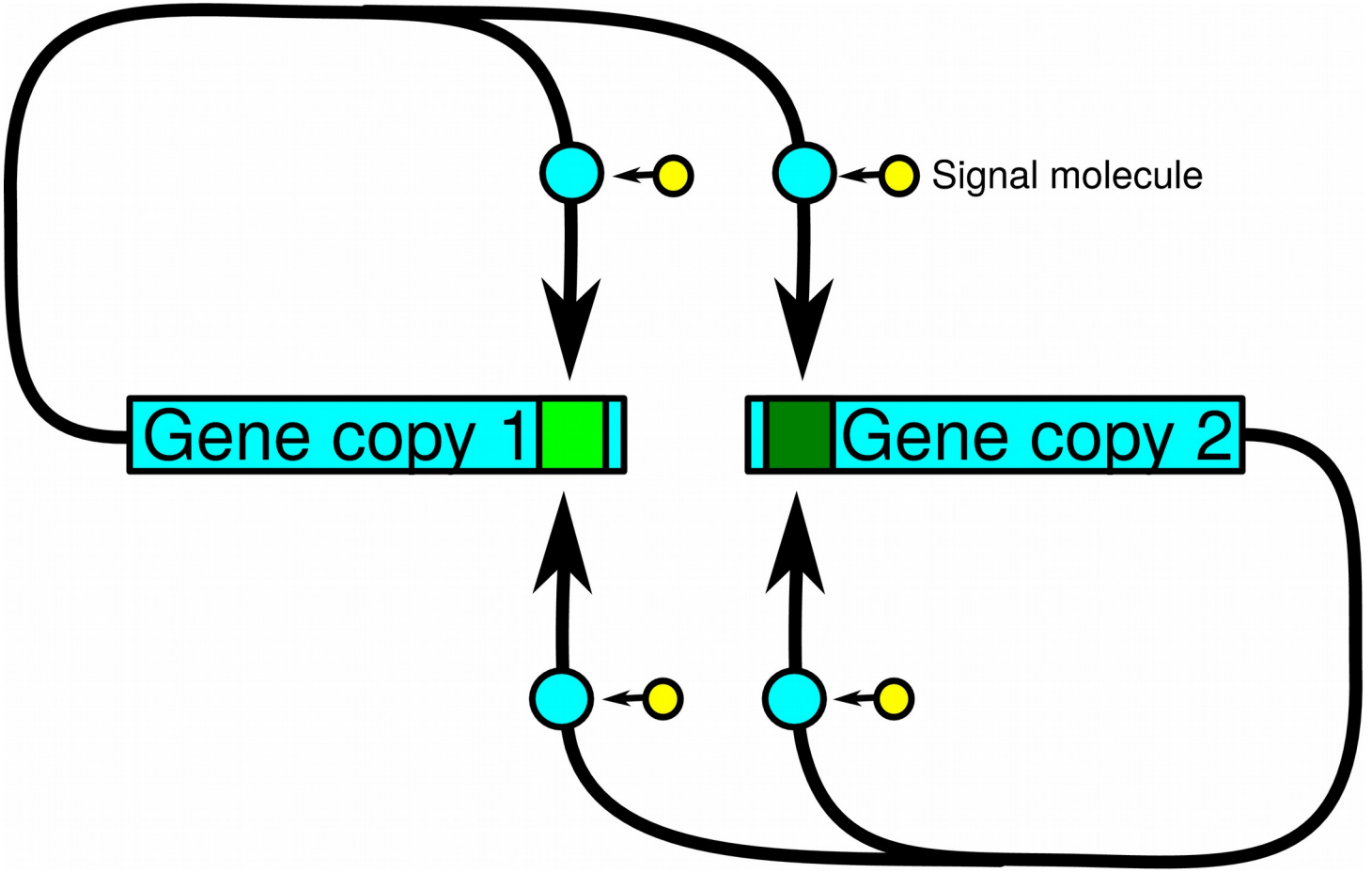
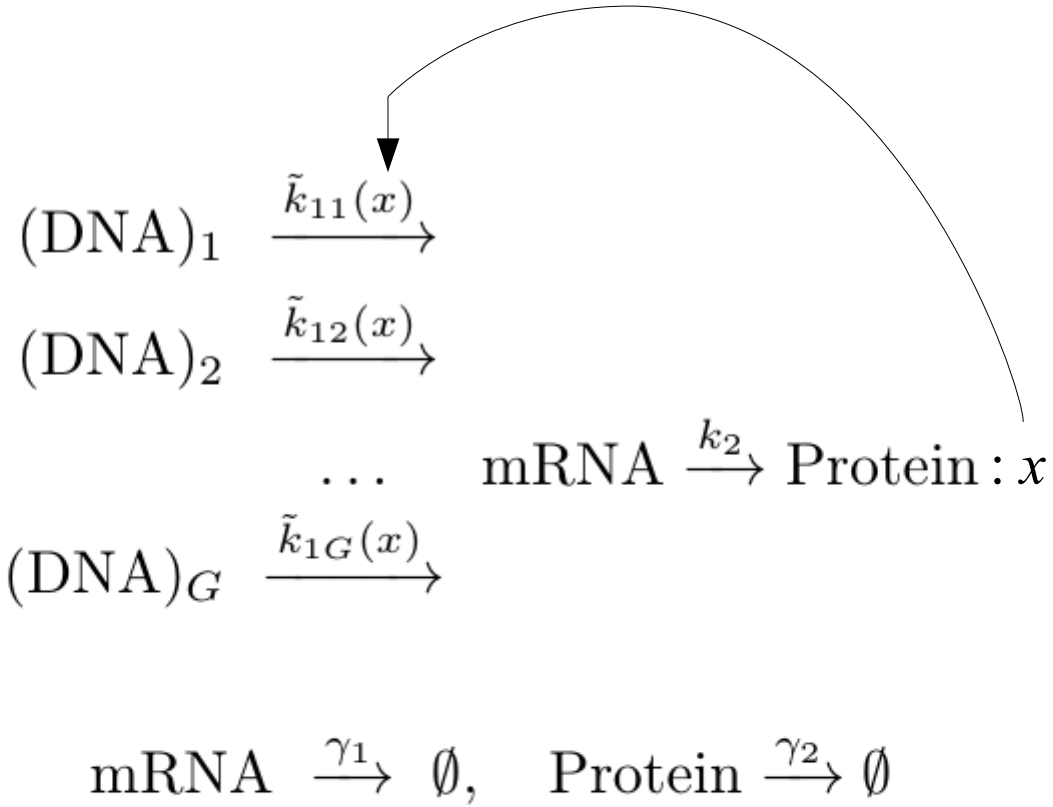
Gene multiplication:
A simple phenomenon that may cause non-intuitive effects

Anna Ochab-Marcinek,
Jakub Jędrak

Institute of Physical Chemistry
Polish Academy od Sciences



J. Jędrak and A. Ochab-Marcinek, Journal of Theoretical Biology **408**, 222 (2016).



G gene copies.

Transcription rates depend on protein concentration x.

Effective rate constants: $\tilde{k}_{1j}(x) = k_{1j} h_j(x)$.

Transcription factor binding:

$$_{\mathrm{j}}\mathrm{O} + \mathrm{nX} \overset{K_j}{\rightleftharpoons} {}_{\mathrm{j}}\mathrm{OX}_{\mathrm{n}}$$

mRNA synthesis/degradation:

Repressor

$$_{\mathrm{j}}\mathrm{O} \xrightarrow{k_{1\mathrm{j}}} \mathrm{Y} + {}_{\mathrm{j}}\mathrm{O}$$

$$_{\mathrm{j}}\mathrm{OX}_{\mathrm{n}} \xrightarrow{k_{1\mathrm{j}\epsilon}} \mathrm{Y} + {}_{\mathrm{j}}\mathrm{OX}_{\mathrm{n}}$$

Activator

$$_{\mathrm{j}}\mathrm{O} \xrightarrow{k_{1\mathrm{j}\epsilon}} \mathrm{Y} + {}_{\mathrm{j}}\mathrm{O}$$

$$_{\mathrm{j}}\mathrm{OX}_{\mathrm{n}} \xrightarrow{k_{1\mathrm{j}}} \mathrm{Y} + {}_{\mathrm{j}}\mathrm{OX}_{\mathrm{n}}$$

$$\mathrm{Y} \xrightarrow{\gamma_1} \varnothing$$

Transcription factor synthesis/degradation:

$$\mathrm{Y} \xrightarrow{k_2} \mathrm{X} + \mathrm{Y}$$

$$\mathrm{X} \xrightarrow{\gamma_2} \varnothing$$

TABLE I: Kinetic scheme. X: protein, Y: mRNA, k_{1j}: rate of mRNA synthesis from the operator of the j-th gene copy in the active state, $k_{1j\epsilon}$: rate of mRNA synthesis from the operator of the j-th gene copy in the inactive state (leakage), γ_1: rate of mRNA degradation, k_2: rate of protein synthesis, γ_2: rate of protein degradation.

- Transfer function:

$$h_j(x) = (1 - \epsilon_j)H_j(x) + \epsilon_j, \qquad j = 1, 2, \ldots, G$$

- Transcriptional leakage: $\epsilon_j = k_{1j\epsilon}/k_{1j}$

- Hill kinetics:

$$H_j(x) \;=\; \left[1 + \left(\frac{x}{K_j}\right)^{n_j}\right]^{-1}$$

$$\frac{\partial p(x,t)}{\partial t} = \gamma_2 \sum_{j=1}^{G} a_j \int_0^x w(x-x') h_j(x') p(x',t) dx' + \gamma_2 \frac{\partial}{\partial x} \left[x p(x,t) \right]$$

$$a_j \equiv \frac{k_{1j}}{\gamma_2}, \qquad b \equiv \frac{k_2}{\gamma_1}$$

- $x \geq 0$ is a continuous variable
- u: burst size, $w(u) = \nu(u) - \delta(u)$ [1]
- Probability distribution of burst sizes (identical for each gene copy):

$$\nu(u) = (1/b) \exp(-u/b)$$

[1] N. Friedman, L. Cai, and X. S. Xie, Phys. Rev. Lett. **97**, 168302 (2006).

$$p(x) = Ax^{-1}e^{-x/b}\prod_{j=1}^{G}\exp\left[a_j\int\frac{h_j(x)}{x}dx\right]$$

$$p(x) = Ax^{-1}e^{-x/b}\prod_{j=1}^{G}x^{a_j}H_j(x)^{\frac{a_j(1-\epsilon_i)}{n_j}}$$

- Bursting of each gene copy is a Poisson process, independent from the bursting of all other copies.
- Thus, their protein production rates are coupled only by the common pool of proteins that regulate the genes as their TFs.

Identical gene copies

- $h_j(x) = h(x),\ a_j = a$
- The maximum burst frequency scales linearly with gene copy number

$$a \to Ga$$

$$p_G(x) \quad = \quad Ax^{Ga-1}e^{-x/b}\left[1+\left(\frac{x}{K}\right)^n\right]^{-\frac{Ga(1-\epsilon)}{n}}$$

Two standard quantitative measures of gene expression noise are used interchangeably in literature:

- Fano factor $F = \sigma^2/\mu_1$
- Coefficient of variation $\eta = \sigma/\mu_1$

Non-regulated gene – Volfson et al. [2]:

$$\langle x \rangle \sim G^1, \quad \eta \sim G^{-\frac{1}{2}}, \quad F \sim G^0$$

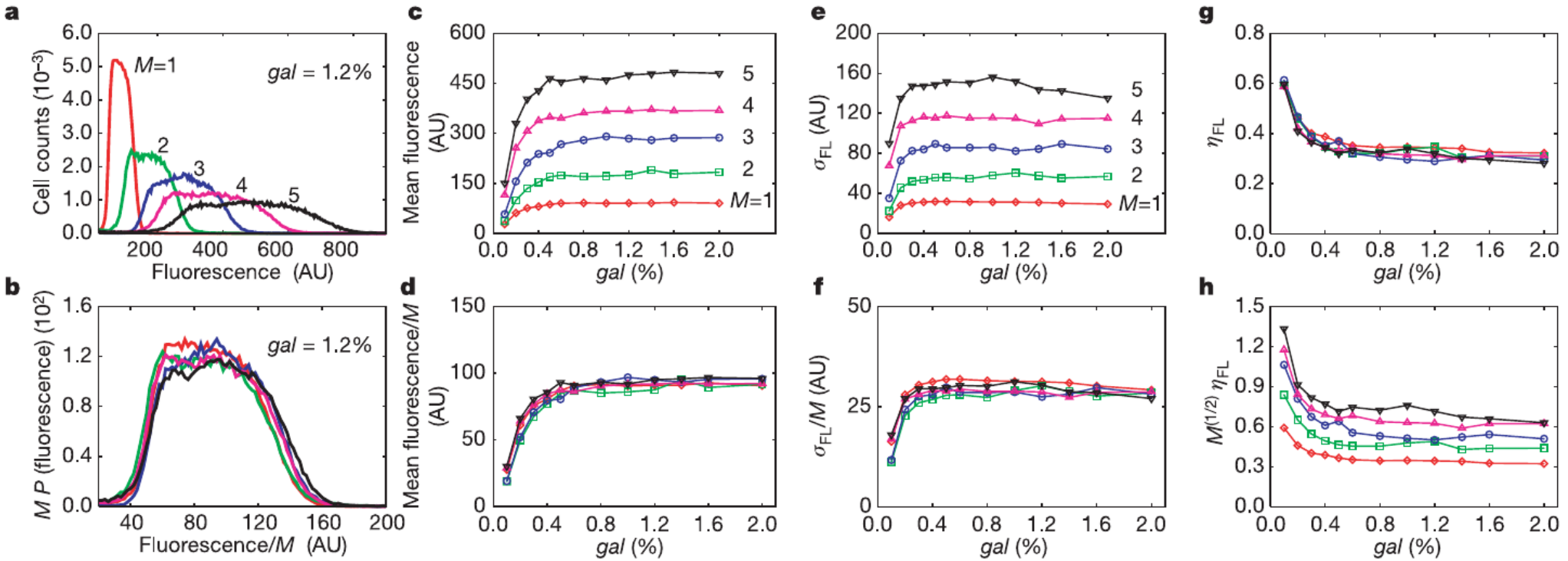
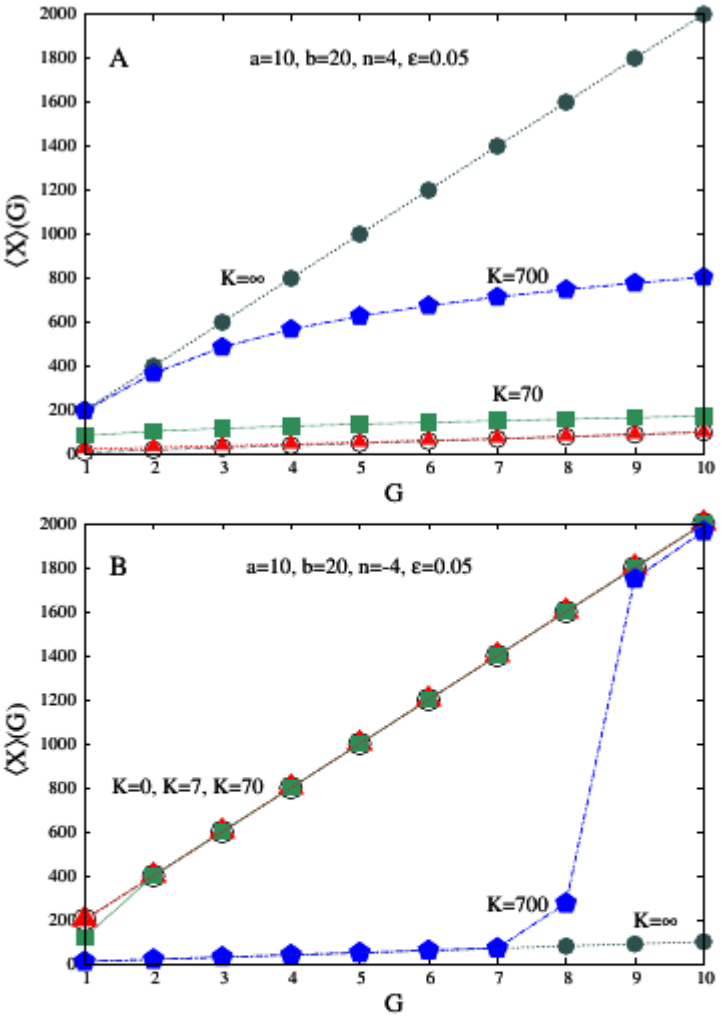


Figure 1 | Experimental results for GFP expression for different copy numbers and galactose concentrations. **a**, Histogram of GFP measurements for copy numbers from $M = 1$ to $M = 5$ above the saturation ($gal = 1.2\%$). **b**, The collapse of GFP distributions under the transformation $F \rightarrow F/M$, $P(F) \rightarrow MP(F/M)$ implies an extrinsic source of variability. AU, arbitrary units. **c**, Induction curves for copy numbers from $M = 1$ to $M = 5$. **d**, Collapse of the induction curves implies that transcription from each promoter is independent. **e**, Standard deviations of GFP corresponding to induction curves. **f**, The collapse of the standard deviation implies an extrinsic source of variability. **g**, The collapse of the coefficient of variation for different copy number implies an extrinsic source of variability. **h**, Lack of collapse implies that the variability is not of intrinsic origin.

[2] D. Volfson, J. Marciniak, W. J. Blake, N. Ostroff, L. S. Tsimring, and J. Hasty, Nature **439**, 861 (2006).

- Self-regulating gene:

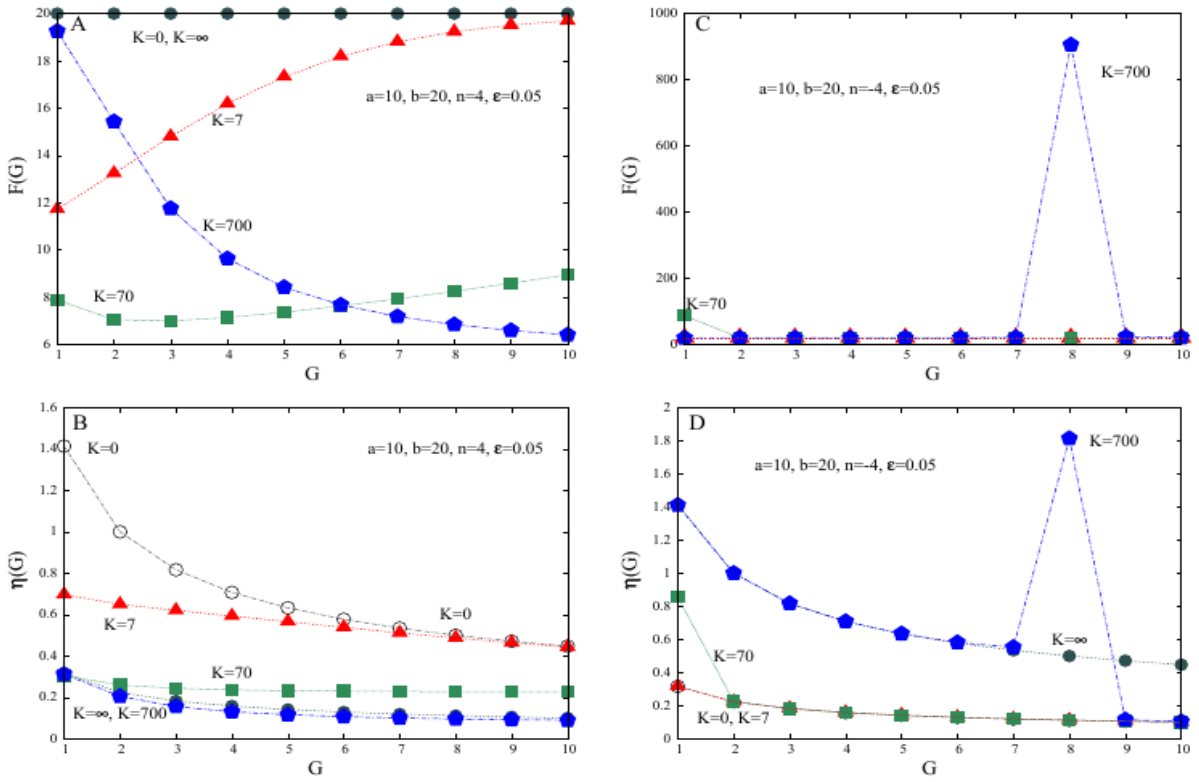
  Mean does not depend linearly on the number of gene copies G.



Average protein number may depend on gene copy number in a nonlinear manner in self-regulating genes. A: Negative auto-regulation ($n = 4$). B: Positive auto-regulation ($n = -4$). The abrupt increase for $K = 700$ and $G = 8$ is due to the transition of the protein number distribution through bimodality Feedback strength parameter $K = 0$ (empty circles), $K = 7$ (triangles), $K = 70$ (squares), $K = 700$ (pentagons), and $K = \infty$ (full circles). Maximum mean burst frequency $a = 10$. Mean burst size $b = 20$. Leakage $\epsilon = 0.05$. Lines provide guide for the eye only.

J. Jędrak and A. Ochab-Marcinek, Journal of Theoretical Biology **408**, 222 (2016).

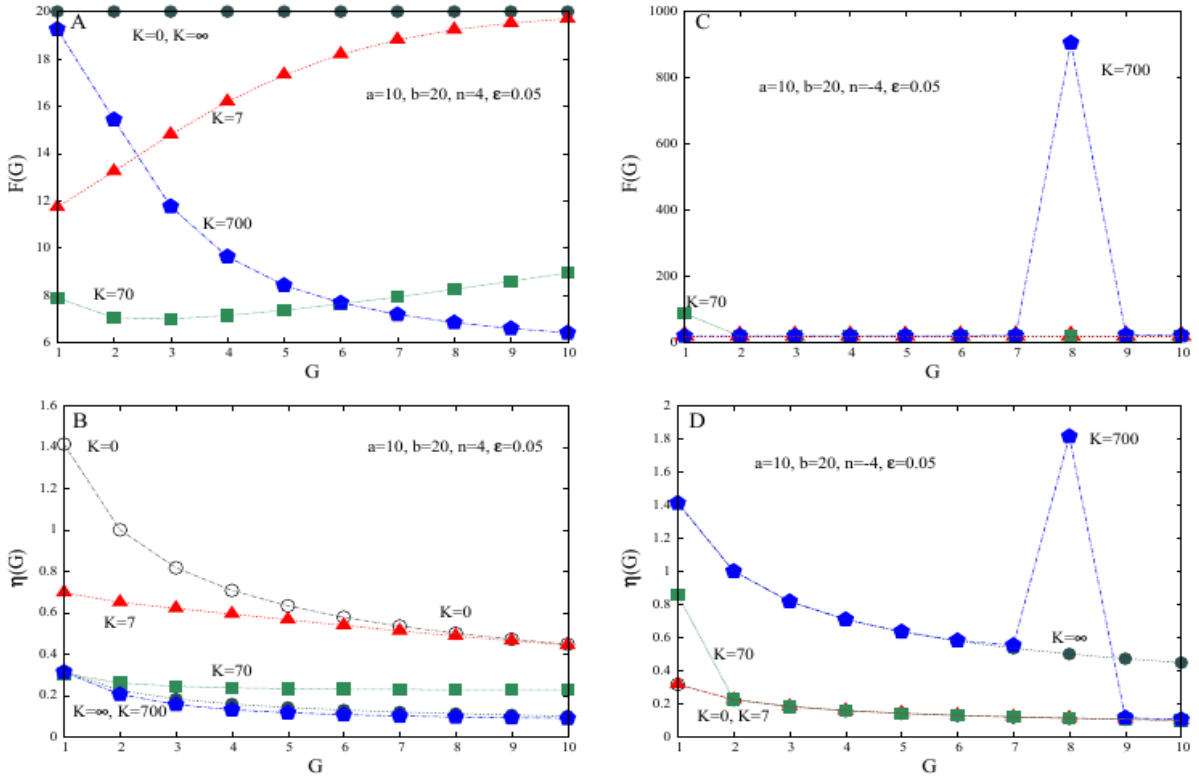
Self-regulating gene: Fano Factor and CV vary in a different manner as G is varied.



In self-regulating genes, Fano factor and by coefficient of variation may depend on gene copy number in a qualitatively different manner. A, B: Negative auto-regulation, $n = 4$. A: Depending on the feedback strength parameter K, Fano factor $F = \sigma^2/\langle x \rangle$ may both decrease, increase or vary in a non-monotonous manner as gene copy number G is varied. B: Coefficient of variation $\eta = \sigma/\langle x \rangle$ is a monotonically decreasing function of gene copy number G. C, D: Positive auto-regulation, $n = -4$. Here, for $K = 700$, Fano factor $F(G)$ has just one maximum (C), whereas the coefficient of variation $\eta(G)$ has two clear maxima (D). The sharp maximum for $K = 700$ and $G = 8$ is due to the transition of the protein number distribution through bimodality. In absence of gene regulation ($K = 0$ and $K = \infty$), $F = b$ and $\eta \sim G^{-1/2}$. For negatively self-regulating genes, $F(G) < b$ and for positive auto-regulation, $F(G) > b$.

J. Jędrak and A. Ochab-Marcinek, Journal of Theoretical Biology **408**, 222 (2016).

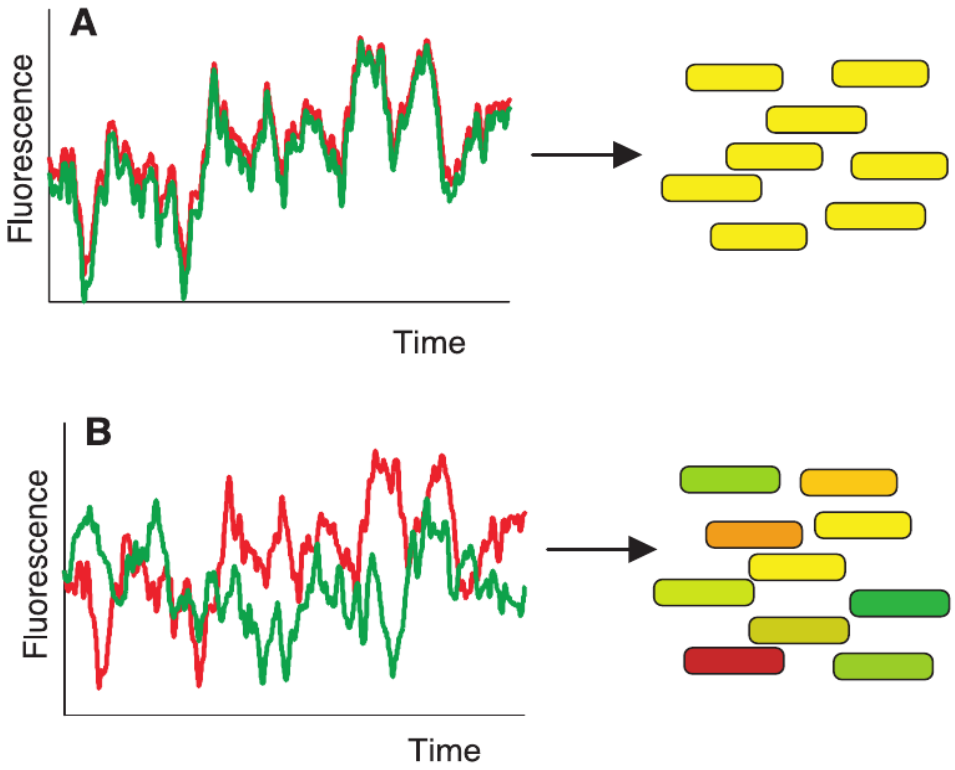
Self-regulating gene: Fano Factor and CV vary in a different manner as G is varied.

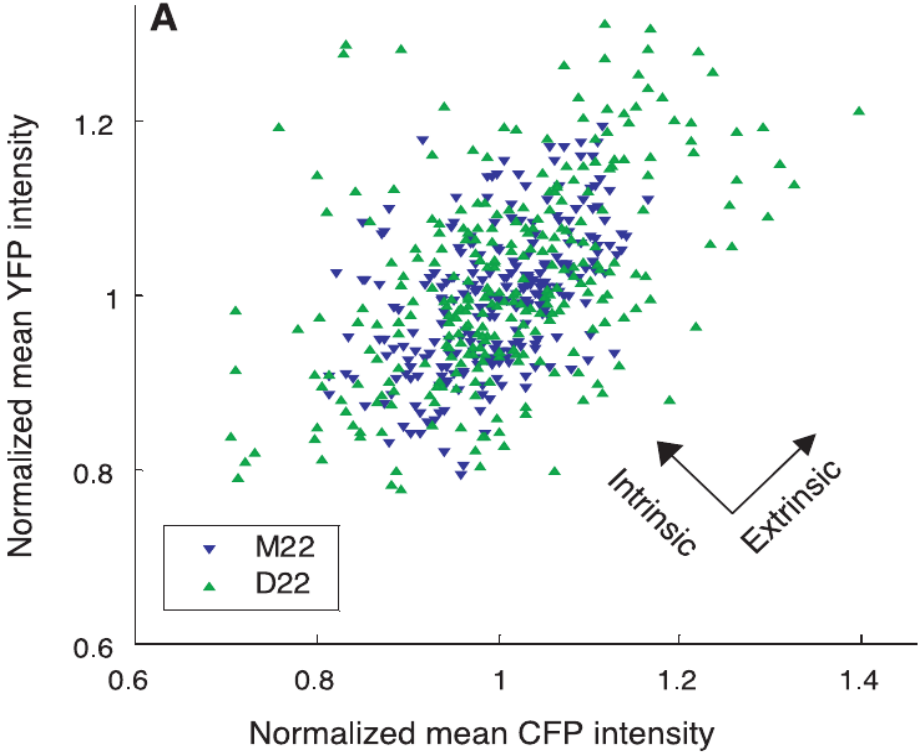


This demonstrates that experimental assessments of the influence of gene expression noise on cell fitness may be ambiguous because they are dependent on the particular function used to quantify noise.

J. Jędrak and A. Ochab-Marcinek, Journal of Theoretical Biology **408**, 222 (2016).

2-color assay to determine intrinsic and extrinsic contributions to gene expression noise

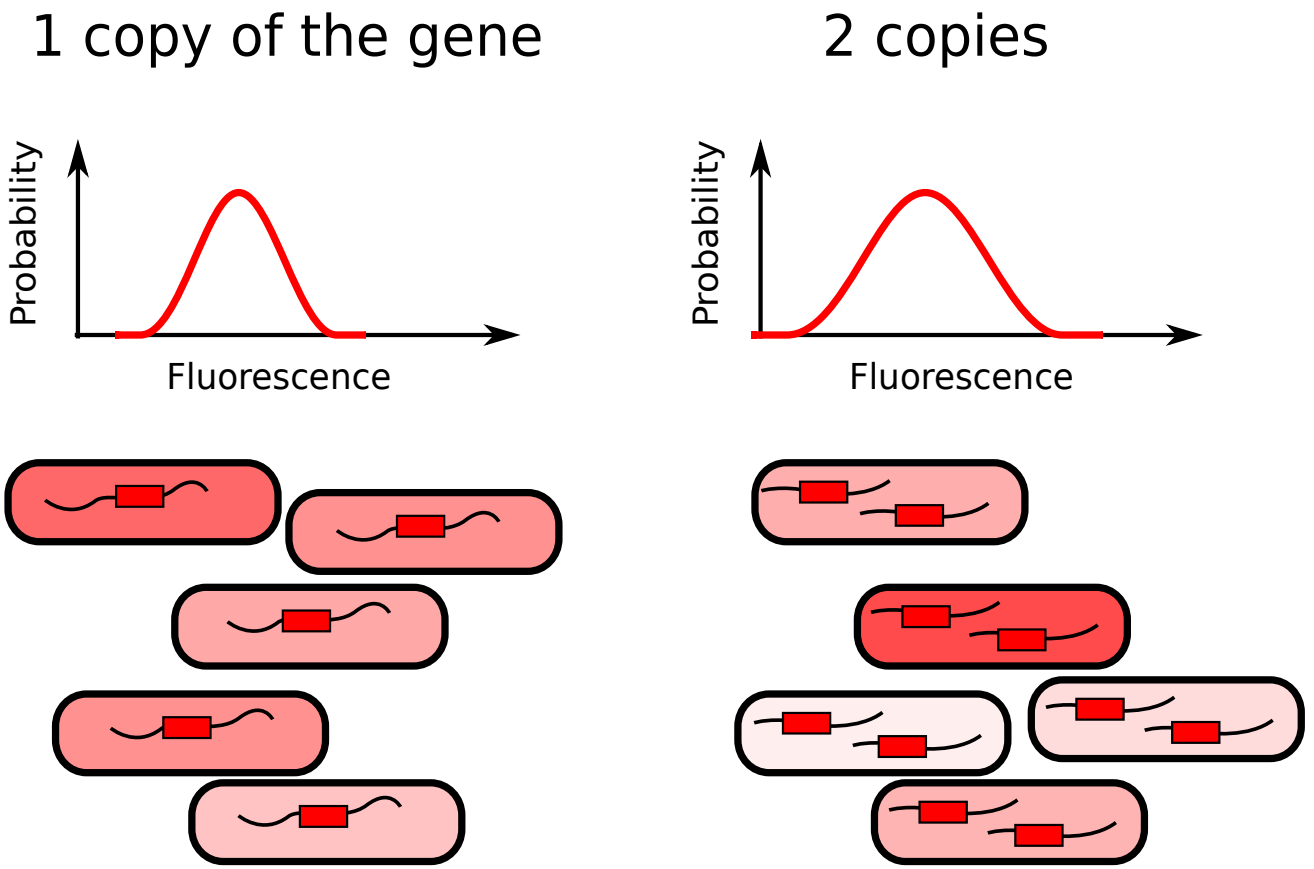




Problem: Equivalence between the two fluorescent proteins

[3] M. B. Elowitz, A. J. Levine, E. D. Siggia, and P. S. Swain, Science **297**, 1183 (2002).

1-color assay



Stewart-Ornstein et al:
Compare the variability in gene expression between these two cell populations

[4] J. Stewart-Ornstein, J. S. Weissman, and H. El-Samad, Molecular cell **45**, 483 (2012).

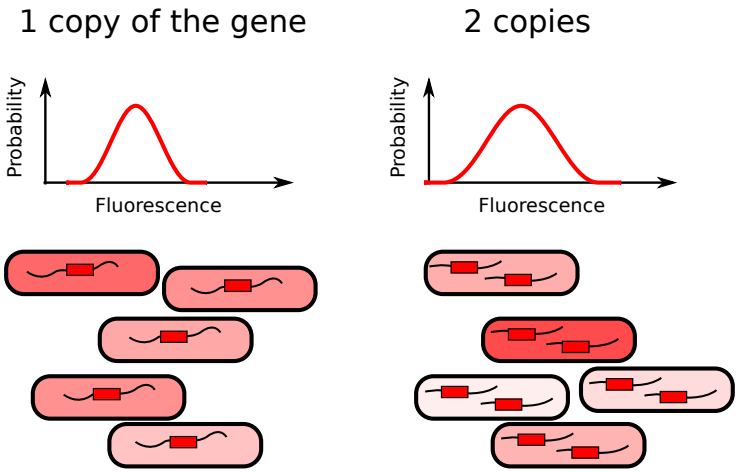
1-color assay

Extrinsic noise was calculated from the one- and two-FP strain measurements using the following formula:

$$Var(a_1 + a_2) = Var(a_1) + Var(a_2) + 2^{*}Cov(a_1, a_2)$$
$$Cov(a_1, a_2) = [Var(a_1 + a_2) - 2^{*}Var(a_1)]/2$$

Here, a_1 and a_2 are indistinguishable alleles of the same gene. $Var(a_1) = Var(a_2)$ is measured in the one-FP strains. $Var(a_1 + a_2)$ is measured in the two-FP strains. Extrinsic noise is then given by the normalized covariance

$$\sqrt{\frac{Cov(a_1, a_2)}{\bar{a}_1 {}^{*} \bar{a}_2}} = Noise_{ext}$$



[4] J. Stewart-Ornstein, J. S. Weissman, and H. El-Samad, Molecular cell **45**, 483 (2012).

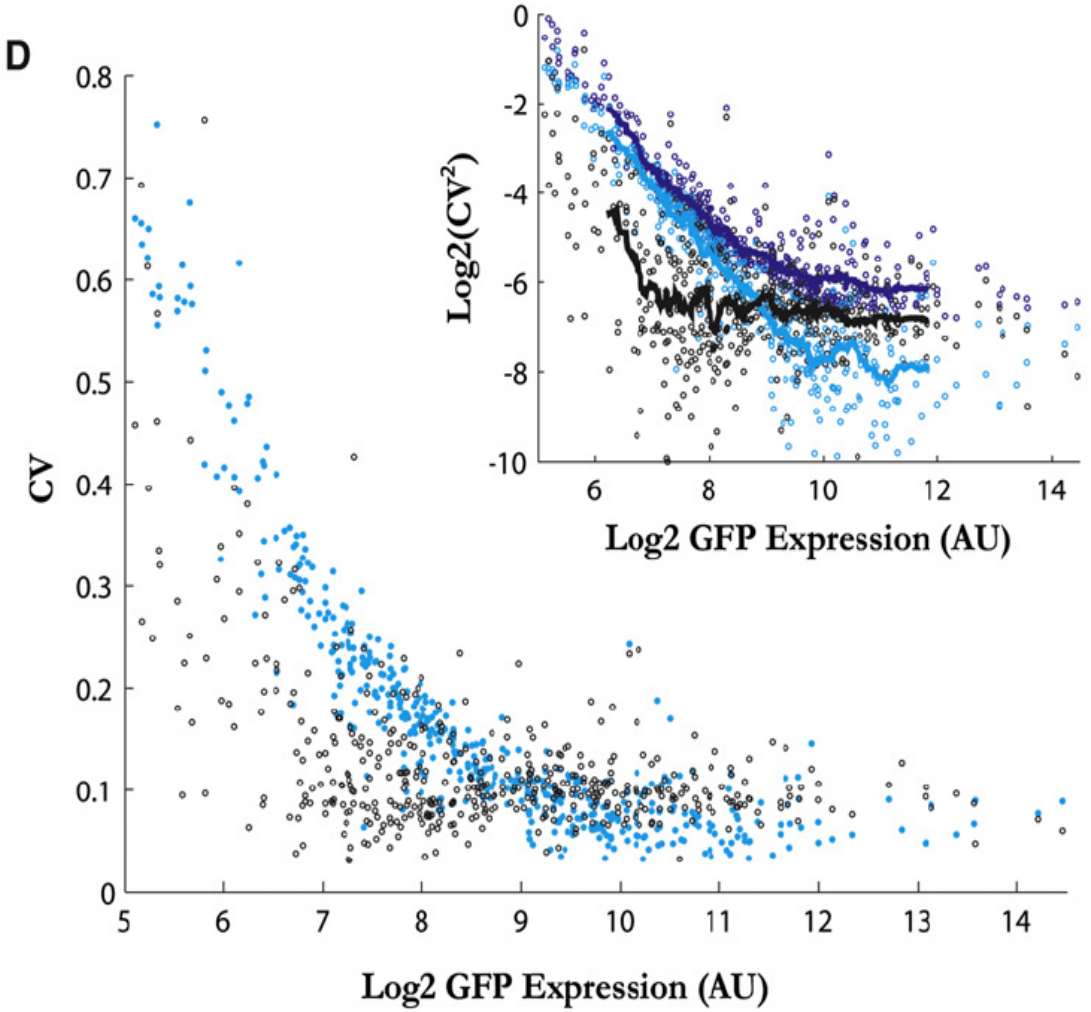
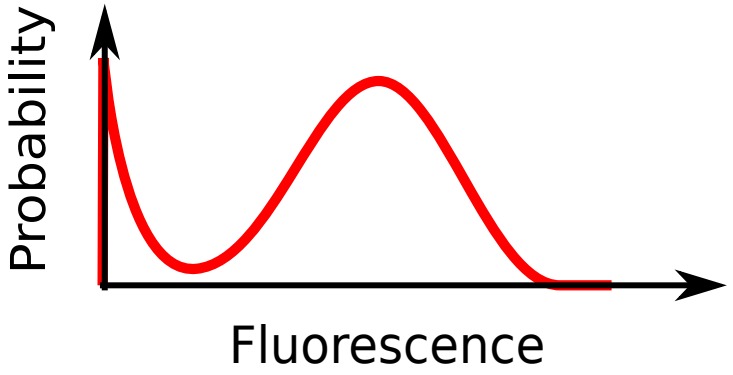


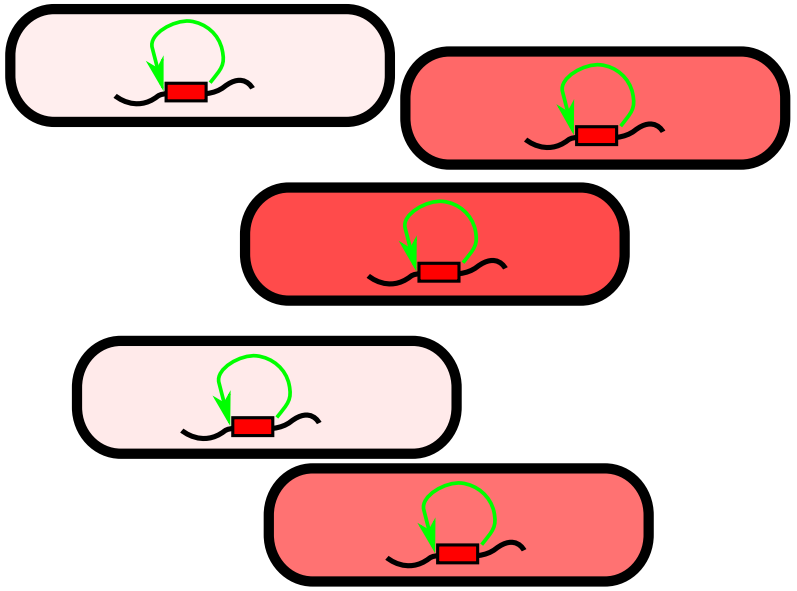
Figure 1. Extrinsic Versus Intrinsic Noise Decomposition across the Proteome

(D) Intrinsic (cyan) and extrinsic(black) noise plotted against log2 mean expression for 465 genes. Inset: log2(CV^2) plotted against log2(mean), running means (smoothing window of 30) for intrinsic(cyan), extrinsic(black), total (dark blue) noise.

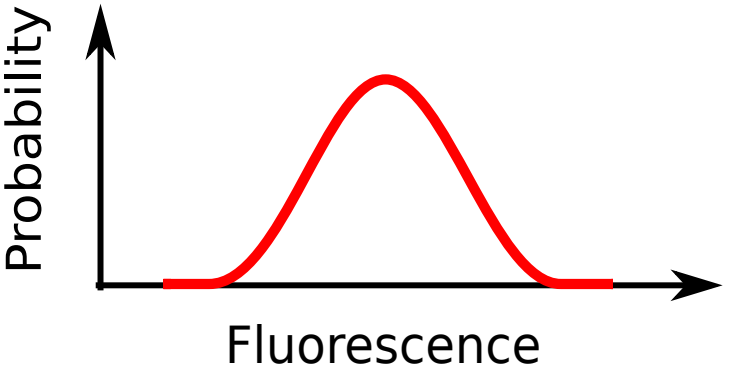
[4] J. Stewart-Ornstein, J. S. Weissman, and H. El-Samad, Molecular cell **45**, 483 (2012).

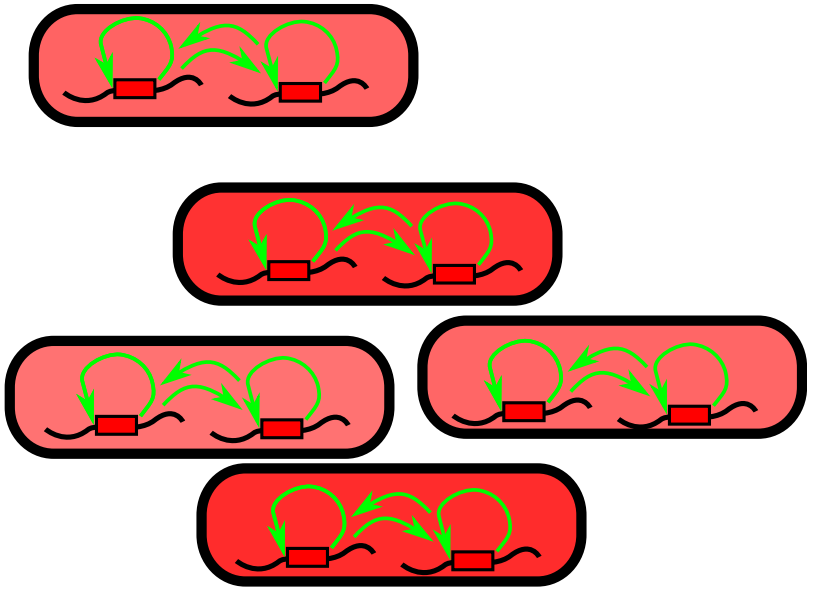
1 copy of the gene



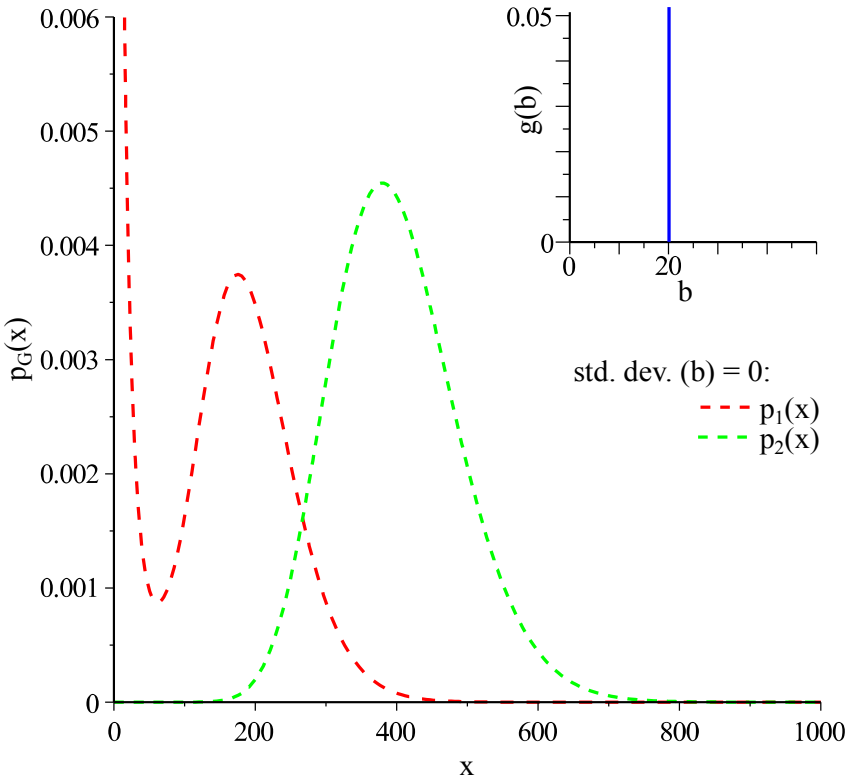


2 copies



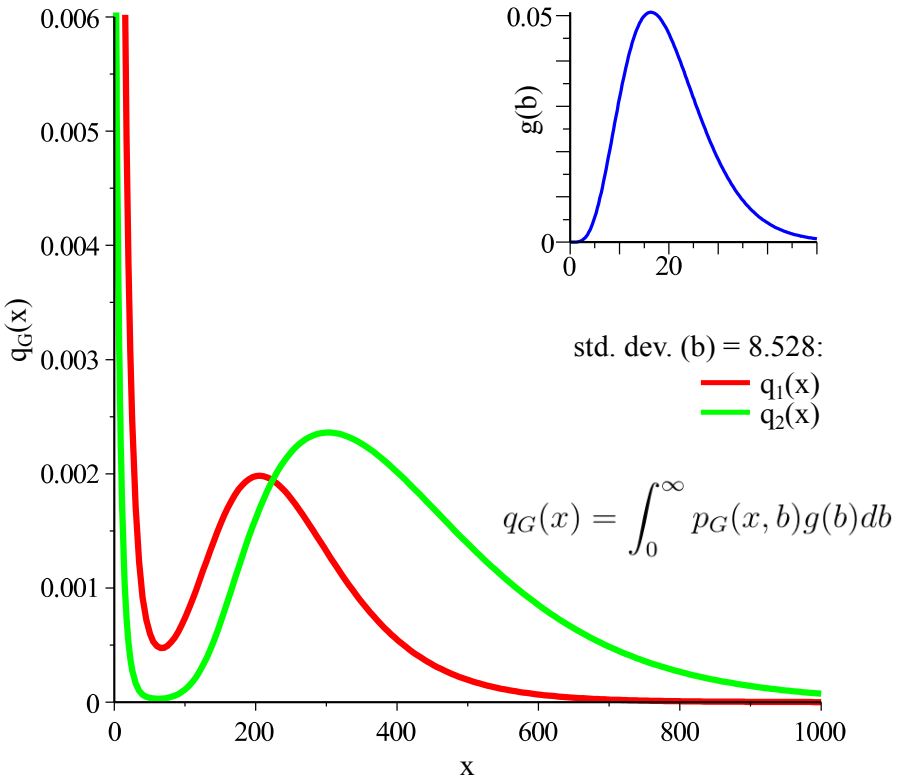


Mean burst size is
the same in each cell



$$\frac{\mathrm{var}(x_1 + x_2) - 2\ \mathrm{var}(x_1)}{\langle x_1^2 \rangle} < 0$$

Mean burst size
differs from cell to cell
(extrinsic noise)



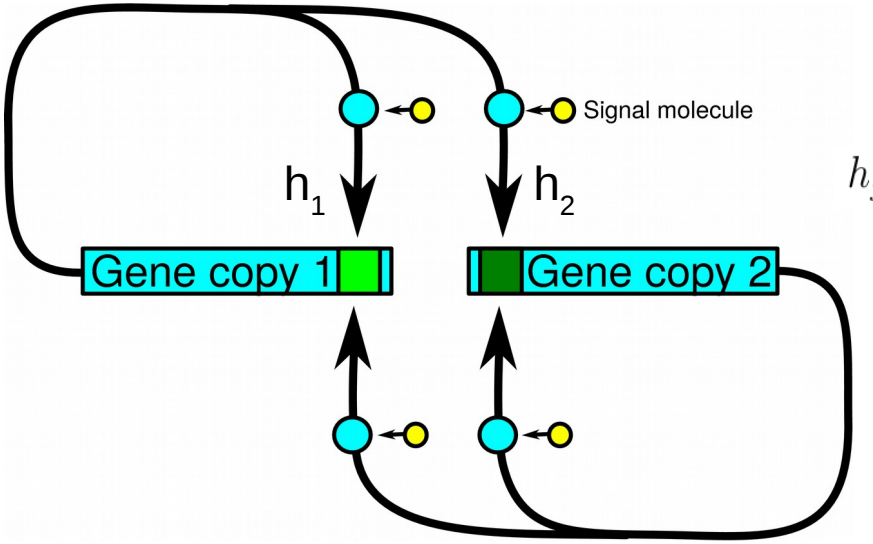
$$q_G(x) = \int_0^\infty p_G(x,b)g(b)db$$

$$\frac{\mathrm{var}(x_1 + x_2) - 2\ \mathrm{var}(x_1)}{\langle x_1^2 \rangle} = 0$$

If interpreted according to Stewart-Ornstein et al., this would mean that extrinsic noise is zero, which is not true!

J. Jędrak and A. Ochab-Marcinek, Journal of Theoretical Biology **408**, 222 (2016).

Two non-identical gene copies (imperfect duplication)

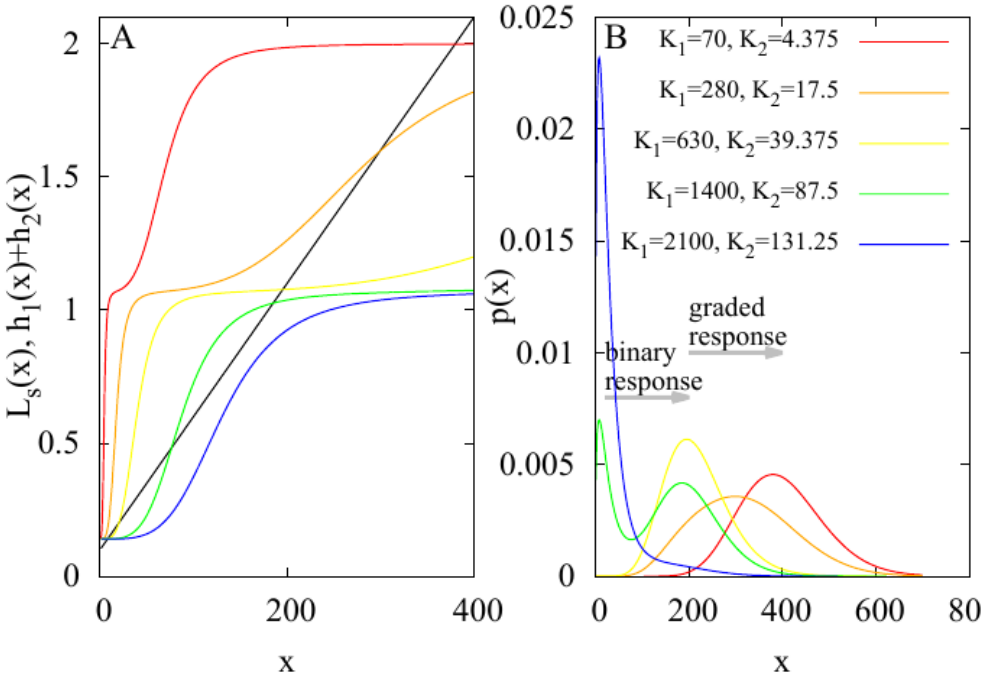


$$h_j(x) = (1-\epsilon_j)H_j(x) + \epsilon_j$$

$$p(x) = Ae^{-x/b}x^{a_1+a_2-1}[H_1(x)]^{\frac{a_1(1-\epsilon_1)}{n_1}}[H_2(x)]^{\frac{a_2(1-\epsilon_2)}{n_2}}$$

Mixed, binary+graded response to signal

$$\frac{1}{ab}x + \frac{1}{a} = h_1(x) + h_2(x)$$



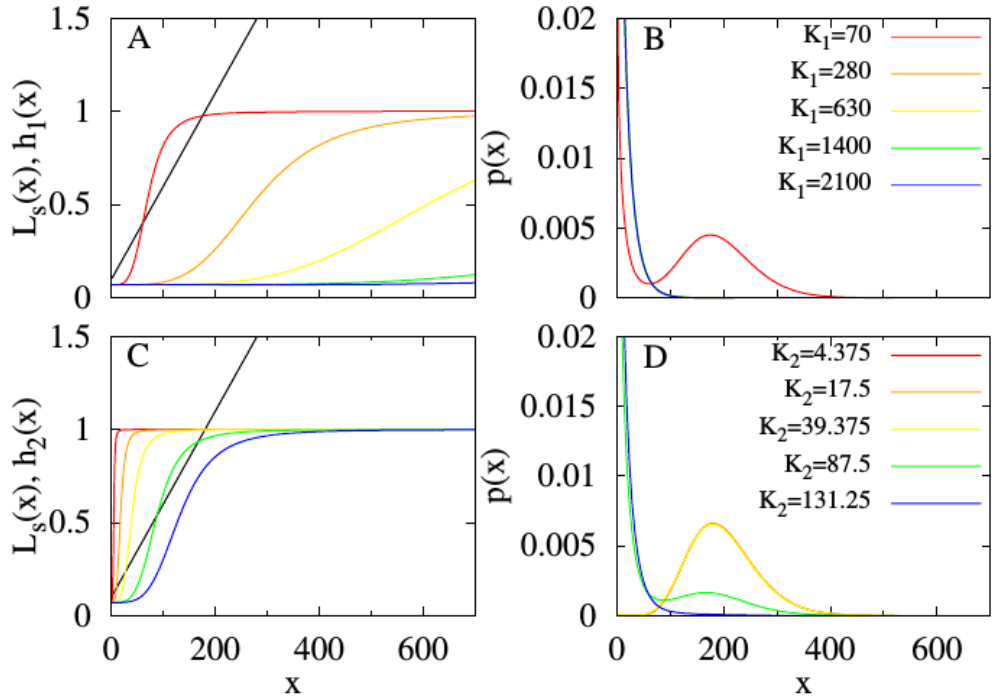


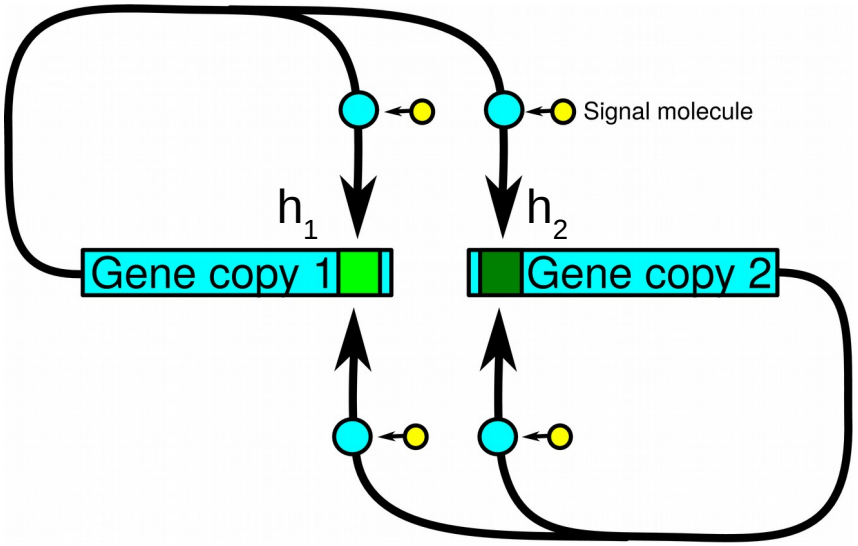
FIG. 1 When two positively self-regulating gene copies have different sensitivities to TF, the geometric construction (A) may predict a mixed, binary+graded, response (B). Binary response is seen for the distribution peaks in the range $0 < x < 200$, and graded response for $200 < x < 400$. Parameters: $n = -4$, $a = 10$, $b = 20$, $\epsilon_1 = \epsilon_2 = 0.07$.

FIG. 2 Each of the genes, whose collective behaviour was shown in Fig. 1 has a binary response when present in the cell in a single copy. Parameters are the same as in Fig. 1 In Fig. B, the orange, yellow, green and blue curves overlap. In Fig. D, the yellow, orange and red curves overlap.

J. Jędrak and A. Ochab-Marcinek, Journal of Theoretical Biology **408**, 222 (2016).

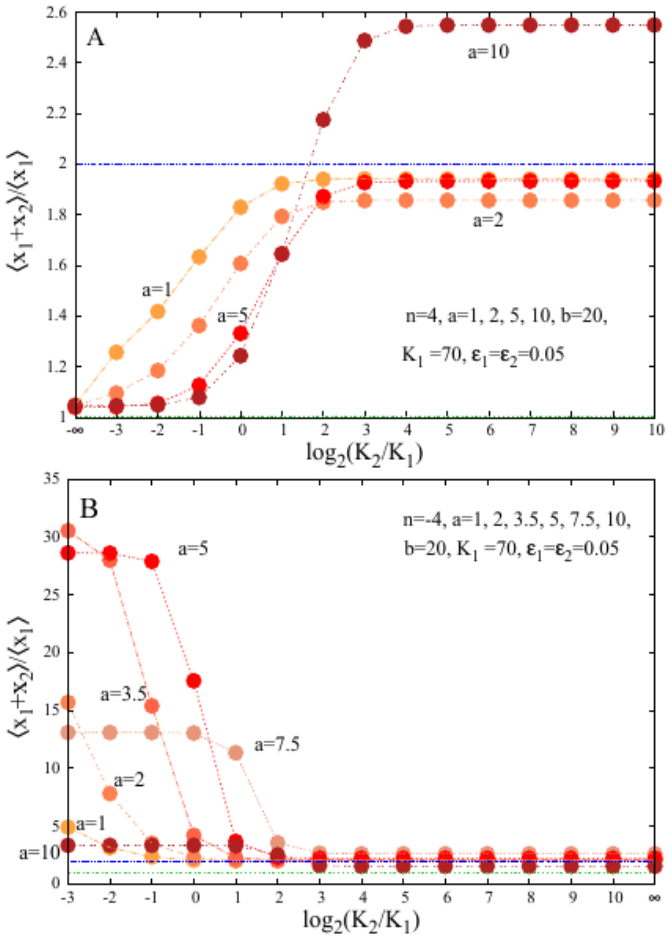
Relative change in average protein number due to gene duplication, depending on how the genes differ in the regulation strength

- $1/K_1$, $1/K_2$: regulation strength



$$h_j(x) = (1-\epsilon_j)H_j(x) + \epsilon_j$$

$$H_j(x) = \left[1 + \left(\frac{x}{K_j}\right)^{n_j}\right]^{-1}$$

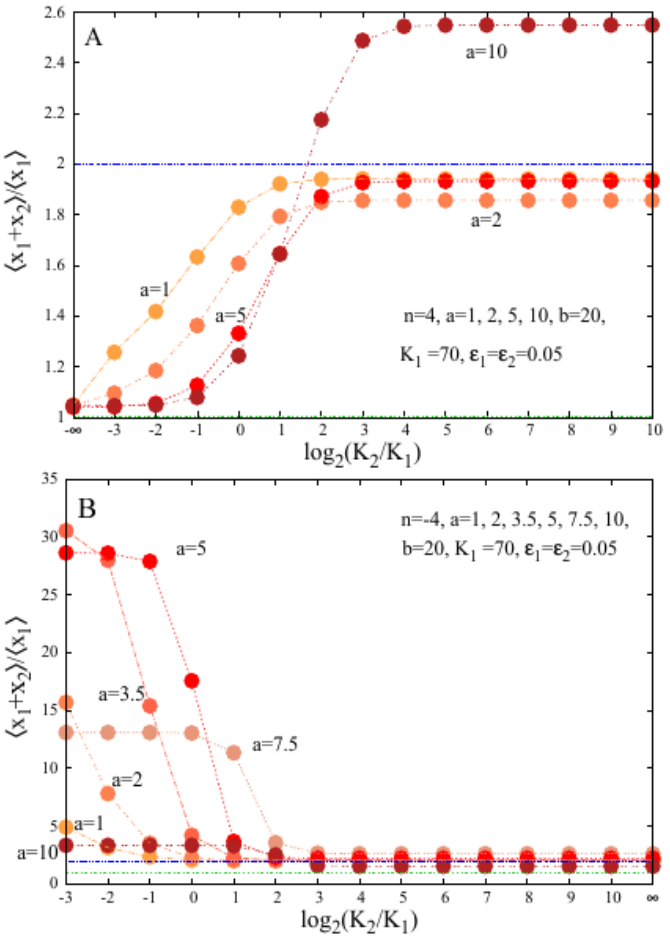


Relative change in the average protein concentration before and after gene duplication, as a function relative affinity of both genes for TF, K_2/K_1. A: Negative auto-regulation, $n = 4$. B: Positive auto-regulation, $n = -4$. Parameters: $b = 20$, $n = 4$, $K_1 = 70$, and $\epsilon_1 = \epsilon_2 = 0.05$. Horizontal dashed lines mark the level of 1 (green) and 2 (blue) for comparison.

J. Jędrak and A. Ochab-Marcinek, Journal of Theoretical Biology **408**, 222 (2016).

Relative change in average protein number due to gene duplication, depending on how the genes differ in the regulation strength

- $1/K_1$, $1/K_2$: regulation strength
- Gene expression is not necessarily twice as high
- Large changes may be detrimental
- Mutants with small changes may survive
- The relative change depends on the burst frequency, a, in a non-monotonic manner

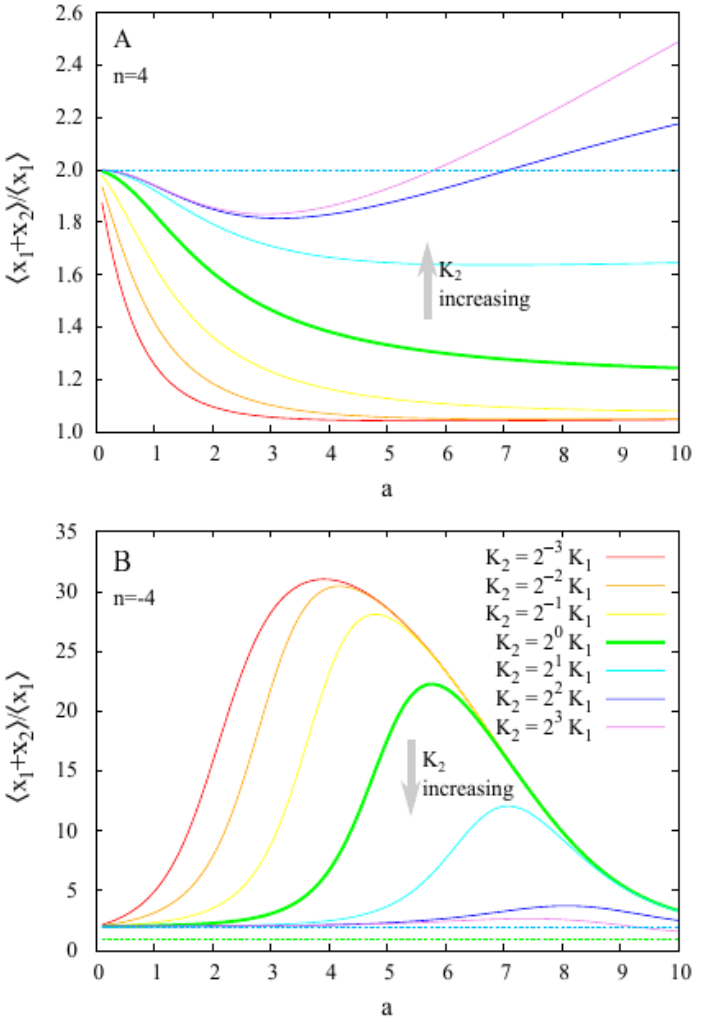


Relative change in the average protein concentration before and after gene duplication, as a function relative affinity of both genes for TF, K_2/K_1. A: Negative auto-regulation, $n = 4$. B: Positive auto-regulation, $n = -4$. Parameters: $b = 20$, $n = 4$, $K_1 = 70$, and $\epsilon_1 = \epsilon_2 = 0.05$. Horizontal dashed lines mark the level of 1 (green) and 2 (blue) for comparison.

J. Jędrak and A. Ochab-Marcinek, Journal of Theoretical Biology **408**, 222 (2016).

Relative change in average protein number due to gene duplication, depending on how the genes differ in the regulation strength

- $1/K_1$, $1/K_2$: regulation strength
- Gene expression is not necessarily twice as high
- Large changes may be detrimental
- Mutants with small changes may survive
- The relative change depends on the burst frequency, a, in a non-monotonic manner

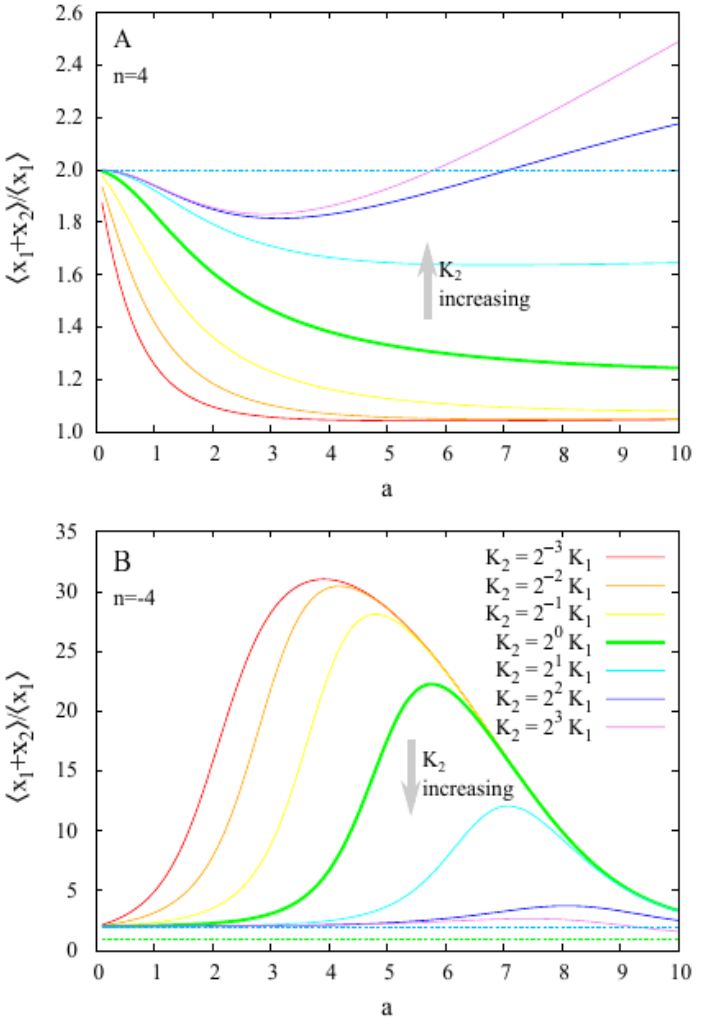


Relative change in the average protein concentration before and after gene duplication, as a function of maximum mean burst frequency a, for various values of K_2/K_1. A: Negative auto-regulation, $n = 4$. B: Positive auto-regulation, $n = -4$.

J. Jędrak and A. Ochab-Marcinek, Journal of Theoretical Biology **408**, 222 (2016).

Relative change in average protein number due to gene duplication, depending on how the genes differ in the regulation strength

- $1/K_1$, $1/K_2$: regulation strength
- Gene expression is not necessarily twice as high
- Large changes may be detrimental
- Mutants with small changes may survive
- The relative change depends on the burst frequency, a, in a non-monotonic manner
- $K_2 > K_1$: Gene 2 has a weaker regulation



Relative change in the average protein concentration before and after gene duplication, as a function of maximum mean burst frequency a, for various values of K_2/K_1. A: Negative auto-regulation, $n = 4$. B: Positive auto-regulation, $n = -4$.

J. Jędrak and A. Ochab-Marcinek, Journal of Theoretical Biology **408**, 222 (2016).

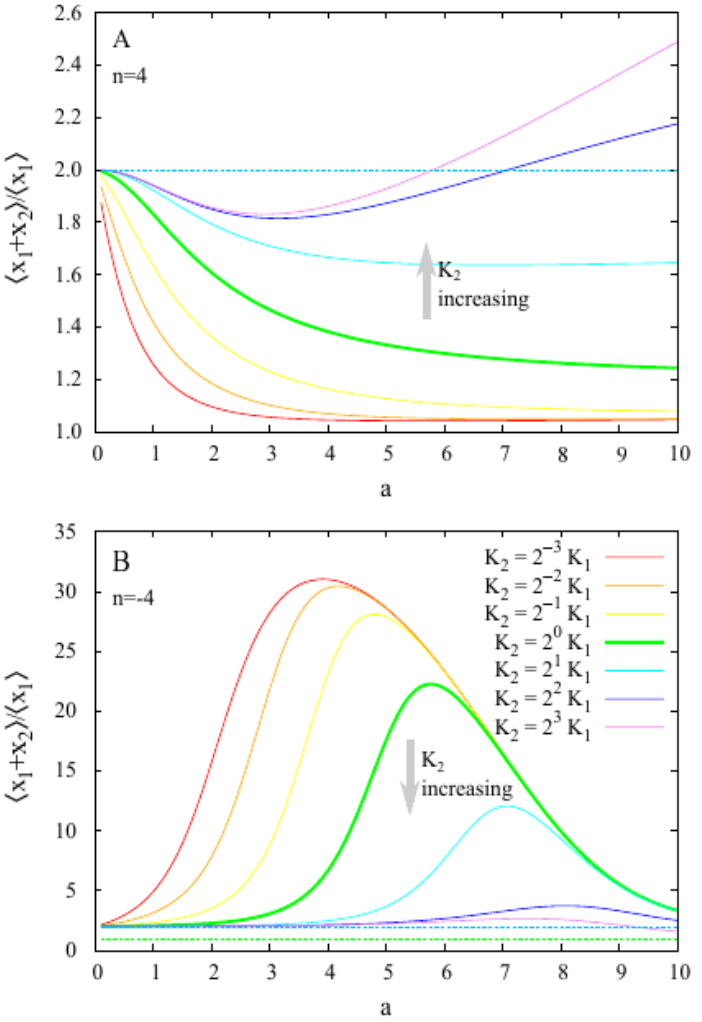
Relative change in average protein number due to gene duplication, depending on how the genes differ in the regulation strength

Negative autoregulation:

- Gene 2 has a stronger regulation than Gene 1 : Accumulation of gene duplications is more probable at larger *a* (smaller noise)
- Gene 2 has a weaker regulation than Gene 1 : Accumulation of gene duplications may be most probable at an optimal *a*

Positive autoregulation:

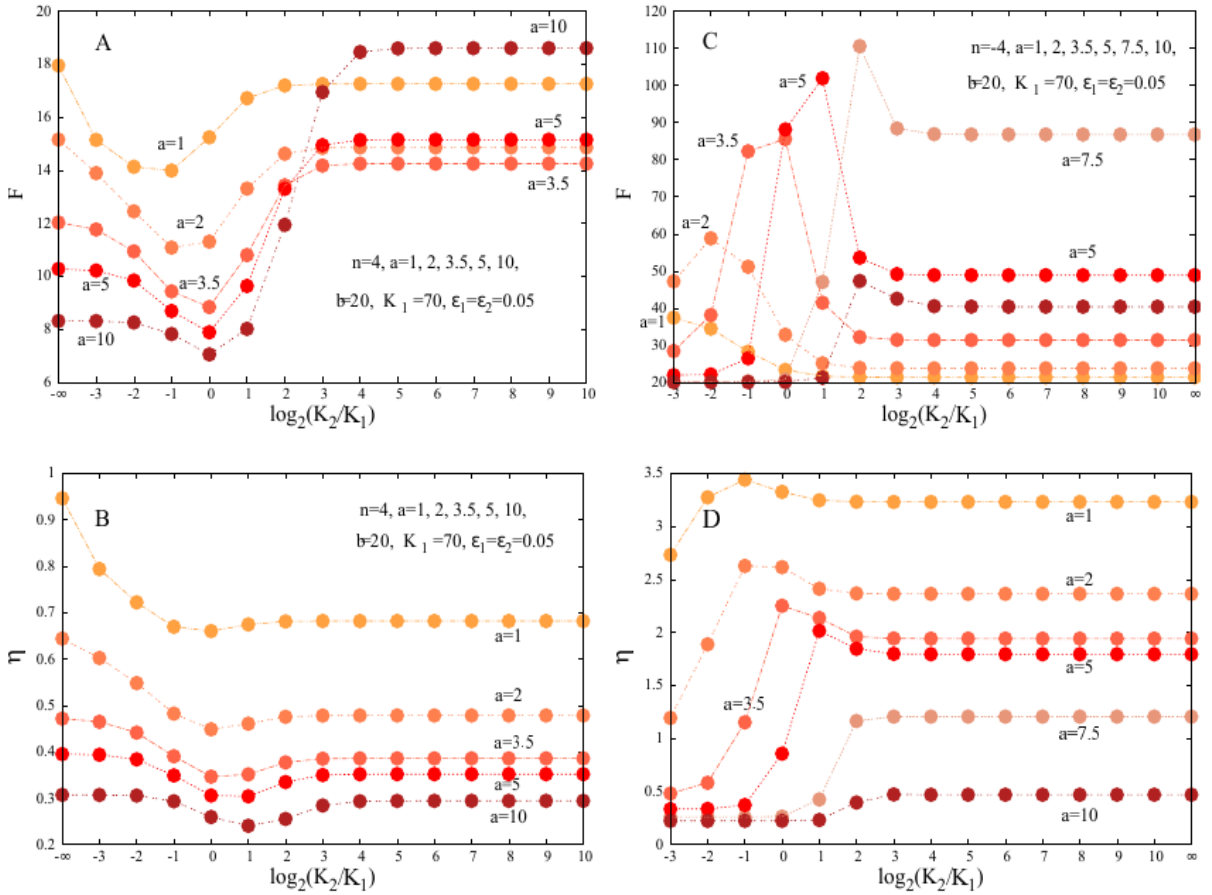
- Intermediate values of *a* give the greatest change in gene expression (one gene is uninduced, but two genes induce each other)
- Accumulation of gene duplications more probable for very small or very large *a*



Relative change in the average protein concentration before and after gene duplication, as a function of maximum mean burst frequency a, for various values of K_2/K_1. A: Negative auto-regulation, $n = 4$. B: Positive auto-regulation, $n = -4$.

J. Jędrak and A. Ochab-Marcinek, Journal of Theoretical Biology **408**, 222 (2016).

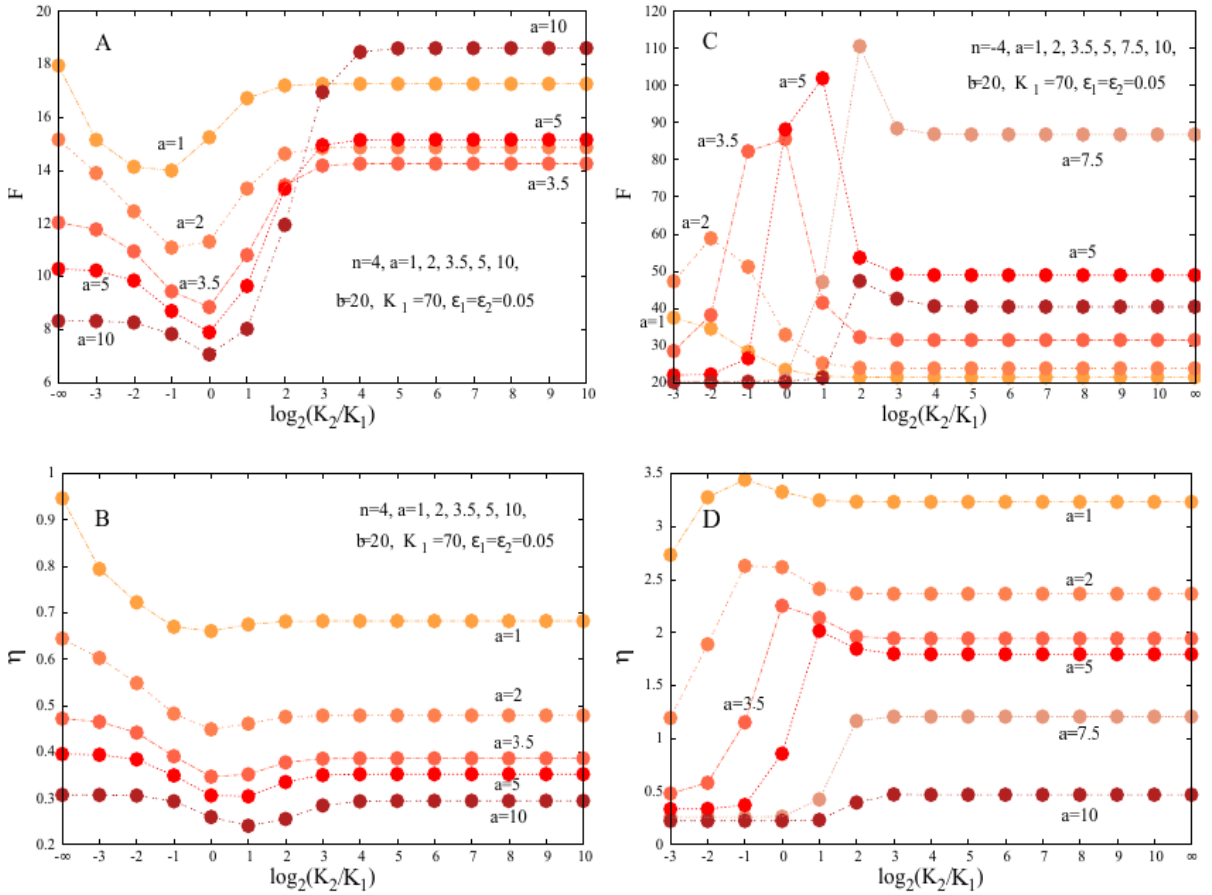
Fano factor and CV vary in a different manner as the relative difference in regulation strengths is varied



Two non-equivalent copies of a negatively (A,B) and positively (C,D) self-regulating gene: Different measures of noise, Fano factor F and coefficient of variation η, may show differences in their behaviour as functions of the relative sensitivity K_2/K_1 of both promoters to auto-regulation. For negative auto-regulation, $n = 4$ (A,B), the positions and depth of minima are different for F and η. For positive auto-regulation, $n = -4$ (C,D), the maxima of both measures of noise roughly correspond to the transition through bimodal distributions. The exact positions and height of the maxima are, however, different for F and η. Additionally, for both positive and negative auto-regulation, F varies non-monotonically with a, whereas the dependence of η on a is monotonic. Parameters: $b = 20$, $K_1 = 70$, $\epsilon_1 = \epsilon_2 = 0.05$.

J. Jędrak and A. Ochab-Marcinek, Journal of Theoretical Biology **408**, 222 (2016).

Fano factor and CV vary in a different manner as the relative difference in regulation strengths is varied



This demonstrates that experimental assessments of the influence of gene expression noise on cell fitness may be ambiguous because they are dependent on the particular function used to quantify noise.

J. Jędrak and A. Ochab-Marcinek, Journal of Theoretical Biology **408**, 222 (2016).

By the way...

PHYSICAL REVIEW E **94**, 032401 (2016)

Time-dependent solutions for a stochastic model of gene expression with molecule production in the form of a compound Poisson process

Jakub Jędrak* and Anna Ochab-Marcinek

Single gene, not regulated

Single gene, not regulated

Time evolution of cumulants of $p(x,t)$

$$\kappa_r(t) = \kappa_r(0)e^{-r\gamma t} + a(1 - e^{-r\gamma t})\frac{m_r}{r}$$

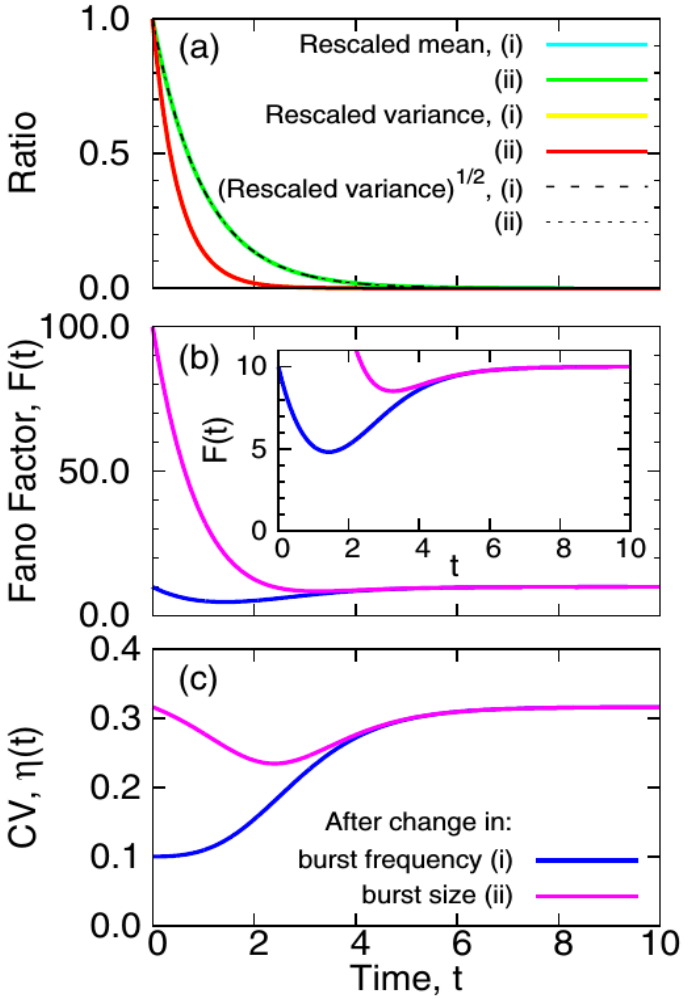


FIG. 3. Different measures of noise have different transient behaviors in time after an abrupt reduction of the mean burst frequency a (i), or the mean burst size b (ii). The mean protein concentration $\langle x \rangle = ab$ was decreased from 10^3 to 10^2. (a) Fractional change of mean, $K_1(\tau)$, fractional change of variance, $K_2(\tau)$, and its square root, $K_2(\tau)^{1/2}$. The curves for the cases (i) and (ii) overlap. (b) Fano factor $F(\tau)$, inset: zoom to show the minima. (c) Coefficient of variation, $\eta(\tau)$.

J. Jędrak and A. Ochab-Marcinek, Physical Review E **94**, 032401 (2016).

The gene was at a certain level of expression

Then, the expression is reduced:

- By the reduction of mean burst size

  or

- By the reduction of mean burst frequency

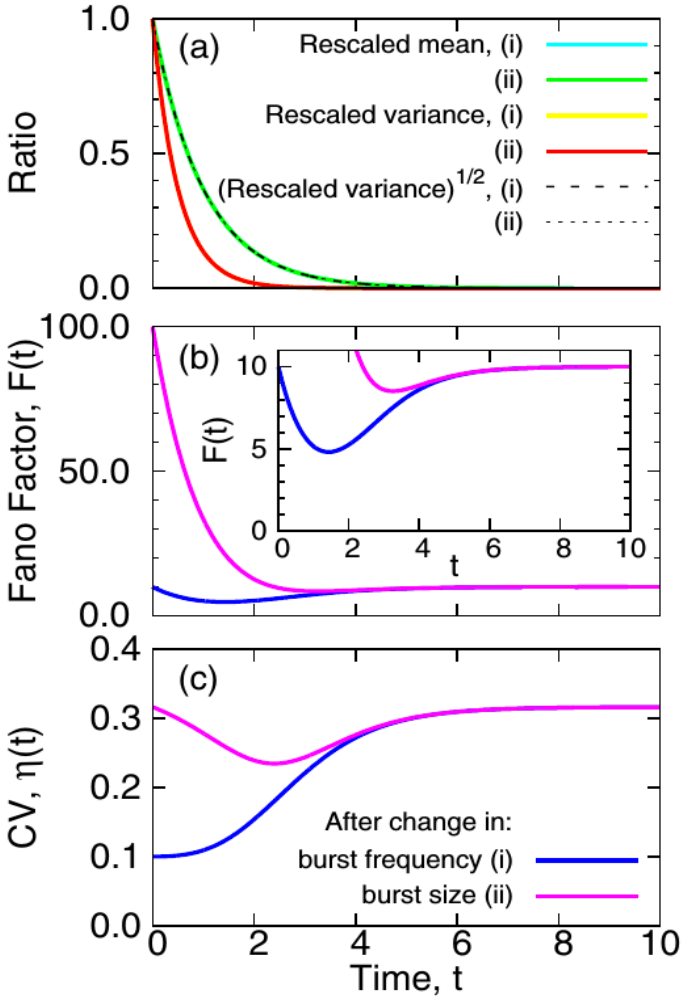


FIG. 3. Different measures of noise have different transient behaviors in time after an abrupt reduction of the mean burst frequency a (i), or the mean burst size b (ii). The mean protein concentration $\langle x \rangle = ab$ was decreased from 10^3 to 10^2. (a) Fractional change of mean, $K_1(\tau)$, fractional change of variance, $K_2(\tau)$, and its square root, $K_2(\tau)^{1/2}$. The curves for the cases (i) and (ii) overlap. (b) Fano factor $F(\tau)$, inset: zoom to show the minima. (c) Coefficient of variation, $\eta(\tau)$.

J. Jędrak and A. Ochab-Marcinek, Physical Review E **94**, 032401 (2016).

Top figure:

Relative change in mean/variance:

$$K_r(t) \equiv \frac{\kappa_r(t) - \kappa_r(\infty)}{\kappa_r(0) - \kappa_r(\infty)} = e^{-r\gamma t}$$

where cumulants are:

$$\kappa_r(t) = \kappa_r(0)e^{-r\gamma t} + a(1 - e^{-r\gamma t})\frac{m_r}{r}$$

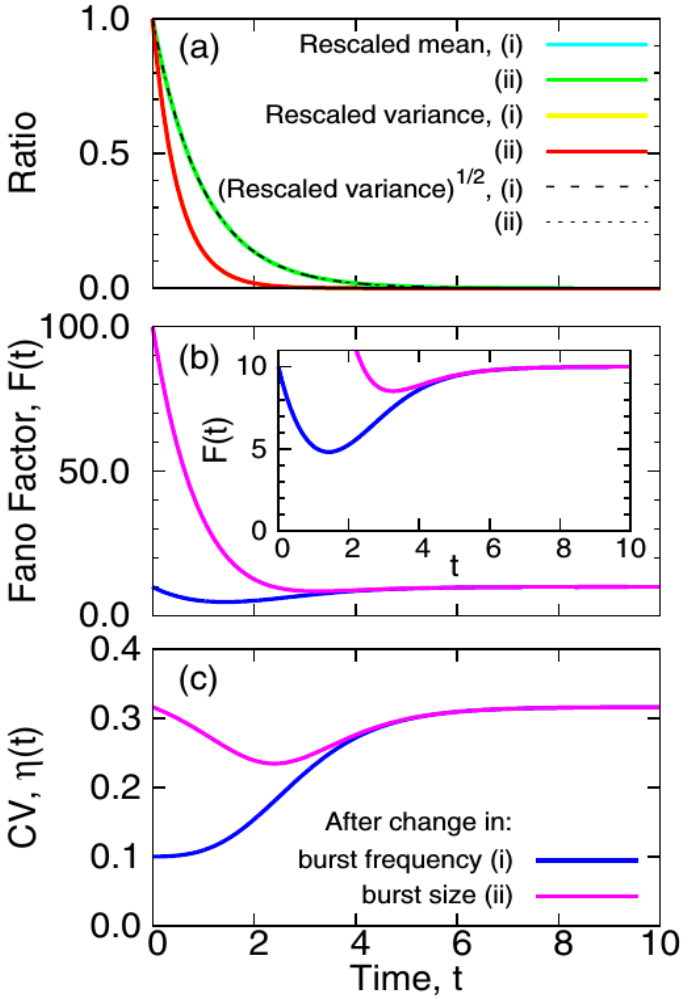


FIG. 3. Different measures of noise have different transient behaviors in time after an abrupt reduction of the mean burst frequency a (i), or the mean burst size b (ii). The mean protein concentration $\langle x \rangle = ab$ was decreased from 10^3 to 10^2. (a) Fractional change of mean, $K_1(\tau)$, fractional change of variance, $K_2(\tau)$, and its square root, $K_2(\tau)^{1/2}$. The curves for the cases (i) and (ii) overlap. (b) Fano factor $F(\tau)$, inset: zoom to show the minima. (c) Coefficient of variation, $\eta(\tau)$.

J. Jędrak and A. Ochab-Marcinek, Physical Review E **94**, 032401 (2016).

The gene was at a certain level of expression

Then, the expression is reduced:

- By the reduction of mean burst size

  or

- By the reduction of mean burst frequency

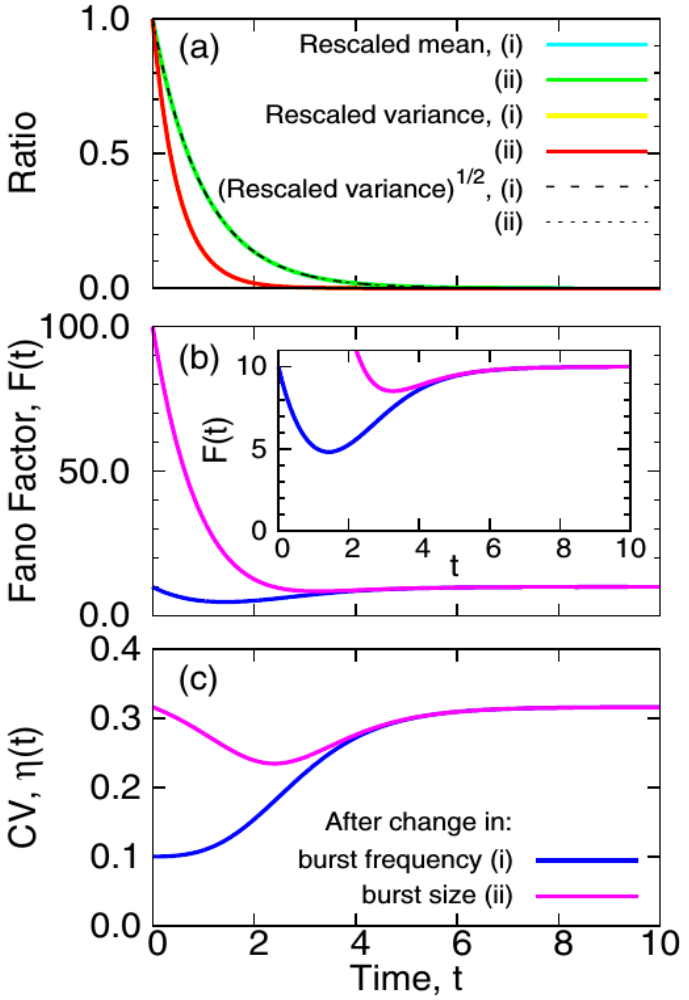


FIG. 3. Different measures of noise have different transient behaviors in time after an abrupt reduction of the mean burst frequency a (i), or the mean burst size b (ii). The mean protein concentration $\langle x \rangle = ab$ was decreased from 10^3 to 10^2. (a) Fractional change of mean, $K_1(\tau)$, fractional change of variance, $K_2(\tau)$, and its square root, $K_2(\tau)^{1/2}$. The curves for the cases (i) and (ii) overlap. (b) Fano factor $F(\tau)$, inset: zoom to show the minima. (c) Coefficient of variation, $\eta(\tau)$.

J. Jędrak and A. Ochab-Marcinek, Physical Review E **94**, 032401 (2016).

Single gene, not regulated

- Again, Fano factor and CV vary in a different manner

- If we propose any biological hypothesis about optimization with respect to noise, we need to justify why the specific measure of noise has been chosen.

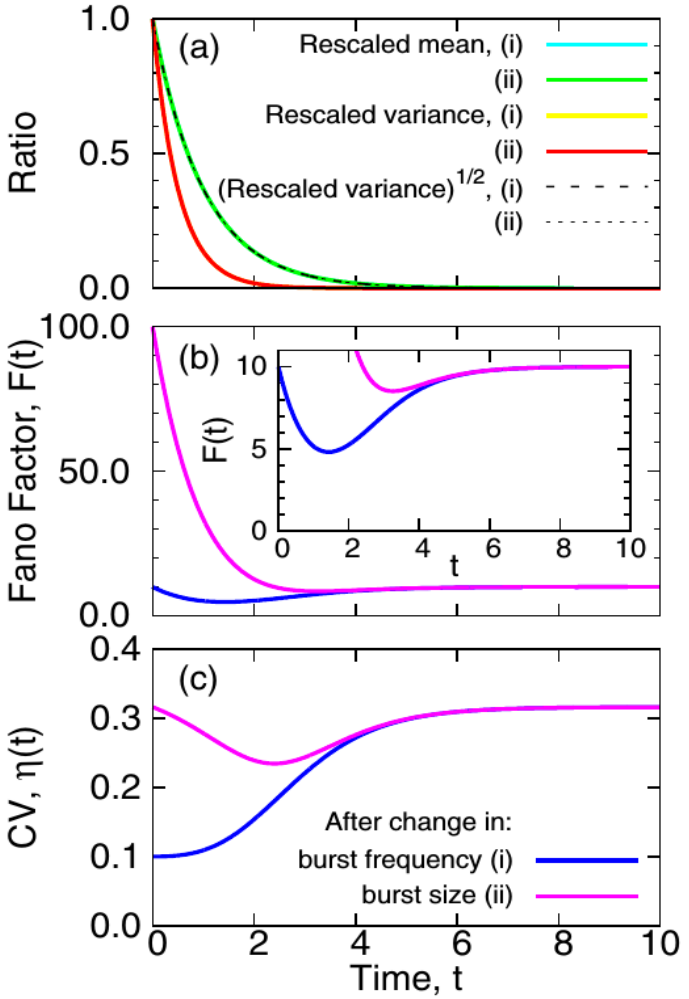


FIG. 3. Different measures of noise have different transient behaviors in time after an abrupt reduction of the mean burst frequency a (i), or the mean burst size b (ii). The mean protein concentration $\langle x \rangle = ab$ was decreased from 10^3 to 10^2. (a) Fractional change of mean, $K_1(\tau)$, fractional change of variance, $K_2(\tau)$, and its square root, $K_2(\tau)^{1/2}$. The curves for the cases (i) and (ii) overlap. (b) Fano factor $F(\tau)$, inset: zoom to show the minima. (c) Coefficient of variation, $\eta(\tau)$.

J. Jędrak and A. Ochab-Marcinek, Physical Review E **94**, 032401 (2016).

Conclusions

- Stochastic model of an autoregulated gene, present in multiple copies.
- One-reporter assay may not measure correctly the extrinsic noise in self-regulated genes.
- Imperfect duplication of auto-activated gene: mixed, binary+graded response possible.
- Accumulation of gene duplications: non-trivial dependence on inherent noisiness of gene.
- Measurement of noise using Fano Factor or coefficient of variation is ambiguous.

Open problem

- If the amount of noise in gene expression is optimized by the evolution, how is it measured?