Chapter 14

Stochastic Modular Analysis for Gene Circuits: Interplay Among Retroactivity, Nonlinearity, and Stochasticity

Kyung Hyuk Kim and Herbert M. Sauro

Abstract

This chapter introduces a computational analysis method for analyzing gene circuit dynamics in terms of modules while taking into account stochasticity, system nonlinearity, and retroactivity.

(1) *Analog electrical circuit representation for gene circuits*: A connection between two gene circuit components is often mediated by a transcription factor (TF) and the connection signal is described by the TF concentration. The TF is sequestered to its specific binding site (promoter region) and regulates downstream transcription. This sequestration has been known to affect the dynamics of the TF by increasing its response time. The downstream effect—retroactivity—has been shown to be explicitly described in an electrical circuit representation, as an input capacitance increase. We provide a brief review on this topic.

(2) *Modular description of noise propagation*: Gene circuit signals are noisy due to the random nature of biological reactions. The noisy fluctuations in TF concentrations affect downstream regulation. Thus, noise can propagate throughout the connected system components. This can cause different circuit components to behave in a statistically dependent manner, hampering a modular analysis. Here, we show that the modular analysis is still possible at the linear noise approximation level.

(3) *Noise effect on module input–output response*: We investigate how to deal with a module input–output response and its noise dependency. Noise-induced phenotypes are described as an interplay between system nonlinearity and signal noise.

Lastly, we provide the comprehensive approach incorporating the above three analysis methods, which we call *"stochastic modular analysis."* This method can provide an analysis framework for gene circuit dynamics when the nontrivial effects of retroactivity, stochasticity, and nonlinearity need to be taken into account.

Key words Stochastic process, Modularity, Synthetic biology, Noise propagation, Fan-out, Retroactivity

1 Introduction

Modularity is an important concept for engineering a composite system from a priori characterized components [1]. This concept enables efficient composition with predictability and

Mario Andrea Marchisio (ed.), *Computational Methods in Synthetic Biology*, Methods in Molecular Biology, vol. 1244,
DOI 10.1007/978-1-4939-1878-2_14, © Springer Science+Business Media New York 2015

module-interchangeability. In a real engineered system a connection between two components can cause interference, resulting in the loss of modularity. To minimize such interference, interfaces are designed to isolate any interfering effects. Synthetic biology is no exception. Connections between two biological components are mediated by transcription factors (TFs), which "bind" and "unbind" from their DNA specific binding sites. Thus, a circuit signal is "used" to establish a connection. In addition, gene circuits are significantly noisy due to the fact that TFs are often found in low copy numbers within single cells. Thus circuit signals—intracellular TF concentrations—can fluctuate in time significantly. The signal fluctuations—here we call "noise"—can propagate to the downstream gene-circuit components, making the signals in upstream and downstream statistically dependent, i.e., correlated.

In this chapter, we will provide a "modular" analysis method for gene circuits. First, we provide a brief review on our previous work on a gene circuit representation in terms of electrical circuit components, where component interference is represented explicitly by an increase in an input capacitance. Second, we describe noise propagation in a modular fashion. Third, the effect of noise on module properties is explained as an interplay between nonlinearity and stochasticity in the module. Lastly, these three different topics will be considered as a whole, to provide a modular framework for gene circuit dynamics that incorporate the effect of retroactivity, stochasticity, and system nonlinearity. This theoretical framework will provide a step toward understanding noise-induced phenotypes in terms of modular responses and will help exploit noise to enhance circuit performance.

2 Alternative Gene Circuit Representation in Terms of Electrical Circuit Components

This section represents gene circuits in terms of analog electrical circuit components: resistors and capacitors. To start with, we describe a circuit interface as a set of biological reactions that involve TFs that bind and unbind from their specific promoter sites to regulate downstream transcription processes. The TFs can be multimers and can self-regulate. Thus, the circuit interface can include a set of reactions of multimerization, TF binding and unbinding, transcription, and translation, as well as feedback reactions. For all these various types of interfaces, it was shown that the gene circuit reaction models can be mathematically mapped to analog electrical circuits [2] under the assumption that the binding and unbinding reactions are fast enough to be in equilibrium and bound TFs degrade much more slowly than the free TFs.

The effect of the downstream regulation by TFs has been experimentally shown to exist in protein signaling networks [3, 4] and gene regulatory networks [5–7]. Regulation can change the response speed of the regulating proteins as well as their concentration levels. These effects can be understood in terms of biological reactions. With respect to the speed of response, experimental verification in gene regulatory networks is still underway. The effect can be intuitively understood as follows: When bound TFs degrade more slowly than free TFs, perhaps because of the limited access of proteases, the degradation of the "total" TFs becomes slower. However, we have to note that if the bound TFs degrade with the same rate as the free form, the response speed will not change but the levels of the free TFs are just scaled down by the same factor for all transient time. This effect on the response speed was originally termed "retroactivity" [8] and the degree of tolerance "fan-out" [2].

The electrical circuit representation presented in [2] is more informative than the reaction model since the retroactive effect is explicitly shown in the former representation while the latter does not. Let us briefly review the work in [2]. We consider a protein synthesis and degradation process. Synthesis and degradation can be represented as a resistor–capacitor (RC) circuit and when the protein is allowed to regulate the downstream promoter sites, the RC circuit can be modified by adding an extra capacitor to the existing one in parallel as shown in Fig. 1. This parallel connection of the extra capacitor increases the total capacitance (that is the sum of the two capacitances), leading to a longer time to charge both the capacitors. Therefore, the response time increases. For a complex gene regulatory network the same mapping to the electrical circuit representation can be performed by adding an extra capacitor for each TF regulation. Thus, the retroactive effect in gene circuits can be systematically taken into account.

As an example, we can consider the repressilator, the first synthetic genetic oscillator. It is composed of three genes, *tetR*, *lacI*, and *cI*, which sequentially inhibit one another, making a negative feedback loop. This delayed negative feedback leads to oscillation in gene expression. Since there are three regulation reactions, three interfaces exist. For each interface, we need to add a capacitor in parallel with the existing one. Here we note that the additional capacitance is dependent on the voltage—TF concentration—applied and thus that the dependency of retroactivity on the TF concentration is taken care of. Thus, the electrical circuit representation of gene circuits can be illustrated as in Fig. 1.

In summary, when a gene circuit module makes a connection to another, the input capacitance of the downstream module will be increased depending on the concentration of TFs and the number of TF binding sites in the downstream module as shown

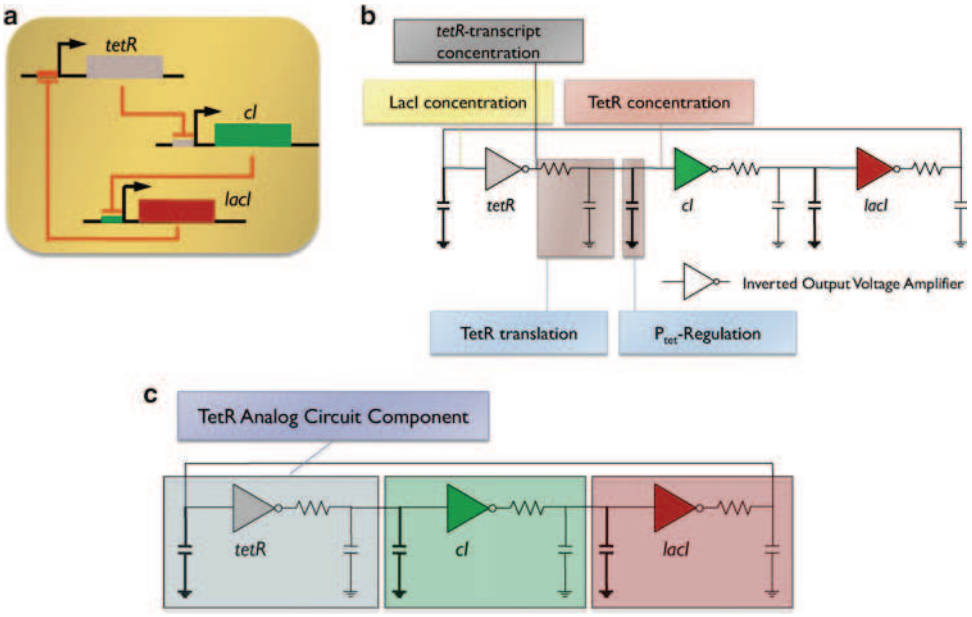

Fig. 1 Analog electrical circuit representation: (**a**) The repressilator is the first synthetic genetic oscillator. (**b**) It can be represented in terms of analog electrical circuit components such as resistors, capacitors, and inverted output voltage amplifiers. Here the inverted amplifiers represent the inhibitory genetic regulation of TetR, cI, and LacI, by inverting, scaling, and shifting the input voltages. Retroactivity can be explicitly described as an additional capacitor that depends on voltage applied. (**c**) Each gene expression and its regulation can be represented as an electrical circuit block or module, and retroactivity as an input capacitance increase of the module

in Figs. 1 and 2. Further analysis shows that the dependence on the number of TF binding sites (P_T in Fig. 2) turns out to be linear, which will be discussed in the next section.

3 Gene Circuit Fan-Out

For the case when retroactivity is significant, how can we quantify the degree of the retroactivity and furthermore provide certain limits in the downstream load to cap the retroactive effect? Electrical circuit components cannot tolerate unlimited connections. This is because with a finite amount of output current it is impossible to drive all the connected components. A similar phenomenon can occur in gene circuits. Thus, it is important to know how many downstream components can be connected, a number we call *fan-out* [2].

For example, consider an upstream module that is an oscillator, which needs to produce oscillatory signals faster than a certain frequency due to a biological limitation or desired design requirement. This maximum operating frequency will limit the number of the connections from the oscillator to downstream circuits because

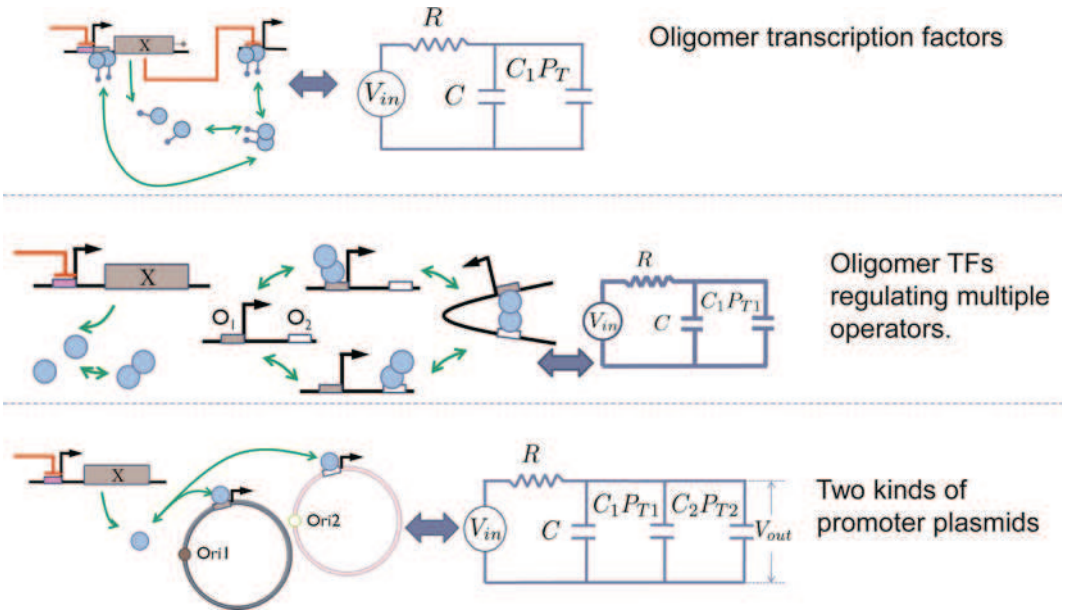

Fig. 2 Gene circuit representation in a wide range of circuit interfaces. *Top* Dimer transcription factors self-regulate their transcription processes, while regulating the downstream promoters as well. This interface can be transformed to an RC-circuit. Here the additional capacitance C_1P_T represents the input capacitance of the downstream circuit. Note that this capacitance is proportional to the number of the downstream promoters, P_T. *Middle* Dimer transcription factors regulate multiple operators (O_1 and O_2). *Bottom* Transcription factors regulate two types of plasmids having different copy numbers, P_{T1} and P_{T2}. In all the above cases and even the combinations of these, the additional capacitors are proportional to the number of the downstream promoters (P_T, P_{T1}, P_{T2}). This linearity can provide efficient characterization of gene circuit fan-out

as the number of the downstream interactions increase, the response time of the regulating protein increases due to retroactivity.

How can we measure the fan-out? Obviously the measurement requires the trend information on the response time vs. number of connections. If we have to repeat experiments for many different numbers of downstream connections, the characterization would be very tedious and perhaps even impractical. However, it was shown that the response time of the TFs show a certain linear relationship to the number of the downstream TF connections. This linearity was shown to be universal if each individual downstream promoter acts as an independent sequestrator of the TFs. Such linearity can be established for a wide range of circuit interface such as multimer TFs, feedback, multiple kinds of promoters, and multiple operator sites as shown in Fig. 2. Thus, fan-out can be characterized very efficiently due to the linearity.

What kinds of experiments need to be performed to measure fan-out? Two kinds, one at the population level using micro-plate readers and the other at the single cell level using fluorescence microscopy. The population level experiments can be performed, for example, by varying the strength of the downstream

connections of the TFs either by adding small molecules that interact with the TFs or by changing the number of promoter sites that are specific to the TFs [5–7]. Single-cell level experiments can be performed by tagging TFs with fluorescent markers to directly monitor the fluctuations of the TF concentrations and by measuring the fluctuation speed (more precisely correlation time). For discussion on experimental characterization of response speed by using gene expression noise, refer to [9], and for more theoretical discussion, refer to [2, 10, 11].

4 Modular Description of Signal Noise

As briefly discussed in the previous section, gene circuits are noisy. Even within an isogenic cell population, each individual cell can show significantly different protein concentration levels. This is because transcription and translation events occur as a series of random bursts and the cells undergo division, differentiation, and apoptosis at random and even extra-cellular environments can show significant fluctuations. Due to the random nature of cellular dynamics and environmental effects, gene expression systems are often described as "stochastic" processes.

One important property of noisy systems is that noise passing through one system component can propagate to another and can lead to nontrivial "interference" between the components. Sometimes, this interference can be significant enough that the system properties are drastically changed, resulting in "noise-induced phenotypes."

As a first step to understand noise-induced phenotypes, we decompose the noise propagation in a modular way so that the phenotypes can be understood in terms of each module component effect. For this task, we propose the following modular analysis (*see* Subheading 5). This approach is based on a very simple fact that when noise is sufficiently small, the noise dominantly sees the linear components of nonlinear systems. In this case, a sinusoidal wave input (with a certain frequency ω) will propagate to an output with an identical-frequency sinusoidal wave without exciting other-frequency signals (because noise components can see only the "linear" system). This fact implies that in the frequency domain the output signal noise can be fully described by the input signal noise. Therefore, a modular description of noise propagation can be performed at the "linear noise approximation."

We will take the Langevin approach with the linear noise approximation [12, 13]. We will formulate the input–output noise response of a module by investigating self-correlation of each individual input and output noise signal [14, 15]. More precisely, the self-correlation can be mathematically described by an autocorrelation function that quantifies similarity between the values of

signal pairs with a certain time lag. It turns out that the input autocorrelation function is sufficient to describe the output autocorrelation function if the module itself is fully characterized a priori.

5 Modular Description of Noise Propagation

To describe how internal and external (input) noise propagate through networks, we will investigate the properties of the autocorrelation functions of the noise. In particular, it can be shown that if the input noise autocorrelation function is known, the output noise autocorrelation function can be fully described by the input noise autocorrelation function for the stationary state of the internal dynamics, which are described by linear Langevin equations [16]. Furthermore, by taking the Fourier transforms of the autocorrelation functions, it can be shown the input–output noise response can be described by a transfer-like matrix equation. This analysis leads to modular analysis of complex stochastic networks.

The input signals of a module are denoted as $\boldsymbol{X}$, its output as $\boldsymbol{\Upsilon}$, and internal variables of the module as $\boldsymbol{Z}$. Internal dynamics is presumed to be described by

$$\frac{d\boldsymbol{Z}}{dt} = \boldsymbol{J} \cdot \boldsymbol{Z} + \xi + \boldsymbol{X}, \tag{1}$$

where ξ is a Gaussian white noise vector satisfying $\langle \xi_t \xi_{t'}^T \rangle = \boldsymbol{D}\delta(t - t')$ with $\boldsymbol{D}$ diffusion coefficient matrix and $\boldsymbol{J}$ is the Jacobian matrix. We assume that the correlation between the input noise and the internal noise (ξ) is negligible. We introduce an autocorrelation function to quantify the similarity between signal pairs separated by a time lag τ:

$$\boldsymbol{G}_Z(\tau) = \lim_{t_0 \to \infty} \langle \boldsymbol{Z}_{t_0} \boldsymbol{Z}_{t_0+\tau}^T \rangle. \tag{2}$$

Previous research [14, 15, 17, 18] has been focused on the properties of the diagonal terms of the above matrix. This is why the general structure of input–output noise relationship has been absent. The solution of Eq. 1 is obtained as

$$\boldsymbol{Z}_t = \exp[\boldsymbol{J}t] \int_0^t ds \ \exp[-\boldsymbol{J}s] \cdot (\xi_s + \boldsymbol{X}_s). \tag{3}$$

The autocorrelation matrix, expressed in terms of the correlation matrix of ξ and $\boldsymbol{X}$ by substituting Eq. 3 to Eq. 2,

can be simplified after Fourier transformations as:

$$\tilde{\boldsymbol{G}}_Z(\omega) = \left(\boldsymbol{J} + i\omega\boldsymbol{I}\right)^{-1} \cdot \left(\boldsymbol{D}/(2\pi) + \tilde{\boldsymbol{G}}_X(\omega)\right) \cdot \left(\boldsymbol{J}^T - i\omega\boldsymbol{I}\right)^{-1},$$

using $\langle(\xi_s + \boldsymbol{X}_s)\cdot(\xi_{s'} + \boldsymbol{X}_{s'})^T\rangle = \boldsymbol{D}\delta(s-s') + \boldsymbol{G}_X(s-s')$. $\tilde{\boldsymbol{G}}(\omega)$ is the Fourier transform of $\boldsymbol{G}(t)$. Finally, the output noise autocorrelation function can be expressed, by using a projection matrix $\boldsymbol{P}$ mapping the state space $\{\boldsymbol{Z}\}$ onto the output subspace $\{\boldsymbol{\Upsilon}\}$, as

$$\tilde{\boldsymbol{G}}_\Upsilon(\omega) = \boldsymbol{P}\cdot\tilde{\boldsymbol{G}}_Z(\omega) = \boldsymbol{P}\cdot\left(\boldsymbol{J}+i\omega\boldsymbol{I}\right)^{-1}\cdot\left(\frac{\boldsymbol{D}}{2\pi}+\tilde{\boldsymbol{G}}_X(\omega)\right)\cdot\left(\boldsymbol{J}^T-i\omega\boldsymbol{I}\right)^{-1}. \tag{4}$$

Equation 4 is important because it shows that the output autocorrelation function can be fully described by the input autocorrelation function and that there is a simple matrix relationship through the transfer-like function. Equation 4 also succinctly shows how output noise is affected by internal noise ($\boldsymbol{D}$) and input noise ($\tilde{\boldsymbol{G}}_X(\omega)$). This allows a systematic approach for noise propagation in complex gene and signaling networks using the *modularity* concept.

To take into account the effect of retroactivity in this modular analysis, we can consider an additional capacitance that corresponds to the effect of downstream loads and analyze the noise propagation with the input capacitance taken into account as illustrated in Fig. 1.

In this analysis method, we can understand the input–output noise response of a module for different cases such as (a) non-stochastic input and stochastic internal noise ($\tilde{\boldsymbol{G}}_X(\omega)=0$); (b) stochastic input and non-stochastic internal noise ($\boldsymbol{D}=0$ and $\tilde{\boldsymbol{G}}_X(\omega)\neq 0$); (c) both internal and external noise ($\boldsymbol{D}\neq 0$ and $\tilde{\boldsymbol{G}}_X(\omega)\neq 0$).

This approach can also be used to devise noise (de-)amplifiers and filters by taking into account different network topology and kinetics, for which the information can be found in the matrices $\boldsymbol{D}$ and $\boldsymbol{J}$. If the matrices are suitably chosen, noise (de-)amplifiers and filters can be designed.

6 Noise Effect on Nonlinear Responses

In the previous section, noise propagation was described in a modular fashion with the approximation that the first and second moments of a signal noise are sufficient to describe the system behaviors. Under the same level of approximation, we will describe the effect of noise on a module input–output transfer curve. First, consider a linear system with an input (i) and an output (o), satisfying the following relationship:

$$o = c\ i,$$

with c a proportionality constant. The mean values of i and o satisfy the same relationship:

$$\langle o \rangle = c\langle i \rangle.$$

However, when a system is nonlinear, the same relationship does not hold typically, simply because of Jensen's inequality: In the convex function $f(x)$, the function of the mean of the two x values, $f([x_1 + x_2]/2)$, is always smaller than the mean value of the two functional values of x_1 and x_2, $[f(x_1) + f(x_2)]/2$, because this is located on the secant line (*see* Fig. 3). In more general term, the expectation of a convex function of a random variable is always greater than the function of the expectation. Jensen's inequality explains that a sigmoidal response curve can be less sigmoidal (as shown in Fig. 3b) and in certain cases, can be more sigmoidal, resulting in ultrasensitivity [19, 20]. This has been shown that the interplay between system nonlinearity and stochasticity can cause nontrivial noise-induced phenotypes such as noise-induced bistability and noise-induced linearized response [20].

More mathematically, the expectation value of $f(x)$ can be approximated as

$$\langle f(x) \rangle \simeq f(\langle x \rangle) + \frac{f''(x)}{2} Variance(x). \tag{5}$$

This gives a simple explanation of Jensen's inequality by showing the sign dependence of $f''(x)$. The second term in Eq. 5 describes the interplay between nonlinearity and stochasticity: The double derivative $f''(x)$ corresponds to the former and the variance to the latter. The variance and mean values of x can be estimated with a moment closure approximation by neglecting the third and higher moments of x (refer to Appendix in [20]). This approximation is called the mass fluctuation kinetics in mass reaction systems [21].

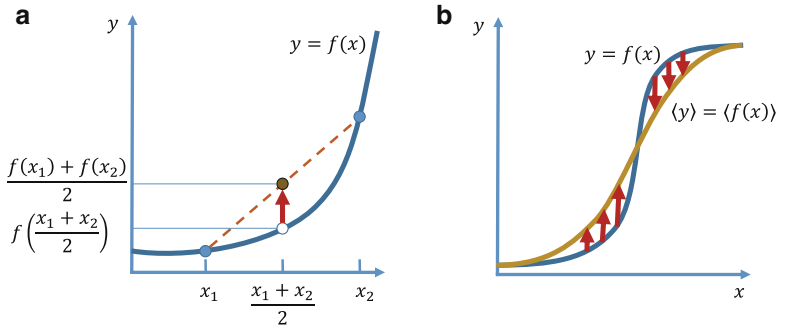

Fig. 3 Interplay between system nonlinearity and stochasticity

7 Stochastic Modular Analysis

We combine the modular noise propagation analysis provided in Subheading 5 to the mass fluctuation kinetics to consider the cases in the stronger noise regime where noise-induced phenotypes can appear. Mass fluctuation kinetics can describe the system behavior with more accuracy than linear noise approximation approaches by taking into account an interplay between system nonlinearity and stochasticity exactly as in Eq. 5. The variance (more in general, covariance) term in Eq. 5 can be described in terms of the input signal values in the modular noise propagation analysis, $\tilde{G}_x(\omega)$, as follows:

$$Variance(x) = \int \tilde{G}_x(\omega)\, d\omega \ .$$

This shows that it is possible to combine the modular noise propagation analysis with the mass fluctuation kinetics (*see* Fig. 4). Therefore, network-level dynamics can be described in terms of modules.

To take into account the effect of retroactivity, an additional input capacitance can be added per a module input that corresponds to the effect of the input to its connected output. Then, the noise propagation can be described in terms of modules as shown in Fig. 4.

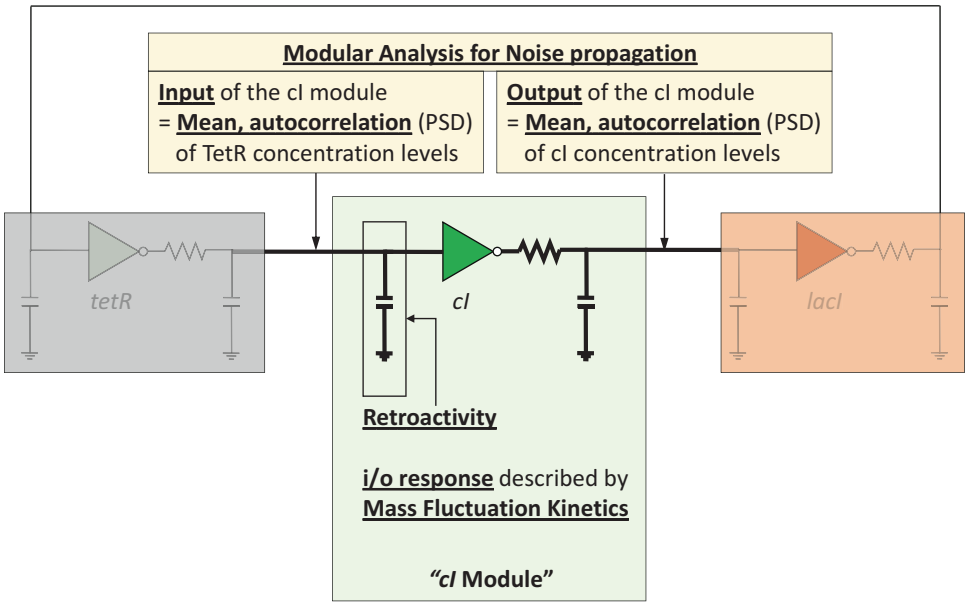

Fig. 4 Stochastic modular analysis: As a simple genetic circuit example, the repressilator is represented in terms of analog circuits. The noisy signals are analyzed in terms of modules by considering its first and second moments: mean and autocorrelation, or equivalently power spectral density (PSD). A module input–output (i/o) response is analyzed by using mass fluctuation kinetics by considering the system nonlinearity and stochasticity. The retroactivity is explicitly described by an additional parallel capacitance connection

Therefore, it is possible to combine the three different analyses that were discussed in the previous sections: (1) analog circuit representation for the explicit description of retroactivity; (2) modular description of noise propagation; (3) the effect of system nonlinearity and stochasticity. This combined global analysis, we name stochastic modular analysis, will provide an analysis framework that takes into account *retroactivity, stochasticity, system nonlinearity, and modularity*, all together in a system-wide level.

References

1. Baldwin CY, Clark KB (2000) Design rules: the power of modularity. MIT Press, Cambridge
2. Kim KH, Sauro HM (2010) Fan-out in gene regulatory networks. J Biol Eng 4:16
3. Jiang P, Ventura AC, Sontag ED, Merajver SD, Ninfa AJ, Del Vecchio D (2011) Load-induced modulation of signal transduction networks. Sci Signal 4(194):ra67
4. Ventura AC, Jiang P, Van Wassenhove L, Del Vecchio D, Merajver SD, Ninfa AJ (2010) Signaling properties of a covalent modification cycle are altered by a downstream target. Proc Natl Acad Sci USA 107:10032–10037
5. Buchler NE, Cross FR (2009) Protein sequestration generates a flexible ultrasensitive response in a genetic network. Mol Syst Biol 5:272
6. Jayanthi S, Nilgiriwala KS, Del Vecchio D (2013) Retroactivity controls the temporal dynamics of gene transcription. ACS Synth Biol
7. Daniel R, Rubens JR, Sarpeshkar R, Lu TK (2013) Synthetic analog computation in living cells. Nature 497(7451):619–623
8. Del Vecchio D, Ninfa AJ, Sontag ED (2008) Modular cell biology: retroactivity and insulation. Mol Syst Biol 4:161
9. Weinberger LS, Dar RD, Simpson ML (2008) Transient-mediated fate determination in a transcriptional circuit of HIV. Nat Genet 40 (4):466–470
10. Kim KH, Sauro HM (2011) Measuring retroactivity from noise in gene regulatory networks. Biophys J 100(5):1167–1177
11. Kim KH, Sauro HM (2012) Measuring the degree of modularity in gene regulatory networks from the relaxation of finite perturbations. In: 2012 I.E. 51st IEEE conference on decision and control (CDC), pp 5330–5335
12. Elf J, Ehrenberg M (2003) Fast evaluation of fluctuations in biochemical networks with the linear noise approximation. Genome Res 13 (11):2475–2484
13. Paulsson J (2004) Summing up the noise in gene networks. Nature 427(6973):415–418
14. Tănase-Nicola S, Warren PB, ten Wolde PR (2006) Signal detection, modularity, and the correlation between extrinsic and intrinsic noise in biochemical networks. Phys Rev Lett 97(6):68102
15. Warren PB, Tanase-Nicola S, ten Wolde PR (2006) Exact results for noise power spectra in linear biochemical reaction networks. J Chem Phys 125(14):144904
16. Kwakernaak H, Sivan R (1972) Linear optimal control systems. Wiley-Interscience, New York
17. Simpson ML, Cox CD, Sayler GS (2003) Frequency domain analysis of noise in autoregulated gene circuits. Proc Natl Acad Sci USA 100(8):4551
18. Austin DW, Allen MS, McCollum JM, Dar RD, Wilgus JR, Sayler GS, Samatova NF, Cox CD, Simpson ML (2006) Gene network shaping of inherent noise spectra. Nature 439:608–611
19. Paulsson J, Berg OG, Ehrenberg M (2000) Stochastic focusing: fluctuation-enhanced sensitivity of intracellular regulation. Proc Natl Acad Sci USA 97(13):7148–7153
20. Kim KH, Qian H, Sauro HM (2013) Nonlinear biochemical signal processing via noise propagation. arXiv:1309.2588 [q-bio.QM]
21. Gómez-Uribe CA, Verghese GC (2007) Mass fluctuation kinetics: capturing stochastic effects in systems of chemical reactions through coupled mean-variance computations. J Chem Phys 126(2):24109