

Supporting Online Materials for Pedraza *et al.*

Noise propagation in gene networks

MATERIALS AND METHODS

Construction of the synthetic network

The network shown in Fig. 1A is implemented in *Escherichia coli* using the plasmid pJM31 (Fig. S1). This plasmid contains all components of the network except the lactose repressor gene, *lacI*, that is integrated in the *E. coli* genome under the strong promoter P_{lacIq} (*I*). The fluorescent proteins were obtained from the plasmids pECFP, pEYFP, and pDsRed-Express (Clontech). The promoters P_{tet} and P_{lac} were amplified from the plasmids pZE21-MCS2 and pZE12-luc (2), respectively (gift of R. Lutz and H. Bujard). The tetracycline repressor gene *tetR* was obtained from pIKE108 (3, gift of T. S. Gardner and J. J. Collins). The strong constitutive promoter P_L and the transcription terminator *aspA* were obtained from plasmid pLEX (Invitrogen). The T1T2 transcription terminator, the ampicillin gene, and the *colE1* origin of replication were obtained from PZE12-luc (2).

Strains, growth conditions and media

E. coli strain JM101 (Stratagene, *supE thi-1 (lac-proAB) [F' traD36 proAB lacIqZ.M15]*) was transformed with plasmid pJM31 and grown in M9 minimal media supplemented with 1 mM glucose and 50 μ g/ml of ampicillin. Expression levels were controlled by induction with IPTG (0 - 2 mM). Cells were grown overnight, rediluted and grown to an $OD_{600} \approx 0.005$ at 37°C. After concentrating by filtering and centrifugation, cells were observed on microscope slides at room temperature.

Data acquisition and analysis

Cells were observed using a Nikon TE-2000E microscope with automated stage (Prior) and focus using a cooled back-thinned CCD camera (Micromax, Roper Scientific). Typically 1000 to 3000 cells were measured per sample. Data analysis was performed using Metamorph (Universal Imaging). Objects with red fluorescence smaller than 0.5% or larger than 700% of the mean were discarded. Auto-fluorescence was quantified using JM101 cells lacking the fluorescent reporters. The experimentally obtained correlations were calculated by:

$$C_{ij} = \frac{\langle I_M J_M \rangle - \langle I_0 J_0 \rangle}{(\langle I_M \rangle - \langle I_0 \rangle)(\langle J_M \rangle - \langle J_0 \rangle)} - 1, \quad [1]$$

where I_M corresponds to the measured fluorescence of reporter I and I_0 corresponds to the background fluorescence of reporter I as measured in the JM101 cells without plasmid. Filter leak-through between the difference colors were determined using cells expressing only one of the three fluorescent reporters. Single cell fluorescence values were corrected for leak-through. Average expression in CFP and YFP were normalized to RFP average. The statistical errors were estimated using bootstrapping. Matlab (The Mathworks Inc.) was used for data fitting.

Having the circuit on a plasmid provides a higher signal to auto-fluorescence ratio, and reduces the relative contribution of the intrinsic noise without reducing the transmitted and global effects.

The assumption that the fluctuations in CFP accurately represent the fluctuations in TetR is based on the fact that intrinsic noise is determined by the events of production and destruction of mRNA (9). The contribution of the individual protein events is smaller by a factor of $\frac{1}{b}$ where b is the average number of proteins produced per mRNA, which from our fits is ~ 40 .

Analytical Langevin model

Our model is based on the Langevin technique (4-8), in which noise terms of appropriate strength are added to the deterministic equations. The result is equivalent to that obtained from the matrix formulation of the Fluctuation-Dissipation Theorem (9), since both methods are obtained from an Ω expansion on the master equation (7). We favor the Langevin approach because the different sources of fluctuations can be added in an intuitive manner and for a cascade it can be solved sequentially.

The concentration of lactose repressors y_0 is given by:

$$\dot{y}_0 = k - \gamma y_0 + \mu_0 + \xi_{0G}, \quad [2]$$

where k is the average rate of protein creation and γ is the effective protein decay rate, which is assumed to be dominated by the dilution rate due to cell growth. μ_0 represents the intrinsic fluctuations and ξ_G represents the fluctuations in rate constants due to global variations in the concentration of cellular components. The statistical properties of the stochastic terms are determined by

$$\begin{aligned} \langle \mu_0 \rangle &= \langle \xi_{0G} \rangle = 0, \\ \langle \mu_0(t) \mu_0(t + \tau) \rangle &= 2\gamma \tilde{b}_0 \bar{y}_0 \delta(\tau), \\ \langle \xi_{0G}(t) \xi_{0G}(t + \tau) \rangle &= 2\gamma \eta_G^2 \bar{y}_0^2 \delta(\tau), \end{aligned} \quad [3]$$

where b_o is the average numbers of proteins produced per mRNA ($\tilde{b}_i \equiv b_i + 1$), and η_G is the effective size of the global fluctuations. $\delta(t)$ denotes the delta function. The first autocorrelation is a good approximation to the intrinsic fluctuations in a single gene (10) in the limit where many proteins are produced per mRNA and the decay time for the mRNA is shorter than that of the protein. The fluctuations coming from the binding and unbinding of repressors to the DNA are small if the binding, unbinding, or mRNA decay rates are large compared to the rate of transcription (5). In cases where the other possible sources of noise are not negligible, the analysis that follows carries through except for the magnitude of the intrinsic noise which would include all sources of noise that are uncorrelated for two copies of the same gene. The second autocorrelation comes from assuming that the fluctuating global variables appear as multiplicative factors in the rates. The steady state fluctuations can be obtained by subtracting the steady state mean from [2], solving in Fourier space, squaring, averaging, and transforming back according to the Wiener-Khintchine theorem (hats denotes variables evaluated in Fourier space, bars denotes concentrations in steady-state):

$$\begin{aligned} \delta\tilde{y}_0 &= \frac{\hat{\mu}_0 + \hat{\xi}_{0G}}{i\omega + \gamma} \Rightarrow \langle \delta\tilde{y}_0^2 \rangle = \frac{\langle \hat{\mu}_0^2 \rangle + \langle \hat{\xi}_{0G}^2 \rangle}{\omega^2 + \gamma^2} = \frac{2\gamma\tilde{b}_0\bar{y}_0 + 2\gamma\eta_G^2\bar{y}_0^2}{\omega^2 + \gamma^2} \\ \Rightarrow \langle \delta y_0^2 \rangle &= \tilde{b}_0\bar{y}_0 + \eta_G^2\bar{y}_0^2 \Rightarrow \eta_0^2 = \frac{\tilde{b}_0}{\bar{y}_0} + \eta_G^2 \equiv \eta_{0\text{int}}^2 + \eta_G^2. \end{aligned} \quad [4]$$

η_0 and $\eta_{0\text{int}}$ denote the total and intrinsic noise of gene 0 respectively. Upon induction, the concentration of active repressors is $y_{0A} = f_0([IPTG])y_0$, where f_0 is assumed to be a Hill type function. Assuming fast binding of IPTG to the repressor and concentrations higher than nanomolar, this does not introduce significant additional fluctuations. Since the standard deviation scales as the mean, the noise in the number of active repressors is the same as the noise in the total number of repressors irrespective of IPTG levels.

For the other genes, it is necessary to consider the effects of plasmid copy number fluctuations. We are interested in time scales longer than the equilibration time after creation of a new plasmid and neglect cell division effects. Plasmid copy number is treated as a multiplicative, fluctuating factor on the creation rate with decay time equal to the cell doubling time. This results in the equation

$$\dot{y}_1 = Nf_1(y_{0A}) - \gamma y_1 + \mu_1 + \xi_{1G}, \quad [5]$$

where N is average plasmid copy number. Linearizing around steady state, we obtain

$$\delta\dot{y}_1 = \delta N\bar{f}_1 + Nc_1\delta y_{0A} - \gamma\delta y_1 + \mu_1 + \xi_{1G}, \quad [6]$$

where $\bar{f}_1 = f_1(\bar{y}_{0A})$ and $c_1 = \frac{df_1(y_{0A})}{dy_{0A}} \Big|_{\bar{y}_{0A}}$. This approximation restricts the model to the bulk of the distribution, because the tails of wide input distributions explore a large part of f_1 's domain. Proceeding as described for the previous gene, using [4] for the variation in y_{0A} , and the assumption that the global fluctuations are fully correlated, we can obtain an explicit expression for the noise in y_1 (CFP concentration). In terms of the logarithmic gain, $H_{10} \equiv -\frac{\partial \ln(y_1)}{\partial \ln(y_{0A})} \Big|_{\bar{y}_{0A}}$, we obtain:

$$\begin{aligned} \delta \tilde{y}_1 &= \frac{\bar{f}_1 \delta \hat{N} + \bar{N} c_1 \delta \tilde{y}_{0A} + \hat{\mu}_1 + \hat{\xi}_{1G}}{i\omega + \gamma} \Rightarrow \\ \eta_1^2 &= \frac{\tilde{b}_1}{\bar{y}_1} + \frac{1}{2} H_{10}^2 \frac{\tilde{b}_0}{\bar{y}_0} + \frac{1}{2} \eta_N^2 + \eta_G^2 \left(1 + \frac{1}{2} H_{10}^2 - H_{10} \right) \\ &= \eta_{\text{int}}^2 + \frac{1}{2} H_{10}^2 \eta_0^2 + \frac{1}{2} \eta_N^2 + \eta_G^2 (1 - H_{10}) \end{aligned} \quad [7]$$

where η_1 and $\eta_{1\text{int}}$ denote the total and intrinsic noise of gene 1 respectively and η_N denotes noise due to plasmid copy number fluctuations. The terms that determine the effect of the global noise on gene 1 correspond to its direct effect, the transmitted effect from the repressor and a negative correction term which arises because the direct and transmitted effects are strongly correlated. For gene 2 we similarly obtain

$$\begin{aligned} \eta_2^2 &= \frac{\tilde{b}_2}{\bar{y}_2} + \frac{1}{2} H_{21}^2 \frac{\tilde{b}_1}{\bar{y}_1} + \frac{3}{8} H_{21}^2 H_{10}^2 \frac{\tilde{b}_0}{\bar{y}_0} + \eta_N^2 \left(\frac{1}{2} + \frac{3}{8} H_{21}^2 - \frac{3}{4} H_{21} \right) \\ &+ \eta_G^2 \left(1 + \frac{1}{2} H_{21}^2 + \frac{3}{8} H_{21}^2 H_{10}^2 - H_{21} - \frac{3}{4} H_{21}^2 H_{10} + \frac{1}{2} H_{21} H_{10} \right) \\ &= \eta_{2\text{int}}^2 + \frac{1}{2} H_{21}^2 \eta_1^2 + \frac{1}{8} H_{21}^2 H_{10}^2 \eta_0^2 + \eta_N^2 \left(\frac{1}{2} + \frac{1}{8} H_{21}^2 - \frac{3}{4} H_{21} \right) \\ &+ \eta_G^2 \left(1 - H_{21} - \frac{1}{4} H_{21}^2 H_{10} + \frac{1}{2} H_{21} H_{10} \right) \end{aligned} \quad [8]$$

η_2 and $\eta_{2\text{int}}$ denote the total and intrinsic noise of gene 2 respectively. It is important to note that the expression for η_2 cannot be rewritten as the transmission of the total steady state noise in gene 1 only. The noise coming from gene 0 (*lacI*) appears explicitly even though η_1 includes its effect, because the time averaging coefficient is different (9,11) as the transmission changes the frequency spectrum of the noise. This means that it is not possible to determine the steady state noise pairwise by iterating the expression for two genes and the entire network has to be considered. This is because the noise η_1 is a steady

state measure of the fluctuations and does not include information on the temporal correlations. A full dynamic description of the fluctuations in the previous gene would be required to determine the downstream noise without information on the rest of the network.

The numerical prefactors in all terms come from time averaging due to the finite protein life-time. It is assumed that the effective protein and plasmid life-times are dominated by dilution due to cell growth and therefore are determined by the cell doubling time. The coefficients presented here correspond to static global noise. We presented our results using this simple assumption to make the formulas tractable. For more details on the coefficients, see (9). It is easy to incorporate a time dependence on the global noise, but this requires knowledge about the dynamics of global noise sources. This changes the coefficients but does not change any of the conclusions.

In the case of the second gene the global noise **[8a]** has one direct term, two transmission terms (positive), two correction terms (negative) and one extra term (positive) that compensates for over-correction. For both genes 1 and 2, this can result in a smaller global noise component than if the gene were unregulated. The modulation of global noise can be seen more clearly if the expression for η_2 is rewritten in terms of the total noise in the previous genes **[8b]**. This shows a possible design principle: the transmitted noise would be the same for repression or activation with the same sensitivity, but the global noise would be reduced under repression and amplified under activation.

The correlation coefficients, $C_{ij} = \frac{\langle y_i y_j \rangle - \langle y_i \rangle \langle y_j \rangle}{\langle y_i \rangle \langle y_j \rangle}$, can be obtained from the product of the expressions for the fluctuations in Fourier space using the same procedure as above. We obtain

$$\begin{aligned}
C_{12} &= -\frac{1}{2} H_{21} \eta_{\text{int}}^2 - \frac{3}{8} H_{21} H_{10}^2 \eta_{0\text{int}}^2 + \eta_N^2 \left(\frac{1}{2} - \frac{3}{8} H_{21} \right) \\
&+ \eta_G^2 \left(1 - \frac{1}{2} H_{21} - \frac{3}{8} H_{21} H_{10}^2 - \frac{1}{2} H_{10} + \frac{3}{4} H_{21} H_{10} \right) \\
C_{03} &= \eta_G^2 \\
C_{13} &= \frac{1}{2} \eta_N^2 + \eta_G^2 \left(1 - \frac{1}{2} H_{10} \right) \\
C_{23} &= \eta_N^2 \left(\frac{1}{2} - \frac{3}{8} H_{21} \right) + \eta_G^2 \left(1 - \frac{1}{2} H_{21} + \frac{1}{4} H_{21} H_{10} \right).
\end{aligned} \tag{9}$$

The fact that C_{I3} is not a constant as a function of IPTG (Fig. 2E, Fig. S3B) is a direct demonstration of the presence of global noise. Since gene 0 is integrated in the *E. coli* chromosome this effect cannot be explained by plasmid copy number fluctuations.

Fitting parameters

The interactions between genes are determined by the transfer functions:

$$\begin{aligned}
 f_0([IPTG]) &= \frac{A_0}{1 + \left(\frac{[IPTG]}{B_0}\right)^{h_0}} + basal_0, \\
 f_1(y_{0A}) &= \frac{A_1}{1 + \left(\frac{y_{0A}}{B_1}\right)^{h_1}} + basal_1, \\
 f_2(y_1) &= \frac{A_2}{1 + \left(\frac{y_{1A}}{B_2}\right)^{h_2}} + basal_2.
 \end{aligned}
 \tag{10}$$

The intrinsic noise for each gene is determined by the burst size b_i . The global and plasmid noise are characterized by the parameters η_G and η_N . It is possible to fit the transfer functions from the means, and to use those parameters to fit the noises with the remaining parameters. However, the lack of a direct reporter for the repressor number makes the fitting from just the means undetermined. We therefore simultaneously fit the means, noises and correlations. In the present case, this extra information can be used to obtain a better determined fit than with the means or means and noises alone. Fitting with just the means implies determination of all parameters in the three Hill functions with two sets of data. Including the noise adds constrains but also requires four more parameters. The correlations, however, contain no new parameters, and therefore significantly constrain the fits. The errors were estimated by the range for each parameter that allowed a fit with a variation of less than 10 % of the best P value. The obtained parameter values including their errors are given in Table S1.

The fraction of activated TetR as determined by the concentration of ATC is given by:

$$f_a([ATC]) = \frac{A_a}{1 + \left(\frac{[ATC]}{B_a}\right)^{h_a}} + basal_a. \quad [11]$$

The parameters in this expression were obtained from the experimentally measured values of YFP fluorescence versus ATC at full IPTG induction (Fig. S3A). The obtained parameter values including their errors are given in Table S2. The parameters were used to predict the noises and correlations as the concentration of ATC is changed (Fig. 4C-F).

Supporting Figure S1

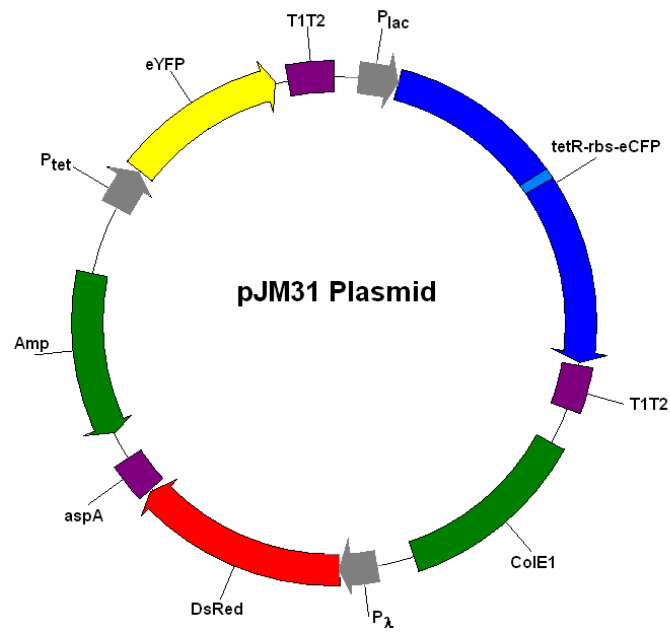


Figure S1. Plasmid map

Supporting Figure S2

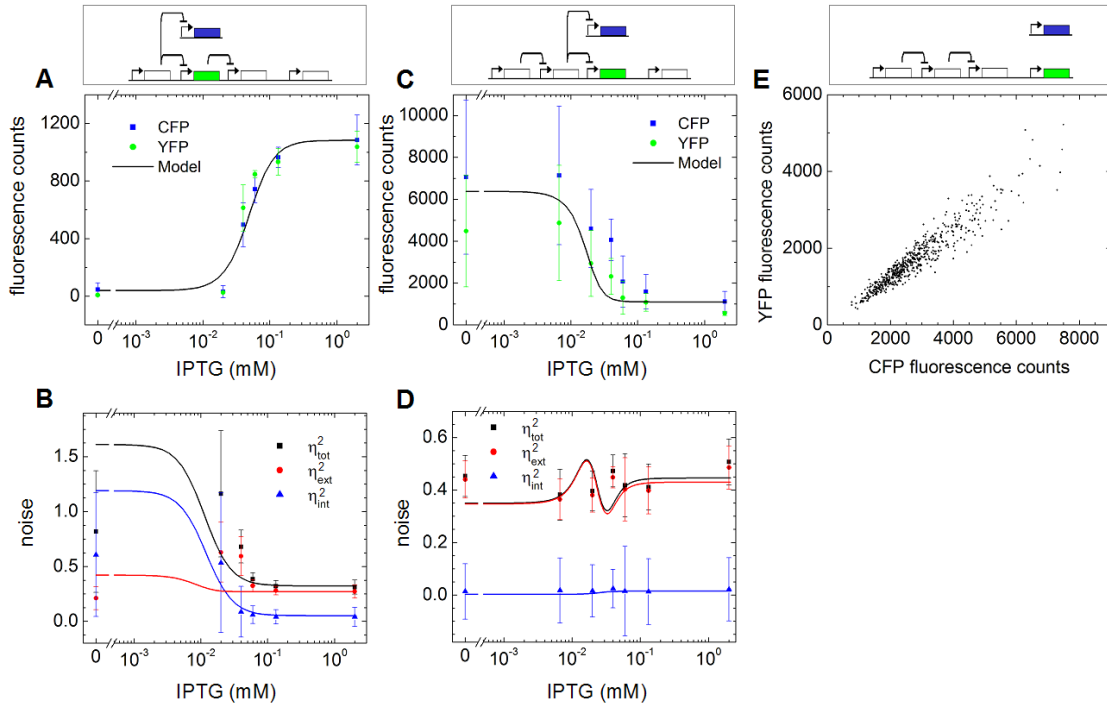


Figure S2. Direct measurements of intrinsic and extrinsic noise. **(A,B)** Two identical copies of the lac promoter are driving CFP and YFP. The average fluorescence is obtained from about 2000 single cells **(A)**. Intrinsic and extrinsic noise as a function of IPTG **(B)** are calculated according to Elowitz et al. (13). The solid lines represent prediction by the stochastic model. The relatively large error bars on the noise at low IPTG concentrations are due to the fact that the CFP and YFP levels are close to the auto fluorescence levels. **(C,D)** Two identical copies of the tet promoter are driving CFP and YFP. The average fluorescence **(C)** and noise **(D)** are plotted as a function of IPTG. The solid lines represent prediction by the stochastic model. The total noise behaves differently compared to Fig. 2B. This is due to the doubling of the number of binding sites for the tet-repressor resulting in an increased basal expression. The model parameters used are those in Table S1, except for the basal transcription which was adjusted to the measured value. Duplicated measurements are averaged. The error bars reflect the standard deviation of run to run differences and the error within each measurement as determined by bootstrapping. **(E)** Two identical copies of the lambda promoter driving CFP and YFP. A strong correlation between single cell expression values of CFP and YFP is observed: $\frac{\eta_{ex}^2}{\eta_{ex}^2 + \eta_{in}^2} \sim 95\%$. Therefore the intrinsic noise of gene 3 can be neglected.

Supporting Figure S3

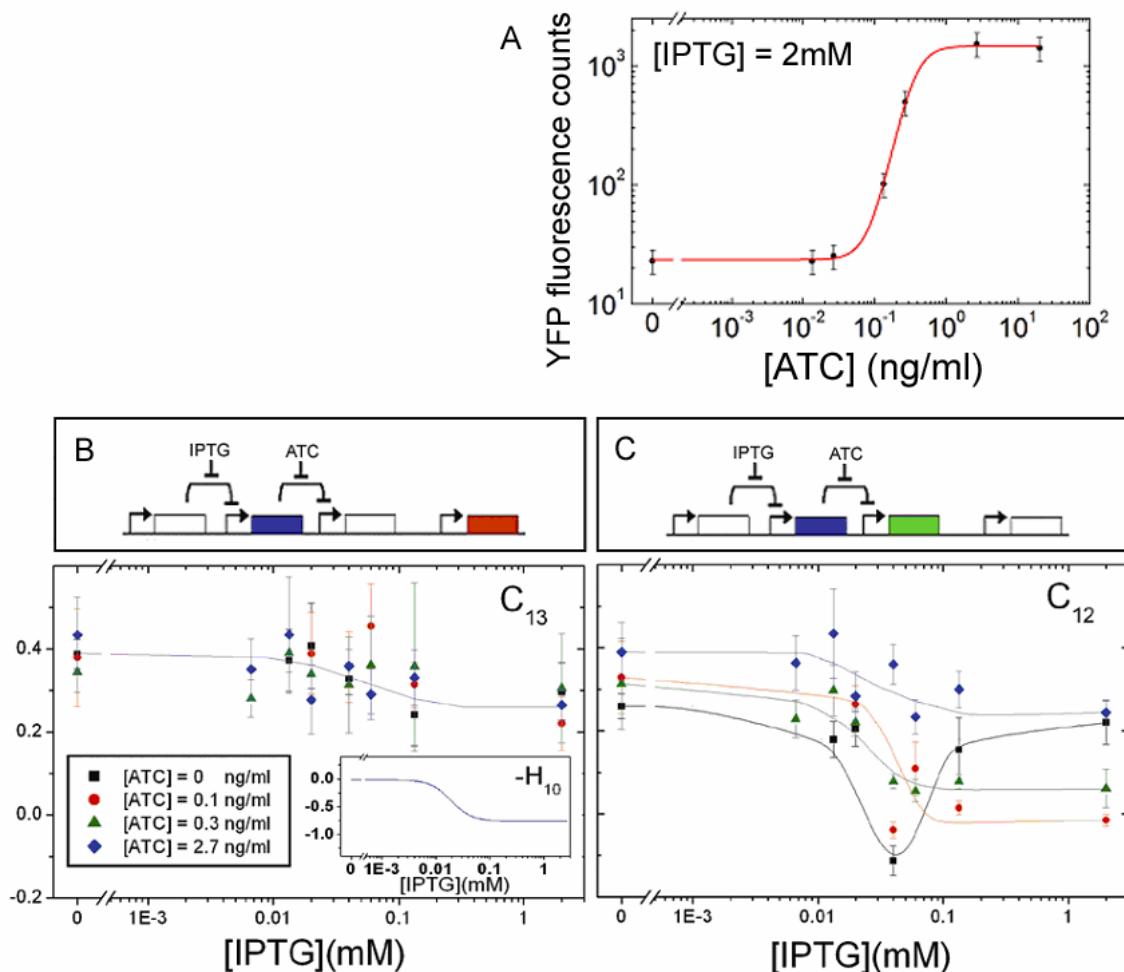


Figure S3. (A) ATC response at full IPTG induction. (B,C) Correlations C_{13} (B) and C_{12} (C) as a function of IPTG for different concentrations of ATC. The correlations C_{13} and C_{23} depend on the global noise experienced by genes 1 and 2, respectively. The correlations C_{12} and C_{23} (data not shown) increase as the interaction between gene 1 and gene 2 is weakened by addition of ATC while the correlation C_{13} is unchanged. Similarly, for maximal ATC induction genes 1 and 2 are almost decoupled ($H_{21} \approx 0$), and therefore C_{12} at full ATC induction should match C_{13} , as is experimentally observed. The remaining IPTG-dependence is due to global noise propagation from gene 0, as determined by the logarithmic gain H_{10} (inset).

Table S1. Model parameters.

	<i>Value</i>	<i>Error</i>	<i>Referen ce</i>
A_0	100		(1)
B_0	0.04	± 0.02	
h_0	2.3	± 0.3	
$basal_0$	0.1		(1)
A_1	184	-120 / +557	
B_1	0.007	-0.006 / +0.025	
h_1	0.89	± 0.15	
$basal_1$	1.98	± 0.07	
A_2	57.6	± 4.3	
B_2	160	± 6	
h_2	3.7	± 0.2	
$basal_2$	0.52	± 0.06	
b_0	10		(1)
b_1	43	± 7	
b_2	6	± 6	
η_N	0	+0.15	
η_G	0.652	± 0.16	

Table S2. Parameters for ATC induction.

	<i>Value</i>	<i>Error</i>
A_a	0.90	± 0.13
B_a	0.072	± 0.034
h_a	1.9	± 1.4
$basal_a$	0.105	± 0.086

Supplementary references and notes:

1. B. Müller-Hill, L. Crapo, W. Gilbert, *Proc. Natl. Acad. Sci. USA* **59**, 1259 (1968).
The parameters were obtained for *E. coli* strain i^d in M56 minimal media with glycerol as carbon source; they should be taken as estimates only.
2. R. Lutz, H. Bujard, *Nucleic Acids Research* **25**, 1203 (1997).
3. T. S. Gardner, C. R. Cantor, J. J. Collins, *Nature* **403**, 339 (2000).
4. M. Thattai, A. van Oudenaarden, *Biophys. J.* **82**, 2943 (2002).
5. T. B. Kepler, T. C. Elston, *Biophys. J.* **81**, 3116 (2001).
6. J. Hasty, *et al.*, *Proc. Natl. Acad. Sci. USA* **97**, 2075, (2000).
7. N. G. van Kampen, *Stochastic Processes in Physics and Chemistry* (North-Holland, Amsterdam, 1981).
8. P. B. Detwiler, S. Ramanathan, A. Sengupta, B. Shraiman, *Biophys. J.* **79**, 2801 (2000).
9. J. Paulsson, *Nature* **427**, 415 (2004).
10. M. Thattai, A. van Oudenaarden, *Proc. Natl. Acad. Sci. U.S.A.* **98**, 8614 (2001).
11. M. L. Simpson, C. D. Cox, G. S. Sayler, *Proc. Natl. Acad. Sci. U.S.A.* **100**, 4551 (2003).
12. D.T. Gillespie, *J. Phys. Chem.* **81**, 2340 (1977)
13. M. B. Elowitz, A. J. Levine, E. D. Siggia, P. S. Swain, *Science* **297**, 1183 (2002).