Extrinsic noise of the target gene governs abundance pattern of feed-forward loop motifs

Md Sorique Aziz Momin* and Ayan Biswas†
Department of Chemistry, Bose Institute, 93/1 A P C Road, Kolkata 700009, India

(Received 4 November 2019; revised manuscript received 2 March 2020; accepted 7 May 2020; published 28 May 2020)

Feed-forward loop (FFL) is found to be a recurrent structure in bacterial and yeast gene transcription regulatory networks. In a generic FFL, transcription factor (TF) S regulates production of another TF X while both of these TFs regulate production of final gene-product Y. Depending upon the regulatory programs (activation or repression), FFLs are grouped into two broad classes: coherent (C) and incoherent (I), each class containing four distinct types (C1–C4 and I1–I4). These FFL types are experimentally observed to occur with varied frequencies, C1 and I1 being the abundant ones. Here we present a stochastic framework singling out the absolute value of the normalized covariance of X and Y to be the determining factor behind the abundance of FFLs while considering differential promoter activities of X and Y. Our theoretical construct employs two possible signal integration mechanisms (additive and multiplicative) to synthesize Y while steady-state population level of S remains fixed or becomes tunable reflecting two possible environmental signaling scenarios. Our model categorically points out that abundant FFLs exhibit higher amount of the designated metric which has a biophysical connotation of extrinsic noise for the target gene Y. Our predictions emanating from an overarching analytical expression utilizing biologically plausible parametric conditions are substantiated by stochastic simulation.

DOI: 10.1103/PhysRevE.101.052411

I. INTRODUCTION

A. Noise in living systems

Living systems are always flooded with varied environmental signals often embedded in noisy profiles. These stochastic signals hold the pivotal roles of initiating shedload of biochemical reactions which are essential for survival of these organisms. However, inside a living cell, biochemical noise itself is capable of enabling useful functionalities. This latter type of noise has been ubiquitously associated with the limited amount of molecules produced in and controlling probabilistic cellular processes, e.g., the gene products constituting a gene transcription regulatory network (GTRN). The randomized production and degradation of these small numbers of biochemical species elicit considerable fluctuations in their concentrations and has been termed as intrinsic noise. The associated variance normalized by the corresponding squared expression level (say N) has been found to scale as $\sim N^{-1}$. This may lead to evolvable trait of phenotypic variability in isogenic population. Noise is also helpful to widen stochastic strategies concerning differentiation and cell-fate determination in population level across uni- and multicellular living machines [1–8]. Noise level is also found to differentially affect proteins depending upon the physiological processes they are involved in. Essential proteins and those tending to form complexes and involved in synthesizing proteins, proteasomal components are hit by relatively low noise whereas proteins interacting with environmental signals and fostering appropriate responses, e.g., stress-related gene expressions are high in noise [9–11]. A pioneering study in *Escherichia coli* using two identical and independent reporters have showed that apart from the intrinsic noise, there exists another type of noise manifested by the shared environment resulting in correlation between the two promoters. This is referred as the extrinsic noise measured by the normalized covariance of the reporters. In that study, the extrinsic noise production was attributed to a pathway-specific repressor [12]. It has been shown that in genetic networks of *E. coli*, apart from the intrinsic noise other types of fluctuations, e.g., those having been originated in upper-level genes of the network tend to get transmitted in downstream genes whereas there is also noise of global nature affecting all the constituent genes in the network [13]. Concentration fluctuations of ribosomes, polymerases, etc., contribute to extrinsic noise [14]. Results have been also obtained to showcase invariant modeling of noise both in prokaryotes and eukaryotes [3]. In Results and Discussion, we will get back to the concepts of intrinsic and extrinsic noise from the perspective of mathematical modeling [14–17] to make sense of implications emanating from our analytical proposition.

It is to be noted that the chemical kinetics of biological processes can be effectively described by writing the chemical master equation but unfortunately exact analytical solution can be obtained only in a limited number of cases. However, stochastic simulation which is exact happens to be increasingly expensive in computational sense as the number of interactions grows. Hence, developing approximate analytic methods to deal with these systems is a hotbed of research. Interested readers may consult a self-contained article by

*soriqueaziz@jcbose.ac.in
†Corresponding author: ayanbiswas19@gmail.com

Schnoerr *et al.* where they tirelessly discuss basic concepts of deterministic and stochastic modeling approaches, exact numerical methods and compare several approximation methods with respect to the former marking their up- and down-sides. As an added bonus, they also discuss Bayesian inference problem from an empirical standpoint and review recent literature [18].

B. Gene transcription and its network-based interpretation

Now, we exemplify the concept of a typical system-environment interaction by invoking the case of a bacterium performing chemotaxis in search for food [7]. In this process, it has to reliably sense the appropriate external chemical cue so that it can synthesize appropriate enzymes in right proportions to chew up the food. In this very searching phase, it has to swim across the natural medium using a motored flagella which is also produced depending upon the external signaling conditions. Both of the enzyme and the flagella motor protein are synthesized when the appropriate signals interact with the bacterial genes which hold the informational blueprints of various essential life-products [19]. This whole process is broadly composed of a transcription step (making mRNA from DNA) and a translation step (producing enzymes and proteins from mRNA). While some of the proteins are directly recruited in physiological processes, e.g., making body parts, others, namely, TFs are responsible for mediating the interaction between the intra or extracellular signals and the genes, coding regions of DNA. TFs play a crucial role in controlling the rate at which mRNA is produced from the gene. TFs can increase (decrease) the production rate of mRNA by enabling (disabling) the efficiency of RNA polymerase in reading off the genetic information and therefore are known as activators (repressors). An activator in the presence of suitable inducer, becomes active and binds to the promoter region which is a noncoding stretch of a DNA, situated upstream of a gene and thereby up-regulate its rate of transcription. However, a repressor depending upon the signaling scenario, may get deactivated and falls off the genetic promoter so that the mRNA production rate can be enhanced. The resulting gene-products may also act as TFs targeting other genes and therefore create a network (i.e., GTRN) composed of genes as nodes and transcriptional regulatory modes as edges [19]. In a typical bacterial cell, genes in their productive forms are found in maximum one copy, mRNA in a few copies, proteins in $\sim$100 copies and therefore low copy number generated fluctuations are inherent in GTRN. In a GTRN, $S \rightarrow X$ symbolizes that product of gene S acts as an activator for gene X. However, $S \dashv X$ means gene-product S acts as a repressor for gene X. The elaborate repertoire of bacterial GTRN involves cascades, loops which may involve auto-regulation where the TF activates or represses its own production. Due to enormity of interactions, reductionism had little role to play while analyzing GTRN until recently when researchers found out that some of its interconnections occur with higher frequencies compared to a random network. These recurring network patterns, known as "network motifs" actually function as basic building blocks in the vast structural scheme of a GTRN [20–22].

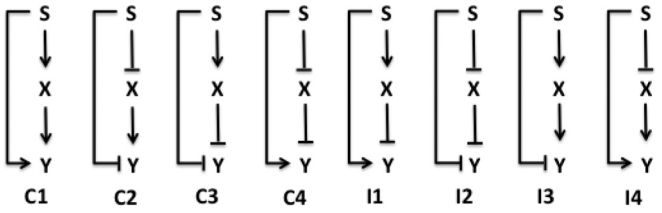

FIG. 1. Schematic diagrams of coherent (C1–C4) and incoherent (I1–I4) FFL motifs. Here S regulates X and both of them regulate Y [19,23].

C. FFL motif—An essential gene-transcriptional machinery

FFL is found to be the only network motif in GTRN of *E. coli* and *Saccharomyces cerevisiae* with respect to the ensemble of total thirteen subgraphs that can possibly exist in a random network consisting of three nodes with directed edges [20]. To form the FFL, a master regulator (S) has to activate or repress another TF (X) which works as a coregulator with S for the target gene Y, up or down-regulating synthesis of its gene-product. Depending upon the two regulatory modes (activation or repression) for each of the three edges, a total of $2^3 = 8$ combinations arise, designated as eight distinct types of the FFL [23]. Assigning $+$ $(-)$ signature for activation (repression), the effective regulatory signature of the indirect branch (from S to Y via X) is determined by product of the signatures of its two steps. If it matches with the signature of the direct branch (from S to Y), then the FFL is categorized as a coherent (C) type, the other possibility being an incoherent (I) type. Thus, FFLs are subdivided into four coherent (C1–C4) and four incoherent (I1–I4) types (see schematic Fig. 1) [19,23]. Regulation mediated by the master regulator (S) and intermediate regulator (X) have been found to converge upon the effector gene (Y) using diverse mechanisms, e.g., OR, AND, SUM gates, etc. Analysis of GTRN databases in model systems of *E. coli* and yeast has unearthed the abundance statistics of all these FFL types which clearly shows that C1 and I1 FFLs are the most abundant types [19,24]. C1 FFL has three activation edges whereas in I1 FFL, only the edge between X and Y is inhibitory in nature. The arabinose system of *E. coli* makes use of C1 FFL where signal cyclic adenosine monophosphate (cAMP) is sensed by cAMP receptor protein or CRP (S). Another signal, L-arabinose which is an alternate energy source for the bacterium, triggers TF araC (X). The target operon *araBAD, araFGH* synthesize enzymes and proteins for utilizing and transporting arabinose, respectively. The system utilizes an AND logic [25]. C1 FFL with an OR gate is found to operate in *E. coli* where the effector genes *fliLMNOPQR* (Y) produce the proteins that build the motor of the bacterial flagella. The upper- and intermediate-level activators are FlhDC (S) and FliA (X), respectively. Different factors like limiting concentration of glucose, pressure built up due to osmosis, and the temperature act as potential signals to S. Production of first few motors generates signaling check point for X [26]. Example of I1 FFL in *E. coli* is its galactose utilization system where CRP (S) induced by cAMP activates GalS (X) which is inactivated by galactose signal. In the absence of galactose, GalS represses target genes *galETK* which is activated by

CRP. The products of the target genes derive energy and carbon for the bacterium out of galactose [24]. Besides these core experimental results, some prominent theoretical analysis on FFL include Ref. [27] dealing with genetic nonloop types of FFL focusing upon optimization of information and Ref. [28] where the filtering capacity of FFL in response to temporal variation of the signal has been studied.

D. Previous deliberations on FFL

The sheer difficulty in understanding the FFL abundance statistics has been multifaceted and substantial. To exemplify, exploration of FFLs in GTRN has been proved to be a challenging task and the corresponding database is continuously getting updated. A number of groups have meticulously studied these recurrent subgraphs from an evolutionary point of view, whereas others concentrated on their dynamical behaviors. Before proposing our current theoretical construct we briefly review these interconnected themes to showcase the conceptual difficulty placed in the way of getting any all-inclusive single physical principle to explain this relative abundance statistics problem of FFLs.

E. Network-level understanding of FFL

One of the important issues in this regard is the identification of any direct impact that structure of the network motif bears to its functionality. Ingram *et al.* in their elaborate study concentrating upon bi-fan motifs (which are more complex than FFLs) have pointed out that depending upon kinetic rate parameters, same topological motif can exhibit a range of dynamical behaviors. The problem is compounded by the fact that the detailed experimental knowledge about these key parameters are limited. This lack of information may also underestimate the interaction that the motifs may have with the rest of a larger network within which they are embedded [29]. The preceding point is in fact a major roadblock in articulating the precise role that a particular type of modular structure plays as some sort of building block for the larger networks. Lack of integration among different databases and literature concerning detailed mapping of interactions in GTRN has also deterred accurate prediction on functional motifs. Ma and colleagues have pointed out that in *E. coli*, detailed exploration of the GTRN have altered the relative abundance statistics of different types of coherent and incoherent FFLs while it is still true that type-1 in both the above mentioned classes of FFLs are over-represented than the rest of FFL types. Interestingly, they have taken a note of the situation where definitional aspects of a node which, may represent either a gene or an operon can potentially result in under or overestimation of the FFLs present in the network which shows massive clustering of these patterns. Apart from this, another surprising finding from them was the occurrence of regulation by more than one FFL and even other nonmodular regulators. Now in this situation, the dynamics shown by the composite regulatory unit is bound to be significantly different from that of an isolated FFL [30]. In the context of FFL clustering, a recent empirical work by Gorochowski and colleagues [31] have intensely focused upon the concept of architecture driven functionality where the environmental conditions that a cell experiences and tries to get adapted with, in principle dictate the topological clustering rules. Their studies on metabolic network of *E. coli*, have identified the nodes which play major roles in FFL aggregates. These nodes while undergoing duplication and divergence type of environmental interventions are naturally placed under stringent selection pressure, albeit subjected to the adaptability criteria in dynamic settings.

F. Studies on the FFL evolution

Coming to the evolutionary standpoint to understand abundance of FFLs in bacterial GTRN, Tsoy *et al.* in their genomic analysis of *E. coli*, concluded that it is the combination of different aspects of gene regulation, e.g., TFs, genes, and TF binding sites that can determine conservation of network motifs while acted upon by evolutionary force. Their efficient approach predicted that C1 FFL is possibly less conserved than I1 FFL because in the former case, both global and local regulation being of the same modality, the motif is endowed with functional redundancy [32]. In Ref. [33], the authors use a sufficiently detailed mutation-based model involving delays and stochasticity in GTRN to show that a C1 FFL evolves to filter out short spurious signals. A closely related four-node diamond pattern also evolves in this situation but in the absence of spurious signals, only the latter gets the evolutionary benefits. Their findings hint at the fact that the functionality of a network motif is adjudged by its dynamical properties one of the important influencer of which is the motif's architecture.

G. FFL as a dynamical system

A number of key papers have concentrated specifically on the dynamical behavior of the FFLs and successfully linked it to their abundance. It has been shown that in the *ara* system of *E. coli*, C1 FFL with an AND logic generates delayed response only to the ON step of cAMP [25]. I1 FFL in the same bacterium accelerates the time taken by the *gal* genes to respond to high level of cAMP when glucose is absent in the medium [24]. I1 FFL also generates pulses [23] and detects fold change [34] in the incoming signal. Other incoherent FFLs also can act as sign-sensitive response accelerator, I4 does it for ON steps of master regulator's signal just like I1 whereas I2 and I3 do it for OFF steps of the similar signal. An AND logic is present in each of these cases. On the coherent front with similar logic, C3 and C4 are also able sign-sensitive delay generators like C1. But in reality, FFLs other than C1 and I1 are rare varieties. Uri Alon has ascribed this rarity to the fact that C3, C4, I3, and I4 systems with AND logic are responsive to only one of the input signals. But, I3 and I4 types implementing OR gate do respond to both signals [19]. More recent work like Ref. [35] is devoted to establish nonequilibrium thermodynamic principle as a strong candidate explaining correctly the relative abundance of all the eight types of FFL for both OR and AND logic. Using the multiobjective pareto-optimality framework, they proposed that the energetic cost incurred in FFLs in resourceful environmental conditions are minimal for the abundant C1 and I1 FFLs compared to other relatively nonabundant FFL

types. Their articulated metric, in this respect, namely, the specific dissipation energy (SDE) marks order-of-magnitude difference while calculated for abundant and nonabundant FFLs, the former contributing in lesser magnitude. In *E. coli* and *S. cerevisiae*, this metric can efficiently segregate the abundant motifs from the statistically uncommon motifs, even if these motifs show similar dynamical characteristics. Similar multiobjective technique has been also successfully used to produce optimal oscillators from activation-repression type interactions embedded within C4 and I3 FFLs, where the competing factors involved in the tradeoff are tunability and stability of the resulting oscillation [36]. The Response times of the FFL gene-products have been also shown to be a prominent quantifier of FFL abundance phenomenon. Murugan [37] has employed a detailed combination of deterministic and stochastic modeling to identify a key kinetic parameter, namely, the ratio of mRNA lifetime of the three constituent genes of FFL to the target protein's lifetime. When the binding conditions of the regulators with the corresponding promoters are strong, the author has been able to show that the response times of C1 and I1 FFLs show relatively reduced sensitivity towards variation in the aforementioned metric. On top of that, C1 FFL also exhibits robustness to variability in additional systemic parameters. These features strengthen the abundant nature of the motifs in perspective. Mayo *et al.* have been able to extract conclusive signature of robustness of the regulated protein level dynamics in FFL system when they identified the rise (induction) time of the top-level regulator as the controlling factor. When the latter is fast, the regulated gene-product remains insensitive to the perturbations flowing downstream. But in case of slow top-level induction to optimal state, the protein rise time becomes significantly dependent on it. Crossover in between these slow and fast domains, as has been documented in their paper, are dependent on three parameters related to the direct regulatory branch and these are TF cooperativity, dissociation constant of the TF-DNA interaction and the relative concentration of the controlling protein [38].

H. Our present approach to understand the FFL abundance problem

In this article, we propose a stochastic framework which can encapsulate in it both additive and multiplicative integration schemes at the target promoter. Additive mechanism may be regarded as a graded form of the OR gate. Our formalism also allows for constant as well as varying steady-state population level of master regulator whereas that of the intermediate regulator and the target gene-product remain constant. In all of these diverse cases, 32 in total, C1 and I1 are shown to be distinguished from rest of the FFL patterns. Previous stochastic treatments of FFLs have shown the importance of biological noise in influencing the choice of motif architecture in a case-specific basis. Kittisopikul and colleague have pointed out that for the purpose of anaerobic metabolism, *E. coli* uses FFL with on states noisier than off states whereas stress response scenario shows the opposite trend [39]. In Ref. [40], the fact that C1 FFL is least noisy among coherent FFLs, has been suggested to be the reason of its abundance whereas no such conclusion could be derived for the incoherent FFLs. Having this varied body of literature on FFLs in perspective, we venture to propose a fresh prescription with a clear focus on covariance of the intermediate TF and the final gene-product to dictate the abundance statistics of FFLs. It is a plausible assumption that correlated behavior between different gene-products in the FFL motif should bear the signature of its functional specialities. In a recent work [41], we have showed that the Shannon information fractions of two-variable mutual information (that each of the input and intermediate species shares with the target species) with respect to corresponding three-variable mutual information are maximized for a broad parametric range pertaining to a C1 FFL when compared with a two-step cascade and a bifurcation network. This finding may possibly rationalize why a C1 FFL is a network motif in *E. coli* but a two-step cascade and a bifurcation network are not [19]. The present article also takes into account stochastic effects in FFL while addressing the larger question that why C1 and I1 FFLs are abundant in the ensemble of total eight different types of FFL.

It is to be particularly noted that multiobjective pareto-optimality framework has been utilized as a sufficiently powerful treatment to analyze the abundance rationale of FFL motifs by Elchiver and colleagues [35] and in a disjoint study led by another group [36] on oscillatory properties of FFL with additional interactions. This technique crucially deals with simultaneous probing of a number of important network performance criteria which are involved in a tug of war. But depending upon the specifics of the dynamical attributes of a network subject to varying environmental conditions, choice of these objective functions may be tricky and even subjective. Consider a case where instead of resource allocation, the principle phenomenon at hand is to garner parametric situations facilitating oscillations displayed by the concentrations of biochemical species forming the network motif. In the former case, transcription rates form a tradeoff with SDE while latter case introduces period tunability and stability of limit cycle trying to get maximized or minimized with respect to each other. Additionally in the former case, definitions of energetic costs are seen to have a strong influence on the optimal performance of the FFLs. In lieu of SDE, when the authors calculated the total heat dissipation rate (HDR), it was found to exhibit positive correlation with transcription rate for simple and two-input regulation. This is in sharp contrast to what was observed for SDE but can be rationalized when we take into consideration that SDE is HDR normalized by the input mass flux in the reaction network. In Ref. [36], the pareto-optimality was used to obtain robust oscillations due to additional negative regulatory feedback with an FFL backbone. Interestingly these FFLs as mentioned earlier are C4 and I3 type which are actually relatively less represented in bacterial and yeast GTRN. Besides, the backbone FFL transforms itself from C4 to I3 as the oscillations become more stable at the cost of reduced period-tunability. These points put stress upon the fact that selection of the evolutionary aim towards which the optimization technique is directed plays an enormous role in the resultant topology coming out of the simulation process. Apart from this reverse-engineering method, there is a key fact in their forward analysis for the choice of modeling approaches. Their findings in this domain

clearly suggest that deterministic and stochastic description of the underlying biological phenomena may produce results not conforming with each other all the time. This again brings us to earlier study in Ref. [35] dealing with deterministic models which may result in misrepresentation of the actual stochastic chemical kinetics of elementary and/or coarse-grained biochemical events. Here, we also want to invoke multiobjective optimization of FFL architectures as in Ref. [42], where a tradeoff is absent between dynamic response and extrinsic noise tolerance. The authors feel that adherence to both the objectives may be challenging to implement due to the issue of high dimensional nature of the problem at hand. To circumvent the problem, they opined deterministic-stochastic combined modeling approach to be an efficient route.

In light of these distinctive features of previous analysis, we favored extrinsic noise as our metric of interest so that we become able to capture the stochastic attributes of biochemical processes constituting FFLs. Noise undoubtedly plays more fundamental role than previously described competing objectives in paretocentric analysis in the sense that these objectives are always prone to stochasticity. To exemplify, the transcription rate along with HDR and SDE are all function or functional of copy numbers of FFL gene-products which in turn display variability. Hence, we believe our choice in this regard has been significantly less subjective than the previous attempts to understand FFL abundance phenomenon.

II. THE MODEL

A. Dynamics of a generic FFL—A Langevin description

For a generic FFL, the dynamics of the copy numbers of its constituent gene-products S, X, and Y are well represented with the help of Langevin formalism [28,43–46],

$$\frac{dS}{dt} = f_s - \mu_s S + \xi_s(t), \tag{1a}$$

$$\frac{dX}{dt} = f_x(S) - \mu_x X + \xi_x(t), \tag{1b}$$

$$\frac{dY}{dt} = f_y(S,X) - \mu_y Y + \xi_y(t). \tag{1c}$$

In other words, gene-products S, X, and Y are synthesized with rates f_s, f_x, and f_y, respectively, whereas these species are lost (degraded and diluted) from the system with rates proportional to their individual population levels in a unit effective cellular volume. The corresponding synthesis (relaxation) rate parameters are k_s, k_x, and k_y (μ_s, μ_x, and μ_y), respectively. While $f_s = k_s$, f_x, and f_y involve nonlinear Hill functions representing the probability of various promoter occupancies by the respective regulatory molecules [19,28,47]. Their functional forms depend on the regulatory nature of individual edges, e.g., activation ($\rightarrow$) and repression ($\dashv$) in an FFL (see schematic Fig. 1 for eight possible patterns). $S \rightarrow X$ has $f_x =: k_{sx}[S^n/(S^n + K_{sx}^{+n})]$ and $S \dashv X$ is represented by $f_x =: k_{sx}[K_{sx}^{-n}/(S^n + K_{sx}^{-n})]$, where "+" and "−" denote activation and repression, respectively. Here, "n" is the Hill coefficient and K_{sx}^{+} (K_{sx}^{-}) is the activation (repression) coefficient for up-regulation (down-regulation) of gene X by S. The synthesis rate parameter is symbolically represented by k_{sx}. The synthesis rate for Y is a two-variable function integrating upstream TFs using either additive (A) or multiplicative (M) mechanism. As examples, C1 FFL ($S \rightarrow Y, X \rightarrow Y$) has an additive input function, $f_y^A =: k_{sy}^A[S^n/(S^n + K_{sy}^{+n})] + k_{xy}^A[X^n/(X^n + K_{xy}^{+n})]$. The multiplicative counterpart is $f_y^M =: k_y^M[S^n/(S^n + K_{sy}^{+n})][X^n/(X^n + K_{xy}^{+n})]$. In case of I1 FFL ($S \rightarrow Y$, $X \dashv Y$), $f_y^A =: k_{sy}^A[S^n/(S^n + K_{sy}^{+n})] + k_{xy}^A[K_{xy}^{-n}/(X^n + K_{xy}^{-n})]$, and $f_y^M =: k_y^M[S^n/(S^n + K_{sy}^{+n})][K_{xy}^{-n}/(X^n + K_{xy}^{-n})]$. Other types of FFL can be modeled in an analogous way. To avoid cumbersome notations for synthesis rate parameters $k_{..}$, we do not specify "+" and "−" signs for them but in principle, these are different numbers for activation and repression for a specific promoter. All the synthesis rate parameters, activation (repression) coefficients, and relaxation rate parameters are expressed in units of $(c.n.)\ min^{-1}$, $c.n.$, and, min^{-1}, respectively, where "c.n." stands for copy number. However, it should be borne in mind that copy numbers of eukaryotic TFs which are produced in the cytoplasm and transported to the nucleus, are exposed to an effective cellular volume instead of only the cytoplasmic or nuclear volume [27]. This is a point of difference with respect to bacterial TFs. Hence, concentration of eukaryotic TFs and their synthesis rate parameters should involve this effective cellular volume in the respective units. For our case, we resort to a simplifying assumption considering this effective volume to be unity and continue not to explicitly invoke volume in any of the relevant units. It appears that Eqs. 1(a)–1(c) is also independent of this volumetric consideration.

Anyway, this transport of TFs from cytoplasm to nucleus has been observed to be rapid so that these newly synthesized TFs in effect instantaneously become enabled to up or down-regulate their target gene(s). Hence, we do not include any time delays in our model to capture this phenomenon [33,48]. Moreover, we avoid incorporating mRNA dynamics in our model because of the implicit timescale separation constraint which is basically the faster dynamics of the DNA transcripts with respect to considerably slower protein species. Considering the latter entities to be at steady states automatically guarantees the mRNAs to be at steady states [49]. This assumption holds well in the absence of bursts of protein production as has been rationalized in Refs. [14,50] and used in Refs. [35,42,44–46]. From the perspective of our stochastic simulation analysis utilized here, it is worth-mentioning that the end-points of the ensemble of independent time-courses of biochemical species are situated sufficiently later than the deterministic time-point where the FFL dynamics become stationary. Hence, noninclusion of time-delays in our model built to analyze single time-point steady-state and not temporal or transcient response of FFLs is sufficiently fit. Although for evolutionary simulation based models made to investigate selection advantage of FFLs, entire temporal response of the dynamical system is essential as in Ref. [33] which naturally includes time-delays due to transcription initiation, mRNA elongation, starting of regulation after transcription completion etc. This particular study consists of both deterministic and stochastic modeling approaches. While transition in between chromatin states, transcription initiation and mRNA degradation are simulated using Gillespie stochastic simulation algorithm (SSA) [51,52], protein dynamics are treated using deterministic ordinary differential equations whereas the requisite

time-delays are also maintained at a constant value by deterministic means.

Coming back to our present model, $\xi_i(t)$ represents independent Gaussian-distributed noise processes with $\langle \xi_i(t) \rangle = 0$, $\langle \xi_i(t)\xi_j(t') \rangle = \langle |\xi_i^2| \rangle \delta_{ij}\delta(t-t')$, where $i, j \in \{s, x, y\}$ and $\langle \cdots \rangle$ symbolizes ensemble average at steady state [28,43–46,53,54]. The total noise strength comes from both synthesis and loss events and thus $\langle |\xi_s^2| \rangle = \langle f_s \rangle + \mu_s \langle S \rangle = 2\langle f_s \rangle = 2\mu_s \langle S \rangle$ (at steady state) and likewise for X and Y. We note that this usage of constant noise strengths at steady state is a judicious approximation made to simplify our analysis [28,44–46].

B. Variances and covariances of the FFL gene-products

Now, we present the formalism to obtain the second moments of the gene-products S, X, and Y of an FFL in closed analytic forms using the technique of small-noise approximation (SNA) [55–58]. Under its purview, we consider first-order fluctuations in the copy numbers of S, X, and Y in a unit effective cellular volume around their individual steady states, i.e., $Z = \langle Z \rangle + \delta Z$ where $\delta Z/\langle Z \rangle \ll 1$ where $Z \in \{S, X, Y\}$. Accordingly, the linearized version of Eqs. 1(a)–1(c) in the last subsection is obtained as follows:

$$\frac{d\delta \boldsymbol{Z}}{dt} = \boldsymbol{J}_{z=\langle z \rangle} \delta \boldsymbol{Z}(t) + \boldsymbol{\Xi}(t). \tag{2}$$

$\delta \boldsymbol{Z}$, $\boldsymbol{J}$, and $\boldsymbol{\Xi}$ are the fluctuation matrix, the steady-state Jacobian matrix, and the noise matrix, respectively,

$$\delta \boldsymbol{Z} = \begin{pmatrix} \delta S \\ \delta X \\ \delta Y \end{pmatrix}, \quad \boldsymbol{J} = \begin{pmatrix} -\mu_s & 0 & 0 \\ \langle f'_{x,s} \rangle & -\mu_x & 0 \\ \langle f'_{y,s} \rangle & \langle f'_{y,x} \rangle & -\mu_y \end{pmatrix},$$

$$\text{and} \quad \boldsymbol{\Xi} = \begin{pmatrix} \xi_s \\ \xi_x \\ \xi_y \end{pmatrix},$$

where $\langle f'_{x,s} \rangle =: (\frac{df_x(S)}{dS})_{S=\langle S \rangle}$ and so on. At steady state of the system, fluctuations and dissipation can be combined in the Lyapunov equation [14,16],

$$\boldsymbol{J}\boldsymbol{\Sigma} + \boldsymbol{\Sigma}\boldsymbol{J}^T + \boldsymbol{D} = \boldsymbol{0}. \tag{3}$$

Here $\boldsymbol{\Sigma}, \boldsymbol{J}^T$, and $\boldsymbol{D} =: \langle \boldsymbol{\Xi}\, \boldsymbol{\Xi}^T \rangle$ are, respectively, the second moment matrix, transposed Jacobian matrix, and matrix of noise strengths, all in steady state. The form of the $\boldsymbol{\Sigma}$ matrix is as follows:

$$\boldsymbol{\Sigma} = \begin{pmatrix} \Sigma(S) & \Sigma(S,X) & \Sigma(S,Y) \\ \Sigma(X,S) & \Sigma(X) & \Sigma(X,Y) \\ \Sigma(Y,S) & \Sigma(Y,X) & \Sigma(Y) \end{pmatrix}.$$

Solving Eq. (3) for $\boldsymbol{\Sigma}$ while noting that covariance is symmetric in its two variables (i.e., $\Sigma(X,S) = \Sigma(S,X)$, etc.), we obtain the following analytic expressions for variances and covariances of S, X, and Y,

$$\Sigma(S) = \langle S \rangle, \tag{4a}$$

$$\Sigma(S,X) = \frac{\langle f'_{x,s} \rangle \Sigma(S)}{(\mu_s + \mu_x)}, \tag{4b}$$

$$\Sigma(S,Y) = \frac{\langle f'_{y,s} \rangle \Sigma(S) + \langle f'_{y,x} \rangle \Sigma(S,X)}{(\mu_s + \mu_y)}, \tag{4c}$$

$$\Sigma(X) = \langle X \rangle + \frac{\langle f'_{x,s} \rangle \Sigma(S,X)}{\mu_x}, \tag{4d}$$

$$\Sigma(X,Y) = \frac{\langle f'_{y,s} \rangle \Sigma(S,X) + \langle f'_{x,s} \rangle \Sigma(S,Y) + \langle f'_{y,x} \rangle \Sigma(X)}{(\mu_x + \mu_y)}, \tag{4e}$$

$$\Sigma(Y) = \langle Y \rangle + \frac{\langle f'_{y,s} \rangle \Sigma(S,Y) + \langle f'_{y,x} \rangle \Sigma(X,Y)}{\mu_y}. \tag{4f}$$

These closed-form analytical expressions of second moments will be invoked while addressing the FFL abundance statistics in the following sections.

III. RESULTS AND DISCUSSION

A. The proposed metric

In our analysis, we observe that it is the magnitude (|..|) of covariance between X and Y, normalized by the product of their steady-state ensemble-averaged population levels ($|\Sigma^N(X,Y)| =: |\Sigma(X,Y)/(\langle X \rangle \langle Y \rangle)|$), that distinguishes all the eight types of FFL performing either additive or multiplicative signal integration in a situation where steady-state ensemble-averaged population of input TF (S) remains constant or varies. In all of these cases, we keep $\langle X \rangle = \langle Y \rangle = 100$ copies so that different FFL architectures may be compared on an equal footing, i.e., equal expression levels of intermediate and target genes. The normalization process provides a metric for associated fluctuations of X and Y while keeping unnecessary population size effects at bay [14]. Rewriting Eq. (4e) of $\Sigma(X,Y)$ in terms of systemic parameters and then normalizing (denoted by superscript "N") by the product of $\langle X \rangle$ and $\langle Y \rangle$ in all the FFL types for additive and multiplicative signal integration mechanisms, we obtained total sixteen expressions. The following equation unifies all of those cases in a simple and efficient manner:

$$\begin{aligned} \Sigma^N(X,Y) = D\Bigg[& \left(nA\left\langle P_{sx}^{\text{off}} \right\rangle\right)\left(nB\left\langle P_{sy}^{\text{off}} \right\rangle\right) \frac{[\Phi_{xs}\Phi_{yx} + \Phi_{ys}\Phi_{xy}]}{\langle S \rangle} \\ & + (nA\langle P_{sx}^{\text{off}} \rangle)^2 \left(nC\left\langle P_{xy}^{\text{off}} \right\rangle\right) \frac{[\Phi_{xs}(\Phi_{ys}\Phi_{xy} + \Phi_{yx})]}{\langle S \rangle} \\ & + \left(nC\left\langle P_{xy}^{\text{off}} \right\rangle\right) \frac{\Phi_{yx}}{\langle X \rangle} \Bigg]. \end{aligned} \tag{5}$$

The right-hand side in Eq. (5) comprises various steady-state probabilities of promoter sites in their off states ($\langle P_{..}^{\text{off}} \rangle$), filtering functions $\Phi_{ij} =: [\mu_i/(\mu_i + \mu_j)]$ where $i, j(\neq i) \in \{s, x, y\}$, Hill coefficient n, regulatory signs (A, B, and C), $\langle S \rangle$, $\langle X \rangle$, and signal integration factor D. In the expression of Φ_{ij}, $\mu_i = \tau_i^{-1}$ where τ_i denotes the average life-time of the gene-product "i." As the name suggests, these filtering functions Φ_{ij} conduct time-averaging operation on the propagated fluctuations between genes "i" and "j." Each of A, B, and C designates regulatory nature of edges from S to X, from S to Y, and from X to Y, respectively, with values +1 (activation) or −1 (repression). D = 0.5 (1) represents additive (multiplicative)

signal integration mechanism. The triplet of (A, B, C) for different FFL types are: C1 (+1, +1, +1), C2 (−1, −1, +1), C3 (+1, −1, −1), C4 (−1, +1, −1), I1 (+1, +1, −1), I2 (−1, −1, −1), I3 (+1, −1, +1), and I4 (−1, +1, +1). An "off" state for gene activation (repression) signifies that the activator (repressor) molecule is not bound (bound) to the corresponding promoter site. An "on" state in contrast implies the opposite binding condition. Associating "+" or "−" sign to promoter states (both "off" and "on") designates whether those states are for activation or repression, respectively. It is to be mentioned that $\langle P_{..}^{\text{off}}\rangle$ is connected to $\langle P_{..}^{\text{on}}\rangle$ by the relation $\langle P_{..}^{\text{off}}\rangle = 1 - \langle P_{..}^{\text{on}}\rangle$ since these are characteristic occurrence probabilities of alternate states "off" and "on" for genetic promoter sites. The functional forms of $\langle P_{..}^{\text{off}}\rangle$ terms corresponding to different promoter sites are as follows [19]: $\langle P_{sx}^{\text{off}}\rangle^{+} = [K_{sx}^{+n}/(\langle S\rangle^n + K_{sx}^{+n})]$, $\langle P_{sx}^{\text{off}}\rangle^{-} = [\langle S\rangle^n/(\langle S\rangle^n + K_{sx}^{-n})]$, $\langle P_{sy}^{\text{off}}\rangle^{+} = [K_{sy}^{+n}/(\langle S\rangle^n + K_{sy}^{+n})]$, $\langle P_{sy}^{\text{off}}\rangle^{-} = [\langle S\rangle^n/(\langle S\rangle^n + K_{sy}^{-n})]$, $\langle P_{xy}^{\text{off}}\rangle^{+} = [K_{xy}^{+n}/(\langle X\rangle^n + K_{xy}^{+n})]$, and $\langle P_{xy}^{\text{off}}\rangle^{-} = [\langle X\rangle^n/(\langle X\rangle^n + K_{xy}^{-n})]$. Therefore, combinations of A, B, C, and $\langle P_{..}^{\text{off}}\rangle$ terms generate C1–C4 and I1–I4 FFL motifs.

B. $|\Sigma^N(X,Y)|$: The extrinsic noise interpretation

Equation (4f) suggests that $\Sigma(X,Y)$ acts as a source of extrinsic noise for the target gene Y. Equation (5) clearly shows the contributions of each of these external factors which are S and X. These biochemical species S and X interact with fluctuating environmental signals and convey their implications in corresponding reaction rates which ultimately enslave the production and degradation of target gene-product Y. To paraphrase the previous statement, the input TF and its coregulator behave as gateways for different environmental states to come in contact with output gene Y which is responsible for the production of physiologically important biochemical products ultimately sustaining the animate system. We were motivated to use normalized covariance of X and Y as a metric for extrinsic noise by the elegant dual-reporter proteins experiments by Elowitz *et al.* [12]. In a related theoretical framework, effective decomposition of the total noise in gene expression into intrinsic and extrinsic parts was done and the conditional dominance of transcription and translation on the intrinsic noise was pointed out [15]. The conditions under which intrinsic and extrinsic noise can be modeled separately with respect to each other while the biological network and its environment change on similar timescales, have been also reported [17]. Usage of |..| removes the resulting effective "−" signs from the $\Sigma^N(X,Y)$ in case of some FFL types thus facilitating comparison of extrinsic noise contributions from different FFLs on an equal footing. The first two terms in Eq. (5) explicitly contain $\langle S\rangle$ whereas the third term comprises of $\langle X\rangle$. Various steady-state probabilities of promoter sites in their off states behave as coefficients for these external sources of fluctuations. These coefficients also include single filtering function and combinations of several of them, dictating the temporal averaging on the fluctuations of S and X. The first term signifies S mediated regulation of X and Y while the second term is simply the two-step cascade contribution, i.e., S regulating X regulating Y. Hence, these two terms signify how on average the steady-state intrinsic fluctuations in S can affect Y in a pathway-dependent modified form. Similarly, the last term captures the manner by which on average the steady-state intrinsic noise of X enslaves Y in a modulated form.

$\langle P_{..}^{\text{off}}\rangle$ may be linked to the logarithmic gain (H) in Paulsson's unified framework of noise summation [14,16]. The mathematical connection can be easily computed and we found it to be encapsulated in the following expressions. $\langle H_{xs}\rangle^{\pm} = \mp n\langle P_{sx}^{\text{off}}\rangle^{\pm}$, $\langle H_{ys}\rangle^{\pm} = \mp n\langle P_{sy}^{\text{off}}\rangle^{\pm}\langle Y_s\rangle^{f}$, $\langle H_{yx}\rangle^{\pm} = \mp n\langle P_{xy}^{\text{off}}\rangle^{\pm}\langle Y_x\rangle^{f}$. Here, $\langle Y_s\rangle^{f} =: (\langle Y_s\rangle/\langle Y\rangle)$, $\langle Y_s\rangle$ being the copy number of Y produced by S at steady state in additive signal integration and "f" stands for fraction. Similar definition also holds for $\langle Y_x\rangle^{f}$. For multiplicative signal integration, $\langle H_{ys}\rangle^{\pm}$ ($\langle H_{yx}\rangle^{\pm}$) is free of $\langle Y_s\rangle^{f}$ ($\langle Y_x\rangle^{f}$). Hence, combinations of different $\langle P_{..}^{\text{off}}\rangle$ as they appear in various terms of Eq. (5), are related to susceptibility factors in Paulsson's framework [14,16]. To be rigorous, the ultimate influence of extrinsic species S and X on target Y should involve another coupling factor $\langle f'_{y,x}\rangle$ and ${\mu_y}^{-1}$ multiplied with $\Sigma(X,Y)$ as may be understood from consulting Eq. (4f). But we have checked that this latter modified extrinsic noise factor fails to explain FFL abundance spanning all the possibilities considered here (data not shown).

C. Our strategy: Differential promoter activity

The steady-state distinguishability of all types of FFL has been possible by introducing differences in promoter occupancy probabilities depending upon activation and repression. Otherwise, different FFLs show clustering for the designated metric of fluctuations. To this end, our fixed $\langle S\rangle$ construct has $K_{sx}^{+} = 1000$ copies for S → X. Taking $n = 1$ for all the promoters and according to the expressions of $\langle P_{..}^{\text{off}}\rangle$ in subsection (A) of Results and Discussion, $\langle P_{sx}^{\text{off}}\rangle^{+} \approx 0.99$. This signifies X promoter remains mostly empty which makes S mediated synthesis of X nearly a first-order kinetics, characteristic of linear activation. However, for $S \dashv X$, we choose $K_{sx}^{-} = \langle S\rangle = 10$ copies which makes $\langle P_{sx}^{\text{off}}\rangle^{-} = 0.5$, i.e., X promoter remains half of the time empty of the repressor molecule, thereby facilitating synthesis of gene-product X. For $S \to Y$ ($S \dashv Y$), we set $K_{sy}^{+} = K_{sx}^{+}$ ($K_{sy}^{-} = K_{sx}^{-}$) resulting in $\langle P_{sy}^{\text{off}}\rangle^{+} = \langle P_{sx}^{\text{off}}\rangle^{+}$ ($\langle P_{sy}^{\text{off}}\rangle^{-} = \langle P_{sx}^{\text{off}}\rangle^{-}$). In case of $X \to Y$ ($X \dashv Y$), we choose K_{xy}^{+} (K_{xy}^{-}) = 1000 ($\langle X\rangle$) copies, making $\langle P_{xy}^{\text{off}}\rangle^{+}$ ($\langle P_{xy}^{\text{off}}\rangle^{-}$) ≈ 0.91 (0.5). Given these values, now we can rank different off promoter states as follows:

$$\left\langle P_{sx}^{\text{off}}\right\rangle^{+} = \left\langle P_{sy}^{\text{off}}\right\rangle^{+} > \left\langle P_{xy}^{\text{off}}\right\rangle^{+} > \left\langle P_{sx}^{\text{off}}\right\rangle^{-} = \left\langle P_{sy}^{\text{off}}\right\rangle^{-} = \left\langle P_{xy}^{\text{off}}\right\rangle^{-}. \tag{6}$$

For the case of constant population of input TF, we maintain $\langle S\rangle = 10$ copies by increasing k_s and μ_s proportionately so that addition and depletion rates of gene-product S is balanced at steady state of the system for each pair of (k_s, μ_s). Otherwise, $\langle S\rangle$ increases by one copy at each instant with increasing k_s while μ_s is held fixed, thereby implementing net gain in the cellular reservoir of gene-product S at steady state for each k_s. Even then, $\langle S\rangle$ remains much lower than the population of downstream gene-products. When, $\langle S\rangle$ increases, we still use same relations of activation (repression) coefficients from the fixed $\langle S\rangle$ case, the only difference being

that the repression coefficients $K_{sx}^{-} = K_{sy}^{-} = \langle S \rangle$ now increases with $\langle S \rangle$. This makes the strengths of S mediated activations to increase within the chosen variational range of $\langle S \rangle$, whereas S mediated repression strengths remain unchanged at their values from fixed $\langle S \rangle$ case. Hence, relation Eq. (6) still holds for this case. Thus, our theoretical construct can successfully distinguish different types of coherent and incoherent FFLs at steady state, utilizing linear activation and half-maximal repression. One biological perspective behind this framework may suggest that in this case, the activators are minimally inducible whereas repressors also remain inactive (although less often than activators) half of the time. In recent times, synthetic techniques have been developed by which dynamic ranges of bacterial genetic promoters may be tuned at ease. These hybrid promoters closely associated with *E. coli* are observed to interact with TFs which are induced by ligands. In this process, the researchers were also able to construct efficient gene-transcriptional logic gates with multiple inputs [59]. This specific endeavor makes us hopeful regarding experimental implementation of our present approach of differential promoter activities.

It is specifically important to note that our choices of activation and repression coefficients being much greater than unity make the binding and unbinding rates of TFs to the genetic promoter sites unequal to each other thus shifting the kinetics of the TFs-DNA interaction far from chemical equilibrium while their dynamical attributes are being calculated at steady states. Hence, by the merit of differential promoter activity, our model enables us to study FFLs in the nonequilibrium steady-state (NESS) condition in accordance with biological realism. Reference [35] also deals with NESS but their choices of activation and repression coefficients are additionally motivated by constraints imposed upon reaction fluxes and nature of the signal integration mechanisms (AND, OR) whereas our model gives equal footing to each of the 32 FFL case studies presented here with only one constraint to maintain throughout, i.e., they adhere to the same set of population levels of biochemical species. To reiterate, this step is motivated by the phenomenological fact that many biochemical processes are optimally executed at their steady states [19]. Considering the concept of binding energy loosely expressed by $F = -k_B T \ln(K_{..}^{\pm})$ (in the units of Joule) [27,60] where k_B, T, and ln denote the Boltzmann constant, temperature in absolute scale, and natural logarithm, respectively, we can understand that $K_{..}^{+}$ being larger than $K_{..}^{-}$, activation possesses higher magnitude of binding energy than repression. Therefore, differential promoter activity allows differential binding energy expenditure to be incurred by various regulation processes constituting FFLs so that different FFL types can be distinguished properly. As an important point, we should mention that the mutational model by [33] which establishes superiority of AND-gated C1 FFL as a filter of spurious signal, is enriched in genetic promoters with weak affinities for their respective TFs. Binding affinity being inversely proportional to activation or repression coefficient depending upon the regulatory mode considered, the situation turns out to be far from chemical equilibrium like our present model. The authors point out that these weak-affinity TF binding sites appearing by chance factor alone get increasingly stronger after being selected by the force of evolution. These special type of genetic promoter sites are capable of simulating duplication and deletion of genes. These important evolutionary observations also strengthen our current choice of activation and repression coefficients to be greater than unity. As a related side note, we want to invoke our previous work on information processing in diamond motif where we showed weak TF binding affinity promotes information sharing between gene-products whereas this sharing quantified by information redundancy occurs with simultaneous increment in signal-to-noise ratio, i.e., betterment of signal fidelity in the GTRN [45].

D. Signaling scenario I: Fixed input TF population

Figures 2(a) and 2(b) show profiles of $|\Sigma^N(X, Y)|$ for C1–C4 FFLs employing additive and multiplicative integration mechanism, respectively, considering populations of S, X, and Y to remain fixed while k_s increases. Figures 2(c) and 2(d) are for additive and multiplicative I1-I4 FFLs, respectively. For fixed $\langle S \rangle$ case, filtering functions Φ_{xs} and Φ_{ys} which contain μ_s in their denominators, decrease as μ_s has to increase along with k_s to keep $\langle S \rangle$ fixed according to the constraint $k_s = \langle S \rangle \mu_s$. In Eq. (5), the leading order contributions in the denominators of different terms comes $\sim {k_s}^1$ and ${k_s}^2$. As a result, $|\Sigma^N(X, Y)|$ decreases with increasing k_s for all FFLs. Equal contributions from both direct and indirect paths in additive FFLs makes $D = 0.5$ whereas multiplicative FFLs having nonseparable path contributions for regulation of Y, has $D = 1$. This is reflected in doubled magnitude of $|\Sigma^N(X, Y)|$ for multiplicative integration compared to their additive counterparts in all types of FFL [compare Fig. 2(b) with Fig. 2(a) and also Fig. 2(d) with Fig. 2(c)]. In Figs. 2(a) and 2(b), C1-A (C1-M) gives higher value for the metric than rest of the coherent additive (multiplicative) FFLs. Same is true for incoherent FFLs, separating I1-A (I1-M) from the other FFLs belonging to the respective groups (additive and multiplicative), as shown in Figs. 2(c) and 2(d). This trend can be rationalized by noting relation Eq. (6) among different steady-state probabilities of promoters in their off states.

Utilizing this aforementioned differential promoter activities along with $A.C = B$, and $A^2 = 1$ to compute $|\Sigma^N(X, Y)|$ from Eq. (5), C1 can be clearly shown to exhibit more amount of normalized fluctuations associating X and Y than C2, C3, and C4 both for additive and multiplicative cases. In addition to relation Eq. (6), while analyzing the incoherent cases, one should also consider $A.C = -B$, $A^2 = 1$, multiplicativity property of absolute value, and the fact that in Eq. (5), temporal-averaging factor for the first term is larger than that of the second term. To paraphrase the last requirement in terms of combinations of filtering functions, we have $[\Phi_{xs}\Phi_{yx} + \Phi_{ys}\Phi_{xy}] > [\Phi_{xs}(\Phi_{ys}\Phi_{xy} + \Phi_{yx})]$. It is to be mentioned that due to increment in μ_s along with k_s in this construct, these two terms in square brackets do decrease. By considering these conditions, I1 can be shown to produce more $|\Sigma^N(X, Y)|$ than I2, I3, and I4.

E. Signaling scenario II: Increasing input TF population

Figures 3(a) and 3(b) show the responses of $|\Sigma^N(X, Y)|$ to increasing $\langle S \rangle$ in coherent FFLs utilizing additive and

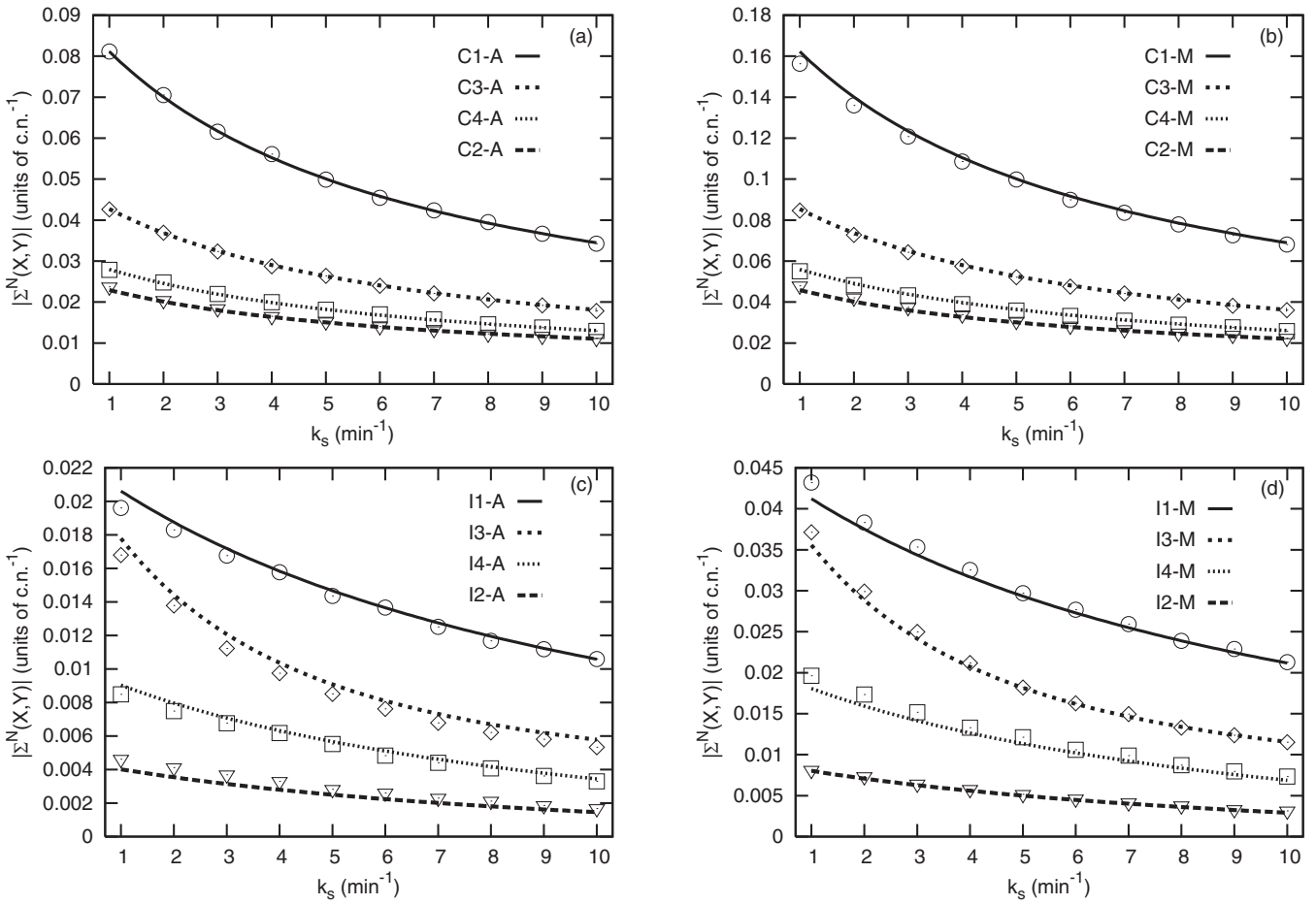

FIG. 2. Absolute value of the normalized covariance of X and Y ($|\Sigma^N(X,Y)| =: |\Sigma(X,Y)/(\langle X\rangle\langle Y\rangle)|$) in units of c.n.$^{-1}$ for (a) C1–C4 (additive), (b) C1–C4 (multiplicative), (c) I1–I4 (additive), and (d) I1–I4 (multiplicative) with respect to variation of k_s (in units of min^{-1}). We maintain $\langle S\rangle = 10$, $\langle X\rangle = \langle Y\rangle = 100$ copies in a unit effective cellular volume. Relaxation rate parameters are $\mu_x = 0.5$ min^{-1}, $\mu_y = 1$ min^{-1}, and $\mu_s = k_s/\langle S\rangle$. The downstream synthesis parameters for C1 FFL are as follows: $k_{sx} = \mu_x\langle X\rangle[\langle S\rangle^n/(\langle S\rangle^n + K_{sx}^{+n})]^{-1}$. For additive signal integration, $k_{sy}^A = 0.5\mu_y\langle Y\rangle[\langle S\rangle^n/(\langle S\rangle^n + K_{sy}^{+n})]^{-1}$ and $k_{xy}^A = 0.5\mu_y\langle Y\rangle[\langle X\rangle^n/(\langle X\rangle^n + K_{xy}^{+n})]^{-1}$ ensure equal contributions from direct and indirect paths to produce total $\langle Y\rangle$ copies. In the multiplicative case, $k_y^M = \mu_y\langle Y\rangle[\langle S\rangle^n/(\langle S\rangle^n + K_{sy}^{+n})]^{-1}[\langle X\rangle^n/(\langle X\rangle^n + K_{xy}^{+n})]^{-1}$. Synthesis rate parameters for other FFL types were chosen according to the corresponding combinations of regulatory edges. For the activating edges we keep the activation coefficient $K_{..}^+ = 1000$ copies. Otherwise, the operating points are set to ensure half-maximal repression by setting $K_{sx}^- = K_{sy}^- = \langle S\rangle = 10$, and $K_{xy}^- = \langle X\rangle = 100$ copies. $n = 1$ is used for all the regulatory edges in different FFLs. Lines are steady-state analytical results supported by simulation data from Gillespie's SSA [51,52] (symbols) with ensemble averaging over 10^5 independent samples at steady state.

multiplicative signal integration schemes, respectively. Figures 3(c) and 3(d) portray responses from additive and multiplicative incoherent FFL types, respectively. The rationale behind greater metric value for multiplicative case than additive case is just the same from previous scenario of fixed $\langle S\rangle$. In the present construct, filtering functions ($\Phi_{..}$) are held unchanged as μ_s does not change. Using the functional forms of $\langle P_{sx}^{\text{off}}\rangle^{\pm}$ and $\langle P_{sy}^{\text{off}}\rangle^{\pm}$ (see subsection (A) in Results and Discussion), it can be shown that the denominators of different terms in Eq. (5) have contributions from $\langle S\rangle$ with a leading order of 1–3. Therefore, $|\Sigma^N(X,Y)|$ decreases with increasing $\langle S\rangle$. Here also, in both the additive and multiplicative cases, C1 is distinguished from C2–C4, by contributing higher values for the metric. I1-A and I1-M are also distinguished by their higher values for the metric compared to other member FFLs of the corresponding groups. To explain this noticeable distinctiveness of C1 and I1 from other FFL types in additive and multiplicative categories, one can take recourse to the strategies from fixed $\langle S\rangle$ case. We noticed $\langle P_{sx}^{\text{off}}\rangle^+$ and $\langle P_{sy}^{\text{off}}\rangle^+$ to decrease with increasing $\langle S\rangle$ in such a way, so that relation Eq. (6) still holds. On a side note, the distinguishability of I1-A (I1-M) is more prominent for low values of $\langle S\rangle$. For $\langle S\rangle \approx 15$ copies onward, I3-A (I3-M) contributes higher $|\Sigma^N(X,Y)|$ than I1-A (I1-M) case (data not shown). Hence, our model predicts that low input TF population is a key issue governing abundance of I1-A (I1-M). Here, it is also to be noted that magnitude-wise, the results coming out of Gillespie SSA tend to differ slightly with predictions stemming from SNA as the input TF population ($\langle S\rangle$) starts to have relatively fewer copies although both exact and approximate data sets adhere to the same characteristic pattern distinguishing I1 FFL from rest of the incoherent FFLs.

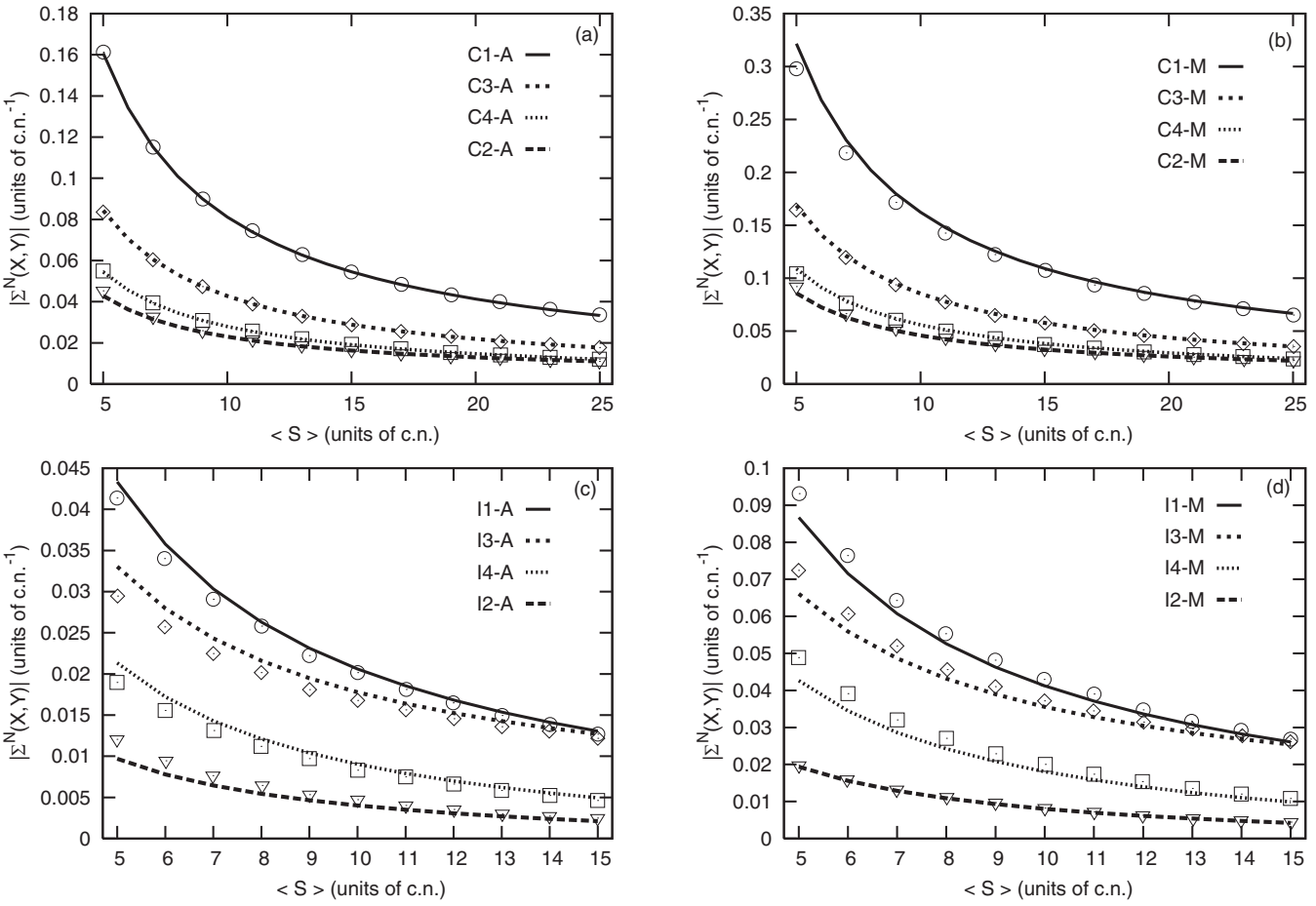

FIG. 3. Absolute value of the normalized covariance of X and Y ($|\Sigma^N(X,Y)| =: |\Sigma(X,Y)/(\langle X\rangle\langle Y\rangle)|$) in units of c.n.$^{-1}$ for (a) C1–C4 (additive), (b) C1–C4 (multiplicative), (c) I1–I4 (additive), and (d) I1–I4 (multiplicative) with respect to variation of $\langle S\rangle$ (in units of c.n.). We maintain $\langle X\rangle = \langle Y\rangle = 100$ copies in a unit effective cellular volume. Keeping $\mu_s = 0.1$ min^{-1} and varying k_s, we ensure $\langle S\rangle$ to increase by 1 copy at each instant. Other parametric conditions are identical with Fig. 2. Analytical results are designated by lines while results from Gillespie SSA [51,52] are presented with symbols. These latter data sets are obtained with ensemble averaging over 10^5 independent samples at steady state.

This discrepancy may be attributed to the low-copy number effects [61,62].

F. Contextualization of our findings in light of the gene regulation program

We have found that only with differential promoter activities, distinct FFL types can be shown to display difference in their steady-state values of $|\Sigma^N(X,Y)|$. If all the regulatory operating points happen to be set at their respective half-maximal thresholds, then all the FFL types can not be fairly distinguished on the basis of $|\Sigma^N(X,Y)|$ values. The fact that the gene expression levels at the input, intermediate, and output layers are constrained to be unique across all FFL types contributes towards this aforementioned indistinguishability. Values of the terms in relation Eq. (6) show that all the genetic activations are of linear type whereas repressions are relatively nonlinear as these are half of the way closer to full saturation of the respective genetic promoters (which is similar to zero-order kinetics of enzymatic reactions). Each of X and Y gene has to produce constant amount of their products irrespective of the corresponding promoter being in their on states for shorter or longer periods of time across different FFL types. For example $\langle P_{sx}^{\text{off}}\rangle^+$ in $S \to X$ is of much higher value than $\langle P_{sx}^{\text{off}}\rangle^-$ in $S \dashv X$. This implies that $\langle P_{sx}^{\text{on}}\rangle^-$ is much higher than $\langle P_{sx}^{\text{on}}\rangle^+$. Hence, in the activated case gene X has to produce a lot more products per unit time compared to the production per unit time in the repressed case so that in both the cases the condition of $\langle X\rangle = 100$ copies gets fulfilled. Therefore, $S \to X$ produces much more fluctuations in synthesis of X than $S \dashv X$ does. Similar arguments follow for other promoters also. Hence, in our construct, each of the on promoters are noisier in genetic activation than repression. Keeping this point in mind, now we turn to the FFL architectures.

On mindful inspection of the FFL topologies in Fig. 1, one can observe that the regulatory edges emanating from S and ending upon X and Y are same for similarly numbered types of coherent and incoherent FFLs. To exemplify, both C1 and I1 have $S \to X$ and $S \to Y$ whereas in each of C3 and I3, one can find $S \to X$ and $S \dashv Y$ and so on for other types. Hence, to suspect that the abundance property may remain with the role S plays in regulating its coregulator X and target gene Y, should come at no surprise. In both of the C1 and I1 FFL, S plays similar role (activation) in regulating X and Y and therefore one may conclude that noise flows

in abundance from S to X and Y. Hence, if any or both of the outgoing edges from S includes a repression, then this S mediated noise flow will be reduced. This point is also valid for the X mediated control of Y but in this case, one has to also keep track of the proportion of noise flowing from S to X. Having said so, one should also notice that the expression level of S is much lower than that of X and as a consequence S brings in more copy-number fluctuations than what X does for Y. This is reflected in greater magnitudes of the first two terms with respect to third term in Eq. (5) (data not shown but can be easily computed using values of various systemic parameters and steady-state promoter activities mentioned in the preceding part of this article). Now, we compare C1 and C2 both of which have $X \to Y$ but the former has $S \to X$, $S \to Y$ while the latter possesses $S \dashv X$, $S \dashv Y$. The extrinsic noise contribution from C1 is the highest and C2 produces least of it as depicted in Figs. 2(a) and 2(b), Figs. 3(a) and 3(b). S mediated regulations also flip from activation to repression in I1 and I2, whereas $X \dashv Y$ remains a common feature. As a result of which, I1 and I2 displays highest and lowest extrinsic noise, respectively, as profiled in Figs. 2(c) and 2(d), Figs. 3(c) and 3(d). Combinations of activation and repression in S mediated control over X and Y in C3 (I3) and C4 (I4) place their extrinsic noise levels intermediate with respect to C1 (I1) and C2 (I2). Comparing C3 with C4 shows that both have $X \dashv Y$ but $S \to X, S \dashv Y$ in C3 change to $S \dashv X, S \to Y$ in C4. It appears that successive repressions in the two-step cascade ($S \dashv X \dashv Y$) places C4 below C3 in extrinsic noise profiles. Just like that, successive activations in the main path ($S \to X \to Y$) bestow higher extrinsic noise in I3 than in I4. The direct effect S has on Y while changing $S \dashv Y$ in C3 (I3) to $S \to Y$ in C4 (I4) seems to be not adequately effective in increasing S mediated extrinsic noise of Y with respect to the corresponding changes of noise levels brought in due to topological variation of the respective two-step cascade pathways.

G. Why C1 and I1 should have greater extrinsic noise for the output gene?—A plausible biological advantage

Having rationalized the superiority of C1 and I1 FFLs than the other FFL types in supplying greater amount of extrinsic noise to target gene Y, we now briefly comment on the plausible connection that this noisy affair may share with the biological roles played by C1 and I1 FFL motifs. As surveyed in subsection (C) of the Introduction, C1 and I1 FFLs are indispensable for searching and consuming energy resources in *E. coli*. A hungry bacterium runs and tumbles while trying to sense appropriate chemical signal which tells it when and where food is available and in what amount. High fidelity in chemical sensing enables the microorganism to synthesize the required enzymes in right proportions to metabolize the food and helps it not to overspend its enzyme synthesizing machinery and stored energy. This chemical cue also tells it to produce its swimming apparatus, the flagella. As it turns out in uncertain environmental conditions, the whole searching process involving chemotaxis is impeded with noise. It is not at all surprising that the input TF and its coregulator both of which interact with these external signals, convey the fluctuations extrinsically to the output gene which should sense the environment precisely and accordingly produce the desired enzymes and flagella motor proteins at right time and in right amounts. Therefore, the logic behind the enhanced extrinsic noise supplied to the target gene in C1 and I1 FFLs may possibly be due to the organismal necessity of acquainting the target gene with a broader knowledge of the fluctuating environmental states.

IV. CONCLUSION

To summarize, we developed a stochastic framework within the purview of Gaussian noise processes, to search for appropriate statistical metric which can explain the abundance phenomenon of coherent and incoherent FFL motifs in model systems like *E. coli* and yeast. To this end, we have found that the magnitude of the normalized covariance of the intermediate and final gene-products, can effectively distinguish the coherent and incoherent type 1 FFLs from the other three FFL types belonging to each of the respective classes. The aforementioned covariance is representative of the extrinsic noise for the output gene and we have shown that C1 and I1 have higher amounts of this extrinsic noise than the rest of the FFL types of each of the corresponding classes. Our metric can encompass additive and multiplicative signal integration schemes in both fixed and varying input TF population conditions. Thus, our formalism adequately takes into account much of the biological complexities that an important signal processing motif like FFL may face in diverse environmental conditions riddled with noise. This extrinsic noise along with its intrinsic counterpart which is generated due to the low copy numbers of bio-molecules pose enough challenges for building efficient stochastic mathematical models. We argued that by providing higher amount of extrinsic noise to output gene, C1 and I1 can provide effective knowledge about a diverse environmental sample space to the target gene so that it can produce the essential physiological commodities suitably for the survival of the microorganism. As a cautionary remark, we should mention that usage of Hill mechanism to describe genetic regulation under quasi-steady-state assumption has an in-built disadvantage as this type of regulatory interaction is of nonelementary character for which rigorously derived reaction probabilities to implement the heuristic SSA are lacking. It has been observed that under this circumstance, the latter data may result in underestimating the size of noise [63]. Another point worth discussing is regarding another useful definition of extrinsic noise which correctly captures the effect of dynamic environment on the copy-number fluctuations via a dependency on systemic rate constants. This distinct notion of extrinsic noise has been also successfully applied to study the robustness of FFL motifs [42].

Talking of FFL robustness, we invoke the paper by Mayo *et al.* [38] where they showed rise time of output protein level is robust to faster induction of top-level TF, the degradation rate of the latter being their tuning parameter. This procedure is closely related to our signaling scenario I where k_s ($= \mu_s \langle S \rangle$) increases thereby increasing μ_s so that $\langle S \rangle = 10$ copies is maintained throughout. Defining the response time of S by the time taken by it to reach half of its steady-state value and denoting it by $\tau_{s_{1/2}}$, we obtain its relation to k_s. It turns out to be $\tau_{s_{1/2}} = \ln 2/\mu_s \approxeq 0.693\langle S \rangle / k_s = 6.93/k_s$. As k_s increases

from 1 to 10 min^{-1}, $\tau_{s_{1/2}}$ decreases approximately from 6.93 to 0.693 min whereas $|\Sigma^N(X,Y)|$ decreases for all C and I type FFLs independent of signal integration mechanisms [see Figs. 2(a)–2(d)]. Hence, we infer that for constant signaling population, variations in response time of the input signal affect the extrinsic noise of target gene-product in a similar fashion in all FFL types. In our signaling scenario II, μ_s remains fixed at 0.1 min^{-1} and consequently $\tau_{s_{1/2}}$ also remains unchanged at approximately 6.93 min throughout variations in $\langle S \rangle$. Under this circumstance, we observe distinguishability of I1-A (I1-M) with I3-A (I3-M) is compromised at $\langle S \rangle \approx 15$ copies onward, whereas profiles of coherent FFLs retain distinguishability of C1-A (C1-M) over other coherent types [see Figs. 3(a)–3(d)]. Hence, it is observed that the input response time plays a key role in the characteristics of extrinsic noise. But we should also make a note that our present construct allows equal contributions of direct and indirect pathways to produce the output gene-product whereas in Ref. [38], faster domain of top-level TF's induction is linked with the direct pathway whereas the slower one is associated with the indirect pathway.

V. FUTURE OUTLOOK

Understanding the FFL abundance problem in bacterial model systems may lead to contextualizing their functionalities in higher-level organisms. Our findings may be tested by constructing synthetic biological circuits functioning of which depends on biochemical noise [64]. *In vivo*, FFLs may involve additional network interactions like autoregulatory loops, etc., as in the cases of C1 and I1 FFLs in *E. coli* [24,26]. It would be interesting to check whether our present construct is tenable to tackle those additional architectural complexities. It has been previously shown in experiments that the signal integration mechanism is highly plastic for *E. coli* where a few mutations in the target regulatory site can alter the logic with which multiple upstream signals can affect the target gene expression [65]. Hence, further research should be devoted to test new statistical metrics, e.g., information-theoretic ones [41,46,66–69] which can justify abundance frequencies of FFLs for other logical types of target input functions beyond additive and multiplicative ones.

ACKNOWLEDGMENTS

Md S.A.M. is supported by Department of Science and Technology, Government of India, through INSPIRE research fellowship (DST/INSPIRE Fellowship/2018/IF180056). A.B. acknowledges Bose Institute, Kolkata for research support.

[1] E. M. Ozbudak, M. Thattai, I. Kurtser, A. D. Grossman, and A. van Oudenaarden, Nat. Genet. **31**, 69 (2002).
[2] W. J. Blake, M. KÆrn, C. R. Cantor, and J. J. Collins, Nature **422**, 633 (2003).
[3] J. M. Raser and E. K. O'Shea, Science **304**, 1811 (2004).
[4] J. M. Raser and E. K. O'Shea, Science **309**, 2010 (2005).
[5] E. Kussell and S. Leibler, Science **309**, 2075 (2005).
[6] A. Eldar and M. B. Elowitz, Nature **467**, 167 (2010).
[7] W. Bialek, *Biophysics: Searching for Principles* (Princeton University Press, Princeton, 2012).
[8] L. S. Tsimring, Rep. Prog. Phys. **77**, 026601 (2014).
[9] H. B. Fraser, A. E. Hirsh, G. Giaever, J. Kumm, and M. B. Eisen, PLoS Biol. **2**, e137 (2004).
[10] J. R. Newman, S. Ghaemmaghami, J. Ihmels, D. K. Breslow, M. Noble, J. L. DeRisi, and J. S. Weissman, Nature **441**, 840 (2006).
[11] A. Bar-Even, J. Paulsson, N. Maheshri, M. Carmi, E. O'Shea, Y. Pilpel, and N. Barkai, Nat. Genet. **38**, 636 (2006).
[12] M. B. Elowitz, A. J. Levine, E. D. Siggia, and P. S. Swain, Science **297**, 1183 (2002).
[13] J. M. Pedraza and A. van Oudenaarden, Science **307**, 1965 (2005).
[14] J. Paulsson, Phys. Life Rev. **2**, 157 (2005).
[15] P. S. Swain, M. B. Elowitz, and E. D. Siggia, Proc. Natl. Acad. Sci. USA **99**, 12795 (2002).
[16] J. Paulsson, Nature **427**, 415 (2004).
[17] A. Hilfinger and J. Paulsson, Proc. Natl. Acad. Sci. USA **108**, 12167 (2011).
[18] D. Schnoerr, G. Sanguinetti, and R. Grima, J. Phys. A: Math. Theor. **50**, 093001 (2017).
[19] U. Alon, *An Introduction to Systems Biology: Design Principles of Biological Circuits* (CRC Press, Boca Raton, FL, 2006).
[20] R. Milo, S. Shen-Orr, S. Itzkovitz, N. Kashtan, D. Chklovskii, and U. Alon, Science **298**, 824 (2002).
[21] S. S. Shen-Orr, R. Milo, S. Mangan, and U. Alon, Nat. Genet. **31**, 64 (2002).
[22] U. Alon, Nat. Rev. Genet. **8**, 450 (2007).
[23] S. Mangan and U. Alon, Proc. Natl. Acad. Sci. USA **100**, 11980 (2003).
[24] S. Mangan, S. Itzkovitz, A. Zaslaver, and U. Alon, J. Mol. Biol. **356**, 1073 (2006).
[25] S. Mangan, A. Zaslaver, and U. Alon, J. Mol. Biol. **334**, 197 (2003).
[26] S. Kalir, S. Mangan, and U. Alon, Mol. Syst. Biol. **1**, 2005.0006 (2005).
[27] A. M. Walczak, G. Tkačik, and W. Bialek, Phys. Rev. E **81**, 041905 (2010).
[28] W. H. de Ronde, F. Tostevin, and P. R. ten Wolde, Phys. Rev. E **86**, 021913 (2012).
[29] P. J. Ingram, M. P. Stumpf, and J. Stark, BMC Genomics **7**, 108 (2006).
[30] H. W. Ma, B. Kumar, U. Ditges, F. Gunzer, J. Buer, and A. P. Zeng, Nucleic Acids Res. **32**, 6643 (2004).
[31] T. E. Gorochowski, C. S. Grierson, and M. di Bernardo, Sci. Adv. **4**, eaap9751 (2018).
[32] O. V. Tsoy, M. A. Pyatnitskiy, M. D. Kazanov, and M. S. Gelfand, BMC Evol. Biol. **12**, 200 (2012).

[33] K. Xiong, A. K. Lancaster, M. L. Siegal, and J. Masel, Nat. Commun. **10**, 2418 (2019).
[34] L. Goentoro, O. Shoval, M. W. Kirschner, and U. Alon, Mol. Cell **36**, 894 (2009).
[35] M. Avila-Elchiver, D. Nagrath, and M. L. Yarmush, Mol. BioSyst. **8**, 511 (2012).
[36] I. Otero-Muras and J. R. Banga, PLoS ONE **11**, e0166867 (2016).
[37] R. Murugan, PLoS ONE **7**, e41027 (2012).
[38] M. Mayo, A. Abdelzaher, E. J. Perkins, and P. Ghosh, Phys. Rev. E **90**, 032706 (2014).
[39] M. Kittisopikul and G. M. Suel, Proc. Natl. Acad. Sci. USA **107**, 13300 (2010).
[40] B. Ghosh, R. Karmakar, and I. Bose, Phys. Biol. **2**, 36 (2005).
[41] M. S. A. Momin, A. Biswas, and S. K. Banik, Phys. Rev. E **101**, 022407 (2020).
[42] E. M. Keizer, B. Bastian, R. W. Smith, R. Grima, and C. Fleck, Phys. Rev. E **99**, 052417 (2019).
[43] S. Tănase-Nicola, P. B. Warren, and P. R. ten Wolde, Phys. Rev. Lett. **97**, 068102 (2006).
[44] A. Biswas and S. K. Banik, Phys. Rev. E **93**, 052422 (2016).
[45] A. Biswas and S. K. Banik, Chaos **28**, 103102 (2018).
[46] A. Biswas, Chaos **29**, 063108 (2019).
[47] L. Bintu, N. E. Buchler, H. G. Garcia, U. Gerland, T. Hwa, J. Kondev, and R. Phillips, Curr. Opin. Genet. Dev. **15**, 116 (2005).
[48] F. van Drogen, V. M. Stucke, G. Jorritsma, and M. Peter, Nat. Cell Biol. **3**, 1051 (2001).
[49] E. Dekel, S. Mangan, and U. Alon, Phys. Biol. **2**, 81 (2005).
[50] V. Shahrezaei and P. S. Swain, Proc. Natl. Acad. Sci. USA **105**, 17256 (2008).
[51] D. T. Gillespie, J. Comput. Phys. **22**, 403 (1976).
[52] D. T. Gillespie, J. Phys. Chem. **81**, 2340 (1977).
[53] D. T. Gillespie, J. Chem. Phys. **113**, 297 (2000).
[54] P. B. Warren, S. Tănase-Nicola, and P. R. ten Wolde, J. Chem. Phys. **125**, 144904 (2006).
[55] J. Keizer, *Statistical Thermodynamics of Nonequilibrium Processes* (Springer-Verlag, Berlin, 1987).
[56] J. Elf and M. Ehrenberg, Genome Res. **13**, 2475 (2003).
[57] N. G. van Kampen, *Stochastic Processes in Physics and Chemistry*, 3rd ed. (North-Holland, Amsterdam, 2007).
[58] R. Grima, Phys. Rev. E **92**, 042124 (2015).
[59] Y. Chen, J. M. L. Ho, D. L. Shis, C. Gupta, J. Long, D. S. Wagner, W. Ott, K. Josić, and M. R. Bennett, Nat. Commun. **9**, 64 (2018).
[60] G. Tkačik, A. M. Walczak, and W. Bialek, Phys. Rev. E **80**, 031920 (2009).
[61] E. Ziv, I. Nemenman, and C. H. Wiggins, PLoS ONE **2**, e1077 (2007).
[62] F. J. Bruggeman, N. Blüthgen, and H. V. Westerhoff, PLoS Comput. Biol. **5**, e1000506 (2009).
[63] P. Thomas, A. V. Straube, and R. Grima, BMC Syst. Biol. **6**, 39 (2012).
[64] M. B. Elowitz and S. Leibler, Nature **403**, 335 (2000).
[65] Y. Setty, A. E. Mayo, M. G. Surette, and U. Alon, Proc. Natl. Acad. Sci. USA **100**, 7702 (2003).
[66] L. Faes, A. Porta, G. Nollo, and M. Javorka, Entropy **19**, 5 (2017).
[67] J. Lizier, N. Bertschinger, J. Jost, and M. Wibral, Entropy **20**, 307 (2018).
[68] D. Marinazzo, L. Angelini, M. Pellicoro, and S. Stramaglia, Phys. Rev. E **99**, 040101(R) (2019).
[69] M. Li, Y. Han, M. J. Aburn, M. Breakspear, R. A. Poldrack, J. M. Shine, and J. T. Lizier, PLoS Comput. Biol. **15**, e1006957 (2019).