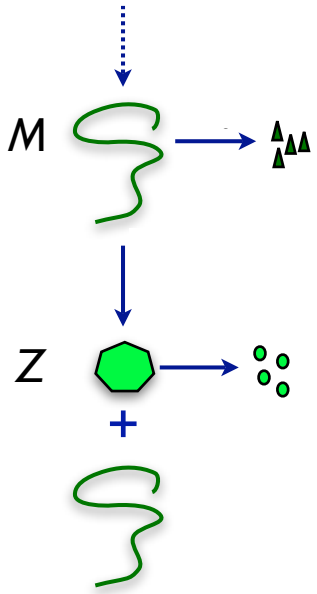
Extrinsic variation is therefore the extra variation created by the interaction of our system of interest with other stochastic systems in the cell and its environment.

Swain *et al.*, 2002
Elowitz *et al.*, 2002
Hilfinger & Paulsson, 2011
Bowsher & Swain, 2012

Example: Decomposing intrinsic variation into transcriptional and translational components to see which is more "noisy".

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

We condition on the history of the levels of mRNA, *M*, and on all stochastic variables extrinsic to gene expression.

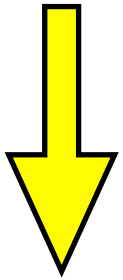


Unpacking translational variation

Translational variation is the additional variation in Z generated on average once the history of levels of mRNA and of extrinsic variables is given:

$$E\Big\{V[Z|(M,Y_e)^{\mathcal{H}}]\Big\}$$

first expectations calculated over all stochastic variables except M^H and Y_e^H



$$E\Big\{\text{variance of Z given the history of } M \text{ and } Y_e\Big\}$$

last expectation calculated over M^H and Y_e^H

Unpacking transcriptional variation

Transcriptional variation is the additional variation in Z generated by fluctuating levels of mRNA once the history of fluctuations in the extrinsic variables is given:

$$E\left\{V\left[E[Z|(M,Y_e)^{\mathcal{H}}]\middle|Y_e^{\mathcal{H}}\right]\right\}$$

first expectation calculated over all stochastic variables except M^H and Y_e^H

$$E\left\{V\left[e_Z(M^{\mathcal{H}},Y_e^{\mathcal{H}})\middle|Y_e^{\mathcal{H}}\right]\right\}$$

variance calculated over M^H

$$E\left\{\text{variance of } e_Z \text{ for fluctuating } M^{\mathcal{H}} \text{ (given } Y_e^{\mathcal{H}})\right\}$$

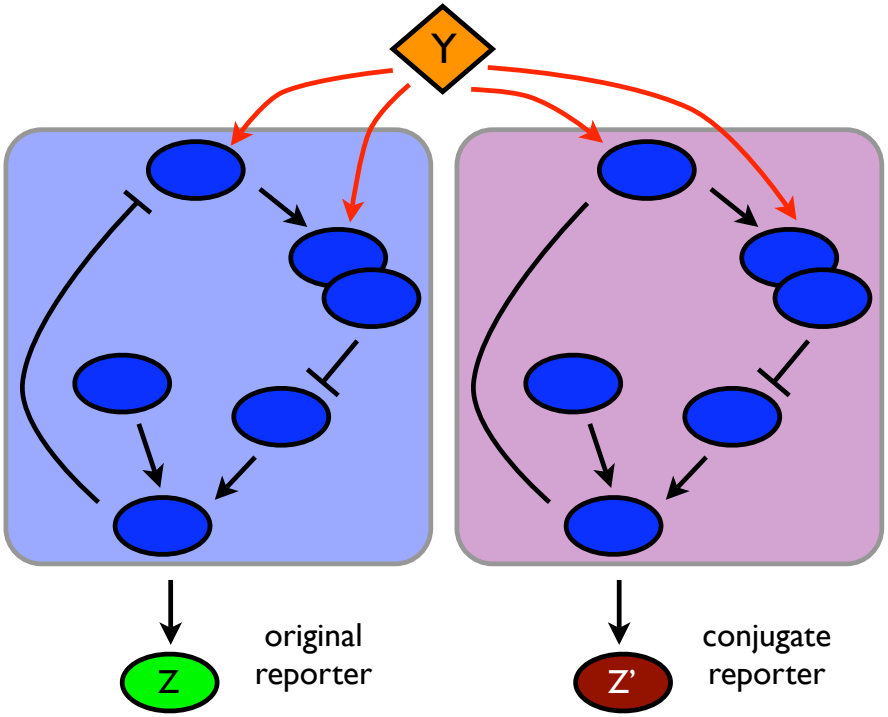
last expectation calculated over Y_e^H

We can experimentally estimate all terms in the decomposition by measuring the covariance between a reporter for Z and a reporter *conjugate* to Z for each of the Y variables.

A reporter conjugate to Z given the history of Y must:

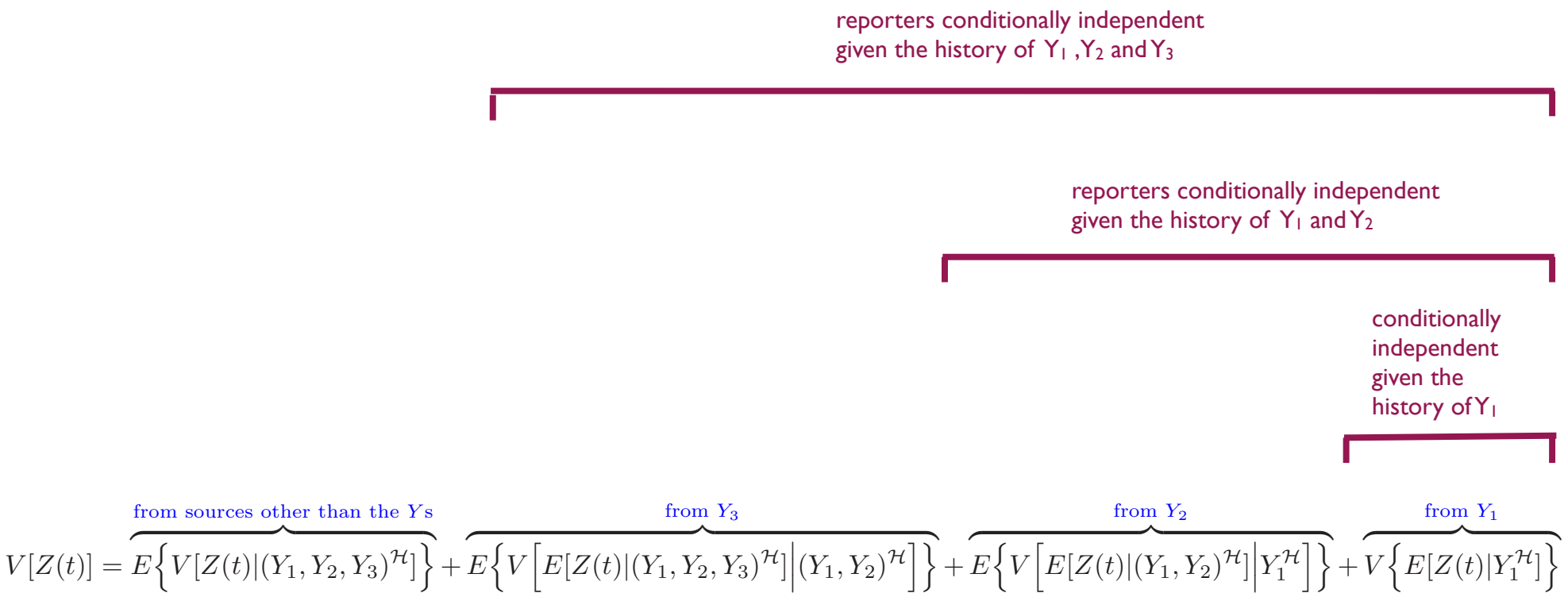
(i) be conditionally independent of Z given the history of Y

(ii) have the same conditional mean as Z given the history of Y



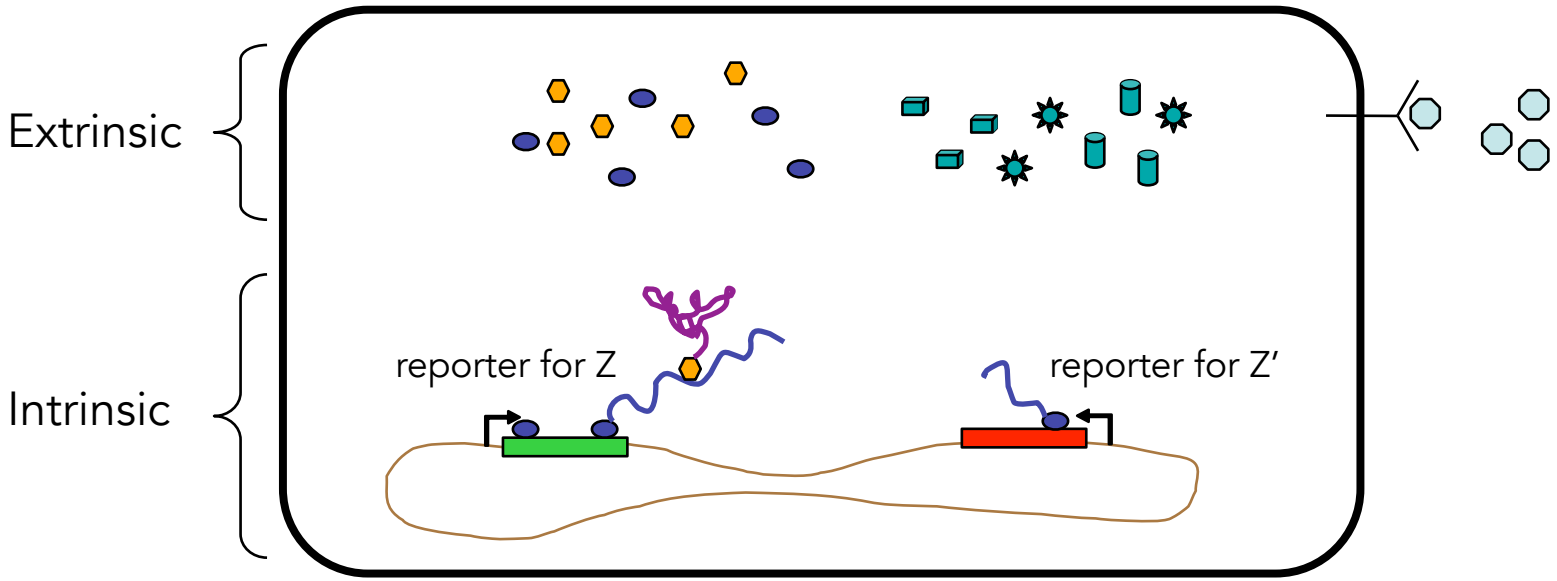
We thus design the conjugate reporter so that it is only fluctuations in Y that cause correlations between Z and the conjugate reporter Z'.

For four terms in the decomposition, we need four reporters.



These reporters can either all be in the same cell or in three different cells, each containing the original reporter for *Z* and one conjugate reporter.

Example: The reporter Z′ must be conjugate given the history of all extrinsic variables to determine intrinsic and extrinsic variation.



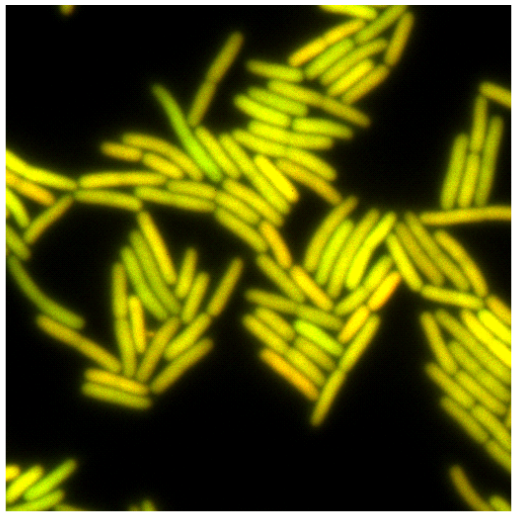
The variance of Z is

$$V[Z(t)] = \overbrace{E\left\{V[Z(t)|Y_e^{\mathcal{H}}]\right\}}^{\text{intrinsic variation}} + \overbrace{V\left\{E[Z(t)|Y_e^{\mathcal{H}}]\right\}}^{\text{extrinsic variation}}$$

Extrinsic variables affect each reporter equally, and only fluctuations in extrinsic variables can cause covariance between the reporters Z and Z′.

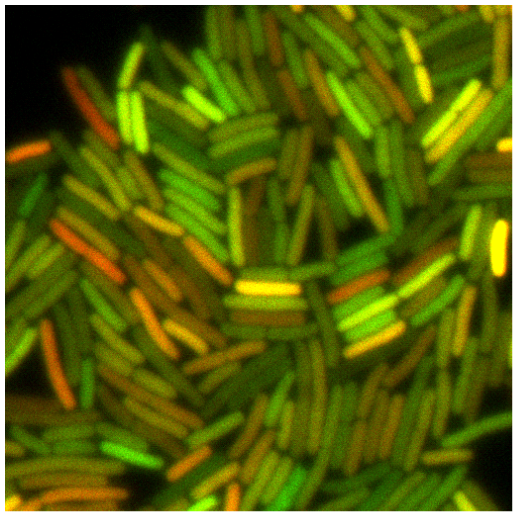
Extrinsic and intrinsic variation have been measured in bacteria and yeast. Extrinsic variation is often greater than intrinsic variation.

WT $_{(RPR37)}$ + IPTG



decrease expression 32-fold

WT $_{(RPR37)}$



Intensity	= 1
η_{int}	= 0.063 (0.058-0.069)
η_{ext}	= 0.098 (0.09-1.1)

Intensity	= 0.03
η_{int}	= 0.25 (0.22-0.27)
η_{ext}	= 0.32 (0.3-0.35)

with η being the coefficient of variation (standard deviation divided by the mean).

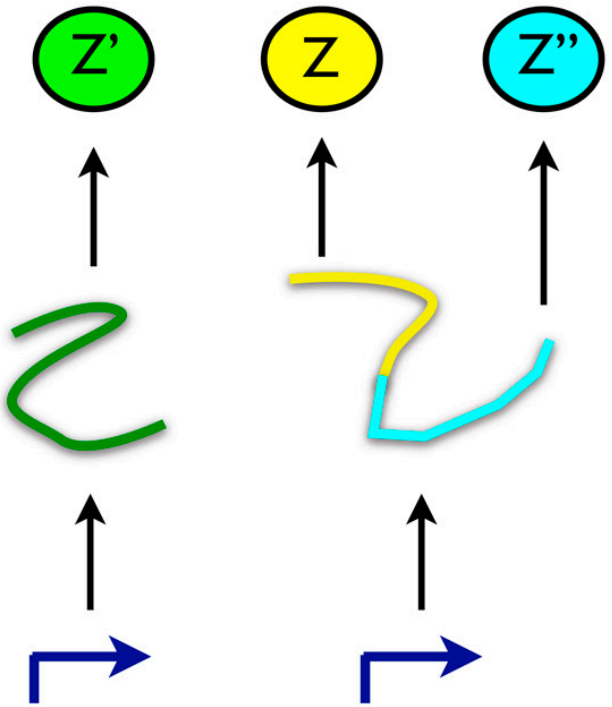
Elowitz *et al.*, 2002

Example: Two conjugate reporters are needed to determine transcriptional and translational variation.

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

We need two conjugate reporters: one conjugate to Z given the history of the extrinsic variables (Z'); the other conjugate to Z given the joint history of the extrinsic variables and mRNA levels (Z'')

Proteins expressed from a bicistronic reporter with two ribosome binding sites (Z and Z'') are conditionally independent given the joint history of the extrinsic variables and mRNA levels.

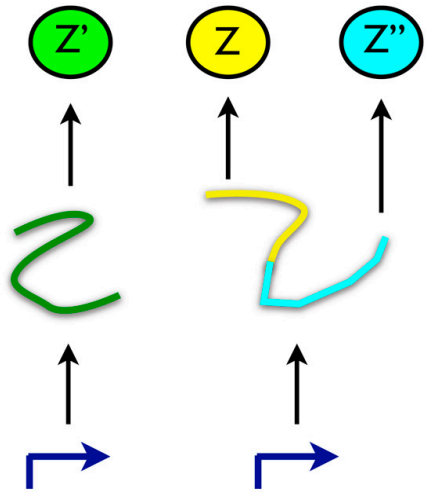


For models of biochemical networks, we can use conjugate reporters in simulations and calculations to make predictions of the magnitude of different variance components.

For example, consider the decomposition of variance generated during gene expression into transcriptional and translational components:

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

Three reporters are needed experimentally, and we can either simulate the reporters or include the reporters in a description of the system by a master equation.



For example, suppose fluctuations in the rate of transcription generate extrinsic fluctuations, then we can augment the master equation with a conjugate reporter to measure extrinsic variation:

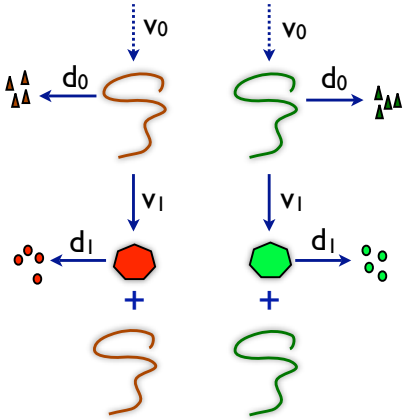
Let the transcription rate fluctuate between 3 states:

The augmented master equation

$$\frac{\partial P^{(i)}}{\partial t} = v_0^{(i)}\left[P^{(i)}_{m_1-1} - P^{(i)}\right] + d_0\left[(m_1+1)P^{(i)}_{m_1+1} - m_1 P^{(i)}\right] + d_1\left[(n_1+1)P^{(i)}_{n_1+1} - n_1 P^{(i)}\right]$$
$$+v_1 m_1\left[P^{(i)}_{n_1-1} - P^{(i)}\right] \qquad \text{first reporter}$$

$$+v_0^{(i)}\left[P^{(i)}_{m_2-1} - P^{(i)}\right] + d_0\left[(m_2+1)P^{(i)}_{m_2+1} - m_2 P^{(i)}\right] + d_1\left[(n_2+1)P^{(i)}_{n_2+1} - n_2 P^{(i)}\right]$$
$$+v_1 m_2\left[P^{(i)}_{n_2-1} - P^{(i)}\right] \qquad \text{conjugate reporter}$$

$$+\begin{cases} \kappa_{10}P^{(1)} - \kappa_{01}P^{(0)} & \text{if } i=0 \\ \kappa_{01}P^{(0)} - (\kappa_{10}+\kappa_{12})P^{(1)} + \kappa_{21}P^{(2)} & \text{if } i=1 \\ \kappa_{12}P^{(1)} - \kappa_{21}P^{(2)} & \text{if } i=2 \end{cases} \qquad \text{extrinsic fluctuations in transcription}$$



where i denotes the state of the transcription rate.

Our calculations imply that transcriptional variation is usually greater than translational variation in *E. coli*.

$$\overbrace{E\Big\{V[Z(t)|(M,Y_e)^{\mathcal{H}}]\Big\}}^{\text{translational}} = E[Z] \qquad \overbrace{V\Big\{E[Z(t)|Y_e^{\mathcal{H}}]\Big\}}^{\text{extrinsic}} = \frac{\tau_e(\tau_z\tau_e+\tau_m\tau_e+\tau_m\tau_z)}{(\tau_m+\tau_z)(\tau_e+\tau_m)(\tau_e+\tau_z)}E[Z]^2\eta_e^2$$

$$\overbrace{E\Big\{V\Big[E[Z(t)|(M,Y_e)^{\mathcal{H}}]|Y_e^{\mathcal{H}}\Big]\Big\}}^{\text{transcriptional}} = \frac{\tau_m}{\tau_m+\tau_z}\frac{E[Z]^2}{E[M]}$$

Assuming an mRNA lifetime of 3 minutes and a cell-cycle time of 50 minutes, then

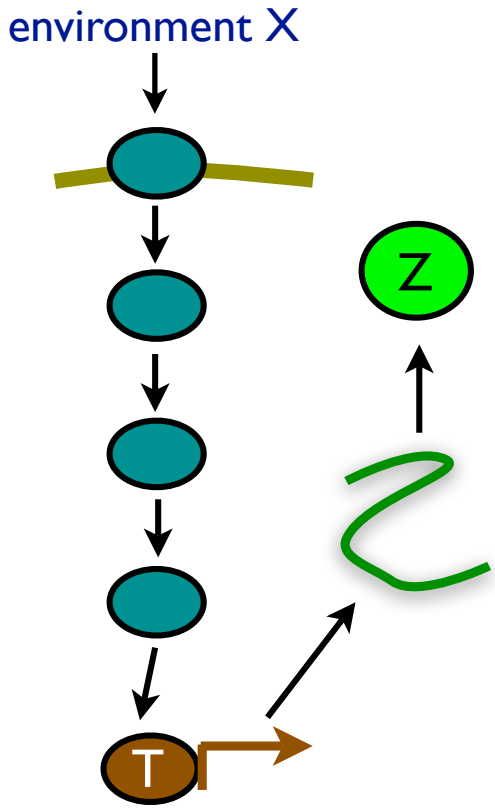
$$E[Z] > 18E[M]$$

if transcriptional variation is to be bigger than translational variation in *E. coli*.

The average number of proteins per mRNA is approximately 540: transcriptional variation dominates translational variation.

Example: For a signalling network, we can identify variation from gene expression, from signal transduction, and informational variation.

The environmental input, X, determines the level of activation of the signalling pathway and so nuclear localization of a transcription factor, T, that activates expression of the output Z.



A four-way decomposition of the variance in Z gives:

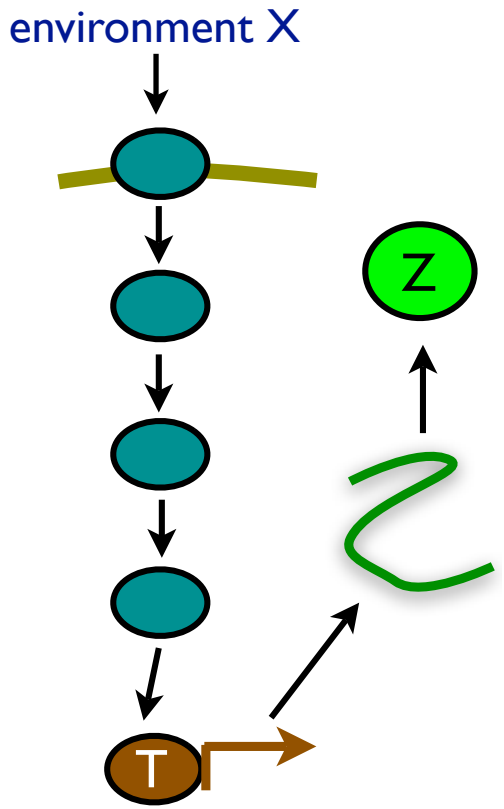
$$V[Z] = \overbrace{E\Big\{V[Z|(Y_{e\setminus T},T)^{\mathcal{H}},X]\Big\}}^{\text{from gene expression}} + \overbrace{E\Big\{V\Big[E[Z|(Y_{e\setminus T},T)^{\mathcal{H}},X]\Big|Y^{\mathcal{H}}_{e\setminus T},X\Big]\Big\}}^{\text{from signal transduction}}$$
$$+ \overbrace{E\Big\{V\Big[E[Z|Y^{\mathcal{H}}_{e\setminus T},X]\Big|X\Big]\Big\}}^{\text{from other extrinsic effects}} + \overbrace{V\Big\{E[Z|X]\Big\}}^{\text{from input signals}}$$

We assume that the environment can be described by the probability of its different states and that the system responds sufficiently quickly that it reaches steady-state before the environment changes again.

For a given environment, X, only a three-way decomposition is necessary.

$$V[Z|X] = \overbrace{E\Big\{V[Z|(Y_{e\setminus T}, T)^{\mathcal{H}}, X]|X\Big\}}^{\text{from gene expression}} + \overbrace{E\Big\{V\Big[E[Z|(Y_{e\setminus T}, T)^{\mathcal{H}}, X]\Big|Y^{\mathcal{H}}_{e\setminus T}, X\Big]\Big|X\Big\}}^{\text{from signal transduction}} + \overbrace{V\Big\{E[Z|Y^{\mathcal{H}}_{e\setminus T}, X]|X\Big\}}^{\text{from other extrinsic effects}}$$

This decomposition describes laboratory experiments that are performed in just one of the possible states of the environment.



Unpacking transductional variation

Transcriptional variation is the additional variation in Z generated by fluctuating levels of mRNA once the history of fluctuations in the extrinsic variables is given:

$$E\Big\{V\Big[E[Z|(Y_{e\setminus T},T)^{\mathcal{H}},X]\Big|Y_{e\setminus T}^{\mathcal{H}},X\Big]\Big|X\Big\}$$

first expectation calculated over all stochastic variables except T^H and $Y_{e\setminus T}{}^H$ and the input X=x

$$E\Big\{V\Big[e_Z(Y_{e\setminus T}^{\mathcal{H}},T^{\mathcal{H}},X)\Big|Y_{e\setminus T}^{\mathcal{H}},X\Big]\Big|X\Big\}$$

variance calculated over T^H

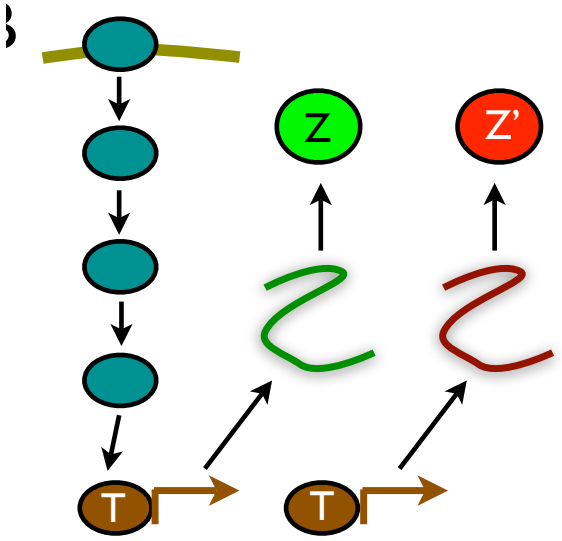
$$E\Big\{\text{variance of } e_Z \text{ for fluctuating } T^{\mathcal{H}} \text{ (given } Y_{e\setminus T}^{\mathcal{H}} \text{ and } X)\Big|X\Big\}$$

last expectation calculated over $Y_{e\setminus T}{}^H$ given X=x

To determine variation arising from gene expression, we need a reporter conjugate to Z given the history of levels of the transcription factor and all other variables extrinsic to gene expression.

$$V[Z|X] = \overbrace{E\Big\{V[Z|(Y_{e\setminus T},T)^{\mathcal{H}},X]|X\Big\}}^{\text{from gene expression}} + \overbrace{E\Big\{V\Big[E[Z|(Y_{e\setminus T},T)^{\mathcal{H}},X]\Big|Y^{\mathcal{H}}_{e\setminus T},X\Big]\Big|X\Big\}}^{\text{from signal transduction}} + \overbrace{V\Big\{E[Z|Y^{\mathcal{H}}_{e\setminus T},X]|X\Big\}}^{\text{from other extrinsic effects}}$$

The reporter, Z', can be constructed by copying the gene for Z, and measuring expression for a given X, then



$$\frac{1}{2}E\Big[(Z-Z')^2\Big|X\Big] = \overbrace{E\Big\{V[Z|(Y_{e\setminus T},T)^{\mathcal{H}},X]\Big|X\Big\}}^{\text{from gene expression given } X}$$

and

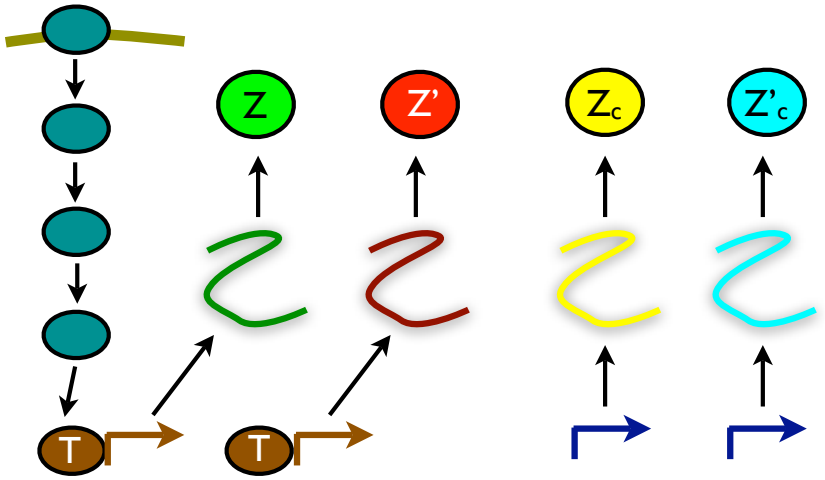
$$\mathrm{Cov}[Z,Z'|X] = \overbrace{E\Big\{V\Big[E[Z|(Y_{e\setminus T},T)^{\mathcal{H}},X]\Big|Y_{e\setminus T},X\Big]\Big|X\Big\}}^{\text{from signal transduction given } X} + \overbrace{V\Big[E[Z|Y^{\mathcal{H}}_{e\setminus T},X]\Big|X\Big]}^{\text{from other extrinsic effects given } X}$$

We cannot directly measure variation from extrinsic fluctuations, but can find a lower bound.

$$\overbrace{V\left[E[Z|Y^{\mathcal{H}}_{e\setminus T}, X]\middle|X\right]}^{\text{from other extrinsic effects}}$$

We would need a reporter conjugate given only the history of variables extrinsic to gene expression (other than T).

Instead, consider a constitutively expressed reporter, Z_c.



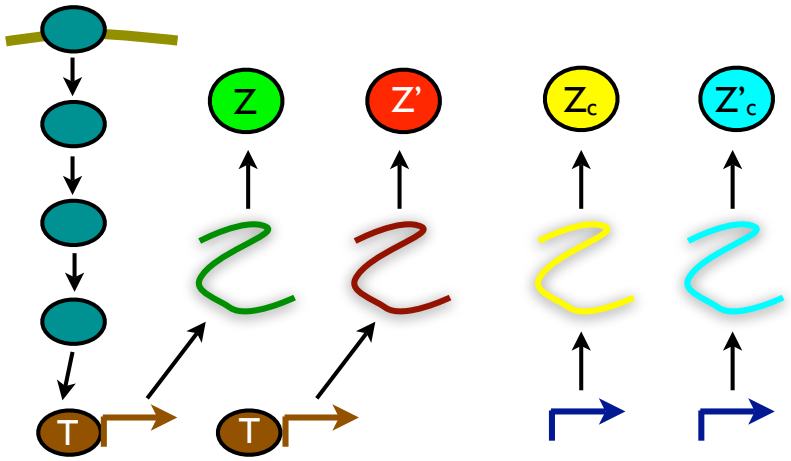
We can then give a lower bound

$$\overbrace{V\left[E[Z|Y^{\mathcal{H}}_{e\setminus T}, X]\middle|X\right]}^{\text{from other extrinsic effects given } X} \geq \frac{\mathrm{Cov}[\mathrm{Z}, \mathrm{Z_c}|\mathrm{X}]}{\mathrm{Cov}[\mathrm{Z_c}, \mathrm{Z'_c}|\mathrm{X}]} \cdot \mathrm{Cov}[Z, Z_c|X]$$

Colman-Lerner *et al.*, 2005
Pedraza & Van Oudenaarden, 2005

We then have an upper bound on the component generated by signal transduction.

$$V[Z|X] = \overbrace{E\Big\{V[Z|(Y_{e\setminus T}, T)^{\mathcal{H}}, X]|X\Big\}}^{\text{from gene expression}} + \overbrace{E\Big\{V\Big[E[Z|(Y_{e\setminus T}, T)^{\mathcal{H}}, X]\Big|Y^{\mathcal{H}}_{e\setminus T}, X\Big]\Big|X\Big\}}^{\text{from signal transduction}} + \overbrace{V\Big\{E[Z|Y^{\mathcal{H}}_{e\setminus T}, X]|X\Big\}}^{\text{from other extrinsic effects}}$$



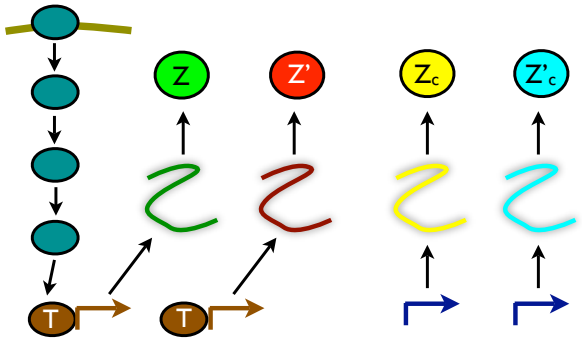
We find that

$$\overbrace{E\Big\{V\Big[E[Z|(Y_{e\setminus T}, T)^{\mathcal{H}}, X]\Big|Y^{\mathcal{H}}_{e\setminus T}, X\Big]\Big|X\Big\}}^{\text{from signal transduction given } X} \leq \mathrm{Cov}[Z, Z'|X] - \frac{\mathrm{Cov}[\mathrm{Z}, \mathrm{Z_c}|\mathrm{X}]^2}{\mathrm{Cov}[\mathrm{Z_c}, \mathrm{Z'_c}|\mathrm{X}]}$$

Example: pheromone response in budding yeast

Regulated cell-to-cell variation in a cell-fate decision system

Alejandro Colman-Lerner[1]*, Andrew Gordon[1]*, Eduard Serra[1], Tina Chin[1], Orna Resnekov[1], Drew Endy[2], C. Gustavo Pesce[1] & Roger Brent[1]

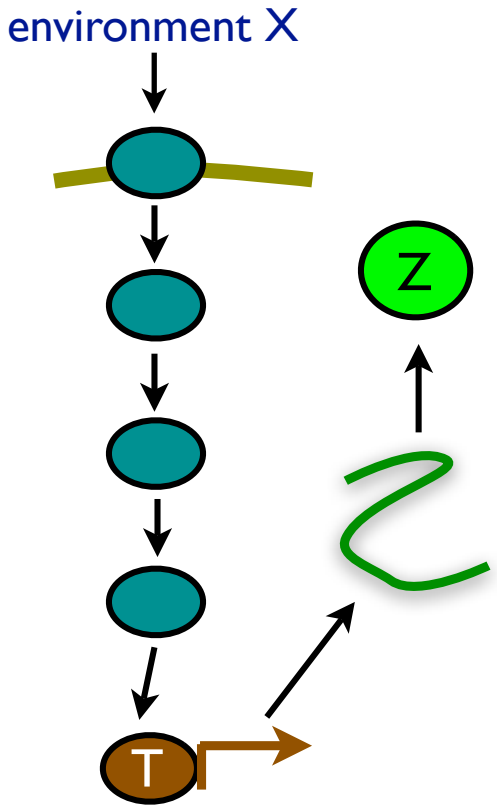


Z driven by pheromone-responsive promoter PRM1
Z' driven by pheromone-responsive promoter PRM1
Z_c driven by the (constitutive) promoter for actin
Z_c' driven by the (constitutive) promoter for actin

Re-analyzing their data, we find that gene expression generates around 10% of variation in Z, that processes extrinsic to gene expression generate at least 50%, and that signal transduction generates less than 40% of the variation for cells exposed to 1.25 nM pheromone.

We can identify the part of the variation of Z that informs on the environment: the informational variation.

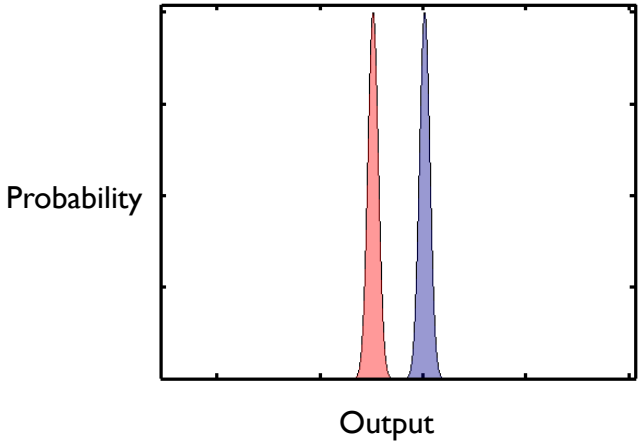
$$V[Z] = \overbrace{E\Big\{V[Z|(Y_{e\setminus T},T)^{\mathcal{H}},X]\Big\}}^{\text{from gene expression}} + \overbrace{E\Big\{V\Big[E[Z|(Y_{e\setminus T},T)^{\mathcal{H}},X]\Big|Y^{\mathcal{H}}_{e\setminus T},X\Big]\Big\}}^{\text{from signal transduction}}$$
$$+ \overbrace{E\Big\{V\Big[E[Z|Y^{\mathcal{H}}_{e\setminus T},X]\Big|X\Big]\Big\}}^{\text{from other extrinsic effects}} + \overbrace{V\Big\{E[Z|X]\Big\}}^{\text{from input signals}}$$



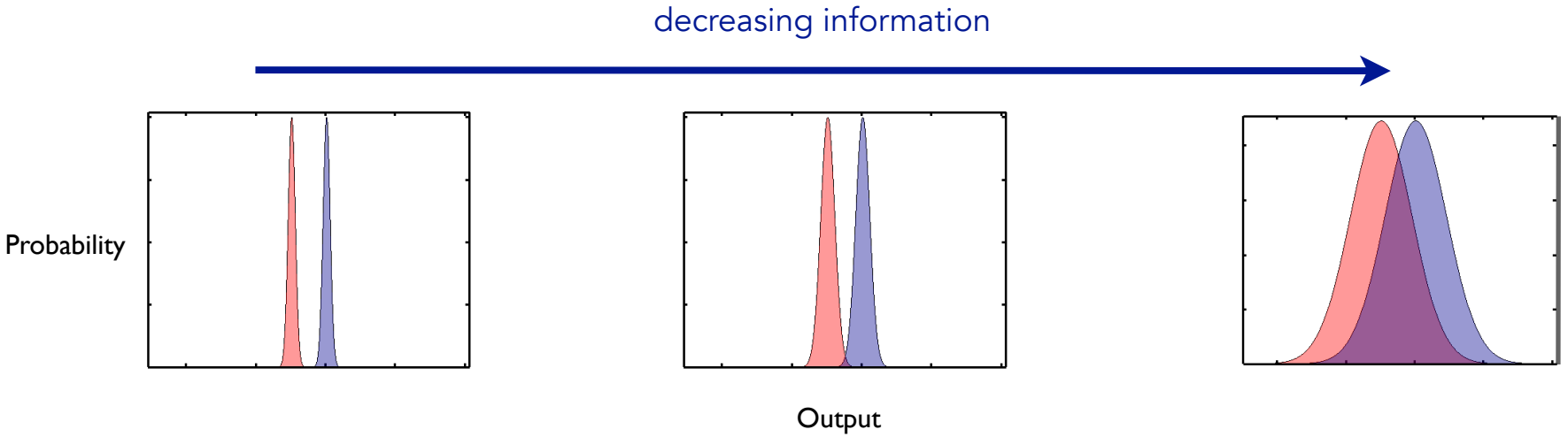
Mathematically, information is a measure of the ambiguity of a signal.

With higher information between the input and the output, it easier to distinguish if the output comes from the red or blue state of the input.

An example with two states of the input and a continuous value of the output.

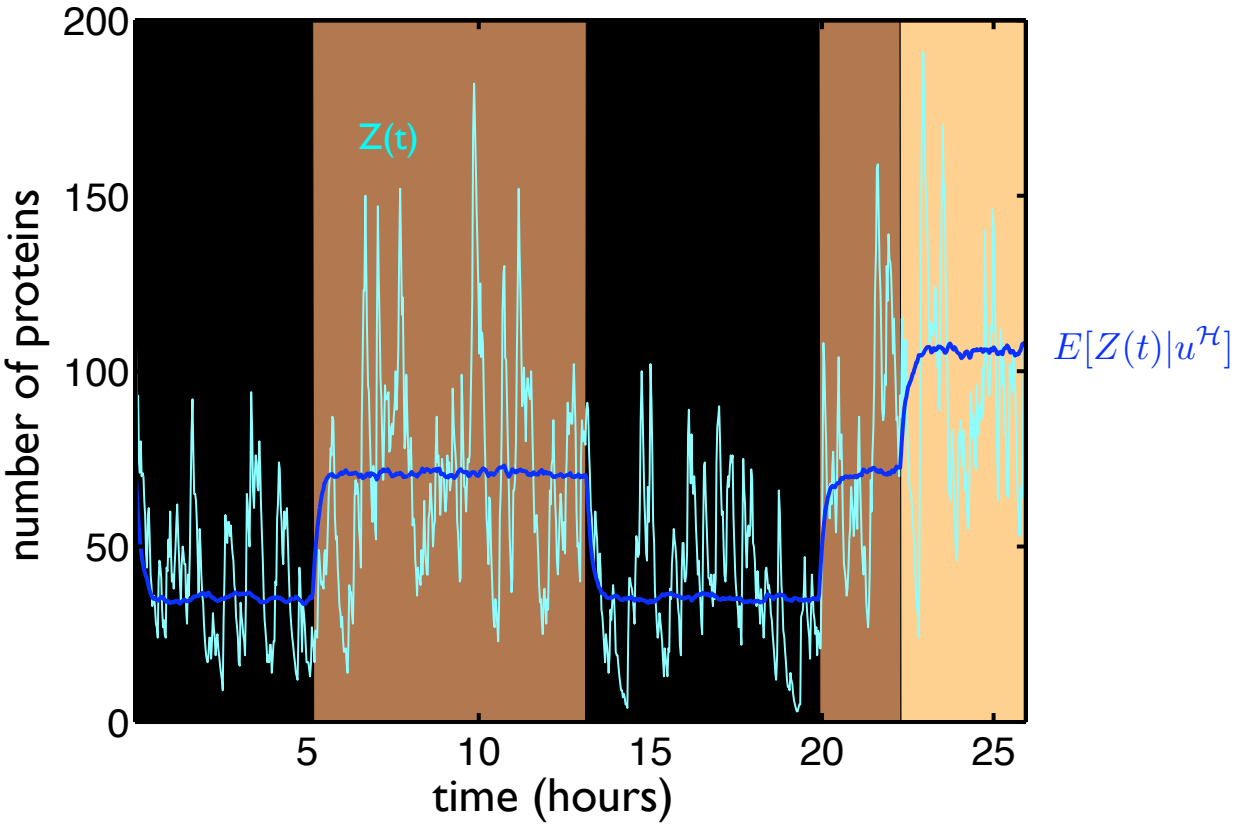


As the transduction mechanism becomes more noisy, the conditional output distributions broaden and information between the input and output decreases.



The expectation of the output conditional on the input tracks changes in the input: its variation is the informational variation.

Let the input, u, the number of active transcription factors, transition between **low**, **medium**, or **high** values.

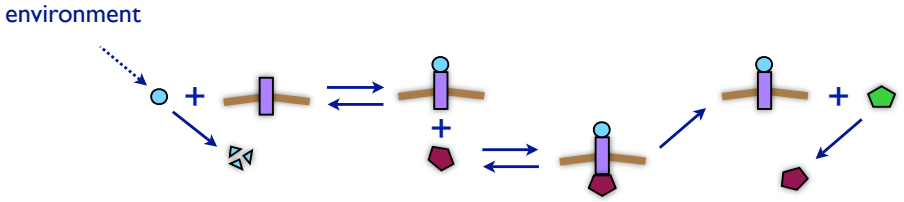


$E[Z(t)|u^{\mathcal{H}}]$ unambiguously tracks changes in the input and consequently conveys information on those changes and so

$$\text{informational variation} = V\Big\{E[Z|u^{\mathcal{H}}]\Big\}$$

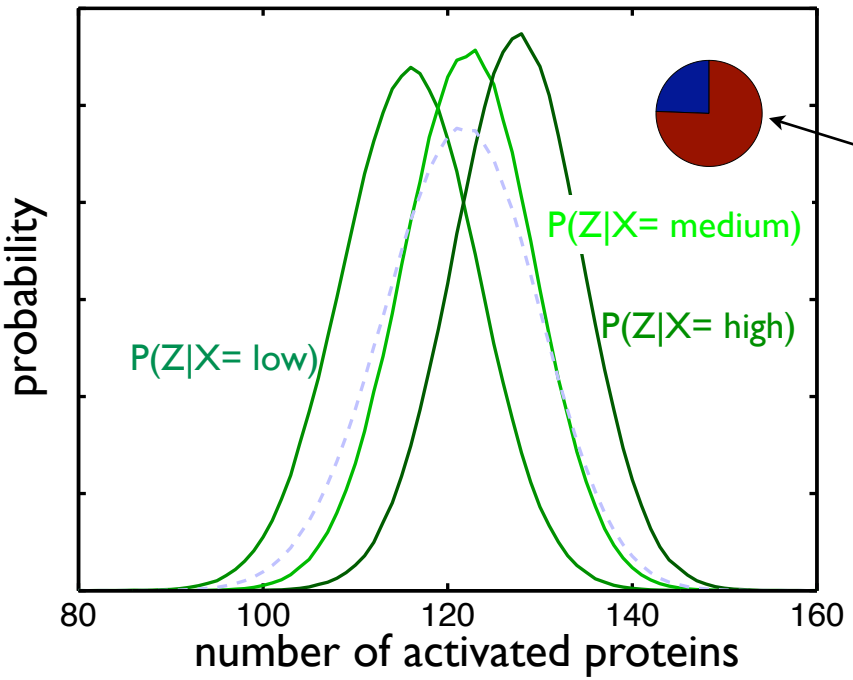
Changes in the informational fraction of variance predict changes in the mutual information between a network's input and output.

Consider an environment with three states and a signalling pathway.



A signalling protein is activated.

I[Z;X] = 0.14 nats

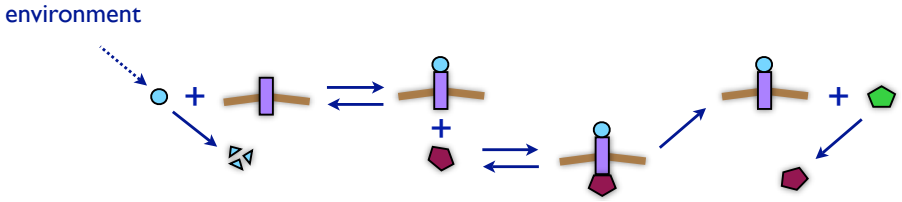


decomposition of the variation in output into informational and transductional components

As the ratio of the informational to the transductional components of the variance increases so too does the mutual information I[Z;X].

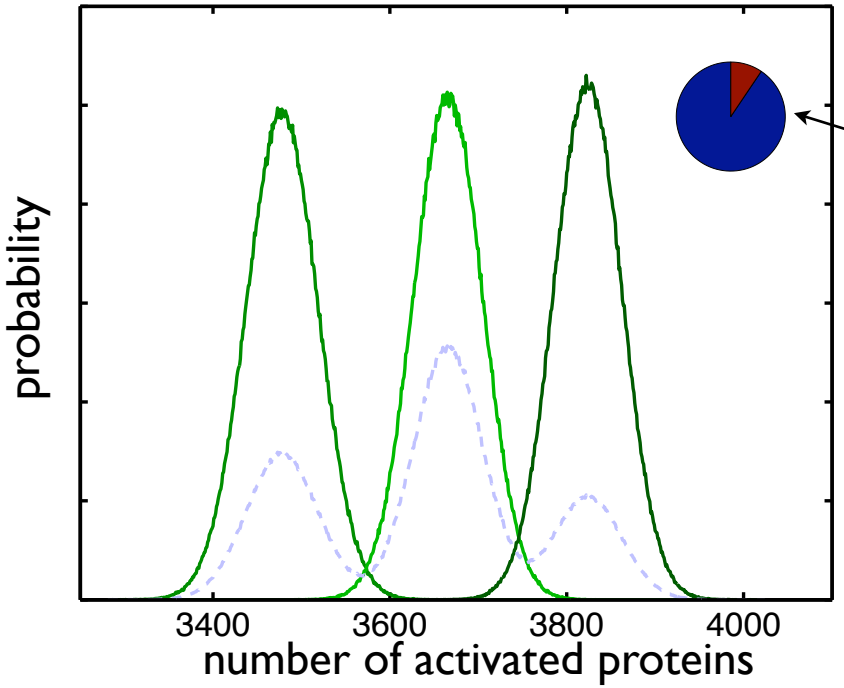
Changes in the informational fraction of variance predict changes in the mutual information between a network's input and output.

Consider an environment with three states and a signalling pathway.



A signalling protein is activated.

I[Z;X] = 0.98 nats

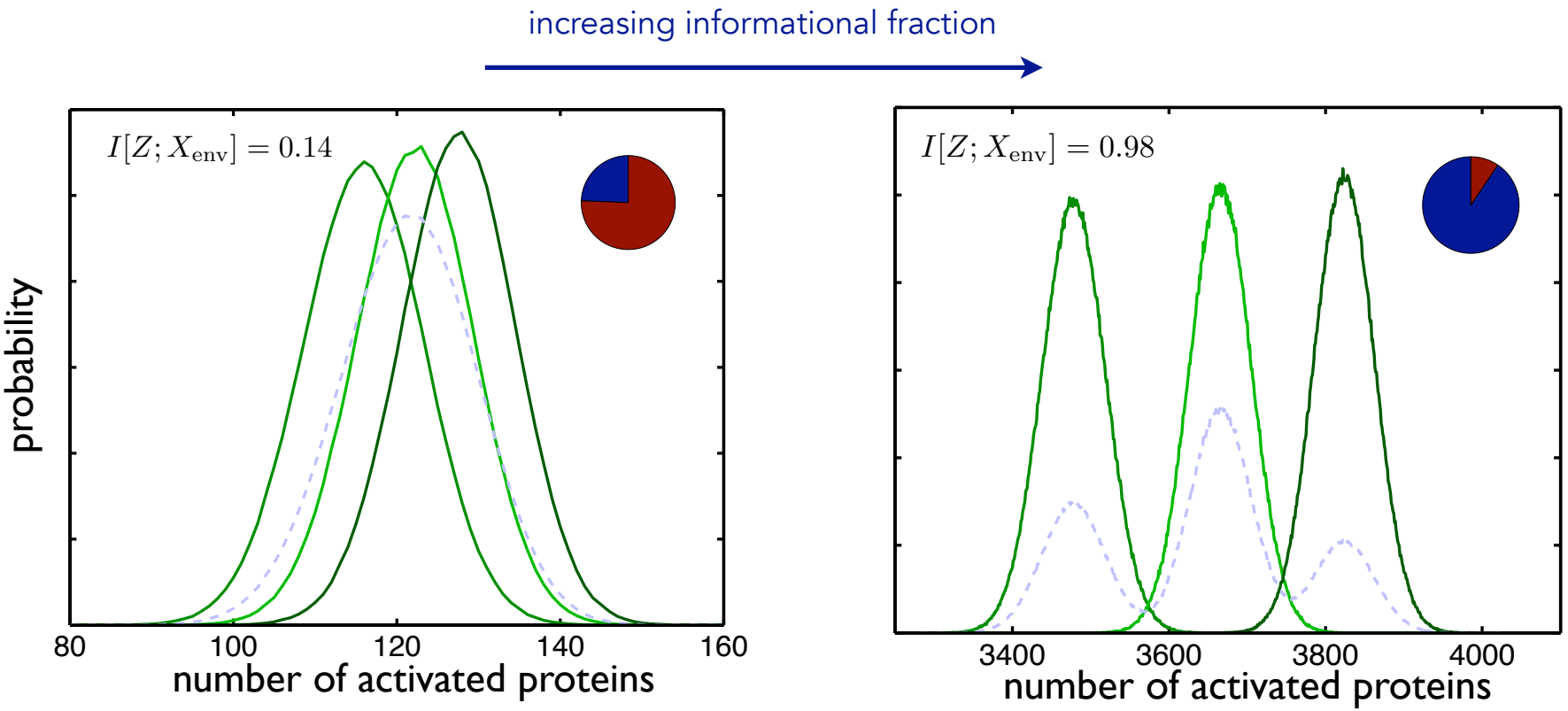


decomposition of the variation in output into informational and transductional components

As the ratio of the informational to the transductional components of the variance increases so too does the mutual information I[Z;X].

Increasing the information fraction typically causes the conditional output distributions to "separate" and so increases a network's information flow.

$$\text{informational fraction} = \frac{V\left\{E[Z|u^{\mathcal{H}}]\right\}}{V[Z]}$$

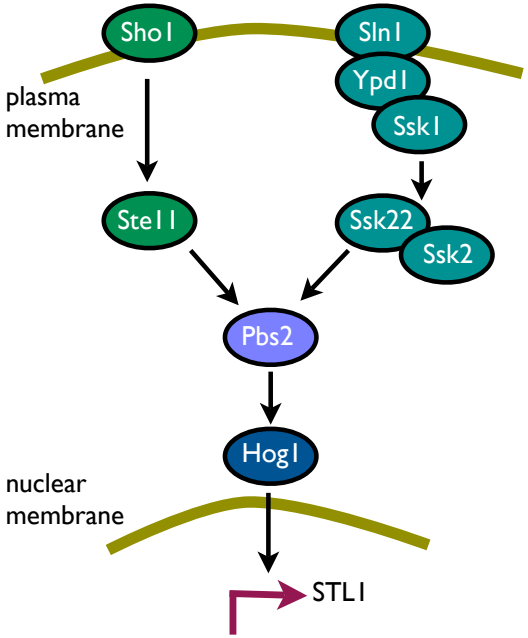


We can use the informational fraction for “inverse” ecology – to determine the probability distribution of input most favoured by a sensing network.

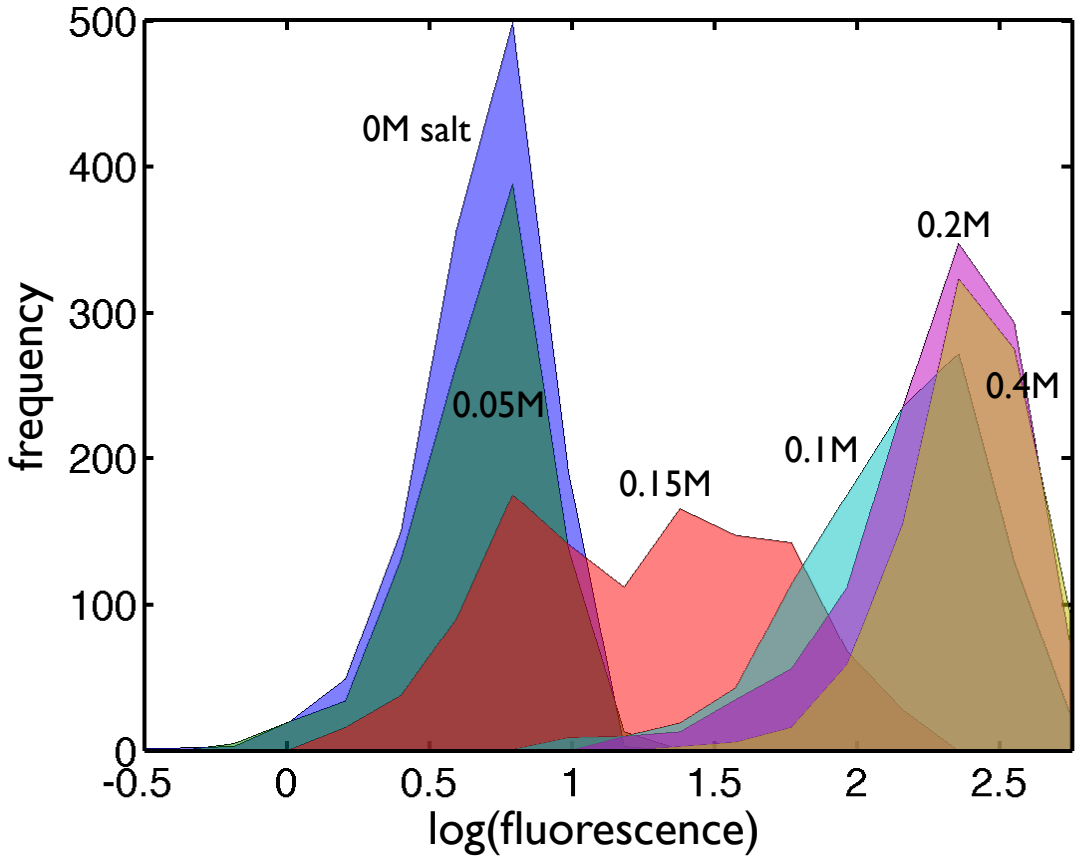
Hyperosmotic stress is sensed by two pathways in budding yeast. Pelet *et al*. used the promoter of STL1 to drive a fluorescent protein reporter of the network’s response in different concentrations of extracellular salt.

Transient Activation of the HOG MAPK Pathway Regulates Bimodal Gene Expression

Serge Pelet,[1*] Fabian Rudolf,[1†] Mariona Nadal-Ribelles,[2] Eulàlia de Nadal,[2] Francesc Posas,[2] Matthias Peter[1*]



From the data of Pelet *et al.*, we can calculate the informational fraction.



Pelet *et al.*, Science 2011;332: 732

Given a probability distribution for X, the informational fraction is

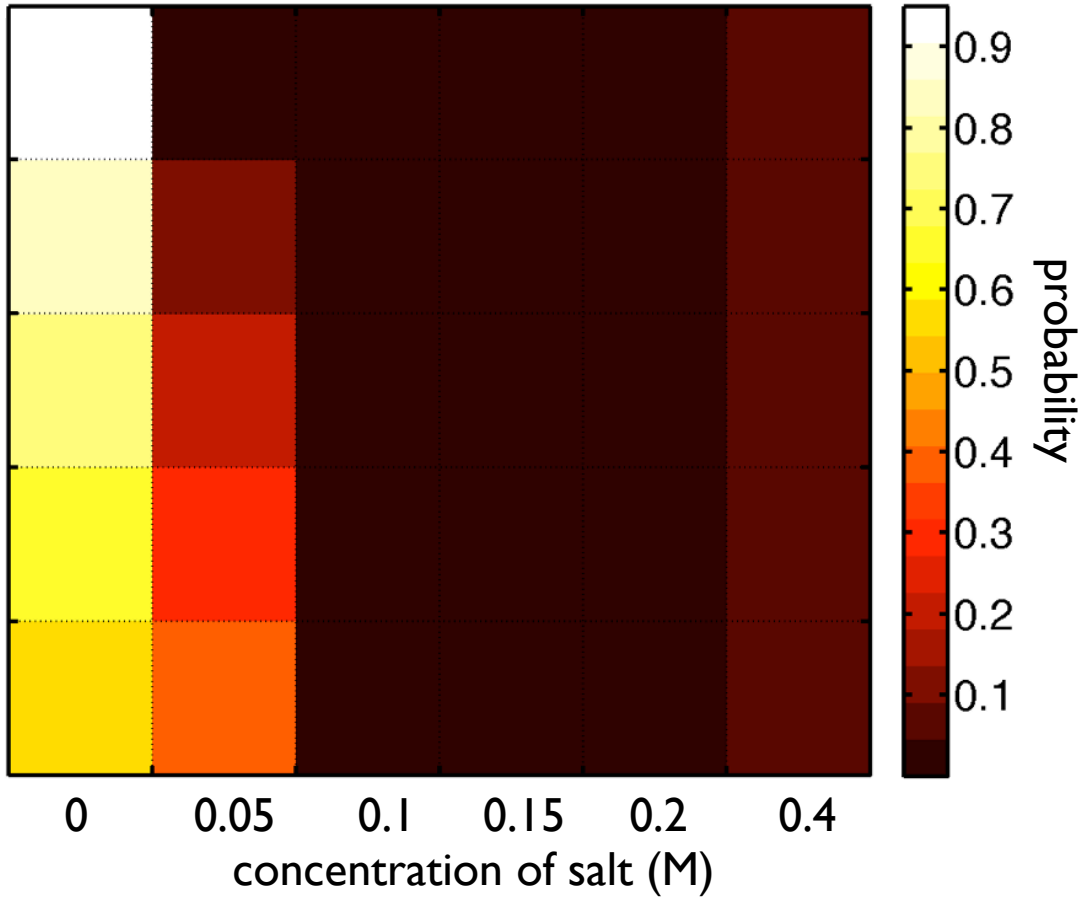
extracellular salt

$$\frac{V\Big\{E[Z|X]\Big\}}{V[Z]} = \frac{V\Big\{E[Z|X]\Big\}}{E\Big\{E[Z^2|X]\Big\} - \Big(E\Big\{E[Z|X]\Big\}\Big)^2}$$

fluorescence from STL1

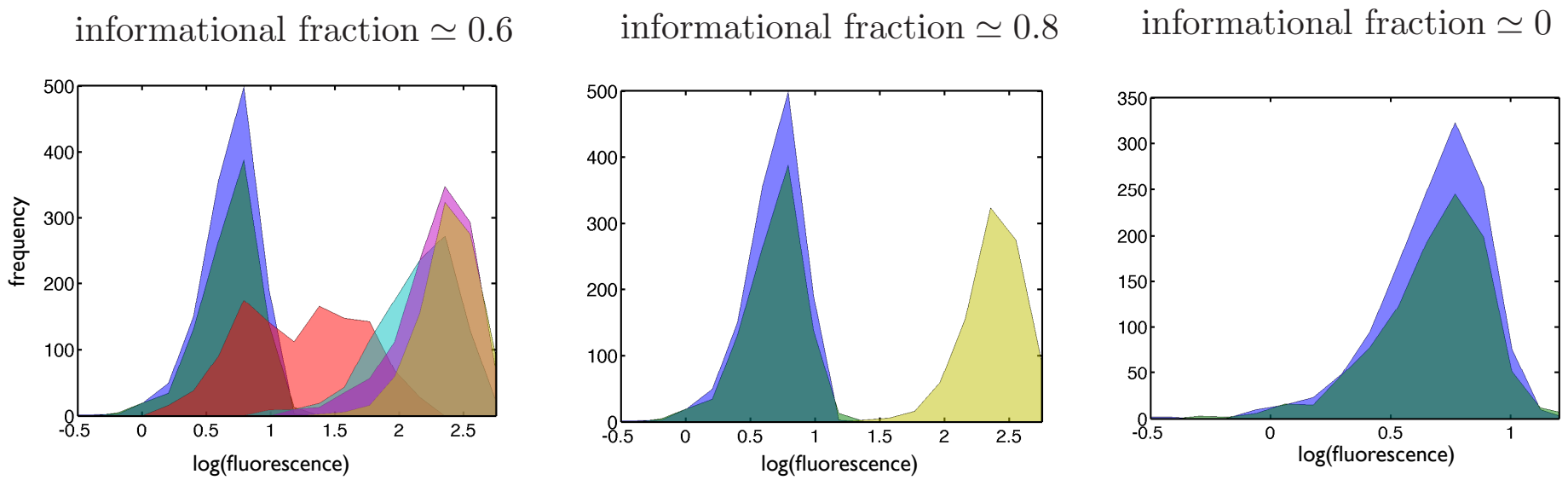
By considering all possible probability distributions of extracellular salt, we can find those that maximize the informational fraction.

The informational fraction varies continuously from 0.8 to 0.



Inverse ecology: yeast “expect” frequent low levels of osmotic stress interspersed with rare high levels.

Increasing the informational fraction decreases the overlap between the output distributions for each salt concentration.



Conclusions

1. We have a general decomposition of variation that holds for all dynamic systems at all times.
2. We can specify conditions that conjugate reporters should satisfy to measure each component of the decomposition.
3. We can distinguish information flow from noise.
4. We can use conjugate reporters in models to calculate the magnitude of the components of the decomposition.

Clive Bowsher
University of Bristol

Margaritis Voliotis
University of Bristol

Bowsher & Swain, 2012
Bowsher *et al.*, 2013
swainlab.bio.ed.ac.uk

SULSA
Scottish Universities Life Sciences Alliance

SynthSys
SYNTHETIC & SYSTEMS BIOLOGY