Deciphering noise amplification and reduction in open chemical reaction networks

Fabrizio Pucci, Marianne Rooman

Department of BioModeling, BioInformatics & BioProcesses,
Department of Theoretical Physics,
Université Libre de Bruxelles, Roosevelt Ave. 50, 1050 Brussels, Belgium;
(e-mail: fapucci@ulb.ac.be, mrooman@ulb.ac.be)

Abstract

The impact of random fluctuations on the dynamical behavior a complex biological systems is a longstanding issue, whose understanding would shed light on the evolutionary pressure that nature imposes on the intrinsic noise levels and would allow rationally designing synthetic networks with controlled noise. Using the Itō stochastic differential equation formalism, we performed both analytic and numerical analyses of several model systems containing different molecular species in contact with the environment and interacting with each other through mass-action kinetics. These systems represent for example biomolecular oligomerization processes, complex-breakage reactions, signaling cascades or metabolic networks. For chemical reaction networks with zero deficiency values, which admit a detailed- or complex-balanced steady state, all molecular species are uncorrelated. The number of molecules of each species follow a Poisson distribution and their Fano factors, which measure the intrinsic noise, are equal to one. Systems with deficiency one have an unbalanced non-equilibrium steady state and a non-zero S-flux, defined as the flux flowing between the complexes multiplied by an adequate stoichiometric coefficient. In this case, the noise on each species is reduced if the flux flows from the species of lowest to highest complexity, and is amplified is the flux goes in the opposite direction. These results are generalized to systems of deficiency two, which possess two independent non-vanishing S-fluxes, and we conjecture that a similar relation holds for higher deficiency systems.

Keywords: Dynamical modeling, Stochastic modeling, Differential equations, Biological systems, Random fluctuations

1. Introduction

The identification and understanding of the principles that guide the modulation of intrinsic noise in biological processes is a major goal of systems biology. Indeed, a wide range of biological phenomena such as the biochemical reactions and the transcription and translation machineries are of a random nature, and

fluctuations play frequently a pivotal role in their dynamics (1; 2; 3; 4). Biological systems appear to have evolved over time to tune the noise level, in some phenomena to reduce and tolerate the fluctuations while in others to utilize the heterogeneity to their advantage (5; 6).

One of the classical examples in which the cell systems use fluctuations to obtain a selective advantage is related to the cellular decision-making processes. Indeed, intrinsic noise can allow the diversification of the phenotype of identical cells that live in the same environmental conditions and thus facilitate the transitions between various cellular states. Multiple examples of the important role of the fluctuations in the cellular decision mechanisms in organisms of different levels of complexity - from virus and bacteria to mammalian cells - have been thoroughly analyzed in the literature (see (3) and references therein).

In contrast, in many other biological systems, stability and robustness criteria basically require the suppression of the fluctuations and a wide series of different mechanisms are used to ensure this attenuation. A simple and common example is the negative feedback loop in gene regulatory networks, in which the protein that is expressed from a given gene inhibits its own transcription (7; 9). This mechanism indeed tends to suppress the noise while reducing the metabolic cost of protein production, and speeds up the rise-times of transcription units (8).

The comprehension of how the modulation of noise is achieved is very important for basically two reasons. The first is of fundamental nature and involves answering open questions about why natural evolution designs specific networks and functional mechanisms and about the role played by fluctuations (10; 11). The second reason concerns the application to synthetic biology with the aim of engineering and assembling biological components into synthetic devices with a controlled level of intrinsic noise (12; 13).

Despite the many valuable advances in the field of the last two decades, the mechanisms employed to amplify or to suppress the fluctuation levels need to be further understood and clarified. Indeed, the huge complexity of biological systems, their dependence on a large number of variables and the system-to-system variability make the unraveling of these issues, whether using experimental or computational approaches, a highly non-trivial task.

More specifically, while the noise control is relatively well understood for small and simple networks (*e.g.* negative or positive feedback loops), it is still far from clear how the fluctuations propagate through more general and complicated networks and what is the link of the network topology and complexity with the noise buffering or amplification. Different investigations address these issues from various perspectives, for example by characterizing the stochastic properties of the chemical reaction networks (CRNs) and studying the propagation of the fluctuations (14; 15; 16). From a physics-oriented perspective, the authors of (17; 18; 19) analyzed the connection between the non-equilibrium thermodynamic properties of the network and the noise level. Finally, it has been shown in (20) that the increase of the network complexity tends to decrease the intrinsic noise and also to reduce the effect of the extrinsic noise for some multistable model systems, whereas a dependence of the noise reduction

or amplification on the system parameters has been found in (21).

In this paper we expand the results presented in (22), where we investigated the relation between the total level of noise in various classes of CRNs with the networks' structural characteristics. This was done by studying systems with different degrees of complexity using the Itō stochastic differential equations formalism, and by fully exploring, both analytically and numerically, the huge parameter space of the models. Moreover, the present computational investigation goes beyond the results that we presented in (22), since we analyzed here not only the total level of noise but also the modulation of the fluctuations for each molecular species involved in the CRNs.

2. Chemical Reaction Networks

In this section we review some of the basic notions of CRN theory in order to set up our conventions; for more details, see for example (24; 25; 26). CRNs are systems of reactions between (bio)chemical species and are characterized by triplets $[\mathcal{S}, \mathcal{C}, \mathcal{R}]$. $\mathcal{S}$ represents the ensemble of all chemical species involved in the network, $\mathcal{C}$ is the set of complexes and $\mathcal{R}$ the ensemble of biochemical reactions. Let us consider for example the network described by:

$$2A_1 \leftrightarrow A_2 \leftrightarrow A_3 + A_1. \tag{1}$$

In this case $\mathcal{S} = \{A_1, A_2, A_3\}$, $\mathcal{C} = \{2A_1, A_2, A_3 + A_1\}$ and $\mathcal{R}$ are the four reactions indicated by arrows. The reaction vectors $\subset \mathbb{R}^3$ are here equal to $[2, -1, 0], [-2, 1, 0], [-1, 1, -1], [1, -1, 1]$, where the entries of the reaction vector i (with $1 \leq i \leq$card$(\mathcal{R})$) are equal to the stoichiometry of the molecular species j (with $1 \leq j \leq$card$(\mathcal{S})$) in the complexes formed or broken by the ith reaction; by convention, a positive sign is associated to the products of a reaction and a negative sign to the reactants. For open systems, in which the molecular species are produced from or degraded to the environment, the environment is not considered as a species but as a complex with vanishing stoichiometry coefficients.

Three main notions have been introduced to characterize the CRNs. The first is the deficiency δ of the network defined as:

$$\delta = \text{card}(\mathcal{C}) - \mathcal{L} - \mathcal{X}, \tag{2}$$

where $\mathcal{L}$ is the number of linkage classes, namely the number of connected components of the CRN, and $\mathcal{X}$ is the dimension of the stoichiometry subspace, namely the rank of the network. For example the network in Eq. (1) contains three complexes, one linkage class and its rank is equal to two, which yields $\delta = 0$.

The second notion is the reversibility of the CRN. A network is said to be *reversible* if for each reaction connecting complex x to y there is an inverse reaction from y to x. The CRN is only *weakly reversible* if the existence of a reaction path from complex x to y implies the existence of a, possibly indirect, path from y to x.

The last notion is the complex balance. A network is complex balanced, if, for each complex y, the sum of the mean reaction rates for the reactions $r \subset \mathcal{R}$ for which y is a reactant complex is equal to the sum of the mean reaction rates for $r' \subset \mathcal{R}$ for which y is a product complex at the steady state U:

$$\sum_{j \in r} \mathbf{E}(a_j(\mathbf{U})) = \sum_{j \in r'} \mathbf{E}(a_j(\mathbf{U})). \tag{3}$$

Detailed balanced CRNs are a subclass of complex balanced CRNs for which this relation holds separately for each pair of forward and inverse reactions linking two complexes. Detailed balanced steady states correspond to thermodynamic equilibrium states, whereas the others are non-equilibrium steady states (NESS).

In this paper we considered mass-action CRNs, for which the the rate of a chemical reaction is proportional to the product of the concentrations (or the number of molecules) of the reactants raised to powers that are equal to their stoichiometric coefficients. It has been shown that such CRNs are complex-balanced *if and only if* they are of deficiency zero and weakly reversible. This is known as the zero deficiency theorem.

Higher deficiency CRNs correspond to systems for which δ independent conditions on the rate constants have to be satisfied in order for the system to be complex balanced. In a certain sense, δ measures the "distance" of the network from complex balancing.

3. Itō stochastic modeling

To describe the time evolution of stochastic bioprocesses, we used the chemical Langevin equation (CLE), which corresponds to Itō stochastic differential equations (SDE) (34; 35) driven by multidimensional Wiener processes. Itō SDEs are equivalent to the Fokker-Planck and master equation formalisms under some mild conditions and they are well suited for studying biochemical reaction networks (35; 36; 37). Here we focused on systems containing several species, which can be produced from or degraded to the environment and interact with each other to form biomolecular complexes. These systems mimic the interaction between molecular species, such as proteins, DNA or ligands that assemble into protein oligomers or protein-ligand and protein-DNA complexes, but also more complex interactions between, for example, different cell types that coexist in the same tissue.

We start considering the open chemical reaction network depicted in Fig. 1, which models for example the process by which n protein monomers x assemble into a homooligomer z, which in turn disassembles into n monomers x. Both the monomers and oligomers can be produced from the environment or be degraded.

The system of Itō SDEs that describes the dynamics of this reaction network as a function of a continuous time parameter $t \in [0, T]$ reads as:

$$\begin{aligned} dX(t) &= dP_x(X,t) - dR_x(X,t) + n\left[dG_{xz}(X,Z,t) - dF_{xz}(X,Z,t)\right] \\ dZ(t) &= dP_z(Z,t) - dR_z(Z,t) + dF_{xz}(X,Z,t) - dG_{xz}(X,Z,t) \end{aligned} \tag{4}$$

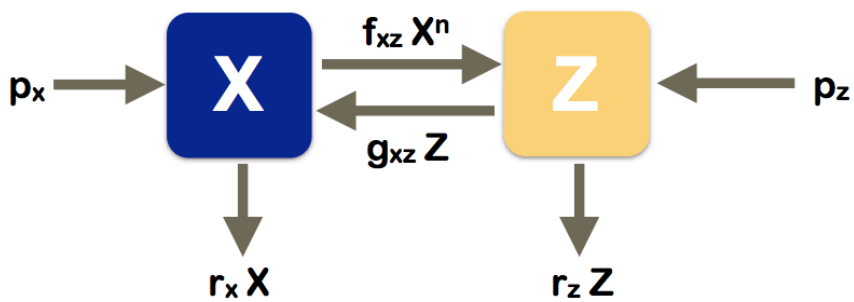


Figure 1: **Schematic picture of the reaction network representing homooligomerization**: $n\,\mathrm{X} \leftrightarrow \mathrm{Z} \leftrightarrow \varnothing \leftrightarrow \mathrm{X}$.

where $X(t)$ and $Z(t)$ are the number of molecules of types x and z, dP_x and dP_z represent the production rates for the corresponding molecular species, dR_x and dR_z their degradation rate, and dF_{xz} and dG_{xz} the interconversion terms.

We chose the production rates to be constant, the degradation rates to be proportional to the number of molecules, and the interconversion rates to satisfy mass-action kinetics, thus to be proportional to the product of the number of molecules of the reacting species raised to the powers of their stoichiometric coefficients. This yields the following relations, each expressed as the sum of a deterministic and a stochastic term:

$$\begin{aligned}
dP_x(X,t) &= p_x\,dt + \alpha_{p_x}\sqrt{p_x}\,dW^{P_x}(t) \\
dR_x(X,t) &= r_x X(t)\,dt + \alpha_{r_x}\sqrt{r_x X(t)}\,dW^{R_x}(t) \\
dP_z(Z,t) &= p_z\,dt + \alpha_{p_z}\sqrt{p_z}\,dW^{P_z}(t) \\
dR_z(Z,t) &= r_z Z(t)\,dt + \alpha_{r_z}\sqrt{r_z Z(t)}\,dW^{R_z}(t) \\
dF_{xz}(X,Z,t) &= f_{xz} X(t)^{(n)}\,dt + \alpha_{f_{xz}}\sqrt{f_{xz} X(t)^{(n)}}\,dW^{F_{xz}}(t) \\
dG_{xz}(X,Z,t) &= g_{xz} Z(t)\,dt + \alpha_{g_{xz}}\sqrt{g_{xz} Z(t)}\,dW^{G_{xz}}(t)
\end{aligned} \tag{5}$$

where $X(t)^{(n)} \equiv X(t)(X(t)-1)\ldots(X(t)-n+1)$. The six $W(t)$ functions stand for independent Wiener processes, satisfying $W(0) = 0$ with $W(t) - W(t')$ following a $\mathcal{N}(0, t-t')$ normal distribution for all (t, t'). Note that these processes have continuous-valued realizations and are thus appropriate when the number of molecules is large enough to be approximated as a continuous variable. In this regime we have also that $X(t)^{(n)} \simeq X(t)^n$. In what follows, we will thus always consider this approximation. The six parameters α that appear in front of the stochastic terms measure the degree of stochasticity of the associated processes. For a simple birth-death process involving only one species

x, the process is purely deterministic when $\alpha_{r_x} = 0 = \alpha_{p_x}$, the fluctuations follow a Poisson distribution when $\alpha_{r_x} = 1 = \alpha_{p_x}$. The stochasticity of the process is increased (super-Poissonian) when $\alpha_{r_x} > 1$ and $\alpha_{p_x} > 1$ and decreased (sub-Poissonian) when both parameters are smaller than one.

In order to solve the system using either analytical or numerical techniques, we approximated the continuous-time SDEs given by Eqs (4,5) by discrete-time SDEs. Therefore, the time interval $[0, T]$ was divided into Ξ equal-length intervals $0 = t_0 < \ldots < t_\Xi = T$, with $t_\tau = \tau \Delta t$ and $\Delta t = T/\Xi$. Using the Euler-Maruyama discretization scheme (38), the discrete-time SDEs read as:

$$\begin{aligned} X_{\tau+1} &= X_\tau + \Delta P_x(X_\tau) - \Delta R_x(X_\tau) + n\left[\Delta G_{xz}(X_\tau, Z_\tau) - \Delta F_{xz}(X_\tau, Z_\tau)\right] \\ Z_{\tau+1} &= Z_\tau + \Delta P_z(Z_\tau) - \Delta R_z(Z_\tau) + \Delta F_{xz}(X_\tau, Z_\tau) - \Delta G_{xz}(X_\tau, Z_\tau) \end{aligned} \quad (6)$$

for all positive integers $\tau \in [0, \Xi]$, with the discretized reaction rates given by:

$$\begin{aligned} \Delta P_x(X_\tau) &= p_x\,\Delta t + \alpha_{p_x}\sqrt{p_x}\,\Delta W_\tau^{P_x} \\ \Delta R_x(X_\tau) &= r_x X_\tau\,\Delta t + \alpha_{r_x}\sqrt{r_x X_\tau}\,\Delta W_\tau^{R_x} \\ \Delta P_z(Z_\tau) &= p_z\,\Delta t + \alpha_{p_z}\sqrt{p_z}\,\Delta W_\tau^{P_z} \\ \Delta R_z(Z_\tau) &= r_z Z_\tau\,\Delta t + \alpha_{r_r}\sqrt{r_z Z_\tau}\,\Delta W_\tau^{R_z} \\ \Delta F_{xz}(X_\tau, Z_\tau) &= f_{xz} X_\tau^n\,\Delta t + \alpha_{f_{xz}}\sqrt{f_{xz} X_\tau^n}\,\Delta W_\tau^{F_{xz}} \\ \Delta G_{xz}(X_\tau, Z_\tau) &= g_{xz} Z_\tau\,\Delta t + \alpha_{g_{xz}}\sqrt{g_{xz} Z_\tau}\,\Delta W_\tau^{G_{xz}} \end{aligned} \quad (7)$$

The independent Wiener processes satisfy $W_\tau = W(t_\tau)$ and $\Delta W_\tau = W_{\tau+1} - W_\tau$, so that in particular $W_0 = 0$, $\mathbf{E}(\Delta W_\tau) = 0$ and $\mathbf{Var}(\Delta W_\tau) = \Delta t$. This system converges towards a steady state in the long-time limit, obtained by first taking the limit $T = \Xi \Delta t \to \infty$ followed by $\Delta t \to 0$. The values of the variables at the steady state will be represented without subscript, *e.g.* $X_\tau \to X$.

To solve analytically these SDEs, and get the mean, variances, and covariances of the different variables at the steady state as a function of the parameters, *i.e.* $\mathbf{E}(X)$, $\mathbf{Var}(X)$, $\mathbf{E}(Z)$, $\mathbf{Var}(Z)$, and $\mathbf{Cov}(X, Z)$, we take the mean of Eqs (6), the mean of their squares, and the mean of the square of well-chosen combinations. The system closes for $n = 1$ but not for $n > 1$, and in this case we thus need to make approximations. We used the moment closure approximation (41) yielding for example:

$$\begin{aligned} \mathbf{E}(X^{n+1}) &\approx \mathbf{E}(X^n)\left(\mathbf{E}(X) + n\frac{\mathbf{Var}(X)}{\mathbf{E}(X)}\right) \\ \mathbf{E}(ZX^n) &\approx \mathbf{E}(X^n)\left(\mathbf{E}(Z) + n\frac{\mathbf{Cov}(X, Z)}{\mathbf{E}(X)}\right) \end{aligned} \quad (8)$$

As usual in reaction networks, the intrinsic noise on the molecular species x and z is quantified through their Fano factors $\mathbf{F}(X)$ and $\mathbf{F}(Z)$, defined as:

$$\mathbf{F}(X) = \frac{\mathbf{Var}(X)}{\mathbf{E}(X)} \quad , \quad \mathbf{F}(Z) = \frac{\mathbf{Var}(Z)}{\mathbf{E}(Z)} \quad (9)$$

If the X and Z follow a Poisson distribution, their Fano factor $\mathbf{F}$ is equal to one. When $\mathbf{F}$ is larger than one, the intrinsic noise affects more strongly the variable concentration, and the distribution is called super-Poissonian. The distribution is called sub-Poissonian when $\mathbf{F} < 1$.

The Itō SDEs described above can be easily generalizable to model more complex processes of interest, for example those described in section 5.

4. Homooligomerization

The homooligomerization system that is schematically depicted in Fig. 1 is a reversible CRN for non-zero interconversion terms f_{xz} and g_{xz}. Its deficiency is equal to one when $n > 1$ and the two species are connected to the environment, and to zero otherwise. The Itō SDEs that describe it are given in Eqs (6, 7). They can be solved analytically, using the moment closure approximation of Eq. (8). For sake of simplicity, we assumed the equality of all stochasticity parameters: $\alpha_{r_x} = \alpha_{r_z} = \alpha_{p_x} = \alpha_{p_z} = \alpha_f = \alpha_g = \alpha$.

In this way, we obtained the Fano factors of X and Z and the covariance at the steady state expressed as a function of J_{xz}, the flux that flows between the two molecular species x and z multiplied by the difference in stoichiometry coefficients between the reacting complexes:

$$J_{xz} = (n-1)\left(f_{xz}\mathbf{E}(X^n) - g_{xz}\mathbf{E}(Z)\right) \tag{10}$$

In what follows, we will call this flux the S-flux. It is zero when $n = 1$, in which case the steady state is complex balanced, or when $f_{xz}\mathbf{E}(X^n) = g_{xz}\mathbf{E}(Z)$, in which case it is detailed balanced. In the other cases, where the S-flux does not vanish, the system admits a unbalanced nonequilibrium steady state. Moreover, it is positive when the flux flows towards the complex of highest complexity, defined as the one of highest stoichiometry. In terms of the S-flux, the Fano factors and the covariance are expressed as:

$$\begin{aligned}
\mathbf{F}(X) &= \alpha\left[1 - J_{xz}\gamma_x\right] \\
\mathbf{F}(Z) &= \alpha\left[1 - J_{xz}\gamma_z\right] \\
\mathbf{Cov}(X,Z) &= -\alpha\, J_{xz}\gamma_{xz}
\end{aligned} \tag{11}$$

with

$$\begin{aligned}
\gamma_x &= \frac{n}{D\mathbf{E}(X)}\left(r_z(n^2 f_{xz}\frac{\mathbf{E}(X^n)}{\mathbf{E}(X)} + g_{xz} + r_x + r_z) + g_{xz}(g_{xz} + r_x + r_z)\right) \geq 0 \\
\gamma_y &= \frac{n}{D\mathbf{E}(Z)} n^2 f_{xz}^2 \frac{\mathbf{E}(X^n)^2}{\mathbf{E}(X)^2} \geq 0 \\
\gamma_{xz} &= \frac{1}{D} n^2 f_{xz}\frac{\mathbf{E}(X^n)}{\mathbf{E}(X)}(g_{xz} + r_z) \geq 0
\end{aligned} \tag{12}$$

and

$$D = 2\left(n^2 r_z f_{xz}\frac{\mathbf{E}(X^n)}{\mathbf{E}(X)} + r_x g_{xz} + r_x r_z\right)\left(n^2 f_{xz}\frac{\mathbf{E}(X^n)}{\mathbf{E}(X)} + g_{xz} + r_x + r_z\right)$$

The remaining equations yield the mean number of molecules in terms of the parameters:

$$\begin{aligned}(g_{xz}+r_z)\mathbf{E}(Y) &= p_z + f_{xz}\mathbf{E}(X^n) \\ nf_{xz}\mathbf{E}(X^n)+r_x\mathbf{E}(X) &= p_x + ng_{xz}\mathbf{E}(Y)\end{aligned} \tag{13}$$

First, we observe that the Fano factors and the covariance are proportional to the stochastic parameter α, and thus that they vanish for deterministic systems, as expected. Furthermore, the covariance is equal to minus the S-flux multiplied by a positive coefficient. This means that when the flux flows towards the complex of highest complexity, the covariance is negative, and when it flows towards the complex of lowest complexity, it is positive. Finally, both Fano factors $\mathbf{F}(X)$ and $\mathbf{F}(Z)$ are equal to α minus the S-flux multiplied by a positive coefficient. This means that when the S-flux is positive, the noise on both species x and z is reduced, whereas it is amplified when the S-flux is negative. When $\alpha = 1$, the reduction or amplification is with respect to Poissonian noise. From these equations also follows:

$$\mathbf{F}(X)+\mathbf{F}(Z) = \alpha\,(2 - \ J_{xz}\gamma) \tag{14}$$

with the positive coefficient:

$$\gamma = \frac{n\left((r_zZ + f_{xz}\mathbf{E}(X^n))n^2 f_{xz}\frac{\mathbf{E}(X^n)}{\mathbf{E}(X)} + (g_{xz}+r_z)\mathbf{E}(Z)(g_{xz}+r_x+r_z)\right)}{D\mathbf{E}(X)\mathbf{E}(Z)} \tag{15}$$

We thus recover the general relation obtained in (22) for a system of a rank 2 with deficiency $\delta = 1$ and stochasticity level $\alpha = 1$. We obtained here the additional result that, for the system described by Fig. 1, not only the global intrinsic noise represented by the sum of the Fano factors, but also the noise on the separate species, is amplified or reduced according to the sign of the S-flux.

To obtain the Fano factors, covariances and number of molecules as a function of the parameters only, we had to consider separately the oligomers of different degrees n.

$n = 1$*: the monomeric system*

The simplest case where each molecule z is built from only one molecule x represents for example molecules that undergo a conformational change or move to different cell compartments, without interaction with other biomolecules. In this case, the deficiency is zero, the system is complex balanced and the S-flux J_{xz} vanishes. No approximations are needed to solve the SDE equations. The number of molecules is obtained from Eq. (13) as a function of the parameters:

$$\mathbf{E}(X) = \frac{g_{xz}p_x + g_{xz}p_z + r_zp_x}{g_{xz}r_x + f_{xz}r_z + r_xr_z} \quad , \quad \mathbf{E}(Z) = \frac{f_{xz}p_x + f_{xz}p_z + r_xp_z}{g_{xz}r_x + f_{xz}r_z + r_xr_z} \tag{16}$$

and Eqs (11) reduce to:

$$\mathbf{F}(X) = \alpha \quad , \quad \mathbf{F}(Z) = \alpha \quad , \quad \mathbf{Cov}(X,Z) = 0 \tag{17}$$

In conclusion, the molecules x and z are uncorrelated, and both have a constant level of intrinsic noise, which is Poissonian in the case $\alpha = 1$.

We would like to emphasize that the S-flux vanishes but not the flux: for most parameter values a net flux flows between the molecules and from and towards the environment. However, as both x and z molecules have the same complexity, the S-flux vanishes. The determinant of the covariance matrix is here simply equal to $\alpha^2(\mathbf{E}(X)^2 + \mathbf{E}(Z)^2)$.

$n = 2$: the dimerization system

In the case in which z are dimers formed of two molecules x, the system no longer closes and we have to use the approximations of Eq. (8) to have an analytical solution. The mean number of molecules at the steady state is then obtained as a function of the systems parameters employing Eq. (13). This yields:

$$\begin{aligned} \mathbf{E}(X) &\approx \frac{-r_x(g_{xz} + r_z) + L}{4 f_{xz} r_z} \\ \mathbf{E}(Z) &\approx \frac{4 f_{xz} r_z (p_x + 2p_z) r_z + r_x^2 (g_{xz} + r_z) - r_x L}{8 f_{xz} r_z^2} \end{aligned} \tag{18}$$

with

$$L = \sqrt{r_x^2 (g_{xz} + r_z)^2 + 8 f_{xz} r_z (p_x + 2p_z + p_x p_z)} \tag{19}$$

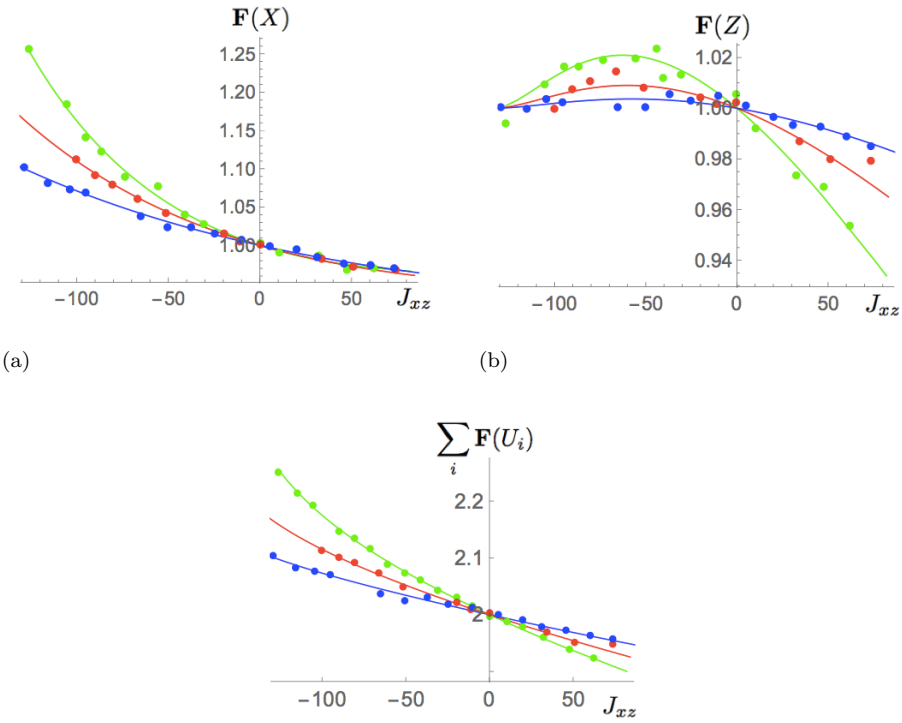
The Fano factors and covariances are given by Eqs (11) with the number of molecules given by Eqs (18).

The S-flux vanishes when

$$\frac{f_{xz}}{g_{xz}} \approx \frac{p_z r_x^2}{p_x^2 r_z} \tag{20}$$

For f_{xz}/g_{xz} values smaller than this threshold, the S-flux is negative while for larger f_{xz}/g_{xz} values it is positive. Note that when the z molecules are not produced or the x molecules not degraded, the S-flux is always positive. In contrast, it is always negative when the x molecules are not produced or the z molecules not degraded.

The Fano factors as a function of the S-flux are depicted in Figs 2a-c, for some parameter values and stochasticity level $\alpha = 1$. We would first like to stress that the numerical and analytical results are very close, which indicates that the moment closure approximation used for the analytical developments is a good approximation, at least for the tested parameter values. We observe a noise reduction for all species and parameter values when the S-flux is positive, and a noise increase for negative S-flux, as expected. We also note that decreasing the production rate of the x molecules, thus lowering the total number of molecules in the system, amplifies this noise modulation effect.



(a) (b)

(c)

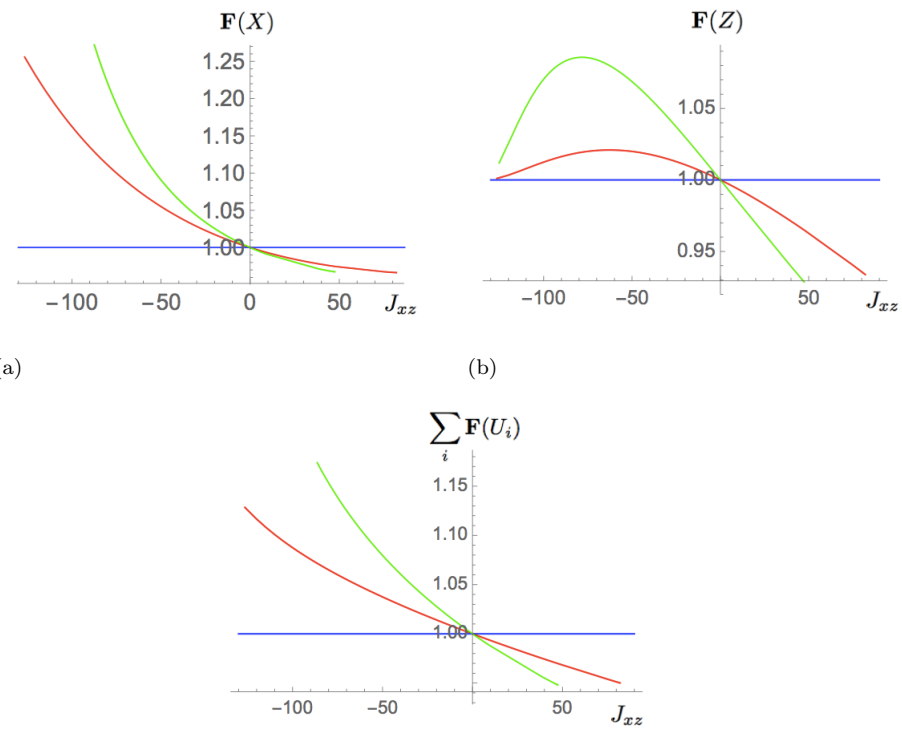
Figure 2: **Stochastic behavior of homodimerization processes, as a function of the interconversion parameter** f_{xz}. The full lines are the analytical results of Eqs (11,18), and the points correspond to results of numerical stochastic simulations. The stochasticity parameter $\alpha = 1$, the oligomerization degree $n = 2$, and the parameters $p_z = 200$, $r_x = r_z = 0.001$ and $g_{xz} = 0.002$. The parameter p_x is given different values: 200 (green line), 500 (red line) and 1000 (blue line). (a) Fano factor $\mathbf{F}(X)$; (b) Fano factor $\mathbf{F}(Z)$; (c) Sum of Fano factors $\mathbf{F}(X) + \mathbf{F}(Z)$.

$n = 3$ and $n = 4$: the trimerization and tetramerization systems

When z are trimers or tetramers of x molecules, we can use the same procedure as in the $n = 2$ case, and solve the mean number of molecules at the steady state from Eq. (13). The analytical results are given in appendix A1.

The Fano factors of the species involved in the tetramerization process ($n = 4$) are plotted as a function of the S-flux in Figs 3 a-c, and are compared with those of the dimerization ($n = 2$) and the monomeric interconversion ($n = 1$). Clearly, for the same parameter values, the amplification and reduction of the intrinsic noise is increased for higher oligomerization degrees. Note the different behaviors of the Fano factors of the monomers and oligomers. When the S-flux is negative, the noise amplification on the oligomers appears limited, in contrast to the noise on the monomers which continues to grow for decreasing flux values.

Instead, when the S-flux is positive, the fluctuations of the oligomers seems to be suppressed more strongly than those of the monomers.



(a) (b)

(c)

Figure 3: **Stochastic behavior of homooligomerization processes, as a function of the interconversion parameter** f_{xz}. The full lines are the analytical results of Eqs (11,18,33). The stochasticity parameter $\alpha = 1$, and the parameters $p_x = p_z = 200$, $r_x = r_z = 0.001$ and $g_{xz} = 0.002$. The oligomerization degree is given different values: $n = 1$ (blue line), $n = 2$ (red line) and $n = 4$ (green line). (a) Fano factor $\mathbf{F}(X)$; (b) Fano factor $\mathbf{F}(Z)$; (c) Sum of Fano factors $\mathbf{F}(X) + \mathbf{F}(Z)$.

5. Oligomerization reactions with intermediate steps

We now turn to the more complex systems schematically depicted in Fig. 4. They describe a wide range of biological systems such as monomeric proteins that tetramerize through an intermediate step of dimerization (31) or that undergo amyloid formation through oligomeric intermediates (32).

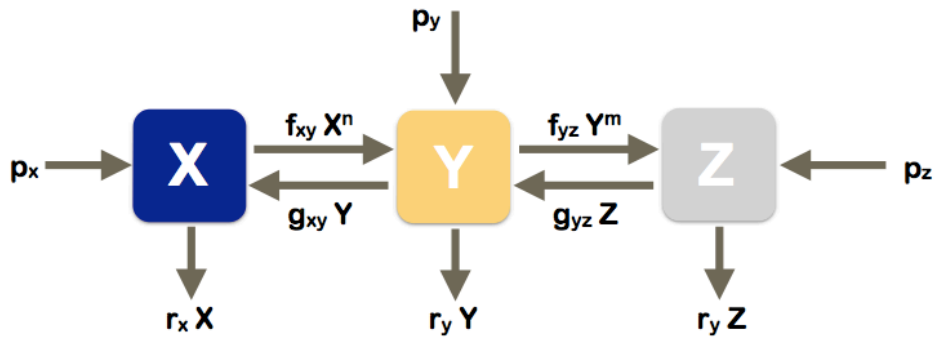


Figure 4: **Schematic picture of the reaction network representing homooligomerization with an intermediate state:** $n\,\mathbf{X} \leftrightarrow \mathbf{Y}$, $m\,\mathbf{Y} \leftrightarrow \mathbf{Z}$, $\mathbf{Z} \leftrightarrow \varnothing, \mathbf{X} \leftrightarrow \varnothing, \mathbf{Y} \leftrightarrow \varnothing$.

Such systems are reversible CRNs for non-zero values of the interconversion parameters f_{xy}, f_{yz}, g_{xy} and g_{yz}, with deficiency values up to $\delta = 2$. In particular, when all the species are connected to the environment, we have $\delta = 2$ for $n > 1$ and $m > 1$, $\delta = 1$ when either $(n > 1, m = 1)$ or $(n = 1, m > 1)$, and $\delta = 0$ for $n = 1 = m$. These CRNs admit a non-equilibrium steady state which is complex or detailed balanced if $\delta = 0$.

To model these systems, we used the same formalism as in the previous section, namely discrete-time Itō SDEs with an Euler-Maruyama discretization scheme (38). The system of non-linear coupled SDEs reads as:

$$\begin{aligned}
X_{\tau+1} &= X_\tau + \Delta P_x(X_\tau) - \Delta R_x(X_\tau) + n\,[\Delta G_{xy}(Y_\tau) - \Delta F_{xy}(X_\tau)] \\
Y_{\tau+1} &= Y_\tau + \Delta P_y(Y_\tau) - \Delta R_y(Y_\tau) + [\Delta F_{xy}(X_\tau) - \Delta G_{xy}(Y_\tau)] + m\,[\Delta G_{yz}(Z_\tau) - \Delta F_{yz}(Y_\tau)] \\
Z_{\tau+1} &= Z_\tau + \Delta P_z(Z_\tau) - \Delta R_z(Z_\tau) + \Delta F_{yz}(Y_\tau) - \Delta G_{yz}(Z_\tau)
\end{aligned} \tag{21}$$

for all positive integers $\tau \in [0, \Xi]$. The discretized reaction rates are given by:

$$\begin{aligned}
\Delta P_x(X_\tau) &= p_x\,\Delta t + \alpha_{p_x}\sqrt{p_x}\,\Delta W_\tau^{P_x} \\
\Delta R_x(X_\tau) &= r_x X_\tau\,\Delta t + \alpha_{r_x}\sqrt{r_x X_\tau}\,\Delta W_\tau^{R_x} \\
\Delta P_y(Z_\tau) &= p_y\,\Delta t + \alpha_{p_y}\sqrt{p_y}\,\Delta W_\tau^{P_y} \\
\Delta R_y(Z_\tau) &= r_y Y_\tau\,\Delta t + \alpha_{r_y}\sqrt{r_y Y_\tau}\,\Delta W_\tau^{R_y} \\
\Delta P_z(Z_\tau) &= p_z\,\Delta t + \alpha_{p_z}\sqrt{p_z}\,\Delta W_\tau^{P_z} \\
\Delta R_z(Z_\tau) &= r_z Z_\tau\,\Delta t + \alpha_{r_z}\sqrt{r_z Z_\tau}\,\Delta W_\tau^{R_z} \\
\Delta F_{xy}(X_\tau) &= f_{xy} X_\tau^n\,\Delta t + \alpha_{f_{xy}}\sqrt{f_{xy} X_\tau^n}\,\Delta W_\tau^{F_{xy}} \\
\Delta G_{xy}(Y_\tau) &= g_{xy} Y_\tau\,\Delta t + \alpha_{g_{xy}}\sqrt{g_{xy} Y_\tau}\,\Delta W_\tau^{G_{xy}} \\
\Delta F_{yz}(Y_\tau) &= f_{yz} Y_\tau^m\,\Delta t + \alpha_{f_{yz}}\sqrt{f_{yz} Y_\tau^m}\,\Delta W_\tau^{F_{yz}} \\
\Delta G_{yz}(Z_\tau) &= g_{yz} Z_\tau\,\Delta t + \alpha_{g_{yz}}\sqrt{g_{yz} Z_\tau}\,\Delta W_\tau^{G_{yz}}
\end{aligned} \tag{22}$$

where the ten Wiener processes are independent.

These equations can be solved analytically, using the moment closure approximation of Eq. (8). For simplicity, we again assumed the equality of all stochasticity parameters: $\alpha_{r_x} = \alpha_{r_y} = \alpha_{r_z} = \alpha_{p_x} = \alpha_{p_y} = \alpha_{p_z} = \alpha_{f_{xy}} = \alpha_{g_{xy}} = \alpha_{f_{yz}} = \alpha_{g_{yz}} = \alpha$. There are two S-fluxes in this CRN, which are independent and in general non-zero when $\delta = 2$:

$$\begin{aligned} J_{xy} &= (n-1)\left(f_{xy}\mathbf{E}(X^n) - g_{xy}\mathbf{E}(Y)\right) \\ J_{yz} &= (m-1)\left(f_{yz}\mathbf{E}(Y^m) - g_{yz}\mathbf{E}(Z)\right) \end{aligned} \tag{23}$$

We obtained the Fano factors of X, Y and Z and the covariances $\mathbf{Cov}(X,Y)$, $\mathbf{Cov}(X,Z)$ and $\mathbf{Cov}(Y,Z)$ at the steady state expressed as a function of these two S-fluxes:

$$\begin{aligned} \mathbf{F}(X) &= \alpha\left[1 - J_{xy}\gamma_x^{xy} - J_{yz}\gamma_x^{yz}\right] \\ \mathbf{F}(Y) &= \alpha\left[1 - J_{xy}\gamma_y^{xy} - J_{yz}\gamma_y^{yz}\right] \\ \mathbf{F}(Z) &= \alpha\left[1 - J_{xy}\gamma_z^{xy} - J_{yz}\gamma_z^{yz}\right] \\ \mathbf{Cov}(X,Y) &= -\alpha\left[J_{xy}\gamma_{xy}^{xy} + J_{yz}\gamma_{xy}^{yz}\right] \\ \mathbf{Cov}(X,Z) &= -\alpha\left[J_{xy}\gamma_{xz}^{xy} + J_{yz}\gamma_{xz}^{yz}\right] \\ \mathbf{Cov}(Y,Z) &= -\alpha\left[J_{xy}\gamma_{yz}^{xy} + J_{yz}\gamma_{yz}^{yz}\right] \end{aligned} \tag{24}$$

with all γ's positive functions of the parameters and the mean values $\mathbf{E}(X)$, $\mathbf{E}(Y)$ and $\mathbf{E}(Z)$. A corollary result is that the sum of the Fano factors over all species is equal to the rank $\mathcal{X}$ of the system minus a linear combination of the S-fluxes with positive coefficients:

$$\mathbf{F}(X) + \mathbf{F}(Y) + \mathbf{F}(Z) = \alpha\left[\mathcal{X} - J_{xy}\gamma^{xy} - J_{yz}\gamma^{yz}\right] \tag{25}$$

The values of the positive coefficient γ^{xy} and γ^{yz} are explicitly given in appendix A2. We thus recover the result obtained in (22) and generalize it to the Fano factors of each species taken individually.

To get also $\mathbf{E}(X)$, $\mathbf{E}(Y)$ and $\mathbf{E}(Z)$ in terms of the parameters of the system, we need to solve the following relations:

$$\begin{aligned} p_x + n\,p_y + n\,m\,p_z &= r_x\mathbf{E}(X) + n\,r_y\mathbf{E}(Y) + n\,m\,r_z\mathbf{E}(Z) \\ p_x + n\,g_{xy}\mathbf{E}(Y) &= r_x\mathbf{E}(X) + n\,f_{xy}\mathbf{E}(X^n) \\ p_z + n\,f_{yz}\mathbf{E}(Y^m) &= r_z\mathbf{E}(Z) + m\,g_{xy}\mathbf{E}(Z) \end{aligned} \tag{26}$$

which correspond to the mean of Eqs (21) at the steady state. For solving these equations, we have to specify the values of m and n.

As an example, we analyzed the results in the case $n = 2 = m$. More specifically, we plotted the values of the Fano factors of X, Y and Z as well as their sum as a function of the two fluxes J_{xy} and J_{yz} for $\alpha = 1$, for fixed values of $p_x = p_y = p_z$, $r_x = r_y = r_z$ and $g_{xy} = g_{yz}$, leaving the two interconversion terms f_{xy} and f_{yz} as free parameters. As seen in Figs 5, the number of molecules of each species follow a sub-Poissonian distribution ($\mathbf{F}(U_i) < 1$) in the first

quadrant (when both fluxes are positive) in which case the noise is reduced, while in the third quadrant a super-Poisonnian distribution is observed with an increase of the noise level ($\mathbf{F}(U_i) > 1$). The domains of existence of the solutions are investigated in Appendix B.

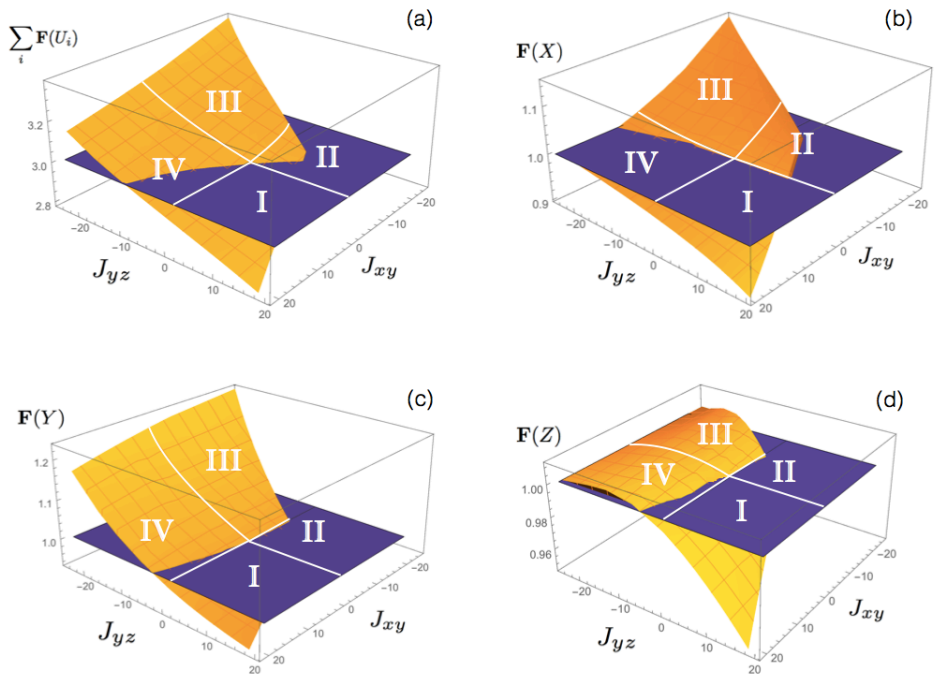


Figure 5: **Stochastic behavior of homotetramerization processes with an intermediate dimerization step, as a function of the interconversion parameters f_{xy} and f_{yz}.** The oligomerization degrees $n = 2 = m$, the stochasticity parameter $\alpha = 1$, and the parameters $p_x = p_y = p_z = 50$ and $r_x = r_y = r_z = g_{xy} = g_{yz} = 0.001$. Fano factors: (b) $\mathbf{F}(X)$, (c) $\mathbf{F}(Y)$, (d) $\mathbf{F}(Z)$), and (a) $\sum_i \mathbf{F}(U_i) = \mathbf{F}(X) + \mathbf{F}(Y) + \mathbf{F}(Z)$ as a function of the S-fluxes J_{xy} and J_{yz}. The orange surfaces correspond to the Fano factors and the blue ones to the constant $z = 1$ plane in (b-d) and $z = 3$ in (a).

6. Discussion

In this paper, we gained important insights into the relation between the complexity of a system and its intrinsic noise, even though the picture is not yet complete. On the basis of the results obtained for the open model CRNs depicted in Figs 1 and 4, we propose the following conclusions and tentative generalizations:

- The modulation of noise in a mass-action CRN is related to its deficiency, and can be expressed as a function of the S-fluxes.

- For δ=0 we have:

$$\mathbf{F}(U_i) = \alpha \tag{27}$$

where U_i is the number of molecules of species i. When α=1, the number of molecules of each species thus follows a Poisson distribution (14). Note that this is only true for open systems. For closed systems, we showed in (23) the weaker result $\sum_i \mathbf{F}(U_i) = \chi$ since in this case, the number of molecules follow a multinomial distribution constrained by the conservation of the total number of molecules (14).

- For δ=1, there is one independent S-flux J_1, and the Fano factors of the different species i are expressed as:

$$\mathbf{F}(U_i) = \alpha \left[1 - J_1 \gamma_i^1\right] \tag{28}$$

where all the γ coefficients are positive functions of the parameters. The noise is thus amplified when the S-flux is negative, which means that the flux flows towards the species of smallest complexity. The noise is reduced when the S-flux is positive, thus when the flux flows towards the species of highest complexity. Note that in the case of several dependent fluxes, the positivity of the γ coefficient is not ensured; this case will be considered in a forthcoming publication.

- For δ=2, there are two independent S-fluxes J_1 and J_2, and the Fano factors satisfy the relations:

$$\mathbf{F}(U_i) = \alpha \left[1 - J_1 \gamma_i^1 - J_2 \gamma_i^2\right] \tag{29}$$

with positive γ values. When the two S-fluxes are positive, and thus the two fluxes flow towards the highest complexity species, we have noise reduction on all species. When the two S-fluxes are negative and the fluxes flow towards lowest complexity, we observe noise amplification on all species. When one S-flux is positive and the other negative, the result depends on the relative value of the associated γ values.

- We argue that these trends remain valid for any value of δ, and that we have:

$$\mathbf{F}(U_i) = \alpha \left[1 - \sum_j J_j \gamma_i^j\right] \tag{30}$$

where the sum is over the internal S-fluxes of the CRN, and all γ coefficients are positive functions of the parameters.

In addition to rigorously demonstrating this conjecture for complex CRNs with generic δ values, we would like to investigate two other points. The first is the extension of our study to systems with generalized kinetic schemes. Indeed, mass-action kinetics is only valid in the case of elementary processes occurring

in homogenous solutions. *In vivo* biomolecular reactions are usually not elementary and are affected by macromolecular crowding. Their description thus requires a modification of the rate law (27; 28).

The second point is related to the modeling of noise in systems described using model reduction techniques. Indeed, systems such as metabolic or signaling networks are far too complex to be mathematically described with full details, which would require a huge number of parameters. To cope with this issue, different reduction techniques have been introduced (29; 30), such as the quasi steady-state approximation (QSSA) in which the fast variables are separated from the slow variables, and only the latter are considered as dynamical. Another reduction technique is the variable lumping method in which the vector of the reactants is dimensionally reduced to a vector of pseudospecies, in such a way that the kinetic equations are easier to solve, and fewer parameters need to be determined. However, it is not trivial to deal with the fluctuations in such reduced models. Indeed, while in elementary processes the fluctuations can be considered to follow Poisson-type distributions with all stochasticity parameters $\alpha = 1$, for non-elementary reduced variables this cannot be assumed *a priori*.

Acknowledgments

We thank Mitia Duerinckx for useful discussions. FP is Postdoctoral researcher and MR Research Director at the Belgian Fund for Scientific Research (FNRS). We declare that there is no conflict of interest regarding the publication of this manuscript.

References

[1] M. Elowitz, A. Levine, E. Siggia, and P. Swain, Stochastic gene expression in a single cell, Science 297, 1183-1186 (2002).

[2] A Raj, A van Oudenaarden, Nature, nurture, or chance: stochastic gene expression and its consequences, Cell 135, 216-226 (2008).

[3] G. Balazsi, A. van Oudenaarden, J.J. Collins, Cellular Decision-Making and Biological Noise: From Microbes to Mammals, Cell, 144, 910-925 (2011).

[4] D.J. Wilkinson, Stochastic modelling for quantitative description of heterogeneous biological systems,Nature Reviews Genetics 10, 122-133 (2009).

[5] J Paulsson, Summing up the noise in gene networks, Nature 427, 415-418 (2004).

[6] N Barkai, BZ Shilo, Variability and robustness in biomolecular systems, Mol Cell 28, 755-60 (2007).

[7] Alon U, Network motifs: theory and experimental approaches, Nat. Rev. Genet. 8, 450-461 (2007).

[8] Rosenfeld N, Elowitz MB, Alon U, Negative autoregulation speeds the response times of transcription networks, J Mol Biol 323, 785-793 (2002).

[9] CV Rao, DM. Wolf, AP. Arkin, Control, exploitation and tolerance of intracellular noise, Nature 420, 231-7 (2002).

[10] M Richard, G Yvert, How does evolution tune biological noise?, Front Genet. 5, 374 (2014).

[11] B Lehner, Selection to minimise noise in living systems and its implications for the evolution of gene expression, Mol. Syst. Biol. 4, 170 (2008).

[12] KF Murphy, RM Adams, X Wang, G Balazsi, JJ Collins, Tuning and controlling gene expression noise in synthetic gene networks, Nucleic Acids Res. 38, 2712-26 (2010).

[13] DA Oyarzun, JB Lugagne, GB Stan, Noise Propagation in Synthetic Gene Circuits for Metabolic Control, ACS Synth. Biol 4, 116-125 (2015).

[14] DF Anderson, G Craciun, TG Kurtz, Product-form stationary distributions for deficiency zero chemical reaction networks, Bulletin of Mathematical Biology, 72, 1947-1970 (2010).

[15] DF Anderson, JC Mattingly, HF Nijhout, MC Ree, Propagation of Fluctuations in Biochemical Systems, I: Linear SSC Networks, Bulletin of Mathematical Biology, 69,1791-1813 (2007).

[16] DF Anderson, JC Mattingly, Propagation of fluctuations in biochemical systems, II: Nonlinear chains, IET Syst Biol 1,313-25 (2007).

[17] T Schmiedl, U. Seifert, Stochastic thermodynamics of chemical reaction networks, The Journal of Chemical Physics 126, 044101 (2007).

[18] M Polettini, A Wachtel, M Esposito, Dissipation in noisy chemical networks: The role of deficiency, The Journal of Chemical Physics 143, 184103 (2015).

[19] R Rao, M Esposito, Nonequilibrium Thermodynamics of Chemical Reaction Networks: Wisdom from Stochastic Thermodynamics, Phys Rev X 6, 041064 (2016).

[20] L Cardelli, A Csikász-Nagy, N Dalchau, M, M Tschaikowski, Noise Reduction in Complex Biological Switches, Scientific Reports 6, 20214 (2016).

[21] M Rooman, J Albert, M Duerinckx. Stochastic noise reduction upon complexification: Positively correlated birth-death type systems, Journal of Theoretical Biology 354, 113-123 (2014).

[22] F Pucci, M Rooman. Insights into the relation between noise and biological complexity, submitted, arXiv:1709.00883 [q-bio.MN] (2017).

[23] M Rooman, F Pucci. Intrinsic noise modulation in closed oligomerization-type systems, submitted (2017).

[24] M. Feinberg, Chemical reaction network structure and the stability of complex isothermal reactors I. The de?ciency zero and de?ciency one theorems, Chemical Engineering Science, 42, 2229-2268 (1987).

[25] M. Feinberg. Chemical reaction network structure and the stability of complex isothermal reactors II. Multiple steady states for networks of de?ciency one. Chemical Engineering Science, 43, 1?25 (1988).

[26] D Angeli, A Tutorial on Chemical Reaction Network Dynamics, European Journal of Control 3-4, 398-406 (2009).

[27] S Mller and G Regensburger, Generalized Mass Action Systems: Complex Balancing Equilibria and Sign Vectors of the Stoichiometric and Kinetic-Order Subspaces, SIAM J. Appl. Math, 72, 1926-1947)2012)

[28] DF Anderson, SL Cotter, Product-form stationary distributions for deficiency zero networks with non-mass action kinetics, Bulletin of mathematical biology 78, 2390-2407 (2016)

[29] O Radulescu, AN Gorban, A Zinovyev, V Noel, Reduction of dynamical biochemical reaction networks in computational biology. Front Genet. 3,00131 (2012).

[30] S Rao, A van der Schaft, K van Eunen, BM Bakker,1B Jayawardhana, A model reduction method for biochemical reaction networks, BMC Syst Biol. 8, 52 (2014).

[31] MH Ali, B Imperiali, Protein oligomerization: how and why, Bioorg Med Chem. 13, 5013-20 (2005).

[32] M Fandrich, Oligomeric Intermediates in Amyloid Formation: Structure Determination and Mechanisms of Toxicity, Journal of Molecular Biology, 421, 24 (2012).

[33] M Hoffmann, HH Chang, S Huang, DE Ingber, M, Loeffler, J Galle, Noise-driven stem cell and progenitor population dynamics, PLoSOne 3,e2922 (2008).

[34] K Itō. Stochastic integral. Proc. Imperial Acad. Tokyo, 20, 519524 (1944).

[35] E Allen. Modeling with Itō Stochastic Differential Equations, Springer, the Netherlands (2007).

[36] W Moon, J S Wettlaufer. On the interpretation of Stratonovich calculus. New Journal of Physics, 6, 055017 (2014)

[37] DT Gillespie. The chemical Langevin equation, Journal of Chemical Physics, 113, 297-306 (2000).

[38] PE Kloeden, E Platen. Numerical Solution of Stochastic Differential Equations, Springer, Berlin (1992).

[39] M Feinberg, Complex balancing in general kinetic systems, Arch. Rational Mech. Anal. 49, 187-194 (1972).

[40] Fritz J. M. Horn, Necessary and sufficient conditions for complex balancing in chemical kinetics, Arch. Rat. Mech. Anal. 49, 172-186 (1972).

[41] C Kuehn, Moment Closure - A Brief Review, Control of Self-Organizing Complex Systems, Springer (2016).

[42] A Whitty, Cooperativity and biological complexity, Nature Chemical Biology 4, 435 - 439 (2008).

Appendix A : Analytical Results

A1. Oligomerization without intermediate step

For the CRN depicted in Fig. 1, with oligomerization degree n, the number of molecules of type x and z can be obtained as a function of the parameters by solving Eq. (13). For $n = 1$ and $n = 2$, the solution is given in Eqs (16,18). For $n = 3$, we get:

$$\begin{aligned}\mathbf{E}(X) &= \frac{1}{D}\left(-2f_{xz}r_xr_z(g_{xz}+r_z)+2^{1/3}\left(9f_{xz}^2r_z^2\left(g_{xz}(p_x+3\,p_z)+p_xr_z\right)+L\right)^{2/3}\right)\\ \mathbf{E}(Y) &= \frac{p_z+f_{xz}\mathbf{E}(X)^3}{g_{xz}+r_z}\end{aligned} \tag{31}$$

with

$$
\begin{aligned}
L &= \sqrt{f_{xz}^3 r_z^3 \left(4 r_x^3 (g_{xz}+r_z)^3 + 81 f_{xz} r_z (g_{xz}(p_x+3p_z)+p_x r_z)^2\right)} \\
D &= 2^{2/3}\, 3\, f_{xz} r_z \left(9 f_{xz}^2 r_z^2 \left(g_{xz}(p_x+3\,p_z)+p_x r_z\right)+L\right)^{1/3}
\end{aligned} \tag{32}
$$

In the case $n=4$, we have:

$$
\begin{aligned}
\mathbf{E}(X) &= \frac{1}{2\;2^{5/6}3^{1/3}}\left(\sqrt{\frac{K}{D}-\frac{D}{f_{xz}r_z}+\frac{6\sqrt{2}\,r_x(g_{xz}+r_z)\sqrt{D}}{\sqrt{f_{xz}r_z}\sqrt{D^2-f_{xz}r_zK}}}-\sqrt{\frac{D^2-f_{xz}r_zK}{f_{xz}r_zD}}\right) \\
\mathbf{E}(Y) &= \frac{p_z+f_{xz}\mathbf{E}(X)^4}{g_{xz}+r_z}
\end{aligned} \tag{33}
$$

with

$$
\begin{aligned}
L &= \sqrt{3}\sqrt{f_{xz}^2 r_z^2 (27\, r_x^4 (g_{xz}+r_z)^4 + 1024\, g_{xz}^4 f_{xz} r_z (g_{xz}(p_x+4p_z)+p_x r_z)^3)} \\
D &= \left(9\, f_{xz} r_x^2 r_z (g_{xz}+r_z)^2 + L\right)^{1/3} \\
K &= 8\;6^{1/3}(g_{xz}(p_x+4p_z)+p_x r_z)
\end{aligned} \tag{34}
$$

A2. Oligomerization with intermediate step

For the CRNs depicted in Fig. 4, with oligomerization degrees n and m, we obtained the Fano factors as a function of the S-fluxes multiplied by positive functions of the parameters (Eqs (24,25)). In particular, the positive coefficients γ_{xy} and γ_{xz} that appear in the sum of Fano factors, Eq. (25) are equal to:

$\gamma_{xy} =$

$$
\begin{aligned}
& n\,\Big((g_{xy}+r_y)\,(g_{xy}+r_x+r_y)\,(g_{yz}+r_z)\,(g_{yz}+r_x+r_z)\,(g_{xy}+g_{yz}+r_y+r_z)\,\mathbf{E}[X]^2\,\mathbf{E}[Y]^3\,\mathbf{E}[Z] + \\
& \quad n^4 f_{xy}^2\,(g_{yz}+r_z)\,(g_{yz}+r_y+r_z)\,\mathbf{E}\left[X^n\right]^2\mathbf{E}[Y]^2\left(f_{xy}\,\mathbf{E}\left[X^n\right]+r_y\,\mathbf{E}[Y]\right)\mathbf{E}[Z] + \\
& \quad n^2 f_{xy}\,(g_{yz}+r_z)\,\mathbf{E}[X]\,\mathbf{E}\left[X^n\right]\mathbf{E}[Y]^2\Big(f_{xy}\,(g_{yz}+r_x+r_z)\,(g_{xy}+g_{yz}+r_y+r_z)\,\mathbf{E}\left[X^n\right]+\Big(g_{xy}^2\,(g_{yz}+r_y+r_z)+ \\
& \qquad r_y\,(g_{yz}+r_y+r_z)\,(g_{yz}+2\,r_x+r_y+r_z)+g_{xy}\left(2\,r_x\,r_y+2\,r_y^2+g_{yz}\,(r_x+3\,r_y)+r_x\,r_z+3\,r_y\,r_z\right)\Big)\,\mathbf{E}[Y]\Big)\,\mathbf{E}[Z] + \\
& \quad m^2 f_{yz}\Big(g_{xy}^2\left(g_{yz}^2+3\,r_z\,(r_x+r_z)+2\,g_{yz}\,(r_x+2\,r_z)\right)+g_{yz}^2\,(r_x\,(2\,r_y+r_z)+r_y\,(r_y+2\,r_z))+ \\
& \qquad g_{yz}\left(r_x^2\,(r_y+r_z)+4\,r_y\,r_z\,(r_y+r_z)+2\,r_x\left(r_y^2+3\,r_y\,r_z+r_z^2\right)\right)+ \\
& \qquad r_z\left(r_x^2\,(2\,r_y+r_z)+r_y\,r_z\,(3\,r_y+2\,r_z)+r_x\left(3\,r_y^2+4\,r_y\,r_z+r_z^2\right)\right)+ \\
& \qquad g_{xy}\left(2\,g_{yz}^2\,(r_x+r_y+r_z)+2\,r_z\,(r_x+r_z)\,(r_x+3\,r_y+r_z)+g_{yz}\left(r_x^2+4\,r_x\,r_y+6\,r_x\,r_z+8\,r_y\,r_z+4\,r_z^2\right)\right)\Big) \\
& \quad \mathbf{E}[X]^2\,\mathbf{E}[Y]^2\,\mathbf{E}\left[Y^m\right]\mathbf{E}[Z]+m^4 f_{yz}^2\Big(r_z\,(r_x+r_z)\,(r_x+3\,r_y+r_z)+g_{yz}\left(r_x\,r_y+2\,r_x\,r_z+2\,r_y\,r_z+r_z^2\right)+ \\
& \qquad g_{xy}\,(3\,r_z\,(r_x+r_z)+g_{yz}\,(r_x+2\,r_z))\Big)\,\mathbf{E}[X]^2\,\mathbf{E}[Y]\,\mathbf{E}\left[Y^m\right]^2\mathbf{E}[Z] + \\
& \quad m^4 n^4 f_{xy}^2 f_{yz}^2 r_z\,\mathbf{E}\left[X^n\right]^2\mathbf{E}[Y]\,\mathbf{E}\left[Y^m\right]^2\mathbf{E}[Z]+m^6 f_{yz}^3 r_z\,(r_x+r_z)\,\mathbf{E}[X]^2\,\mathbf{E}\left[Y^m\right]^3\mathbf{E}[Z] + \\
& \quad m^6 n^2 f_{xy} f_{yz}^3 r_z\,\mathbf{E}[X]\,\mathbf{E}\left[X^n\right]\mathbf{E}\left[Y^m\right]^3\mathbf{E}[Z]+m^4 n^2 f_{xy} f_{yz}^2\,\mathbf{E}[X]\,\mathbf{E}\left[X^n\right]\mathbf{E}\left[Y^m\right]^2 \\
& \qquad \Big(f_{xy} f_{yz}\,\mathbf{E}\left[X^n\right]\mathbf{E}\left[Y^m\right]+(g_{yz}\,(r_y+2\,r_z)+r_z\,(2\,g_{xy}+2\,r_x+3\,r_y+2\,r_z))\,\mathbf{E}[Y]\,\mathbf{E}[Z]\Big)+ \\
& \quad m^2 n^4 f_{xy}^2 f_{yz}\,\mathbf{E}\left[X^n\right]^2\mathbf{E}[Y]\,\mathbf{E}\left[Y^m\right]\Big((g_{yz}\,(r_y+r_z)+r_z\,(2\,r_y+r_z))\,\mathbf{E}[Y]\,\mathbf{E}[Z]+f_{xy}\,\mathbf{E}\left[X^n\right]\left(f_{yz}\,\mathbf{E}\left[Y^m\right]+r_z\,\mathbf{E}[Z]\right)\Big)+ \\
& \quad m^2 n^2 f_{xy} f_{yz}\,\mathbf{E}[X]\,\mathbf{E}\left[X^n\right]\mathbf{E}[Y]\,\mathbf{E}\left[Y^m\right]\Big(\Big(g_{xy}^2\,r_z+g_{yz}^2\,(2\,r_y+r_z)+2\,g_{yz}\left(r_y^2+3\,r_y\,r_z+r_z^2+r_x\,(r_y+r_z)\right)+ \\
& \qquad r_z\left(3\,r_y^2+4\,r_y\,r_z+r_z^2+2\,r_x\,(2\,r_y+r_z)\right)+g_{xy}\,(g_{yz}\,(r_x+2\,r_y+3\,r_z)+r_z\,(2\,r_x+4\,r_y+3\,r_z))\Big)\,\mathbf{E}[Y]\,\mathbf{E}[Z]+ \\
& \qquad f_{xy}\,\mathbf{E}\left[X^n\right]\Big(f_{yz}\,(g_{xy}+g_{yz}+r_x+r_y+r_z)\,\mathbf{E}\left[Y^m\right]+\left(g_{yz}^2+2\,g_{yz}\,r_z+r_z\,(r_x+r_z)\right)\mathbf{E}[Z]\Big)\Big)\Big)
\end{aligned}
$$

$\gamma_{yz} =$

$$
\begin{aligned}
&m\Big(r_x\,(g_{xy}+r_x+r_y)\,(g_{yz}+r_z)\,(g_{yz}+r_x+r_z)\,(g_{xy}+g_{yz}+r_y+r_z)\,\mathbf{E}[X]^3\,\mathbf{E}[Y]^2\,\mathbf{E}[Z]\,+\\
&\quad n^6\,f_{xy}^3\,(g_{yz}+r_z)\,(g_{yz}+r_y+r_z)\,\mathbf{E}\big[X^n\big]^3\,\mathbf{E}[Y]^2\,\mathbf{E}[Z]+n^4\,f_{xy}\,(g_{yz}+r_z)\,\mathbf{E}[X]\,\mathbf{E}\big[X^n\big]\,\mathbf{E}[Y]^2\\
&\quad\Big(f_{xy}\,(g_{xy}\,(g_{yz}+r_x+r_z)+(g_{yz}+r_y+r_z)\,(g_{yz}+3\,r_x+r_y+r_z))\,\mathbf{E}\big[X^n\big]+g_{xy}^2\,(g_{yz}+r_y+r_z)\,\mathbf{E}[Y]\Big)\,\mathbf{E}[Z]\,+\\
&\quad n^2\,(g_{yz}+r_z)\,\mathbf{E}[X]^2\,\mathbf{E}[Y]^2\,\Big(f_{xy}\,\Big(g_{yz}^2\,(2\,r_x+r_y)+g_{xy}\,\big(2\,r_x^2+2\,r_x\,r_y+g_{yz}\,(3\,r_x+r_y)+3\,r_x\,r_z+r_y\,r_z\big)+\\
&\qquad (r_y+r_z)\,\big(3\,r_x^2+r_y\,r_z+2\,r_x\,(r_y+r_z)\big)+g_{yz}\,\big(3\,r_x^2+4\,r_x\,(r_y+r_z)+r_y\,(r_y+2\,r_z)\big)\Big)\,\mathbf{E}\big[X^n\big]+\\
&\qquad g_{xy}^2\,(g_{yz}+r_x+r_z)\,(g_{xy}+g_{yz}+r_y+r_z)\,\mathbf{E}[Y]\Big)\,\mathbf{E}[Z]+m^2\,n^6\,f_{xy}^3\,f_{yz}\,\mathbf{E}\big[X^n\big]^3\,\mathbf{E}[Y]\,\mathbf{E}\big[Y^m\big]\\
&\quad\big(f_{yz}\,\mathbf{E}\big[Y^m\big]+r_z\,\mathbf{E}[Z]\big)+m^4\,f_{yz}^2\,r_x\,(r_x+r_z)\,\mathbf{E}[X]^3\,\mathbf{E}\big[Y^m\big]^2\,\big(f_{yz}\,\mathbf{E}\big[Y^m\big]+r_z\,\mathbf{E}[Z]\big)\,+\\
&\quad m^4\,n^2\,f_{xy}\,f_{yz}^2\,(2\,r_x+r_z)\,\mathbf{E}[X]^2\,\mathbf{E}\big[X^n\big]\,\mathbf{E}\big[Y^m\big]^2\,\big(f_{yz}\,\mathbf{E}\big[Y^m\big]+r_z\,\mathbf{E}[Z]\big)\,+\\
&\quad m^4\,n^4\,f_{xy}^2\,f_{yz}^2\,\mathbf{E}[X]\,\mathbf{E}\big[X^n\big]^2\,\mathbf{E}\big[Y^m\big]^2\,\big(f_{yz}\,\mathbf{E}\big[Y^m\big]+r_z\,\mathbf{E}[Z]\big)+m^2\,f_{yz}\,r_x\,\mathbf{E}[X]^3\,\mathbf{E}[Y]\,\mathbf{E}\big[Y^m\big]\\
&\quad\Big(f_{yz}\,(g_{xy}+r_x+r_y)\,(g_{yz}+r_x+r_z)\,\mathbf{E}\big[Y^m\big]+\Big(g_{yz}^2\,(r_x+r_z)+r_z\,(r_x+r_z)\,(r_x+2\,r_y+r_z)\,+\\
&\qquad g_{xy}\,(2\,r_z\,(r_x+r_z)+g_{yz}\,(r_x+2\,r_z))+g_{yz}\,(2\,r_z\,(r_y+r_z)+r_x\,(r_y+3\,r_z))\Big)\,\mathbf{E}[Z]\Big)\,+\\
&\quad m^2\,n^4\,f_{xy}\,f_{yz}\,\mathbf{E}[X]\,\mathbf{E}\big[X^n\big]\,\mathbf{E}[Y]\,\mathbf{E}\big[Y^m\big]\,\Big(g_{xy}^2\,r_z\,\mathbf{E}[Y]\,\mathbf{E}[Z]+f_{xy}\,\mathbf{E}\big[X^n\big]\,\\
&\qquad\Big(f_{yz}\,(g_{xy}+g_{yz}+3\,r_x+r_y+r_z)\,\mathbf{E}\big[Y^m\big]+\Big(g_{yz}^2+g_{yz}\,(r_y+3\,r_z)+r_z\,(3\,r_x+2\,(r_y+r_z))\Big)\,\mathbf{E}[Z]\Big)\Big)\,+\\
&\quad m^2\,n^2\,f_{yz}\,\mathbf{E}[X]^2\,\mathbf{E}[Y]\,\mathbf{E}\big[Y^m\big]\,\Big(g_{xy}^2\,\Big(g_{yz}^2+2\,g_{yz}\,r_z+r_z\,(r_x+r_z)\Big)\,\mathbf{E}[Y]\,\mathbf{E}[Z]\,+\\
&\qquad f_{xy}\,\mathbf{E}\big[X^n\big]\,\Big(f_{yz}\,\big(2\,g_{xy}\,r_x+3\,r_x^2+2\,r_x\,r_y+g_{yz}\,(2\,r_x+r_y)+2\,r_x\,r_z+r_y\,r_z\big)\,\mathbf{E}\big[Y^m\big]\,+\\
&\qquad\Big(g_{yz}^2\,(2\,r_x+r_z)+g_{xy}\,(g_{yz}\,(r_x+r_z)+r_z\,(2\,r_x+r_z))\,+\\
&\qquad\quad r_z\,\big(3\,r_x^2+4\,r_x\,(r_y+r_z)+r_z\,(2\,r_y+r_z)\big)+2\,g_{yz}\,(r_z\,(r_y+r_z)+r_x\,(r_y+3\,r_z))\Big)\,\mathbf{E}[Z]\Big)\Big)\Big)
\end{aligned}
$$

with $D =$

$$
\begin{aligned}
&2\,\Big(n^6\,f_{xy}^3\,\mathbf{E}\big[X^n\big]^3\,\mathbf{E}[Y]\\
&\quad\Big(r_y\,(g_{yz}+r_z)\,(g_{yz}+r_y+r_z)\,\mathbf{E}[Y]^2+m^2\,f_{yz}\,(g_{yz}\,(r_y+r_z)+r_z\,(2\,r_y+r_z))\,\mathbf{E}[Y]\,\mathbf{E}\big[Y^m\big]+m^4\,f_{yz}^2\,r_z\,\mathbf{E}\big[Y^m\big]^2\Big)\,+\\
&\quad r_x\,\mathbf{E}[X]^3\,\Big((g_{xy}+r_x+r_y)\,(g_{yz}+r_x+r_z)\,\mathbf{E}[Y]+m^2\,f_{yz}\,(r_x+r_z)\,\mathbf{E}\big[Y^m\big]\Big)\\
&\quad\Big((g_{xy}+r_y)\,(g_{yz}+r_z)\,(g_{xy}+g_{yz}+r_y+r_z)\,\mathbf{E}[Y]^2\,+\\
&\qquad m^2\,f_{yz}\,(g_{yz}\,(r_y+r_z)+r_z\,(2\,r_y+r_z)+g_{xy}\,(g_{yz}+2\,r_z))\,\mathbf{E}[Y]\,\mathbf{E}\big[Y^m\big]+m^4\,f_{yz}^2\,r_z\,\mathbf{E}\big[Y^m\big]^2\Big)+n^4\,f_{xy}^2\,\mathbf{E}[X]\,\mathbf{E}\big[X^n\big]^2\,\\
&\quad\Big((g_{yz}+r_z)\,\Big(r_y\,(g_{yz}+r_y+r_z)\,(g_{yz}+3\,r_x+r_y+r_z)+g_{xy}\,\big(2\,r_x\,r_y+r_y^2+g_{yz}\,(r_x+2\,r_y)+r_x\,r_z+2\,r_y\,r_z\big)\Big)\,\mathbf{E}[Y]^3\,+\\
&\qquad m^2\,f_{yz}\,\Big(g_{yz}^2\,(2\,r_y+r_z)+r_z\,\big(3\,r_y^2+4\,r_y\,r_z+r_z^2+3\,r_x\,(2\,r_y+r_z)\big)\,+\\
&\qquad\quad g_{xy}\,(2\,r_z\,(r_x+r_y+r_z)+g_{yz}\,(r_x+r_y+2\,r_z))+g_{yz}\,\big(3\,r_x\,(r_y+r_z)+2\,\big(r_y^2+3\,r_y\,r_z+r_z^2\big)\big)\Big)\,\mathbf{E}[Y]^2\,\mathbf{E}\big[Y^m\big]\,+\\
&\qquad m^4\,f_{yz}^2\,(g_{yz}\,(r_y+2\,r_z)+r_z\,(g_{xy}+3\,r_x+3\,r_y+2\,r_z))\,\mathbf{E}[Y]\,\mathbf{E}\big[Y^m\big]^2+m^6\,f_{yz}^3\,r_z\,\mathbf{E}\big[Y^m\big]^3\Big)\,+\\
&\quad n^2\,f_{xy}\,\mathbf{E}[X]^2\,\mathbf{E}\big[X^n\big]\,\Big((g_{yz}+r_z)\,\Big(g_{xy}^2\,\big(r_x^2+g_{yz}\,(2\,r_x+r_y)+r_y\,r_z+2\,r_x\,(r_y+r_z)\big)+g_{xy}\,\Big(g_{yz}^2\,(r_x+r_y)\,+\\
&\qquad\quad 2\,r_x^2\,(2\,r_y+r_z)+r_y\,r_z\,(2\,r_y+r_z)+r_x\,\big(4\,r_y^2+6\,r_y\,r_z+r_z^2\big)+2\,g_{yz}\,\big(r_x^2+r_y\,(r_y+r_z)+r_x\,(3\,r_y+r_z)\big)\Big)\,+\\
&\qquad r_y\,\Big(g_{yz}^2\,(2\,r_x+r_y)+(r_y+r_z)\,\big(3\,r_x^2+r_y\,r_z+2\,r_x\,(r_y+r_z)\big)+g_{yz}\,\big(3\,r_x^2+4\,r_x\,(r_y+r_z)+r_y\,(r_y+2\,r_z)\big)\Big)\Big)\,\\
&\qquad\mathbf{E}[Y]^3+m^2\,f_{yz}\,\Big(g_{xy}^2\,(g_{yz}\,(r_x+r_z)+r_z\,(2\,r_x+r_z))+g_{yz}^2\,(2\,r_x\,(2\,r_y+r_z)+r_y\,(r_y+2\,r_z))\,+\\
&\qquad\quad g_{yz}\,\big(3\,r_x^2\,(r_y+r_z)+4\,r_y\,r_z\,(r_y+r_z)+4\,r_x\,\big(r_y^2+3\,r_y\,r_z+r_z^2\big)\big)\,+\\
&\qquad\quad r_z\,\big(3\,r_x^2\,(2\,r_y+r_z)+r_y\,r_z\,(3\,r_y+2\,r_z)+2\,r_x\,\big(3\,r_y^2+4\,r_y\,r_z+r_z^2\big)\big)+g_{xy}\,\Big(g_{yz}^2\,(2\,r_x+r_y+r_z)\,+\\
&\qquad\qquad r_z\,\big(4\,r_x^2+8\,r_x\,r_y+6\,r_x\,r_z+4\,r_y\,r_z+r_z^2\big)+g_{yz}\,\big(2\,r_x^2+5\,r_x\,r_y+8\,r_x\,r_z+5\,r_y\,r_z+2\,r_z^2\big)\Big)\Big)\,\mathbf{E}[Y]^2\,\mathbf{E}\big[Y^m\big]\,+\\
&\qquad m^4\,f_{yz}^2\,\Big(g_{yz}\,\big(2\,r_x\,r_y+4\,r_x\,r_z+2\,r_y\,r_z+r_z^2\big)+r_z\,\big(3\,r_x^2+6\,r_x\,r_y+4\,r_x\,r_z+3\,r_y\,r_z+r_z^2\big)\,+\\
&\qquad\quad g_{xy}\,(g_{yz}\,(r_x+r_z)+2\,r_z\,(2\,r_x+r_z))\Big)\,\mathbf{E}[Y]\,\mathbf{E}\big[Y^m\big]^2+m^6\,f_{yz}^3\,r_z\,(2\,r_x+r_z)\,\mathbf{E}\big[Y^m\big]^3\Big)\,\mathbf{E}[Z]
\end{aligned}
$$

The γ values that appear as coefficients in the individual Fano factors (Eqs (24)) are obtained in a similar manner. They are not given here due to their complexity.

Appendix B : Numerical Investigations

The oligomerization reactions with an intermediate state of lower oligomeric order depend on a wide range of parameters, which makes them complex to analyze even if the analytical solution is known. To limit the parameter space, we made the choice of considering only some parameters to be free and fixing the others. In particular, we analyzed in section 5 a CRN with fixed oligomerization degree, *i.e.* $n = 2$ and $m = 2$, which describes for example the tetramerization of monomeric proteins occurring through an intermediate dimerization step.

For analyzing numerically the mean and variances of the stochastic variables, we also fixed all production rates $p_x = p_y = p_z$ to be equal to 50 and all degradation rates $r_x = r_y = r_z$ and $g_{xy} = g_{yz}$ to be equal to 0.001. Other values have also been tested but did not lead to substantial differences in the interpretation of the results. The free parameters considered were f_{xy} and f_{yz} that describe the strength of the two dimerization reactions.

The first step consisted in analyzing the domain of existence of the analytic solutions at the steady state, where the mean numbers of molecules and variances are positive for all molecular species ($\mathbf{E}(X)$, $\mathbf{E}(Y)$, $\mathbf{E}(Z)$, $\mathbf{Var}(X)$, $\mathbf{Var}(Y)$, $\mathbf{Var}(Z) > 0$). The result of this analysis, namely the domain of existence of a solution in the $f_{xy} - f_{yz}$ plane, is plotted in Fig. (6) using logarithmically rescaled f-values.

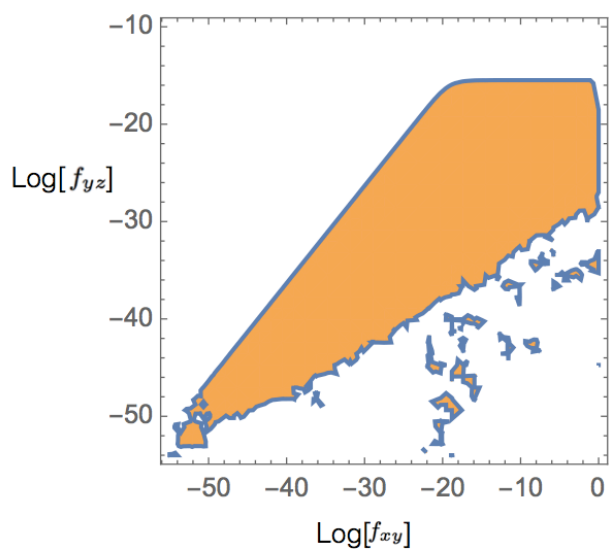
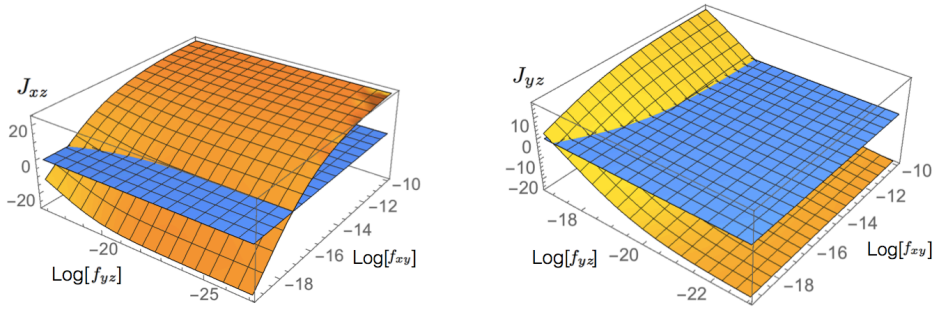


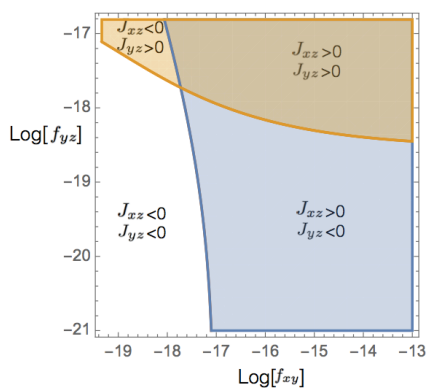
Figure 6: **Domain of existence of the steady state solution as a function of the logarithm of f_{xy} and f_{yz} for $p_x = p_y = p_z = 50$ and $r_x = r_y = r_z = g_{xy} = g_{yz} = 0.001$.**

The second step consisted in analyzing the sign of the two independent S-

fluxes, J_{xy} and J_{yz}, in this domain. The results are shown in Fig. 7a.



(a)



(b)

Figure 7: **Sign of the S-fluxes J_{xy} and J_{yz}.** (a) J_{xy} and J_{yz} as a function of the free parameters f_{xy} and f_{yz}; the blue surface is the plane that satisfies the equation $z = 0$, and the orange planes are the two S-fluxes. (b) Projection of the two surfaces onto the $z = 0$ plane. This projection divides the plane in four areas according to the signs of the S-fluxes.

By projecting the S-fluxes $J_{xy}(f_{xy}, f_{yz})$ and $J_{xy}(f_{xy}, f_{yz})$ onto the plane $z = 0$, we can see for which parameter values the sign of the fluxes are equal or differ (Fig. 7b). The four areas delimited in this way are referenced in the main text and in Fig. 5 as the four quadrants using Roman numerals.